



US 20190202893A1

(19) **United States**

(12) **Patent Application Publication**
Falkner et al.

(10) **Pub. No.: US 2019/0202893 A1**

(43) **Pub. Date: Jul. 4, 2019**

(54) **VIRAL VECTORS ENCODING
RECOMBINANT FVIII VARIANTS WITH
INCREASED EXPRESSION FOR GENE
THERAPY OF HEMOPHILIA A**

(60) Provisional application No. 62/255,317, filed on Nov. 13, 2015.

Publication Classification

(71) Applicants: **Baxalta Incorporated**, Bannockburn, IL (US); **Baxalta GmbH**, Zug (CH)

(51) **Int. Cl.**
C07K 14/755 (2006.01)
A61K 48/00 (2006.01)
C12N 15/86 (2006.01)

(72) Inventors: **Falko-Günter Falkner**, Orth (AT);
Franziska Horling, Gaenserndorf (AT);
Johannes Lengler, Vienna (AT);
Hanspeter Rottensteiner, Vienna (AT);
Friedrich Scheifflinger, Vienna (AT)

(52) **U.S. Cl.**
CPC **C07K 14/755** (2013.01); **A61K 48/0066**
(2013.01); **A61K 48/0058** (2013.01); **C12N**
2800/22 (2013.01); **C12N 15/86** (2013.01);
C12N 2840/007 (2013.01); **C12N 2750/14143**
(2013.01); **A61K 48/0008** (2013.01)

(21) Appl. No.: **16/211,201**

(57) **ABSTRACT**

(22) Filed: **Dec. 5, 2018**

The present disclosure provides, among other aspects, codon-altered polynucleotides encoding Factor VIII variants for expression in mammalian cells. In some embodiments, the disclosure also provides mammalian gene therapy vectors and methods for treating hemophilia A.

Related U.S. Application Data

Specification includes a Sequence Listing.

(63) Continuation of application No. 15/349,930, filed on Nov. 11, 2016, now Pat. No. 10,189,888.

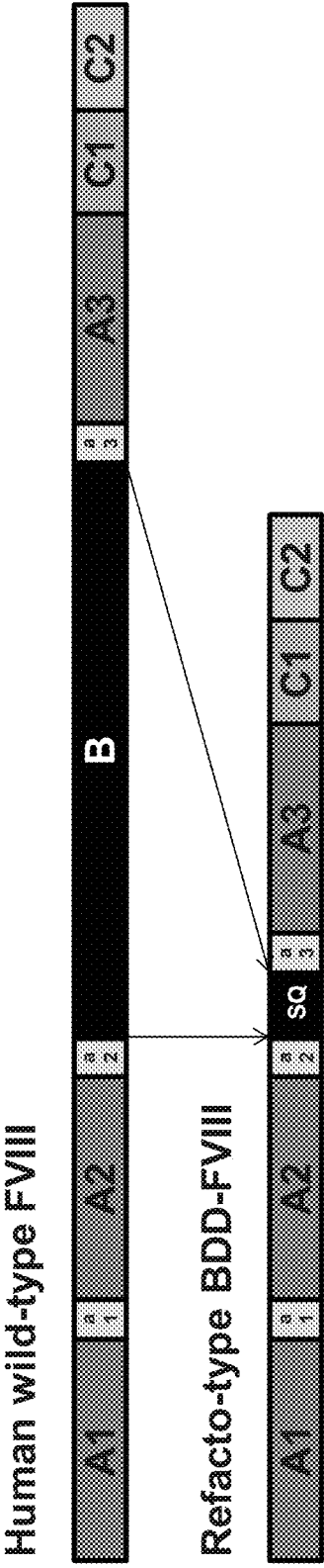


Figure 1

CS04-FL-NA

atgcagattgagotgagcacctgcttctctctctgctgctgctgaggttctgcttctctctgcccaccagga
gatactacctggggctgtggagctttctctgggactacatgcagctctgacctggggagctgctgt
ggatgccaggttcccaccagagtgccccaaatccttcccattcaacacctctgtggtctacaagaag
acctctttgtggagtcaactgaccacctgttcaacattgccaaaaccaggccacctggatgggac
tctctgggaaccaccattcaggctgaggtgatgaoactgtggtcatcacctcaagaacatggcctc
ccacctgtgagcctgcatgctgtggggctcagctactggaaggcctctgagggggctgagtatgat
gaccagacctcccagaggagaaggaggatgacaaaagtgttccctgggggagccacacctatgtgt
ggcaggtcctcaaggagaatggccccatggcctctgacctctgacctactcctaccttct
tcatgtggacctggtcaaggacctcaactctggactgattggggcctgctggtgtgagggagggc
tccctggccaaagagaagaccagacctgcacaagtctcattctctgtttgctgtctttgatgagg
gcaagagctggcactctgaaaccaagaactcctgatgcaggacagggatgctgcctctgccagggc
ctggcccaagatgcacactgtgaatggctatgtgaacaggagcctgcctggactcattggctgccac
aggaatctgtctactggcatgtgattggcatggggacaaccctgaggtgcactccatcttctctgg
agggccacaccttctgtgaggaaccacagacagggccagcctggagatcagccccatcaccttct
cactgcccagacctgctgatggacctcggacagttcctgctgttctgccacatcagctcccaccag
catgatggcatggaggcctatgtcaaggtggacagctgcctgaggagccacagctcaggatgaaga
acaatgaggaggtgaggactatgatgatgacctgactgactctgagatggatgtggctcctttga
tगतgacaacagccatccttcaatcaagatcaggtctgtggccaagaaacccccagacctgggtg
cactacattgtctgctgaggaggagactgggactatgccccactggctcctggccccctgatgacgga
gctacaagagccagtaacctcaacaatggcccacagaggattggacgcaagtacaagaaagttaggtt
catggcctacactgatgaaaacctcaagaccagggaggccattcagcatgagctctggcatcctgggc
ccactcctgtatggggaggtgggggacacctgctcatctcaagaaccaggcctccaggccct
acaacatctaccacatggcatcactgatgtcaggccccctgtacagccgagggctgccaaaggggt
gaaacacctcaaggacttccccattctgctggggagatcttcaagtacaagtggactgtcactgtg
gaggatggaccaacaaaatctgaccccagggtgctcaccagatactactccagctttgtgaacatgg
agagggacctggcctctggcctgatggccccactgctcatctgctacaaggagtctgtggaccagag
gggaaaccagatcatgtctgacaagaggaatgtgattctgttctctgtctttgatgagaacaggagc
tggtagcctgactgagaacattcagccttctgccccaccctgctgggggtgcagctggaggacctg
agttccaggccagcaacatcatgcactccatcaatggctatgtgtttgacagcctccagctttctgt
ctgctgcatgaggtggcctactggtagacttctttctattggggcccagactgacttcttctgtc
tctctctctggctacaccttcaaacacaagatgggtgatgaggacacctgacctcttcccattct
ctggggagactgtgttcatgagcatgggaaacctggcctgtggattctgggatgccacaactctga
cttccgcaacagggcatgactgcctgctcaagctctcctcctgtgacaagaacactggggactac
tatgaggacagctatgaggacatctctgctacctgctcagcaagaacaatgccattgagcccagga
gcttcagccagaatccacctgtcctgaaacgcccaccagaggagatcaccaggaccacctccagtc
tgaccaggagagatgactatgatgacaccttctctgtggagatgaagaagaggactttgacatc
tatgacgaggacgagaaccagagccccaggagcttccagaagaagaccaggcactacttctgtg
ctgtggagcctgtgggaotatggcatgagctccagccccatgtcctcaggaacagggcccagtc
tggctctgtgccacagttcaagaaagtggctctcaagaggttcaactgatggcagcttccccagccc
ctgtacagaggggagctgaatgagcacctgggactcctgggcccatacatcagggctgaggtggagg
acaacatcatggtagcttccgcaaccaggcctccagggcctacagcttctacagctcctcatcag
ctatgagGaggaccagagggcaggggctgagccacgcaagaactttgtgaaacccaatgaaaccaag
acctacttctgaaagtccagcaccacatggcccccccaaggatgagttgactgcaaggcctggg

(Continued)

Figure 2A

ootactttctctgatgtggacctggagaaggatgtgcactctggcctgattggcccactcctgggtctg
 ccacaccaacacccctgaacccctgcccctggaaggcaagtgactgtgcaggagtttgcccctctctcttc
 accatcttttgatgaaaccaagagctggtaacttcaactgagaacatggagcgcgaactgcagggccccc
 gcaacattcagatggaggacccccacttcaaagagaactaccgcttccatgccatcaatggctacat
 catggacacccctgocctgggcttgcacatggcccaggaccagaggatcaggtggtaacctgctttctatg
 ggctccaatgagaacattcaactccatccacttctctgggcatgtcttcaactgtgocgcaagaaggagg
 agtacaagatggccctgtacaacctctaccctgggggtctttgagactgtggagatgctgcccctcaa
 agctggcatctggagggtggagtgcctcattggggagcacctgcacatgctggcatgagcaccctgttc
 ctggctctacagcaacaagtgcacagacccccctgggaatggcctctggccacatcagggacttccaga
 tcaactgcctctggccagtatggccagtgggcccccaagctggccaggctccactactctggatccat
 caatgcctggagcaccaaggagccattcagctggatcaaagtggacctgctggcccccctatgatcctc
 catggcatcaagacccagggggccaggcagaagttctccagcctgtacatcagccagttcatcatca
 tgtacagcctggatggcaagaaatggcagacctacagaggcaactccactggaacactcatggtctt
 ctttggcaatgtggacagctctggcatcaagcacaacatcttcaaccccccaatcatcgccagatac
 atcaggctgcacccccaccactacagcatccgcagcaccctcaggatggagctgatgggctgtgacc
 tgaactcctgcagcatgcccctgggcatggagagcaaggccatttctgatgcccagatcactgcctc
 cagctacttcaccaacatgtttgccacctggagcccgaaggccaggctgcacctccaggggaagg
 agcaatgcctggaggccccaggtcaacaacccaaaggagtggtgcagggtggacttccagaagacca
 tgaaggctcactggggtgaccacccaggggtcaagagcctgctcaccagcatgtatgtgaaggagtt
 cctgatcagctccagccaggatggccaccagtggaacctcttcttccagaatggcaagggtcaagggtg
 ttccagggcaaccaggacagcttcaacCectgtggtgaacagcctggacccccccctcctgaccagat
 acctgaggattcaccaccagagctgggtccaccagattgcctgaggatggagggtcctgggatgtga
 ggcccaggacctgtactga (SEQ ID NO:1)

Figure 2B

CS04-FL-AA

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDAR' FPPRVPKSFPFNNTSVVYK
 KTLFVEFTDHLFENIAKPRPPWMGLLGPTTQAEVYDVTVVITLKNMASHPVSLHAVGVSYWKASEGAEY
 DDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVRDLNSGLIGALLVCRE
 GSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGC
 HRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSH
 QHDGMEAYVKVDSCEPEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSVAKKHFKTW
 VHYIAAEEDWDYAPLVLAAPDDRSYKSQYLLNNGPQRIGRKYKKVRFEMAYTDETFKTRAIQHESGIL
 GFLLYGEVGDTHLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVT
 VEDGPTKSDPRCLTRYSSFVNMERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVI LFSVFDENR
 SWYL TENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLS
 VFFSGYTFKHKMVEYEDTLTLFFPFSGETVFMSMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGD
 YYEDSYEDI SAYLLSKNNAIEPRSFQNPVLRHQRREITRTTLQSDQEEIDYDDTISVEMKKEDFD
 IYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSPHVLRNRAQSGSVQFKKVVFEFTDGSFTQ
 PLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEQDQGAEPKRNFKPNET
 KTYFWKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALF
 FTIFDETksWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQIRWYLLS
 MGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTL
 FLVYSNKQOTPLGMASGHIRDFQITASGQYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI
 IHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIRHNI FNPP IAR
 YIRLHPHTHYSIRSTLRMELMGCDLNSCSMPLGMESKAI SDAQITASSYFTNMFATWSPSKARLHLQG
 RSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLSMYVKEFLISSQDGHQWTLFFQNGKVK
 VFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY (SEQ ID NO:2)

Figure 3

CS04-HC-NA

```

                                gcc
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg
ggggagctgc ctgtgggatgc caggttccca cccagagtgc ccaaatecctt cccattcaac
acctctgtgg tctacaagaa gacctctttt gtggagttca ctgaccacct gttcaacatt
gocaaaccca ggccaccctg gatgggaactc ctgggaccca ccattcaggc tgaggtgtat
gacactgtgg tcatcacctt caagaacatg gcctcccacc ctgtgagcct gcatgctgtg
ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg
gagaaggagg atgacaaagt gttccctggg ggcagocaca cctatgtgtg gcaggtcctc
aaggagaatg gccccatggc ctctgaccca ctctgocctga cctactccta cctttctcat
gtggacctgg tcaaggacct caactctgga ctgattgggg cctgctggt gtgcaggagg
ggctccctgg ccaaagagaa gaccocagacc ctgcacaagt tcattctcct gtttgcctgc
tttgatgagg gcaagagctg gcaactctgaa accaagaact cctgatgca ggacagggat
gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagg
ctgcctggac tcattggctg ccacaggaaa totgtctact ggcattgtgat tggcatgggg
acaacccctg aggtgcactc cattttcctg gagggocaca ccttccctggt caggaaccac
agacaggcca gcctggagat cagccccatc accttctcda ctgccagac cctgctgatg
gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cgcctttgat
gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca cccaagacc
tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gccacagag gattggacgc
aagtacaaga aagtcaagtt catggcctac actgatgaaa ccttcaagac cagggaggcc
attcagcatg agtctggcat cctgggcccc ctctgtatg gggagggtgg ggacacctg
ctcatcatct tcaagaacca ggctccagg ccctacaaca totaccaca tggcatcact
gatgtcaggc cctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc
cccattctgc ctggggagat cttcaagtac aagtggactg tcaactgtga ggatggacca
accaaactctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg
gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag
aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag
aacaggagct ggtacctgac tgagaacatt cagoccttcc tgcccaacc tgctggggtg
cagctggagg acctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg
tttgacagcc tcagcttctc tgtctgocctg catgaggtgg cctactggta cattctttct
attggggccc agaotgactt cctttctgtc ttcttctctg gotacacctt caaacacaag
atgggtgatg aggacacct gacctctctc ccattctctg gggagactgt gttcatgagc
atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc
atgactgccc tgcctcaaagt ctctctctgt gacaagaaca ctggggacta ctatgaggac
agctatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gccagg
(SEQ ID NO:3)

```

Figure 4

CS04-LC-NA

```

                                g agatcaccag gaccaccctc
cagtctgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag
gactttgaca tctatgacga ggacgagaac cagagcccaa ggagcttcca gaagaagacc
aggcactact tcattgetgc tgtggagcgc ctgtgggact atggcatgag ctccagcccc
catgtcctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtgggtotc
caagagttca ctgatggcag cttcaccag cccctgtaca gaggggagct gaatgagcac
ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgacctc
cgcaaccagg cctccaggcc ctacagcttc tacagctccc tcatcagcta tgaggaggac
cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac
ttctggaaag tccagcacca catggccccc accaaggatg agtttgactg caaggcctgg
gectacttct ctgatgtgga cctggagaag gatgtgcact ctggcctgat tggcccactc
ctgggtctgcc acaccaaac cctgaaccct gcccatggaa ggoaagtgac tgtgcaggag
tttgccctct tcttcaccat ctttgatgaa accaagagct ggtacttoac tgagaacatg
gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaac
taccgcttcc atgccatcaa tggctacatc atggacaccc tgccctgggt tgtcatggcc
caggaccaga ggatcaggtg gtacctgctt tctatgggt ccaatgagaa cttcactcc
atccacttct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggccctg
tacaacctct acctggggt ctttgagact gtggagatgc tgccctcaa agctggcactc
tggagggtgg agtgccatc tggggagcac ctgcatgtg goatgagcac cctgttctctg
gtctacagca acaagtgcc gaacccccctg ggaatggcct ctggccacat cagggacttc
cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac
tctggatcca tcaatgcctg gagcaccaag gagccattca gctggatcaa agtggacctg
ctggccccca tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc
ctgtacatca gccagttcat catcatgtac agcctggatg gcaagaaatg gcagacctac
agaggcaact ccaactggaac actcatggtc ttctttggca atgtggacag ctctggcactc
aagcacaaca tcttoaacco cccaatcctc gccagataca tcaggctgca ccccaccac
tacagcatcc gcagcaccct caggatggag ctgatgggt gtgacctgaa ctctgcagc
atgcccctgg gcatggagag caaggccatt tctgatgcc agatcactgc ctccagctac
ttcaccaaca tgtttgcac ctggagccca agcaaggcca ggctgcacct ccagggagg
agcaatgcct ggaggcccc ggtcaacaac ccaaaggagt ggctgcaggt ggaactccag
aagaccatga aggtcactgg ggtgaccacc caggggtca agagcctgct caccagcatg
tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctcttcttc
cagaatggca aggtcaaggt gttccagggc aaccaggaca gcttcacccc tgtggtgaa
agcctggacc cccccctct gaccagatac ctgaggatto acccccagag ctgggtccac
cagattgccc tgaggatgga ggtcctggga tgtgaggccc aggaacctga c
(SEQ ID NO:4)

```

Figure 5

BDL001 - agc ttctctaga atcaccctgt ectgaagaga caccagaga (SEQ ID NO:5)

BDL004 - agc ttcagccaga atccacctgt cctgaaacgc caccagagg (SEQ ID NO:6)

BDL023 - agc ttcagccaga accccccctg gctgaagag caccagagg (SEQ ID NO:7)

BDLNG1 - agcttcagccagaatGTGAGCAACAATGTGAGCACAATGGCCCAATAATGCTACCACAaccacctgtcctgaaacgccaccagagg (SEQ ID NO:36)

BDLNG4 - agcttcagccagaatGTGAGCAACAATGCCACCAACAATGTGAGCAACcccacctgtcctgaaacgccaccagagg (SEQ ID NO:37)

BDLNG5 - agcttcagccagaatGTGAGCAATAATGCCCACCAACAaccacctgtcctgaaacgccaccagagg (SEQ ID NO:38)

BDLNG6 - agcttcagccagaatGTGAGCAATAATccacctgtcctgaaacgccaccagagg (SEQ ID NO:39)

BDLNG9 - agcttcagccagaatAGGAGCCCTGcccacctgtcctgaaacgccaccagagg (SEQ ID NO:40)

BDLNG10 - agcttcagccagaatGCCACTAATGIGCTAACAACCTCTGTACTCTGCTGAGGcccacctgtcctgaaacgccaccagagg (SEQ ID NO:41)

BDLNG16 - agcttcagccagaatGCCACCAACTAATGTGAACAGGAGCCcaccacctgtcctgaaacgccaccagagg (SEQ ID NO:42)

BDLNG17 - agcttcagccagaatGCCACCAACTATGTGAACAGGAGCCcctgtctgccacccctgtctgtgacctgctgtgagccAGAAAccacctgtcctgaaacgccaccagagg (SEQ ID NO:43)

BDLNG18 - agcttcagccagaatGTGAGCAACAATAATGTGAGCAATGCTGTGCTGCTcaccacctgtcctgaaacgccaccagagg (SEQ ID NO:44)

BDLNG19 - agcttcagccagaatATCAGTGTGGCCCTGTGCCCTTACAATCACTGTGGCCCTCTGCTGACcccacctgtcctgaaacgccaccagagg (SEQ ID NO:45)

BDLNG20 - agcttcagccagaatATCACTGTGACCAACAATCACTGTGACTGCCcccacctgtcctgaaacgccaccagagg (SEQ ID NO:46)

BDLNG21 - agcttcagccagaatCAGACTGTGACCAACAATCACTGTGACTGGCCcccacctgtcctgaaacgccaccagagg (SEQ ID NO:47)

BDLNGV - agcttcagccagaatGCCACTAATGIGCTAACAACAGCAACACCAGCAATGCTGTcaccacctgtcctgaaacgccaccagagg (SEQ ID NO:48)

Figure 6

CS04-AV-NA

1 togcgcggttt cggatgatgac ggtgaaaacc totgacacat gcagctcccg gagacgggtca
61 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg
121 ttggcgggtg tcgggctgg ctttaactatg cggcatcaga gcagattgta ctgagagtgc
181 accatattgg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggogcc
241 attcggcatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat
301 tacggccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta accccagggt
361 tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cctcgagatt taaatgacgt
421 tggccactcc ctctctgcgc gctcgcctgc toactgaggg cgggcgacca aaggtcggcc
481 gacgcccggg ctttgcccgg gccgcctcag tgagcagcgc agcgcgcaga gaggagtg
541 ccaactccat cactaggggt tcctgagttt aaacttcgtc gacgattoga gcttgggctg
601 caggctcgagg gcaactgggag gatgttgagt aagatggaaa actactgatg acccttgca
661 agacagagta ttaggacatg tttgaacagg gccggggcga tcagcaggta gctctagagg
721 atccccgtct gtctgcacat ttctgtagagc gagtgttcgc atactctaata ctcccaggc
781 aaggttcata tttgtgtagg ttaacttattc tccttttgtt gactaagtca ataatoagaa
841 tcagcagggt ttgagtcagc ttggcaggga tcagcagcct gggttggaag gagggggtat
901 aaaagcccct tcaccaggag aagccgtcac acagactagg cgcgcaccgc ccaccatgca
961 gattgagctg agcacctgct tcttcctgtg cctgctgagg ttctgctctt ctcccaccag
1021 gagatactac ctgggggctg ttgagcttcc ttgggactac atgcagctg acctggggga
1081 gctgcctgtg gatgccagggt tcccaccag agtgccaaa tccttcccat tcaacacctc
1141 tgtggtctac aagaagacc cttttgtgga gttcactgac cacctgttca acattgcca
1201 acccaggcca cctggatgg gactcctggg acccaccatt caggctgagg tgtatgacac
1261 tgtggtcact accctcaaga acatggcctc ccaccctgtg agcctgcatg ctgtgggggt
1321 cagctactgg aaggcctctg agggggctga gtatgatgac cagacctccc agagggagaa
1381 ggaggatgac aaagtgttcc ctgggggag ccacacctat gtgtggcagg tctcaagga
1441 gaatggcccc atggcctctg acccactctg cctgacctac tctaccttt ctcatgtgga
1501 cctggtcaag gacctaact ctggactgat tggggccctg ctggtgtgca gggagggtc
1561 cctggccaaa gagaagacc agaccctgca caagtccatt ctctgtttg ctgtctttga
1621 tgagggcaag agctggcact ctgaaaccaa gaactccctg atgcaggaca gggatgctgc
1681 ctctgccagg gctggcccca agatgcacac tgtgaatggc tatgtgaaca ggagcctgcc
1741 tggactcatt ggctgccaca ggaatctgt ctactggcat gtgattggca tggggacaac
1801 cctgagggtg cactccattt tctggaggg ccacacctc ctggtcagga accacagaca
1861 ggccagcctg gagatcagcc ccactcactt cctcactgoc cagacctgc tgatggacct
1921 cggacagttc ctgctgttct gccacatcag cteccaccag catgatggca tggaggccta
1981 tgtcaagggt gacagctgcc ctgaggagcc acagctcagg atgaagaaca atgaggaggc
2041 tgaggactat gatgatgacc tgactgactc tgagatggat gtggtccgct ttgatgatga
2101 caacagccca tcttccattc agatcaggtc tgtggccaaag aaacacccca agacctgggt
2161 gcactacatt gctgctgagg aggaggactg ggactatgcc ccaactggtcc tggcccctga
2221 tgacaggagc tacaagagcc agtacctcaa caatggccca cagaggattg gacgcaagta
2281 caagaaagtc aggttcattg cctacactga tgaaaccttc aagaccaggg agggcattca
2341 gcatgagtct ggcactctgg gccactcct gtatggggag gtgggggaca cctgctcat
2401 catcttcaag aaccaggcct ccaggcccta caacatctac ccacatggca tcaactgatgt
2461 caggcccctg tacagccgca ggtgcocaaa gggggtgaaa caactcaagg acttccccat

Figure 7A

2521 totgcoctggg gagatcttca agtacaagtg gactgtcact gtggaggatg gaccaaccaa
 2581 atctgacccc aggtgcctca ccagatacta ctccagcttt gtgaacatgg agagggacct
 2641 ggcoctctggc ctgattggcc cactgctcat ctgctacaag gactctgtgg accagagggg
 2701 aaaccagatc atgtctgaca agaggaatgt gattctgttc tctgtctttg atgagaacag
 2761 gagctggtac ctgactgaga acattcagcg ctccctgccc aacctgctg ggggtgcagct
 2821 ggaggacctt gagttccagg ccagcaacat catgcactcc atcaatggct atgtgtttga
 2881 cagcctccag ctttctgtct gcctgcatga ggtggcctac tggtaacatt tttctattgg
 2941 ggcccagact gacttccttt ctgtctcttt ctctggctac acctcaaac acaagatggg
 3001 gtatgaggac acctgacct tcttcccatt ctctggggag actgtgttca tgagcatgga
 3061 gaacctggc ctgtggatc tgggatgcca caactctgac tccgcaaca ggggcatgac
 3121 tgccctgctc aaagtctcct cctgtgacaa gaacctggg gactactatg aggacagcta
 3181 tgaggacatc tctgcctacc tctcagcaa gaacaatgcc attgagccca ggagcttcag
 3241 ccagaatcca cctgtcctga aacgccacca gaggagatc accaggacca cctccagtc
 3301 tgaccaggag gagattgact atgatgacac catttctgtg gagatgaaga aagaggactt
 3361 tgacatctat gacgaggacg agaaccagag cccaaggagc tccagaaga agaccaggca
 3421 ctacttcatt gctgctgtgg agcgcctgtg ggactatggc atgagctcca gccccatgt
 3481 cctcaggaac agggcccagt ctggctctgt gccacagttc aagaaagtgg tottccaaga
 3541 gttcaactgat ggcagcttca cccagcccct gtacagaggg gagctgaatg agcacctggg
 3601 actcctgggc ccatacatca gggctgaggt ggaggacaac atcatggtga ccttccgcaa
 3661 ccagggcctcc agggccctaca gcttctacag ctccctcctc agctatgagg aggaccagag
 3721 gcagggggct gagccacgca agaactttgt gaaacccaat gaaaccaaga cctacttctg
 3781 gaaagtccag caccacatgg cccccacaa ggatgagttt gactgcaagg cctgggccta
 3841 cttctctgat gtggacctgg agaaggatgt gcactctggc ctgatgggc cactcctggt
 3901 ctgccacacc aacaccctga acctgccc a tggaggcaa gtgactgtgc aggagtttgc
 3961 cctctctctc accatctttg atgaaaccaa gagctggtac ttcactgaga acatggagcg
 4021 caactgcagg gcccctatgca acatccagat ggaggacccc acctcaaaag agaactaccg
 4081 cttccatgcc atcaatggct acatcatgga cacctgcoct gggcttgtca tggcccagga
 4141 ccagaggatc aggtggtacc tgotttctat gggctccaat gagaacattc actccatcca
 4201 cttctctggg catgtcttca ctgtgcgcaa gaaggaggag tacaagatgg cootgtacaa
 4261 cctctaccct ggggtctttg agactgtgga gatgctgccc tccaaagctg goatctggag
 4321 ggtggagtgc ctcatgggg agcaccctgca tgcctggcat agcaccctgt tootggtcta
 4381 cagcaacaag tgccagacc cctgggaat ggctctgtgc cacatcaggg acttccagat
 4441 cactgcoctc ggccagtat gccagtgggc cccaagctg gccaggctcc actactctgg
 4501 atccatcaat gcctggagca ccaaggagcc attcagctgg atcaaaagtgg acctgctggc
 4561 ccccatgata atccatggca tcaagaccca gggggccagg cagaagttct ccagcctgta
 4621 catcagccag ttcacatca tgtacagcct ggatggcaag aaatggcaga cctacagag
 4681 caactccact ggaacactca tggctctctt tggcaatgtg gacagctctg gcatcaagca
 4741 caacatcttc aaccccccaa tcatcgccag atacatcagg ctgcaaccca cccactacag
 4801 catccgcagc acctcagga tggagctgat gggctgtgac ctgaaactcct gcagcatgcc
 4861 cctgggcatg gagagcaagg ccatttctga tgcccagatc actgcoctca gctacttcc
 4921 caacatgttt gccacctgga gcccagcaa ggccaggctg cactccagg gaaggagcaa
 4981 tgctggagg ccccaggtca acaacccaaa ggagtggctg caggtggact tccagaagac

Figure 7B

5041 catgaaggct actgggggtga ccaccagggt ggtoaagagc ctgctcacca goatgtatgt
5101 gaaggagttc ctgatoagct ccagccagga tggccaccag tggaccctct tttccagaa
5161 tggcaaggct aagggtgtcc agggcaacca ggacagcttc acccctgtgg tgaacagcct
5221 ggaccccccc ctctgacca gatacctgag gattcacccc cagagctggg tccaccagat
5281 tgccotgagg atggaggtcc tgggatgtga ggcccaggac ctgtaactgat gacgaguggc
5341 cgcctottagt agcagtatcg ataataaaag atctttatct tcattagatc tgtgtgttgg
5401 ttttttgtgt gttaattaag ctgcggaagg aacccttagt gatggagttg gccactccct
5461 ctctgcgcgc tcgctcgcct actgaggccg ggcgacaaa ggtcgcccga cgcocgggct
5521 ttgcccgggc ggccctcagt agcgagcgag cgcgcagaga gggagtggcc aagacgattt
5581 aatgacaag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc
5641 tcacaattcc acacaacata cgagccggaa gataaaagtg taaagcctgg ggtgcctaata
5701 gagtgagcta actcacatta attgctgttc gctcaactgcc cgctttccag tggggaaacc
5761 tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggg ttgogtattg
5821 ggcgctcttc cgcttcctcg ctcaactgact cgctgcgcctc ggtcgttcgg ctgcggcgag
5881 cggatcagc tcactcaaaag ggggtaatac ggttatccac agaactcagg gataacgcag
5941 gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc
6001 tggcgttttt ccataggctc cgcocccctg acyagcatca caaaaatcga cgtcaagtc
6061 agaggtggcg aaacccgaca ggactataaa gataccaggc gtttccccc ggaagctccc
6121 tcgtgcgcct cctgttccg accctgcgc ttaaccggata cctgtccgc tttctccctt
6181 cgggaagcgt ggcgctttct catagctcac gctgtaggta tctcagttcg ggttaggtcg
6241 ttcgctccaa gctgggtgt gtgcacgaac ccccgcttca gcccgaccgc tgcgccttat
6301 ccgtaacta tcgtcttgag tccaaccgg taagacacga cttatcgcca ctggoagcag
6361 ccaactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt
6421 ggtggcctaa ctacggctac actagaagaa cagtatttgg tatctgcgct ctgctgaagc
6481 cagttacott cggaaaaaga gttggtagct cttgatccgg caaacaacc accgctggta
6541 ggggtggttt tttgtttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag
6601 atcccttgat ctttctacg gggcttgac ctcactggaa cgaaaactca cgttaaggga
6661 ttttggctat gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa
6721 gtttttaaat aatctaaagt atatatgagt aaacttggct tgacagttac caatgcttaa
6781 tcagtgagge acctatctca gcatctgtc tatttctgtc atccatagtt gctgactcc
6841 ccgctcgtga gataactacg atacgggagg gottaccatc tggccccagt gctgcaatga
6901 tacgcgcgaga ccaacgctca ccggctccag atttatcagc aataaaccag ccagocggaa
6961 gggccgagcg cagaagtggc cctgcaactt tatccgcctc catccagctc attaattgtt
7021 gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgocattg
7081 ctacaggcat cgtggtgtca cgtcgtcgt ttggtatggc ttcattcagc tccggttccc
7141 aacgatcaag gcgagttaca tgatccccc tgttgtgcaa aaaagcgggt agctccttcg
7201 gtoctcogat cgttctcaga agtaagttgg ccgagtggt atcaactcag gttatggcag
7261 cactgcataa ttctcttact gtcattgcat ccgtaagatg cttttctgtg actggtgagt
7321 actcaaccaa gtcattctga gaatagtgtg tgcggcgacc gagttgctct tgcccggcgt
7381 caatacggga taataccgcg ccacatagca gaactttaaa agtgctcctc attggaaaac
7441 gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt togatgtaac
7501 ccaactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt totgggtgag
7561 caaaaacagg aaggcaaaaat gccgcaaaaa agggaataag ggcgacacgg aaatggtgaa
7621 tactcactact ctctcttttt caatattatt gaagcattta tcagggttat tgtctcatga
7681 gccgatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc
7741 ccgaaaaagt gccacctgac gtetaagaaa ccattattat catgacatta acctataaaa
7801 ataggcgtat cacgagccc tttcgtc (SEQ ID NO:8)

Figure 7C

CS01m1-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGGA
GATACTACCTGGGGGCTGTGGAACCTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCTGT
GGATGCCAGGTTCCCACCCAGAGTGCCTCAAGTCTTCCCATTCAACACCTCTGTGGTCTACAAGAAG
ACACTCTTTGTGGAATTCACCTGACCACCTGTTCAACATTCGAAAACCCAGACCACCTGGATGGGAC
TCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCTCAAGAACATGGCATC
CCACCTGTGTCTCTGCATGCTGTGGGAGTCTCATACCTGGAAAGCCTCTGAAGGGCTGAGTATGAT
GACCAGACATCCCAGAGAGAGAAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACCTATGTGT
GGCAAGTCCCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACCTTACCTTTC
TCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCTGCTGGTGTGCAGGGAAAGGA
TCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGGTGTCTTTGATGAGG
GCAAGTCTTGGCACTCTGAAAACAAAAGAACTCCCTGATGCAAGACAGGGATGCTGCCCTTGCCAGGGC
ATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTCATTGGCTGCCAC
AGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCCTCCATTTTCCFGG
AGGGACACACCTTCCCTGGTCAGGAACCCACAGACAAGCCTCTCTGGAGATCTCTCCCATCACCTTCT
CACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTCTGCCACATCTCTTCCCACCAG
CATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCCACAGCTCAGGATGAAGA
ACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGGTCCAGATTTGA
TGATGACAACCTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAACACCCCAAGACATGGGTG
CACTACATTTGCTGCTGAGGAAGAGGACTGGGACTATGCACCCTGGTCCCTGGCCCCGATGACAGGA
GCTACAAGTCTCAGTACCTCAACAATGGCCCCACAAAGAATGGAAGAAAGTACAAGAAAGTCAGATT
CATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAGCATGAGTCTGGCATTCTGGGA
CCACTCCTGTATGGGGAAAGTGGGAGACACCTTCTCATCATCTTCAAGAACCAGGCTCCAGGCCCT
ACAACATCTACCCACATGGCATCACTGATGTGAGGCCCCGTACAGCAGGAGACTGCCAAAAGGGGT
GAAACACCTCAAGGACTTCCCCATTTGACCTGGAGATCTTCAAGTACAAGTGGACTGTCACTGTG
GAGGATGGACCAACAAGTCTGACCCAGGTGCCTCACCAGATACTACTCCTCTTTTGTGAACATGG
AGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCACTCTGCTACAAGGAGTCTGTGGACCAGAG
AGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTGTTCTCTGTCTTTGATGAGAACAGATCA
TGGTACCTGACTGAGAACAATTCAGAGATTCCTGCCCAACCCTGCTGGGGTGCACCTGGAAGACCCCTG
AGTTCCAGGCAAGCAACATCATGCCTCCATCAATGGCTATGTGTTTGGACTCTCTCCAGCTTTCTGT
CTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTTCTATTGGGGCACAACTGACTTCCCTTTCTGT
TTCTTCTCTGGATACACCTTCAAGCACAAAGATGGTGTATGAGGACACCCCTGACACTCTTCCCATTCT
CTGGGGAAACTGTGTTTTCATGAGCATGGAGAACCCTGGACTGTGGATTCCTGGGATGCCACAACCTG
CTTCAGAAAACAGGGGAATGACTGCCTGCTCAAAGTCTCCTCCTGTGACAAGAACTGGGGACTAC
TATGAGGACTCTTATGAGGACATCTCTGCCTACCTGCTCAGCAAGAACAATGCCATTGAGCCCAGAA
GCTTCTCTCAGAATCCACCTGTCTGAAAGAGACACCAGAGAGAGATCACCAGGACAACCCCTCCAGTC
TGACCAGGAAGAGATTGACTATGATGACACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATC
TATGATGAGGACGAGAACCAGTCTCCAAGATCAATCCAGAAGAAGACAAGACACTACTTCATTGCTG
CTGTGGAAAGACTGTGGGACTATGGCATGTCTTCTCTTCCCATGTCTCAGGAACAGGGCACAGTC
TGGCTCTGTGCCACAGTTCAGAAAGTGGTCTTCCAGGAGTTCACTGATGGCTCAATTCACCCAGCCC
CTGTACAGAGGGGAACCTGAATGAGCACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAG
ACAACATCATGGTGCATTCAGAAAACAGGGCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAG
CTATGAGGAAGACCAGAGACAAGGGGCTGAGCCAAGAAAGAACTTTGTGAAAACCCAAATGAAAACCAAG
ACCTACTTCTGGAAAGTCCAGCACACATGGCACCCACCAAGGATGAGTTTACTGCAAGGCCCTGGG

(Continued)

Figure 8A

CATACTTCTCTGATGTGGACCTGGAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTG
CCACACCAACACCCCTGAACCCCTGCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGGCCCTCTTCTTC
ACCATCTTTGATGAAACCAAGTCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCAT
GCAACATTCAGATGGAAGACCCCACTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACAT
CATGGACACCCCTGCCTGGGCTTGTTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATG
GGATCCAATGAGAACATTCACCTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGG
AATACAAGATGGCCCTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGGCCCTCCAA
AGCTGGCATCTGGAGGGTGGAAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCCTGTTT
CTGGTCTACAGCAACAAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGA
TCACTGCCTCTGGCCAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCAT
CAATGCATGGTCAACCAAGGAGCCATTCTCTTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTT
CATGGCATCAAGACACAGGGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCA
TGTACTCTCTGGATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCCTCATGGTCTT
CTTTGGCAATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCTCCCATCATTGCCAGATAC
ATCAGGCTGCACCCCACTACTCAATCAGATCAACCCTCAGGATGGAACGATGGGGATGTGACC
TGAATCCTGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCAGATCCTGCATC
CTCTTACTTCACCAACATGTTTGGCACCTGGTCAACATCAAAAGCCAGGCTGCACCTCCAGGGGAAGA
AGCAATGCCTGGAGACCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAA
TGAAAGTCACTGGGGTGACAACCCAGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTT
CCTGATCTCTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTG
TCCAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACCTCACTGGACCCCCCTCCTGACAAGAT
ACCTGAGAAATTCACCCCACTCTTGGGTCCACCAGATTTGCCCTGAGAATGGAAGTCTTGGGATGTGA
GGCACAAGACCTGTACTGA (SEQ ID NO:49)

Figure 8B

CS04A(760-1667) - CS04-SCI-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCTGCTGCTGAGGTTCTGCTTCTTCTGCCACCAGGAGATAC
TACCTGGGGGCTGTFGGAGCTTTCTTGGGACTACATGCACTCTGAOCTGGGGGAGCTGCCTGTGGATGCCAGG
TTCCCACCCAGAGTGCCTAAATCCTTCCCATTCAACACCTCTGTTGGTCTACAAGAAGACCCTCTTTGTGGAG
TTCACTGACCACCTGTTCAACATTTGCCAAABCCAGGCCACCCTGGATGGGACTCTGGGACCCACCATTAG
GCTGAGGTGATGACACTGTGGTTCATCACCTCAAGAACATGGCCTCCACCCCTGTGAGCCTGCATGCTGTG
GGGGTACGCTACTGGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGACCTCCAGAGGGAGAAGGAGGAT
GACAAAGTGTTCCTTGGGGGCAGCCACACCTATGTTGGCAGGTCTCAAGGAGAAATGGCCCCATGGCCTCT
GACCCACTCTGCCTGACCTACTCTACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATT
GGGGCCCTGCTGGTGTGACAGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCCTGCACAAGTTCAATTCTC
CTGTTTGTCTGTCTTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGAT
GTCCTCTGCCAGGGCCTGGCCCAAGATGCACACTGTGATGGCTATGTGAACAGGAGCCTGCCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGACAAACCCCTGAGGCTGACTCCATT
TTCTTGGAGGGCCACACCTTCTGTGTCAGGAACCCAGACAGGCCAGCCTGGAGATCAGCCCCATCACTTCTC
CTCACITGCCAGACCCTGCTGATGGACCTCGGACAGTTCTGCTGTTCTGCCACATCAGCTCCACCAGCAT
GATGGCATGGAGGCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCACAGCTCAGGATGAAGAACAATGAG
GAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGGTCCGCTTTGATGATGACAACAGC
CCATCTTTCATTGATCAGGCTGTGGCCAAAGAAACACCCCAAGACCTGGGTGCACTACATTGCTGCTGAG
GAGGAGGACTGGGACTATGCCCCACTGGTCTTGGCCCTGATGACAGGAGCTACAAGAGCCAGTACTCAAC
AATGGCCACAGAGGATTTGGACGCAAGTACAAGAAAGTCAAGTTTCATGGCCTACACTGATGAAACCTTCAAG
ACCAGGGAGGCCAFTCAGCATGAGTCTGGCATCTTGGGCCACTCCTGTATGGGGAGGTGGGGGACACCCTG
CTCATCATCTTCAAGAACCAGGCCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAAGGCC
CTGTACACCCGCGAGGCTGCCAAAGGGGTGAAACACCTCAAGGACTTCCCCATTTCTGCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAAATCTGACCCAGGTGCTCACCAGATACTAC
TCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTTGGCCACTGCTCATCTGCTACAAGGAG
TCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTGTTCTCTGTCTTTGATGAG
AACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCTGCCCCAACCTGCTGGGGTGCAGCTGGAGGAC
CCTGAGTTCCAGGCCAGCAACATTCATGCACTCCATCAATGGCTATGTGTTTACAGCCTCCAGCTTCTGTGTC
TGCTTGCATGAGGTGGCTACTGGTACATTTCTTCTATTGGGGCCAGACTGACTTCCCTTTCTGCTTCTTCTC
TCTGGCTACACCTTCAAACACAAGATGGTGTATGAGGACACCCCTGACCCTCTTCCCATTTCTCTGGGGAGACT
GTGTTTATGAGCATGGAGAACCCTGGCCTGTGGATTTCTGGGATGCCACAACCTTGACTTCCGCAACAGGGGC
ATGACTGCCCTGCTCAAAGTCTCTCTCTGTGACAAGAACACTGGGACTACTATGAGGACAGCTATGAGGAC
ATCTCTGCCTACCTGCTCAGCAAGAACAATGCCATTGAGCCAGGGAGATCACCAGGACACCCCTCCAGTCT
GACCAGGAGGAGATTGACTATGATGACACCATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGAC
GAGGACGAGAACCAGAGCCCAAGGAGCTTCCAGAAGAAGACCAGGCACTACTTCAATTGCTGCTGTGGAGCGC
CTGTGGGACTATGGCATGAGCTCCAGCCCCATGTCTCAGGAACAGGGCCAGTCTGGCTCTGTGCCACAG
TTCAAAGAAAGTGGTCTTCCAAGAGTTCACTGATGGCAGCTTCAACCAGCCCTGTACAGAGGGGAGCTGAAT
GAGCACCTGGGACTCTTGGGCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTACCTTCCGCAAC
CAGGCTTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAG
CCACGCAAGAACCTTTGTGAAACCCAAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCACCATGCCCCC

(Continued)

Figure 9A

ACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAGGATGTGCACTCT
GGCCTGATTTGGCCCCTCCTGGTCTGCCACACCAACACCCTGAACCCTGCCCATGGAAGGCAAGTACTGTG
CAGGAGTTTGGCCCTCTTCTCACCATCTTTGATGAAACCAAGAGCTGGTACTTCACAGAGAACATGGAGCGC
AACTGCAGGGCCCATGCAACATTCAGATGGAGGACCCACCTTCAAAGAGAACFACCGCTTCCATGCCATC
AATGGCTACATCATGGACACCCTGCCTGGGCTTGTTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTT
TCTATGGGCTCCAATGAGAACATTCACCTCCATCCACTTCTCTGGGCATGTCTTCACAGTGGCCAAGAAGGAG
GAGTACAAGATGGCCCTGTACAACCTCTACCCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCT
GGCATCTGGAGGGTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCCTGTTCTTGGTCTAC
AGCAACAAGTGGCAGACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCCTGGC
CAGTATGGCCAGTGGGCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCCTGGAGCACCAAG
GAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCATGATCATCCATGGCATCAAGACCCAGGGGGCC
AGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGATGGCAAGAAATGGCAG
ACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGCAATGTGGACAGCTCTGGCATCAAGCAC
AACATCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTGCACCCACCCACTACAGCATCCGCAGCACC
CTCAGGATGGAGCTGATGGGCTGTGACCTGAACTCCTGCAGCATGCCCTGGGCATGGAGAGCAAGGCCATT
TCTGATGCCAGATCACTGCCCTCCAGCTACTTCACCAACATGTTTGGCACCTGGAGCCCAAGCAAGGCCAGG
CTGCACCTCCAGGGAAGGAGCAATGCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGSTGGAC
TTCCAGAAGACCATGAAGTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTG
AAGGAGTTCCTGATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCCTCTTCTTCCAGAATGGCAAGGTCAAG
GTGTTCCAGGGCAACCAGGACAGCTTACCCTGTGGTGAACAGCCTGGACCCCCCTCTGACCAGATAC
CTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCTGGGATGTGAGGCCCCAG
GACCTGTACTGA (SEQ ID NO:9)

Figure 9B

CS04A(760-1667) - CS04-SC1-AA

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPFRVPKSFPFNTSVVYKK
TLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVS YWKASEGAEYD
DQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREG
SLAKEKTQTLHKFILLFAVFDGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCH
RKSVMYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLMLDLGQFLLFCHISSHQ
HDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSVAKKHPTWV
HYIAAEEDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILG
PLLYGEVGDTELLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTY
EDGPTKSDPRCLTRYSSFVNMERDLASGLIGPLLIICYKESVDQRGNQIMSDKRNVI LFSVFDENRS
WYLTENIQRF LPNPAGVQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLSV
FFSGYTFKHKMVEYEDTLTLFPFSGETVFMSEMPGLWLILGCHNSDFRNRGMFALLKVSSCDKNTG DY
YEDSYEDISAYLLSKNNAIEPREITRITLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQK
KTRHYFIAAVERLWDYGMSSSPHVLNRNAQSGSVQFKKVVVFQEF TDG SFTQPLYRGELNEHLGLLG
PYIRAEVEDNIMVTFRNQASRPYSFYSSLI SYEEDQRQGAEPKRFVKNPNETKTYFWKVQHMAPTK
DEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTEN
MERNCRAPCNTQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGH
VFTVRKKEEYKMALYNLYPGVFETVEMLP SKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMA
SGHIRDFQITASGQYQWAPK LARLHYSGSINAWSTKEPFSWIKVDLLAPMI IHGIKTQGARQKFSS
LYISQFIIMYSLDGKKWQTYRCNSTGTLMVFFGNVDSSG IKHNI FNPP IARYIRLHPHYSIRSTL
RMELMGCDLNSCSMPLMESKAISDAQITASSYFTNM FATWSPSKARLHLQGRSNAWRPQVNNPK EW
LQVDFQKTMKVTGVTTQGVKSLTSMYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNS
LDPPLLTRYLR IHPQSWVHQIALRMEVLGCEA QDLY (SEQ ID NO:10)

Figure 10

CS04A(772-1667) - CS04-SC2-NA

ATGCAGATTTGAGCTGAGCACCTGCTTCTTCCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCCTCTTTGTGGAGTTCACCTGACCACCTGTTCACATTTGCCAAACCCAGGCCACCTGGATG
GGACTCCTGGGACCCACCATTTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCCTCAAGAACATG
GCCFCCACCCCTGTGAGCCATGCATGCTGTGGGGGTGAGTACTGGAAGGCCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAAGTGTTCCTGGGGGACGCCACACC
TATGTGTGGCAGGTTCTCAAGGAGAATGGCCCCATGGCCCTCTGACCCACTCTGCCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCCTGCACAAGTTCATTCTCCTGTTTGCTGTG
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCCTGCCTGGACTC
ATTTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCTGGAGGGCCACACCTTCCCTGGTCAGGAACCCAGACAGGCCAGCCCTGGAGATCAGC
CCCATCACCTTCTCACTGCCAGACCCCTGCTGATGGACCTCGGACAGTTTCTGCTGTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCACAGAGGATTTGGAACG
AAGTACAAGAAAGTTCAGGTTTCATGGCCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCACTCCTGTATGGGGAGGTGGGGGACACCCCTGCTCATCATCTTC
AGAACCAGGCCCTCCAGGCCCTACAACATCTACCCACATGGGCATCACTGATGTCAGGCCCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCAGGTGCCTCACCCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCTGCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCCTGAGTCCAGGCCAGCAACATCATGCACCTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTTCT
ATTTGGGUCACAGACTGACTTCCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCCTGACCCCTCTTCCCATTTCTCTGGGGAGACTGTGTTTCATGAGCATGGAGAACCC
GGCTGTGGATTTCTGGGATGCCACAACCTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTTAGCCUCCAGGAGCTTCAGCCAGAATTTCCAGACACCCAGCACC
AGGGAGATCACCAGGACCACCCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGACACCATTTCT
GTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCAAGGAGCTTC

(Continued)

Figure 11A

CAGAAGAAGACCAGGCAC TACTTCA TTTGCTGCTG TGGAGCGCC TGTGGGACTATGGCATGAGCTCC
AGCCCCCATGTCC TCAGGAACAGGGCCAGTCTGGCTCTGTGCCACAGTTCAAGAAAAGTGGTCTTC
CAAGAGTTACATGATGGCAGC TTCACCCAGCCCC TGTACAGAGGGGAGCTGAAATGAGCACCTGGGA
CTCCTGGGCCCATACATCAGGGCTGAGGTTGGAGGACAACATCATGGTGACCTTCCGCAACCAGGCC
TCCAGGCCCTACAGC TTTCTACAGC TCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAG
CCACGCAAGAACTTTGTGAAACCCAAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCACCCACATG
GCCCCACCAAGGATGAGT TTTGACTGCAAGGCCCTGGGCC TACTTCTCTGATGTGGACCTGGAGAAG
GATGTGCACCTCTGGCCTGAT TTTGGCCAC TCC TGGTCTGCCACACCAAC ACCCTGAACCTGCCAT
GGAAGGCAAGTGACTGTGCAGGAGTTTGCCTCTTCTTACCATCTTTGATGAAACCAAGAGCTGG
TACTTCACTGAGAACATGGAGCGCAAC TGCAGGGCCCCATGCAACATTCAGATGGAGSACCCCACC
TTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGGCTTGT
ATGGCCCAGGACCAGAGGATCAGGTGGTACC TGTCTTCTATGGGCTCCAATGAGAACATTCACTCC
ATCCACTTCTCTGGGCATGCTTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCCCTGTACAAC
CTCTACCC TGGGGTCT TTTGAGACTG TGGAGATGCTGCCCTCAAAGCTGGCA TCTGGAGGGTGGAG
TGCCTCA TTTGGGGAGCACCTGCATGCTGGCATGAGCACCCCTGTTCTGGTCTACAGCAACAAGTGC
CAGACCCCCCTGGGAATGGCC TCTGGCCACATCAGGGACTTCCAGATCACTGCCCTCTGGCCAGTAT
GGCCAGTGGGCCCCCAAGCTGGCCAGSCTCCACTACTCTGGATCCATCAATGCCTGGAGCAACCAAG
GAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAGACCCAG
GGGGCCAGGCAGAAAG TTTCTCCAGCCTGTACATCAGCCAGTTCAATCATCATGTACAGCCTGGATGGC
AAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGCAATGTGGAC
AGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAAFCATCGCCAGATACATCAGGCCTGCACCCC
ACCCACTACAGCATCCGCAGCACCCCTCAGGATGGAGCTGATGGGCTGTGACCTGAACTCCTGCAGC
ATGCCCTTGGGCATGGAGAGCAAGGCCATTTCTGATGCCAGATCACTGCC TCCAGCTACTTTACC
AACATGTTTGGCCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGCAATGCCTGG
AGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTCACT
GGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTCTTGATCAGC
TCCAGCCAGGATGGCCACCAGTGGACCCCTCTTCTTCCAGAATGGCAAGGTCAAGGFTTCCAGGGC
AACCAGGACAGCTTACCCCTSTGGTGAACAGCCTGGACCCCCCTCCTGACCAGATACCTGAGG
ATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCTCTGGGATGTGAGGCCAG
GACCTGTACTGA (SEQ ID NO:11)

Figure 11B

CS04A{772-1667} - CS04-SC2-AA

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPPNTSVVYKKTLEFVEF
TDHLFNIAKPRPPWMGLLGPTIQAEVYDVTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDK
VFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVLDLVKDLNSGLIGALLVCREGSLAKEKTQTLRKFILLFA
VFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTPEVHSIFLEG
HTFLVRNHRQASLEISPIITFLTAQTLLMDLGGFLFCHISSHQHDGMEAYVKVDSCEPEPQLRMKNNEEAEDY
DDDLDSEMDVVRPDDDNSPFSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLAPDDRSYKSQYLNNGPQRI
GRKYKVRFMAYTDETFKTREAIQHESGILGPLYGEVGDTLIIIFKNQASRPYNIYPHGITDVRPLYSRRLP
KGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYSSSFVNMERDLASGLIGPLLYCYKESVDQRGNQ
IMSDKRNVLFSVFDENRSWYLTEINQRFPLNPAGVQLEDPEFQASNIMHSINGYVFDLQLSVCLHEVAYWY
ILSIGAQTDFLSVFFSCYTFKHKMVYEDTLLPFPFSGETVFMSEMPGLWILGCHNSDFRNRGMTALLKVSSC
DKNTGDYEDSYEDISAYLLSKNNAIEPRFSQNSRHPSTREITRRTTLQSDQEIIDYDDTISVEMKKEDFDIY
DEBENQSFERSFQKKTRHYFIAAVERLWDYGMSSSPHVLNRRAQSGSVFQFKKVVFEFTDGSFTQPLYRGELN
EHLGLLGFYIRAEVEDNIMVTFRNQASRPYSFYSSLSIYEEDQRQGAEPKRFVKNFNETKTYFWKVQHMMAPT
KDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERN
RAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYK
MALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQFPLGMASGHIRDFQITASGQYGG
WAPKLARLHYSGSINAWSTKEFFSWIKVDLLAPMI IHGIKTQGARQKFS SLYISQFIIMYSLDGKKWQTYRGN
STGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQIT
ASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISS
SQDGHQWTLFFQNGKVVVFQGNQDSFTPVVNSLDPPLTRYLRIPHQSWVHQIALRMEVLGCBAQDLY
(SEQ ID NO:12)

Figure 12

NG1: V S N N V S N N A T N A T N (SEQ ID NO:51)
GTG AGC AAC AAT GTG AGC AAC AAT GCC ACC AAT AAT GCT ACC AAC (SEQ ID NO:50)

NG4: V S N N A T N N V S N (SEQ ID NO:53)
GTG AGC AAC AAT GCC ACC AAC AAT GTG AGC AAC (SEQ ID NO:52)

NG5: V S N N A T N (SEQ ID NO:55)
GTG AGC AAT AAT GCC ACC AAC (SEQ ID NO:54)

NG6: V S N N (SEQ ID NO:57)
GTG AGC AAT AAT (SEQ ID NO:56)

NG9: R S L (SEQ ID NO:59)
AGG AGC CTG (SEQ ID NO:58)

NG10: A T N V S N N S A T S A D S A V S (SEQ ID NO:61)
GCC ACT AAT GTG TCT AAC AAC TCT GCT ACC TCT GCT GAC TCT GCT GTG AGC (SEQ ID NO:60)

NG16: A T N Y V N R S L (SEQ ID NO:63)
GCC ACC AAC TAT GTG AAC AGG AGC CTG (SEQ ID NO:62)

Figure 13A

NG17: A T N Y V N R S L S A T S A D S A V S Q N (SEQ ID NO:65)
GCC ACC AAC TAT GTG AAC AGG AGC CTG TCT GCC ACC TCT GCT GAC TCT GCT GTG AGC CAG AAT (SEQ ID NO:64)

NG18: V S N N V S N A V S A V S A (SEQ ID NO:67)
GTG AGC AAC AAT GTG AGC AAT GCT GTG TCT GCT GTG TCT GCT (SEQ ID NO:66)

NG19: I T V A S A T S N I T V A S A D (SEQ ID NO:69)
ATC ACT GTG GCC TCT GCC ACC TCT AAC ATC ACT GTG GCC TCT GCT GAC (SEQ ID NO:68)

NG20: I T V T N I T V T A (SEQ ID NO:71)
ATC ACT GTG ACC AAC ATC ACT GTG ACT GCC (SEQ ID NO:70)

NG21: Q T V T N I T V T A (SEQ ID NO:73)
CAG ACT GTG ACC AAC ATC ACT GTG ACT GCC (SEQ ID NO:72)

NGV: A T N V S N N S N T S N D S N V S (SEQ ID NO:75)
GCC ACT AAT GTG TCT AAC AAC AGC AAC ACC AGC AAT GAC AGC AAT GTG TCT (SEQ ID NO:74)

Figure 13B



Figure 14

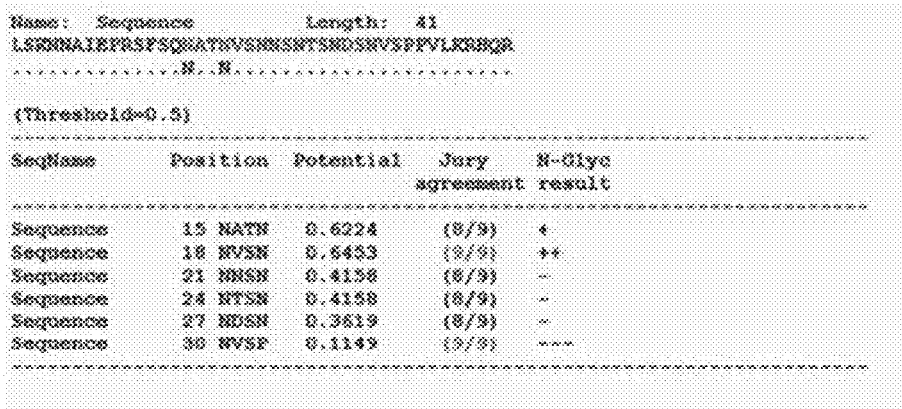


Figure 15

CS01-FL-NA

ATGCAGATTGAGCTGTCCAACCTGCTTCTTTCTGTGCTGCTGAGATTTCTGCTTCTGCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACCTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCUACCCAGAGTGCCCAAGTCTTCCUATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCAGTACCACCTGTTCAACATTTGCAAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCAATTCAGGCTGAGGTGTATGACACTGTGGTTCATCACCCCTCAAGAACATG
GCATCCCACCCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAAAGCCTCTGAAGGGGCTGAG
TATGATGACCAGACATUCCAGAGAGAGAAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACC
TATGTGTGGCAAGTCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAAACCCAGACACTGCACAAGTTCAATCTCTCTGTTTGCTGTG
TTTGATGAGGGCAAGTCTTGGCACTCTGAAAACAAAAGAACTTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCTGGAGGGACACACCTTCTGGTCAAGAACCCAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCTCACTGCACAGACACTGCTGATGGACCTTTGGACAGTTCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACCTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAAGAAA
CACCCCAAGACATGGGTGCACTACATTTGCTGCTGAGGAAAGAGGACTGGGACTATGUACCACTGGT
CTGGCCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCAAAAAGAAATTTGGAAGA
AAGTACAAGAAAGTCAAGTTTATGGCCCTACACTGATGAAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTTCTGGGACCACTCCTGTATGGGGAAAGTGGGAGACACCCCTGCTCATCATCTTC
AAGAACCAGGCCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTGAGGCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAAACACCTCAAGGACTTCCCCATTTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAATCTGACCCCAAGGTGCCTCACCCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGGCAACCAGATCATGTCTGACAAGAGAAAATGTGATTTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGTGCAACTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCAC'TCCATCAATGGC
TATGTGTTTGGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTTCT
ATTTGGGGCACAAACTGACTTCCCTTTCTGTCTTTCTCTGGATACACCTTCAAGCACAAAGATGGT
TATGAGGACACCCCTGACACTCTTCCATTTCTCTGGGGAAACTGTGTTTCAATGAGCATGGAGAACCCT
GGACTGTGGATTTCTGGGATGCCACAACCTCTGACTTCAAGAAAACAGGGGAAATGACTGCACCTGTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTTGAGCCCAGAAGCTTCTTUTCAGAAATCCACCTGTCTGAAAGAGA
CACCAGAGAGAGATCACCCAGGACAACCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAGGAGGACTTTTACATCTATGATGAGGACGAGAACCAGTCTTCAAGA
TCATTTCCAGAAGAAGACAAGACACTACTTCAATGCTGCTGTGGAAAAGACTGTGGGACTATGGCATG
TCTTCTCTCCCCATGTCTCAGGAAACAGGGCACAGTCTGGCTCTGTGCCACAGTTTCAAGAAAAGTG
GTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCCTGTACAGAGGGGAACTGAATGAGCAC

(Continued)

Figure 16A

CTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTTCAGAAAC
CAGGCCFCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGG
GCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCCTGGGCATACTTCTCTGATGTGGACCTG
GAGAAAGATGTGCACTCTGGCCTGATTGGCCACTCTGGTCTGCCACACCAACACCCTGAACCCCT
GCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTACCATCTTTGATGAAAACCAAG
TCATGGTACTTCACTGASAACATGGAGAGAACTGCAGAGCACCATGCAACATTTCAGATGGAAGAC
CCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGG
CTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAACATF
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAATGCCFCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCCTGGTCTACAGCAAC
AAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCA
ACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATCAAG
ACACAGGGGGCAAGACAGAAATCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTG
GATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCCTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATACATCAGGCTG
CACCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAAGTGGGATGTGACCTGAACTCC
TGCTCAATGCCCTTGGGAATGGAGAGCAAGGCCATTTCTGATGCCAGATCACTGCATCCTCTTAC
TTCACCAACATGTTTGGCACCTGGTCAACATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGCAAT
GCCTGGAGACCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATGAAA
GTCACTGGGGTGACAACCCAGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTCTTG
ATCTCTTCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTGTTT
CAGGGCAACCAGGACTCTTTACACCTGTGGTGAACACTCACTGGACCCCCCTCCTGACAAGATAC
CTGAGAAATTCACCCCAAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCTGGGATGTGAG
GCACAAGACCTGTACTGA (SEQ ID NO:13)

Figure 16B

CS08-FL-NA

ATGCAGATCGAACTGAGCACTTGCTTCTTCCTGTGTCTCCTGCGCTTTTGGCTTCTCCGCCACAAGG
AGATACTATCTCGGTGCCGTGGAGCTCAGCTGGGACTACATGCAGAGCGACTTGGGTGAACTFGCT
GTGGACGCCAGGTTTCCACCCCGCGTGCCTAAGAGTTTCCCGTTCAACACCAGTGTCTGTGACAAG
AAAACCTTCTTCGTGGAATTCACCGAACCTGTTCACATCGCCAAAACCGCGCCCTCCCTGGATG
GGGCTGCTCGGCCCGACGATCCAGGCTGAGGTCTATGACACGGTGGTGATTACCTCAAGAACATG
GCTAGCCACCCGGTGAGCTGCACGCCGTGGGCGTGTCTTATTGGAAAGCGTCCGAGGGTGC GGAG
TACGATGACCAGACTTCACAGCGGGAGAAGGAAGACGACAAAAGTGTTCCTCCGGGGTTCACACACC
TATGTCTGGCAGGTCTGAAGGAGAATGGTCTATGGCTCCGACCCATTGTGCCTCACCTACTCT
TACCTAAGCCATGTGGATCTCGTCAAGGACCTGAACTCGGGGCTGATCGGGCCCTGCTCGTGTGC
CGGGAGGGCTCACTGGCCAAGGAGAAGACCCAACTCTGCACAAGTTCATCCTGCTGTTCGGCGTA
TTCGACGAGGGGAAGTCCCTGGCACTCCGAGACCAAGAACAGCCTGATGCAGGACCCGACGCGACCC
TCGGCCCGTGCCTGGCCAAAGATGCACACCGTGAACGGCTACGTTAACAGGAGCCTACCCGGCCTG
ATCGGCTGCCACCGCAAATCGGTCTACTGGCATGTGATCGGAATGGGCACAACGCCCGAGGTCCAC
AGTATCTTCCTCGAGGGCCACACTTTCTGGTCCGGAATCACCGCCAGGCCAGCCTGGAGATCAGC
CCCATAACCTTTCTGACGGCGCAGACCTTACTCATGGATCTCGGCCAGTTCTCTCTGTTCCTGCCAC
ATTTCTGCCACCAGCAGCATGGGATGGAAGCATATGTGAAAGTGGACTCCTGCCCGGAGGAAACC
CAGCTTAGGATGAAGAACAATGAGGAGGCCGAGGACTACGACGATGACCTTACCATTGAGAAATG
GACGTAGTACGCTTTGACGACGACAACCTCTCCATCCTTCATACAGATTCGCTCCGTCGCCAAGAAG
CACCTAAGACTTGGGTGCACTACATCGCGGCCGAGGAGGAGGACTGGGATTTATGCTCCCTGGTG
CTGGCCCCGACGACCCGAGCTACAAGAGCCAGTACCTGAATAACGGGCCCCAGCGCATCGGCCGG
AAGTACAAGAAAGTGCGGTTCATGGCTTACACGGACGAGACCTTCAAGACCCGGGAGGCTATCCAG
CATGAGAGCGGCATCTTGGGGCCCTCCTGTACGGCGAAGTGGGAGACACACTGCTGATCATCTTC
AAGAACCAGGCGAGCAGGCCCTACAACATCTACCCACCGCATTACCGATGTCCGGCCGTTGTAC
AGCCGACGGCTGCCAAAGGGCGTGAAGCACCTGAAGGACTTTCGGATCCTGCGGGGCGAGATCTTC
AAGTACAAGTGGACTGTGACCGTGGAGGATGGGCCGACCAAGAGCGATCCGCGCTGCCPGACCCGT
TACTACTCCAGCTTTGTCAATATGGAGCGCGACCTCGCTAGCGGCTTGATTTGGCCCTCTGCTGATC
TGCTACAAGGAGTCCGTGGACCAGAGGGGGAATCAGATCATGAGTGACAAGAGGAACGTGATCCTG
TTCTCCGTGTTFCGACGAAAACCGCAGCTGGTATCTCAACGAGAATATCCAGCGCTTCCTGCCAAC
CCGGCCGGTGTGCAGCTGGAGGACCCCGAGTTTCAGGCCAGCAACATCATGCATTCATCAACGGA
TATGTGTTTGTATCCCTGCAGCTCTCAGTGTGTCTGCACGAGGTGCCTACTGGTATATCCTCAGC
ATTTGGGGCACAGACCGACTTCCTGAGCGTGTTCCTCTCCGGGTATACCTTCAAGCACAAAGATGGTG
TACGAGGATACCCTGACCCTGTTCCCTTTAGCGGGGAAAACCGTGTTFATGCTATGGAGAACCC
GGGCTCTGGATCCTTGGCTGCCATAACTCCGACTTCCGCAACCGCGGAATGACCGCGCTCCTGAAA
GTGTCCGAGTTGTGACAAGAACACCGCGCACTATTACGAGGACAGTTACGAGGACATCTCTGCGTAC
CTCCTTAGCAAGAATAACGCCATCGAGCCAAGATCCTTCAGCCAGAACCCCGAGTGTGGAAGAGG
CATCAGCGGGAGATCACCCGACGACCTGCAGTCCGATCAGGAGGAGATFGATTACGACGACAG
ATCAGTGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAAGATGAAAACCGTCCCTCGG
TCCTTCCAAAAGAAGACCCGGCACTACTTCATCGCCGCTGTGGAACCGCTGTGGGACTATGGAATG

(Continued)

Figure 17A

TCTTCTAGCCCTCACGTTTTGAGGAACCGCGCCAGTFCGGGCAGCGTFCCCCCAGTTCAAGAAAGTG
GTGTTCCAGGAGTTCACCGACGGCTCCTTCACCCAGCCACTTTACCGGGGCGAGCTCAATGAACAT
CTGGGCCTGCTGGGACCCACATFCAGGGCTGAGGTGGAGGACAACATFCATGGTGACATTCGGGAAT
CAGGCCAGCAGACCAFCACAGTTTCTACAGTTCACTCATCTCCTACGAGGAGGACCAGCGCCAGGGG
GCTGAACCCCGTAAGAACTTCGTGAAGCCAAACGAAACAAAGACCTACTTCTGGAAGGTCCAGCAC
CACATGGCACCTACCAAGGACGAGTTFCGATTTGCAAGGCCCTGGGCCCTACTTCTCCGACGTGGACCTG
GAGAAAGATGTGCACAGCGGCCCTGATTGGCCCTCTGCTGGTGTGTACACGAAACACACTCAACCCT
GCACACGGGGCGGCAGGTCACTGTGCAGGAATTCGCCCTGTTCTTTACCATCTTTGATGAGACGAAG
TCCTGGTATTTACCGGAAAACATGGAGAGGAACTGCCGCGCACCCCTGCAACATCCAGATGGAAGAT
CCGACATTC AAGGAGAACTACCGGTTCCATGCCATCAATGGCTACATFCATGGACACCCTGCCCTGGC
CTCGTGATGGCCCAAGACCAGCGTATCCGCTGGTATCTGCTGTGCGATGGGCTCCAACGAGAACATC
CATAGTATCCACTTCAGCGGGCATGTCTTCACGGTGGAGAAAAGGAGGAGTACAAGATGGCACTG
TACAACCTCTATCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCTCCAAGGCCGGCATCTGGAGA
GTGGAATGCCTGATCGGCGAGCACCTCCACGCTGGGATGTCCACGCTGTTCCCTCGTTTACAGCAAT
AAGTGCCAGACCCCTCTGGGCAFGGCGAGCGGCCACATCCGCGACTTCCAGATTACAGCCAGCGGC
CAGTACGGTFCAGTGGGCTCCAAAGCTGGCCCGTCTGCACTACTCCGGATCCATCAACGCCCTGGTCC
ACCAAGGAACCGTTCCTGGATCAAAGTAGACCTGCTAGCCCCCATGATCATTFCACGGCATCAAG
ACACAAGGCGCCCGACAGAAAGTTCCTCGAGCCTCTATATCTCCCAGTTCATFCATCATGTATAGCCTG
GACGGAAAGAAGTGGCAGACTTACCGCGGAAACTCGACAGGGACCCTGATGGTATTCTTCGGTAAC
GTGGACAGCTCCGGAATCAAGCACAACATCTTCAACCCACCCATTATCGCCCGCTACATCCGCTG
CACCCCACTCACTATAGCAFTTAGGTCCACCCTGCGAATGGAGCTCATGGGCTGTGACCTGAACAGC
TGTAGCATGCCCCFCGGCATGGAGTCTAAGGCGATCTCCGACGCACAGATAACGGCATCATCTTAC
TTTACCAACATGTTGCTACCTGGTCCCCCTCCAAGGCCCGACTCCACCTGCAAGGGAGATCCAAC
GCCTGGCGGCCACAGGTCAACAATCCCAAGGAGTGGCTGCAAGTGGACTTTCAGAAAACATATGAAA
GTCACCGGAGTGACCACACAGGGAGTGAAGTCTCTGCTGACCAGCATGTACGTGAAGGAGTTCCTC
ATCTCCAGTTCGCAGGATGGCCACCAGTGGACGTTGTTCTTCCAAAACGGTAAAGTCAAAGTCTTC
CAAGGGAACCAGGACAGCTTTACACCCGTCGTGAACCTCCCTGGACCCCCGCTTCTCACTAGATAC
CTCCGCATCCACCTFCAGAGCTGGGTGCACCAGATTGCCCTGCGCATGGAGGTTCTGGGGTGTGAA
GCCCAGGACCTGTACTAA (SEQ ID NO:14)

Figure 17B

CS10-FL-NA

ATGCAGATTGAGCTCTCCACCTGCTTCTTTCTCTGCCTTCTTCGCTTCTGCTTTTCTGCCACACGC
AGGTACTAFTTTGGGAGCAGTGGAACTGAGCTGGGATTACATGCAGAGTGACCTTTGGTGAACFTCCF
GTGGACGCTCGTTTTCCACCTAGAGTTCCCAAGTCCTTCCCCTTCAACACCTCAGTGGTCTACAAG
AAAACGCTGTTTTGTGGAGTTCAGTACCACCTCTTCAACATTGCCAAAACCAAGACCCCTTGGATG
GGATTGCTGGGACCCACAATACAAGCAGAAGTCTACGACACGGTGGTGAATTACCCTGAAGAACATG
GCGTCACACCCTGTTTTCACTTTCAGCTGTTGGGGTTCAGTTATTGGAAAGCCTCAGAGGGTGCGGAA
TACGATGATCAAACCAGCCAGAGGGAGAAGGAAGATGACAAGGTCTTTTCTGGGGGTAGCCATACC
TATGTTTTGGCAGGTGCTGAAAGAGAATGGGCCTATGGCCTCTGATCCCTTTGTGCCTCACATACTCT
TACCTGAGTCACGTCGACCTGGTGAAGACCTGAATAGCGGTCTGATTTGGTGCAGTCTGTTTTGT
AGAGAGGGGAGTTTTGGCCAAGGAGAAAACCTCAGACTCTCCACAAGTTTATCCTCCTGTTTTGCTGTG
TTCCGACGAGGGCAAGTCTTGGCAGCTCTGAAACAAAAGAACTCCCTGATGCAGGACAGAGATGCTGCA
TCTGCAAGGGCTTGGCCAAAATGCACACAGTGAACGGCTATGTGAATCGATCACTGCCAGGACTG
ATAGGCTGTTCATCGCAAGTCAGTGTATTGGCACGTTATCGGGATGGGAACAACCTCCAGAAGTGCAC
AGCAFTCTCCFTTGGGGCCACACTTTTCTGTTTCGGAATCATAGACAGGCCAGCCTTTGAGATCAGC
CCAATCACCTTTCTGACTGCCCAAACCTTGGCTGATGGATCTGGGACAGTTCCCTCCTGTTTTGTAC
ATCTCCTCCCAACATGACGGGATGGAGGCTTATGTGAAGGTGCGATAGCTGTCCGGAGGAACCA
CAACTGAGGATGAAGAACAACGAAGAGGCAGAGGACTATGACGACGATCTGACTGACAGTGAATG
GACGTGGTTCGSTTCGACGATGACAATTTCTCTTCAFTTATCCAGATCCGTTCCGTGGCCAAGAAG
CACCCCAAGACTTTGGGTTCAFTACATCGCTGCTGAGGAGGAGGATTGGGACTACCGCCCTTGGTG
TTGGCCCCAGACGATCGCTCATAACAAGAGCCAGTACCTTAACAATGGTCCACAAAAGGATCGGCCGG
AAGTACAAGAAGGTTAGATTTATGGCTTATACCGACGAGACTTTTAAAACCTAGGGAAGCAATTCAG
CATGAAAGTGGCATTCTTGGACCCTGCTGTATGGCGAGGTTGGCGACACCCTGCTGATTTATCTTT
AAGAACCAGGCAAGCCGGCCCTACAACATCTACCCGCACGGCATAACCGATGTACGACCCCTGTAC
AGTCGCAGACTTCCATAAAGGGGTGAAACACCTGAAGGACTTCCCAATTTCTGCCCGGGGAGATCTTC
AAGTATAAATGGACCCTGACGGTTGAGGATGGTCCCACAAAAGTCCGATCCGAGATGCCCTTACCCGA
TATTATTCAGCTTCGTGAACATGGAAAGGGACCTGGCCAGCGGGCTGATTGGCCCACTGCTGATTT
TGTTACAAGGAGTCTGTGATCAAAGAGGAAACCAATAATGAGCGACAAAACGTAACGTCATCCTG
TTTACGGTCTTTGATGAGAAATAGAAGCTGGTACCTCACAGAAAATATTTACGCGGTTTTCTGCCTAAC
CCCGCAGGGCTCCAGCTGGAAGATCCCGAGTTCCAAGCCTCAAACATCATGCATAGCATCAACGGA
TACGTTTCGATAGCCTGCAGCTGTCCGTCTGTCTCCATGAAGTGGCATAATTGGTACATCCTGAGT
ATCGGGGCGCAGACCGACTTCTGAGCGTGTCTTTTCTGGATAACAGTTCAAACACAAAATGGTC
TATGAAGATACCCTGACTCTGTTTTCCATTTCTCAGGAGAGACAGTCTTTATGAGTATGGAAAATCCF
GGACTGTGGATCCTGGGCTGTCACAATTTGATTTTCGGAACAGAGGCATGACAGCCCTGCTTAAA
GTGAGCTCATGCGACAAGAACACCGGTGATTACTACGAAGATAGCTATGAGGACATCAGTGGGTAT
TTGCTCTCAAGAACAACGCTATCGAGCCACGGTCTTTTTCAGTCAGAAATCCTCCCGTTCTGAAGCGG
CATCAGCGCGAAATAACACGCACAACCTTTCAGTCAGACCAAGAGGAAATCGACTACGATGATACT
ATCTCTGTGGAGATGAAGAAGGAGGATTTTCGACATTTACGACGAGGACGAGAATCAGTCCCCAAGG
AGCTTTCAGAAGAAAACAAGACACTATTTTCAATTGCCCGCGTGGAGCGACTGTGGGACTACGGCATG

(Continued)

Figure 18A

TCTAGCTCTCCGCATGTACTTAGAAATAGGGCACAAAGCGGATCCGTGCCTCAGTTTAAGAAAGTT
GTCTTTTCAGGAGTTTACAGATGGCTCCCTTCACCCAGCCCCTTGTATCGCGGGGAACCTCAATGAACAC
CTGGGCCCTCCTGGGTCCCTTATATTAGGGCCGAAGTCGAGGACAATATCATGGTGACCTTTAGGAAC
CAGGCATCTAGAACCTTACTCTTTCTACTCCTCCCTGATATCCTATGAGGAGGACCAGCGGCAAGGC
GCTGAGCCTCGGAAGAACTTTGTGAAGCCAAATGAAACCAAACATACTTTTGGAAAGTTTCAGCAC
CACATGGCTCCACGAAGGACGAATTTGACTGTAAAGCCTGGGCCTACTTCTCAGATGTAGATCTC
GAGAAAGACGTGCACTCAGGGCTCATTTGGTCCCTTCTGGTCTGTCTACTAATAACCCCTCAATCCA
GCACACGGACGTCAGGTAACCGTCCAGGAATTTGCCCTGTTCTTTACCATTTTTCGATGAGACTAAA
TCCFTGGTACTTTTACCAGAAAACATGGAGAGGAAATTCAGAGCCCCATGCAACATCCAGATGGAGGAC
CCTACCTTCAAAGAGAACTATCGCTTCCATGCCATTAACGGTTACATTATGGATACTCTCCAGGA
CTTGTGTATGGCACAGGATCAGCGGATAAGATGGTATCTGTTGAGCATGGGCTCCAACGAGAATATT
CACAGCATCCATTTCTCCGGTCACTGTTTACAGTGAGAAAGAAAGAAGAGTACAAGATGGCTCTG
TATAATCTCTATCCAGGGCTATTCGAAACGGTGGAGATGTTGCCTAGCAAGGCCGGCATTTGGCGA
GTAGAAATGCCCTTATCGGGGAACATCTGCATGCCGGAATGAGCACGCTCTTCTGGTGTATAGTAAC
AAGTGGCAGACTCCGCTGGGCATGGCATCTGGCCATATACGGGACTTTTCAGATTACGGCTAGCGGG
CAGTATGGGCAGTGGGCACCCAAACTTGCAGGACTGCCTATTTCAGGCTCTATCAATGCATGGTCC
ACCAAGGAACCCCTTCTCTTGGATTAAGGTGGACCTTTTGGCGCCCATGATAATCCATGGGATCAAA
ACCCAGGGCGCTCGTCAGAAATTTCTCACTACTCTACATCTCTCAGTTCAATAATAATGTATTCACTG
GATGGGAAGAAATGGCAGACTTACAGAGGAAACAGCACCGGGACGCTGATGGTGTCTTTGGCAAC
GTGGACAGCAGCGGCATCAAACACAACATCTTCAATCCTCCCATTATTGCCCGTATATTAGACTG
CATCCCACCTACTACTCTATACGCAGCACACTTAGGATGGAGCTCATGGGATGCGACCTGAACAGT
TGTAGTATGCCCTTGGGGATGGAGTCCAAAGCTATAAGCGACGCACAAATTACAGCTAGCTCTTAC
TTTACGAATAFGTTCCGCCACGTGGAGCCCAAGCAAAGCCCGGCTGCATTTGCAGGGTTCGGAGTAAT
GCTTGGCGCCCAACAGGTGAATAACCCCTAAGGAATGGTTGCAAGTAGATTTCCAGAAAACATGAAG
GTAAACCGGCTCACTACACAGGGAGTCAAGTCCCTCTTGACCTCTATGTACGTCAAGGAGTTCCCTG
ATTAGCAGCAGTCAGGATGGGCACCAATGGACACTGTTCTTCCAGAAATGGGAAAGTTAAAGTATTT
CAGGGTAACAGGACTCCTTTACACCTGTGGTGAATAGCCTCGACCCACCCCTGCTGACACGATAC
CTCCGCATCCACCCCTCAGTCTTGGGTGCATCAAATTCGCCCTGCGAATGGAGGTGTTGGGATGCGAA
GCTCAGGACCTCTACTGA (SEQ ID NO:15)

Figure 18B

CS11-FL-NA

ATGCAGATCGAACTCTCTACTTGGCTTCTTCTGTGCCCTTCTGAGGTTCTGCTTCTCTGCCACTCGC
CGATATTACCTCGGGCCGTGGAGTTGAGTTGGGACTACATGCAATCAGATCTGGGCCAACTCCCT
GTGGATGCCCGATTCCCACCGCGCGTGGCCAAAGTCTTCCCATTAAATACTTCTGTGGTGTACAAG
AAGACATTGTTTGTGGAGTTTACCGATCACCTGTTCAACATCGCCAAACCGCGGCCCCCATGGATG
GGTCTGCTTGGGCCACCATTCAAGCGGAGGTCTATGATACAGTGGTGATAACGCTTAAGAACAATG
GCGAGCCACCAGTGTCTCTGCATGCCGTTGGTGTATCATATTGGAAGGCCAGCGAAGGAGCGGAG
TACGATGACCAGACCTCTCAGAGAGAGAAGGAAGACGATAAGGTTTTTCCCTGGCGGAAGTCATACA
TATGTATGGCAGGTCTTGAAGAGAATGGCCGATGGCTTCTGACCCCTTTGCTTACCTATAGT
TATCTGAGCCACGTGGACCTGGTCAAGGACCTCAACAGTGGTCTGATTGGGGCTCTGCTTGTGTTGT
AGAGAGGGTAGCTTGGCTAAGGAGAAAACCCAAACACTCCATAAGTTTCATTTTGTGTTTCGCGGTG
TTCGACGAGGGAAAGAGTTGGCACAGCGAAACAAAGAATTCAGTGTGCAAGACAGGGACGCCGCT
TCCGCAAGGGCTTGGCCTAAGATGCATACGGTGAATGGGTATGTGAACCGGAGCCTCCCGGGCTG
ATCGGGTGGCCATCGCAAGTCTGTTTACTGGCACGTCAATGGGAATGGGGACAACGCCAGAGGTACAT
AGTATATTTCTTGAAGGCCACACGTTCCCTCGTACGGAACCACCGACAGGCTTCCCTGGAGATAAGC
CCCATTACCTTTCTGACCGCTCAGACTCTGCTGATGGACCTTGGCCAGTTTCTCCTGTTCTGCCAT
ATFAGCAGCCACCAGCACGACGGTATGGAAGCATACGTGAAAGTCGATAGCTGTCCFGAGGAGCCT
CAGCTCAGAATGAAGAACAACGAGGAGGCCGAAGACTATGACGATGACCTTACAGATFCCGAGATG
GACGTGGTGGCTTTGACGACGATAACAGTCTTAGTTTCATTCAAATCAGATCCGTAGCCAAAAG
CATCCAAAGACATGGGTGCATTACATTGCAGCCGAAGAGGAGGATTGGGATTTATGCGCCCTTGT
CTGGCTCCAGATGACAGGAGCTATAAGTCCCAGTACTTGAACAACGGGCCACAGCGAATCGGTAGA
AAATATAAGAAGGTAAGATTTCATGGCCCTACACTGACGAAACATTTAAAACCAGGGAAGCTATCCAA
CACGAATCTGGAATTTCTCGGCCCTCTGCTCTACGGTGGAGGTGGGGGACACCTTGGCTGATCATTTTC
AAAAATCAGGCATCCAGGCCTTACAACATATACCCCATGGCATCACCGATGTCCGGCCCGCTGTAT
TCCAGAAGACTCCCCAAGGGAGTGAAACATCTGAAAGATTTTCCCATCCTGCCGGGCGGAGATCTTT
AAATACAAATGGACTGTGACTGTAGAGGACGGGCCCTACAAAATCAGACCCACGGTGCCTGACAAGG
TATTACAGTAGCTTTCGTCAACATGGAACCGGACCTCGCCAGCGGACTCAATGGCCCACTGTTGATC
TGTTACAAAGAGTCAGTGGATCAGAGGGGAAATCAGATCATGAGCGATAAGAGAAAACGTTATCCTG
TTTAGTGTCTTCGACGAGAACCGGTCTTGGTACCTTACTGAGAACATCCAGAGGTTCCCTGCCGAAT
CCGGCTGGCGTTCAGCTCGAGGACCCAGAGTTCCAGGCCAGTAATATAATGCACCTCAATCAACGGT
TATGTGTTTCGATAGCCTGCAGCTGAGCGTCTGCCTCCACGAGGTAGCCTATTGGTACATATTGTCC
ATCGGGGCTCAGACCGATTTTCTGTCCGTGTTCTTTAGCGGGTATACCTTTAAACATAAAATGGTC
TATGAAGACACCTGACCCTGTTCCATTTCTCCGTTGAGACTGFGTTCATGTPCCATGGAGAACCCA
GGGCTGTGGATCCTGGGGTGTCACAATAGTACTTTAGGAATCGGGGAATGACGGCACTGCTGAAG
GTGAGTTCCTGCGATAAAAATACAGGAGATTACTATGAGGATAGTTACGAGGATATCAGTGCCTAT
CTGCTTTCAAAAAACAACGCAATGAGCCCCGGTCTTTCTCACAAAACCCCCGGTGTGAGCGC
CACCAGCGGAAATTAACCCGACAACCTTGCAGTCCGACCAGGAGGAAATCGATTTATGACGATACT
ATCAGTGTAGAAATGAAAAAGGAGGATTTTGATATTTACGACGAAGACGAGAACCAGTCTCCGCGA

(Continued)

Figure 19A

AGTTTTTCAGAAGAAAACGCGACACTACTTTTATAGCTGCCGTTGGAACGACTCTTGGGATTATGGCATG
TCCTCCAGCCCTCATGTCTTAGGAATCGAGCGCAGAGTGGCTCTGTGCCCTCAGTTCAAAAAGGTT
GTGTTCCAGGAATTCACCGACGGCTCAFTTACCCAGCCGCTGTACAGAGGCGAACTCAACGAACAC
CTTGGGCTGCTTGGGCCATATATTGAGCAGAGGTGGAAGATAATAFCATGGTAACCTTTAGAAAC
CAGGCGTCAAGACCCATTCTCTTCTACAGTTCTCTGATCAGCTACGAGGAGGACCAAAGACAGGGA
GCTGAACCCAGGAAGAACTTTGTGAAACCTAATGAGACCAAGACCTACTTCTGGAAGGTCCAGCAC
CATATGGCCCCAACTAAAGATGAATTCGATTGCAAGGCCCTGGGCTTATTTACGCGACGTGGATCTC
GAAAAGGATGTGCACAGCGGGTTGATCGGACCGCTTTTGGTGTGCCACACAAAATACCCCTCAATCCT
GCCACCGGGCGCAGGTCACAGTTCAAGAGTTTGCACCTCTTCTTTACAATATTTGACGAGACAAAG
TCATGGTATTTTTACAGAGAATATGGAGAGAAATGTGCGCACCTTGCAACATTCAGATGGAGGAC
CCCACATTTAAGGAGAATTACAGATTTTCATGCTATCAATGGGTACATTAATGGATACTCTGCCPGGT
CTGGTCATGGCCAGGATCAGCGCATAAGGTGGTACTTGCTGAGCATGGGATCTAATGAGAATATA
CACAGCATTCACTTCAGTGGCCACGTTTTTTACTGTTAGAAAAGAAGGAGGAGTACAAAATGGCGCTC
TACAACCTTTACCCGGGTGTGTTTGAGACAGTGGAGATGCTGCCAAGCAAGGCAGGCATCTGGAGG
GTTGAGTGTCTTATTGGGGAGCATCTGCATGCTGGAATGTUCACCCCTCTTTCTTGTGTACAGCAAT
AAGTGCCAGACACCGCTTGGCATGGCCAGCGGCCACATTAGGGACTTTCAGATAACTGCCAGTGGGA
CAGTACGGCCAGTGGGCTCCCAAGCTTGCAAGACTCCACTACTCCGGAAGCATAAAACGCATGGAGC
ACCAAGGAACCCCTTCTCTTGGATTAAAGGTGGACCTGCTGGCGCCAATGATCATTACGGCCATAAAA
ACCCAAGGGGCACGACAGAAATTTTCACTCTTTGTATATTAGTCAGTTTTATCATCAATGTACAGCTTG
GATGGAAAAGAAGTGGCAGACGTACAGGGGCAATTTACAGGAACACTTATGGTGTFTTTTTGGGAAT
GTCGATTCCAGCGGGATCAAACATAACATCTTCAATCCTCTCTAATTATCGCCCGATATAATCCGCTG
CACCCCTACGCATTAATCCATCAGGTCCACATTTGAGAATGGAACCTGATGGGGTGGGACCTGAATAGT
TGTAGTATGCCACTGGGCATGGAGTCTAAAGCCATCAGCGATGCACAGATCACTGCCAGCTCTTAC
TTCACCAACATGTTTTGCAACTTGGTCCCCCTCTAAAGCTCGCCTGCATCTGCAGGGACGCTCAAAT
GCATGGCGACCACAGGTGAACAATCCAAAAGAGTGGCTCCAGGTGACTTTFCAGAAGCAATGAAG
GTAACAGGAGTGACAACCCAGGGTGTAAAAAGCCTCCTTACGAGTATGTACGTTAAGGAGTTTCTG
ATTTCTAGCTCCCAGGACGGACACCAGTGGACTCTGTTCTTCCAGAACGGCAAAGTGAAGGATTTT
CAGGGAAACCAGGATTTCTTTTACCCCGGTAGTGAATAGCCTGGATCCACCCTTCTGACCCGCTAT
CTGAGAATTCATCCACAATCCTGGGTGCATCAGATTCGCCCTCCGGATGGAAGTGTCTGGCTGTGAA
GCTCAGGATCTGTATTAG (SEQ ID NO:16)

Figure 19B

CS40-FL-NA

ATGCCAAATAGAGCTCTCCACCTGCTTCTTTCTGTGCTTTTGGGATTTCTGCTTTAGTGCCACCAGA
AGA'ACTACCTGGGTGCAGTGGAACTG'CATGGGACTATATGCAAAGTGA'CTCGGTGAGCTGCC'
GTGGACGCAAGATTTCC'PCTAGAGTGCCAAAATCTTTTCCATTC AACACCTCAGTCGTGTACAAA
AAGACTCTG'PTGTAGAA'PTACGGATCACCTTTTCAACATCGCTAAGCCAAGGCCACCTTGGATG
GGTCTGCTAGGTCTACCATCCAGGCTGAGGTTTATGATACAGTGGTCA'ATTACACTTAAGAACATG
GCT'FCCCATCT'GTCAGTCT'PCATGCTG'PTGGTGTATCCTAC'FGGAAAGC'PTCTGAGGGAGCTGAA
TATGATGATCAGACCAGTCAAAGGGAGAAAAGAAGATGATAAAAGTCTTCCC'GGTGGAAAGCCATACA
TATGTCTGGCAGGTCTT'GAAAAGAGAATGGTCCAATGGCTCTGACCCACTGTGCC'TTACCTACTCA
TATCT'PTCTCA'GTGGACC'GGTAAAAGACT'GAA'PCAGGCC'CAT'GGAGCCCTACT'AGTATG'
AGAGAAGGGAGTCTGGCC'AAGGAAAAGACACAGACCTTGCACAAA'TTAFACTACTTTTGGCTGTA
T'PTGATGAAGGGAAAAG'PTGGCAC'PCAGAAAACAAAAGAAC'PCTTGA'FGCAGGA'FAGGGATGC'FGCA
TCTGCTCGGGCCCTGGCTAAAATGCACACAGTCAATGGTTATGTAACAGGTCTCTGCCAGGTCTG
AT'FGGATGCCACAGGAAATCAGTCTAT'FGGCATG'FGATTGGAAT'GGGCACC'ACTCC'FGAAGTGCAC
TCAATATTCCTCGAAGGTACACATTTCT'GTGAGGAACCATCGCCAGGCGTCTTGGAAATCTCG
CCAATAAC'PTTCT'FACTGCTCAAACACTCT'FGATGGACC'PTGGACAG'PTTCT'ACTGTT'PTG'PCAT
ATCTCTTCCCACCAACATGATGGCATGGAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCC
CAACTACGAATGAAAAATAA'GAAGAAGCGGAAGACTATGATGATGATCT'FACTGA'PTGAAAATG
GATGTGGTCAAGT'PTGATGATGACAACCTCTCTCTCTTATCCAAATTCGCTCAGT'FGCCAAAGAAG
CATCTTAAAAC'PTGGT'ACAT'FACA'PTGCT'GCT'GAAGAGGAGGAC'FGGGACTA'FGCTCCC'PTAGTC
CTCGCCCCCGATGACAGAAGT'ATAAAAAGTCAATAT'PTGAACAA'PTGGCC'PCAGCGGAT'GGTAGG
AAG'FACAAAAAAG'PCCGAT'PTATGGCA'FACACAGATGAAACC'PTAAGAC'PCGTGAAGCTATTCAG
CATGAATCAGGAATCT'PTGGACC'PTACT'PTATGGGGAAGT'PTGGAGACACAC'FG'PTGAT'ATAATTT
AAGAATCAAGCAAGCAGACCA'FA'FAACATCT'ACCCTCACGGAA'CACT'GATG'PCCGTCC'PT'FGTAT
TCAAAGGAGATTACCAAAGGTGTAAAACATTTGAAGGATTTTCCAAATCTCGCCAGGAGAAAATATTC
AAA'FA'AAAATGGACAGT'GTAGAAAGATGGGCCAACTAAA'FCAGATCC'PCGGTGGCT'FGACCCGC
TATFACTCTAGTTTCTGTTAATA'FGGAGAGAGATCTAGCT'PCAGGACTCA'PTGGCCCTCTCTCTCATC
TGCTACAAAGAATCTG'PAGATCAAAGAGGAAAACCAGATFAATGTCAAGCAAGAGGAATGTCA'CTCTG
TT'PTCTGTAT'PTGATGAGAACCGAAGCT'GGTACC'PCACAGAGAA'ATACAAACGCT'PTCTCCCCAA'
CCAGCTGGAGTGCAGCT'FGAGGATCCAGAGTTCOAAGCTCCAAACATCAT'FGCACAGCATCAATGGC
TATGTT'PT'PTGATAGT'PTGCAGT'PTG'PCAGT'PTG'PTTGCAT'FGAGGTGGCATACT'GGTACATTTAAGC
ATTGGAGCACAGACTGACTTCC'PTCTGTCTCTCTCTCTGGATATACCTTCAAACACAAAAATGGTC
TATGAAGACACACTCACCC'PAT'PCCAT'PTC'PCAGGAGAAAC'FGTCTTCA'FGTCSAT'GGAAAAACCCA
GGTCTATGGAT'CTGGGGTGCACAACTCAGACTTTCGGAACAGAGGCCAT'GACCGCTTACTGAAG
G'PTTCTAG'PTGTGACAAGAACACT'GGTGA'PT'PTACGAGGACAGT'FA'PGAAGA'FA'PTT'GAGCATAC
TTGCTGAGTAAAAACAATGCCAT'FGAACCAAGAAGCTTCTCCCAGAA'PCCACCAGTCTTGAACGC
CATCAACGGGAAA'AACTCGTACTACTCT'PTCAGT'CATGATCAAGAGGAAA'PTGACTA'FGATGATACC
ATATCAGT'FGAAATGAAGAAGGAAGATTTTGACATTTATGATGAGGATGAAAATCAGAGCCCCCGC
AGCTTTCAAAGAAAACACGACACT'ATTT'FA'PTGCTGCAGTGGAGAGGCTCT'GGGATTA'FGGGATG

(Continued)

Figure 20A

AGTAGCTCCCCACATGTTCTAAGAAAACAGGGGCTCAGAGTGGCAGTGTCCCTCAGTTCAAGAAAAGTT
GTTTTCCAGGAATTTACTGATGGCTCCTTTACTCAGCCCTTATAACCGTGGAGAACTAAATGAACAT
TTGGGACTCCTGGGGCCATATATAAGAGCAGAAGTTGAAGATAAATATCATGGTAACTTTTCAGAAAT
CAGGCCTCTCGTCCCTATTCCCTTCTATTCTAGCCTTATTTCTTATGAGGAAGATCAGAGGCAAGGA
GCAGAACCTAGAAAAAACTTTGTCAAGCCTAATGAAACCAAAACTTACTTTTTGGAAAAGTGCAACAT
CATATGGCACCCACTAAAAGATGAGTTTGACTGCAAAGCCTGGGCTTATTTCTCTGATGTTGACCTG
GAAAAAGATGTGCACTCAGGCCTGATTGGACCCCTTCTGGTCTGCCACACTAACACACTGAACCCCT
GCTCATGGGAGACAAGTGACAGTACAGGAATTTGCTCTGTTTTTTCACCATCTTTTGATGAGACCAAAA
AGCTGGTACTTCACTGAAAATATGGAAAGAACTGCAGGGCTCCCTGCAATATCCAGATGGAAGAT
CCCCTTTTAAAGAGAAATTAATCGCTTCCATGCAATCAATGGCTACATAAATGGATACACTACCTGGC
TTAGTAATGGCTCAGGATCAAAGGATTCGATGGTATCTGCTCAGCATGGGCAGCAATGAAAACATC
CATTTCTATTCATTTTCAGTGGACATGTGTTCACTGTACGAAAAAAAGAGGAGTATAAAAATGGCACTG
TACAACTCTATCCAGGTGTTTTTGGAGACAGTGGAAATGTTACCATCCAAAGCTGGAATTTGGCGG
GTGGAATGCCTTATTGGCGAGCATCTACATGCTGGGATGAGCACACTTTTTCTGGTGTACAGCAAT
AAGTGTGAGACTCCCTGGGAATGGCTTCTGGACACATTAGAGATTTTCAGATTACAGCTTCAGGA
CAATATGGACAGTGGGCCCCAAAGCTGGCCAGACTTCATTATTCCGGATCAATCAATGCCTGGAGC
ACCAAGGAGCCCTTTTTCTTGGATCAAGGTGGATCTGTTGGCACCAATGATTAATTCACGGCATCAAG
ACCCAGGGTGGCCGTCAGAAAGTTCTCCAGCCTCTACATCTCTCAGTATCATCATGTATAGTCTT
GATGGGAAGAAGTGGCAGACTTATCGAGGAAATCCACTGGAACCTTAAATGGTCTTTCTTTGGCAA
GTGGATTCATCTGGGATAAAAACACAATATTTTTAAACCCTCCAATTATGCTCGATACATCCGTTTG
CACCAACTCATTATAGCATTCCGAGCACTCTTCGCATGGAGTTGATGGGCTGTGATTTAAATAGT
TGCAGCATGCCATTGGGAATGGAGAGTAAAGCAATATCAGATGCACAGATTAAGTCTTCATCCTAC
TTTACCAATATGTTTGGCACCTGGTCTCTTTCAAAAAGCTCGACTTCACCTCCAAGGGAGGAGTAA
GCCTGGAGACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAAGTGGACTTCCAGAAGACAATGAAA
GTCACAGGAGTAACTACTCAGGGAGTAAAATCTCTGCTTACCAGCATGTATGTGAAGGAGTTCCCTC
ATCTCCAGCAGTCAAGATGGCCATCAGTGGACTCTCTTTTTTTCAGAATGGCAAAGTAAAGGTTTTT
CAGGGAAATCAAGACTCCTTCACACCTGTGGTGAAGTCTCTAGACCCACCGTTACTGACTCGCTAC
CTTCGAATTCACCCCCAGAGTTGGGTGCACCAGATTTGCCCTGAGGATGGAGGTTCTGGGCTGCGAG
GCACAGGACCTCTACTGA (SEQ ID NO:17)

Figure 20B

CH25-FL-NA

ATGCAGATCGAGCTGTCCACATGCTTTTTTCTGTGCCTGCTGCGGTTCTFGCTTCAGCGCCACCCGG
CGGTACTACCTGGGCGCCGTTGGAGCTGTCTCTGGGACTACATGCAGAGCGACCTGGGCGAGCTGCC
GTGGACGCCCCGTTCCCCCCCAGAGTGCCTCAAGAGCTTCCCTTCAACACCAGCGTGGTGTACAAG
AAAACCTGTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCCTGGATG
GGCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGTATCACCTGAAGAACATG
GCCAGCCACCCCGTGGAGCTGCACGCGCTGGGCGTGGAGCTACTGGAAGGCCTCCGAGGGCGCCGAG
TAGGACGACCAGACCAGCCAGCGGGAGAAAGAGGACGACAAAGTCTTTCCTGGCGGCAGCCACACC
TACGTGTGGCAGGTCTGAAAGAAAACGGCCCCATGGCCTCCGACCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACAGCGGGCTGATGGGGCCCTGCTGGTCTGC
CGGGAGGGCAGCCTGGCCAAAGAGAAAACCCAGACCCTGCACAAGTTCATCCTGCTGTTCCGCGTG
TTCGACGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACCGGGACGCCGCC
TCTGCCAGAGCCTGGCCAAAGATGCACACCGTGAACGGCTACGTGAACAGAAGCCTGCCCGGCTG
ATTGGCTGCCACCGAAGAGCGTGTACTGGCAGCTGATCGGCATGGGCACCACACCCGAGGTGCAC
AGCATCTTTCTGGAAGGGCACACCTTTCTGGTGGGAACCACCGGCAGGCCAGCCTGGAAATCAGC
CCTATACCTTCCCTGACCGCCAGACACTGCTGATGGACCTGGGCCAGTTCCTGCTGTTTTGCCAC
ATCAGCTCTCACCAGCAGCAGCGCATGGAAGCCTACGTGAAGGTGGACTCCTGCCCGAGGAAACC
CAGCTGCGGATGAAGAACAACGAGGAAGCCGAGGACTACGACGACGACCTGACCGACAGCGAGATG
GACGTGGTGGGTTTCGACGACGACAACAGCCCCAGCTTCATCCAGATCAGAAGCGTGGCCAAGAAG
CACCCAAAGACCTGGGTGCACTACATCGCCCGCAGGAAAGAGGACTGGGACTACGCCCCCTGGTG
CTGGCCCCGACGACAGAAGCTACAAGAGCCGTACCTGAACAATGGCCCCAGCGGATCGGCCGG
AAGTACAAGAAAGTGGGTTTCATGGCCTACACCGACGAGACCTCAAGACCCGGGAGGCCATCCAG
CAGGACCGCCATCCTGGGCCCCCTGCTGTACGGCGAAGTGGGCGACACACTGCTGATCATCTTC
AAGAACCAGCCAGCCCGCCCTACAACATCTACCCCCACGGCATCACCGACGTGGGCCCCCTGTAC
AGCAGGCGGCTGCCCAAGGGCCTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGACGGCCCCACCAAGAGCGACCCAGATGCCTGACCCGG
TACTACAGCAGCTTCGTGAACATGGAACGGGACCTGGCCTCCGGCTGATCGGACCTCTGTGATC
TGCTACAAAGAAAGCGTGGACCAGCGGGCAACCAGATCATGAGCGACAAGCGGAACGTGATCTTG
TTCAGCGTGTTCGATGAGAACCGGTCTGTTATCTGACCGAGAACATCCAGCGGTTTCTGCCAAC
CCTGCCGGGGTGCAGCTGGAAGATCCCGAGTTCAGGCCAGCAACATCATGCACTCCATCAATGGC
TACGTGTTTCGACAGCCTGCAGCTGTCCGTGTGTCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCAGACCGACTTCCCTGAGCGTGTCTTTCAGCGGCTACACCTTCAAGCACAAAGATGGT
TACGAGGACACCCTGACCCTGTTCCCTTTTTCAGCGGCGAGACCGTGTTCATGAGCATGGAACACCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCCGGAACCGGGCATGACCGCCCTGCTGAAG
GTGTCCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGATATCAGCGCTTAC
CTGCTGTCCAAGAACAACGCCATCGAGCCCAGAAGCTTCAGCCAGAACCCCTGTGCTGAAGCGG
CACCAGAGAGAGATCACCCGGACCACCTGCAGTCCGACCAGGAAGAGATCGATTACGACGACACC

(Continued)

Figure 21A

ATCAGCGTGGAGATGAAAAAAGAAGATTTTCGACATCTACGACGAGGACGAGAACCAGAGCCCCCGG
TCCTTCCAGAAGAAAAACCCGGCACTACTTTATFCGCCGCCGTGGAGCGGCTGTGGGACTACGGCATG
AGCAGCAGCCCCACGTCGCTGCGGAACCGGGCCAGAGCGGCAGCGTCGCCACAGTTCAAGAAAAGTG
GTGTTCCAGGAATTCACCGACGGCAGCTTCACCCAGCCCCGTACCCGGGGCGAGCTGAACGAGCAC
CTGGGGCTGCTGGGGGCCCTACATCAGGGCCGAAAGTGGAGGACAACATCATGGTGAACCTTCCGGAAAT
CAGGCCAGCAGACCCCTACTCCCTTCTACAGCAGCCTGATCAGCTACGAAGAGGACCAGCGGCAGGGC
GCTGAACCCCGGAAGAACCTTCGTCGAAGCCCAATGAGACCAGAGCTACTTCTGGAAAAGTGCAGCAC
CACATGGCCCCCACCAGGACGAGTTCGACTGCAAGGCCGTTGGGCTACTTTCAGCGACSTGGATCTG
GAAAAGGACGTCGACTCTGGACTGATTTGGCCCTCTGCTGGTGTGTCACACCCAAACACCCCTGAACCCC
GCCACGGCCCGGCAGGTGACCGTCGAGGAAATTCGCCCCGTTCTTCACCATCTTCGACGAGACCAAG
TCCTGGTACTTTCACCCGAGAATAATGGAACGGAACTCGAGAGCCCCCTGCAACATCCAGATGGAAGAT
CCTACTTTCAAAAGAGAACCTACCGGTTCACCGCCATCAACGGCTACATCATGGACACCCCTGCCCTGGC
CTGGTGGATGGCCCAGGACCCAGAGGATCCGGTGGTATCTGCTGTCATGGGGCAGCAACGAGAATAATC
CACAGCATCCACTTCAGCGGCCACGTCGTCACCGTCGAGGAAGAAAGAGTACAAGATGGCCCTG
TACAACCTGTACCCCGGCGTGTTCGAGACCGTCGGAGATGCTGCCCAGCAAGGCCGGCATCTGCGGG
GTGGAGTGTCTGATCGGGCAGCACCTGCAATGCCGGGATGAGCACCCCTGTTTTCGGTGTACAGCAAC
AAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCCGGGACTTCCAGATCAACCGCTCCGGC
CAGTACGGCCAGTGGGCCCCCCAAGCTGGCCCGGCTGCACCTACAGCGGCAGCACTAACGCCCTGGTCC
ACCAAAGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCATGATCATCCACGGCATTAAG
ACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCTGTACAGCCCTG
GACGGCAAGAAGTGGCAGACCTACCGGGCAACAGCACCGGCACCCCTGATGGTGTTCCTCGGGCAAC
GTGGACAGCAGCGCATCAAGCACAAACATCTTTCAACCCUCCCCATCATCGCCCGGTACATCCUGGCTG
CACCCACCCACTACAGCATCAGATCCACCCCTGCGGATGGAACTGATGGGGTGGCAGCTGAACTCC
TGCAGCATGCCCTCTGGGCATGGAAAAGCAAGGCCATCAGCSACGCCAGATCACAGCCAGCAGCTAC
TTCACCAACATGTTTCGCCACCTGGTCCCCCTCCAAGGCCAGGCTGCACTGCAGGGCCGGTCCAAC
GCTTGGCGGCCCTCAGGTGAACAACCCCAAAGAAATGGCTGCAGGTGGACTTTCAGAAAACCAFGAAG
GTGACCGGGCTGACCACCCAGGGCGTGAAGAAGCCCTGCTGACCAGCATGTACGTGAAAAGATTTCTG
ATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCCTGTTCTTTTCAGAACGGCAAGGTGAAAAGTGTTC
CAGGGCAACCAGGACTCCTTCACCCCGCTGGTGAACCTCCCTGGACCCCCCCCCCTGCTGACCCGCTAC
CTGCGGATCCACCCCCAGTCTTGGGTGCACCCAGATCGCCCCGAGGATGGAAAGTGTGGGATGTPGAG
GCCCAGGATCTGTACTGA (SEQ ID NO:18)

Figure 21B

EVIII-EL-AA

mqielstcfff lcllrfcfsa trryylgave lswdymqsdg gelpvdarfp prvpksfpfn
 tsvvykktlf veftdhlfni akprppwmgj lgptiqaevy dtvvitiknm ashpvsilhav
 gvsvwkaseg aeyddqtsqr ekeddkvfpg gshtyvwqvl kengpmasdp lcltysylsh
 vdlvkdinsg ligallvcre gslakektqt lkhfillfav fdegkswhse tknslmqdrd
 aasarawpkm htvngyvnrslp gligchrk svywhvigmg ttpevhsifl eghtflvrnh
 rgasleispi tfltaqtllm dlqqfllfch isshqhdgme aykvvdscpe epqlrmkne
 eaedydddltd semdvvrfd ddnspsfiqi rsvakkhpk wvhyiaaeed dwdyaplvla
 pddrsyksqy lnnqpqrigr kykkvrfmay tdetfktrea iqhesgilgp llygevgdtl
 liifknqasr pyniypggit dvrplysrll pkgvkhkdf pilpgeifky kwtvtvedgp
 tkdprcltr yyssfvnmer dlasgligpl licykesvdq rgnqimsdkr nvilfsvfde
 nrswyiteni qrflpnpagv qledpefqas nimhsingyv fdsiqslvcl hevaywyils
 igaqtdflsv ffsqytfkhh mvyedtltilf pfsgetvfms menpglwilg chnsdfrnrg
 mtallkvssc dkntgdyyed syedisayll sknnaieprs fsqnsrhpst rqkqfnatti
 pendiektqp wfahrtpmpk ignvssdill mlrqsptph glslsdlqea kyetfsddps
 pgaidsnsl semthfrpqi hhsqdmvftp esglqlrline klgttaatel kkldfkvsst
 snnlistips dnlaagtdnt ssigppsmvp hydsqldttl fgkksppite sggplslsee
 nndskillesg lmsqesswg knvsstesgr lfkgkrahgp alltkdnalf kvsisliktn
 ktsnnsatnr kthidgpsll ienspsvwqn ilesdtefkk vtplihdrml mdknatalrl
 nhmsnkttss knmemvqqkk egpipdaqn pdmsffkmlf lpesarwiqr thgknslnsg
 qgspkqlvs lgpeksvegq nflseknkvv vgkgeftkdv glkemvfps rnlfltnldn
 lhennthnqe kkiqeeiekk etliqenvvl pqihtvtgtk nfmknlflls trqnvegsyd
 gayapvlqdf rslndstnrt kkhtahfskk geeenleglg nqtkqiveky acttrispnt
 sqqnfvtrqs kralkqfrlp leetelekri ivddtstqws knmkhltpst ltqidyneke
 kgaitqspis dcltrshsip qanrspipia kvssfpsirp iyltrvlfqd nsshlpaasy
 rkkdsgvqes shflqgakkn nlsalittle mtgdqrevgs lgtsatnsvt ykkventvlp
 kpdlpktsgk vellpkvhiy qkdlfptets ngspghldlv egslqgteg aikwneanrp
 gkvpflrvat essaktpski ldplawdnhy gtqipkeewk sqekspekta fkkkdtllsl
 nacesnhaia ainegqnkpe ievtwakqgr terlcsqnp vlkrhqreit rttlqsdqee
 idyddtisve mkkedfdiye edenqsprsf qktrhyfia averlwdygm sssphvlrnr
 aqsgsvppfk kvvfqeftdg sftqplyrge lnehlgllgp yiraevedni mvtfrnqasr
 pysfysslis yeedqrggae prknfvkpne tktyfwkvqh hmaptkdefd ckawayfsdv
 dlekdvhsgl igpllchtn tlnpahgrqv tvqefalfft ifdetkswyf tenmerncra
 pcniqmedpt fkenyrfhai ngyimdtlpg lvmaqdrir wyllsmgsne nihsihfsg
 vftvrkkeey kmalynlypg vftvempls kagiwrvecl igeihhags tflflvsnko
 qtplgmasgh irdfqitasg qyqwapkla rihygsina wstkepfswi kvdlapmii
 hgiktqgarq kfsslyisqf iimysldgkk wqtyrgnstg tlmvffgnvd ssgikhnifn
 ppiaryirl hpthysirst lrmelmgodl nscsmplgme skaisdaqit assyftnmfa
 twspskarlh lqgrsnawrp qvnnpkewlq vdfqktmkvt gvttqgvksl ltsmyvkefl
 isssqdghqw tlffqngkvk vfqgnqdsft pvvnslpppl ltrylrihpq swvhqialrm
 evlqceaadl y (SEQ ID NO:19)

Figure 22

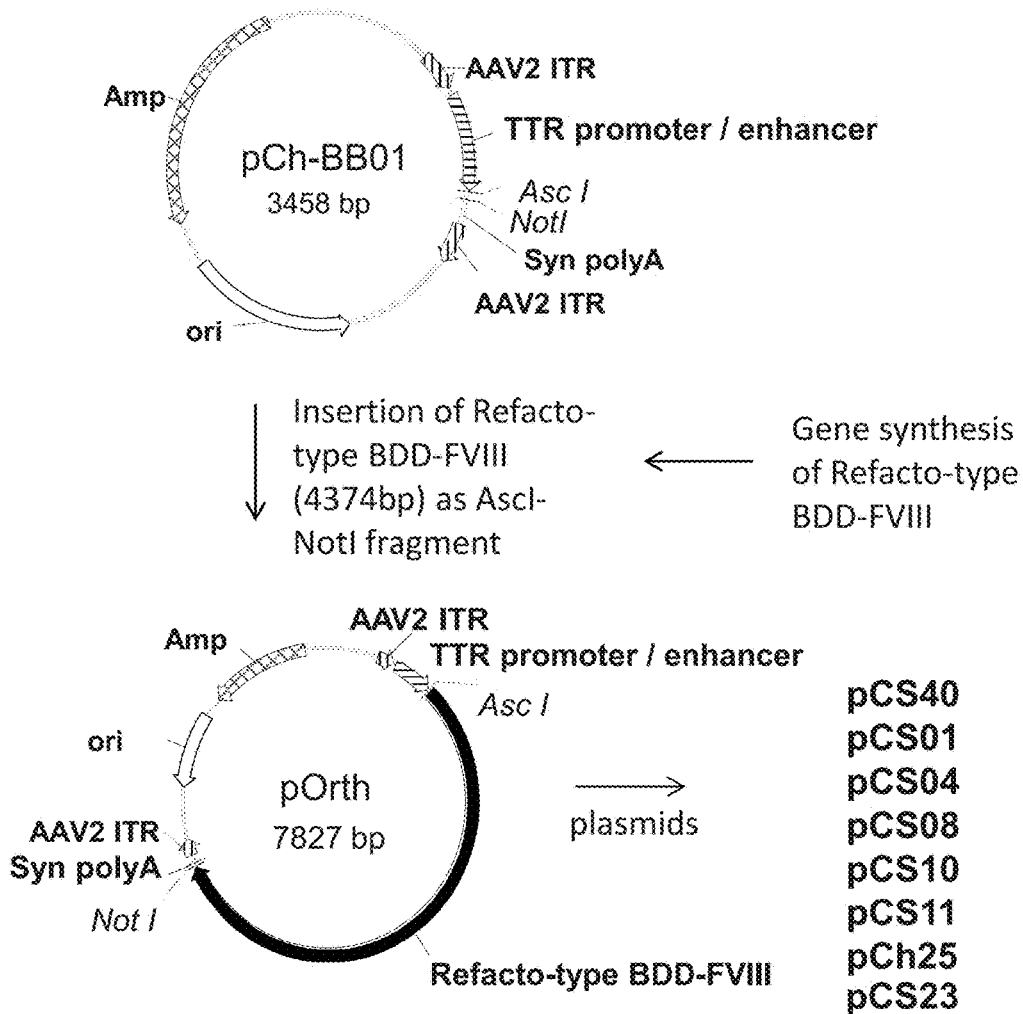


Figure 23

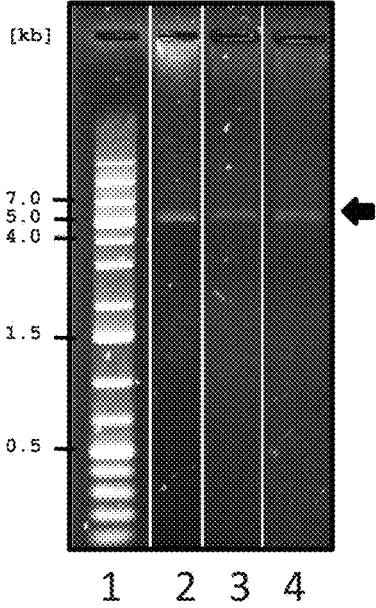


Figure 24

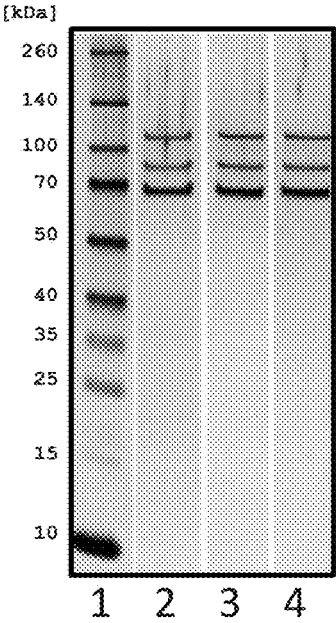


Figure 25

CS23-FL-NA

atgcagattgagctgagcaccctgcttcttccctgtgcctgctgaggttctgcttctctgccaccagg
agatactacctgggocgctggagctgagctgggactacatgcagctotgacctgggagagctgct
gtggacgcccagggttccccccagagtgcccaagagcttccccctcaacacctcagtggtgtacaag
aagaccctgttctgtggagttcacccagaccctgttcaacatcgccaagcccaggccccctggatg
ggcctgctgggccccaccatccagggccgaggtgtacgacaccctgggtgatcaccctgaagaacatg
gccagccaccccgtagcctgcaacgcccgtgggctgtagctactggaaggcctctgagggcggccgag
tatgacgaccagaccagccagaggggagaaggaggacgacaagggtgttccccggggcagccacacc
tacgtgtggcagggtgctgaaggagaacggccccatggccagccagccccctgtgcctgacctacagc
tacctgagccacgctggaccctggtgaaggacctgaactctggcctgatcgggcccctgctggtgtgc
agggagggcagcctggccaaggagaagaccagaccctgcacaagttcatcctgctgttggccgtg
t.tcgatgagggcaagagctggccacagccagaccaagaacagccctgatgcaggacagggatgcccgc
tctgccagggcctggccccaaagatgacacaccctggaacggctacgtgaacaggagcctgcccggcctg
atcgctgcccacaggaagctctgtgtactggcacgtgatcggcatgggcaccacccccggaggtgcac
agcatcttccctggagggccacaccttccctgggtgaggaaccacaggcagycaccagcctggagatcagc
cccataaccttccctgaccgcccagaccctgctgatggacctgggcccagttccctgctgttctgccac
atcagcagccaccagcaacgacggcatggaggccctacgtgaagggtggacagctgcccagggagccc
cagctgaggatgaagaacaacgaggaggccgaggactatgatgatgacctgaccgactctgagatg
gacgtggtgaggtttgatgatgacaacagccccagcttcatccagatcaggtctgtggccaagaag
cccccaagaacctgggtgcaactacatcgccggcagggaggaggactgggactacgccccctgggtg
ctggcccccgacgacaggagctacaagagccagtaacctgaacaacggccccagaggatcggcagg
aagtacaagaaggctcagattcatggcctacaccgacgagaccttcaagaccagggaggccatccag
cacgagctctggcatcctgggccccctgctgtacggcaggggtgggacacacctgctgatcatctt
aagaaccaggccagcaggccctacaacatctacccccagggcatcaccgatgtgaggccccctgtac
agcaggaggctgcccaggggcgtgaagcacctgaaggacttccccatcctgcccggcgagatctt
aagtacaagtggaccctgaccgtggaggatggccccaccaagctctgaccccaggtgcctgaccagg
tactacagcagcttctgtaacatggagagggacctggcctctggcctgatcggccccctgctgatc
tgctacaaggagagcgtggaccagaggggcaaccagatcatgtctgacaagaggaacgtgatcctg
ttctctgtgttctgatgagaacaggagctgggtatctgaccgagaacatccagaggttccctgcccac
ccggccggcgtgcagctggaggacccccaggttccaggccagcaacatcatgcacagcatcaacggc
tacgtgttcgacagcctgcagctgtctgtgtgcctgcacgaggtggcctactggtaacatcctgagc
atcgggcccagaccgacttccctgtctgtgttcttctctggctacaccttcaagcacaagatgggtg
tacgaggacacctgacctgttccccctcagcggcgagaccctgttcatgagcatggagaacccc
ggcctgtggatcctgggctgccacaacagcgacttcagggaacaggggcatgaccgcccctgctgaaa
gtcagcagctgcgacaagaacaccggcgactactacgaggacagctaogaggacatcagcgctac
ctgctgagcaagaacaacggccatcgagcccaggagcttcagccagaacccccctgtgctgaagg
caccagagggagatcaccaggaccacctgacagagcagaccaggaggagatcagctatgatgacacc

(Continued)

Figure 26A

atcagcgtggagatgaagaaggaggacttcgacatctacgacgaggacgagaaccagagccccagg
agcttccagaagaagaccaygcaactacttcaatcgccyccgtggagaggctgtggygactatggcatg
agcagcagccccacgtgctgaggaaacagggcccagagcggcagcgtgccccagttcaagaagggtg
gtgttccaggagttcacogacggcagottcacccagccccctgtacagagggcagagctgaaocagagcac
ctgggocctgctgggccccctacatcagggccgagggtggaggacaacatcatggtgaccttcaggaac
caggccagcaggccctacagcttctacagcagocctgatcagctacyaggagggaaccagaggcagggc
gcccagcccaggaagaacttcgtgaagcccaacgagaccaagacctacttctggaagggtgcagcac
cacatggcccccccaaggacgagttcgaactgcaaggccctgggocctacttctctgatgtggacctg
gagaaggacgtgcaacagcggccctgatcggccccctgctgggtgtgcccacccaacacccctgaacccc
gcccacggcaggcagggtgacogtgcaggagttcgccccctgttcttcaacctcttcgacgagaccaag
agctggtaacttcaocgagaaacatggagaggaaactgcagggccccctgcaacatccagatggaggac
cccaccttcaaggagaactacaggttccaacgcatcaacggctacatcatggacacccctgcccggc
ctgggtgatggcccaggaccagaggatcaggtgggtatctgctgagcatgggcagcaacgagaacatc
cacagcatccacttcagcggccacogtgttcaocgtgaggaagaaggaggagtacaagatggccctg
tacaacotgtaccccggcgtgttcgagaccogtggagatgctgcccagcaaggccggcatctggagg
gtggagtgocctgatcggcagcaccctgcaacgcccggcatgagcaccctgttccctgggtgtacagcaac
aagtggcagacccccctgggcatggccagcggcccaacatcagggacttcagatcaaccgocctotggc
cagtaacggccagtgggcccccaagctggccaggctgcaactacagcggcagcatcaacgocctggagc
accaaggagcccttcagctygatcaagggtggaocctgctggcccccatgatcatcaacggcatcaag
accagggcggccaggccagaagttcagcagcctgtacatcagccagttcatcatcatgtacagcctg
gacggcaagaagtgccagacctaagggggcaacagcaaccggccacocctgatgggtgttcttoggcaac
gtggacagcagcggcatcaagcaacaacatcttcaacccccccatcatcgccagggtacatcaggctg
caccocccccactacagcatcaggagcaacocctgcggtatggaaactgatgggctgcaocctgaaacagc
tgacagcatgcccctggycatggagagcaaggccatctctgaacgcccagatcaaccgcccagcagctac
ttcaaccaacatgttcgcaocctggagccccagcaaggccaggctgcaocctgcaaggggcaggagcaac
gocctggaggccccaggtgaacaacccccaggagtggtgcaggtggacttccagaagaaccargaag
gtgacggcgtgacaccccaggggcgtgaagagocctgctgaccagcatgtacgtgaaggagttcctg
atcagcagcagccaggacggccaccagtggaocctgttcttccagaacgggcaagtggaaggtgttc
cagggcaaccaggacagcttcaCcccgtgggtgaaacagcctggacccccccctgctgacccaggtat
ctgaggatccacccccagagctgggtgaccagatggccctgagaatggaagtgctgggatgagag
gcccaggacctgtactga (SEQ ID NO:20)

Figure 26B

CS23-FL-AA

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDAR' FPPRVPKSFFPNTSVVYK
KTLFVEFTDHLFNIAPRPPWMLLGPFIQAEVYDVTVVITLKNMASHFVSLHAVGVSYWKASEGAEY
DDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCRE
GSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGC
HRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLMDLGGQFLLFCHISSH
QHDGMEA YVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSVAKKHPKTW
VHYIAAEEEDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHEGIL
GPLLYGEVGD TLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVT
VEDGPTKSDPRCLTRYYSFVNMERDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVILFSVFDENR
SWYL TENIQRF LNPAGVQLEDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYILSIGAQTDFLS
VFFSGYTFKHKMVEYEDTLTLFPFSGETVFMSENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGD
YYEDSYEDI SAYLLSKNNAIEPRSFQNPVFLKRHQREITR TLLQSDQEEIDYDDTISVEMKKE DFD
IYDE DENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVPFKKVVFQEF TDGSFTQ
PLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLLISYEEDQRQGAEPKRNFKPNET
KTYFWKVQHMMAPT KDEFDC KAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALF
FTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLS
MGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTL
FLVYSNKCQTPLG MASGHIRD FQITASGQYQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI
IHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIENPPIIAR
YIRLHPHYSIRSTLRMELMGC DLNSCSMPLGMESKAISDAQITASSYFTNM FATWSPSKARLHLQG
RSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWTLFFQNGKVK
VFQGNQDSFTPVVNSLDP LLLTRYLRIFQSWVHQIALRMEVLGCEAQDLY (SEQ ID NO:21)

Figure 27

CS23-HC-NA

```

                                                                    gcc
accaggagat actacctggg cgccgtggag ctgagctggg actaatgca gtctgacctg
ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac
acctcagtgg tgtacaagaa gacctgttcc gtggagtcca cagaccacct gttcaacatc
gccaaagcca ggccccctg gatgggcctg ctgggccccca ccatccaggc cgaggtgtac
gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg
ggcgtgagct actggaaggc ctctgagggc gccaggtatg acgaccagac cagccagagg
gagaaggagg acgacaaggt gttccccggc ggccagccaca cctacgtgtg gcaggtgctg
aaggagaacg gccccatggc cagcagcccc ctgtgcctga cctacagcta cctgagccac
gtggacctgg tgaaggacct gaactctggc ctgatcggcg cctgctggt gtgcagggag
ggcagcctgg ccaaggagaa gaccagacc ctgcacaagt tcatcctgct gttcgcctg
ttcgtatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat
gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc
ctgcccgcc tgatcggtg ccacaggaag tctgtgtact ggcactgat cggcatgggc
accacccccg aggtgcacag catcttctctg gagggccaca ccttctggt gaggaaccac
aggcaggcca gcctggagat cagccccatc accttctctga ccgcccagac cctgctgatg
gacctgggccc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag
gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag
gagggcaggg actatgatga tgacctgacc gactctgaga tggactggt gaggtttgat
gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca cccaagacc
tgggtgcact acatgccccg cgaggaggag gactgggact aogccccct ggtgctggcc
cccagcagca ggagctacaa gagccagtac ctgaacaacg gccccagag gatcggcagg
aagtacaaga aggtcagatt catggcctac accgacgaga ccttaagac cagggaggcc
atccagcaag agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg
ctgatcatct tcaagaacca ggccagcagg cctacaaca tctaccccc cggcatcacc
gatgtgaggg cctgttacag caggaggctg cccaagggcg tgaagcacct gaaggacttc
ccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc
accaagtctg accccaggtg cctgaccagg tactacagca gcttctgtaa catggagagg
gacctggcct ctggcctgat cggccccctg ctgatctgct acaaggagag cgtggaccag
aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgtatgag
aacaggagct ggtatctgac cgagaacatc cagaggttcc tgccccaccc cggcggcgtg
cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg
ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc
atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag
atggtgtacg aggacaccct gacctgttc cccttcagcg gcgagaccgt gttcatgagc
atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc
atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac
agctacgagg acatcagcgc ctacctgctg agcaagaaca acgccatcga gccccagg
(SEQ ID NO:22)

```

Figure 28

CS23-LC-NA

```

                                g agatcaccag gaccaccctg
cagagcgacc aggaggagat cgactatgat gacaccatca gcggtggagat gaagaaggag
gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc
aggcaactact tcatcgccgc cgtggagagg ctgtgggact atggcatgag cagcagcccc
caogtgotga ggaacagggc ccagagcggc agcgtgcccc agttcaagaa ggtgggtgttc
caggagttca ccgacggcag ctcaaccag cccctgtaca gaggcgagct gaacgagcac
ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc
aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac
cagaggcagg gcgcccagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac
ttotggaagg tgcagcacca catggccccc accaaggacg agttogactg caaggcctgg
gootactttct ctgatgtgga cctggagaag gacgtgcaca gggcctgat cggccccctg
ctgggtgtgcc acaccaaac cctgaacccc gccacaggca ggcaggtgac cgtgcaggag
ttcgccctgt tcttcacct cttcgacgag accaagagct ggtacttcac cgagaacatg
gagaggaaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac
tacaggttcc acgccatcaa cggctacatc atggaacccc tgcccggcct ggtgatggcc
caggaccaga ggtcaggtg gtatctgtct agcatgggca gcaacgagaa catccacagc
atccacttca gcggccacgt gttcaccgtg aggaagaagg aggagtaca gatggccctg
tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggcgggctc
tgagggggtg agtgccctgat cggcgagcac ctgcacggcg gcatgagcac cctggttctg
gtgtacagca acaagtgcca gacccccctg ggcattggca cgggcccacat cagggacttc
cagatcaccg cctctggcca gtaaggccag tggccccca agctggccag gctgcactac
agcggcagca tcaacgcctg gagaccaag gagcccttca gctggatcaa ggtggacctg
ctggccccc tgatcatcca cggcatcaag acccagggcg ccaggcagaa gttcagcagc
ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac
aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcctc
aagcacaaca tcttcaaccc ccccatcatc gccaggtaca tcaggctgca ccccaccac
tacagcatca gtagcaccct gcggatggaa ctgatgggct ggcacctgaa cagctgcagc
atgcccctgg gcatggagag caaggccatc tctgacgccc agatcacccg cagcagctac
ttaccaaca tgttcgccac ctggagcccc agcaaggcca gctgacacct gcagggcagg
agcaacgcct ggaggcccca ggtgaacaac cccaaggagt gctgacaggt ggacttccag
aagaccatga aggtgaccgg cgtgaccacc cagggcgtga agagcctgct gaccagcatg
taogtgaagg agttcctgat cagcagcagc caggacggcc accagtggac cctggttctc
cagaacggca aagtgaaggt gttccagggc aaccaggaca gottcaccac cgtgggtgaa
agcctggacc cccccctgct gaccaggtat ctgaggatcc accccagag ctgggtgcaac
cagatcgccc tgagaatgga agtgcctggga tgcgaggccc aggacctgta c
(SEQ ID NO:23)

```

Figure 29

CS01m13-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACCTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAAATCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCTCGGATG
GGACTCCTGGGACCCACCATTCAAGGCTGAGGTGTATGACACTGTGGTCATCACCTTCAAGAACATG
GCATCCCACCCCTGTGTCTCTGCAFGCTGTGGGAGTCTCATACTGGAAAGCCTCTGAAAGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAAGAGGATGACAAGGTGTCCCTGGGGGATCTCACACC
TATGTGTGGCAAGTCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCFGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCTGTTTGTCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAAGTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCAFGGCCCAAGATGCACACTGTGAATGGCTATGFGAACAGATCACTGCCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCTTGGAGGGACACACCTTCCTGGTCAAGAACACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCTCACFGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACCTCTCCATCCTTCAATCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACCTGGTC
CTGGCCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCCAAAGAAATFGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCATTCTGCCFGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACCTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGCAACCCAGATCATGTCTGACAAGAGAAAATGTGATTTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACAATTCAGAGATTCCTGCCAAC
CCTGCTGGGGTGCAACTGGAAGACCCTGAGTTCAGGCAAGCAACATCATGCACCTCCATCAATGGC
TATGTGTTTGGACTCTCCAGCTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTTCT
ATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAAGATGGTG
TATGAGGACACCCFGACACTCTTCCCATTCTCTGGGGAAACTGFGTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCFGGGATGCCACAACCTGACTTCAGAAACAGGGGAATGACTGCACCTGCTCAAA
GTCTCCTCCTGTGACAAGAACAATACCACTACGTGAACCGCTCCCTGTCTCAGAATCCACCTGTCTGAAG
AGACACCAGAGAGAGATCACCCAGGACAACCTCCAGTCTGACCAGGAAGAGATFGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCA
AGATCATTCAGAAGAAGACAAGACACTACTTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGC
ATGTCTTCTCTCCCATGTCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(Continued)

Figure 30A

GTGGTCTTCCAGGAGTTCACTGATGGCTCATTACCCAGCCCCGTACAGAGGGGAACTGAATGAG
CACCTGGGACTCCTGGGACCATAACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGA
AACCAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAA
GGGGCTGAGCCAAGAAAGAACTTTGTGAAACCCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCACCACCAAGGATGAGTTTGAAGTGCAGGGCTGGGCATACTTCTCTGATGTGGAC
CTGGAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCACATGGAAGGCAAGTGAAGTGTGCAGGAGTTTGGCCCTCTTCTTCAACATCTTTGATGAAACC
AAGTCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAA
GACCCACCTTCAAGGAGAACTACAGGTTCATGCCATCAATGGCTACATCATGGACACCCCTGCCT
GGGCTTGTCTATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCATGGGATCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCCTCTGG
AGGGTGGAAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCACCCCTGTTCCTGGTCTACAGC
AACAAAGTGCCAGACACCCCTGGGAATGGCCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCCATGG
TCAACCAAGGAGCCATTTCTTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATC
AAGACACAGGGGGCAAGACAGAAATCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCT
CTGGATGGCAAGAAGTGGCAGACATAAGAGGCAACTCCACTGGCACCCCTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCATCATGGCCAGATACATCAGG
CTGCACCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACCTGAAC
TCCTGCTCAATGCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCT
TACTTACCAACATGTTTGGCACCTGGTACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGC
AATGCCTGGAGACCCCAAGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATG
AAAGTCACTGGGGTGAACAACCCAGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTC
CTGATCTCTTCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTG
TTCCAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACCTCACTGGACCCCCCTCCTGACAAGA
TACCTGAGAATTCACCCCAAGTCTTGGGTCCACCAGATFGCCCTGAGAATGGAAGTCTGGGATGT
GAGGCACAAGACCTGTACTGA (SEQ ID NO:90)

Figure 30B

CS01m23-FL-NA

ATGCAGATFGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTGCCCACCAGG
AGATACTACCFGGGGGCTGTGGAACCTTTCTTTGGGACTACATGCAGTCTGACCFGGGAGAGCTGCGCT
GTGGATGCCAGGTTCCCAACCCAGAGTGCCCAAGTCTTTCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAAATTCAGTACCACCTGTTCAACATTTGCAAAACCCAGACCACCTTGGATG
GGACTCCTGGGACCCACCATTTCAGGCTGAGGTGTATGACACTGTGGTCTGTCACCCCAAGAACATG
GCATCCCACCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAATCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGAAGTCTCACACC
TATGTGTGGCAAGTCCCAAGGAGAAATGGACCCACTGCATCTGACCCACCTTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTTCTCTGPTTGGCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCTGGAGGGACACACCTTCCCTGGTCAGGAACCCAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCAC
ATCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTGAGATCT
GATGTGGTCAGATTTGATGATGACAACCTCTCCATCCTTCATTGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCCTGGTCT
CTGGCCCCGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCAAAAGAAATFGGAAGA
AAGTACAAGAAAGTCAAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAAGTGGGAGACACCCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATCTGCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAGTCTGACCCAGGTGCCCTCACCCAGA
TACTACTCCTCTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGTGCACCTGGAAGACCTGAGTTCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTCT
ATTGGGGCACAACCTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCCTGACACTCTTCCATTCTCTGGGGAAACTGTGTTTATGAGCATGGAGAACCCCT
GGACTGTGGATTCCTGGGATGCCACAACCTCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAA
GTCTCCTCTGTGACAAGAACAATACCACCTACGTGAACCGCTCCCTGTCTCAGAATCCACCTGTCTTGAAG
AGACACCAGAGAGAGATCACAGGACAACCCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCA
AGATCATTCAGAAGAAGACAAGACACTACTTCAATTGCTGCTGTGGAAAGACTGTGGGACTATGGC
ATGTCTTCTCTCCCCATGCTCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(Continued)

Figure 31A

GTGGTCTTCCAGGAGTTCACCTGATGGCTCAFTTCACCCAGCCCCTGTACAGAGGGGAACCTGAATGAG
CACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGA
AACCAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAA
GGGGCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAAGTCCAG
CACCACATGGCACCACCAAGGATGAGTTTACTGCAAGGCCTGGGCATACTTCTCTGATGTGGAC
CTGGAGAAAGATGTGCACTCTGGCCCTGATTTGGCCCACTCCTGGTCTGCCACACCAACACCCCTGAAC
CCTGCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGGCCCTCTTCTTACCATCTTTGATGAAAACC
AAGTCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAA
GACCCACCTTCAAGGAGAACTACAGGTCCATGCCATCAATGGCTACATCATGGACACCCCTGCCT
GGGCTTGTTCATGGCACAGGACCAGAGAAFCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCC
CTGTACAACCTCTACCCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTTCTGGTCTACAGC
AACAAAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGG
TCAACCAAGGAGCCATTTCTCTTGGATCAAGGTGGACCTGCTGGCACCCTATGATCATTCATGGCATC
AAGACACAGGGGGCAAGACAGAAATTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCT
CTGGATGGCAAGAAGTGGCAGACATAACAGAGGCAACTCCACTGGCACCCCTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAAACATCTTCAACCCTCCCATCATTTGCCAGATACATCAGG
CTGCACCCCAACCCACTACTCAATCAGATCAACCCTCAGGATGGAACCTGATGGGATGTGACCTGAAC
TCCTGCTCAATGCCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCAGATCACTGCATCCTCT
TACTTACCAACATGTTTGGCCACCTGGTCACCATCAAAGCCAGGCTGCACCTCCAGGGAAGAAGC
AATGCCTGGAGACCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATG
AAAGTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTC
CTGATCTCTTCCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTG
TTCCAGGGCAACCAGGACTCTPTTACACCTGTGGTGAACCTCACTGGACCCCCCTCCTGACAAGA
TACCTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCTTGGGATGT
GAGGCACAAGACCTGTACTGA (SEQ ID NO: 91)

Figure 31B

CS01m3--FL--NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACPTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTCGAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTCAAGGCTGAGGPGTATGACACTGTGGTCATCACCCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAAGCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACC
TATGTGTGGCAAGTCCTCAAGGAGAAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGTCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAGAAGTCCCTGATGCAAGACAGGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTPGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCTGGAGGGACACACCTTCCCTGGTCAGGAACCAAGCAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACCTTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTC
CTGGCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCAAAAGAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCCTCCTGTATGGGGAAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAGTCTGACCCCAAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCAAC
CCTGCTGGGGTGCACCTGGAAGACCCCTGAGTTCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTACTCTCTCCAGCTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTCT
ATTTGGGGCACAACTGACTTCCCTTCTGTCTTCTTCTTCTGGATACACCTTCAAGCACAAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTTCTCTGGGGAAACTGTGTTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAA
GTCTCCTCCTGTGACAAGAACAATACCACCTACGTGAACCGCTCCCTGTCTCAGAATCCACCTGTCTGAAG
AGACACCAGAGAGAGATCACCAGGACAACCCCTCAGTCTGACCAGGAAGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAGGAGGACTTTCAGATCTATGATGAGGACGAGAACCAGTCTCCA
AGATCATTCAGAGAAGAAGACAAGACACTACTTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGC
ATGTCTTCCCTCTCCCATGTCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(Continued)

Figure 32A

GTGGTCTTCCAGGAGTTCACCTGATGGCTCATTACCCAGCCCCTGTACAGAGGGGAAGTGAATGAG
CACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGA
AACCAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAA
GGGGCTGAGCCAAGAAAGAAGCTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCACCACCAAGGATGAGTTTACTGCAAGGCCCTGGGCATACTTCTCTGATGTGGAC
CTGGAGAAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGGCCCTCTTCTTACCATCTTTGATGAAACC
AAGTCATGGTACTTCACTGAGAACATGGAGAGAACTGCAGAGCACCATGCAACATTCAGATGGAA
GACCCACCTTCAAGGAGAAGTACAGGTTCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTATGGCACAGGACCAGAGAAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAAATGCCTCATTTGGGGAGCACCTGCATGCTGGCACAATCAACCTGTTCTTGGTCTACAGC
AACAAAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGG
TCAACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCTATGATCATTCATGGCATC
AAGACACAGGGGGCAAGACAGAAATTTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCT
CTGGATGGCAAGAAGTGGCAGACATAACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATGCCAGATACATCAGG
CTGCACCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAAGTGTGGGATGTGACCTGAAC
TCCTGCTCAATGCCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCAGATCACTGCATCCTCT
TACTTCACCAACATGTTTGGCACCTGGTCAACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGC
AATGCCTGGAGACCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATG
AAAGTCACTGGGGTGACAACCCAGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTC
CTGATCTCTTCTTCCAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTG
TTCCAGGGCAACCAGGACTCTTTTACACCTGTGGTGAAGTCACTGGACCCCCCTCCTGACAAGA
TACCTGAGAATTCACCCCACTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCTGGGATGT
GAGGCACAAGACCTGTACTGA (SEQ ID NO:92)

Figure 32B

CS01m2-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTCTGTGTGCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACTTTCCTFGGGACTACATGCAGTCTGACCTGGGAGAGCTGCOCT
GTGGATGCCAGGTTCCUACCCAGAGTGCCTCAAGTCTTCCCAATFCAACACCTCTGTGGTCTACAAAG
AAGACACTCTTTTGTGGAAATTCACCTGACCACCTGTTCACATTCGAAAACCCAGACCACCCCTGGATG
GGACTCCCTGGGACCCACCAATTCAGGCTGAGGGTGTATGACACTGTGGTCTGTCACCCFCAAGAACAATG
GCATCCACCCCTGTCTCTGCAATGCTGTGGGAGTCTCATACTGGAAATCCCTCTGAAGGGGCTGAG
TATGATGACUAGACATCCAGAGAGAGAAAGAGUAFGACAAGGTFPTCCCTGGGAAGTCTCACACC
TATGTGTGGCAAGTCTCAAGGAGAAATGGACCCACTGCATCTGACCCACCCCTGCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATFGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCCTCTGTTTGTCTGTC
TTTGTATGAGGGCAAGTCTFGGCACTCTGAAAACAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAAATGGGCTATGTFGAACAGATCACTGCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATFGGCATGGGGACAACCCCTGAAGTGCAC
TCCAATTTCTGGAGGGACACACCTTCCCTGGTCAAGAACACAGACAAGCCCTCTCTGGAGATCTCT
CCCATCACCTTCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCAATGATGGCAATGGAAGCCATATGTCAGGTFGGACTCAATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTFGGTCAATTTGATGATGACAACCTCTCCATCCCTTCATTCAGATCAGGCTCTGTFGGCAAAGAAA
CACCCCAAGACATFGGGTGCACACTACATGCTGCTGAGGAAGAGGACTGGGACTATGCACCACCTGGTFC
CTGGCCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCCACAAGAATFPGAAGA
AAGTACAAGAAAGTCAATTCATGGCCCTACACTGATGAAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCAATCTGGGACCACTCCCTGTFATGGGGAAATGGGAGACACCCCTGCTCATCATCTTC
AAGAACCAGGCCCTCCAGGCCCTACAACATCTACCCACATFGGCATCACTGATGTCAGGCCCTCTGAC
AGCAGGAGACTFGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCCTGCTGGAGAGATCTTC
AAGTACAAGTFGGACTGTCACCTGTGGAGGATGGACCAACAAGTCTGACCCCAAGGTGCCCTCACCAGA
TACTACTCTCTTTTGTGAACAATGGAGAGAGACCTGGCATCTGGACTGATFSSACCACCTGCTCATFC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGTGAACFPGAAGACCCCTGAGTTCAGGCAAGCAACATCATGCACCTCCATCAATGGC
TATGTGTTTACTCTCTCCAGCTTTCTGTCTGCCCTGCATGAGGTGGCCCTACTGGTACATTCCTTCT
ATTGGGGCACAAACTGACTTCCCTTCTGTCTCTCTCTCTGATAACCTTCAAGCACAAAGATGGTGTG
TATGAGGACACCCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTTATGAGCATGGAGAACCCCT
GGACTGTGGATTTCTGGGATGCCACAACCTGACTTFCAGAAACAGGGGAATGACTGCACCTGCTCAAA
GTCTCTCTCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCCTAC
CTGCTCAGCAAGAACAATGCCATFAGAGCCCAAGACTTCTCTCAGAAATCCACCTGFCCTGAAGAGA
CACCAGAGAGAGATCACCAGGACAACCCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAGGAGGACTTTGACAATCATGATGAGGACGAGAACCAGTCTCCAAGA
TCATTCAGAGAAGAAGACAAGACACTACTTCATTTGCTGCTGTGGAAAGACTGTGGGACTATGGCATG
TCTTCTCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAGAAGAAATG

(Continued)

Figure 33A

GTCTTCCAGGAGTTCACTGATGGCTCATTACCCAGCCCTGTACAGAGGGGAACTGAATGAGCAC
CTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGAAAC
CAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGG
GCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGACCTG
GAGAAAGATGTGCACTCTGGCCTGATTTGGCCCACTCCTGGTCTGCCACACCAACACCCCTGAACCT
GCACATGGAAGGCAAGTGACTGTGCAGGAGTTGGCCCTCTTCTTACCATCTTTGATGAAAACCAAG
TCATGGTACTTCACTGAGAACATGGAGAGAACTGCAGAGCACCATGCAACATTCAGATGGAAGAC
CCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCCTGCCCTGGG
CTTGTCTATGGCACAGGACCAGAGAAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAATGCCTCATTTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCTGGTCTACAGCAAC
AAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCCTGGC
CAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCA
ACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTTCATGGCATCAAG
ACACAGGGGGCAAGACAGAAATTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTG
GATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCCTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATAACATCAGGCTG
CACCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAAGTGTGGGATGTGACCTGAACTCC
TGCTCAATGCCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCAGATCACTGCATCTCTTAC
TTCACCAACATGTTTGGCACCTGGTCAACATCAAAGCCAGGCTGCACCTCCAGGGAAGAAGCAAT
GCCTGGAGACCCCAAGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAAATGAAA
GTCCTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTCCTG
ATCTCTTCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTGTT
CAGGGCAACCAGGACTCTTTCACACCTGTGGTGAAGTCACTGGACCCCCCTCTGACAAGATAC
CTGAGAAATTCACCCCAAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCTTGGGATGTGAG
GCACAAGACCTGTACTGA (SEQ ID NO:93)

Figure 33B

CS04m2-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCTGCTGCTGAGGTTCTGCTTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCACCCAGAGTGCCCAAATCCTTCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCCTCTTFTGTGGAGTTCACCTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCCTGGATG
GGACTCCTGGGACCCACCATTGAGGCTGAGGTTGTATGACACTGTGGTCTGTCACCCCTCAAGAACAATG
GCCTCCACCCCTGTGAGCCTGCATGCTGTGGGGGTCAGCTACTGGAAGTCCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCTGGGAAGAGCCACACC
TATGTGTGGCAGGTTCTCAAGGAGAATGGCCCCACTGCTTCTGACCCACCCGCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGCTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCCTGCACAAGTTCATTCTCCTGTTTGTCTGT
TTTGTATGAGGGCAAGAGCTGGCACCTCTGAAAACCAAGAATCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCCTGGCCCAAGATGCACACTGTGAAATGGCTATGTGAACAGGAGCCTGCCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCTTGGAGGGCCACACCTTCTTGGTCTAGGAACACAGACAGGCCAGCCCTGGAGATCAGC
CCCATCACCTTCTTACTGCCCAGACCCCTGCTGATGGACCTCGGACAGTTCCTGCTGTTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGSAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCAITTCAGATCAGGTCFTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACACATTTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTCT
CTGGCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCACAGAGGATTTGGACGC
AAGTACAAGAAAAGTCAAGTTCATGGCTTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGACTCTGGCATCCTGGGGCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAACACCTCAAGGACTTCCCCATTTCTGCCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAATCTGACCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCCTGTGGCCCTGATTTGGCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAAGATCATGCTGACAAGAGGAATGTGATCTG
TTCTCTGCTTTTGTGATGAGAACAGGAGCTGGTACCTGACTGAGAACAATTCAGCGCTTCTGCCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCCTGAGTTCAGGCCAGCAACATCATGCACCTCCATCAATGGC
TATGTGTTTTGACAGCCCTCCAGCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
ATTGGGGCCAGACTGACTTCCCTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCCTGACCTCTTCCATTCTCTGGGGAGACTGTGTTTCTGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACCTCTGACTTCCGCAACAGGGGGCATGACTGCCCTGCTCAA
GTCTCTCTCTGTGACAAGAACAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCCTAC
CTGCTCAGCAAGAACAATGCCATTTGAGCCAGGAGCTTCCAGCCAGAAATCCACCTGTCTCTGAAACGC
CACCAGAGGGAGATCACCAGGACCACCCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCAAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCAATTGCTGCTGTGGAGCGCCTGTGGGACTATGGCATG
AGCTCCAGCCCCCATGTCTCAGGAACAGGGCCAGTCTGGCTCTGTGCCACAGTTCAAGAAAAGTG

(Continued)

Figure 34A

GTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAGCAC
CTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTFCCGCAAC
CAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGG
GCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCCCCCACCAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTG
GAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCTT
GCCCATGGAAGGCAAGTGACTGTGCAAGGAGTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAG
AGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAGGAC
CCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGG
CTTGTCATGGCCCAGGACCAGAGGATCAGTGGTACCTGCTTTCTATGGGCTCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCTGTTCCCTGGTCTACAGCAAC
AAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGGAGC
ACCAAGGAGCCATTCACTGGATCAAAGTGGAOCTGCTGGCCCCATGATCATCCATGGCATCAAG
ACCCAGGGGGCCAGGCAGAAAGTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAAACTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTG
CACCCACCCACTACAGCATCCGCAGCACCTCAGGATGGAGCTGATGGGCTGTGACCTGAACCTCC
TGCAGCATGCCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCCTCCAGCTAC
TTCACCAACATGTTTTGCCACCFTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGCAAT
GCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTCCTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTCCTG
ATCAGCTCCAGCCAGGATGGCCACCAGTGGACCTCTTCTTCCAGAATGGCAAGGTCAAGGTGTTT
CAGGGCAACCAGGACAGCTTACCCCTGTGGTGAACAGCCTGGACCCCCCCCCTCCTGACCAGATAC
CTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCTGGGATGTGAG
GCCAGGACCTGTACTGA (SEQ ID NO:94)

Figure 34B

CS04m3--FL-NA

ATGCAGATTGAGCTFGAGCACTFCGCTTCTTCCFTGTGCTFCGCTGAGGTTCTGCTTCTCTGCCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCOCT
GFGGATGCCAGGTTCCCACCCAGAGTGGCCAAAFCCTTCCCAFFCAACAUCCTCTGTGGFCTACAAG
AAGACCCCTCTTTGTGGAGTTCACTGACCACCTGTTC AACATTGCCAAAACCAGGCCACCCCTGGATG
GGACTCCFTGGGACCCACCATTFCAGGCTGAGGTTGATGACACTGTGGFCATCACCCCTCAAGAACA TG
GCCTCCCACCCCTGTGAGCCTGCATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAAGTTCCTCCFTGGGGGACGCCACACC
TATGTGTGGCAGGTTCTCAAGGAGAAATGGCCCCATGGCCTCTGACCCACTCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGFCAAGGACCTCAACTCTFGGACTGATFGGGGGCCCTGCTGGTGTGFC
AGGGAGGGCTCCCFTGGCCAAAGAGAAGACCCAGACCCFTGCACAAGTTCATTCTCCTGTTTGTGTCT
TTTGTGATGAGGGCAAGAGCTFGGCACTCTFGAAAACAAGAACTCCCTFGATGCAGGACAGGGATGCTGCC
CTCGCCAGGGCCCTGGCCCAAGATGCACACTGTGAAATGGCTATGTGAAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATFGGCATGGGGACAACCCCTGAGGTTGCAC
TCCATTTTCTGGAGGGCCACAACCTTCTGGTCAAGAACACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCTCTACTFGCCAGACCTTGTGATFGGACCTCGGACAGTTCCTGCTGTCTCTGCCAC
ATCAGCTCCCACCAGCATGATGSCATGGAGGCCATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCCTTTGATGATGACAACAGCCCATCTTCATTGAGATCAGGCTCTGTGGCCAAGAAA
CACCCCAAGACTGGGTGCCTACTATFTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTU
CTGGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCACAGAGGATFGGACGC
AAGTACAAGAAAGTCAAGGTTCAFGGCCCTACACTGATGAAACCTTCAAGACCAAGGGAGGCCATTCAG
CATGAGTCTGGCATCTGGGCCACTCTCTGTATGGGGAGGTGGGGGACACCCCTGCTCATCATCTTC
AGAACCAGGGCTCCAGGCCCTACAACAFTCAUCCACATGGCAFCACTGATGTCAGGCCCTGTAC
AGCCGCAGGGCTGCCAAAGGGGGTGAACACCTCAAGGACTTCCCATTCTGCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCTACTFGGAGGATFGACCAACCAAATCTGACCCCAAGGFGCCTCAACCAGA
TACTACTCCAGCTTTGTGAACAATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGFTGACAAGAGGAATGFGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCFTGCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCCTGAGTTCAGAGCCAGCAACAFCATGCACCTCCATCAAFFGGC
TATGTGTTTTGACAGCCTCCAGCTTTCTGTCTGCTGCAATGAGGTGGCCCTACTGGTACATTCTTTCT
ATFGGGGGCCAGACTGACTTCCCTTTCTGFTCTTCTTCTCTGCTTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCCTGACCCCTCTTCCCAFTCTCTGGGGAGACTGTGTTCAFGAGCATGGAGAACCCT
GGCCTGTGGATTCTFGGGATGCCACAACCTCTGACTTCCGCAACAGGGGCATGACTFGCCCTGCTCAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATFACCACCTAGGTTGAACCGCTCCCTGAGCCAGAAFFCCACCTGFTCTGAAA
CGCCACCAGAGGGAGATCACCAGGACCCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAAGAGGACTTTTGACATCTATGACGAGGACGAGAACCAGAGCCCA
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTTCATTGCTGTGGAGCGCCTGTGGGACTATFGGC
ATGAGCTCCAGCCCCCAFTGCTCAGGAACAGGGCCAGTCTFGGCTCTGTGCCACAGTTC AAGAAA

(Continued)

Figure 35A

GTGGTCTTCCAAGAGTTCACCTGATGGCAGCTTCACCCAGCCCTGTACAGAGGGGAGCTGAATGAG
CACCTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGAACCTTCCGC
AACCAGGCCCTCCAGGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAG
GGGGCTGAGCCACGCAAGAACTTTGTGAAACCCAAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCCCCACCAAGGATGAGTTTGACTGCAAGGCTGGGCCCTACTTCTCTGATGTGGAC
CTGGAGAAGGATGTGCACCTGGCCCTGATTTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGGCCCTCTTCTTCCACCATCTTTGATGAAACC
AAGAGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCCAGATGGAG
GACCCACCTTTCAAAGAGAACFACCGCTTCCAATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTCAATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAAC
ATTCACCTCCATCCACTTCTCTGGGCATGTCCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTCTACCCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAGTGCCTCAATGGGGAGCACCTGCATGCTGGCATGAGCACCCCTGTTCCCTGGTCTACAGC
AACAAAGTGCCAGACCCCCCTGGGAATGGCCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCCACTACTCTGGATCCATCAATGCCTGG
AGCACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATC
AAGACCCAGGGGGCCAGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAAACATCTTCAACCCCCCAATCATCGCCAGATAACATCAGG
CTGCACCCCCAUCCTACAGCATCCCGCAGCACCTCAGGATGGAGCTGATGGGCTGTGACCTGAAC
TCTTGCAGCATGCCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCAGATCACTGCCTCCAGC
TACTTCACCAACATGTTTGGCCACCTGGAGCCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGC
AATGCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTC
CTGATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTG
TTCCAGGGCAACCAGGACAGCTTCAACCCCTGTGGTGAACAGCCTGGACCCCCCTCTGACCAGA
TACCTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCTGGGATGT
GAGGCCAGGACCTGTACTGA (SEQ ID NO:95)

Figure 35B

CS04m23-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCAACCCAGAGTGGCCAAATCCTTCCCATTC AACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCCTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTTCAGGCTGAGGTGTATGACACTGTGGTCTGACCCCTCAAGAACATG
GCCTCCACCCCTGTGAGCCTGCATGCTGTGGGGTTCAGCTACTGGAAGTCCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAAGTGTTCCTGGGAAGAGCCACACC
TATGTGTGGCAGGTCCCTCAAGGAGAATGGCCCCACTGCTCTGACCCACCCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCTGACAAAAGTTCAATTCCTGTTTGTCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCTGGAGGGCCACACCTTCTGGTCAAGAACACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCTCACTGCCCAGACCCCTGCTGATGGACCTCGGACAGTTCCTGCTGTTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCATATGTC AAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCCACAGAGGATTGGACGC
AAGTACAAGAAAGTCAGGTTTCATGGCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTTCAG
CATGAGTCTGGCATCCTGGGCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACATGATGTCAGGCCCTGTAC
AGCCCGAGGCTGCCAAAGGGGGTGAACACCTCAAGGACTTCCCCATTCGCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCCTGGCCCTGATTGGCCCCTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAAACCAGATCATGTCGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCTGCCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTCTGTCTGCCTGCATGAGGTGGCCCTACTGGTACATTTCTTTCT
ATTGGGGCCAGACTGACTTCCCTTTCTGTCTTCTTCTGCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTTCTCTGGGGAGACTGTGTTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACAACCTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATAACACCTACGTGAACCGCTCCCTGAGCCAGAATCCACCTGTCTGAAA
CGCCACCAGAGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCA
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATTTGCTGCTGTGGAGCGCCTGTGGGACTATGGC
ATGAGCTCCAGCCCCCATGTCTCAGGAACAGGGCCCACTGCTGCTGTGCCACAGTTCAAGAAA

(Continued)

Figure 36A

GTGGTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAG
CACCTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGC
AACCAGGCCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAG
GGGGCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCCCCACCAAGGATGAGTTTACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGAC
CTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGGCCCTCTTCTTCACCATCTTTGATGAAACC
AAGAGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAG
GACCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAGTGCCTCAATGGGGAGCACCTGCATGCTGGCATGAGCACCCCTGTTCTGGTCTACAGC
AACAAAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGG
AGCACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATC
AAGACCCAGGGGGCCAGGCAGAAGTTTCTCCAGCCTGTACATCAGCCAGTTTCAATCATCATGTACAGC
CTGGATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAAACATCTTCAACCCCCAATCATCGCCAGATACATCAGG
CTGCACCCCACTACTACAGCATCCGCAGCACCCCTCAGGATGGAGCTGATGGGCTGTGACCTGAAC
TCTTGCAGCATGCCCTGGGCATGGAGAGCAAGGCCATTTCTPGATGCCCAGATCACTGCCTCCAGC
TACTTCAACCAACATGTTTGGCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGC
AATGCCTGGAGGCCCCAGSTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTC
CTGATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTG
TTCCAGGGCAACCAGGACAGCTTCAACCCTGTGGTGAACAGCCTGGACCCCCCTCCTGACCAGA
TACCTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTTGCCCTGAGGATGGAGGTCTGGGATGT
GAGGCCCAGGACCTGTACTGA (SEQ ID NO:96)

Figure 36B

CS04m1-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCTGCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACC'TGGGGGAGCTGCCT
GTGGATGCCAGGTTCCACCCAGAGTGCACAAATCCTTCCCATFCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCAGTACACCTGTTCACATTCGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTTCAGGCTGAGGTGTATGACACTGTGGTTCATCACCCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCTGGGGGCAGCCACACC
TATGTGTGGCAGGTCCCTCAAGGAGAATGGCCCCATGGCCTCTGACCCACTCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCTGCACAAGTTCATTCTCCTGTTTGTCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAATCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCTGGAGGGCCACACCTTCTGGTCAAGAACACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCTCACTGCCAGACCTGCTGATGGACCTCGGACAGTTCCTGCTGTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCATCCTTCATTTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACC'TCAACAATGGCCACAGAGGATTTGGACGC
AAGTACAAGAAAGTCAGGTTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTTCAG
CATGAGTCTGGCATCCTGGGCCCCTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCCACAACATCTACCCACATGGCATCACTGATGTCAGGCCCTGTAC
AGCCCGCAGGCTGCCAAAGGGGGTGAACACCTCAAGGACTTCCCATTTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAATCTGACCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCGACAAGAGGAATGTGATTTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCTGCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTTCAGACCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTTCT
ATTGGGGCCAGACTGACTTCCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTTCTGGGATGCCACAACCTTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAA
GTCTCCTCCTGTGACAAGAACAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAATCCACCTGTCTGAAACGC
CACCAGAGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCAAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGCATG
AGCTCCAGCCCCATGTCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG

(Continued)

Figure 37A

GTCTTCCAAGAGTTTCACTGATGGCAGCTTTCACCCAGCCCCCTGTACAGAGGGGAGCTGAATGAGCAC
CTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGCAAC
CAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGG
GCTGAGCCACGCCAAGAACTTTGTGAAACCCCAATGAAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCCCCCACCAGGATGAGTTTGACTTGCAAGGCTGGGCCACTTCTCTGATGTGGACCTG
GAGAAGGATCTGCACTCTGGCCTGATTTGGCCCACTCCTGGTCTGCCACACCAACACCCCTGAACCCCT
GCCCATGGAAGGCAAGTACTGTGCAGGAGTTTGGCCCTCTTCTTACCATCTTTGATGAAACCAAG
AGCTGGTACTTCACTGAGAACAATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAGGAC
CCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCCTGCCTGGG
CTTGTTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTTCTATGGGCTCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCCAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACTTCACTTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAGTGCCCTCATTTGGGGAGCACCTGCATGCTGGCATGAGCACCCCTGTTCCCTGGTCTACAGCAAC
AAGTGGCAGACCCCCCTGGGAATGGCTCTGCCCACATCAGGGACTTCCAGATCACTGCCCTCTGGC
CAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGGCCGGAGC
ACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAG
ACCCAGGGGGCCAGGCAGAAATTTCTCCAGCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACTCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTG
CACCCACCCACTACAGCATCCGCAGCACCCCTCAGGATGGAGCTGATGGGCTGTGACCTGAACCTC
TGCAGCATGCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCAGATCACFGCCCTCCAGCTAC
TTCACCAACATGTTTGGCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGCAAT
GCCTGGAGGCCCCAGGTCAAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTCACTGGGGTGGACCACCAGGGGGTCAAGAGCTGCTCACCAGCAFGTATGTGAAGGAGTTCCCTG
ATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCCTCTTCTCCAGAAATGGCAAGGTCAAGGTGTTT
CAGGGCAACCAGGACAGCTTACCCCTGTGGTGAACAGCCTGGACCCCCCTCCTGACCAGATAC
CTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTTGCCCTGAGGATGGAGGTCTGGGATGTGAG
GCCAGGACCTGTACTGA (SEQ ID NO:97)

Figure 37B

CS04m13-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCAUCCAGAGTGUCCAAATCCTTCCCAATTC AACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTTCAGGCTGAGGTTGATGACACTGTGGTTCATCACCCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCUAGAGGGAGAAGGAGGATGACAAAAGTGTCCCTGGGGGCAGCCACACC
TATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCATGGCCTCTGACCCACTCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCAATCTCCTGTTTGTGTGTC
TTTGTATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGUTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTTGCAC
TCCATTTTCTGGAGGGCCACACCTTCTGGTCAAGAACACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCTCTACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCTTGCTGTCTTGCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCAATTCAGATCAGGTTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACFACATTTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTFC
CTGGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCACAGAGGATTTGGACGC
AAGTACAAGAAAGTCAAGTTTCTGGCCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCAATTCAG
CATGAGTCTGGCACTCCTGGGCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTFACAACATCTACCCACATGGCATCACTGATGTGAGGCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAACACCTCAAGGACTTCCCCATTTCTGCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAAATCTGACCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTTGTGAACATGGAGAGGGACCTGGCCCTGGCCCTGATTGGCCCACTGCTCATC
TGCTACAAGGASTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCTGCUCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCCTGAGTTCCAGGCCAGCAACATCATGCCTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTTCT
ATTTGGGGCCAGACTGACTTCCTTTCTGTCTTTCTTCTCTGGCTACACCFTCAAACACAAGATGGTGT
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTTATGAGCATGGAGAACCCT
GGCCTGTGGATTTCTGGGATGCCACAACFTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACAATAACCACCTACGTGAACCGCTCCCCTGAGCCAGAATCCACCFTGCTCTGAAA
CGCCACCAGAGGGAGATCAACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCA
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTCAATGCTGCTGTGGAGCGCCTGTGGGACTATGGC
ATGAGCTCCAGCCCCATGTCTTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(Continued)

Figure 38A

GTGGTCTTCCAAGAGTTCACCTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAAATGAG
CACCTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGC
AACCAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAG
GGGGCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCCCCCACCAGGATGAGTTTGGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGAC
CTGGAGAAGGATGTGCACTCTGGCCTGATTTGGCCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGGCCCTCTTCTTACCATCTTTGATGAAACC
AAGAGCTGGTACTTTCACCTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAG
GACCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTTGTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTTCTATGGGCTCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCCTTACCCCTGGGGTCTTTTGGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAGTGCCTCAATGGGGAGCACCTGCATGCTGGCATGAGUACCCTGTTCCCTGGTCTACAGC
AACAAAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCASTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGG
AGCACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATC
AAGACUCAGGGGGUCAGGCAGAAAGTTCTCCAGCCTGTACATCAGUCAGTTCATCATCATGTACAGC
CTGGATGGCAAGAAAATGGCAGACCTACAGAGGCAACTCCACTGGAAACACTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACAFTTCAACUCCCCAATCATCGCCAGATACATCAGG
CTGCACCCCACCCACTACAGCATCCGCAGCACCCCTCAGGATGGAGCTGATGGGCTGTGACCTGAAC
TCCTGCAGCATGCCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCTCCAGC
TACTTACCAACATGTTTGGCCACCTGGAGSCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGC
AATGCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAASACCATG
AAGGTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTC
CTGATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCCTCTTCTTCCAGAATGGCAAGGTCAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCTCCTGACCAGA
TACCTGAGGATTCACCCUCCAGAGCTGGGTCCACCAGATTTGCCCTGAGGATGGAGGTCTGGGATGT
GAGGCCCAGGACCTGTACTGA (SEQ ID NO:98)

Figure 38B

CS23m13-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCC
GTGGACGCCAGGTTCCCCCCCAGAGTGCCTCAAGAGCTTCCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCTGTTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGATCACCTGAAGAACATG
GCCAGCCACCCCGTGAACCTGCACGCGGTGGGCGTGAAGTACTGGAAGGCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGCGGCAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGCGAACCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTTATCCTGCTGTTCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGGCCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCAACCCCGAGGTGCAC
AGCATCTTCCCTGGAGGGCCACACCTTCCCTGGTGAAGAACACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCCTGACCGCCAGACCCTGCTGATGGACCTGGGCCAGTTCTTGCTGTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGAAGTTTGGATGATGACAACAGCCCCAGCTTTCATCCAGATCAGGTTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCCGGAGGAGGAGGACTGGGACTACGCCCCCTGGTG
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTGAGATTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGCGGAGGTGGGCGACACCCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCTTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCTGCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTTCGACAGCCTGCAGCTGTCTGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCAGACCAGCTTCCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCCTTCCAGCGGCGAGACCCTGTTTCATGAGCATGGAGAACCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCACAAGAACAACACCCTACGTGAACCGCTCCCTGAGCCAGAACCCCCCTGCTGAG
AGGCACCAGAGGGAGATCACAGGACCACCTGCAGAGCGACCAGGAGGAGATCGACTATGATGAC
ACCATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCC
AGGAGCTTCCAGAAGAAGACCAGGCCTACTTTCATCGCCCGCGTGGAGAGGCTGTGGGACTATGGC
ATGAGCAGCAGCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTCCCCAGTTCAAGAAG

(Continued)

Figure 39A

GTGGTGTTCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCTGTACAGAGGGCAGCTGAACGAG
CACCTGGGCTGCTFGGGCCCCACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGG
AACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAG
GGCGCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAG
CACCACATFGGCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCCTACTTCTCTGATGTGGAC
CTGGAGAAGGACGTGCACAGCGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCCTGAAC
CCCGCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCCCTGCTTCTTACCATCTTCGACGAGACC
AAGAGCTGGTACTTACCCGAGAACATGGAGAGGAAGTGCAGGGCCCCCTGCAACATCCAGATGGAG
GACCCACCTTCAAGGAGAATACAGGTTCACCGCCATCAACGGCTACATCATGGACACCCCTGCCC
GGCCTGGTGTATGGCCCAGGAACAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAAC
ATCCACAGCATCCACTTCAGCGGCCAGTGTTCACCGTGAGGAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCAGCAAGGCCGGCATCTGG
AGGGTGGAGTGCCTGATCGGGCAGCACCTGCACGCCCGCATGAGCACCCCTGTTCCCTGGTGTACAGC
AACAAAGTGCCAGACCCCCCTGGGCATGGCCAGCGCCACATCAGGGACTTCCAGATCACCGCCTCT
GGCCAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGG
AGCACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTFGGCCCCCATGATCATCCACGGCATC
AAGACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCCTGATGGTGTCTTCGGC
AACGTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCATCATCGCCAGGTACATCAGG
CTGCACCCCAACCACTACAGCATCAGGAGCACCTGCGGATGGAAGTGTGGGCTGCGACCTGAAC
AGCTGCAGCATGCCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGC
TACTTACCAACATGTTCCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGC
AACGCCTGSAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTC
CTGATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGGCAAAGTGAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCTGCTGACCAGG
TATCTGAGGATCCACCCCAAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGTGGGATGC
GAGGCCAGGACCTGTACTGA (SEQ ID NO: 99)

Figure 39B

CS23m3-FL-NA

ATGCAGATTGAGCTGASCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCT
GTGGACGGCAGGTTCCCCCCCAGAGTGCCTCAAGAGCTTCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCCTGTTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCCTGGATG
GGCTTGCTGGGCCCCACCAATCCAGGCCGAGGTGTACGACACCCGTGGTGAFCACCCCTGAAGAACAATG
GCCAGCCACCCCCGTGAGCCTGCACGCCGTGGGCGTGTAGCTACTGGAAGGCCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCGGGCGGCAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGCGACCCCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCCTGCACAAGTTCATCCTGCTGTTTCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCTTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGCCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCCGAGGTGCAC
AGCATCTTCTGGAGGGCCACACCTTCTGGTGGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCTGACCGCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGTTGAGGTTTGGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACCTACATCGCCGCGGAGGAGGAGGACTGGGACTACGCCCCCTGGTG
CTGGCCCCCGAUCAGCAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTGAGATTCATGGCTTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCTTGGGCCCTTGTGTACCGCGAGGTGGGCGACACCCCTGCTGATCATCTTC
AAGAACAGGCCAGCAGGCCCTTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCTGCCCCGCGAGATCTTC
AAGTACAAGTGGACCTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCTGCTGATC
TGCTACAAGSAGAGCCTGSACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TCTCTGTGTTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCTTGCCCAAC
CCCGCCGGGCTGCAGCTGGAGGACCCCGAGTTCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTTCGACAGCTGCAGCTGTCTGTGTGCTGCACGAGGTGGCCTACTGGTACATCTGAGC
ATCGGCGCCAGACCGACTTCTCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCCTGACCCCTGTTCCCTTTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACACCACCTACGTGAACCGCTCCCTGAGCCAGAACCCCCCGTGTGAAG
AGGCACCAGAGGGAGATCACCCAGGACCACCTGCAGAGCGACCAGGAGGAGATCGACTATGATGAC
ACCATCAGCGTGGAGATGAAGAAGGAGGACTTTCGACATCTACGACGAGGACGAGAACCAGAGCCCC
AGGAGCTTCCAGAAGAAGACCAGGCACCTTTCATCGCCGCGCGTGGAGAGGCTGTGGGACTATGGC
ATGAGCAGCAGCCCCACGTGCTGAGGAACAGGGCCAGAGCGGCAGCGTGGCCAGTTCAAGAAG

(Continued)

Figure 40A

GTGGTGTTCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCGTACAGAGGGCGAGCTGAACGAG
CACCTGGGCCCTGCTGGGGCCCCACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGG
AACCAGGCCAGCAGGCCUUTACAGCTTTCACAGCAGCCGTGATCAGCTACGAGGAGGACCAGAGGCAG
GGCGCCGAGCCCGAGGAAGAACTTCGTGAAGCCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAG
CACCACATGGCCCCCACCAGGACGAGTTCGACTGCAAGGCCGTTGGCCCTACTTCTCTGATGTGGAC
CTGGAGAAGGACCTGCACAGCGGCCCTGATCGGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAAC
CCCGCCCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCCGTGTTCTTCACCATCTTCGACGAGACC
AAGAGCTGGTFACTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAG
GACCCACCTTCAAGGAGAACACAGGTTCACCGCCATCAACGGCTACATCATGGACACCCTGCCC
GGCCFTGGTGTATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAAC
ATCCACAGCATCCACTTCAGCGGCCACGTGTTCACCGTGGAGGAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCCAGCAAGGCCGGCATCTGG
AGGGTGGAGTGCCGTGATCGGCGAGCACCTGCAACGCCGGCATGAGCACCCCTGTTCCCTGGTGTACAGC
AACAAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCCTCT
GGCCAGTACGGCCAGTGGGGCCCCAAGCTGGCCAGGCTGCCTACAGCGGCAGCATCAACGCCCTGG
AGCACCAAGGAGCCCCCTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATC
AAGACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCCGGCACCCCTGATGGTGTCTTCGGC
AACGTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCATCATCGCCAGGTACATCAGG
CTGCACCCCCACCCACTACAGCATCAGGAGCACCCCTGCGGATGGAACTGATGGGCTGCGACCTGAAC
AGCTGCAGCATGCCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGC
TACTTCACCAACATGTTCGCCACCCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGC
AACGCCCTGGAGGGCCCCAGGTGAACAACCCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTGACCGGCGTGAACACCCAGGGCGTGAAGAGCCCTGCTGACCAGCATGTACGTGAAGGAGTTC
CTGATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCCTGTTCTTCCAGAACGGCAAAGTGAAGGTG
TTCAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCCTGGACCCCCCCCCCTGCTGACCAGG
TATCTGAGGATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGTGGGATGC
GAGGCCAGGACCTGTACTGA (SEQ ID NO:100)

Figure 40B

CS23m2-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGCGCCGTGGAGCTGAGCTGGGACTACATGCACTGACCTGGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCCTGTTTCGTGGAGTTACCCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCCTGGATG
GGCCTGCTGGGGCCCCACCATCCAGGGCCGAGGTGTACGACACCCGTGGTGGTACCCCTGAAGAATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGGCGTGAGCTACTGGAAGTCCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGCAAGAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACCGGCCCACTGCCAGCGACCCCCCTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACCTCTGGCCGTGATCGGGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCTTGCAACAAGTTCATCCTGCTGTTCCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCCTGGCCCAAGATGCACACCGTGAACGGCTACCTGGAACAGGAGCCTGCCCGGCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCAFGGGCAACCCCGAGGTGCAC
AGCATCTTCTGGAGGGCCACACCTTCTGGTGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCTGACCGCCAGACCCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCACTGTAAGGTGGACAGCTGCCCGGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGGGTTTGTATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCCGCCAGGAGGAGGACTGGGACTACGCCCCCCCTGGT
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCAGAGGATCCGCCAGG
AAGTACAAGAAGGTGAGATTCATGGCTTACCCGACGAGACCTTCAAGACCAGGGGCCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGGACACCCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCCCTGTAC
AGCAGGAGGCTGCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCTCTGGCCTGATCGGCCCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTTCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCUAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTTCGACAGCCTGCAGCTGTCTGTGTGCCFTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCAGACCGACTTCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCTTTCAGCGCGGAGACCGTGTTCATGAGCATGGAGAACCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACGCCATCGAGCCCAGGAGCTTCAGCCAGAACCCCCCTGCTGAAGAGG
CACCAGAGGGAGATCACCAGGACCACCCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGACACC
ATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCCAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCATCGCCCGCGTGGAGAGGCTGTGGGACTATGGCATG
AGCAGCAGCCCCACGTGCTGAGGAACAGGGCCCCAGAGCGGCAGCGTGGCCAGTTCAGAAGGTTG

(Continued)

Figure 41A

GTGTTCCAGGAGTTACCCGACGGCAGCTTCAACCAGCCCCGTACAGAGGCGAGCTGAACGAGCAC
CTGGGCCTGCTGGGCCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGGAAC
CAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAGGGC
GCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAGCAC
CACATGGCCCCCACCAGGACGAGTTCGACTGCAAGGCCTGGGCTACTTCTCTGATGTGGACCTG
GAGAAGGACGTGCACAGCGGCCTGATCGGCCCTGCTGGTGTGCCACACCAACACCCTGAACCCC
GCCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCTTACCATCTTCGACGAGACCAAG
AGCTGGTACTTCAACGAGAACATGGAGAGGAACTCAGGGCCCCCTGCAACATCCAGATGGAGGAC
CCCACCTTCAAGGAGAACTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCCGGC
CTGGTGTGAGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGCCAGCAACGAGAACATC
CACAGCATCCACTTCAGCGGCCACGTGTTTACCCTGAGGAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCAGCAAGGCCGGCATCTGGAGG
GTGGAGTGCCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCCTGTTCCCTGGTGTACAGCAAC
AAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCTGGC
CAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACCTACAGCGGCAGCATCAACGCCTGGAGC
ACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATCAAG
ACCCAGGGCGCCAGGCAGAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCCACCCTGATGGTGTCTTCGGCAAC
GTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCATCATCGCCAGGTACATCAGGCTG
CACCCACCCACTACAGCATCAGGAGCACCCCTGCCGATGGAACTGATGGGCTGCGACCTGAACAGC
TGCAGCATGCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGCTAC
TTCACCAACATGTTTCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAAC
GCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTGACCGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTCCTG
ATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCCTGTTCTTCCAGAACGGCAAAGTGAAGGTGTTT
CAGGGCAACCAGGACAGCTTACCCCCGTTGGTGAACAGCCTGGACCCCCCTGCTGACCAGGTAT
CTGAGGATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAAATGGAAGTGTGGGATGCGAG
GCCCAGGACCTGTACTGA (SEQ ID NO:101)

Figure 41B

CS23m1-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCTGTTCTGTTGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGSCCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGTATCACCTGAAGAACATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGCGTGAGCTACTGGAAGGCCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGGCGGACGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGCGACCCCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAAGGAGAAGACCCAGACCCCTGCACAAGTTTATCCTGCTGTTCCCGT
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACCTGAACAGGAGCCTGCCGGCCG
ATCGGCTGCCACAGAAAGTCTGTGTACTGGCAGCTGATCGGCATGGGCACCCACCCCGAGGTGCAC
AGCATCTTCCCTGGAGGGCCACACCTTCCCTGGTGAAGGAACACAGCCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCCTGACCGCCAGACCCCTGCTGATGGACCTGGGCCAGTTCCCTGCTGTCCCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGAGGTTTTGATGATGACAACAGCCCCAGCTTTCATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCCGGAGGAGGAGGACTGGGACTACGCCCCCCCTGGTG
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTGAGATTTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCUCACGGCATCACCGATGTGAGGCCCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGCACCGTGGAGGATGGCCCCACCAAGTCTGACCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCCCTGCCAAC
CCCCCGGGCGTGCAGCTGGAGGACCCCGAGTTCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTTCGACAGCCTGCAGCTGCTGTGTGCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCCAGACCGACTTCCCTGCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGSACACCCCTGACCCCTGTTCCCTTCAGCGGGGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACAACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACGCCATCGAGCCCAGGAGCTTCAGCCAGAACCCCCCGTGTGAAGAGG
CACCAGAGGGAGATCACCAGGACCACCTGCAGAGCGACCAGGAGGAGATCGACTATGATGACACC
ATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCCAGG
AGCTTCCAGAAGAAGACCAGGCACACTTCATCGCCCGCGTGGAGAGGCTGTGGGACTATGGCATG
AGCAGCAGCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTGCCTCAGTTCAAGAAGGTG

(Continued)

Figure 42A

GTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCGTACAGAGGGGAGCTGAACGAGCAC
CTGGGCCTGCTGGGCCCCFACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGGAAC
CAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAGGGC
GCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAGCAC
CACATGGCCCCCACCAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTG
GAGAAGGACGTGCACAGCGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAACCCC
GCCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCCTCACCATCTTCGACGAGACCAAG
AGCTGGTACTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAGGAC
CCCACCTTCAAGGAGAACTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCCGGC
CTGGTGATGGCCAGGACCAGGATCAGGTGGTATCTPGCTGAGCATGGGCAGCAACGAGAATC
CACAGCATCCACTTCAGCGGCCACGTGTTCAACCTGAGGAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACCTGTACCCCGCGTGTTCGAGACCGTGGAGATGCTGCCAGCAAGGCGGCATCTGGAGG
GTGGAGTGCCTGATCGGCGAGCACCTGCACGCCCGCATGAGCACCCCTGTTCCCTGGTGTACAGCAAC
AAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCAACGCCTCTGGC
CAGTACGGCCAGTGGGCCCCCCAAGCTGGCCAGGCTGCACACTACAGCGGCAGCATCAACGCCTGGAGC
ACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCAATCCACGGCATCAAG
ACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTCTTCGGCAAC
GTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCATCATCGCCAGGTACATCAGGCTG
CACCCCACCCACTACAGCATCAGGAGCACCCCTGCCGATGGAAGTGTATGGGCTGCCACCTGAACAGC
TGCAGCATGCCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCAGATCACCGCCAGCAGCTAC
TTCACCAACATGTTCCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAAC
GCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTCCTG
ATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCCTGTTCTTCCAGAACGGCAAAGTGAAGGTGTTTC
CAGGGCAACCAGSACAGCTTACCCCCGTGGTGAACAGCCTGGACCCCCCCTGCTGACCAGGTAT
CTGAGGATCCACCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTCTGGGATGCGAG
GCCAGGACCTGTACTGA (SEQ ID NO:102)

Figure 42B

CS23m23-FL-NA

ATGCAGATTTGAGCTGAGCACCTGCTTCTTCCCTGTGCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGCCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCCAGAGTGCCTAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCTGTTCGTGGAGTTCACCGACCACCTGTTC AACATCGCCAAGCCCAGGCCCCCCTGGATG
GGCTGTGGGGCCCCACCATCCAGGCCGAGGTGTACGACACCCGTGGTGGTCACCCCTGAAGAACATG
GCCAGCCACCCCTGAGCCTGCACGCCGTGGGCGTGGCTACTGGAAGTCCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGAACGACAAGGTGTTCCCCGGCAAGAGCCACACC
TACGTGTGGCAGGTTGCTGAAGGAGAACGGCCCCACTGCCAGCGACCCCCCTGCCGTGACCTACAGC
TACCTGAGCCACGTTGGACCTGCTGAAGGACCTGAACTCTGGCCTGATCGGGCGCCCTGCTGCTGTG
AGGGAGGGCAGCCTGGCCAAAGGAGAAGACCCAGACCCCTGCACAAGTTCATCCTGCTGTTCGCCCTG
TTCGATGAGGGCAAGAGCTTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCCGTGAACGGCTACGTGAACAGGAGCCTGCCGCCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCAGTGTATCGGCATGGGCACCACCCCCGAGGTGCAC
AGCATCTTCTGGAGGGCCACACCTTCCCTGGTGGAGAACACAGCCAGCCAGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCACTACCTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGCTGAGTTTGTATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAAGAAG
CACCCCAAGACCTGGGTGCATACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCCCCTGGTG
CTGGCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGTTCAGATTCATGGCCACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCTGGGCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCUCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTCCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTTCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTCCGACAGCCTGCAGCTGTCTGTGTGCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGGCCCCAGACCGACTTCCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCGTGACCTGTTCCTTCCCTTTCAGCGGGCGAGACCGTGTTCATGAGCATGGAGAACCC
GGCCTGTGGATCCTGGGCCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACACCACCTACGTGAACCGCTCCCTGAGCCAGAACCCCCCGTGTGAAG
AGGCACCAGAGGGAGATCACCAGGACCACCCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGAC
ACCATCAGCGTGGAGATGAAGAAGGAGGACTTTCGACATCTACGACGAGGACGAGAACCAGAGCCCC
AGGAGCTTCCAGAAGAAGACCAGGCCTACTTCATCGCCGCCGTGGAGAGGCTGTGGGACTATGGC
ATGAGCAGCAGCCCCACGTGCTGAGGAACAGGGCCCCAGAGCGGCAGCGTGCCCCAGTTCAAGAAG

(Continued)

Figure 43A

GTGGTGTTCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCCTGTACAGAGGGCAGCTGAACGAG
CACCTGGGCCTGCTGGGCCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGG
AACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAG
GGCGCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAG
CACCATGGCCCCCAAGGACGAGTTCGACTGCAAGGCCCTGGGCTACTTCTCTGATGTGGAC
CTGGAGAAGGACGTGCACAGCGGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAAC
CCCCCCCAGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCTTCACCATCTTCGACGAGACC
AAGAGCTGGTACTTCACCGAGAACATGGAGAGGAAGTGCAGGGCCCCCTGCAACATCCAGATGGAG
GACCCACCTTCAAGGAGAAGTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCC
GGCCTGGTGTATGGCCCAGGACCAAGAGGATCAGGTGGTATCTGCTGAGCATGGCAGCAACGAGAAC
ATCCACAGCATCCACTTCAGCGGCCACGTGTTCAACCGTGGAGGAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTGTACCCCCGGCTGTTTCGAGACCTGGAGATGCTGCCAGCAAGGCCGGCATCTGG
AGGTTGGAGTGCCTGTATCGGCGAGCACCTGCACGCCGGCATGAGCACCTGTTCTTGGTGTACAGC
AACAAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCT
GGCCAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGG
AGCACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATC
AAGACCCAGGGCGCCAGGCAGAAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTCTTCGGC
AACGTGGACAGCAGCGGCATCAAGCACAAACATCTTCAACCCCCCATCATCGCCAGGTACATCAGG
CTGCACCCACCCACTACAGCATCAGGAGCACCCCTGCGGATGGAAGTGTATGGGCTGCGACCTGAAC
AGCTGCAGCATGCCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGC
TACTTCACCAACATGTTTCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGC
AACGCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTC
CTGATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGSCAAAGTGAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCTGCTGACCAGG
TATCTGAGGATCCACCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGTGGGATGC
GAGGCCAGGACCTGTACTGA (SEQ ID NO:103)

Figure 43B

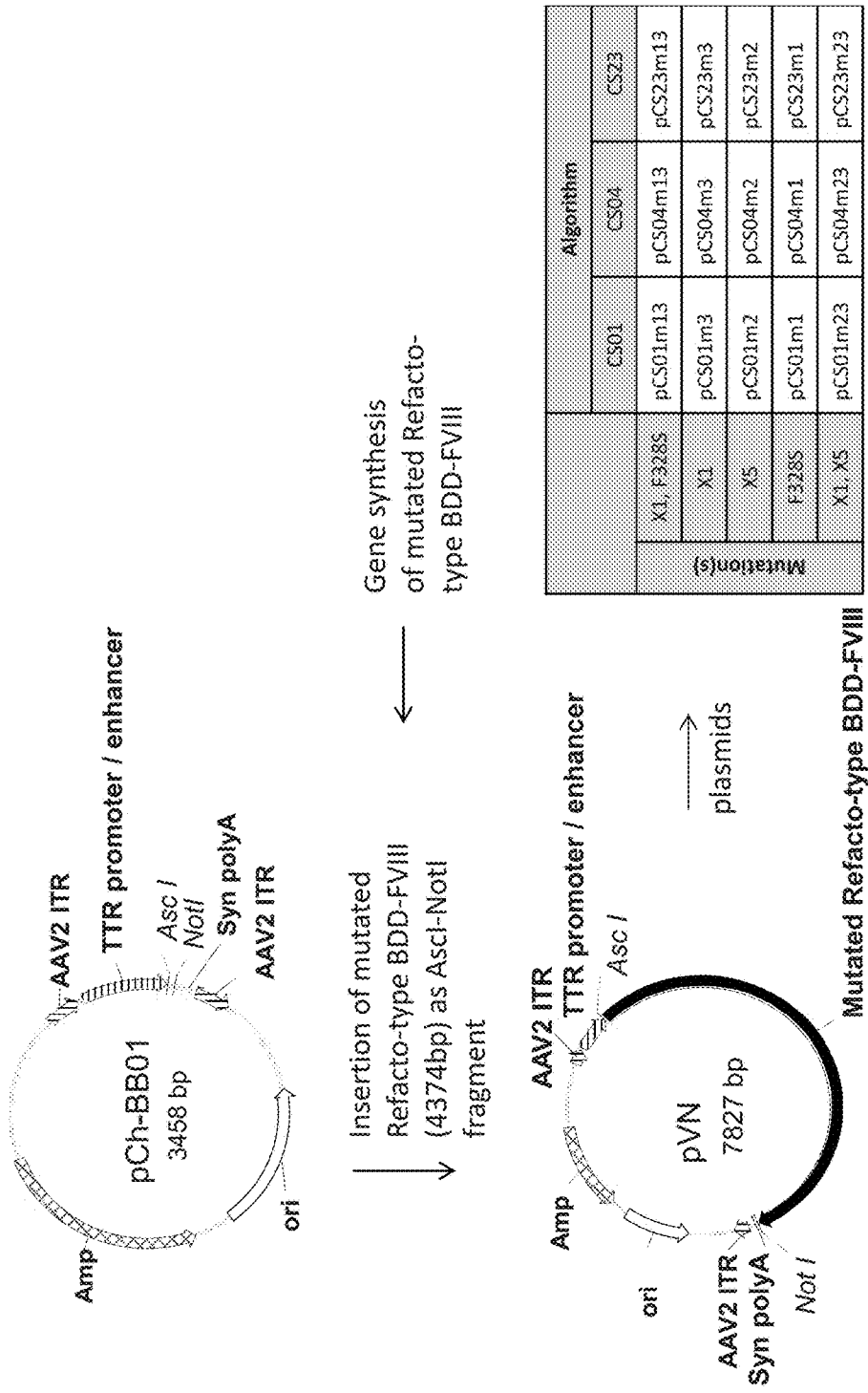


Figure 44

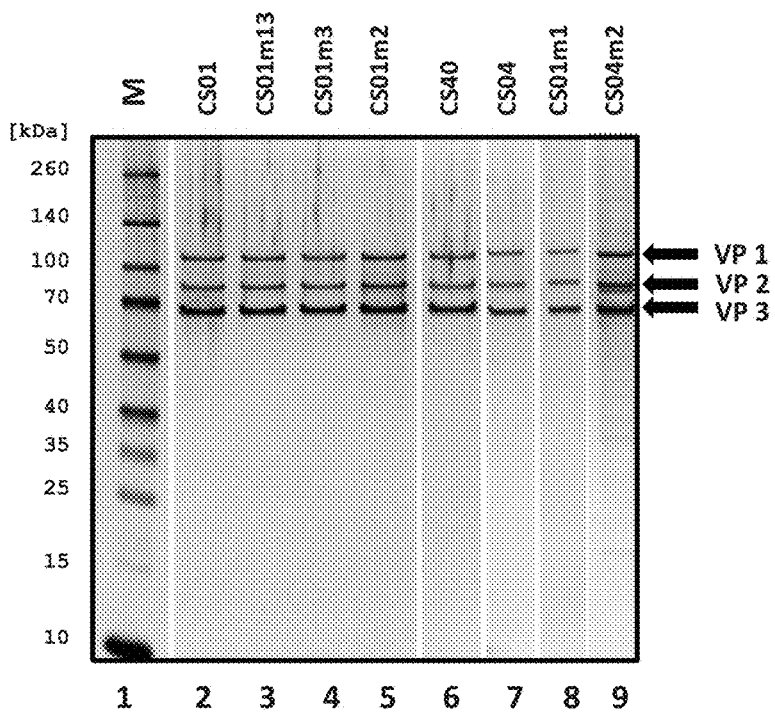


Figure 45

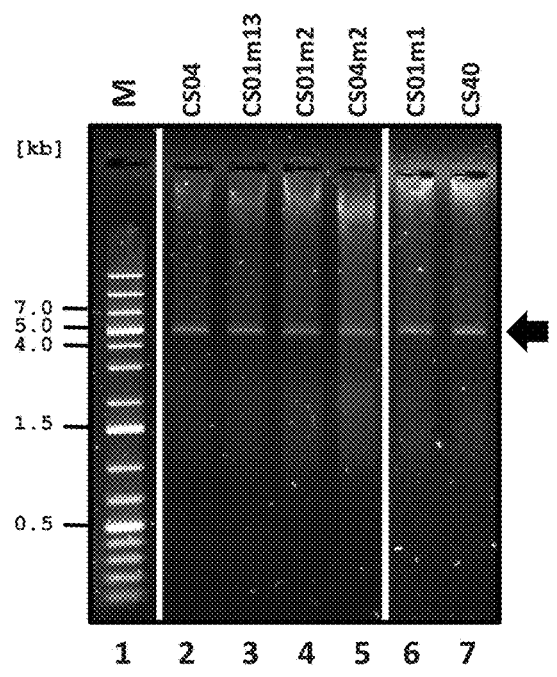


Figure 46

CS01-HC-NA

```

                                                    gcc
accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg
ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac
acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt
gcaaaaccca gaccacccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat
gacactgtgg tcatcacctt caagaacatg gcaccccacc ctgtgtctct gcattgtgtg
ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga
gagaaagagg atgacaaggc gttccctggg ggatctcaca cctatgtgtg gcaagtcctc
aaggagaaatg gacccatggc atctgacca ctctgcctga catactccta cctttctcat
gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa
ggatccctgg ccaaggagaa aaccagaca ctgcacaagt tcattctcct gtttctgtc
tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat
gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca
ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg
acaacccctg aagtgcactc cttttctctg gagggacaca ccttctctgt caggaaccac
agacaagcct ctctggagat ctctcccctc accttctcct ctgcacagac actgctgatg
gaccttgyac agttcctgct gttctgccac atctottccc accagcatga tggcatggaa
gctatgtca aggtggactc atgcctgag gaaccacagc tcaggatgaa gaacaatgag
gaggtgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat
gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca cccaagaca
tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc
cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccaaaaag aattggaaga
aagtacaaga aagtcaagatt catggcctac actgatgaaa ccttcaagac aagagaagcc
attcagcatg agtctggcat tctgggacca ctctgtatg gggaaagtgg agacaccctg
ctcatcatct tcaagaacca ggctccagg cctacaaca tctaccaca tggcatcact
gatgtcaggc cctgtacag caggagactg caaaaggggg tgaaacacct caaggacttc
cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca
acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga
gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag
agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag
aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaccc tgcctgggtg
caactggaag acctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg
tttgaactctc tccagcttct tgtctgctg catgaggtgg cctactggta cattcttct
attggggcac aaactgactt ctttctctg ttcttctctg gatacacctt caagcacaag
atggtgtatg aggacacct gacactcttc ccattctctg gggaaactgt gttcatgagc
atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga
atgactgca tgcctaaagt ctctctctgt gacaagaaca ctggggacta ctatgaggac
tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccaaga
(SEQ ID NO:24X)

```

Figure 47

CS01-LC-NA

```

                                y agatcaccag gacaaccctc
cagtctgacc aggaagagat tgactatgat gacaccattt ctgtggagat gaagaaggag
gactttgaca tctatgatga ggacgagaac cagtctccaa gatcattcca gaagaagaca
agacactact tcattgctgc tgtggaaga ctgtgggact atggcatgtc ttctctctcc
catgtcctca ggaacagggc acagtctggc tctgtgccac agttcaagaa agtgggtctc
caggagtca ctgatggctc attcaccag ccctgtaca gagggaact gaatgagcac
ctgggactcc tgggaccata catcagggtc gaggtggaag acaacatcat ggtgacattc
agaaaaccagg cctccaggcc ctacagcttc tactcttccc tcatcageta tgaggaagac
cagagacaag gggctgagcc aagaaagaac tttgtgaaac ccaatgaaac caagacctac
ttctggaaaag tccagcacca catggcacco accaaggatg agtttgactg caaggcctgg
gcatacttct ctgatgtgga cctggagaaa gatgtgcaact ctggcctgat tggcccactc
ctggctctgcc acaccaacac cctgaacctc gcacatggaa ggcaagtgac tgtgcaggag
tttgcctct tcttaccat ctttgatgaa accaagtcat ggtacttcac tgagaacatg
gagagaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagaac
tacaggttcc atgccatcaa tggctacatc atggacacco tgcctgggct tgtcatggca
caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cttcactcc
atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggcctg
tacaacctct acctgggggt ctttgagact gtggagatgc tgcctccaa agctggcctc
tggaggggtg aatgcctcat tggggagcac ctgcatgctg gcattgtcaac cctgttctg
gtctacagca acaagtgcc gacacctctg ggaatggcct ctggccacat cagggacttc
cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac
tctggctcca tcaatgcatg gtcaaccaag gagccattct cttggatcaa ggtggacctg
ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctctct
ctgtacatct cacagttcat catcatgtac tctctggatg gcaagaagtg gcagacatac
agaggcaact cactggcac cctcatggtc ttctttggca atgtggacag ctctggcctc
aagcacaaca tottcaacct tccatcatt gccagataca tcaggctgca cccaccccac
tactcaatca gatcaacctt caggatggaa ctgatgggat gtgacctgaa ctctgctca
atgcccctgg gaatggagag caaggccatt tetgatgccc agatcactgc atcctcttac
ttaccaaca tgtttgccac ctggtcacca tcaaaagcca ggctgcaact ccaggggaaga
agcaatgctt ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttccag
aagacaatga aagtcaactg ggtgacaacc caggggggta agtctctgct cacctcaatg
tatgtgaagg agttctgat ctcttctca caggatggcc accagtggac actctcttc
cagaatggca aagtcaaggt gttccagggc aaccaggact ctttcaacc tgtggtgaac
tcactggacc cccccctct gacaagatac ctgagaatto acccccagtc ttgggtccac
cagattgccc tgagaatgga agtctgsgga tgtgaggcac aagacctgta c
(SEQ ID NO:25)

```

Figure 48

CS01A(760-1667) - CS01-SC1-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGGAGATAC
TACCTGGGGGCTGTGGAACTTTCTTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCTGTGGATGCCAGG
TTCCCACCCAGAGTGCCTCAAGTCTTCCCATTCAACACCTCTGTGGTCTACAAGAAGACACTCTTTGTGGAA
TTCCTGACCACCTGTCAACATTGCAAAACCCAGACCACCTGGATGGGACTCCTGGGACCCACCATTGAG
GCTGAGGTGTATGACACTGTGGTCAACCCCTCAAGAACATGGCATCCCACCTGTGCTCTGCATGCTGTG
GGAGTCTCATACTGGAAAGCCTCTGAAGGGGCTGAGTATGATGACCAGACATCCCAGAGAGAGAAAGAGGAT
GACAAGGTGTTCCCTGGGGGATCTCACACCTATGTGTGGCAAGTCCCTCAAGGAGAATGGACCCATGGCATCT
GACCCACTCTGCCTGACATACTCCTACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATT
GGGGCACTGCTGGTGTGCAGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTTATTCTC
CTGTTTGTCTCTTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGAT
GCTGCCCTCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCAFTGATTGGCATGGGACAACCCCTGAAGTGCCTCCATT
TTCCFTGGAGGGACACACTTCTCTGGTCAGGAACCCAGACAGCAAGCCTCTCTGGAGATCTCTCCCATCACCTC
CTCAGTGCACAGACACTGCTGATGGACCTTGGACAGTTCCCTGCTGTTCTGCCACATCTCTCCACCAGCAT
GATGGCATGGAAAGCCTATGTCAAGGTGGACTCATGCCCCCTGAGGAACCCAGCTCAGGATGGAAGAACAATGAG
GAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGGTCAGATTTGATGATGACAACCTCT
CCATCCTTCATTGAGATCAGGTCTGTGGCAAAGAAAACCCCCAAGACATGGGTGCACTACATTGCTGCTGAG
GAAGAGGACTGGGACTATGCACCCTGGTCCCTGGCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAAC
AATGGCCCAAAAGAATTTGAAGAAAGTACAAGAAAGTCAAGATTCATGGCTACACTGATGAAACCTTCAAG
ACAAGAGAAGCCATTGAGTCTGGCATTCTGGGACCCTCTGATGGGGAAGTGGGAGACACCCCTG
CTCATCATCTTCAAGAACCAGGCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCC
CTGTACAGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTCCTCACCAGATACTAC
TCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCCTGCTCATCTGCTACAAGGAG
TCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTTCTGTTCTCTGTCTTTGATGAG
AACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCTTGCCCAACCCTGCTGGGGTSCAACTGGAAGAC
CCTGAGTTCCAGGCAAGCAACATCATGCACATCCATCAATGGCTATGTGTTTGGACTCTCTCCAGCTTTCTGTG
TGCTGGCATGAGGHTGGCTACTGGTACATTTCTTATTGGGGCACAACTGACTTCTCTGCTCTTCTTCTC
TCTGGATACACCTTCAAGCACAAGATGGTGTATGAGGACACCCTGACACTCTTCCATTCTCTGGGGAAACT
GTGTTTATGAGCATGGAGAACCCTGGACTGTGGATTCTGGGATGCCACAACCTCTGACTTCAGAAACAGGGGA
ATGACTGCACTGCTCAAAGTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGAC
ATCTCTGCCTACCTGCTCAGCAAGAACAATGCCATTTGAGCCCAGAGAGATCACCAGGACAACCCTCCAGTCT
GACCAGGAAGAGATTGACTATGATGACACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGAT
GAGGACGAGAACCAGTCTCCAAGATCATTCAGABAAGAAGACAAGACACTACTTCAATTGCTGCTGTGGAAAGA
CTGTGGGACTATGGCATGTCTTCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAG
TTCAAGAAAGTGGTCTTCCAGGAGTTCACTGATGGCTCATTACCCAGCCCCCTGTACAGAGGGGAACTGAAT
GAGCACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGCATTCAGAAAC
CAGGCCCTCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGGGCTGAG
CCAAGAAAGAACTTTGTGAAACCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCACCACATGGCACCC

(Continued)

Figure 49A

ACCAAGGATGAGFTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGACCTGGAGAAAGATGTGCACTCT
 GGCCTGATFPGCCCACTCCTGGTCTGCCACACCAACACCCCTGAACCCTGCACATGGAAGGCAAGTACTGTG
 CAGGAGTTFGCCCTTCTTCCACCATCTTTGATGAAACCAAGTCATGGTACTTCACTGAGAACATGGAGAGA
 AACTGCAGAGCACCATGCAACATTCAGATGGAAGACCCACCTTCAAGGAGAACTACAGGTTCCATGCCATC
 AATGGCTACATCAFGGACACCCCTGCCTGGGCTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTT
 TCTATGGGATCCAATGAGAACATTCCTCCATCCACTTCTCTGGGCATGCTTCACTGTGAGAAAGAAGSAG
 GAATACAAGATGGCCCTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCOCTCCAAAGCT
 GGCATCTGGAGGGTGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCCTGTTCTGGTCTAC
 AGCAACAAGTGGCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCTCTGGC
 CAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCAACCAAG
 GAGCCATTCTCTGGATCAAGGEGGACCTGCTGGCACCCATGATCATTTCATGGCATCAAGACACAGGGGGCA
 AGACAGAAATCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTGGATGGCAAGAAGTGGCAG
 ACATACAGAGGCAACTCCACTGGCACCCCTCATGGTCTTCTTTGGCAATGTGGACAGCTCTGGCATCAAGCAC
 AACAFCTTCAACCCCTCCCATCATTTGCCAGATACATCAGGCTGCACCCACCCACTACTCAATCAGATCAACC
 CTCAGGATGGAACGTGATGGGATGTGACCTGAACCTCCTGCTCAATGCCCTGGGAATGGAGAGCAAGGCCATT
 TCTGATGCCCAGATCACTGCATCCTCTTACTTCACCAACATGTTTGGCCACTGGTCACCATCAAAAGCCAGG
 CTGCACCTCCAGGGGAAGCAATGCCTGGAGACCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGAC
 TTCCAGAAGACAATGAAAGTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTG
 AAGGATTCCTGATCTCTTCTCACAGGATGGCCACCAGTGGACACTCTTCTCCAGAAATGGCAAAGTCAAG
 GTGTTCCAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACCTCACTGGACCCCCCTCTGACAAGATAC
 CTGASAATTCACCCCAAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCTGGGATGTGAGGCACAA
 GACCTGTACTGA (SEQ ID NO:26)

Figure 49B

CS01A(772-1667) - CS01-SC2-NA

ATGCAGATTTGAGCTGTCCAACCTGCTTCTTTCTGTGCCCTGCTGAGATTTCTGCTTCTCTGCCACCAGG
AGATACTACCTFGGGGGCTGTGGAACCTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCAACCCAGAGTGCCCAAGTCCTTCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCAGTACCACCTGTTCAACATTGCAAAAACCCAGACCACCTGGATG
GGACTCCTGGGACCCACCATTGAGGCTGAGGTGTATGACACTGTGGTCATCACCCCTCAAGAACATG
GCATCCCAACCTGTGTCTCTGCAATGCTGTGGGAGTCTCATACTGGAAAGCCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACC
TATGTGTFGGCAAGTCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACTCC
TACCTTTCTCATGTGGACTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTTCTCTGTTTGTGTGC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCTGGAGGGACACACCTTCTGGTCAAGGACCCACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTFCAGATTTGATGATGACAACCTCTCCATCCTTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCCTGGTC
CTGGCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCAAGAAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTGAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCCTGCTCATCATCTTC
AAGAACCAGGCCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCTGTAC
AGCAGGAGACTGCCAAAAGGGTGAACACCTCAAGGACTTCCCCATTCTGCTTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCAAC
CCTGCTGSSGTGCAACTGGAAGACCCCTGAGTTCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTCT
ATTTGGGGCACAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAAGATGGTG
TATGAGGACACCCCTGACACTCTTCCATTCTCTGGGAAACTGTGTTTCATGAGCATGGAGAACCCT
GGACTGTGGATTTCTGGGATGCCACAACCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAAA
GTCTCCTCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTTGAGCCCAAGCTTCTCTCAGAATTCCAGACACCCCAAGCACC
AGGGAGATCACAGGACAACCCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGACACCATTCT
GTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCAAGATCATTC

(Continued)

Figure 50A

CAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGCATGTCTTCC
TCTCCCCATGTCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCFAAGAAAGTGGTCTTC
CAGGAGTTCACTGATGGCTCATTACCCAGCCCCGTGACAGAGGGGAACTGAATGAGCACUTGGGA
CTCCTGGGACCATAACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGAAAACCAGGCC
TCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGGGCTGAG
CCAAGAAAGAACTTTGTGAAACCCAAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCACCATG
GCACCCACCAAGGATGAGTTTGAAGTCAAGGCCTGGGCATACTTCTCTGATGTGGACCTGGAGAAA
GATGTGCACTCTGGCCTGATTGGCCACTCCTGGTCTGCCACACCAACACCCTGAACCCTGCACAT
GGAAGGCAAGTGAAGTGTGAGGAGTTTGGCCCTCTTCTCACCATCTTTGATGAAACCAAGTCATGG
TACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAAGACCCACC
TTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGGCTTGT
ATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAAATGAGAACATTCACTCC
ATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATAACAAGATGGCCCTGTACAAC
CTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGGGTGGAA
TGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCTGGTCTACAGCAACAAGTGC
CAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGCCAGTAT
GGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCAACCAAG
GAGCCATTCTCTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATCAAGACACAG
GGGGCAAGACAGAAATTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTGGATGGC
AAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCCTCATGGTCTTCTTTGGCAATGTGGAC
AGCTCTGGCATCAAGCACAAACATCTTCAACCCTCCCATCATFGCCAGATACATCAGGCTGCACCCC
ACCCACTACTCAATCAGATCAACCCTCAGGATGGAAGTGTGGATGTGACCTGAACTCCTGCTCA
ATGCCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCAGATCACTGCATCCTCTTACTTCACC
AACATGTTTGGCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGCAATGCCTGG
AGACCCCAAGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATGAAAGTCACT
GGGGTGACAACCCAGGGGTCAAGTCTCTGCTCACCTCAATGATGATGAAGGAGTTCCTGATCTCT
TCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAAATGGCAAAGTCAAGGTGTCCAGGGC
AACCAGGACTCTTTCACACCTGTGGTGAACACTGAGACCCCCCTCCTGACAAGATACCTGAGA
ATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAAATGGAAGTCTTGGGATGTGAGGCACAA
GACCTGTACTGA (SEQ ID NO:27)

Figure 50B

CS23A(760-1667) - CS23-SC1-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTCCTGTGCCCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGATAC
TACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCTGTGGACGCCAGG
FTCCCCCCCAGAGTGCCCAAGAGCTTCCCTTCAACACCTCAGTGGTGTACAAGAAGACCCTGTTCGTGGAG
TTCACCGACCACCTGTTCAACATCGCCAAGCCAGGCCCCCTGGATGGGCTGCTGGGCCCCACCATCCAG
GCGGAGGTGTACGACACCGTGGTGTACCCCTGAAGAACATGGCCAGCCACCCCGTGAGCCTGCACGCCGTG
GGCGTGACTACTGSAAGGCCCTGTGAGGGCGCCGAGTATGACGACCAGACCAGCCAGAGGSAGAAGGAGGAC
GACAAGGTTTCCCCGGCCGAGCCACACCTACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGC
GACCCCTGTGCTGACCFACAGCTACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACCTCTGGCCTGATC
GGCCCTGTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCAGACCCCTGCACAAGTTCATCCTG
CTGTTCCGCTGTTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGAT
GCCGCTCTGCCAGGGCCTGGCCAAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGCCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCGAGGTGCACAGCATC
TTCTGGAGGGCCACACCTTCTGGTGAAGAACACAGCGAGGCCAGCCTGGAGATCAGCCCCATCACCTTC
CTGACCGCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTCTGCCACATCAGCAGCCACCAGCAC
GACGGCATGGAGGCCATACGTGAAGGTGGACAGCTGCCCGAGGAGCCACAGCTGAGGATGAAGAACAACGAG
GAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATGGACCTGGTGAAGTTGATGATGACAACAGC
CCCAGCTTCAACAGATCAGTCTGTGGCCAAGAAGCACCCCAAGACCTGGGTGCACTACATCGCCGCGGAG
GAGGAGGACTGGGACTACGCCCCCTGGTCTGGCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAAC
AACGGCCCCCAGAGGATCGGCAGGAAGTACAAGAAGGTGAGATTTCATGGCCTACACCGACGAGACCTTCAAG
ACCAGGGAGGCCATCCAGCAGAGTCTGGCATCCTGGGCCCTGCTGTACGGGAGGTTGGGCGACACCCCTG
CTGATCATCTCAAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCAACGGCATCACCGATGTGAGGCC
CTGTACAGCAGGAGGCTGCCCAAGGGCCTGAAGCACCTGAAGGACTTCCCATCCTGCCCGGCGAGATCTT
AAGTACAAGTGGACCGTGAACCTGGAGAGGACCTGGCCTCTGSCCTGATCGGCCCTGCTGATCTGTACAAGGAG
AGCAGCTTCTGTAACATGGAGAGGGACCTGGCCTCTGSCCTGATCGGCCCTGCTGATCTGTACAAGGAG
AGCCTGGACCCAGAGGGGCAACCAGATCATGCTGACAAGAGGAACGTGATCCTGTCTCTGTGTTCGATGAG
AACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCTGCCCAACCCCGCGGCTGCAGCTGGAGGAC
CCCGAGTCCAGGCCAGCAACATCATGCACAGCATCAACGGCTACGTGTTGACAGCCTGCAGCTGTCTGTG
TGCTGCACGAGGTGGCCTACTGGTACATCCTGAGCATCGGGCCCCAGACCGACTTCTGTCTGTGTTCTTC
TCTGGCTACACCTTCAAGCACAAGATGGTGTACGAGGACACCCCTGACCTGTTCCTCCTCAGCGGCCGAGACC
GTGTTATGAGCATGGAGAACCCCGGCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGC
ATGACCGCCCTGCTGAAAGTCAAGCAGCTGCGACAAGAACACCGGGGACTACTACGAGGACAGCTACGAGGAC
ATCAGCGCCTACCTGCTGAGCAAGAACAACGCCATCGAGCCAGGGAGATCACCAGGACCACCCCTGCAGAGC
GACCAGGAGGAGATCGACTATGATGACACCATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGAC
GAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATCGCCCGCTGGAGAGG
CTGTGGACTATGGCATGAGCAGCAGCCCCACGTGCTGAGGAACAGGGCCCCAGAGCGGCAGCGTGCCTCAG
TTCAAGAAGGTGGTGTCCAGGAGTTCACCGACGGCAGCTTCAACCCAGCCCCCTGTACAGAGGCGAGCTGAAC
GAGCACCTGGGCTGCTGGGCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTACCTTCAGGAAC
CAGGCCAGCAGCCCTACAGCTTCTACAGCAGCTGATCAGCTACGAGGAGGACCAGAGGACAGGGCCGCGGAG

(Continued)

Figure 51A

CCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAGCACCACATGGCCCCC
ACCAAGGACGAGTTCGACTGCAAGGCTGGGCTACTTCTCTGATGTGGACCTGGAGAAGGACGTGCACAGC
GGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCTGAACCCCGCCACGGCAGGCAGGTGACCGTG
CAGGAGTTCGCCCTGTTCTCACCATCTTCGACGAGACCAAGAGCTGGTACTTCACCGAGAACATGGAGAGG
AACTGCAGGGCCCCCTGCAACATCCAGATGGAGGACCCACCTTCAAGGAGAACTACAGGTCCACGCCATC
AACGGCTACATCATGGACACCCTGCCCGCCTGGTGTGGCCAGGACCAGAGCATCAGGTGGTATCTGCTG
AGCATGGGCAGCAACGAGAACATCCACAGCATCCACTTCAGCGGCCACGTSTTCACCGTGAGGAAGAAGGAG
GAGTACAAGATGGCCCTGTACAACCTGTACCCCGGGCTGTTCGAGACCGTGGAGATGCTGCCAGCAAGGCC
GGCATCTGGAGGGTGGAGTGCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCCTGTTCTCTGGTGTAC
AGCAACAAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCTGGC
CAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCCTGGAGCACCAAG
GAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCATGATCATCCAGGCATCAAGACCCAGGGCGCC
AGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGACGGCAAGAAGTGGCAG
ACCTACAGGGGCAACAGCACCGGCACCTGATGGTGTCTTCGGCAACGTGGACAGCAGCGGCATCAAGCAC
AACATCTCAACCCCCCATCATCGCCAGGTACATCAGGCTGCACCCACCCACTACAGCATCAGGAGCACC
CTGCGGATGGAAGTGTATGGGCTGCGACCTGAACAGCTGCAGCATGCCCTGGGCATGGAGAGCAAGGCCATC
TCTGACGCCCAGATCACCGCCAGCAGCTACTTCACCAACATGTTTCGCCACCTGGAGCCCCAGCAAGGCCAGG
CTGCACCTGCAGGGCAGGAGCAACGCCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGAC
TTCCAGAAGACCATGAAGGTGACCGGCTGACCACCCAGGGCTGAAGAGCCTGCTGACCAGCATGTACGTG
AAGGAGTTCCTGATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCAGAACGGCAAGTGAAG
GTGTTCCAGGGCAACCAGGACAGCTTCACCCCGTGGTGAACAGCCTGGACCCCCCCTGCTGACCAGGTAT
CTGAGGATCCACCCCAAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGTCTGGGATGCGAGGCCAG
GACCTGTACTGA (SEQ ID NO:28)

Figure 51B

CS23A(772-1667) - CS23-SC2-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCTGTTCGTGGAGTTTCAACGACCACCTGTTCACATCGCCAAGCCCAGGCCCCCCCTGGATG
GGCTGTGGGGCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGTATCACCCTGAAGAACATG
GCCAGCCACCCCGTGGCCTGCACGCCGTGGGCGTGGAGCTACTGGAAGGCCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGGCGGCAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGCGACCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCTGATCGGGGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTCCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGCCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCAGTGTATCGGCATGGGCACCACCCCGAGGTGCAC
AGCATCTTCCTGGAGGGCCACACCTTCTGGTGGAGAACACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTGACCGCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCGGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGGGTTTGTATGATGACAACAGCCCCAGCTTCAATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCCGGAGGAGGACTGGGACTACGCCCCCTGGTGTG
CTGGCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTGAGATTCATGGCCTACACCGACGAGACCTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGGCCCCCTGCTGATC
TGCTACAAGGAGAGCCTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTTCGACAGCCTGCAGCTGTCTGTGTGCTGCACGAGGTGGCTACTGGTACATCCTGAGC
ATCGGGCCCCAGACCGACTTCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTGT
TACGAGGACACCCTGACCCTGTTCCCCCTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGGACAAGAACAACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACGCCATCGAGCCCAGGAGCTTCAGCCAGAACTCCAGACACCCAGCACC

(Continued)

Figure 52A

AGGGAGATCACCAGGACCACCCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGACACCATCAGC
GTFGGAGATGAAGAAGGAGGACTTCGACATCFAAGACGAGGACGAGAACCAGAGCCCCAGGAGCTTC
CAGAAGAAGACCAGGCACACTTTCATCGCCGCCGTFGGAGAGGCTGTGGGACTATGGCATGAGCAGC
AGCCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCCGCCCCAGTTCAAGAAGGTGGTGTTC
CAGGAGTTCACCGACGGCAGCTTCAACCAGCCCCGTACAGAGGGCAGCTGAACGAGCACCCTGGGC
CTFGCTGGGCCCCFACATCAGGGCCGAGGTGGAGGACAAATCATGCTGACCTTCAGGAACCAGGCC
AGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAGGGCCCGAG
CCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCFPGAAGGTGCAGCACCACATG
GCCCCACCAAGGACGAGTTCGACTGCAAGGCCFGGGCCACTTCCTCTGATGTGGACCTGGAGAAG
GACGTGCACAGCGGCCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCCTGAACCCCCGCCAC
GGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCCTTCACCATCTTCGACGAGACCAAGAGCTGG
TACTTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAGGACCCACC
TTCAAGGAGAACTACAGGTTCACGCCATCAACGGCTACATCAFGGACACCCFPGCCGGCCFGGTG
ATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAACATCCACAGC
ATCCACTTCAGCGGCCACGTGTTCCACCGTGAAGGAAGGAGGAGTACAAGATGGCCCCGTACAAC
CTGTACCCCGGGCTGTTCGAGACCGTGGAGATGCTGCCAGCAAGGCCGGCATCTGGAGGGTGGAG
TGCCCTGATCGGCCGAGCACCTGCACGUCGGCATGAGCACCCTGTTCTGGTGTACAGCAACAAGTGC
CAGACCCCCCTGGGCAFGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCCTCTGGCCAGTAC
GGCCAGTGGGCCCCCCAAGCTGGCCAGGCTGCACCTACAGCGGCAGCATCAACGCCCTGGAGCACCAAG
GAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGGCCCCATGATCAATCCACGGCATCAAGACCCAG
GGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGACGGC
AAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCCACCCCTGATGGTGTCTTCFPGGCAACCTGGAC
AGCAGCGGCATCAAGCACAAACATCTTCAACCCCCCATCATCGCCAGGTACATCAGGCTGCACCCCC
ACCCACTACAGCATCAGGAGCACCCCTGCGGATGGAACTGATGGGCTGCGACCTGAACAGCTGCAGC
ATGCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGCTACTTCCACC
AACATGTFPGCCACCFTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAACGCCCTGG
AGGCCCCAGGTGAACAACCCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAAGACCATGAAGGTGACC
GGCCTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACCTGAAGGAGTTCCTGATCAGC
AGCAGCCAGGACGGCCACCAGTGGACCCCTGTTCTTCCAGAACGGCAAGTGAAGGTGTTCAGGGC
AACCAGGACAGCTTCAACCCCCGTGGTGAACAGCCTGGACCCCCCCCCCTGCTGACCAGGTATCTGAGG
ATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAAATGGAAGTGTCTGGGATGCGAGGCCAG
GACCTGTACTGA (SEQ ID NO:29)

Figure 52B

CS01m23-FL-AA (SEQ ID NO: 104)

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDAREFPPRVPKSFFPN
TSVVYKKTFLVFEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVVTLKNMASHFVSLHAV
GVSYWKSSEGAEYDDQTSQREKEDDKVFPGKSHTYVWQVLKENGPTASDPPCLFYSYLSH
VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRD
AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNH
RQASLEISPIITFLTAQTLMLDLGQFLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNE
EAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSVAKKHFKTWVHYIAAEEEDWDYAPLVLA
PDDRSYKSQYLNNQPQRIGRKYKVRFMAYTDETFKTREAIQHESGILGPLYGEVGDTL
LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP
TKSDPRCLTRYYSFVNMERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVILFSVFDE
NRSWYL TENIQRF LNPAGVQLEDPEFQASNIMHSINGYVFDLQLSVCLHEVAYWYILS
IGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSENPGLWILGCHNSDFRNRG
MTALLKVSSCDKNTGDYEDSYEDI SAYLLSKNNTTYVNRSLSQNPPVLKRHQREITRPT
LQSDQEEIDYDDTISVEMKKEFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSS
PHVLRNRAQSGSVPQFKKVVFEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVT
FRNQASRPYSFYSSLSIYEEDQRQGAEPKRNFKPNETKTYFWKVQHMAPTKDEFDCKA
WAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTEN
MERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQIRWYLLSMGSNENIH
SIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLF
LVYSNKCQTPGLMASGHIRDFQITASGOYGQWAPKLARLHYSGSINAWSTKEPFSWIKVD
LLAPMI IHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSG
IKHNIFNPPIIARYIRLHPHTYSISIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASS
YFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTCVTTQGVKSLITS
MYVKEFLISSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIPHQSWV
HQIALRMEVLGCEAQDLY

Figure 53

CS04m3-FL-AA (SEQ ID NO: 105)

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFN
TSVYKKTLEFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAV
GVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSH
VDLVKDLNSGLIGALLVCREGLAKEKTQTLHKFILLFAVFDEGKSWHSETKNLSMQDRD
AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNH
RQASLEISPIITFLTAQTLLMDLGQFLLFCHISSHQHDGMEAYVKVDSCEEPQLRMKNNE
EAEDYDDDLTDSEMDVVRFDNNSPFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLA
PDDRSYKSYQLNNGPQRIQRKYKVRFMAYTDETFKTREAIQHESGILGPELLYGEVGDTL
LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP
TKSDPRCLTRYSSFVNMRDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVI LFSVFDE
NRSWYL TENIQRF LNPAGVQLEDPEFQASNIMHSINGYVFDLQLSVCLHEVAYWYILS
IGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFM SMENPGLWILGCHNSDFRNRG
MTALLKVSSCDKNTGDYEDSYEDI SAYLLSKNNTTYVNRSLSQNPPVLKRHQREITRRT
LQSDQEEIDYDDTISVEMKKEDEFDIYDEDENQSPRSEFQKKTRHYFIAAVERLWDYGMSSS
PHVLRNRAQSGSVPOFKKVVVFQEF TDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVT
FRNQASRPYSFYSSLSIYEEDQRQGAEPKRFVKNPNETKTYFWKVQHMAPTKDEFDCKA
WAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDET KSWYFTEN
MERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIH
SIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPKAGIWRVECLIGEHLHAGMSTLF
LVYSNKCQTPLGMASGHIRDFQITASGOYQWAPKLARLHYSGSINAWSTKEPFSWIKVD
LLAPMIHGIKTQGARQKFSSLYISQFTIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSG
IKHNIFNPPIIARYIRLHPHYSIRSTLRMELMGC DLNSCSMPLGMESKAISDAQITASS
YFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLS
MYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRHPQSWV
HQIALRMEVLGCEAQDLY

Figure 54

CS01-FL-AAm12 (SEQ ID NO: 106)

MQIELSTCFFLCLLRFCFSATTRYYLGAVELSWDYMQSDLGELPVDARFPPRPVPKSFPFNTSVVYK
KTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVVTLKNMASHPVSLHAVGVSYWKSSEGAE
YDQTSQREKEDDKVFPGKSHTYVWQVLKENGPTASDPPCLTYSYLSHVDLVKDLNSGLIGALLVC
REGSLAKEKTQTLHKFILLFAVFDECKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGL
IGCHRKSVYVHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPIITFLTAQTLLMDLGQFLLSCH
ISSHQHDGMEAYVKVDSCEPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNPSFSFIQIRSVAKK
HPKTWVHYIAAEEEEWDYAFLVLA PDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTTREAIQ
HESGILGPLLYGEVGD TLLIIFKNQASRPYNIYPHGI TDVRPLYSRRLPKGVKHLKDFPILPGEIF
KYKWTVTVEDGPTKSDPRCLTRYSSFVNMERDLASGLIGPLLYCYKESVDQRGNQIMSDKRNVL
FSVFDENRSWYL TENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYLLS
IGAQTDFLSVFFSGYTFKHKMVYEDTTLTLPFSGETVFMSENPGLWILGCHNSDFERNRGM TALLK
VSSCDKNTGDYEDSYEDISAYLLSKNNAIEPRSFQSNPPVLKRHQREITR TTTLOSDQEEIDYDDT
ISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPFKKV
VFQEF TDGSFTQPLYRGELNEHLGLLGPYIRAEVDNIMVTFRNQASRPYSFYSSLI SYEEDQRQG
AEPRKNEVKPNETKTYFWKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNP
AHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPG
LVMAQDQIRWYLLSMGSENENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLP SKAGIWR
VECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYQWAPKLARLHYSGSINAWS
TKEPFSWIKVDLLAPMI IHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTG TLMVFFGN
VDSSGIKHNI FNPPILARYIRLHPTHYSIRSTLRMELMGC DLNCSMPLGME SKAISDAQITASSY
FTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQRTMKVTGVTTQGVKSLLTSMYVKEFL
ISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCE
AQDLY

Figure 55

CS04-FL-AAm12 (SEQ ID NO: 107)

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPPNTSVVYK
KTLFVEFTDHLFNIAKPRFPWMGLLGPTIQAEVYDTVVVTLKNMASHPVSLHAVGVSYWKSSEGAE
YDDQTSQREKEDDKVFPFGKSHTYVWQVLKENGPTASDPPCLTYSYLSHVDLVKDLNSGLIGALLVC
REGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGL
IGCHRKSVYWHVIGMGTTPPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQFLMDLGQFLLSCH
ISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDNNSPSFIQIRSVAKK
HPKTWVHYIAAEEDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQ
HESGILGPLYGEVGDITLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIF
KYKWTVTVEDGPTKSDPRCLTRYSSFVNMERDLASGLIGPLLI CYKESVDQRGNQIMSDKRNVI L
FSVFDENRSWYL TENIQRF LPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAWYIIS
IGAQTDFLSVFFSGYTFKHKMVEYEDTLTLFFFSGETVFMSEMPGLWILGCHNSDFRNRMGTALLK
VSSCDKNTGDYYEDSYEDI SAYLLSKNNAIEPRSFQNPVLRKHQREITRTTLQSDQEEIDYDDT
ISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLNRNRAQSGSVQPKKV
VFQEF TDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQG
AEPRKNFVKPNETKTYFWKVQHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNP
AHGRQVTVQEFALFFTIIFDETKSWYFTENMERNCRAPCNIQMEDPTFRENRYRFHAINGYIMDTLPG
LVMAQDQIRIRWYLLSMGNSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFEFVEMLPKAGIWR
VECLIGEHLHAGMSTLFLVYSNKCQTPLGMA SGHIRDFQITASGQYQWAPKLARLHYSGSINAWS
TKEPFSWIKVDLLAPMI IHGIKTQGARQKFS SLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGN
VDSSGIKHNI FNPP IARYIRLHPHYSIRSTLRMELMGC DLNSCSMPLGMESKAISDAQITASSY
FTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTQGVKSL L TSMYVKEFL
ISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLFRYLRIHPQSWVHQIALRMEVLGCE
AQDLY

Figure 56

CS01-FL-NA12 (SEQ ID NO: 108)

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACCTTTCTTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCTCAAGTCTTTCCCATTTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACCTGACCACCTGTTCAACATTGCAAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTTCAGGCTGAGGTGTATGACACTGTGGTCTGTCACCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAATUCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAAGAGGATGACAAGGTGTTCCCTGGGAAGTCTCACACC
TATGTGTGGCAAGTCTTCAAGGAGAATGGACCCACTGCATCTGACCCACCCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTTCTCCTGTTTGTCTGTC
TTTGATGAGGGCAAGTCTTTGGCACTCTGAAACAAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCTCCTGGAGGGACACACCTTCTGGTCAAGGAACACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTTCTGCTGTCTGCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAAGTTTGTGATGATGACAACCTCTCCATCCTTCAATTCAGATCAGGTCTGTGGCAAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCCTGGTCT
CTGGCCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCAAAAAGAATTGGAAGA
AAGTACAAGAAAGTCAAGTTTCAAGGCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTGAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCAGGTGCCTCACCCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCCTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAAGAGATTCTGCCCCAAC
CCTGCTGGGGTGAACCTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCACAACCTGACTTCTTTCTGTCTTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTTATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACCTCTGACTTCAGAAAACAGGGGAATGACTGCACCTGCTCAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTTGAGCCAGAAGCTTCTCTCAGAAATCCACCTGTCTGAAGAGA
CACCAGAGAGAGATCACCAGGACAACCTTCCAGTCTGACCAGGAAGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCAAGA
TCATTTCCAGAAGAAGACAAGACACTACTTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGCATG
TCTTCCCTCTCCCATGTCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG

(Continued)

Figure 57A

GTCTTCCAGGAGTTCACCTGATGGCTCATTACCCAGCCCCGTACAGAGGGGAACGAATGAGCAC
CTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGAAAC
CAGGCCFCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGG
GCTGAGCCAAGAAAGAACTTTGTGAAACCCAAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCACCCACCAAGGATGAGTTTACTGCAAGGCCGGGCATACTTCTCTGATGTGGACCTG
GAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGGCCACACCAACACCCCTGAACCC
GCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGGCCCTCTCTTCACCATCTTTGATGAAACCAAG
TCATGGTACTTCACCTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAAGAC
CCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCCTGCCGGG
CTTGTCTATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAACAT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCCCTG
TACAACCTCTACCCFGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAATGCCCTCATTTGGGGAGCACCTGCATGCTGGCATGTCACCCCTGTTTCTGGTCTACAGCAAC
AAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCCTCTGGC
CAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCA
ACCAAGGAGCCATTCCTTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTTCATGGCATCAAG
ACACAGGGGGCAAGACAGAAATTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTG
GATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCCTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAAACATCTTCAACCCCTCCCATCATTTGCCAGATACATCAGGCTG
CACCCACCCACTACTCAATCAGATCAACCCCTCAGGATGGAACCTGATGGGATGTGACCTGAACCTC
TGCTCAATGCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCTTAC
TTCACCAACATGTTTGGCACCTGGTCAACATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGCAAT
GCCTGGAGACCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATGAAA
GTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTCCTG
ATCTCTTCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTGTTT
CAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACCTACTGGACCCCCCTCTGACAAGATAC
CTGAGAAATCACCCCAAGTCTTGGGTCCACCAGATTTGCCCTGAGAAATGGAAGTCTTGGGATGTTGAG
GCACAAGACCTGTACTGA

Figure 57B

CS04-FL-NAml2 (SEQ ID NO: 109)

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCAGTACCACCTGTTCAACATTGCCAAACCCAGGCCACCCCTGGATG
GGACTCCTGGGACCCACCATTACAGGCTGAGGTGTATGACACTGTGGTTCGTCACCCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTGAGCTACTGGAAGTCCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCTTGGGAAGAGCCACACC
TATGTGTGGCAGGTCTCAAGGAGAATGGCCCCACTGCUCTCTGACCCACCCCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCCTGCACAAGTTCATTCCTCTGTTGTCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTTCTGGAGGGCCACACCTTCCCTGGTCAGGAACCCACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTCACTGCCAGACCCCTGCTGATGGACCTGGGACAGTTCCTGCTGTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCACAGAGGATTTGGACGC
AAGTACAAGAAAGTCAGGTTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCCCACTCCTGTATGGGGAGGTGGGGGACACCCCTGCTCATCATCTTC
AAGAACCAGGCCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAATCTGACCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCCTCTGGCCTGATTGGCCCACCTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTTCTTTCT
ATTTGGGGCCAGACTGACTTCCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCCTGACCCCTCTTCCCATTCTCTGGGGAGACTGTGTTTATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACCTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACAACCTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAATCCACCTGTCTGAAACGC
CACCAGAGGGAGATCACCAGGACCACCCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACCAGGACGAGAACCAGAGCCCAAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGCATG
AGCTCCAGCCCCCATGTCTCAGGAACAGGGCCAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG
GTCTTCCAAGAGTTCAGTGTGAGGAGCTTACCCAGCCCTGTACAGAGGGGAGCTGAATGAGCAC
CTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGCAAC

(Continued)

Figure 58A

CAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGG
GCTGAGCCACGCAAGAACCTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCCCCCACCAGGATGAGTTTGGACTGCAAGGCCCTGGGCCTACTTCTCTGATGTPGACCTG
GAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCT
GCCCATGGAAGGCAAGTGAAGTGTGCAGGAGTTTGGCCCTCTTCTTCCACCATCTTTGATGAAACCAAG
AGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAGGAC
CCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCCTGCCTGGG
CTTGTTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAACAT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGGCACAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAGTGCCTCATTTGGGGAGCACCTGCATGCTGGCATGAGCACCCCTGTTCCTGGTCTACAGCAAC
AAGTCCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGGAGC
ACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCATGATCATCCATGGCATCAAG
ACCCAGGGGGCCAGGCAGAAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GATGGCAAGAAATGGCAGACCTACAGAGGCCAACTCCACTGGAAACTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTG
CACCCACCCACTACAGCATCCGCAGCACCCCTCAGGATGGAGCTGATGGGCTGTGACCTGAACTCC
TGCAGCATGCCCTTGGGCATGGAGAGCAAGGCCATTTCTGATGCCAGATCACTGCCTCCAGCTAC
TTCACCAACATGTTTGGCACCTGGAGCCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAAGGAGCAAT
GCCTGGAGGGCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTCATGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTCCTG
ATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCCTCTTCTTCCAGAATGGCAAGGTCAAGGTGTT
CAGGGCAACCAGGACAGCTTCACCCCGTGTGGTGAACAGCCTGGACCCCCCTCTTGACCAGATAC
CTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCTTGGGATGTPGAG
GCCCAGGACCTGTACTGA

Figure 58B

**VIRAL VECTORS ENCODING
RECOMBINANT FVIII VARIANTS WITH
INCREASED EXPRESSION FOR GENE
THERAPY OF HEMOPHILIA A**

CROSS REFERENCE TO RELATED
APPLICATIONS

[0001] This application is a Continuation of U.S. application Ser. No. 15/349,930, filed Nov. 11, 2016, which claims priority to U.S. Provisional Patent Application No. 62/255,317, filed Nov. 13, 2015, the content of which are hereby incorporated by reference in its entirety for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Nov. 9, 2016, is named 008073_5107_US01_Sequence_Listing.txt and is 345 KB bytes in size.

BACKGROUND OF THE DISCLOSURE

[0003] Blood coagulation proceeds through a complex and dynamic biological pathway of interdependent biochemical reactions, referred to as the coagulation cascade. Coagulation Factor VIII (FVIII) is a key component in the cascade. Factor VIII is recruited to bleeding sites, and forms a Xase complex with activated Factor IX (FIXa) and Factor X (FX). The Xase complex activates FX, which in turn activates prothrombin to thrombin, which then activates other components in the coagulation cascade to generate a stable clot (reviewed in Saenko et al., *Trends Cardiovasc. Med.*, 9:185-192 (1999); Lenting et al., *Blood*, 92:3983-3996 (1998)).

[0004] Hemophilia A is a congenital X-linked bleeding disorder characterized by a deficiency in Factor VIII activity. Diminished Factor VIII activity inhibits a positive feedback loop in the coagulation cascade. This causes incomplete coagulation, which manifests as bleeding episodes with increased duration, extensive bruising, spontaneous oral and nasal bleeding, joint stiffness and chronic pain, and possibly internal bleeding and anemia in severe cases (Zhang et al., *Clinic. Rev. Allerg. Immunol.*, 37:114-124 (2009)).

[0005] Conventionally, hemophilia A is treated by Factor VIII replacement therapy, which consists of administering Factor VIII protein (e.g., plasma-derived or recombinantly-produced Factor VIII) to an individual with hemophilia A. Factor VIII is administered prophylactically to prevent or reduce frequency of bleeding episodes, in response to an acute bleeding episode, and/or perioperatively to manage bleeding during surgery. However, there are several undesirable features of Factor VIII replacement therapy.

[0006] First, Factor VIII replacement therapy is used to treat or manage hemophilia A, but does not cure the underlying Factor VIII deficiency. Because of this, individuals with hemophilia A require Factor VIII replacement therapy for the duration of their lives. Continuous treatment is expensive and requires the individual to maintain strict compliance, as missing only a few prophylactic doses can have serious consequences for individuals with severe hemophilia A.

[0007] Second, because Factor VIII has a relatively short half-life in vivo, conventional prophylactic Factor VIII replacement therapy requires administration every second or

third day. This places a burden on the individual to maintain compliance throughout their life. While third generation “long-acting” Factor VIII drugs may reduce the frequency of administration, prophylactic Factor FVIII replacement therapy with these drugs still requires monthly, weekly, or more frequent administration in perpetuity. For example, prophylactic treatment with ELOCTATE™ [Antihemophilic Factor (Recombinant), Fc Fusion Protein] requires administration every three to five days (ELOCTATE™ Prescribing Information, Biogen Idec Inc., (2015)). Moreover, the long-term effects of chemically modified biologics (e.g., pegylated polypeptides) are not yet fully understood.

[0008] Third, between 15% and 30% of all individuals receiving Factor VIII replacement therapy form anti-Factor VIII inhibitor antibodies, rendering the therapy inefficient. Factor VIII bypass therapy (e.g., administration of plasma-derived or recombinantly-produced prothrombin complex concentrates) can be used to treat hemophilia in individuals that form inhibitor antibodies. However, Factor VIII bypass therapy is less effective than Factor VIII replacement therapy (Mannucci P. M., *J Thromb Haemost.*, 1(7):1349-55 (2003)) and may be associated with an increased risk of cardiovascular complication (Luu and Ewenstein, *Haemophilia*, 10 Suppl. 2:10-16 (2004)).

[0009] Somatic gene therapy holds great promise for the treatment of hemophilia A because it would remedy the underlying under-expression functional Factor VIII activity (e.g., due to missense or nonsense mutations), rather than provide a one-time dose of Factor VIII activity to the individual. Because of this difference in the mechanism of action, as compared to Factor VIII replacement therapy, one-time administration of a Factor VIII gene therapy vector may provide an individual with Factor VIII for several years, reducing the cost of treatment and eliminating the need for continued patient compliance.

[0010] Coagulation Factor IX (FIX) gene therapy has been used effectively to treat individuals with hemophilia B, a related blood coagulation condition characterized by diminished Factor IX activity (Manno C. S., et al., *Nat Med.*, 12(3):342-47 (2006)). However, Factor VIII gene therapy presents several unique challenges. For example, the full-length, wild-type Factor VIII polypeptide (2351 amino acids; UniProt accession number P00451) is five times larger than the full-length, wild-type Factor IX polypeptide (461 amino acids; UniProt accession number P00740). As such, the coding sequence of wild-type Factor VIII is 7053 base pairs, which is too large to be packaged in conventional AAV gene therapy vectors. Further, reported recombinant expression of B-domain deleted variants of Factor VIII (BDD-FVIII) has been poor. As such, several groups have attempted to alter the codon usage of BDD-FVIII constructs, with limited success.

BRIEF SUMMARY OF DISCLOSURE

[0011] Accordingly, there is a need for Factor VIII variants whose coding sequences are more efficiently packaged into, and delivered via, gene therapy vectors. There is also a need for synthetic, codon-altered nucleic acids which express Factor VIII more efficiently. Such Factor VIII variants and codon-altered nucleic acids allow for improved treatment of Factor VIII deficiencies (e.g., hemophilia A). The above deficiencies and other problems associated with the treat-

ment of Factor VIII deficiencies (e.g., hemophilia A) are reduced or eliminated by the disclosed codon-altered Factor VIII variants.

[0012] In accordance with some embodiments, the present disclosure provides nucleic acids encoding Factor VIII variants that have high sequence identity to the disclosed codon-altered sequences of the Factor VIII heavy chain (e.g., CS01-HC-NA, CS04-HC-NA, or CS23-HC-NA) and light chain (CS01-LC-NA, CS04-LC-NA, or CS23-LC-NA). In some embodiments, these nucleic acids further include a sequence encoding a linker sequence that replaces the native Factor VIII B-domain (e.g., a linker sequences comprising a furin cleavage site), between the sequences coding for the Factor VIII heavy and light chains.

[0013] In one aspect, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS04-HC-NA (SEQ ID NO: 3). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS04-LC-NA (SEQ ID NO: 4). The polypeptide linker comprises a furin cleavage site.

[0014] In one embodiment of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO04 (SEQ ID NO: 6).

[0015] In one aspect, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS01-HC-NA (SEQ ID NO: 24). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS01-LC-NA (SEQ ID NO: 25). The polypeptide linker comprises a furin cleavage site.

[0016] In one embodiment of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO01 (SEQ ID NO: 5).

[0017] In one aspect, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS23-HC-NA (SEQ ID NO: 22). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS23-LC-NA (SEQ ID NO: 23). The polypeptide linker comprises a furin cleavage site.

[0018] In one embodiment of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO23 (SEQ ID NO: 7).

[0019] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the

heavy chain of the Factor VIII polypeptide has at least 96% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 96% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0020] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 97% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 97% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0021] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 98% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 98% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0022] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0023] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99.5% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99.5% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0024] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99.9% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99.9% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0025] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS04-HC-NA (SEQ ID NO: 3), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS04-LC-NA (SEQ ID NO: 4).

[0026] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS01-HC-NA (SEQ ID NO: 24), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS01-LC-NA (SEQ ID NO: 25).

[0027] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS23-HC-NA (SEQ ID NO: 22), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS23-LC-NA (SEQ ID NO: 23).

[0028] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0029] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0030] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0031] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 96% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0032] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 97% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0033] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 98% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0034] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0035] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0036] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.9% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0037] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS04-FL-NA (SEQ ID NO: 1).

[0038] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS01-FL-NA (SEQ ID NO: 13).

[0039] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS23-FL-NA (SEQ ID NO: 20).

[0040] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 95% identity to CS04-FL-AA (SEQ ID NO: 2).

[0041] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 96% identity to CS04-FL-AA (SEQ ID NO: 2).

[0042] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 97% identity to CS04-FL-AA (SEQ ID NO: 2).

[0043] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 98% identity to CS04-FL-AA (SEQ ID NO: 2).

[0044] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99% identity to CS04-FL-AA (SEQ ID NO: 2).

[0045] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99.5% identity to CS04-FL-AA (SEQ ID NO: 2).

[0046] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99.9% identity to CS04-FL-AA (SEQ ID NO: 2).

[0047] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising the amino acid sequence of CS04-FL-AA (SEQ ID NO: 2).

[0048] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-SC1-NA (SEQ ID NO: 9), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0049] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-SC2-NA (SEQ ID NO: 11), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0050] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-SC1-NA (SEQ ID NO: 26), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0051] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-SC2-NA (SEQ ID NO: 27), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0052] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-SC1-NA (SEQ ID NO: 28), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0053] In one aspect, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-SC2-NA (SEQ ID NO: 29), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0054] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 96% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0055] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 97% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0056] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 98% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0057] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0058] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0059] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.9% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0060] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS04-SC1-NA (SEQ ID NO: 9).

[0061] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS04-SC2-NA (SEQ ID NO: 11).

[0062] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS01-SC1-NA (SEQ ID NO: 26).

[0063] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS01-SC2-NA (SEQ ID NO: 27).

[0064] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS23-SC1-NA (SEQ ID NO: 28).

[0065] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS23-SC2-NA (SEQ ID NO: 29).

[0066] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 95% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0067] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 96% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA,

CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0074] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide comprises a glycosylation polypeptide positioned between two consecutive amino acids.

[0075] In one embodiment of the polynucleotides described above, the encoded polypeptide linker includes a glycosylation peptide with an amino acid sequence having at least 92% identity to a glycosylation peptide selected from the group consisting of NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA and NGV-AA.

[0076] In one embodiment of the polynucleotides described above, the encoded polypeptide linker comprises a glycosylation peptide with an amino acid sequence selected from the group consisting of NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA and NGV-AA.

[0077] In one embodiment of the polynucleotides described above, the glycosylation peptide is encoded by a polynucleotide with a nucleotide sequence having at least 95% identity to a sequence selected from the group consisting of NG1-NA, NG4-NA, NG5-NA, NG6-NA, NG7-NA, NG9-NA, NG10-NA, NG16-NA, NG17-NA, NG18-NA, NG19-NA, NG20-NA, NG21-NA and NGV-NA.

[0078] In one embodiment of the polynucleotides described above, the glycosylation peptide is encoded by a polynucleotide with a nucleotide sequence selected from one of NG1-NA, NG4-NA, NG5-NA, NG6-NA, NG7-NA, NG9-NA, NG10-NA, NG16-NA, NG17-NA, NG18-NA, NG19-NA, NG20-NA, NG21-NA and NGV-NA.

[0079] In one embodiment of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to a sequence selected from the group consisting of BDLNG1-NA, BDLNG3-NA, BDLNG5-NA, BDLNG6-NA, BDLNG9-NA, BDLNG10-NA, BDLNG16-NA, BDLNG17-NA, BDLNG18-NA, BDLNG19-NA, BDLNG20-NA and BDLNG21-NA.

[0080] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes an F328S (SPI, F309S SPE) amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0081] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes I105V, A127S, G151K, M166T, and L171P (SPI; I86V, A108S, G132K, M147T, and L152P, SPE, respectively) amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0082] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and b) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some embodiments (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0083] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S (SPI; F309S SPE) amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), and b) C1918G and C1922G (SPI; C1899G and C1903 SPE, respectively) amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0084] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S (SPI; F309S SPE) amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), and b) I105V, A127S, G151K, M166T, and L171P (SPI; I86V, A108S, G132K, M147T, and L152P, SPE, respectively) amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0085] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), b) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and c) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some embodiments (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0086] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), b) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and c) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some embodiments (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0087] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), b) C1918G and C1922G amino acid substitutions, relative to FVIII-FL-AA (SEQ ID

NO: 19), and c) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0088] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), b) C1918G and C1922G amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), c) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and d) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some embodiments (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0089] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), b) an F328S amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), c) C1918G and C1922G amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), d) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and e) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some embodiments (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0090] In one embodiment of the polynucleotides described above, the polynucleotide also includes a promoter element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0091] In one embodiment of the polynucleotides described above, the polynucleotide also includes an enhancer element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0092] In one embodiment of the polynucleotides described above, the polynucleotide also includes a polyadenylation element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0093] In one embodiment of the polynucleotides described above, the polynucleotide also includes an intron operatively linked to the nucleotide sequence encoding the Factor VIII polypeptide.

[0094] In one embodiment of the polynucleotides described above, the intron is positioned between a promoter element and the translation initiation site (e.g., the first coding ATG) of the nucleotide sequence encoding a Factor VIII polypeptide.

[0095] In another aspect, the disclosure provides a mammalian gene therapy vector including a polynucleotide as described above.

[0096] In one embodiment of the mammalian gene therapy vector described above, the mammalian gene therapy vector is an adeno-associated virus (AAV) vector.

[0097] In one embodiment of the mammalian gene therapy vector described above, the AAV vector is an AAV-8 vector.

[0098] In another aspect, the disclosure provides a method for treating hemophilia A including administering, to a patient in need thereof, a mammalian gene therapy vector as described above.

[0099] In another aspect, the disclosure provides a mammalian gene therapy vector as described above for treating hemophilia A.

[0100] In another aspect, the disclosure provides the use of a mammalian gene therapy vector as described above for the manufacture of a medicament for treating hemophilia A.

[0101] In another aspect, the disclosure provides a Factor VIII polypeptide including a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-HC-AAm23. The light chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-LC-AAm23. The polypeptide linker of the Factor VIII polypeptide includes a furin cleavage site. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0102] In another aspect, the disclosure provides a Factor VIII polypeptide including a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-HC-AAm123. The light chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-LC-AAm123. The polypeptide linker of the Factor VIII polypeptide includes a furin cleavage site. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19), and iv) an F328S amino acid substitution.

[0103] In another aspect, the disclosure provides a Factor VIII polypeptide including a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-HC-AAm234. The light chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-LC-AAm234. The polypeptide linker of the Factor VIII polypeptide includes a furin cleavage site. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19), and iv) F328S/C1918G/C1922G amino acid substitutions.

[0104] In one embodiment of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 96% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 96%

identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[0105] In one embodiment of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 97% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 97% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[0106] In one embodiment of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 98% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 98% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[0107] In one embodiment of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 99% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 99% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[0108] In one embodiment of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 99.5% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 99.5% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[0109] In one embodiment of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[0110] In one embodiment of the Factor VIII polypeptides described above, the polypeptide linker has at least 95% identity to BDL-SQ-AA (SEQ ID NO: 30).

[0111] In one embodiment of the Factor VIII polypeptides described above, the polypeptide linker has the amino acid sequence of BDL-SQ-AA (SEQ ID NO: 30).

[0112] In one embodiment of the Factor VIII polypeptides described above, the polypeptide linker includes a glycosylation peptide with an amino acid sequence having at least 92% identity to a glycosylation peptide selected from the group consisting of NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA and NGV-AA.

[0113] In one embodiment of the Factor VIII polypeptides described above, the polypeptide linker includes a glycosylation peptide selected from the group consisting of NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA and NGV-AA.

[0114] In one embodiment of the Factor VIII polypeptides described above, the polypeptide linker has an amino acid

sequence having at least 95% identity to a sequence selected from the group consisting of BDLNG1-AA, BDLNG3-AA, BDLNG5-AA, BDLNG6-AA, BDLNG9-AA, BDLNG10-AA, BDLNG16-AA, BDLNG17-AA, BDLNG18-AA, BDLNG19-AA, BDLNG20-AA and BDLNG21-AA.

[0115] In one embodiment of the Factor VIII polypeptides described above, the polypeptide linker has an amino acid sequence selected from the group consisting of BDLNG1-AA, BDLNG3-AA, BDLNG5-AA, BDLNG6-AA, BDLNG9-AA, BDLNG10-AA, BDLNG16-AA, BDLNG17-AA, BDLNG18-AA, BDLNG19-AA, BDLNG20-AA and BDLNG21-AA.

[0116] In another aspect, the disclosure provides a Factor VIII polypeptide having an amino acid sequence with at least 95% identity to CS40-FL-AAm23 (SEQ ID NO: 104). The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0117] In another aspect, the disclosure provides a Factor VIII polypeptide having an amino acid sequence with at least 95% identity to CS40-FL-AAm123. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19), and iv) an F328S amino acid substitution.

[0118] In another aspect, the disclosure provides a Factor VIII polypeptide having an amino acid sequence with at least 95% identity to CS40-FL-AAm234. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19), and iv) F328S/C1918G/C1922G amino acid substitutions.

[0119] In one embodiment of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 96% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[0120] In one embodiment of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 97% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[0121] In one embodiment of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 98% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[0122] In one embodiment of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 99% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[0123] In one embodiment of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 99.5% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[0124] In one embodiment of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

BRIEF DESCRIPTION OF DRAWINGS

[0125] FIG. 1 shows schematic illustrations of the wild-type and ReFacto-type human Factor VIII protein constructs.

[0126] FIGS. 2A and 2B show the CS04 codon-altered nucleotide sequence (SEQ ID NO: 1) encoding a Factor VIII variant in accordance with some embodiments (“CS04-FL-NA” for full-length coding sequence).

[0127] FIG. 3 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 2) encoded by the CS04 codon-altered nucleotide sequence in accordance with some embodiments (“CS04-FL-AA” for full-length amino acid sequence).

[0128] FIG. 4 shows the portion of the CS04 codon-altered nucleotide sequence (SEQ ID NO: 3) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments (“CS04-HC-NA”).

[0129] FIG. 5 shows the portion of the CS04 codon-altered nucleotide sequence (SEQ ID NO: 4) encoding the light chain of a Factor VIII variant in accordance with some embodiments (“CS04-LC-NA”).

[0130] FIG. 6 shows exemplary coding sequences (SEQ ID NOS 5-7 and 36-48, respectively, in order of appearance) for B-domain substituted linkers in accordance with some embodiments. BDLO01 (SEQ ID NO: 5), BDLO04 (SEQ ID NO: 6), and BDLO23 (SEQ ID NO: 7) are the respective portions of the CS01, CS04, and CS23 codon-altered nucleotide sequences that encode a B-domain substituted linker, respectively.

[0131] FIGS. 7A, 7B, and 7C show an AAV vector sequence (SEQ ID NO: 8) containing an CS04 codon-altered nucleotide sequence in accordance with some embodiments (“CS04-AV-NA”).

[0132] FIGS. 8A and 8B show the CS01m1 codon-altered nucleotide sequence (SEQ ID NO: 49) encoding a Factor VIII variant with an F328S amino acid substitution in accordance with some embodiments (“CS01m1-FL-NA”).

[0133] FIGS. 9A and 9B show the CS04Δ(760-1667) (SPI; CS04Δ(741-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 9) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS04-SC1-NA”).

[0134] FIG. 10 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 10) encoded by the CS01Δ(760-1667) (SPI; CS01Δ(741-1648), SPE), CS04Δ(760-1667) (SPI; CS04Δ(741-1648), SPE), and CS23Δ(760-1667) (SPI; CS23Δ(741-1648), SPE) codon-altered nucleotide sequences in accordance with some embodiments (“CS01-SC1-AA,” “CS04-SC1-AA,” and “CS23-SC1-AA,” respectively).

[0135] FIGS. 11A and 11B show the CS04Δ(772-1667) (SPI; CS04Δ(753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 11) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS04-SC2-NA”).

[0136] FIG. 12 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 12) encoded by the CS01Δ(772-1667) (SPI; CS01Δ(753-1648), SPE), CS04Δ(772-1667) (SPI; CS04Δ(753-1648), SPE), and CS23Δ(772-1667) (SPI;

CS23Δ(753-1648), SPE) codon-altered nucleotide sequence in accordance with some embodiments (“CS01-SC2-AA,” “CS04-SC2-AA,” and “CS23-SC2-AA,” respectively).

[0137] FIGS. 13A and 13B show amino acid and nucleotide sequences for exemplary glycosylation peptides that are inserted into the B-domain substituted linker in accordance with some embodiments. “NG1” or NG1-AA” is the code for the amino acid sequence, shown in the top line. “NG1-NA” is the code for the nucleic acid sequence, shown in the bottom line for each set. FIGS. 13A and 13B disclose the amino acid sequences as SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, and the nucleotide sequences as SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, all respectively, in order of appearance.

[0138] FIG. 14 shows the results of in silico prediction of in vivo N-glycosylation of the wild-type Factor VIII B-domain. Figure discloses SEQ ID NOS 76 and 76-82, respectively, in order of appearance.

[0139] FIG. 15 shows the results of in silico prediction of in vivo N-glycosylation of the V3 peptide linker. Figure discloses SEQ ID NOS 83 and 83-89, respectively, in order of appearance.

[0140] FIGS. 16A and 16B show the CS01 codon-altered nucleotide sequence (SEQ ID NO: 13) encoding a Factor VIII variant in accordance with some embodiments (“CS01-FL-NA”).

[0141] FIGS. 17A and 17B show the CS08 codon-altered nucleotide sequence (SEQ ID NO: 14) encoding a Factor VIII variant in accordance with some embodiments (“CS08-FL-NA”).

[0142] FIGS. 18A and 18B show the CS10 codon-altered nucleotide sequence (SEQ ID NO: 15) encoding a Factor VIII variant in accordance with some embodiments (“CS10-FL-NA”).

[0143] FIGS. 19A and 19B show the CS11 codon-altered nucleotide sequence (SEQ ID NO: 16) encoding a Factor VIII variant in accordance with some embodiments (“CS11-FL-NA”).

[0144] FIGS. 20A and 20B show the CS40 wild-type ReFacto coding sequence (SEQ ID NO: 17), in accordance with some embodiments (“CS40-FL-NA”).

[0145] FIGS. 21A and 21B show the CH25 codon-altered nucleotide sequence (SEQ ID NO: 18) encoding a Factor VIII variant in accordance with some embodiments (“CH25-FL-NA”).

[0146] FIG. 22 shows a wild-type human Factor VIII amino acid sequence (SEQ ID NO: 19), in accordance with some embodiments (“FVIII-FL-AA”).

[0147] FIG. 23 illustrates the scheme for cloning the pCS40, pCS01, pCS04, pCS08, pCS10, pCS11, and pCh25 constructs, by inserting synthetic Refacto-type BDD-FVIII DNA sequences into the vector backbone pCh-BB01 via *AscI* and *NotI* restriction sites.

[0148] FIG. 24 shows the integrity of AAV vector genome preparations, as analyzed by agarose gel electrophoresis. Lane 1, DNA marker; lane 2, vCS40; lane 3, vCS01; lane 4, vCS04. The AAV vectors have all the same-sized genomes, migrating at approximately 5 kb (arrow, right side). The scale on the left side indicates size of the DNA fragments in kilobases (kb).

[0149] FIG. 25 shows the protein analysis of AAV vector preparations by PAGE and silver staining. Lane 1, protein marker (M); lane 2, vCS40, lane 3, vCS01; and lane 4, vCS04. The constructs all have the same AAV8 capsids

consisting of VP1, VP2, and VP3 (arrows right side). The scale on the left side indicates size of the protein marker in kilodaltons (kDa).

[0150] FIGS. 26A and 26B show the CS23 codon-altered nucleotide sequence (SEQ ID NO: 20) encoding a Factor VIII variant in accordance with some embodiments (“CS23-FL-NA”).

[0151] FIG. 27 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 21) encoded by the CS23 codon-altered nucleotide sequence in accordance with some embodiments (“CS23-FL-AA”).

[0152] FIG. 28 shows the portion of the CS23 codon-altered nucleotide sequence (SEQ ID NO: 22) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments (“CS23-HC-NA”).

[0153] FIG. 29 shows the portion of the CS23 codon-altered nucleotide sequence (SEQ ID NO: 23) encoding the light chain of a Factor VIII variant in accordance with some embodiments (“CS23-LC-NA”).

[0154] FIGS. 30A and 30B show the CS01m13 codon-altered nucleotide sequence (SEQ ID NO: 90) encoding a Factor VIII variant with m1 (F328S) and m3 amino acid substitutions in accordance with some embodiments (“CS01-FL-NA-m13”).

[0155] FIGS. 31A and 31B show the CS01m23 codon-altered nucleotide sequence (SEQ ID NO: 91) encoding a Factor VIII variant with the m2 and m3 mutation sets in accordance with some embodiments (“CS01-FL-NA-m23”).

[0156] FIGS. 32A and 32B show the CS01m3 codon-altered nucleotide sequence (SEQ ID NO: 92) encoding a Factor VIII variant with m3 amino acid substitutions in accordance with some embodiments (“CS01-FL-NA-m3”).

[0157] FIGS. 33A and 33B show the CS01m2 codon-altered nucleotide sequence (SEQ ID NO: 93) encoding a Factor VIII variant with the m2 mutation set (I105V/A127S/G151K/M166T/L171P (SPI)) amino acid substitutions in accordance with some embodiments (“CS01-FL-NA-m2”).

[0158] FIGS. 34A and 34B show the CS04m2 codon-altered nucleotide sequence (SEQ ID NO: 94) encoding a Factor VIII variant with the m2 mutants (I105V/A127S/G151K/M166T/L171P (SPI)) amino acid substitutions in accordance with some embodiments (“CS01-FL-NA-m2”).

[0159] FIGS. 35A and 35B show the CS04m3 codon-altered nucleotide sequence (SEQ ID NO: 95) encoding a Factor VIII variant with m3 amino acid substitutions in accordance with some embodiments (“CS04-FL-NA-m3”).

[0160] FIGS. 36A and 36B show the CS04m23 codon-altered nucleotide sequence (SEQ ID NO: 96) encoding a Factor VIII variant with the m2 mutant set (I105V/A127S/G151K/M166T/L171P (SPI)) and m3 amino acid substitutions in accordance with some embodiments (“CS04-FL-NA-m23”).

[0161] FIGS. 37A and 37B show the CS04m1 codon-altered nucleotide sequence (SEQ ID NO: 97) encoding a Factor VIII variant with an m1 (F328S) amino acid substitution in accordance with some embodiments (“CS04-FL-NA-m1”).

[0162] FIGS. 38A and 38B show the CS04m13 codon-altered nucleotide sequence (SEQ ID NO: 98) encoding a Factor VIII variant with m1 and m3 amino acid substitutions in accordance with some embodiments (“CS04-FL-NA-m13”).

[0163] FIGS. 39A and 39B show the CS23m13 codon-altered nucleotide sequence (SEQ ID NO: 99) encoding a

Factor VIII variant with m1 and m3 amino acid substitutions in accordance with some embodiments (“CS23m13-FL-NA”).

[0164] FIGS. 40A and 40B show the CS23m3 codon-altered nucleotide sequence (SEQ ID NO: 100) encoding a Factor VIII variant with m3 amino acid substitutions in accordance with some embodiments (“CS23-FL-NA-m3”).

[0165] FIGS. 41A and 41B show the CS23m2 codon-altered nucleotide sequence (SEQ ID NO: 101) encoding a Factor VIII variant with the m2 mutant set (I105V/A127S/G151K/M166T/L171P) amino acid substitutions in accordance with some embodiments (“CS23-FL-NA-m2”).

[0166] FIGS. 42A and 42B show the CS23m1 codon-altered nucleotide sequence (SEQ ID NO: 102) encoding a Factor VIII variant with an m1 (F328S) amino acid substitution in accordance with some embodiments (“CS23-FL-NA-m1”).

[0167] FIGS. 43A and 43B show the CS23m23 codon-altered nucleotide sequence (SEQ ID NO: 103) encoding a Factor VIII variant with the m2 mutant set (I105V/A127S/G151K/M166T/L171P) and m3 amino acid substitutions in accordance with some embodiments (“CS23-FL-NA-m23”).

[0168] FIG. 44 depicts cloning of the pCS constructs, done by inserting synthetic Refacto-type BDD-FVIII carrying different mutations (see inserted table) into the vector backbone pCh-BB01 via *AscI* and *NotI* restriction sites.

[0169] FIG. 45 depicts the protein analysis of AAV vector preparations by PAGE and silver staining. Lane 1, protein marker (M); lane 2, vCS01, lane 3, vCS17; lane 4, vCS19; lane 5, vCS20; lane 6, vCS40; lane 7, vCS04; lane 8, vCS17; lane 9, vCS24 construct. The constructs have all the same AAV8 capsids consisting of VP1, VP2 and VP3 (arrows right side). The scale on the left side indicates size of the protein marker in kilo Daltons (kDa).

[0170] FIG. 46 depicts the integrity of AAV vector genome preparations analyzed by agarose gel electrophoresis. Lane 1, DNA marker (M); lane 2, vCS04, lane 3, vCS17; lane 4, vCS20; lane 5, vCS24; lane 6, vCS16; lane 7, vCS40 construct. Vector load is 1.5E10 vg per lane. The AAV vectors have the same-sized genomes, migrating at approximately 5 kb (arrow, right side). The scale on the left side indicates size of the DNA fragments in kilobases (kb).

[0171] FIG. 47 shows the portion of the CS01 codon-altered nucleotide sequence (SEQ ID NO: 24) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments (“CS01-HC-NA”).

[0172] FIG. 48 shows the portion of the CS01 codon-altered nucleotide sequence (SEQ ID NO: 25) encoding the light chain of a Factor VIII variant in accordance with some embodiments (“CS01-LC-NA”).

[0173] FIGS. 49A and 49B show the CS01Δ(760-1667) (SPI; CS01Δ(741-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 26) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS01-SC1-NA”).

[0174] FIGS. 50A and 50B show the CS01Δ(772-1667) (SPI; CS01Δ(753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 27) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS01-SC2-NA”).

[0175] FIGS. 51A and 51B show the CS23Δ(760-1667) (SPI; CS23Δ(741-1648), SPE) codon-altered nucleotide

sequence (SEQ ID NO: 28) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS23-SC1-NA”).

[0176] FIGS. 52A and 52B show the CS23Δ(772-1667) (SPI; CS23Δ(753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 29) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS23-SC2-NA”).

[0177] FIG. 53 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 104) encoded by the CS01m23 codon-altered nucleotide sequence in accordance with some embodiments (“CS01m23-FL-AA”).

[0178] FIG. 54 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 105) encoded by the CS04m3 codon-altered nucleotide sequence in accordance with some embodiments (“CS01m23-FL-AA”).

[0179] FIG. 55 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 106) encoded by the CS01m12 codon-altered nucleotide sequence in accordance with some embodiments (“CS01m12-FL-AA”).

[0180] FIG. 56 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 107) encoded by the CS04m12 codon-altered nucleotide sequence in accordance with some embodiments (“CS04m12-FL-AA”).

[0181] FIGS. 57A and 57B show the CS01m12 codon-altered nucleotide sequence (SEQ ID NO: 108) encoding a Factor VIII variant with m1 (F328S) and m2 amino acid substitutions in accordance with some embodiments (“CS01-FL-NA_{m12}”).

[0182] FIGS. 58A and 58B show the CS04m12 codon-altered nucleotide sequence (SEQ ID NO: 109) encoding a Factor VIII variant with m1 (F328S) and m2 amino acid substitutions in accordance with some embodiments (“CS04-FL-NA_{m12}”).

DETAILED DESCRIPTION OF DISCLOSURE

I. Introduction

[0183] AAV-based gene therapy holds great promise for the treatment of hemophiliacs. For hemophilia B, first clinical data are encouraging in that FIX levels of about 10% can be maintained in at least some patients for more than 1 year. For hemophilia A however, achieving therapeutic expression levels of 5-10% with AAV vectors remains challenging for various reasons. First, the Factor VIII coding sequence is too large for conventional AAV-based vectors. Second, engineered B-domain deleted or truncated Factor VIII constructs suffer from poor expression in vivo, even when codon-optimized. Third, these B-domain deleted or truncated Factor VIII variant constructs have short half-lives in vivo, exacerbating the effects of poor expression. Fourth, even when expressed, FVIII is not efficiently secreted from cells, as are other coagulation factors, such as Factor IX.

[0184] Moreover, these challenges cannot be addressed by simply administering higher doses of the gene therapy construct. According to current knowledge, the vector dose of an AAV-based gene therapy vector should be increased above 2×10^{12} vg/kg bodyweight. This is because at such high doses a T cell immune response is triggered, which destroys transduced cells and, as a consequence, transgene expression is reduced or even eliminated. Therefore, strategies to improve the expression of FVIII are needed to make FVIII gene therapy a viable therapeutic option for hemophilia A patients.

[0185] The present disclosure relates to the discovery of codon-altered Factor VIII variant coding sequences that solve these and other problems associated with Factor VIII gene therapy. For example, the polynucleotides disclosed herein provide markedly improved expression in mammalian cells, and display improved virion packaging due to stabilized packing interactions. In some implementations, these advantages are realized by using coding sequences for the heavy and light chains of Factor VIII with high sequence identity to the codon altered CS01, CS04, and CS23 constructs (e.g., with high sequence identity to one of the CS01-HC, CS04-HC, and CS23-HC heavy chain coding sequences and high sequence identity to one of the CS01-LC, CS04-LC, and CS23-LC light chain coding sequences).

[0186] In some implementations, the Factor VIII molecules encoded by the polynucleotides described herein have been shortened by truncating, deleting, or replacing the wild-type B-domain. As such, the polynucleotides are better suited for expressing Factor VIII via conventional gene therapy vectors, which inefficiently express larger polypeptides, such as the wild-type Factor VIII.

[0187] Advantageously, it is shown herein that the CS01, CS04, and CS23 codon-altered Factor VIII variant coding sequences provide superior expression of a B-domain deleted Factor VIII construct in vivo. For example, it is demonstrated in Example 2 and Example 4 that intravenous administration of AAV-based gene therapy vectors having the CS01 (SEQ ID NO: 13), CS04 (SEQ ID NO: 1), and CS23 (SEQ ID NO: 20) coding sequence provide 18-fold, 74-fold, and 30-fold increases in Factor VIII expression, relative to the corresponding CS40 construct encoded with the wild-type polynucleotide sequence (SEQ ID NO: 17), in Factor VIII knock-out mice (Table 4 and Table 7).

[0188] Further, it also shown herein that the CS01 and CS04 codon-altered Factor VIII variant coding sequences provide superior virion packaging and virus production. For example, it is demonstrated in Example 1 that AAV vector constructs containing the CS01 and CS04 constructs provided 5 to 7-fold greater viral yield, relative to the corresponding CS40 construct encoded with the wild-type polynucleotide sequence, when isolated from the same amount of cell pellet.

[0189] Advantageously, Applicants also found that the improved Factor VIII activity generated from the CS01, CS04, and CS23 codon altered sequences could be further enhanced by introducing mutations into the underlying Factor VIII polypeptide sequence. For example, as demonstrated in Example 4, the F328S, X5, and X1 mutations, alone and in combination with one another, further increased FVIII activity when expressed in vivo in the CS01 or CS04 codon altered background 2 to 7-fold, relative to the wild type, codon altered constructs (Table 7). More strikingly, these codon altered sequences, encoding the mutant Factor VIII mutants, provided up to 246-fold greater increase as compared to the corresponding CS40 construct encoded with the wild-type polynucleotide sequence (Table 7).

II. Definitions

[0190] As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0191] As used herein, the terms “Factor VIII” and “FVIII” are used interchangeably, and refer to any protein with Factor VIII activity (e.g., active FVIII, often referred to as FVIIIa) or protein precursor (e.g., pro-protein or pre-pro-

protein) of a protein with Factor VIII activity, particularly Factor IXa cofactor activity. In an exemplary embodiment, a Factor VIII polypeptide refers to a polypeptide that has sequences with high sequence identity (e.g., at least 70%, 75%, 80%, 85%, 90%, 95%, 99%, or more) to the heavy and light chains of a wild type Factor VIII polypeptide. In some embodiments, the B-domain of a Factor VIII polypeptide is deleted, truncated, or replaced with a linker polypeptide to reduce the size of the polynucleotide encoding the Factor VIII polypeptide. In an exemplary embodiment, amino acids 20-1457 of SEQ ID NO: 2 constitute a Factor VIII polypeptide.

[0192] Non-limiting examples of wild type Factor VIII polypeptides include human pre-pro-Factor VIII (e.g., GenBank accession nos. AAA52485, CAA25619, AAA58466, AAA52484, AAA52420, AAV85964, BAF82636, BAG36452, CAI41660, CAI41666, CAI41672, CAI43241, CA003404, EAW72645, AAH22513, AAH64380, AAH98389, AAI11968, AAI11970, or AAB61261), corresponding pro-Factor VIII, and natural variants thereof; porcine pre-pro-Factor VIII (e.g., UniProt accession nos. F1RZ36 or K7GSZ5), corresponding pro-Factor VIII, and natural variants thereof; mouse pre-pro-Factor VIII (e.g., GenBank accession nos. AAA37385, CAM15581, CAM26492, or EDL29229), corresponding pro-Factor VIII, and natural variants thereof; rat pre-pro-Factor VIII (e.g., GenBank accession no. AAQ21580), corresponding pro-Factor VIII, and natural variants thereof; rat pre-pro-Factor VIII; and other mammalian Factor VIII homologues (e.g., monkey, ape, hamster, guinea pig, etc.).

[0193] As used herein, a Factor VIII polypeptide includes natural variants and artificial constructs with Factor IX cofactor activity. As used in the present disclosure, Factor VIII encompasses any natural variants, alternative sequences, isoforms, or mutant proteins that retain some basal Factor IX cofactor activity (e.g., at least 5%, 10%, 25%, 50%, 75%, or more of the corresponding wild type activity). Examples of Factor VIII amino acid variations (relative to FVIII-FL-AA (SEQ ID NO: 19)) found in the human population include, without limitation, S19R, R22T, Y24C, Y25C, L26P/R, E30V, W33G, Y35C/H, G41C, R48C/K, K67E/N, L69P, E72K, D75E/V/Y, P83R, G89D/V, G92A/V, A97P, E98K, V99D, D101G/H/V, V104D, K108T, M110V, A111T/V, H113R/Y, L117F/R, G121S, E129V, G130R, E132D, Y133C, D135G/Y, T137A/I, S138R, E141K, D145H, V147D, Y155H, V159A, N163K, G164D/V, P165S, C172W, S176P, S179P, V181E/M, K185T, D186G/N/Y, S189L, L191F, G193R, L195P, C198G, S202N/R, F214V, L217H, A219D/T, V220G, D222V, E223K, G224W, T252I, V253F, N254I, G255V, L261P, P262L, G263S, G266F, C267Y, W274C, H275L, G278R, G280D, E284K, V285G, E291G/K, T294I, F295L, V297A, N299I, R301C/H/L, A303E/P, I307S, S308L, F312S, T314A/I, A315V, G323E, L326P, L327P/V, C329F, I331V, M339I, E340K, V345A/L, C348R/S/Y, Y365C, R391C/H/P, S392L/P, A394S, W401G, I405F/S, E409G, W412G/R, K427I, L431F/S, R437P/W, I438F, G439D/S/V, Y442C, K444R, Y450D/N, T454I, F455C, G466E, P470L/R/T, G474E/R/V, E475K, G477V, D478N, T479R, F484C, A488G, R490G, Y492C/H, Y492H, I494T, P496R, G498R, R503H, G513S/V, I522Y, K529E, W532G, P540I, T541S, D544N, R546W, R550C/G/H, S553P, S554C/G, V556D, R560T, D561G/H/Y, I567T, P569R, S577F, V578A, D579A/H, N583S, Q584H/K/R, I585R/T, M586V, D588G/Y,

L594Q, S596P, N601D/K, R602G, S603I/R, W604C, Y605H/S, N609I, R612C, N631K/S, M633I, S635N, N637D/I/S, Y639C, L644V, L650F, V653A/M, L659P, A663V, Q664P, F677L, M681I, V682F, Y683C/N, T686R, F698L, M699T/V, M701I, G705V, G710W, N713I, R717L/W, G720D/S, M721I/L, A723T, L725Q, V727F, E739K, Y742C, R795G, P947R, V1012L, E1057K, H1066Y, D1260E, K1289Q, Q1336K, N1460K, L1481P, A1610S, I1698T, Y1699C/F, E1701K, Q1705H, R1708C/H, T1714S, R1715G, A1720V, E1723K, D1727V, Y1728C, R1740G, K1751Q, F1762L, R1768H, G1769R, L1771P, L1775F/V, L1777P, G1779E/R, P1780L, I1782R, D1788H, M1791T, A1798P, S1799H, R1800C/G/H, P1801A, Y1802C, S1803Y, F1804S, L1808F, M1842I, P1844S, T1845P, E1848G, A1853T/V, S1858C, K1864E, D1865N/Y, H1867P/R, G1869D/V, G1872E, P1873R, L1875P, V1876L, C1877R/Y, L1882P, R1888I, E1894G, I1901F, E1904D/K, S1907C/R, W1908L, Y1909C, A1939T/V, N1941D/S, G1942A, M1945V, L1951F, R1960L/Q, L1963P, S1965I, M1966I/V, G1967D, S1968R, N1971T, H1973L, G1979V, H1980P/Y, F1982I, R1985Q, L1994P, Y1998C, G2000A, T2004R, M2007I, G2013R, W2015C, R2016P/W, E2018G, G2022D, G2028R, S2030N, V2035A, Y2036C, N2038S, 2040Y, G2045E/V, I2051S, I2056N, A2058P, W2065R, P2067L, A2070V, S2082N, S2088F, D2093G/Y, H2101D, T2105N, Q2106E/P/R, G2107S, R2109C, I2117F/S, Q2119R, F2120C/L, Y2124C, R2135P, S2138Y, T2141N, M2143V, F2145C, N2148S, N2157D, P2162L, R2169C/H, P2172L/Q/R, T2173A/I, H2174D, R2178C/H/L, R2182C/H/P, M2183R/V, L2185S/W, S2192I, C2193G, P2196R, G2198V, E2200D, I2204T, I2209N, A2211P, A2220P, P2224L, R2228G/L/P/Q, L2229F, V2242M, W2248C/S, V2251A/E, M2257V, T2264A, Q2265R, F2279C/I, I2281T, D2286G, W2290L, G2304V, D2307A, P2319L/S, R2323C/G/H/L, R2326G/L/P/Q, Q2330P, W2332R, I2336F, R2339T, G2344C/D/S, and C2345S/Y. Factor VIII proteins also include polypeptides containing post-translational modifications.

[0194] Generally, polynucleotides encoding Factor VIII encode for an inactive single-chain polypeptide (e.g., a pre-pro-protein) that undergoes post-translational processing to form an active Factor VIII protein (e.g., FVIIIa). For example, referring to FIG. 1, the wild type human Factor VIII pre-pro-protein is first cleaved to release the encoded signal peptide (not shown), forming a first single-chain pro-protein (shown as "human wild-type FVIII). The pro-protein is then cleaved between the B and A3 domains to form a first polypeptide that includes the Factor VIII heavy chain (e.g., the A1 and A2 domains) and B-domain, and a second polypeptide that includes the Factor VIII light chain (e.g., including the A3, C1, and C3 domains). The first polypeptide is further cleaved to remove the B-domain, and also to separate the A1 and A2 domains, which remain associated with the Factor VIII light chain in the mature Factor VIIIa protein. For review of the Factor VIII maturation process, see Graw et al., *Nat Rev Genet.*, 6(6):488-501 (2005), the content of which is incorporated herein by reference in its entirety for all purposes.

[0195] However, in some embodiments, the Factor VIII polypeptide is a single-chain Factor VIII polypeptide. Single-chain Factor VIII polypeptides are engineered to remove natural cleavage sites, and optionally remove, truncate, or replace the B-domain of Factor VIII. As such, they are not matured by cleavage (other than cleavage of an

optional signal and/or leader peptide), and are active as a single chain. Non-limiting examples of single-chain Factor VIII polypeptides are described in Zollner et al. (Thromb Res, 134(1):125-31 (2014)) and Donath et al. (Biochem J., 312(1):49-55 (1995)), the disclosures of which are hereby incorporated by reference in their entireties for all purposes.

[0196] As used herein, the terms “Factor VIII heavy chain,” or simply “heavy chain,” refers to the aggregate of the A1 and A2 domains of a Factor VIII polypeptide. In an exemplary embodiment, amino acids 20-759 of CS04-FL-AA (SEQ ID NO: 2) constitute a Factor VIII heavy chain.

[0197] As used herein, the term “Factor VIII light chain,” or simply “light chain,” refers to the aggregate of the A3, C1, and C2 domains of a Factor VIII polypeptide. In an exemplary embodiment, amino acids 774-1457 CS04-FL-AA (SEQ ID NO: 2) constitute a Factor VIII light chain. In some embodiments, a Factor VIII light chain excludes the acidic a3 peptide, which is released during maturation in vivo.

[0198] Generally, Factor VIII heavy and light chains are expressed as a single polypeptide chain, e.g., along with an optional B-domain or B-domain substituted linker. However, in some embodiments, a Factor VIII heavy chain and Factor VIII light chain are expressed as separate polypeptide chains (e.g., co-expressed), and reconstituted to form a Factor VIII protein (e.g., in vivo or in vitro).

[0199] As used herein, the terms “B-domain substituted linker” and “Factor VIII linker” are used interchangeably, and refer to truncated versions of a wild type Factor VIII B-domain (e.g., amino acids 760-1667 of FVIII-FL-AA (SEQ ID NO: 19)) or peptides engineered to replace the B-domain of a Factor VIII polypeptide. As used herein, a Factor VIII linker is positioned between the C-terminus of a Factor VIII heavy chain and the N-terminus of a Factor VIII light chain in a Factor VIII variant polypeptide in accordance with some embodiments. Non-limiting examples of B-domain substituted linkers are disclosed in U.S. Pat. Nos. 4,868,112, 5,112,950, 5,171,844, 5,543,502, 5,595,886, 5,610,278, 5,789,203, 5,972,885, 6,048,720, 6,060,447, 6,114,148, 6,228,620, 6,316,226, 6,346,513, 6,458,563, 6,924,365, 7,041,635, and 7,943,374; U.S. Patent Application Publication Nos. 2013/024960, 2015/0071883, and 2015/0158930; and PCT Publication Nos. WO 2014/064277 and WO 2014/127215, the disclosures of which are hereby incorporated by reference, in their entireties, for all purposes.

[0200] Unless otherwise specified herein, the numbering of Factor VIII amino acids refers to the corresponding amino acid in the full-length, wild-type human Factor VIII sequence (FVIII-FL-AA), presented as SEQ ID NO: 19 in FIG. 22. As such, when referring to an amino acid substitution in a Factor VIII variant protein disclosed herein, the recited amino acid number refers to the analogous (e.g., structurally or functionally equivalent) and/or homologous (e.g., evolutionarily conserved in the primary amino acid sequence) amino acid in the full-length, wild-type Factor VIII sequence. For example, a T2105N amino acid substitution refers to a T to N substitution at position 2105 of the full-length, wild-type human Factor VIII sequence (FVIII-FL-AA; SEQ ID NO: 19), a T to N substitution at position 1211 of the Factor VIII variant protein encoded by CS04 (CS04-FL-AA; SEQ ID NO: 2), and a T to N substitution at position 1212 of the Factor VIII variant encoded by CS04m3 (CS04m3-FL-AA; SEQ ID NO: 105).

[0201] As described herein, the Factor VIII amino acid numbering system is dependent on whether the Factor VIII signal peptide (e.g., amino acids 1-19 of the full-length, wild-type human Factor VIII sequence) is included. Where the signal peptide is included, the numbering is referred to as “signal peptide inclusive” or “SPI”. Where the signal peptide is not included, the numbering is referred to as “signal peptide exclusive” or “SPE.” For example, F328S is SPI numbering for the same amino acid as F309S, in SPE numbering. Unless otherwise indicated, all amino acid numbering refers to the corresponding amino acid in the full-length, wild-type human Factor VIII sequence (FVIII-FL-AA), presented as SEQ ID NO: 19 in FIG. 22.

[0202] As described herein, the codon-altered polynucleotides provide increased expression of transgenic Factor VIII in vivo (e.g., when administered as part of a gene therapy vector), as compared to the level of Factor VIII expression provided by a natively-coded Factor VIII construct (e.g., a polynucleotide encoding the same Factor VIII construct using the wild-type human codons). As used herein, the term “increased expression” refers to an increased level of transgenic Factor VIII activity in the blood of an animal administered the codon-altered polynucleotide encoding Factor VIII, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively-coded Factor VIII construct. The activity levels can be measured using any Factor VIII activity known in the art. An exemplary assay for determining Factor VIII activity is the Technochrome FVIII assay (Technoclone, Vienna, Austria).

[0203] In some embodiments, increased expression refers to at least 25% greater transgenic Factor VIII activity in the blood of an animal administered the codon-altered Factor VIII polynucleotide, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively coded Factor VIII polynucleotide. In some embodiments, increased expression refers to at least 50% greater, at least 75% greater, at least 100% greater, at least 3-fold greater, at least 4-fold greater, at least 5-fold greater, at least 6-fold greater, at least 7-fold greater, at least 8-fold greater, at least 9-fold greater, at least 10-fold greater, at least 15-fold greater, at least 20-fold greater, at least 25-fold greater, at least 30-fold greater, at least 40-fold greater, at least 50-fold greater, at least 60-fold greater, at least 70-fold greater, at least 80-fold greater, at least 90-fold greater, at least 100-fold greater, at least 125-fold greater, at least 150-fold greater, at least 175-fold greater, at least 200-fold greater, at least 225-fold greater, or at least 250-fold greater transgenic Factor VIII activity in the blood of an animal administered the codon-altered Factor VIII polynucleotide, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively coded Factor VIII polynucleotide.

[0204] As described herein, the codon-altered polynucleotides provide increased vector production, as compared to the level of vector production provided by a natively-coded Factor VIII construct (e.g., a polynucleotide encoding the same Factor VIII construct using the wild-type human codons). As used herein, the term “increased virus production” refers to an increased vector yield in cell culture (e.g., titer per liter culture) inoculated with the codon-altered polynucleotide encoding Factor VIII, as compared to the vector yield in cell culture inoculated with a natively-coded Factor VIII construct. The vector yields can be measured

using any vector titer assay known in the art. An exemplary assay for determining vector yield (e.g., of an AAV vector) is qPCR targeting the AAV2 inverted terminal repeats (Aurnhammer, *Human Gene Therapy Methods: Part B* 23:18-28 (2012)).

[0205] In some embodiments, increased virus production refers to at least 25% greater codon-altered vector yield, as compared to the yield of a natively-coded Factor VIII construct in the same type of culture. In some embodiments, increased vector production refers to at least 50% greater, at least 75% greater, at least 100% greater, at least 3-fold greater, at least 4-fold greater, at least 5-fold greater, at least 6-fold greater, at least 7-fold greater, at least 8-fold greater, at least 9-fold greater, at least 10-fold greater, at least 15-fold greater, or at least 20-fold greater codon-altered vector yield, as compared to the yield of a natively-coded Factor VIII construct in the same type of culture.

[0206] As used herein, the term “hemophilia” refers to a group of disease states broadly characterized by reduced blood clotting or coagulation. Hemophilia may refer to Type A, Type B, or Type C hemophilia, or to the composite of all three diseases types. Type A hemophilia (hemophilia A) is caused by a reduction or loss of factor VIII (FVIII) activity and is the most prominent of the hemophilia subtypes. Type B hemophilia (hemophilia B) results from the loss or reduction of factor IX (FIX) clotting function. Type C hemophilia (hemophilia C) is a consequence of the loss or reduction in factor XI (FXI) clotting activity. Hemophilia A and B are X-linked diseases, while hemophilia C is autosomal. Conventional treatments for hemophilia include both prophylactic and on-demand administration of clotting factors, such as FVIII, FIX, including Bebulin®-VH, and FXI, as well as FEIBA-VH, desmopressin, and plasma infusions.

[0207] As used herein, the term “FVIII gene therapy” includes any therapeutic approach of providing a nucleic acid encoding Factor VIII to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with hemophilia. The term encompasses administering any compound, drug, procedure, or regimen comprising a nucleic acid encoding a Factor VIII molecule, including any modified form of Factor VIII (e.g., Factor VIII variant), for maintaining or improving the health of an individual with hemophilia. One skilled in the art will appreciate that either the course of FVIII therapy or the dose of a FVIII therapeutic agent can be changed, e.g., based upon the results obtained in accordance with the present disclosure.

[0208] As used herein, the term “bypass therapy” includes any therapeutic approach of providing non-Factor VIII hemostatic agents, compounds or coagulation factors to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with hemophilia. Non-Factor VIII compounds and coagulation factors include, but are not limited to, Factor VIII Inhibitor Bypass Activity (FEIBA), recombinant activated factor VII (FVIIa), prothrombin complex concentrates, and activated prothrombin complex concentrates. These non-Factor VIII compounds and coagulation factors may be recombinant or plasma-derived. One skilled in the art will appreciate that either the course of bypass therapy or the dose of bypass therapy can be changed, e.g., based upon the results obtained in accordance with the present disclosure.

[0209] As used herein, a “combination therapy” including administration of a nucleic acid encoding a Factor VIII

molecule and a conventional hemophilia A therapeutic agent includes any therapeutic approach of providing both a nucleic acid encoding a Factor VIII molecule and a Factor VIII molecule and/or non-Factor VIII hemostatic agent (e.g., bypass therapeutic agent) to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with hemophilia. The term encompasses administering any compound, drug, procedure, or regimen including a nucleic acid encoding a Factor VIII molecule, including any modified form of factor VIII, which is useful for maintaining or improving the health of an individual with hemophilia and includes any of the therapeutic agents described herein.

[0210] The terms “therapeutically effective amount or dose” or “therapeutically sufficient amount or dose” or “effective or sufficient amount or dose” refer to a dose that produces therapeutic effects for which it is administered. For example, a therapeutically effective amount of a drug useful for treating hemophilia can be the amount that is capable of preventing or relieving one or more symptoms associated with hemophilia. The exact dose will depend on the purpose of the treatment, and will be ascertainable by one skilled in the art using known techniques (see, e.g., Lieberman, *Pharmaceutical Dosage Forms* (vols. 1-3, 1992); Lloyd, *The Art, Science and Technology of Pharmaceutical Compounding* (1999); Pickar, *Dosage Calculations* (1999); and Remington: *The Science and Practice of Pharmacy*, 20th Edition, 2003, Gennaro, Ed., Lippincott, Williams & Wilkins).

[0211] As used herein, the term “gene” refers to the segment of a DNA molecule that codes for a polypeptide chain (e.g., the coding region). In some embodiments, a gene is positioned by regions immediately preceding, following, and/or intervening the coding region that are involved in producing the polypeptide chain (e.g., regulatory elements such as a promoter, enhancer, polyadenylation sequence, 5'-untranslated region, 3'-untranslated region, or intron).

[0212] As used herein, the term “regulatory elements” refers to nucleotide sequences, such as promoters, enhancers, terminators, polyadenylation sequences, introns, etc., that provide for the expression of a coding sequence in a cell.

[0213] As used herein, the term “promoter element” refers to a nucleotide sequence that assists with controlling expression of a coding sequence. Generally, promoter elements are located 5' of the translation start site of a gene. However, in certain embodiments, a promoter element may be located within an intron sequence, or 3' of the coding sequence. In some embodiments, a promoter useful for a gene therapy vector is derived from the native gene of the target protein (e.g., a Factor VIII promoter). In some embodiments, a promoter useful for a gene therapy vector is specific for expression in a particular cell or tissue of the target organism (e.g., a liver-specific promoter). In yet other embodiments, one of a plurality of well characterized promoter elements is used in a gene therapy vector described herein. Non-limiting examples of well-characterized promoter elements include the CMV early promoter, the β -actin promoter, and the methyl CpG binding protein 2 (MeCP2) promoter. In some embodiments, the promoter is a constitutive promoter, which drives substantially constant expression of the target protein. In other embodiments, the promoter is an inducible promoter, which drives expression of the target protein in response to a particular stimulus (e.g., exposure to a particular treatment or agent). For a review of designing promoters for AAV-mediated gene therapy, see Gray et al.

(Human Gene Therapy 22:1143-53 (2011)), the contents of which are expressly incorporated by reference in their entirety for all purposes.

[0214] As used herein, the term “vector” refers to any vehicle used to transfer a nucleic acid (e.g., encoding a Factor VIII gene therapy construct) into a host cell. In some embodiments, a vector includes a replicon, which functions to replicate the vehicle, along with the target nucleic acid. Non-limiting examples of vectors useful for gene therapy include plasmids, phages, cosmids, artificial chromosomes, and viruses, which function as autonomous units of replication *in vivo*. In some embodiments, a vector is a viral vehicle for introducing a target nucleic acid (e.g., a codon-altered polynucleotide encoding a Factor VIII variant). Many modified eukaryotic viruses useful for gene therapy are known in the art. For example, adeno-associated viruses (AAVs) are particularly well suited for use in human gene therapy because humans are a natural host for the virus, the native viruses are not known to contribute to any diseases, and the viruses illicit a mild immune response.

[0215] As used herein, the term “CpG island” refers to a region within a polynucleotide having a statistically elevated density of CpG dinucleotides. As used herein, a region of a polynucleotide (e.g., a polynucleotide encoding a codon-altered Factor VIII protein) is a CpG island if, over a 200-base pair window: (i) the region has GC content of greater than 50%, and (ii) the ratio of observed CpG dinucleotides per expected CpG dinucleotides is at least 0.6, as defined by the relationship:

$$\frac{N[CpG] * N[\text{length of window}]}{N[C] * N[G]} \geq 0.6.$$

For additional information on methods for identifying CpG islands, see Gardiner-Garden M. et al., *J Mol Biol.*, 196(2): 261-82 (1987), the content of which is expressly incorporated herein by reference, in its entirety, for all purposes.

[0216] As used herein, the term “nucleic acid” refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, and complements thereof. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, and peptide-nucleic acids (PNAs).

[0217] The term “amino acid” refers to naturally occurring and non-natural amino acids, including amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids include those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Naturally occurring amino acids can include, e.g., D- and L-amino acids. The amino acids used herein can also include non-natural amino acids. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., any carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R

group, e.g., homoserine, norleucine, methionine sulfoxide, or methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refer to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that function in a manner similar to a naturally occurring amino acid. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0218] The nucleotide sequences that encode the mutant Factor VIII constructs herein may be identical to the coding sequence provided herein or may be a different coding sequence, which sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same polypeptides as the coding sequences provided herein. One of ordinary skill in the art will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each variation of a nucleic acid which encodes a same polypeptide is implicit in each described sequence with respect to the expression product, but not with respect to actual gene therapy constructs.

[0219] As to amino acid sequences, one of ordinary skill in the art will recognize that individual substitutions, deletions or additions to a nucleic acid or peptide sequence that alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the disclosure.

[0220] Conservative amino acid substitutions providing functionally similar amino acids are well known in the art. Dependent on the functionality of the particular amino acid, e.g., catalytic, structural, or sterically important amino acids, different groupings of amino acid may be considered conservative substitutions for each other. Table 1 provides groupings of amino acids that are considered conservative substitutions based on the charge and polarity of the amino acid, the hydrophobicity of the amino acid, the surface exposure/structural nature of the amino acid, and the secondary structure propensity of the amino acid.

TABLE 1

Groupings of conservative amino acid substitutions based on the functionality of the residue in the protein.	
Important Feature	Conservative Groupings
Charge/Polarity	1. H, R, and K 2. D and E 3. C, T, S, G, N, Q, and Y 4. A, P, M, L, I, V, F, and W

TABLE 1-continued

Groupings of conservative amino acid substitutions based on the functionality of the residue in the protein.	
Important Feature	Conservative Groupings
Hydrophobicity	1. D, E, N, Q, R, and K 2. C, S, T, P, G, H, and Y 3. A, M, I, L, V, F, and W
Structural/Surface Exposure	1. D, E, N, Q, H, R, and K 2. C, S, T, P, A, G, W, and Y 3. M, I, L, V, and F
Secondary Structure Propensity	1. A, E, Q, H, K, M, L, and R 2. C, T, I, V, F, Y, and W 3. S, G, P, D, and N
Evolutionary Conservation	1. D and E 2. H, K, and R 3. N and Q 4. S and T 5. L, I, and V 6. F, Y, and W 7. A and G 8. M and C

[0221] The terms “identical” or percent “identity,” in the context of two or more nucleic acids or peptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection.

[0222] As is known in the art, a number of different programs may be used to identify whether a protein (or nucleic acid as discussed below) has sequence identity or similarity to a known sequence. Sequence identity and/or similarity is determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith & Waterman, *Adv. Appl. Math.*, 2:482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.*, 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. U.S.A.*, 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.), the Best Fit sequence program described by Devereux et al., *Nucl. Acid Res.*, 12:387-395 (1984), preferably using the default settings, or by inspection. Preferably, percent identity is calculated by FastDB based upon the following parameters: mismatch penalty of 1; gap penalty of 1; gap size penalty of 0.33; and joining penalty of 30, “Current Methods in Sequence Comparison and Analysis,” *Macromolecule Sequencing and Synthesis, Selected Methods and Applications*, pp 127-149 (1988), Alan R. Liss, Inc, all of which are incorporated by reference.

[0223] An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pair wise alignments. It may also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng

& Doolittle, *J. Mol. Evol.* 35:351-360 (1987); the method is similar to that described by Higgins & Sharp *CABIOS* 5:151-153 (1989), both incorporated by reference. Useful PILEUP parameters including a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps. **[0224]** Another example of a useful algorithm is the BLAST algorithm, described in: Altschul et al., *J. Mol. Biol.* 215, 403-410, (1990); Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997); and Karlin et al., *Proc. Natl. Acad. Sci. U.S.A.* 90:5873-5787 (1993), both incorporated by reference. A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul et al., *Methods in Enzymology*, 266:460-480 (1996); <http://blast.wustl.edu/blast/README.html>]. WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span=1, overlap fraction=0.125, word threshold (T)=11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity.

[0225] An additional useful algorithm is gapped BLAST, as reported by Altschul et al., *Nucl. Acids Res.*, 25:3389-3402, incorporated by reference. Gapped BLAST uses BLOSUM-62 substitution scores; threshold T parameter set to 9; the two-hit method to trigger ungapped extensions; charges gap lengths of k a cost of 10+k; Xu set to 16, and Xg set to 40 for database search stage and to 67 for the output stage of the algorithms. Gapped alignments are triggered by a score corresponding to ~22 bits.

[0226] A % amino acid sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the “longer” sequence in the aligned region. The “longer” sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored). In a similar manner, “percent (%) nucleic acid sequence identity” with respect to the coding sequence of the polypeptides identified is defined as the percentage of nucleotide residues in a candidate sequence that are identical with the nucleotide residues in the coding sequence of the cell cycle protein. A preferred method utilizes the BLASTN module of WU-BLAST-2 set to the default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively.

[0227] The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer amino acids than the protein encoded by the sequence of FIG. 2 (SEQ ID NO:1), it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical amino acids or nucleotides in relation to the total number of amino acids or nucleotides. Thus, for example, sequence identity of sequences shorter than that shown in FIG. 2 (SEQ ID NO:1), as discussed below, will be determined using the number of nucleotides in the shorter sequence, in one embodiment. In percent identity calculations relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, etc.

[0228] In one embodiment, only identities are scored positively (+1) and all forms of sequence variation including

gaps are assigned a value of “0”, which obviates the need for a weighted scale or parameters as described below for sequence similarity calculations. Percent sequence identity may be calculated, for example, by dividing the number of matching identical residues by the total number of residues of the “shorter” sequence in the aligned region and multiplying by 100. The “longer” sequence is the one having the most actual residues in the aligned region.

[0229] The term “allelic variants” refers to polymorphic forms of a gene at a particular genetic locus, as well as cDNAs derived from mRNA transcripts of the genes, and the polypeptides encoded by them. The term “preferred mammalian codon” refers a subset of codons from among the set of codons encoding an amino acid that are most frequently used in proteins expressed in mammalian cells as chosen from the following list: Gly (GGC, GGG); Glu (GAG); Asp (GAC); Val (GTG, GTC); Ala (GCC, GCT); Ser (AGC, TCC); Lys (AAG); Asn (AAC); Met (ATG); Ile (ATC); Thr (ACC); Trp (TGG); Cys (TGC); Tyr (TAT, TAC); Leu (CTG); Phe (TTC); Arg (CGC, AGG, AGA); Gln (CAG); His (CAC); and Pro (CCC).

[0230] As used herein, the term codon-altered refers to a polynucleotide sequence encoding a polypeptide (e.g., a Factor VIII variant protein), where at least one codon of the native polynucleotide encoding the polypeptide has been changed to improve a property of the polynucleotide sequence. In some embodiments, the improved property promotes increased transcription of mRNA coding for the polypeptide, increased stability of the mRNA (e.g., improved mRNA half-life), increased translation of the polypeptide, and/or increased packaging of the polynucleotide within the vector. Non-limiting examples of alterations that can be used to achieve the improved properties include changing the usage and/or distribution of codons for particular amino acids, adjusting global and/or local GC content, removing AT-rich sequences, removing repeated sequence elements, adjusting global and/or local CpG dinucleotide content, removing cryptic regulatory elements (e.g., TATA box and CCAAT box elements), removing of intron/exon splice sites, improving regulatory sequences (e.g., introduction of a Kozak consensus sequence), and removing sequence elements capable of forming secondary structure (e.g., stem-loops) in the transcribed mRNA.

[0231] As discussed herein, there are various nomenclatures to refer to components of the disclosure herein. “CS-number” (e.g., “CS04”, “CS01”, “CS23”, etc.) refer to codon altered polynucleotides encoding FVIII polypeptides and/or the encoded polypeptides, including variants. For example, CS01-FL refers to the Full Length codon altered CS01 polynucleotide sequence or amino acid sequence (sometimes referred to herein as “CS01-FL-AA” for the Amino Acid sequence and “CS01-FL-NA” for the Nucleic Acid sequence) encoded by the CS01 polynucleotide sequence. Similarly, “CS01-LC” refers to either the codon altered nucleic acid sequence (“CS01-LC-NA”) encoding the light chain of a FVIII polypeptide or the amino acid sequence (also sometimes referred to herein as “CS01-LC-AA”) of the FVIII light chain encoded by the CS01 polynucleotide sequence. Likewise, CS01-HC, CS01-HC-AA and CS01-HC-NA are the same for the FVIII heavy chain. As will be appreciated by those in the art, for constructs such as CS01, CS04, CS23, etc., that are only codon-altered (e.g. they do not contain additional amino acid substitutions as compared to Refacto), the amino acid sequences will be identical, as

the amino acid sequences are not altered by the codon optimization. Thus, sequence constructs of the disclosure include, but are not limited to, CS01-FL-NA, CS01-FL-AA, CS01-LC-NA, CS01-LC-AA, CS01-HC-AA, CS01-HC-NA, CS04-FL-NA, CS04-FL-AA, CS04-LC-NA, CS04-LC-AA, CS04-HC-AA, CS04-HC-NA, CS23-FL-NA, CS23-FL-AA, CS23-LC-NA, CS23-LC-AA, CS23-HC-AA and CS23-HC-NA.

[0232] This nomenclature also applies to glycosylation peptides as shown in FIG. 13, such that “NGA1-AA” refers to the amino acid sequence and NGA1-NA refers to the nucleic acid sequence.

[0233] The disclosure also includes additional new Factor VIII variants, as described below, with the appropriate nomenclature.

III. Codon-Altered Factor VIII Variants

[0234] In some embodiments, the present disclosure provides codon-altered polynucleotides encoding Factor VIII variants. These codon-altered polynucleotides provide markedly improved expression of Factor VIII when administered in an AAV-based gene therapy construct. The codon-altered polynucleotides also demonstrate improved AAV-virion packaging, as compared to conventionally codon-optimized constructs. As demonstrated in Example 2 and Example 4, Applicants have achieved these advantages through the discovery of three codon-altered polynucleotides (CS01-FL-NA, CS04-FL-NA, and CS23-FL-NA) encoding a Factor VIII polypeptide with human wild-type Factor VIII heavy and light chains, and a short, 14 amino acid, B-domain substituted linker (the “SQ” linker) containing a furin cleavage site to facilitate maturation of an active FVIIIa protein in vivo. As further demonstrated in Example 4, incorporation of various combinations of the F328S, X5, and X1 amino acid mutations into the encoded Factor VIII molecule further increased the in vivo expression of Factor VIII activity.

[0235] In one embodiment, a codon-altered polynucleotide provided herein has nucleotide sequences with high sequence identity to at least the sequences within CS01, CS04, or CS23 (SEQ ID NOS 13, 1, and 20, respectively) encoding the Factor VIII heavy chain and Factor VIII light chains. As known in the art, the B-domain of Factor VIII is dispensable for activity in vivo. Thus, in some embodiments, the codon-altered polynucleotides provided herein completely lack a Factor VIII B-domain. In some embodiments, the native Factor VIII B-domain is replaced with a short amino acid linker containing a furin cleavage site, e.g., the “SQ” linker consisting of amino acids 760-773 of the CS01, CS04, or CS23 (SEQ ID NOS 2, 2, and 21, respectively) constructs. The “SQ” linker is also referred to as BDLO4, (-AA for the amino acid sequence and -NA for the nucleotide sequence shown in FIG. 6).

[0236] In one embodiment, the Factor VIII heavy and light chains encoded by the codon-altered polynucleotide are human Factor VIII heavy and light chains, respectively. In other embodiments, the Factor VIII heavy and light chains encoded by the codon-altered polynucleotide are heavy and light chain sequences from another mammal (e.g., porcine Factor VIII). In yet other embodiments, the Factor VIII heavy and light chains are chimeric heavy and light chains (e.g., a combination of human and a second mammalian sequence). In yet other embodiments, the Factor VIII heavy and light chains are humanized version of the heavy and

light chains from another mammal, e.g., heavy and light chain sequences from another mammal in which human residues are substituted at select positions to reduce the immunogenicity of the resulting peptide when administered to a human.

[0237] The GC content of human genes varies widely, from less than 25% to greater than 90%. However, in general, human genes with higher GC contents are expressed at higher levels. For example, Kudla et al. (PLoS Biol., 4(6):80 (2006)) demonstrate that increasing a gene's GC content increases expression of the encoded polypeptide, primarily by increasing transcription and effecting a higher steady state level of the mRNA transcript. Generally, the desired GC content of a codon-optimized gene construct is equal or greater than 60%. However, native AAV genomes have GC contents of around 56%.

[0238] Accordingly, in some embodiments, the codon-altered polynucleotides provided herein have a GC content that more closely matches the GC content of native AAV virions (e.g., around 56% GC), which is lower than the preferred GC contents of polynucleotides that are conventionally codon-optimized for expression in mammalian cells (e.g., at or above 60% GC). As outlined in Example 1, CS04-FL-NA (SEQ ID NO: 1), which has a GC content of about 56%, has improved virion packaging as compared to similarly codon-altered coding sequences with higher GC content.

[0239] Thus, in some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 60%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is no more than 56%.

[0240] In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 56%. In some embodiments, the

overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 56%.

[0241] In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.5\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.4\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.3\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.2\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.1\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is 56%.

[0242] A. Factor VIII Amino Acid Substitutions

[0243] To further increase the efficiency of AAV-vector based expression of the Factor VIII constructs described herein, amino acid substitutions known to improve secretion, increase specific activity, and/or enhanced the stability of Factor VIII are further incorporated, in some implementations. A number of potential variants were identified that increase the plasma levels of FVIII activity at a given vector dose. These variants include those with a more efficient signal peptide, amino acid substitutions that prevent BiP interactions, amino acid substitutions resembling more efficiently secreted Factor VIII orthologs (e.g., porcine Factor VIII), single-chain Factor VIII variants, and amino acid substitutions that stabilize Factor VIII and/or reduce subunit dissociation.

[0244] Mutation of residues A108, R121, and L2302 (SPE), located at the interface between the A1 and C2 domains, increases the stability of Factor VIII. For example, the A108I amino acid substitution introduces a hydrophobic residue that better fills the inter-domain space, stabilizing the interaction. Likewise, an R121C/L2302C (SPE) double amino acid substitution introduces a disulfide bond spanning the A1-C2 domains, further stabilizing the interaction. Taken together, all three amino acid substitutions increase the thermal stability of Factor VIII by 3 to 4-fold. For review, see Wakabayashi et al., J Biol Chem. 286(29):25748-55 (2011) and Wakabayashi et al., Thromb Haemost. 10(3): 492-95 (2012). Accordingly, in some embodiments, the encoded Factor VIII polypeptide includes A108I and/or R121C/L2302C amino acid substitutions.

[0245] Mutation of E113 (SPE), located within the calcium binding domain of Factor VIII, increases the specific FVIII clotting activity. For example, E113A appears to increase FXase formation through increased FVIII affinity for Factor IXa. Specifically, the E113A amino acid substitution increases specific FVIII clotting activity two-fold and increases affinity for Factor IXa by four-fold (Biochemistry, 41:8485 (2002); J. Biol. Chem., 279:12677 (2004); and Biochemistry, 44:10298 (2005)). Accordingly, in some embodiments, the encoded Factor VIII polypeptides include an E113A amino acid substitution.

[0246] Substitution of one or more amino acid residues surrounding the Factor VIII APC cleavage site (residues 331-341 (SPE)) reduce Factor VIIIa inactivation by activated protein C, without affecting FVIII activity. For example PQL333-335VDQ (SPE) amino acid substitutions reduce Factor VIII inactivation by 16-fold. Likewise, MKN336-339GNQ amino acid substitutions reduce Factor

VIII inactivation by 9-fold. When combined, the two triple amino acid substitutions (e.g., PQLRMKN333-339VDQRGNQ) (SEQ ID NOS 34 and 35, respectively) reduce Factor VIII inactivation by 100-fold (J. Biol. Chem., 282:20264 (2007)). Accordingly, in some embodiments, the encoded Factor VIII polypeptide include PQL333-335VDQ and/or MKN337-339GNQ (SPE) amino acid substitutions.

[0247] Mutations within the A2 domain interface also increase Factor VIII stability. Specifically, mutating charged residues in the A1-A2 and A2-A3 domain interfaces increases stability and retention of the A2 subunit in Factor VIIIa. For example, mutation of D519, E665, and E1984 to V or A yields up to 2-fold increased stability in Factor VIII and up to 5-fold stability in Factor VIIIa. Specifically, D519A/E665V amino acid substitutions provide a 3-fold increase in stability; D519V/E665V amino acid substitutions provide a 2-fold increase in stability, an 8-fold decrease in A2 dissociation, and a 2-4-fold increase in thrombin generation potential; D519V/E1984A amino acid substitutions provide a 2-fold increase in stability; and D519V/E665V/E1984A amino acid substitution provide a 2-fold increase in stability (Blood 112:2761-69 (2008); J. Thromb. Haemost., 7:438-44 (2009)). Accordingly, in some embodiments, the encoded Factor VIII polypeptides include one or more of D519A/V, E665A/V, and E1984A/V amino acid substitutions.

[0248] Of particular relevance to the present disclosure are a number of specific mutations that can be included separately or in combinations with other variants described herein. These variants are coded as sets herein as follows: “m1” refers to a single amino acid change, “m2” is a set of 5 amino acid variants, “m3” is a combination of a deletion of 7 amino acids and an insertion of six amino acids that span the junction between the polypeptide linker and the heavy chain, “m4” is a combination of the m1 single mutation and the m5 double mutation, and “m5” is a set of two cysteine ablations. These mutations are described below. These can be included in any particular construct alone or in combination with other variants, and they are coded accordingly. For example, “m23” is a combination of the m2 and m3 variants onto a particular scaffold, as outlined herein; thus “CS01m23-FL-NA” or “CS01-FL-NAm23” refers to the CS01 codon-altered polynucleotide sequence with the nucleotides encoding the m2 and m3 mutations included, and “CS01m23-FL-AA” or “CS01-FL-AAm23” refers to the amino acid sequence. As CS01 is codon-altered but does not change the amino acid sequence of Refacto, these can be thought of on the amino acid level as mutations as compared to the Refacto amino acid sequence of CS01-FL-AA (SEQ ID NO: 2).

[0249] In many embodiments, the polypeptides of the disclosure are made with the “m1” variant included. Mutations within an 11 amino acid hydrophobic β -sheet in the A1 domain, which interacts with BiP, increase secretion of Factor VIII. For example, an F328S (SPI, F309S SPE) amino acid substitution within the pocket increased Factor VIII secretion 3-fold. The F328S variant is referred to herein as the “m1” mutation and is within the heavy chain. Again, as described herein, the number of the variants can be done inclusive of the signal peptide, “Signal Peptide Inclusive”, or “SPI”, or starting from the processed final protein sequence, “Signal Peptide Exclusive”, or “SPE”. Thus, using SPI numbering, the mutation F328S is the same as the F309 SPE mutant. Generally the specification uses the SPI

numbering, but as will be appreciated by those in the art, either numbering system results in the same mutation(s).

[0250] Accordingly, included in the present disclosure are polypeptides that include the m1 mutation, including CS01-FL-AAm1, CS01-HC-AAm1, CS04-FL-AAm1, CS04-HC-AAm1 CS23-FL-AAm1, CS23-HC-AAm1, CS40-FL-AAm1 and CS40-HC-AAm1 (all of which encode the same corresponding protein sequences).

[0251] In addition, included in the present disclosure are not only polypeptide sequences that include the m1 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m1 mutation, such as CS01-FL-NAm1, CS01-HC-NAm1, CS04-FL-NAm1, CS04-HC-NAm1, CS23-FL-NAm1, CS23-HC-NA-m1, CS40-FL-NAm1 and CS40-HC-NAm1.

[0252] In many embodiments, the polypeptides of the disclosure are made with the “m2” variant set included, which is the I105V/A127S/G151K/M166T/L171P mutations (SPI numbering; (SPE numbering is V861/S108A/K132G/T147M/P152L, respectively). The m2 mutation set is based on the fact that substitution of porcine amino acids 82-176 for the corresponding human amino acids in a B-domain deleted gene therapy construct increased Factor VIII activity when expressed in HEK293 cells (W. Xiao, communication). Id. Back-mutation of single porcine amino acids into the human BDD-FVIII construct identified five amino acids within the A1 domain that contribute to this phenomenon: I105V, A127S, G151K, M166T, and L171P (SPI). Introduction of the combination of these mutations into the human construct recapitulated the improved activity of the larger porcine substitution. Id. Accordingly, in some embodiments, the encoded Factor VIII polypeptides include one or more amino acid substitutions selected from I105V, A127S, G151K, M166T, and L171P, with the entire 5 amino acid set, m2, finding particular use in many embodiments. As for the m1 mutation, the m2 variants are in the heavy chain, and thus the present disclosure includes polypeptides that include the m2 mutation, including CS01-FL-AAm2, CS01-HC-AAm2, CS04-FL-AAm2, CS04-HC-AAm2, CS23-FL-AAm2, CS23-HC-AAm2, CS40-FL-AAm2 and CS40-HC-AAm2 (all of which encode the same corresponding protein sequences).

[0253] In addition, included in the present disclosure are not only polypeptide sequences that include the m2 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m2 mutation, such as CS01-FL-NAm2, CS01-HC-NAm2, CS04-FL-NAm2, CS04-HC-NAm2, CS23-FL-NAm2, CS23-HC-NA-m2, CS40-FL-NAm2 and CS40-HC-NAm2.

[0254] In additional embodiments, the polypeptides and polynucleotides of the disclosure include m3 mutations. m3 is the substitution of seven amino acids for six across the HC-B domain interface that introduces an additional glycosylation site introduced close to the interface. Accordingly, in some embodiments, m3 is the deletion of amino acids AIEPRSF755-761 and the insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19) (e.g., AIEPRSF755-761 TTYVNRSL) (“TTYVNRSL” disclosed as SEQ ID NO: 33). Residues AIEPR755-759, relative to SEQ ID NO: 19, fall within the end of the heavy chain, while residues 5760 and F761 fall within the B-domain. In some embodiments, where the FVIII B-domain is deleted, truncated, or replaced, residues 5760 and F761 may not be present in the underlying

amino acid sequence being mutated. Accordingly, in some embodiments, m3 is the deletion of amino acids AIEPR755-759 and the insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19) (e.g., AIEPR755-759TTYVNRSL (“TTYVNRSL” disclosed as SEQ ID NO: 33)

[0255] The m3 variants are in the junction between the heavy chain and the B domain, and thus the present disclosure includes polypeptides that include the m3 mutation, including CS01-FL-AAm3, CS01-HC-AAm3, CS04-FL-AAm3, CS04-HC-AAm3, CS23-FL-AAm3, CS23-HC-AAm3, CS40-FL-AAm3 and CS40-HC-AAm3 (all of which encode the same corresponding protein sequences).

[0256] In addition, included in the present disclosure are not only polypeptide sequences that include the m3 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m3 mutations, such as CS01-FL-NAm3, CS01-HC-NAm3, CS04-FL-NAm3, CS04-HC-NAm3, CS23-FL-NAm3, CS23-HC-NA-m3, CS40-FL-NAm3 and CS40-HC-NAm3.

[0257] In additional embodiments, the polypeptides and polynucleotides of the disclosure include m4 mutations. Elimination of the C1899-C1903 disulfide bond in Factor VIII also increased secretion. Moreover, the increases in Factor VIII secretion are additive for the combination of F328S (SPI, F309S SPE) and C1918G/C1922G amino acid substitutions (Miao et al., *Blood*, 103:3412-19 (2004); Selvaraj et al., *J. Thromb. Haemost.*, 10:107-15 (2012)). Accordingly, in some embodiments, the encoded Factor VIII polypeptides include m4 mutations, which is the F328S (SPI, F309S SPE) and C1918G/C1922G (SPI) amino acid substitutions. As the F328S variant is in the heavy chain and the two cysteine variants are in the light chain, polypeptide sequences that include m4 mutations are CS01-FL-AAm4, CS01-HC-AAm4, CS01-LC-AAm4, CS04-FL-AAm4, CS04-HC-AAm4, CS04-LC-AAm4, CS23-FL-AAm4, CS23-HC-AAm4 and CS23-LC-AAm4.

[0258] In addition, included in the present disclosure are not only polypeptide sequences that include the m4 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m4 mutations, such as CS01-FL-NAm4, CS01-HC-NAm4, CS01-LC-NAm4, CS04-FL-NAm4, CS04-HC-NAm4, CS04-LC-NAm4, CS23-FL-NAm4, CS23-HC-NAm4, CS23-LC-NAm4, CS40-FL-NA-m4, CS40-HC-NA-m4 and CS40-LC-NA-m4.

[0259] In additional embodiments, the polypeptides and polynucleotides of the disclosure include m5 mutations. As above, elimination of the C1899-C1903 disulfide bond in Factor VIII also increased secretion. C1918G/C1922G (SPI) amino acid substitutions, contained within the light chain, referred to herein as the m5 mutation set.

[0260] The m5 variants are in the light chain, and thus the present disclosure includes polypeptides that include the m5 mutation, including CS01-FL-AAm5, CS01-LC-AAm5, CS04-FL-AAm5, CS04-LC-AAm5, CS23-FL-AAm5, CS23-LC-AAm5, CS40-FL-AAm5 and CS40-LC-AAm5 (all of which encode the same corresponding protein sequences).

[0261] In addition, included in the present disclosure are not only polypeptide sequences that include the m5 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m5 mutations, such as CS01-FL-NAm5, CS01-LC-NAm5, CS04-FL-NAm5, CS04-LC-

NAm5, CS23-FL-NA-m5, CS23-LC-NA-m5, CS40-FL-NA-m5 and CS40-LC-NA-m5.

[0262] In addition to specific constructs (both amino acid and nucleic acid) that include m1, m2, m3, m4 and m5 individually, combinations of mutation sets can be made as outlined herein. As noted herein, these are noted as “m12”, which is the combination of m1 and m2 sets, or “m123” which is the combination of m1, m2 and m3 sets. Thus, included in the disclosure are dual combinations including m12, m13, m14, m15, m23, m24, m25, m34, m35 and m45. Also included are triple combinations, m123, m124, m125, m234, m235 and m345. Further included are quad combinations, m1234, m1235, m1345 and the m12345 combination. Of particular interest in some embodiments are the following mutation sets: m1, m2, m3 and m4, m23, m123, and m234.

[0263] B. Factor VIII B-Domain Substituted Linkers

[0264] In some embodiments, the linkage between the FVIII heavy chain and the light chain (e.g., the B-domain in wild-type Factor VIII) is further altered. Due to size constraints of AAV packaging capacity, B-domain deleted, truncated, and or linker substituted variants should improve the efficacy of the FVIII gene therapy construct. The most conventionally used B-domain substituted linker is that of SQ FVIII, which retains only 14 amino acids of the B domain as linker sequence. Another variant of porcine VIII (“OBI-1,” described in U.S. Pat. No. 6,458,563) is well expressed in CHO cells, and has a slightly longer linker of 24 amino acids. In some embodiments, the Factor VIII constructs encoded by the codon-altered polynucleotides described herein include an SQ-type B-domain linker sequence. In other embodiments, the Factor VIII constructs encoded by the codon-altered polynucleotides described herein include an OBI-1-type B-domain linker sequence.

[0265] In some embodiments, the encoded Factor VIII polypeptides described herein include an SQ-type B-domain linker, including amino acids 760-762/1657-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Sandberg et al. *Thromb. Haemost.* 85:93 (2001)). In some embodiments, the SQ-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the SQ-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the SQ-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in FIG. 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[0266] In some embodiments, the encoded Factor VIII polypeptides described herein include a Greengene-type B-domain linker, including amino acids 760/1582-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Oh et al., *Biotechnol. Prog.*, 17:1999 (2001)). In some embodiments, the Greengene-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the Greengene-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the Greengene-type B-domain linker. In some embodiments, the glycosylation peptide is selected from

those shown in FIG. 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[0267] In some embodiments, the encoded Factor VIII polypeptides described herein include an extended SQ-type B-domain linker (SFSQNPVLRHQR; BDL-SQ-AA; SEQ ID NO: 30), including amino acids 760-769/1657-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Thim et al., *Haemophilia*, 16:349 (2010)). In some embodiments, the extended SQ-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the extended SQ-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the extended SQ-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in FIG. 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[0268] In some embodiments, the encoded Factor VIII polypeptides described herein include a porcine OBI-1-type B-domain linker, including the amino acids SFAQNSRPPSASAPKPPVLRHQR (SEQ ID NO: 31) from the wild-type porcine Factor VIII B-domain (Toschi et al., *Curr. Opin. Mol. Ther.* 12:517 (2010)). In some embodiments, the porcine OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the porcine OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the porcine OBI-1-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in FIG. 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[0269] In some embodiments, the encoded Factor VIII polypeptides described herein include a human OBI-1-type B-domain linker, including amino acids 760-772/1655-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19). In some embodiments, the human OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the human OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the human OBI-1-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in FIG. 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[0270] In some embodiments, the encoded Factor VIII polypeptides described herein include an 08-type B-domain linker, including the amino acids SFSQNSRHQAYRYRRG (SEQ ID NO: 32) from the wild-type porcine Factor VIII B-domain (Toschi et al., *Curr. Opin. Mol. Ther.* 12:517 (2010)). In some embodiments, the porcine OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the porcine OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the porcine OBI-1-type B-domain linker. In some embodiments, the glycosylation peptide is selected

from those shown in FIG. 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[0271] Removal of the B-domain from Factor VIII constructs does not appear to affect the activity of the activated enzyme (e.g., FVIIIa), presumably because the B-domain is removed during activation. However, the B-domain of Factor VIII contains several residues that are post-translationally modified, e.g., by N- or O-linked glycosylation. In silico analysis (Prediction of N-glycosylation sites in human proteins, R. Gupta, E. Jung and S. Brunak, in preparation (2004)) of the wild-type Factor VIII B-domain predicts that at least four of these sites are glycosylated in vivo (FIG. 14). It is thought that these modifications within the B-domain contribute to the post-translational regulation and/or half-life of Factor VIII in vivo.

[0272] While the Factor VIII B-domain is absent in mature Factor VIIIa protein, glycosylation within the B-domain of the precursor Factor VIII molecule may increase the circulating half-life of the protein prior to activation. Thus, in some embodiments, the polypeptide linker of the encoded Factor VIII constructs described herein includes one or more glycosylation sequences, to allow for glycosylation in vivo. In some embodiments, the polypeptide linker includes at least one consensus glycosylation sequence (e.g., an N- or O-linked glycosylation consensus sequence). In some embodiments, the polypeptide linker includes at least two consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least three consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least four consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least five consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least 6, 7, 8, 9, 10, or more consensus glycosylation sequences.

[0273] In some embodiments, the polypeptide linker contains at least one N-linked glycosylation sequence N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least two N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least three N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least four N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least five N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least 6, 7, 8, 9, 10, or more N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T.

[0274] In some embodiments, the polypeptide linker includes a glycosylation peptide with high sequence identity to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in FIGS. 13A-13B. In some embodiments, the glycosylation polypeptide has at least 92% identity to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in FIGS. 13A-13B. In some embodiments, the glycosylation peptide has no more than two amino acid substitutions relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order

amino acid substitutions relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in FIGS. 13A-13B, and a glycosylation peptide having no more than one amino acid substitution relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in FIGS. 13A-13B. In some embodiments, the glycosylation peptide is inserted in the SQ peptide between residues N768 and P769 (relative to CS04-FL-AA; SEQ ID NO: 2).

[0280] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to any one of those shown in FIG. 6 (SEQ ID NOS 5-7 and 36-48, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 95% identity to any one of those shown in FIG. 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 96% identity to any one of those shown in FIG. 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 97% identity to any one of those shown in FIG. 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 98% identity to any one of those shown in FIG. 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 99% identity to any one of those shown in FIG. 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 99.5% identity to any one of those shown in FIG. 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence is identical to any one of those shown in FIG. 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance).

[0281] C. Codon-Altered Polynucleotides Encoding a Factor VIII Variant with a Cleavable Linker

[0282] CS04 Codon Altered Polynucleotides

[0283] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS04-HC-NA (SEQ ID NO: 3), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS04-LC-NA (SEQ ID NO: 4), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site, which

allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[0284] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively.

[0285] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[0286] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 98% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence is identical to CS04-FL-NA (SEQ ID NO: 1).

[0287] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 97% identity to CS04-FL-AA (SEQ ID

NO: 2). In some embodiments, the amino acid sequence has at least 98% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence is identical to CS04-FL-AA (SEQ ID NO: 2).

[0288] In some embodiments, the Factor VIII variant encoded by the CS04 polynucleotide, having high sequence homology to CS04-FL-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0289] In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m5 amino acid substitution.

[0290] In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m35 amino acid substitutions.

[0291] In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m125 amino acid substitutions.

[0292] CS01 Codon Altered Polynucleotides

[0293] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS01-HC-NA (SEQ ID NO: 24), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to

CS01-LC-NA (SEQ ID NO: 25), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site, which allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[0294] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively.

[0295] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[0296] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 95% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 98% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence is identical to CS01-FL-NA (SEQ ID NO: 13).

[0297] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino

acid sequence with high sequence identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 97% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 98% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.9% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence is identical to CS01-FL-AA (SEQ ID NO: 2).

[0298] In some embodiments, the Factor VIII variant encoded by the CS01 polynucleotide, having high sequence homology to CS01-FL-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0299] In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m5 amino acid substitution.

[0300] In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m35 amino acid substitutions.

[0301] In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m125 amino acid substitutions.

[0302] CS23 Codon Altered Polynucleotides

[0303] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS23-HC-NA (SEQ ID NO: 22), which is the portion of CS23-FL-NA

(SEQ ID NO: 20) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS23-LC-NA (SEQ ID NO: 23), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site, which allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[0304] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively.

[0305] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[0306] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 95% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99.9% identity to

CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence is identical to CS23-FL-NA (SEQ ID NO: 20).

[0307] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 97% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 98% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence is identical to CS23-FL-AA (SEQ ID NO: 21).

[0308] In some embodiments, the Factor VIII variant encoded by the CS23 polynucleotide, having high sequence homology to CS23-FL-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0309] In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m5 amino acid substitution.

[0310] In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m35 amino acid substitutions.

[0311] In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m125 amino acid substitutions.

[0312] D. Codon-Altered Polynucleotides Encoding a Single-Chain Factor VIII Protein

[0313] Factor VIII constructs in which the furin cleavage site located at the C-terminal end of the B-domain is removed retain activity as a single chain polypeptide, despite that normal maturation of the Factor VIII molecule

cannot occur (Leyte et al. (1991)). Similarly, a B-domain deleted Factor VIII construct with an attenuated furin site (containing an R1664H amino acid substitution) is more biologically active than the corresponding Factor VIII construct with a wild-type furin cleavage site (Siner et al. (2013)). Accordingly, in some embodiments, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The single-chain Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The polypeptide linker does not include a furin cleavage site.

[0314] Single-Chain CS04 Codon Altered Polynucleotides

[0315] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS04-HC-NA (SEQ ID NO: 3), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS04-LC-NA (SEQ ID NO: 4), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[0316] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively.

[0317] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence

has at least 98% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence is identical to CS04-SC1-NA (SEQ ID NO: 9).

[0318] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 98% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence is identical to CS04-SC2-NA (SEQ ID NO: 11).

[0319] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; HsFVIII Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS04-SC1-AA (SEQ ID NO: 10).

[0320] In some embodiments, the Factor VIII variant encoded by the CS04-SC1 polynucleotide, having high sequence homology to CS04-SC1-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0321] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; HsFVIII Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS04-SC2-AA (SEQ ID NO: 12). In some

embodiments, the amino acid sequence has at least 98% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS04-SC2-AA (SEQ ID NO: 12).

[0322] In some embodiments, the single-chain Factor VIII variant encoded by the CS04-SC2 polynucleotide, having high sequence homology to CS04-SC2-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0323] In one embodiment, the single-chain Factor VIII variant encoded by the CS04 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m5 amino acid substitution.

[0324] In one embodiment, the single-chain Factor VIII variant encoded by the CS04 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m35 amino acid substitutions.

[0325] In one embodiment, the single-chain Factor VIII variant encoded by the CS04 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m125 amino acid substitutions.

[0326] Single-Chain CS01 Codon Altered Polynucleotides

[0327] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS01-HC-NA (SEQ ID NO: 24), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS01-LC-NA

(SEQ ID NO: 25), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[0328] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively.

[0329] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 95% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 98% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence is identical to CS01-SC1-NA (SEQ ID NO: 26).

[0330] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 95% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 98% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 99.9%

identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence is identical to CS01-SC2-NA (SEQ ID NO: 27).

[0331] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; HsFVIII Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS01-SC1-AA (SEQ ID NO: 10).

[0332] In some embodiments, the Factor VIII variant encoded by the CS01-SC1 polynucleotide, having high sequence homology to CS01-SC1-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0333] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; HsFVIII Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 98% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.9% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS01-SC2-AA (SEQ ID NO: 12).

[0334] In some embodiments, the single-chain Factor VIII variant encoded by the CS01-SC2 polynucleotide, having high sequence homology to CS01-SC2-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0335] In one embodiment, the single-chain Factor VIII variant encoded by the CS01 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m4 amino acid substitution. In one

embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m5 amino acid substitution.

[0336] In one embodiment, the single-chain Factor VIII variant encoded by the CS01 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m35 amino acid substitutions.

[0337] In one embodiment, the single-chain Factor VIII variant encoded by the CS01 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m125 amino acid substitutions.

[0338] Single-Chain CS23 Codon Altered Polynucleotides

[0339] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS23-HC-NA (SEQ ID NO: 22), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS23-LC-NA (SEQ ID NO: 23), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[0340] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and

23), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively.

[0341] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 95% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence is identical to CS23-SC1-NA (SEQ ID NO: 28).

[0342] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 95% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence is identical to CS23-SC2-NA (SEQ ID NO: 29).

[0343] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; CS04 Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS23-SC1-AA (SEQ ID NO: 10).

[0344] In some embodiments, the Factor VIII variant encoded by the CS23-SC1 polynucleotide, having high sequence homology to CS23-SC1-AA (e.g., at least 95%,

96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0345] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; HsFVIII Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 98% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS23-SC2-AA (SEQ ID NO: 12).

[0346] In some embodiments, the single-chain Factor VIII variant encoded by the CS23-SC2 polynucleotide, having high sequence homology to CS23-SC2-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[0347] In one embodiment, the single-chain Factor VIII variant encoded by the CS23 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m5 amino acid substitution.

[0348] In one embodiment, the single-chain Factor VIII variant encoded by the CS23 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m35 amino acid substitutions.

[0349] In one embodiment, the single-chain Factor VIII variant encoded by the CS23 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m125 amino acid substitutions.

[0350] E. Factor VIII Expression Vectors

[0351] In some embodiments, the codon-altered polynucleotides described herein are integrated into expression vectors. Non-limiting examples of expression vectors include viral vectors (e.g., vectors suitable for gene therapy), plasmid vectors, bacteriophage vectors, cosmids, phagemids, artificial chromosomes, and the like.

[0352] Non-limiting examples of viral vectors include: retrovirus, e.g., Moloney murine leukemia virus (MMLV), Harvey murine sarcoma virus, murine mammary tumor virus, and Rous sarcoma virus; adenoviruses, adeno-associated viruses; SV40-type viruses; polyomaviruses; Epstein-Barr viruses; papilloma viruses; herpes viruses; vaccinia viruses; and polio viruses.

[0353] In some embodiments, the codon-altered polynucleotides described herein are integrated into a gene therapy vector. In some embodiments, the gene therapy vector is a retrovirus, and particularly a replication-deficient retrovirus. Protocols for the production of replication-deficient retroviruses are known in the art. For review, see Kriegler, M., *Gene Transfer and Expression, A Laboratory Manual*, W.H. Freeman Co., New York (1990) and Murry, E. J., *Methods in Molecular Biology*, Vol. 7, Humana Press, Inc., Clifton, N.J. (1991).

[0354] In one embodiment, the gene therapy vector is an adeno-associated virus (AAV) based gene therapy vector. AAV systems have been described previously and are generally well known in the art (Kelleher and Vos, *Biotechniques*, 17(6):1110-17 (1994); Cotten et al., *Proc Natl Acad Sci USA*, 89(13):6094-98 (1992); Curiel, *Nat Immun*, 13(2-3):141-64 (1994); Muzyczka, *Curr Top Microbiol Immunol*, 158:97-129 (1992); and Asokan A, et al., *Mol. Ther.*, 20(4): 699-708 (2012), each incorporated herein by reference in their entireties for all purposes). Details concerning the generation and use of rAAV vectors are described, for example, in U.S. Pat. Nos. 5,139,941 and 4,797,368, each incorporated herein by reference in their entireties for all purposes. In a particular embodiment, the AAV vector is an AAV-8 vector.

[0355] In some embodiments, the codon-altered polynucleotides described herein are integrated into a retroviral expression vector. These systems have been described previously, and are generally well known in the art (Mann et al., *Cell*, 33:153-159, 1983; Nicolas and Rubinstein, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt, eds., Stoneham: Butterworth, pp. 494-513, 1988; Temin, In: *Gene Transfer*, Kucherlapati (ed.), New York: Plenum Press, pp. 149-188, 1986). In a specific embodiment, the retroviral vector is a lentiviral vector (see, for example, Naldini et al., *Science*, 272(5259): 263-267, 1996; Zufferey et al., *Nat Biotechnol*, 15(9):871-875, 1997; Blomer et al., *J Virol.*, 71(9):6641-6649, 1997; U.S. Pat. Nos. 6,013,516 and 5,994,136).

[0356] A wide variety of vectors can be used for the expression of a Factor VIII polypeptide from a codon-altered polypeptide in cell culture, including eukaryotic and prokaryotic expression vectors. In certain embodiments, a plasmid vector is contemplated for use in expressing a Factor VIII polypeptide in cell culture. In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell are used in connection with these hosts. The vector can carry a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. The plasmid will include the codon-altered polynucleo-

otide encoding the Factor VIII polypeptide, operably linked to one or more control sequences, for example, a promoter.

[0357] Non-limiting examples of vectors for prokaryotic expression include plasmids such as pRSET, pET, pBAD, etc., wherein the promoters used in prokaryotic expression vectors include lac, trc, trp, recA, araBAD, etc. Examples of vectors for eukaryotic expression include: (i) for expression in yeast, vectors such as pAO, pPIC, pYES, pMET, using promoters such as AOX1, GAP, GAL1, AUG1, etc; (ii) for expression in insect cells, vectors such as pMT, pAc5, pIB, pMIB, pBAC, etc., using promoters such as PH, p10, MT, Ac5, OpIE2, gp64, polh, etc., and (iii) for expression in mammalian cells, vectors such as pSVL, pCMV, pRc/RSV, pcDNA3, pBPV, etc., and vectors derived from viral systems such as vaccinia virus, adeno-associated viruses, herpes viruses, retroviruses, etc., using promoters such as CMV, SV40, EF-1, Ubc, RSV, ADV, BPV, and β -actin.

IV. Examples

Example 1—Construction of a Codon Altered Factor VIII Variant Expression Sequence

[0358] Two hurdles had to be overcome in order to create a Factor VIII coding sequence that is effective for gene therapy of hemophilia A. First, because of the genomic size limitations of conventional gene therapy delivery vectors (e.g., AAV virions), the encoded Factor VIII polypeptide had to be shortened considerably. Second, the coding sequence had to be altered to: (i) stabilize packaging interactions within the delivery vector, (ii) stabilize the mRNA intermediary, and (iii) improve the robustness of transcription/translation of the mRNA.

[0359] To achieve the first objective, Applicants started with a B-domain deleted Factor VIII variant construct, referred to herein as “FVIII-BDD-SQ.” In this construct, the B-domain is replaced with a fourteen amino acid sequence referred to as the “SQ” sequence. Recombinant FVIII-BDD-SQ is sold under the trade name REFACTO®, and has been shown to be effective for the management of hemophilia A. However, the native coding sequence for FVIII-BDD-SQ, which includes human wild-type nucleic acid sequences for the Factor VIII heavy and light chains, is ineffectively expressed in gene therapy vectors.

[0360] To address the poor expression of the native FVIII-BDD-SQ, the codon optimization algorithm described in Fath et al. (PLoS ONE, 6:e17596 (2011)), modified as described in Ward et al. (Blood, 117:798 (2011)) and in McIntosh et al. (Blood, 121, 3335-3344 (2013)), was applied to the FVIII-BDD-SQ sequence to create first intermediate coding sequence CS04a. However, Applicants recognized that the CS04a sequence created using the modified algorithm could be improved by further modifying the sequence. Accordingly, Applicants re-introduced CpG dinucleotides, re-introduced the CGC codon for arginine, changed the leucine and serine codon distributions, re-introduced highly conserved codon pairs, and removed cryptic TATA box, CCAAT box, and splice site elements, while avoiding CpG islands and local overrepresentation of AT-rich and GC-rich stretches.

[0361] First, the modified algorithm systematically replaces codons containing CpG-dinucleotides (e.g., arginine codons) with non-CpG-dinucleotide codons, and eliminates/avoids CpG-dinucleotides created by neighboring codons. This strict avoidance of CpG dinucleotides is usu-

ally done to prevent TLR-induced immunity after intramuscular injection of DNA vaccines. However, doing so limits the codon optimization possibilities. For example, the modified algorithm excludes use of the complete set of CGX arginine codons. This is particularly disruptive in the coding of genes for expression in human cells, because CGC is the most frequently used arginine codon in highly expressed human genes. Additionally, avoiding the creation of CpGs by neighboring codons further limits the optimization possibilities (e.g., limits the number of codon pairs that may be used together).

[0362] Because TLR-induced immunity is not expected to be a problem associated with liver-directed, AAV-based gene therapy, codons including CpGs, and neighboring codons creating CpGs, were re-introduced into intermediate coding sequence CS04a, preferentially in the sequence coding for the Factor VIII light chain (e.g., at the 3' end of the FVIII-BDD-SQ coding sequence). This allowed for more frequent use of preferred human codons, particularly those for arginine. Care was taken, however, to avoid creation of CpG islands, which are regions of coding sequence having a high frequency of CpG sites. This is contrary to the teachings of Krinner et al. (Nucleic Acids Res., 42(6):3551-64 (2014)), which suggests that CpG domains downstream of transcriptional start sites promote high levels of gene expression.

[0363] Second, the modified algorithm applies certain codons exclusively, such as CTG for leucine, GTG for valine, and CAG for glutamine. However, this offends the principles of balanced codon use, for example, as proposed in Haas et al. (Current Biology, 6(3):315-24 (1996)). To account for the overuse of preferred codons by the modified algorithm, alternate leucine codons were re-introduced where allowed by the other rules applied to the codon alteration (e.g., CpG frequency and GC content).

[0364] Third, the modified algorithm replaces codon pairs without regard to how conserved they are in nature, when certain criteria (e.g., the presence of CG-dinucleotides) are met. To account for beneficial properties which may have been conserved by evolution, the most conserved codon pairs that were replaced by the algorithm and the most conserved preferred codon pairs, e.g., as described in Tats et al. (BMC Genomics 9:463 (2008)), were analyzed and adjusted where allowed by the other rules applied to the codon alteration (e.g., CpG frequency and GC content).

[0365] Fourth, serine codons used in the intermediate coding sequence were also re-engineered. Specifically, AGC, TCC, and TCT serine codons were introduced into the modified coding sequence with higher frequency, to better match overall for human codon usage (Haas et al., supra).

[0366] Fifth, TATA box, CCAAT box elements, and intron/exon splice sites were screened and removed from the modified coding sequence. When modifying the coding sequence, care was taken to avoid local overrepresentation of AT-rich or GC rich stretches.

[0367] Finally, in addition to optimizing the codon usage within the coding sequence, the structural requirements of the underlying AAV virion were considered when further refining the intermediate coding sequence CS04a. AAV vectors (e.g., the nucleic acid portion of an AAV virion) are packaged as single stranded DNA molecules into their capsids (for review, see, Daya and Berns, Clin. Microbiol Rev., 21(4):583-93 (2008)). The GC content of the vector is therefore likely to influence packaging of the genome and,

thus, vector yields during production. Like many algorithms, the modified algorithm used here creates an optimized gene sequence with a GC content of at least 60% (see, Fath et al., PLoS One, 6(3):e17596 (2011) (erratum in: PLoS One, (6)3 (2011)). However, the AAV8 capsid protein is encoded by a nucleotide sequence having a lower GC content of about 56%. Thus, to better mimic the native AAV8 capsid protein coding sequence, the GC content of the intermediate coding sequence CS04a was reduced to 56%.

[0368] The resulting CS04 coding sequence, shown in FIG. 2, has an overall GC content of 56%. The CpG-dinucleotide content of the sequence is moderate. However, CpG dinucleotides are predominantly present in the downstream portion of the coding sequence, e.g., the portion coding for the Factor VIII light chain. The CS04 sequence has 79.77% nucleotide sequence identity to the corresponding coding sequences in wild-type Factor VIII (Genbank accession M14113).

[0369] For comparison purposes, several other codon-optimized, ReFacto constructs were prepared. CS01 was constructed by applying the codon-optimization algorithm of Fath et al., as modified by Ward et al., as done for CS04. However, unlike CS04, the CS01 construct does not contain any CpG islands. The CS08 ReFacto construct was codon-optimized as described in Radcliff P. M. et al., Gene Therapy, 15:289-97 (2008), the content of which is hereby expressly incorporated by reference herein, in its entirety, for all purposes. The CS10 codon-optimized ReFacto construct was obtained from Eurofins Genomics (Ebersberg, Germany). The CS11 codon-optimized ReFacto construct was obtained from Integrated DNA Technologies, Inc. (Coralville, USA). The CH25 codon-optimized ReFacto construct was obtained from ThermoFischer Scientific's GeneArt services (Regensburg, Germany). The CS40 ReFacto construct consists of the wild type Factor VIII coding sequence. The algorithm used to construct CS23 is based on the JCAT tool (www.jcat.de), an on-line tool for codon-optimizations (Grote et al., 2005; Nucl. Acids Res. W526-31). The sequence was further modified to more reflect the codon usage of the albumin superfamily (Mirsafian et al. 2014; Sc. Word Journal 2014, ID 639682). The sequence identities shared between each of the ReFacto coding sequences is shown in Table 2, below.

TABLE 2

Percent identity matrix for codon-altered Factor VIII constructs.								
	CS01	CS04	CS08	CS10	CS11	CS40	CH25	CS23
CS01	100%							
CS04	93.0%	100%						
CS08	80.7%	82.2%	100%					
CS10	79.1%	79.4%	78.4%	100%				
CS11	78.3%	78.3%	78.1%	77.5%	100%			
CS40	79.6%	79.8%	76.7%	77.6%	75.4%	100%		
CH25	81.3%	85.1%	85.0%	79.9%	79.4%	75.8%	100%	
CS23	84.3%	89.2%	85.1%	80.3%	79.9	76.5%	93.2%	100%

[0370] Plasmids of each construct were constructed by cloning different synthetic DNA fragments into the same vector backbone plasmid (pCh-BB01). DNA synthesis of the Refacto-type BDD-FVIII fragments with flanking *AscI* and *NotI* enzyme restriction sites were done by ThermoFischer Scientific (Regensburg, Germany). The vector backbone contains two flanking AAV2-derived inverted terminal

repeats (ITRs) that encompass a promoter/enhancer sequence derived from the liver-specific murine transthyretin gene, *AscI* and *NotI* enzyme restriction sites for insertion of the respective Refacto-type BDD-FVIII and a synthetic polyA site. After ligation of the prepared vector backbone and inserts via the *AscI* and *NotI* sites, the resulting plasmids were amplified in milligram scale. The Refacto-type BDD-FVIII sequences of the constructs were verified by direct sequencing (Microsynth, Balgach, Switzerland). The cloning resulted in seven different plasmid constructs named pCS40, pCS01, pCS04, pCS08, pCS10, pCS11, and pCh25 (FIG. 23). The constructs have the same vector backbone and encode the same B-domain deleted FVIII protein (Refacto-type BDD-FVIII), but differ in their FVIII coding sequence.

[0371] AAV8-based vectors were prepared by the three plasmid transfection method, as described in Grieger J C, et al. (Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector, Mol Ther., October 6. (2015) doi: 10.1038/mt.2015.187. [Epub ahead of print]), the content of which is hereby expressly incorporated by reference herein, in its entirety, for all purposes. HEK293 suspensions cells were used for plasmid transfections using the corresponding FVIII vector plasmid, the helper plasmid pXX6-80 (carrying adenoviral helper genes), and the packaging plasmid pGSK2/8 (contributing the rep2 and cap8 genes). To isolate the AAV8 constructs, the cell pellets of one liter cultures were processed using iodixanol gradients, as described in Grieger et al. (2015, Supra). The procedure resulted in vector preparations called vCS01, vCS04, vCS08, vCS10, vCS11, and vCH25. Vectors were quantified by qPCR using the universal qPCR procedure targeting the AAV2 inverted terminal repeats (Aurnhammer, Human Gene Therapy Methods: Part B 23:18-28 (2012)). A control vector plasmid carrying AAV2 inverted terminal repeats served for preparing the standard curve. The resulting vCS04 construct is presented as SEQ ID NO: 8 in FIGS. 7A-7C.

[0372] The integrity of the vector genomes was analyzed by AAV agarose gel electrophoresis. The electrophoresis was performed as described in Fagone et al., Human Gene Therapy Methods 23:1-7 (2012). Briefly, AAV vector prepara-

tions were incubated at 75° C. for 10 minutes in the presence of 0.5% SDS and then cooled down to room temperature. Approximately 1.5E10 vector genomes (vg) were loaded per lane on a 1% 1×TAE agarose gel and electrophoresed for 60 min at 7 V/cm of gel length. The gel was then stained in 2× GelRed (Biotium Cat#41003) solution and imaged by ChemiDocTMMP (Biorad). The results

shown in FIG. 24 demonstrate that the vCS01, vCS04, and vCS40 viral vectors have the same-sized genome, indicated by a distinct band in the 5 kb range (FIG. 24, lanes 2-4). Despite a vector size of approx. 5.2 kb, the genome is a homogenous band confirming correct packaging of the somewhat oversized genome (relative to an AAV wild-type genome of 4.7 kb). All other vCS vector preparations show the same genomic size (data not shown).

[0373] In order to confirm the expected pattern of capsid proteins, SDS PAGE followed by silver staining was performed with the vectors vCS01, vCS04, and vCS40 (FIG. 25). As shown in the figure, the downstream purification procedure resulted in highly purified material displaying the expected protein pattern of VP1, VP2 and VP3 (FIG. 25, lanes 2-4). The same pattern was seen with all other viral preparations (not shown). The SDS-PAGE procedure of AAV preparations was done according to standard procedures. Each lane contained 1E10 vg of the respective viral construct, and were separated on a 4-12% Bis-Tris (NuPAGE® Novex, Life Technologies) gel as per manufacturer's instructions. Silver staining was performed with a SilverQuest™ kit (Novex, Life Technologies) according to the manufacturer's instructions.

[0374] Surprisingly, AAV vectors vCS01 and vCS04 had higher virion packaging, measured by higher yields in AAV virus production, as compared to the vCS40 wild-type coding construct and the other codon-optimized constructs. As shown in Table 3, the vCS01 and vCS04 vectors replicated substantially better than vCS40, providing a 5-7 fold yield increase in AAV titer.

TABLE 3

Yields per liter cell culture obtained with AAV vector constructs vCS01, vCS04, and vCD40, as purified from cell pellets.			
Construct	Vector concentration [vg/ml] × 10E12	Yields [vg/liter] × 10E12	Fold increase vs wt
vCS40	2.0	11.0	—
vCS01	9.2	51.4	4.7
vCS04 - Sample 1	17.6	79.2	7.2
vCS04 - Sample 2	15.9	58.8	5.4

Example 2—In Vivo Expression of Codon Altered Factor VIII Variant Expression Sequences

[0375] To test the biological potency of the codon-altered Factor VIII variant sequences, the ReFacto-type FVIII constructs described in Example 1 were administered to mice lacking Factor VIII. Briefly, the assays were performed in C57Bl/6 FVIII knock-out (ko) mice (with 6-8 animals per group) by tail vein injection of 4E12 vector genomes (vg) per kilogram body weight of mouse. Blood was drawn 14 days after injection by retroorbital puncture and plasma was prepared and frozen using standard procedures. Expression levels at day 14 were chosen because there is minimal influence of inhibitory antibodies at this time, which are seen in some animals of this mouse model at later times. FVIII activity in the mouse plasma was determined using the Technochrome FVIII assay performed, with only minor modifications, as suggested by the manufacture (Technoclone, Vienna, Austria). For the assay, the plasma samples were appropriately diluted and mixed with assay reagents, containing thrombin, activated factor IX (FIXa), phospho-

lipids, factor X and calcium. Following FVIII activation by thrombin a complex with FIXa, phospholipids and calcium is formed. This complex activates FX to activated FX (FXa) which in turn cleaves para-nitroanilide (pNA) from the chromogenic substrate. The kinetics of pNA formation is measured at 405 nm. The rate is directly proportional to the FVIII concentration in the sample. FVIII concentrations are read from a reference curve and results are given in IU FVIII/milliliter.

[0376] The results, presented in Table 4 below, demonstrate that the codon-altered sequences designed using commercial algorithms (CS10, CS11, and CH25) provided only a modest increase in BDD-Factor VIII (3-4 fold) as compared to the wild-type BDD-Factor VIII construct (CS40). Similarly, the codon-altered BDD-Factor VIII construct prepared as described in Radcliffe et al. (CS08), only provided a 3-4 fold increase in BDD-FVIII expression. This result is consistent with the results reported in Radcliff et al. Surprisingly, the CS01, CS04, and CS23 constructs provided much higher BDD-FVIII expression in the in-vivo biopotency assays (18-, 74-, and -30-fold increases, respectively).

TABLE 4

Expression of FVIII in the plasma of FVIII-knock-out mice induced by the different AAV vector constructs.					
Construct	Codon Algorithm	Average FVIII Expression at Day 14 [IU/ml]	Standard deviation	Number of mice	Fold increase vs wt
vCS40	Human wild-type	0.03	0.03	12	—
vCS01	Applicants'	0.55	0.28	22	18.3
vCS04	Applicants'	2.21	1.20	55	73.7
vCS08	Radcliffe et al.	0.11	0.01	6	3.6
vCS10	Eurofins	0.09	0.01	7	3.0
vCS11	IDT	0.08	0.02	8	2.7
vCH25	GeneArt	0.13	0.12	18	4.3
vCS23	Applicants'	0.91	0.32	5	30.3

Example 3—Design of Glycosylation Peptides for the B-Domain Substituted Linker

[0377] Others have shown that inclusion of a small peptide (the "V3 peptide") containing six putative N-linked glycosylation sites from the wild-type Factor VIII B-domain, into a B-domain deleted gene therapy construct, increased Factor VIII levels in the plasma of mice (McIntosh et al., Blood 121(17):3335-44 (2013)). However, in order to maintain the small size of the B-domain substituted linker, the glycosylation sites were taken out of the context of the wild-type B-domain. In silico prediction (Gupta et al., Supra) of the linker containing the V3 peptide suggests that only two of these glycosylation sites in the V3 peptide will be modified in vivo (FIG. 15).

[0378] Thus, Applicants attempted to identify alternative glycosylation peptides that would support higher levels of glycosylation in vivo, which matched wild type glycosylation more closely than the V3 peptide. Applicants designed and tested several alternative glycosylation peptides, in silico. Several of these peptides, shown in FIGS. 13A-13B, were predicted to have equal or greater glycosylation in vivo than the V3 peptide, when placed between amino acids N768 and P769 of the B-domain substituted linker in SEQ

ID NO:2. The results of the *in silico* predictions are shown in Table 5, below. Table 5 also reports the results of expression experiments performed for several constructs encoding a Refacto-type Factor VIII protein with a glycosylation peptide incorporated into the B-domain substituted linker, in a CS01 codon-optimized background.

TABLE 5

Prediction of N-glycosylation in B-domain substituted linker peptides and performance of AAV vector constructs <i>in vivo</i> .					
Sequence	Number of Predicted N-glycosylation sites	Day 28 expression [IU/ml]	SD	Number of mice [n]	Fold expression
vCS01	0	0.74	0.52	5	21
vNG1/CS01	4	n.d.	—	—	—
vNG4/CS01	3	1.93	0.57	6	55
vNG5/CS01	2	n.d.	—	—	—
vNG6/CS01	1	0.80	0.67	5	23
vNG9/CS01	1	n.d.	—	—	—
vNG10/CS01	2	2.66	0.52	6	76
vNG16/CS01	2	1.59	0.57	6	45
vNG17/CS01	2	n.d.	—	—	—
vNG18/CS01	2	n.d.	—	—	—
vNG19/CS01	2	0.88	0.25	5	25
vNG20/CS01	2	n.d.	—	—	—
vNG21/CS01	2	n.d.	—	—	—
vCS40	0	0.035	0.030	12	1

[0379] AAV vectors containing the NG variants were constructed as described in Example 1 and tested in FVIII knock-out mice as described in Example 2. All virus vectors (except the control vector vCS40) shown in Table 5 are based on the algorithm as used in vCS01. A parallel set of constructs using the algorithm of vCS04 was also prepared (vNG/CS04 series) and is tested in the mouse model. Results were compared to the expression levels achieved with the wild-type vCS40 construct. The day 28 expression levels were chosen in this example, because expression levels of the majority of construct reached the highest levels at this time point. Three AAV vectors achieved greater than 40-fold FVIII expression levels including vNG4/CS01, vNG10/CS01 and vNG16/CS01 (Table 5). The corresponding constructs vNG4/CS04, vNG10/CS04 and vNG16/CS04 are expected to show even higher expression because they are based on the superior vCS04 algorithm.

[0380] Surprisingly, the AAV vectors of the vNG/CS01 series had higher virion packaging, measured by higher yields in AAV virus production, as compared to the vCS40 wild-type coding construct. As shown in Table 6, the vNG/CS01-based vectors replicated substantially better than vCS40, providing an approximately 3-fold yield increase in AAV titer.

TABLE 6

Yields per liter cell culture obtained with AAV vector constructs as purified from cell pellets.			
Sequence	Vector conc. [vg/ml] × 10 ¹²	Yields [vg/liter] × 10 ¹²	Fold increase vs wild-type
vCS01	9.17	51.35	4.7
vNG1/CS01	2.13	17.04	1.5
vNG4/CS01	5.74	33.01	3.0
vNG5/CS01	6.91	27.29	2.5

TABLE 6-continued

Yields per liter cell culture obtained with AAV vector constructs as purified from cell pellets.			
Sequence	Vector conc. [vg/ml] × 10 ¹²	Yields [vg/liter] × 10 ¹²	Fold increase vs wild-type
vNG6/CS01	7.01	40.66	3.7
vNG9/CS01	6.39	29.39	2.7
vNG10/CS01	8.57	37.71	3.4
vNG16/CS01	5.3	28.36	2.6
vNG17/CS01	4.24	32.22	2.9
vNG18/CS01	6.11	37.88	3.4
vNG19/CS01	9.42	39.56	3.6
vNG20/CS01	4.09	30.27	2.8
vNG21/CS01	n.d.	—	—
vCS40	2.03	11	1.0

Example 4—Construction of Mutant BDD-FVIII Constructs

[0381] Numerous different mutated Refacto-type BDD-FVIII constructs, carrying amino acid mutations within the Factor VIII heavy chain and/or B-domain substituted linker, were cloned and screened. The corresponding vectors, as referred to herein as the “vCS” series of vectors, encode BDD-FVIII variants in the CS01, CS04, and CS23 codon-altered backgrounds. The method used to construct the CS01 and CS04 backgrounds is described in Example 1. The method used to construct CS23 was based on the JCAT tool (www.jcat.de), an on-line tool for codon-optimizations (Grote et al., 2005; Nucl. Acids Res. W526-31). The sequence was further modified to better reflect the codon usage of the albumin superfamily (Mirafian et al., Sc. Word Journal, ID 639682 (2014)), the content of which is hereby expressly incorporated by reference, in its entirety, for all purposes.

[0382] Combinations of three types of mutations were included in the FVIII sequences of the vCS series of constructs. The first amino acid change introduced into the FVIII sequence is the X1 mutation (TTYVNRSL (SEQ ID NO: 33); X. Xiao), which introduces an additional glycosylation site near the B-domain substituted linker. The X1 mutation is also referred to herein as the “m3” mutation. The second amino acid change made in the FVIII sequence includes the F328S (SPI, F309S SPE) mutation, an amino acid change known to improve secretion of FVIII (Swaaroop, J. Biol. Chem., 272:24121-24 (1997)). This mutation is also referred to herein as the “m1” mutation. The third change is the so-called X5 mutation, which is a combination of five amino acid changes in the A1 domain of the heavy chain that improves specific activity and secretion of BDD-FVIII (Cao et al., 2014; ASGCT abstract #460; details of mutations disclosed in oral presentation). The X5 mutation is also referred to herein as the “m2” mutation. Next, combinations of X1 and F328S (SPI, F309S SPE) were made, followed by combinations of X1 and X5, also referred to as “X6,” and yet other combinations of X5 and F328S (SPI, F309S SPE) were made (Table 7).

[0383] Gene Synthesis and Cloning of the Vector Plasmids.

[0384] The plasmids were constructed by cloning different synthetic DNA fragments into the same vector backbone plasmid (pCh-BB01). DNA synthesis of the Refacto-type BDD-FVIII fragments with flanking *AscI* and *NotI* enzyme restriction sites were done by ThermoFischer Scientific

(Regensburg, Germany). The vector backbone contains two flanking AAV2-derived inverted terminal repeats (ITRs) that encompass a promoter/enhancer sequence derived from the liver-specific murine transthyretin gene, *AscI* and *NotI* enzyme restriction sites for insertion of the respective Refacto-type BDD-FVIII, and a synthetic polyA site. After ligation of the prepared vector backbone and insertions via the *AscI* and *NotI* sites, the resulting plasmids were amplified in milligram scale. The Refacto-type BDD-FVIII sequences of the constructs were verified by direct sequencing (Microsynth, Balgach, Switzerland). The cloning resulted in different plasmid constructs, as shown in FIG. 44.

[0385] Small Scale Vector Preparations and Quantification by Quantitative PCR (qPCR).

[0386] AAV8-based vectors were prepared by the three plasmid transfection method essentially as described in Grieger et al. (2015, Supra). HEK293 suspensions cells were used for plasmid transfections using the corresponding FVIII vector plasmid, the helper plasmid pXX6X80 (carrying adenoviral helper genes) and the packaging plasmid pGSK2/8 (contributing the rep2 and cap8 genes). In the downstream process the cell pellet of a one liter culture was processed using iodixanol gradients as described above. The procedure resulted in vector preparations as outlined in Table 8. Vectors were quantified by qPCR using the universal qPCR procedure targeting the AAV2 inverted terminal repeats (Aurnhammer, HUMAN GENE THERAPY METHODS: Part B 23:18-28 (2012)). An accurately quantified vector plasmid carrying AAV2 Inverted terminal repeats served for preparing the standard curve.

[0387] AAV Vector Characterizations.

[0388] The integrity of the vector genome was analyzed by AAV agarose gel electrophoresis. The electrophoresis was done similar as described in Fagone et al. (Human Gene Therapy Methods, 23:1-7 (2012)). AAV vector preparations were incubated at 75° C. for 10 minutes in the presence of 0.5% SDS and then cooled down to room temperature. Approximately 1.5E10 vector genomes (vg) were loaded per lane on a 1% 1×TAE agarose gel and electrophoresed for 60 min at 7 V/cm of gel length. The gel was then stained in 2× GelRed (Biotium Cat#41003) solution and imaged by ChemiDoc™MP (Biorad). The results of a selection of vectors are shown in FIG. 45. The viral vectors vCS04 (control), vCS17, vCS20, vCS24, vCS16 and vCS40 (control) show all the same-sized genome as a distinct band in the 5 kb range (FIG. 45, lanes 2-7; arrow right side). Despite a vector size of approx. 5.2 kb, the genome is a homogenous band confirming correct packaging of the somewhat oversized genome (relative to an AAV wild-type genome of 4.7 kb).

[0389] In order to confirm purity of the vector and the expected pattern of capsid proteins, SDS PAGE followed by silver staining was performed with the vectors, as shown in FIG. 46. As shown in the figure, the downstream purification procedure resulted in highly purified material displaying the expected protein pattern of VP1, VP2 and VP3 (FIG. 46 lanes 2-9; arrows right hand side). The SDS-PAGE procedure of AAV preparations was done according to standard procedures. The amounts of 1E10 vg per lane were separated on a 4-12% Bis-Tris (NuPAGE® Novex, Life Technologies) gel as per manufacturer's instructions. Silver staining was performed with a SilverQuest™ kit (Novex, Life Technologies) according to the instructions of the manufacturer.

[0390] In-Vivo Biopotency Screening of Vectors.

[0391] The different Refacto-type BDD-FVIII constructs were screened in mice. The assay was performed in C57Bl/6 FVIII knock-out (ko) mice (with 6-8 animals per group) by tail vein injection of 4E12 vector genomes (vg) per kilogram body weight of mouse. Blood was drawn 14 days after injection by retroorbital puncture and plasma was prepared and frozen using standard procedures. FVIII activity in mouse plasma was determined with a chromogenic assay from Technoclone with minor modifications (Technochrome FVIII, Technoclone, Vienna, Austria). In brief, the plasma sample was appropriately diluted and mixed with assay reagents, containing thrombin, activated factor IX (FIXa), phospholipids, factor X and calcium. Following FVIII activation by thrombin a complex with FIXa, phospholipids and calcium is formed. This complex activates FX to activated FX (FXa) which in turn cleaves para-nitroanilide (pNA) from the chromogenic substrate. The kinetics of pNA formation is measured at 405 nm. The rate is directly proportional to the FVIII concentration in the sample. FVIII concentrations are read from a reference curve and results are given in IU FVIII/milliliter.

[0392] The results of the mouse biopotency assay (day 14 expression data of FVIII in international units per milliliter [IU/ml] in mouse plasma and fold expression compared to the wild-type vCS40 control) are shown in Table 7. AAV vectors vCS19, vCS26 and vCS32 all contain the X1 glycosylation site in the CS01, CS04, and CS23 codon-altered backgrounds, respectively. As seen in Table 7, surprisingly high expression levels were obtained, as compared to the wild-type construct vCS40 (level defined as 1). vCS26, for instance, expressed 202-fold higher levels compared to the wild-type vCS40 vector. Another control construct for the X1-series of vectors, vCH111, that contains the X1 mutation in the Geneart codon context, showed a more modest increase in expression (12-fold).

[0393] Vectors vCS16, vCS28, and vCS34 all contain the F328S (SPI, F309S SPE) mutation enhancing secretion in the CS01, CS04, and CS23 codon-altered backgrounds, respectively. As seen in Table 7, high expression levels (45-93-fold higher than the wt vCS40 control) were obtained with vCS16 and vCS28.

[0394] Vectors vCS20, vCS24, and vCS33 contain the X5 mutation in the CS01, CS04, and CS23 codon-altered backgrounds, respectively. The best performing variant in the X5 series was vCS20, achieving levels of >3 units/ml after day 14 and a 121-fold increase over the wt vCS40 control.

[0395] Vectors vCS17, vCS29, and vCS31 contain the combination of the X1 and F328S (SPI, F309S SPE) mutations in the CS01, CS04, and CS23 codon-altered backgrounds, respectively (Table 6). The vCS17 and vCS29 constructs achieved very high expression levels in the mouse studies (115 to 246-fold increase over the vCS40 control). Remarkably, in the FVIII KO mouse model used, the majority of mice treated with the vCS17 construct did not develop neutralizing antibodies over time, evidenced by increasing levels of FVIII at later time points (e.g., day 28 and day 42; data not shown). This is an unexpected finding, because in some other constructs the expression levels began to decrease with time due to the formation of neutralizing antibodies. The CS01 background combined with the secretion-enhancing mutations F328S (SPI, F309S SPE) and X1 resulted in low immunogenicity induction.

[0396] Vectors vCS18, vCS27, and vCS35 contain the combination of the X1 and X5 mutations in the CS01, CS04,

and CS23 codon-altered backgrounds, respectively. The combination of these two mutations was also very efficient. A 145-fold increase over the vCS40 control could be achieved with vCS18, for example (Table 7).

[0397] Vectors vCS48 and vCS49 contain the combination of the X5 and F328S (SPI, F309S SPE) mutations in the CS01 and CS04 codon-altered backgrounds, respectively. The combination of these two mutations was also very efficient. One of the largest increases of all mutants, a 239-fold increase over the vCS40 control, could be achieved with vCS49 confirming the special value of the combinations including the F328S (SPI, F309S SPE) mutation.

[0398] A further surprising observation was that the mutant AAV vectors grew substantially better than the vCS40 construct harboring the wild-type BDD-FVIII codons. Sequence optimization resulted in a several-fold yield increase in vector production. In some of the best expressing constructs (e.g., vCS29, vCS17, vCS20, and vCS26) the increase in yields due to codon-alteration and/or mutant sequence was approximately 3-5-fold higher, as compared to the wild-type vector (Table 8).

[0399] Expression of BDD-FVIII in the plasma of FVIII-knock-out mice induced by the different AAV vector constructs is shown in Table 7. The constructs have the same vector backbone, however, encode different types of mutated FVIII, including different codon optimization backgrounds. Expression levels at day 14 were chosen because at this time point there is minimal influence of inhibitory antibodies usually seen in some animals in the mouse model at later times. N.d., not determined.

TABLE 7

In vivo biopotency data of vCS constructs.					
#	Algorithm, mutations	Day 14 expression [IU/ml]	SD	Number of mice [n]	Fold expression
<u>Vector</u>					
1	vCS19 CS01, X1	2.34	1.10	13	78
2	vCS26 CS04, X1	6.07	2.72	12	202
3	vCS32 CS23, X1	n.d.	—	—	—
4	vCS16 CS01, F328S	1.35	0.88	6	45
5	vCS28 CS04, F328S	2.78	0.92	7	93
6	vCS34 CS23, F328S	n.d.	—	—	—
7	vCS20 CS01, X5	3.62	1.96	21	121
8	vCS24 CS04, X5	0.79	0.89	18	26
9	vCS33 CS23, X5	n.d.	—	—	n.d.
10	vCS17 CS01, X1, F328S	3.44	1.92	20	115
11	vCS29 CS04, X1, F328S	7.39	2.64	9	246
12	vCS31 CS23, X1, F328S	n.d.	—	—	n.d.
13	vCS18 CS01, X1 + X5 (X6)	4.34	2.50	6	145
14	vCS27 CS04, X1 + X5 (X6)	8.03	3.97-	6-	268-
15	vCS35 CS23, X1 + X5 (X6)	n.d.	—	—	—
19	vCS48 CS01, X5, F328S	2.54	0.72	8	85
20	vCS49 CS04, X5, F328S	7.17	1.30	7	239
<u>controls</u>					
16	vCS40 Human wild-type	0.03	0.03	12	1
17	vCh25 Geneart	0.13	0.12	18	4
18	vCh111 Geneart + X1	0.37	0.21	17	12

TABLE 8

Yields per liter cell culture (packaging efficiency) obtained with the different AAV vector constructs. The vectors were purified out of the cell pellets;					
con-struct	Algorithm, mutations	Vector conc. [vg/ml] × 10 ¹²	Yields [vg/liter] × 10 ¹²	Fold increase vs wt	
1	vCS19 CS01, X1	9.71	36	3.22	
2	vCS26 CS04, X1	5.93	32	2.87	
3	vCS32 CS23, X1	n.d.	n.d.	n.d.	
4	vCS16 CS01, F328S	6.51	29	2.56	
5	vCS28 CS04, F328S	5.85	32	2.88	
6	vCS34 CS23, F328S	n.d.	n.d.	n.d.	
7	vCS20 CS01, X5	9.90	50	4.48	
8	vCS24 CS04, X5	3.00	16	1.46	
9	vCS33 CS23, X5	n.d.	n.d.	n.d.	
10	vCS17 CS01, X1, F328S	8.94	37	3.34	
11	vCS29 CS04, X1, F328S	7.42	53	4.72	
12	vCS31 CS23, X1, F328S	n.d.	n.d.	n.d.	
13	vCS18 CS01, X1 + X5 (X6)	21.20	53	4.75	
14	vCS27 CS04, X1 + X5 (X6)	4.15	19	1.67	
15	vCS35 CS23, X1 + X5 (X6)	n.d.	n.d.	n.d.	
16	vCS48 CS01, X5, F328S	7.14	42.1	3.77	
17	vCS49 CS04, X5, F328S	8.27	37.2	3.33	
18	vCS40 Human wild-type	2.03	11	1.00	

n.d., not determined.

[0400] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 109

<210> SEQ ID NO 1

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 1

```
atgcagattg agctgagcac ctgcttcttc ctgtgctgc tgaggttctg cttctctgcc      60
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg      120
ggggagctgc ctgtggatgc caggttocca cccagagtgc ccaaatecct cccattcaac      180
acctctgtgg tctacaagaa gacctcttt gtggagtcca ctgaccacct gttcaacatt      240
gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgagggtgat      300
gacactgtgg tcatcacocct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg      360
ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg      420
gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggtctct      480
aaggagaatg gccccatggc ctctgacca ctctgcctga cctactecta cctttctcat      540
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcaggagg      600
ggctccctgg ccaagagaa gaccagacc ctgcacaagt tcattctcct gtttgctgtc      660
tttgatgagg gcaagagctg gcaactctga accaagaact ccctgatgca ggacagggat      720
gctgcctctg ccagggctg gcccaagatg cacactgtga atggctatgt gaacaggagc      780
ctgcctggac tcattggctg ccacagaaa tctgtctact ggcattgtgat tggcatgggg      840
acaaccctg aggtgcactc cttttctctg gagggccaca ccttctctgt caggaaccac      900
agacaggcca gcctggagat cagccccatc accttctca ctgcccagac cctgctgatg      960
gacctcggac agttcctgct gttctgccc atcagctccc accagcatga tggcatggag      1020
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag      1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat      1140
gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca cccaagacc      1200
tgggtgcact acattgtctg tgaggaggag gactgggact atgccccact ggtcctggcc      1260
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gccacagag gattggacgc      1320
aagtacaaga aagtcagggt catggcctac actgatgaaa cttcaagac caggagggcc      1380
attcagcatg agtctggcat cctgggcccc ctctgtatg gggagggtgg ggacaccctg      1440
ctcatcatct tcaagaacca ggctccagg cctacaaca tctaccaca tggcatcact      1500
gatgtcaggc ccctgtacag ccgaggctg ccaagggggg tgaaacacct caaggacttc      1560
cccattctgc ctggggagat cttcaagtac aagtggactg tcaactgtga ggatggacca      1620
accaaactctg accccagggt cctcaccaga tactactcca gctttgtgaa catggagagg      1680
gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag      1740
aggggaaacc agatcatgtc tgacaagagg aatgtgatcc tgttctctgt ctttgatgag      1800
aacaggagct ggtacctgac tgagaacatt cagcgttcc tgcccacccc tgctggggtg      1860
cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggcatatgt      1920
```

-continued

tttgacagcc tccagctttc tgtctgctg catgaggtgg cctactggta cattctttct 1980
attggggccc agactgaatt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040
atggtgtatg aggacacct gacctcttc ccattctctg gggagactgt gttcatgagc 2100
atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160
atgactgccc tgctcaaagt ctctcctgt gacaagaaca ctggggacta ctatgaggac 2220
agctatgagg acatctctgc ctacctgtc agcaagaaca atgccattga gcccaggagc 2280
ttcagccaga atccacctgt cctgaaacgc caccagaggg agatcaccag gaccacctc 2340
cagtctgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag 2400
gactttgaca tctatgacga ggacgagaac cagagcccaa ggagcttcca gaagaagacc 2460
aggcactact tcattgtctg tgtggagcgc ctgtgggact atggcatgag ctccagcccc 2520
catgtcctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtggcttctc 2580
caagagttca ctgatggcag cttcaccag cccctgtaca gaggggagct gaatgagcac 2640
ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgacctc 2700
cgcaaccagg cctccaggcc ctacagctc tacagctccc tcatcagcta tgaggaggac 2760
cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac 2820
ttctgaaag tccagcacca catggccccc accaaggatg agtttgactg caaggcctgg 2880
gcctacttct ctgatgtgga cctggagaag gatgtgact ctggcctgat tggccactc 2940
ctggtctgcc acaccaacac cctgaacct gcccatggaa ggcaagtgac tgtgcaggag 3000
tttgccctct tcttaccat ctttgatgaa accaagagct ggtacttca tgaagaatg 3060
gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaac 3120
taccgcttcc atgccatcaa tggtctacac atggacaccc tgctgggct tgtcatggcc 3180
caggaccaga ggatcagggt gtacctgctt tctatgggct ccaatgagaa cattcactcc 3240
atccacttct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggcctg 3300
tacaacctct acctgggggt ctttgagact gtggagatgc tgccctcaa agctggcatc 3360
tgagggtgg agtgctctat tggggagcac ctgcatgctg gcatgagcac cctgttctg 3420
gtctacagca acaagtgcc gacccccctg ggaatggcct ctggccacat cagggaactc 3480
cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac 3540
tctggatcca tcaatgctg gagcaccaag gagccattca gctggatcaa agtggacctg 3600
ctggccccca tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc 3660
ctgtacatca gccagttcat catcatgtac agcctggatg gcaagaaatg gcagacctac 3720
agaggcaact ccaactggaac actcatggtc ttctttggca atgtggacag ctctggcactc 3780
aagcacaaca tcttcaacc cccaatcact gccagatata tcaggctgca cccaccacc 3840
tacagcatcc gcagcacct caggatggag ctgatgggct gtgacctgaa ctctgcagc 3900
atgcccctgg gcatggagag caaggccatt tctgatgccc agatcactgc ctccagctac 3960
ttcaccaca tgtttgccac ctggagcccc agcaaggcca ggctgcacct ccagggaagg 4020
agcaatgctt ggaggcccc ggtcaacaac ccaaaggagt ggctgcagggt ggacttccag 4080
aagacctga aggtcactgg ggtgaccacc caggggttca agagcctgct caccagcatg 4140
tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctcttctc 4200

-continued

cagaatggca aggtcaaggt gttccagggc aaccaggaca gcttcacccc tgtgggtgaa 4260
 agcctggacc cccccctoct gaccagatac ctgaggattc acccccagag ctgggtccac 4320
 cagattgccc tggagatgga ggtcctggga tgtgaggccc aggacctgta ctga 4374

<210> SEQ ID NO 2
 <211> LENGTH: 1457
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 2

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
 1 5 10 15
 Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
 20 25 30
 Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
 35 40 45
 Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
 50 55 60
 Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
 65 70 75 80
 Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
 85 90 95
 Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
 100 105 110
 His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
 115 120 125
 Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
 130 135 140
 Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
 145 150 155 160
 Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
 165 170 175
 Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
 180 185 190
 Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
 195 200 205
 Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
 210 215 220
 Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
 225 230 235 240
 Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
 245 250 255
 Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
 260 265 270
 Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
 275 280 285
 Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
 290 295 300
 Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met

-continued

305					310						315					320
Asp	Leu	Gly	Gln	Phe	Leu	Leu	Phe	Cys	His	Ile	Ser	Ser	His	Gln	His	
				325					330					335		
Asp	Gly	Met	Glu	Ala	Tyr	Val	Lys	Val	Asp	Ser	Cys	Pro	Glu	Glu	Pro	
			340					345					350			
Gln	Leu	Arg	Met	Lys	Asn	Asn	Glu	Glu	Ala	Glu	Asp	Tyr	Asp	Asp	Asp	
		355					360					365				
Leu	Thr	Asp	Ser	Glu	Met	Asp	Val	Val	Arg	Phe	Asp	Asp	Asp	Asn	Ser	
	370					375					380					
Pro	Ser	Phe	Ile	Gln	Ile	Arg	Ser	Val	Ala	Lys	Lys	His	Pro	Lys	Thr	
	385				390					395					400	
Trp	Val	His	Tyr	Ile	Ala	Ala	Glu	Glu	Glu	Asp	Trp	Asp	Tyr	Ala	Pro	
				405					410					415		
Leu	Val	Leu	Ala	Pro	Asp	Asp	Arg	Ser	Tyr	Lys	Ser	Gln	Tyr	Leu	Asn	
		420					425						430			
Asn	Gly	Pro	Gln	Arg	Ile	Gly	Arg	Lys	Tyr	Lys	Lys	Val	Arg	Phe	Met	
		435					440					445				
Ala	Tyr	Thr	Asp	Glu	Thr	Phe	Lys	Thr	Arg	Glu	Ala	Ile	Gln	His	Glu	
	450					455					460					
Ser	Gly	Ile	Leu	Gly	Pro	Leu	Leu	Tyr	Gly	Glu	Val	Gly	Asp	Thr	Leu	
	465				470					475					480	
Leu	Ile	Ile	Phe	Lys	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Asn	Ile	Tyr	Pro	
				485					490					495		
His	Gly	Ile	Thr	Asp	Val	Arg	Pro	Leu	Tyr	Ser	Arg	Arg	Leu	Pro	Lys	
		500						505					510			
Gly	Val	Lys	His	Leu	Lys	Asp	Phe	Pro	Ile	Leu	Pro	Gly	Glu	Ile	Phe	
		515					520					525				
Lys	Tyr	Lys	Trp	Thr	Val	Thr	Val	Glu	Asp	Gly	Pro	Thr	Lys	Ser	Asp	
	530					535					540					
Pro	Arg	Cys	Leu	Thr	Arg	Tyr	Tyr	Ser	Ser	Phe	Val	Asn	Met	Glu	Arg	
	545				550					555					560	
Asp	Leu	Ala	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Ile	Cys	Tyr	Lys	Glu	
				565				570						575		
Ser	Val	Asp	Gln	Arg	Gly	Asn	Gln	Ile	Met	Ser	Asp	Lys	Arg	Asn	Val	
		580					585						590			
Ile	Leu	Phe	Ser	Val	Phe	Asp	Glu	Asn	Arg	Ser	Trp	Tyr	Leu	Thr	Glu	
		595					600					605				
Asn	Ile	Gln	Arg	Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp	
	610					615					620					
Pro	Glu	Phe	Gln	Ala	Ser	Asn	Ile	Met	His	Ser	Ile	Asn	Gly	Tyr	Val	
	625				630					635					640	
Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	Cys	Leu	His	Glu	Val	Ala	Tyr	Trp	
				645				650						655		
Tyr	Ile	Leu	Ser	Ile	Gly	Ala	Gln	Thr	Asp	Phe	Leu	Ser	Val	Phe	Phe	
		660						665					670			
Ser	Gly	Tyr	Thr	Phe	Lys	His	Lys	Met	Val	Tyr	Glu	Asp	Thr	Leu	Thr	
		675					680					685				
Leu	Phe	Pro	Phe	Ser	Gly	Glu	Thr	Val	Phe	Met	Ser	Met	Glu	Asn	Pro	
	690					695				700						
Gly	Leu	Trp	Ile	Leu	Gly	Cys	His	Asn	Ser	Asp	Phe	Arg	Asn	Arg	Gly	
	705				710					715					720	

-continued

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
740 745 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu
755 760 765

Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln
770 775 780

Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu
785 790 795 800

Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe
805 810 815

Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp
820 825 830

Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln
835 840 845

Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr
850 855 860

Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His
865 870 875 880

Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile
885 890 895

Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser
900 905 910

Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
915 920 925

Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val
930 935 940

Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp
945 950 955 960

Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu
965 970 975

Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His
980 985 990

Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe
995 1000 1005

Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn
1010 1015 1020

Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys
1025 1030 1035

Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr
1040 1045 1050

Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr
1055 1060 1065

Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His Phe
1070 1075 1080

Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met
1085 1090 1095

Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met
1100 1105 1110

-continued

Leu	Pro	Ser	Lys	Ala	Gly	Ile	Trp	Arg	Val	Glu	Cys	Leu	Ile	Gly
1115						1120					1125			
Glu	His	Leu	His	Ala	Gly	Met	Ser	Thr	Leu	Phe	Leu	Val	Tyr	Ser
1130						1135					1140			
Asn	Lys	Cys	Gln	Thr	Pro	Leu	Gly	Met	Ala	Ser	Gly	His	Ile	Arg
1145						1150					1155			
Asp	Phe	Gln	Ile	Thr	Ala	Ser	Gly	Gln	Tyr	Gly	Gln	Trp	Ala	Pro
1160						1165					1170			
Lys	Leu	Ala	Arg	Leu	His	Tyr	Ser	Gly	Ser	Ile	Asn	Ala	Trp	Ser
1175						1180					1185			
Thr	Lys	Glu	Pro	Phe	Ser	Trp	Ile	Lys	Val	Asp	Leu	Leu	Ala	Pro
1190						1195					1200			
Met	Ile	Ile	His	Gly	Ile	Lys	Thr	Gln	Gly	Ala	Arg	Gln	Lys	Phe
1205						1210					1215			
Ser	Ser	Leu	Tyr	Ile	Ser	Gln	Phe	Ile	Ile	Met	Tyr	Ser	Leu	Asp
1220						1225					1230			
Gly	Lys	Lys	Trp	Gln	Thr	Tyr	Arg	Gly	Asn	Ser	Thr	Gly	Thr	Leu
1235						1240					1245			
Met	Val	Phe	Phe	Gly	Asn	Val	Asp	Ser	Ser	Gly	Ile	Lys	His	Asn
1250						1255					1260			
Ile	Phe	Asn	Pro	Pro	Ile	Ile	Ala	Arg	Tyr	Ile	Arg	Leu	His	Pro
1265						1270					1275			
Thr	His	Tyr	Ser	Ile	Arg	Ser	Thr	Leu	Arg	Met	Glu	Leu	Met	Gly
1280						1285					1290			
Cys	Asp	Leu	Asn	Ser	Cys	Ser	Met	Pro	Leu	Gly	Met	Glu	Ser	Lys
1295						1300					1305			
Ala	Ile	Ser	Asp	Ala	Gln	Ile	Thr	Ala	Ser	Ser	Tyr	Phe	Thr	Asn
1310						1315					1320			
Met	Phe	Ala	Thr	Trp	Ser	Pro	Ser	Lys	Ala	Arg	Leu	His	Leu	Gln
1325						1330					1335			
Gly	Arg	Ser	Asn	Ala	Trp	Arg	Pro	Gln	Val	Asn	Asn	Pro	Lys	Glu
1340						1345					1350			
Trp	Leu	Gln	Val	Asp	Phe	Gln	Lys	Thr	Met	Lys	Val	Thr	Gly	Val
1355						1360					1365			
Thr	Thr	Gln	Gly	Val	Lys	Ser	Leu	Leu	Thr	Ser	Met	Tyr	Val	Lys
1370						1375					1380			
Glu	Phe	Leu	Ile	Ser	Ser	Ser	Gln	Asp	Gly	His	Gln	Trp	Thr	Leu
1385						1390					1395			
Phe	Phe	Gln	Asn	Gly	Lys	Val	Lys	Val	Phe	Gln	Gly	Asn	Gln	Asp
1400						1405					1410			
Ser	Phe	Thr	Pro	Val	Val	Asn	Ser	Leu	Asp	Pro	Pro	Leu	Leu	Thr
1415						1420					1425			
Arg	Tyr	Leu	Arg	Ile	His	Pro	Gln	Ser	Trp	Val	His	Gln	Ile	Ala
1430						1435					1440			
Leu	Arg	Met	Glu	Val	Leu	Gly	Cys	Glu	Ala	Gln	Asp	Leu	Tyr	
1445						1450					1455			

<210> SEQ ID NO 3

<211> LENGTH: 2220

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

-continued

polynucleotide

<400> SEQUENCE: 3

gccaccagga gatactacct gggggctgtg gagctttctt gggactacat gcagtctgac	60
ctgggggagc tgcctgtgga tgccaggctc ccaccagag tgcccaaadc ctteccattc	120
aacacctctg tggctotacaa gaagaccctc tttgtggagt tcaactgacca cctgttcaac	180
attgccaaac ccaggccacc ctggatggga ctctctggac ccaccattca ggctgaggtg	240
tatgacctg tggctatcac cctcaagaac atggcctccc accctgtgag cctgcatgct	300
gtgggggtca gctactggaa ggctctgag ggggctgagt atgatgacca gacctccag	360
agggagaagg aggatgacaa agtgttcctt gggggcagcc acacctatgt gtggcaggtc	420
ctcaaggaga atggcccat ggctctgac ccaactctgc tgacctactc ctacctttct	480
catgtggacc tggtaagga cctcaactct ggactgattg gggccctgct ggtgtgcagg	540
gagggtctcc tggccaaaga gaagaccag accctgcaca agttcattct cctgtttgct	600
gtctttgatg agggcaagag ctggcactct gaaaccaaga actccctgat gcaggacagg	660
gatgtgcct ctgccagggc ctggcccaag atgcacactg tgaatggcta tgtgaacagg	720
agcctgcctg gactcattgg ctgccacagg aaatctgtct actggcatgt gattggcatg	780
gggacaacc ctgaggtgca ctccattttc ctggagggcc acaccttctt ggtcaggaac	840
cacagacagg ccagcctgga gatcagcccc atcaccttcc tcaactgcca gacctgctg	900
atggacctg gacagttcct gctgttctgc cacatcagct cccaccagca tgatggcatg	960
gaggcctatg tcaaggtgga cagctgcctt gaggagccac agctcaggat gaagaacaat	1020
gaggaggctg aggactatga tgatgacctg actgactctg agatggatgt ggtccgcttt	1080
gatgatgaca acagcccac cttcattcag atcaggctctg tggccaagaa acaccccaag	1140
acctgggtgc actacattgc tgctgaggag gaggactggg actatgcccc actggtcctg	1200
gccccgatg acaggagcta caagagccag tacctcaaca atggcccaca gaggattgga	1260
cgcaagtaca agaaagttag gttcatggcc tacactgatg aaaccttcaa gaccaggag	1320
gccattcagc atgagtctgg catcctgggc ccaactcctg atggggaggt gggggacacc	1380
ctgctcatca tcttcaagaa ccaggcctcc aggcctaca acatctaccc acatggcatc	1440
actgatgtca gggccctgta cagccgcagg ctgcccagg ggggtgaaaca cctcaaggac	1500
ttccccattc tgcctgggga gatcttcaag tacaagtgga ctgtcactgt ggaggatgga	1560
ccaacccaat ctgacccag gtgcctcacc agatactact ccagctttgt gaacatggag	1620
agggacctgg cctctggcct gattggccca ctgctcatct gctacaagga gtctgtggac	1680
cagaggggaa accagatcat gtctgacaag aggaatgtga ttctgttctc tgtctttgat	1740
gagaacagga gctggtacct gactgagaac attcagcctt cctgcccaca cctgctggg	1800
gtgcagctgg aggacctga gttccaggcc agcaacatca tgcactccat caatggctat	1860
gtgtttgaca gcctccagct ttctgtctgc ctgcatgagg tggcctactg gtacattctt	1920
tctattgggg cccagactga ctctctttct gtcttcttct ctggctacac cttaaacac	1980
aagatggtgt atgaggacac cctgaccctc ttcccattct ctggggagac tgtgttcatg	2040
agcatggaga accctggcct gtgattctg ggatgccaca actctgactt ccgcaacagg	2100
ggcatgactg cctgctcaa agtctcctcc tgtgacaaga aactgggga ctactatgag	2160

-continued

gacagctatg aggacatctc tgcctacctg ctcagcaaga acaatgccat tgagcccagg 2220

<210> SEQ ID NO 4

<211> LENGTH: 2052

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 4

gagatcacca ggaccaccct ccagtctgac caggaggaga ttgactatga tgacaccatt 60
tctgtggaga tgaagaaaga ggactttgac atctatgacg aggacgagaa ccagagccca 120
aggagcttcc agaagaagac caggcactac ttcattgctg ctgtggagcg cctgtgggac 180
tatggcatga gctccagccc ccatgtctc aggaacaggg cccagtctgg ctctgtgcca 240
cagttcaaga aagtggctct ccaagagttc actgatggca gcttcaccca gccctgtac 300
agaggggagc tgaatgagca cctgggactc ctgggcccac acatcagggc tgaggtggag 360
gacaacatca tggtgacctt ccgcaaccag gctccaggc cctacagctt ctacagctcc 420
ctcatcagct atgaggagga ccagaggcag ggggctgagc cacgcaagaa ctttgtgaaa 480
cccaatgaaa ccaagaccta cttctggaaa gtccagcacc acatggcccc caccaaggat 540
gagtttgact gcaaggctg ggctacttc tctgatgtgg acctggagaa ggatgtgac 600
tctggcctga ttggcccact cctggtctgc cacaccaaca ccctgaacc tgccatgga 660
aggcaagtga ctgtgcagga gtttgccctc ttcttcaaca tctttgatga aaccaagagc 720
tggtaactca ctgagaacat ggagcgcaac tgcagggccc catgcaacat tcagatggag 780
gacccacct tcaaagagaa ctaccgctc catgccatca atggctacat catggacacc 840
ctgctgggc ttgtcatggc ccaggaccag aggatcaggt ggtacctgct ttctatgggc 900
tccaatgaga acattcactc catccactc tctgggcatg tcttcaactgt gcgcaagaag 960
gaggagtaca agatggccct gtacaacctc taccctgggg tctttgagac tgtggagatg 1020
ctgcctcca aagctggcat ctggaggggtg gagtgccca ttggggagca cctgcatgct 1080
ggcatgagca ccctgttctc ggtctacagc aacaagtgcc agacccccct gggaatggcc 1140
tctggccaca tcagggactt ccagatcact gcctctggcc agtatggcca gtgggcccc 1200
aagctggcca ggctccacta cttctggatc atcaatgctt ggagcaccaa ggagccattc 1260
agctggatca aagtgaacct gctggcccc atgatcatcc atggcatcaa gaccagggg 1320
gccaggcaga agttctccag cctgtacatc agccagttca tcatcatgta cagcctggat 1380
ggcaagaaat ggcagaccta cagaggcaac tccactggaa cactcatggt cttctttggc 1440
aatgtggaca gctctggcat caagcacaac atcttcaacc cccaatcat cgccagatac 1500
atcaggctgc accccacca ctacagcatc cgcagcacc tcaggatgga gctgatgggc 1560
tgtgacctga actcctgag catgccctg gcatggaga gcaaggccat ttctgatgcc 1620
cagatcactg cctccagcta cttcaccac atgtttgcca cctggagccc aagcaaggcc 1680
aggctgcacc tccaggaag gagcaatgcc tggaggcccc aggtcaacaa cccaaaggag 1740
tggctgcagg tggacttoca gaagaccatg aaggtcactg gggtgaccac ccagggggtc 1800
aagagcctgc tcaccagcat gtatgtgaag gagttcctga tcagctccag ccaggatggc 1860
caccagtgga ccctctctt ccagaatggc aaggtcaagg tgttccaggg caaccaggac 1920

-continued

```

agcttcaccc ctgtggtgaa cagcctggac cccccctcc tgaccagata cctgaggatt 1980
cacccccaga gctgggtcca ccagattgcc ctgaggatgg aggtcctggg atgtgaggcc 2040
caggacctgt ac 2052

```

```

<210> SEQ ID NO 5
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

```

```

<400> SEQUENCE: 5

```

```

agcttctctc agaatccacc tgcctgaa agacaccaga ga 42

```

```

<210> SEQ ID NO 6
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

```

```

<400> SEQUENCE: 6

```

```

agcttcagcc agaatccacc tgcctgaaa cgccaccaga gg 42

```

```

<210> SEQ ID NO 7
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

```

```

<400> SEQUENCE: 7

```

```

agcttcagcc agaaccccc cgtgctgaa aggcaccaga gg 42

```

```

<210> SEQ ID NO 8
<211> LENGTH: 7827
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 8

```

```

tcgcgcttt cggatgatgac ggtgaaaacc tctgacacat gcagctccc gagacggtca 60
cagcttgct gtaagcggat gccgggagca gacaagccc tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accatatggt gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagggccc 240
attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgetat 300
tacgccagct ggcgaaaggg gtagtgctg caaggcgatt aagttgggta acgccagggt 360
tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cctcgagatt taaatgacgt 420
tggccactcc ctctctgcgc gctcctcgc tcaactgaggc cgggcgacca aaggtcgccc 480
gacgcccggg ctttgcccgg gcggcctcag tgagcgagcg agcgcgcaga gagggagtgg 540
ccaactccat cactaggggt tctgagttt aaactctgct gacgattcga gcttgggctg 600

```

-continued

caggtcgagg	gcaactgggag	gatgttgagt	aagatggaaa	actactgatg	acccttgag	660
agacagagta	ttaggacatg	ttgaaacagg	ggccgggca	tcagcagga	gctctagagg	720
atccccgtct	gtctgcacat	ttcgtagagc	gagtgttccg	atactctaat	ctccctaggc	780
aaggttcata	tttgtgtagg	ttacttattc	tccttttggt	gactaagtca	ataatcagaa	840
tcagcaggtt	tggagtcagc	ttggcagga	tcagcagcct	gggttgaag	gagggggat	900
aaaagcccct	tcaccaggag	aagccgtcac	acagactagg	cgcgccaccg	ccaccatgca	960
gattgagctg	agcacctgct	tcttctgtg	cctgctgagg	ttctgcttct	ctgccaccag	1020
gagatactac	ctgggggctg	tggagcttct	ttgggactac	atgcagctctg	acctggggga	1080
gctgcctgtg	gatgccagg	tcccacccag	agtgcacaaa	tccttcccat	tcaacacctc	1140
tgtggtctac	aagaagacc	tctttgtgga	gttcaactgac	cacctgttca	acattgccaa	1200
accaggcca	ccctggatgg	gactcctggg	accacaccatt	caggetgagg	tgtatgacac	1260
tgtggtcatc	accctcaaga	acatggcctc	ccaccctgtg	agcctgcatg	ctgtgggggt	1320
cagctactgg	aaggcctctg	agggggctga	gtatgatgac	cagacctccc	agagggagaa	1380
ggaggatgac	aaagtgttcc	ctgggggag	ccacacctat	gtgtggcagg	tcctcaagga	1440
gaatggcccc	atggcctctg	accactctg	cctgacctac	tcctaccttt	ctcatgtgga	1500
cctggtcaag	gacctcaact	ctggactgat	tggggccctg	ctggtgtgca	gggagggctc	1560
cctggccaaa	gagaagacc	agaccctgca	caagttcatt	ctcctgtttg	ctgtctttga	1620
tgagggaag	agctggcact	ctgaaaccaa	gaactccctg	atgcaggaca	gggatgctgc	1680
ctctgccagg	gcctggccca	agatgcacac	tgtgaatggc	tatgtgaaca	ggagcctgcc	1740
tggactcatt	ggctgccaca	ggaaatctgt	ctactggcat	gtgattggca	tggggacaac	1800
ccctgagggt	cactccattt	tctggagggg	ccacaccttc	ctggtcagga	accacagaca	1860
ggccagcctg	gagatcagcc	ccatcacctt	cctcactgcc	cagaccctgc	tgatggacct	1920
cggacagttc	ctgctgttct	gccacatcag	ctcccaccag	catgatggca	tggaggccta	1980
tgtcaagggt	gacagctgcc	ctgaggagcc	acagctcagg	atgaagaaca	atgaggaggc	2040
tgaggactat	gatgatgacc	tgactgactc	tgagatggat	gtggtccgct	ttgatgatga	2100
caacagccca	tccttcattc	agatcaggtc	tgtggccaag	aaacacccca	agacctgggt	2160
gcactacatt	gctgctgagg	aggaggactg	ggactatgcc	ccactggtcc	tggccccctga	2220
tgacaggagc	tacaagagcc	agtacctcaa	caatggccca	cagaggattg	gacgcaagta	2280
caagaaagtc	aggttcatgg	cctacactga	tgaaaccttc	aagaccaggg	aggccattca	2340
gcatgagtct	ggcatcctgg	gcccactcct	gtatggggag	gtgggggaca	ccctgctcat	2400
catcttcaag	aaccaggcct	ccaggcccta	caacatctac	ccacatggca	tactgatgt	2460
caggccccctg	tacagccgca	ggctgccaaa	gggggtgaaa	cacctcaagg	acttccccat	2520
tctgcctggg	gagatcttca	agtacaagtg	gactgtcact	gtggaggatg	gaccaaccaa	2580
atctgacccc	aggtgcctca	ccagatacta	ctccagcttt	gtgaacatgg	agagggacct	2640
ggcctctggc	ctgattggcc	cactgctcat	ctgctacaag	gagtctgtgg	accagagggg	2700
aaaccagatc	atgtctgaca	agaggaatgt	gattctgttc	tctgtctttg	atgagaacag	2760
gagctggtac	ctgactgaga	acattcagcg	cttctgccc	aacctgctg	gggtgcagct	2820
ggaggacct	gagttccagg	ccagcaacat	catgcactcc	atcaatggct	atgtgtttga	2880

-continued

cagcctccag ctttctgtct gcttgcata ggtggcctac tggtagattc tttctattgg 2940
ggcccagact gacttctttt ctgtcttctt ctctggctac acctcaaac acaagatggt 3000
gtatgaggac acctgaccc tcttcccatt ctctggggag actgtgttca tgagcatgga 3060
gaaccctggc ctgtggatc tgggatgcca caactctgac ttcgcaaca ggggcatgac 3120
tgccctgctc aaagtctct cctgtgacaa gaacctggg gactactatg aggacagcta 3180
tgaggacatc tctgcctacc tgctcagcaa gaacaatgcc attgagccca ggagcttcag 3240
ccagaatcca cctgtctga aacgccacca gagggagatc accaggacca cctccagtc 3300
tgaccaggag gagattgact atgatgacac ctttctgtg gagatgaaga aagaggactt 3360
tgacatctat gacgaggacg agaaccagag cccaaggagc ttccagaaga agaccaggca 3420
ctacttcatt gctgctgtgg agcgcctgtg ggactatggc atgagctcca gccccatgt 3480
cctcaggaac agggcccagt ctggctctgt gccacagttc aagaaagtgg tcttccaaga 3540
gttactgat ggcagcttca cccagcccct gtacagaggg gagctgaatg agcacctggg 3600
actcctgggc ccatacatca gggctgaggt ggaggacaac atcatggtga ccttccgcaa 3660
ccaggcctcc agggcctaca gcttctacag ctcccctatc agctatgagg aggaccagag 3720
gcagggggct gagccacgca agaactttgt gaaacccaat gaaaccaaga cctacttctg 3780
gaaagtccag caccacatgg cccccacaa ggatgagttt gactgcaagg cctgggctca 3840
cttctctgat gtggacctgg agaaggatgt gcaactctggc ctgattggcc cactcctggt 3900
ctgccacacc aacacctga acctgccca tgggaaggcaa gtgactgtgc aggagtgtgc 3960
cctctcttc accatctttg atgaaacaa gagctggtac ttcactgaga acatggagcg 4020
caactgcagg gccccatgca acattcagat ggaggacccc acctcaaag agaactaccg 4080
cttccatgcc atcaatggct acatcatgga caccctgect gggcttctca tggcccagga 4140
ccagaggatc aggtggtacc tgtttctat gggctccaat gagaacattc actccatcca 4200
cttctctggg catgtcttca ctgtgcgcaa gaaggaggag tacaagatgg ccctgtacaa 4260
cctctacctc ggggtctttg agactgtgga gatgctgccc tccaaagctg gcatctggag 4320
gggtgagtg ctcattgggg agcacctgca tgctggcatg agcacctgt tctgtgtcta 4380
cagcaacaag tgccagacc ccttgggaat gccctctggc cacatcaggg acttccagat 4440
cactgcctct ggccagtatg gccagtggc cccaagctg gccaggctcc actactctgg 4500
atccatcaat gcctggagca ccaaggagcc attcagctgg atcaaagtgg acctgctggc 4560
ccccatgatc atccatggca tcaagacca gggggccagg cagaagtctt ccagcctgta 4620
catcagccag ttcacatca tgtacagcct ggatggcaag aatggcaga cctacagagg 4680
caactccact ggaacactca tggctcttct tggcaatgtg gacagctctg gcatcaagca 4740
caacatctc aacccccca tcatgcccag atacatcagg ctgcaccca cccactacag 4800
catccgagc acctcagga tggagctgat gggctgtgac ctgaaactcct gcagcatgcc 4860
cctgggcatg gagagcaagg cctttctga tgcccagatc actgcctcca gctacttcac 4920
caacatgtt gccacctgga gccaagcaa gggcaggctg cacctccagg gaaggagcaa 4980
tgctggagg ccccagtgca acaacccaaa ggagtggctg caggtggact tccagaagac 5040
catgaagtc actgggtgta cccccaggg ggtcaagagc ctgctacca gcatgtatgt 5100
gaaggagt cctgatcagct ccagccagga tggccaccag tggacctct tcttccagaa 5160

-continued

tggcaaggtc aagggtgtcc agggcaacca ggacagcttc acccctgtgg tgaacagcct 5220
ggaccccccc ctctgaacca gatacctgag gattcaccoc cagagctggg tccaccagat 5280
tgccctgagg atggagggtcc tgggatgtga ggcccaggac ctgtactgat gacgagcggc 5340
cgctcttagt agcagtatcg ataataaaaag atctttatctt tcattagatc tgtgtgttgg 5400
ttttttgtgt gttaattaag ctgcggaagg aaccctagt gatggagtgg gccactccct 5460
ctctgcgcgc tcgctcgtc actgaggccg ggccgacaaa ggctgcgccga cgcccgggct 5520
ttgcccgggc ggctcagtg agcagcggag cgccgagaga gggagtggcc aagacgattt 5580
aatgacaag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc 5640
tcacaattcc acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaata 5700
gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc 5760
tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcccgg gagagggcgg ttgcgtattg 5820
ggcgtctctc cgcttcctcg ctactgact cgctcgcctc ggctcgttcgg ctgcggcgag 5880
cggatcagc tcaactcaag gcggtaatac ggttatccac agaatacagg gataacgcag 5940
gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc 6000
tggcgttttt ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc 6060
agaggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct ggaagctccc 6120
tcgtcgcctc tcctgttccg accctgccc ttaccggata cctgtcccgc tttctccctt 6180
cgggaagcgt ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg 6240
ttcgtcccaa gctgggctgt gtcacgaac cccccgttca gcccgaccgc tgcgcttat 6300
ccgtaacta tcgtcttgag tccaaccgg taagacacga cttatcgcca ctggcagcag 6360
ccactggtaa caggattagc agagcgaggat atgtaggcgg tgctacagag ttcttgaagt 6420
gggtgcctaa ctacggctac actagaagaa cagtatttgg tatctgcgct ctgctgaagc 6480
cagttacctt cggaaaaaga gttggtagct cttgatccgg caaacaacc accgctggta 6540
gcggtggttt ttttgttgc aagcagcaga ttacgcgcag aaaaaagga tctcaagaag 6600
atcctttgat cttttctacg gggctgacg ctccagtgaa cgaaaactca cgtaagggga 6660
ttttggatc gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa 6720
gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa 6780
tcagtgaggc acctatctca gcgatctgtc tatttcttc atccatagtt gctgactcc 6840
ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga 6900
taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa 6960
gggcccagcg cagaagtggt cctgcaactt tatccgcctc catccagctc attaattggt 7020
gccgggaagc tagagtaagt agttcccgag ttaatagttt gcgcaacggt gttgccattg 7080
ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc 7140
aacgatcaag gcgagttaca tgatccccca tgttgtgcaa aaaagcgggt agctccttcg 7200
gtcctccgat cgttgtcaga agtaagttgg ccgagctgtt atcactcatg gttatggcag 7260
cactgcataa ttctcttact gtcatgcat ccgtaagatg cttttctgtg actggtgagt 7320
actcaaccaa gtcattctga gaatagtgtg tgcggcgacc gagttgctct tgcccggcgt 7380
caatacggga taataccgag ccacatagca gaactttaa agtgctcatc attgaaaaac 7440

-continued

gttcttcggg gcgaaaactc tcaaggatct taccgctggt gagatccagt tegatgtaac	7500
ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag	7560
caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa	7620
tactcatact ctctcttttt caatattatt gaagcattta tcagggttat tgtctcatga	7680
gcgatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc	7740
ccgaaaagt gccacctgac gtctaagaaa ccattattat catgacatta acctataaaa	7800
ataggcgtat cagcaggccc tttcgtc	7827

<210> SEQ ID NO 9

<211> LENGTH: 4332

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 9

atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc	60
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg	120
ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaactcctt cccattcaac	180
acctctgtgg tctacaagaa gacctcttt gtggagtcca ctgaccacct gttcaacatt	240
gccccaaacca ggccacctg gatgggactc ctgggaccca ccattcagge tgagggtgat	300
gacactgtgg tcatcacct caagaacatg gcctcccacc ctgtgacct gcctgctgtg	360
ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg	420
gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggctctc	480
aaggagaatg gccccatggc ctctgaccca ctctgctga cctactccta ctttctcat	540
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcaggag	600
ggctccctgg ccaaaagaaa gaccagacc ctgcacaagt tcattctcct gtttctgtc	660
tttgatgagg gcaagagctg gacctctgaa accaagaact ccctgatgca ggacagggat	720
gtgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc	780
ctgctctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg	840
acaaccctg aggtgcactc cattttctg gagggccaca cttctctggt caggaaccac	900
agacaggcca gcctggagat cagccccatc accttctca ctgcccagac cctgctgatg	960
gacctcggac agttctctgct gttctgccac atcagctccc accagcatga tggcatggag	1020
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat	1140
gatgacaaca gccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc	1200
tgggtgcact acattgtctg tgaggaggag gactgggact atgccccact ggtcctggcc	1260
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gccacagag gattggacgc	1320
aagtacaaga aagtcagggt catggcctac actgatgaaa cttcaagac caggagggcc	1380
attcagcatg agtctggcat cctgggocca ctctgtatg gggaggtggg ggacaccctg	1440
ctcatcatct tcaagaacca ggccctcagg ccctacaaca tctaccaca tggcatcact	1500

-continued

gatgtcaggc	ccctgtacag	ccgcaggetg	ccaaaggggg	tgaaacacct	caaggacttc	1560
cccattctgc	ctggggagat	cttcaagtac	aagtggactg	tcaactgtga	ggatggacca	1620
accaaactcg	accccaggtg	cctcaccaga	tactactcca	gctttgtgaa	catggagagg	1680
gacctggcct	ctggcctgat	tggcccactg	ctcatctgct	acaaggagtc	tgtggaccag	1740
aggggaaacc	agatcatgtc	tgacaagagg	aatgtgattc	tgttctctgt	ctttgatgag	1800
aacaggagct	ggtagctgac	tgagaacatt	cagcgcttcc	tgcccaaccc	tgctggggtg	1860
cagctggagg	accctgagtt	ccaggccagc	aacatcatgc	actccatcaa	tggetatgtg	1920
tttgacagcc	tccagctttc	tgtctgcctg	catgaggtgg	cctactggta	cattctttct	1980
attggggccc	agactgactt	cctttctgtc	ttcttctctg	gctacacctt	caaacacaag	2040
atggtgatg	aggacacct	gacctcttc	ccattctctg	gggagactgt	gttcatgagc	2100
atggagaacc	ctggcctgtg	gattctggga	tgccacaact	ctgacttccg	caacaggggc	2160
atgactgccc	tgtcaaaagt	ctcctcctgt	gacaagaaca	ctggggacta	ctatgaggac	2220
agctatgagg	acatctctgc	ctacctgtc	agcaagaaca	atgccattga	gcccagggag	2280
atcaccagga	ccaccctcca	gtctgaccag	gaggagattg	actatgatga	caccatttct	2340
gtggagatga	agaagagga	ctttgacatc	tatgacgagg	acgagaacca	gagcccaagg	2400
agcttcaga	agaagaccag	gcactacttc	attgctgctg	tggagcgccct	gtgggactat	2460
ggcatgagct	ccagccccc	tgtcctcagg	aacagggccc	agctctggctc	tgtgccacag	2520
ttcaagaaag	tggtcttcca	agagttcact	gatggcagct	tcaccagacc	cctgtacaga	2580
ggggagctga	atgagcacct	gggactcctg	ggcccataca	tcagggctga	ggtggaggac	2640
aacatcatgg	tgaccttccg	caaccaggcc	tccaggccct	acagcttcta	cagctccctc	2700
atcagctatg	aggaggacca	gaggcagggg	gctgagccac	gcaagaactt	tgtgaaaccc	2760
aatgaaacca	agacctactt	ctggaaagtc	cagcaccaca	tggccccac	caaggatgag	2820
tttgactgca	aggcctgggc	ctacttctct	gatgtggacc	tggagaagga	tgtgcactct	2880
ggcctgattg	gcccactcct	ggtctgccac	accaaacc	tgaaccctgc	ccatggaagg	2940
caagtgactg	tgcaggagtt	tgcctcttc	ttcaccatct	ttgatgaaac	caagagctgg	3000
tacttcactg	agaacatgga	gcgcaactgc	agggcccat	gcaacattca	gatggaggac	3060
cccaccttca	aagagaacta	ccgcttccat	gccatcaatg	gctacatcat	ggacacctg	3120
cctgggcttg	tcatggccc	ggaccagagg	atcaggtggt	acctgcttcc	tatgggtccc	3180
aatgagaaca	ttcactccat	ccacttctct	gggcatgtct	tcaactgtgcg	caagaaggag	3240
gagtacaaga	tggccctgta	caacctctac	cctggggtct	ttgagactgt	ggagatgctg	3300
ccctccaag	ctggcatctg	gaggggtggag	tgctcattg	gggagcacct	gcatgctggc	3360
atgagcacc	tgttctgggt	ctacagcaac	aagtgccaga	ccccctggg	aatggcctct	3420
ggccacatca	gggacttcca	gatcaactgc	tctggccagt	atggccagtg	ggcccccaag	3480
ctggccaggc	tccactactc	tggatccatc	aatgcctgga	gcaccaagga	gccattcagc	3540
tggatcaaag	tggacctgct	ggcccccatg	atcatccatg	gcatcaagac	ccagggggcc	3600
aggcagaagt	tctccagcct	gtacatcagc	cagttcatca	tcatgtacag	cctggatggc	3660
aagaaatggc	agacctacag	aggcaactcc	actggaacac	tcatggtctt	ctttgcaat	3720
gtggacagct	ctggcatcaa	gcacaacatc	ttcaacccc	caatcatcgc	cagatacatc	3780

-continued

```

aggctgcacc ccacccta cagcatccgc agcaccctca ggatggagct gatgggctgt 3840
gacctgaact cctgcagcat gccctgggc atggagagca aggccatttc tgatgccag 3900
atcactgcct ccagctactt caccaacatg ttgcccact ggagcccaag caaggccag 3960
ctgcacctcc agggaaggag caatgcctgg aggccccagg tcaacaaccc aaaggagtgg 4020
ctgcaggtgg acttocagaa gaccatgaag gtcactgggg tgaccacca gggggtaag 4080
agcctgctca ccagcatgta tgtgaaggag ttcctgatca gctccagcca ggatggccac 4140
cagtggacc tcttcttoca gaatggcaag gtcaaggtgt tccagggcaa ccaggacagc 4200
ttcaccctg tggtaacag cctggacccc cccctcctga ccagatacct gaggattcac 4260
ccccagagct gggccaacca gattgcctg aggatggagg tctgggatg tgaggcccag 4320
gacctgtact ga 4332

```

<210> SEQ ID NO 10

<211> LENGTH: 1443

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 10

```

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
 1           5           10          15
Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
 20          25          30
Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
 35          40          45
Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
 50          55          60
Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
 65          70          75          80
Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
 85          90          95
Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
100         105         110
His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
115         120         125
Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
130         135         140
Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145         150         155         160
Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
165         170         175
Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
180         185         190
Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
195         200         205
Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
210         215         220
Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225         230         235         240

```

-continued

Ala	Ala	Ser	Ala	Arg	Ala	Trp	Pro	Lys	Met	His	Thr	Val	Asn	Gly	Tyr
				245						250				255	
Val	Asn	Arg	Ser	Leu	Pro	Gly	Leu	Ile	Gly	Cys	His	Arg	Lys	Ser	Val
			260					265					270		
Tyr	Trp	His	Val	Ile	Gly	Met	Gly	Thr	Thr	Pro	Glu	Val	His	Ser	Ile
		275					280					285			
Phe	Leu	Glu	Gly	His	Thr	Phe	Leu	Val	Arg	Asn	His	Arg	Gln	Ala	Ser
	290					295					300				
Leu	Glu	Ile	Ser	Pro	Ile	Thr	Phe	Leu	Thr	Ala	Gln	Thr	Leu	Leu	Met
305					310					315					320
Asp	Leu	Gly	Gln	Phe	Leu	Leu	Phe	Cys	His	Ile	Ser	Ser	His	Gln	His
				325					330					335	
Asp	Gly	Met	Glu	Ala	Tyr	Val	Lys	Val	Asp	Ser	Cys	Pro	Glu	Glu	Pro
			340					345					350		
Gln	Leu	Arg	Met	Lys	Asn	Asn	Glu	Glu	Ala	Glu	Asp	Tyr	Asp	Asp	Asp
		355					360					365			
Leu	Thr	Asp	Ser	Glu	Met	Asp	Val	Val	Arg	Phe	Asp	Asp	Asp	Asn	Ser
	370					375					380				
Pro	Ser	Phe	Ile	Gln	Ile	Arg	Ser	Val	Ala	Lys	Lys	His	Pro	Lys	Thr
385					390					395					400
Trp	Val	His	Tyr	Ile	Ala	Ala	Glu	Glu	Glu	Asp	Trp	Asp	Tyr	Ala	Pro
				405					410					415	
Leu	Val	Leu	Ala	Pro	Asp	Asp	Arg	Ser	Tyr	Lys	Ser	Gln	Tyr	Leu	Asn
			420					425					430		
Asn	Gly	Pro	Gln	Arg	Ile	Gly	Arg	Lys	Tyr	Lys	Lys	Val	Arg	Phe	Met
		435					440					445			
Ala	Tyr	Thr	Asp	Glu	Thr	Phe	Lys	Thr	Arg	Glu	Ala	Ile	Gln	His	Glu
	450					455					460				
Ser	Gly	Ile	Leu	Gly	Pro	Leu	Leu	Tyr	Gly	Glu	Val	Gly	Asp	Thr	Leu
465					470					475					480
Leu	Ile	Ile	Phe	Lys	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Asn	Ile	Tyr	Pro
				485					490					495	
His	Gly	Ile	Thr	Asp	Val	Arg	Pro	Leu	Tyr	Ser	Arg	Arg	Leu	Pro	Lys
			500					505					510		
Gly	Val	Lys	His	Leu	Lys	Asp	Phe	Pro	Ile	Leu	Pro	Gly	Glu	Ile	Phe
		515					520					525			
Lys	Tyr	Lys	Trp	Thr	Val	Thr	Val	Glu	Asp	Gly	Pro	Thr	Lys	Ser	Asp
	530					535					540				
Pro	Arg	Cys	Leu	Thr	Arg	Tyr	Tyr	Ser	Ser	Phe	Val	Asn	Met	Glu	Arg
545					550					555					560
Asp	Leu	Ala	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Ile	Cys	Tyr	Lys	Glu
				565					570					575	
Ser	Val	Asp	Gln	Arg	Gly	Asn	Gln	Ile	Met	Ser	Asp	Lys	Arg	Asn	Val
			580					585					590		
Ile	Leu	Phe	Ser	Val	Phe	Asp	Glu	Asn	Arg	Ser	Trp	Tyr	Leu	Thr	Glu
		595					600					605			
Asn	Ile	Gln	Arg	Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp
	610					615					620				
Pro	Glu	Phe	Gln	Ala	Ser	Asn	Ile	Met	His	Ser	Ile	Asn	Gly	Tyr	Val
625					630					635					640
Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	Cys	Leu	His	Glu	Val	Ala	Tyr	Trp

-continued

Tyr	Leu	Leu	Ser	Met	Gly	Ser	Asn	Glu	Asn	Ile	His	Ser	Ile	His
1055						1060					1065			
Phe	Ser	Gly	His	Val	Phe	Thr	Val	Arg	Lys	Lys	Glu	Glu	Tyr	Lys
1070						1075					1080			
Met	Ala	Leu	Tyr	Asn	Leu	Tyr	Pro	Gly	Val	Phe	Glu	Thr	Val	Glu
1085						1090					1095			
Met	Leu	Pro	Ser	Lys	Ala	Gly	Ile	Trp	Arg	Val	Glu	Cys	Leu	Ile
1100						1105					1110			
Gly	Glu	His	Leu	His	Ala	Gly	Met	Ser	Thr	Leu	Phe	Leu	Val	Tyr
1115						1120					1125			
Ser	Asn	Lys	Cys	Gln	Thr	Pro	Leu	Gly	Met	Ala	Ser	Gly	His	Ile
1130						1135					1140			
Arg	Asp	Phe	Gln	Ile	Thr	Ala	Ser	Gly	Gln	Tyr	Gly	Gln	Trp	Ala
1145						1150					1155			
Pro	Lys	Leu	Ala	Arg	Leu	His	Tyr	Ser	Gly	Ser	Ile	Asn	Ala	Trp
1160						1165					1170			
Ser	Thr	Lys	Glu	Pro	Phe	Ser	Trp	Ile	Lys	Val	Asp	Leu	Leu	Ala
1175						1180					1185			
Pro	Met	Ile	Ile	His	Gly	Ile	Lys	Thr	Gln	Gly	Ala	Arg	Gln	Lys
1190						1195					1200			
Phe	Ser	Ser	Leu	Tyr	Ile	Ser	Gln	Phe	Ile	Ile	Met	Tyr	Ser	Leu
1205						1210					1215			
Asp	Gly	Lys	Lys	Trp	Gln	Thr	Tyr	Arg	Gly	Asn	Ser	Thr	Gly	Thr
1220						1225					1230			
Leu	Met	Val	Phe	Phe	Gly	Asn	Val	Asp	Ser	Ser	Gly	Ile	Lys	His
1235						1240					1245			
Asn	Ile	Phe	Asn	Pro	Pro	Ile	Ile	Ala	Arg	Tyr	Ile	Arg	Leu	His
1250						1255					1260			
Pro	Thr	His	Tyr	Ser	Ile	Arg	Ser	Thr	Leu	Arg	Met	Glu	Leu	Met
1265						1270					1275			
Gly	Cys	Asp	Leu	Asn	Ser	Cys	Ser	Met	Pro	Leu	Gly	Met	Glu	Ser
1280						1285					1290			
Lys	Ala	Ile	Ser	Asp	Ala	Gln	Ile	Thr	Ala	Ser	Ser	Tyr	Phe	Thr
1295						1300					1305			
Asn	Met	Phe	Ala	Thr	Trp	Ser	Pro	Ser	Lys	Ala	Arg	Leu	His	Leu
1310						1315					1320			
Gln	Gly	Arg	Ser	Asn	Ala	Trp	Arg	Pro	Gln	Val	Asn	Asn	Pro	Lys
1325						1330					1335			
Glu	Trp	Leu	Gln	Val	Asp	Phe	Gln	Lys	Thr	Met	Lys	Val	Thr	Gly
1340						1345					1350			
Val	Thr	Thr	Gln	Gly	Val	Lys	Ser	Leu	Leu	Thr	Ser	Met	Tyr	Val
1355						1360					1365			
Lys	Glu	Phe	Leu	Ile	Ser	Ser	Ser	Gln	Asp	Gly	His	Gln	Trp	Thr
1370						1375					1380			
Leu	Phe	Phe	Gln	Asn	Gly	Lys	Val	Lys	Val	Phe	Gln	Gly	Asn	Gln
1385						1390					1395			
Asp	Ser	Phe	Thr	Pro	Val	Val	Asn	Ser	Leu	Asp	Pro	Pro	Leu	Leu
1400						1405					1410			
Thr	Arg	Tyr	Leu	Arg	Ile	His	Pro	Gln	Ser	Trp	Val	His	Gln	Ile
1415						1420					1425			

-continued

Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1430 1435 1440

<210> SEQ ID NO 11
 <211> LENGTH: 4368
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 11

atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60
 accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120
 ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaactcctt cccattcaac 180
 acctctgtgg tctacaagaa gacctctttt gtggagtcca ctgaccacct gttcaacatt 240
 gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgagggtgat 300
 gacactgtgg tcatcacctc caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360
 ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420
 gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggctctc 480
 aaggagaatg gccccatggc ctctgaccca ctctgctga cctactccta cttttctcat 540
 gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcaggagg 600
 ggctccctgg ccaaaagaaa gaccagacc ctgcacaagt tcattctcct gtttctgtc 660
 tttgatgagg gcaagagctg gactctgaa accaagaact ccctgatgca ggacagggat 720
 gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780
 ctgctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840
 acaaccctg aggtgcactc cattttctct gagggccaca ctttctggt caggaaccac 900
 agacaggcca gcctggagat cagccccatc accttctca ctgcccagac cctgctgatg 960
 gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag 1020
 gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080
 gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140
 gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200
 tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260
 cctgatgaca ggagctacaa gagccagtac ctcaacaatg gccacagag gattggacgc 1320
 aagtacaaga aagtcagggt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380
 attcagcatg agtctggcat cctgggccc a ctctgtatg gggaggtggg ggacacctg 1440
 ctcatcatct tcaagaacca ggctccagg cctacaaca tctaccaca tggcatcact 1500
 gatgtcaggc ccctgtacag ccgaggctg ccaaaggggg tgaaacacct caaggacttc 1560
 cccattctgc ctggggagat cttcaagtac aagtggaactg tcaactgtga ggatggacca 1620
 accaaatctg accccagggt cctcaccaga tactactcca gctttgtgaa catggagagg 1680
 gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740
 aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800
 aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaccc tgctgggggtg 1860

-continued

cagctggagg	accctgagtt	ccaggccagc	aacatcatgc	actccatcaa	tgctatgtg	1920
tttgacagcc	tccagctttc	tgtctgctg	catgagggtg	cctactggta	cattctttct	1980
attggggccc	agactgactt	cctttctgtc	ttcttctctg	gctacacctt	caaacacaag	2040
atggtgtatg	aggacacct	gacctcttc	ccattctctg	gggagactgt	gttcatgagc	2100
atggagaacc	ctggcctgtg	gattctggga	tgccacaact	ctgacttccg	caacaggggc	2160
atgactgccc	tgctcaaagt	ctctctctgt	gacaagaaca	ctggggacta	ctatgaggac	2220
agctatgagg	acatctctgc	ctacctgtc	agcaagaaca	atgccattga	gcccaggagc	2280
ttcagccaga	attccagaca	ccccagcacc	agggagatca	ccaggaccac	cctccagtct	2340
gaccaggagg	agattgacta	tgatgacacc	atctctgtgg	agatgaagaa	agaggacttt	2400
gacatctatg	acgaggacga	gaaccagagc	ccaaggagct	tccagaagaa	gaccaggcac	2460
tacttcattg	ctgctgtgga	gctctgtgg	gactatggca	tgagctccag	ccccatgtc	2520
ctcaggaaca	gggcccagtc	tgctctgtg	ccacagtcca	agaaagtggg	cttccaagag	2580
ttcactgatg	gcagcttcac	ccagccctg	tacagagggg	agctgaatga	gcacctggga	2640
ctcctgggcc	catacatcag	ggctgagggtg	gaggacaaca	tcatggtgac	cttccgcaac	2700
caggcctcca	ggcctacag	cttctacagc	tccctcatca	gctatgagga	ggaccagagg	2760
cagggggctg	agccacgcaa	gaactttgtg	aaaccaatg	aaaccaagac	ctacttctgg	2820
aaagtccagc	accacatggc	ccccaccaag	gatgagtttg	actgcaaggc	ctgggcctac	2880
ttctctgatg	tggacctgga	gaaggatgtg	cactctggcc	tgattggccc	actcctggtc	2940
tgccacacca	acaccctgaa	ccctgcccac	ggaaggcaag	tgactgtgca	ggagtttgcc	3000
ctcttcttca	ccatctttga	tgaaccaag	agctggact	tactgagaa	catggagcgc	3060
aactgcaggg	ccccatgcaa	cattcagatg	gaggacccca	ccttcaaaga	gaactaccgc	3120
ttccatgcca	tcaatggcta	catcatggac	accctgcctg	ggcttgtcat	ggcccaggac	3180
cagaggatca	ggtggtaoct	gctttctatg	ggctccaatg	agaacattca	ctccatccac	3240
ttctctgggc	atgtcttca	tgtgcgcaag	aaggaggagt	acaagatggc	cctgtacaac	3300
ctctaccctg	gggtcttga	gactgtggag	atgctgcct	ccaagctgg	catctggagg	3360
gtggagtgcc	tcattgggga	gcacctgcat	gctggcatga	gcacctgtt	cctggtctac	3420
agcaacaagt	gccagacccc	cctgggaatg	gcctctggcc	acatcagggg	cttccagatc	3480
actgcctctg	gccagtatgg	ccagtgggcc	cccaagctgg	ccaggctcca	ctactctgga	3540
tccatcaatg	cctggagcac	caaggagcca	ttcagctgga	tcaaagtgga	cctgctggcc	3600
cccatgatca	tccatggcat	caagacccag	ggggccaggc	agaagttctc	cagcctgtac	3660
atcagccagt	tcatcatcat	gtacagcctg	gatggcaaga	aatggcagac	ctacagaggc	3720
aactccactg	gaacactcat	ggtcttcttt	ggcaatgtgg	acagctctgg	catcaagcac	3780
aacatcttca	accccccaat	catgcgcaga	tacatcaggc	tgacccccac	ccactacagc	3840
atccgcagca	ccctcaggat	ggagctgatg	ggctgtgacc	tgaactcctg	cagcatgccc	3900
ctgggcatgg	agagcaaggc	catttctgat	gcccagatca	ctgcctccag	ctacttcacc	3960
aacatgtttg	ccacctggag	cccaagcaag	gccaggctgc	acctccaggg	aaggagcaat	4020
gcctggaggc	cccaggtdaa	caacccaaag	gagtggtctg	aggtggactt	ccagaagacc	4080
atgaaggtca	ctggggtgac	caccagggg	gtcaagagcc	tgctcaccag	catgtatgtg	4140

-continued

```

aaggagttcc tgatcagctc cagccaggat gccaccagt ggacctctt cttccagaat 4200
ggcaaggtca aggtgttcca gggcaaccag gacagcttca cccctgtggt gaacagcctg 4260
gacccccccc tcctgaccag atacctgagg attcaccccc agagctgggt ccaccagatt 4320
gccctgagga tggaggtcct gggatgtgag gcccaggacc tgtactga 4368

```

<210> SEQ ID NO 12

<211> LENGTH: 1455

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 12

```

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
 1           5           10          15
Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
 20          25          30
Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
 35          40          45
Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
 50          55          60
Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
 65          70          75          80
Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
 85          90          95
Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
100         105         110
His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
115         120         125
Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
130         135         140
Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145         150         155         160
Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
165         170         175
Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
180         185         190
Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
195         200         205
Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
210         215         220
Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225         230         235         240
Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
245         250         255
Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
260         265         270
Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
275         280         285
Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
290         295         300

```


-continued

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
 305 310 315 320
 Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
 325 330 335
 Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
 340 345 350
 Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
 355 360 365
 Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
 370 375 380
 Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
 385 390 395 400
 Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
 405 410 415
 Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
 420 425 430
 Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
 435 440 445
 Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
 450 455 460
 Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
 465 470 475 480
 Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
 485 490 495
 His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
 500 505 510
 Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
 515 520 525
 Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
 530 535 540
 Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
 545 550 555 560
 Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
 565 570 575
 Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
 580 585 590
 Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
 595 600 605
 Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
 610 615 620
 Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
 625 630 635 640
 Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
 645 650 655
 Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
 660 665 670
 Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
 675 680 685
 Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
 690 695 700
 Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly

-continued

705	710	715	720
Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp 725 730 735			
Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys 740 745 750			
Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro 755 760 765			
Ser Thr Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln Glu Glu 770 775 780			
Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu Asp Phe 785 790 795 800			
Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys 805 810 815			
Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr 820 825 830			
Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly 835 840 845			
Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly 850 855 860			
Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly 865 870 875 880			
Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val 885 890 895			
Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser Leu 900 905 910			
Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg Lys Asn 915 920 925			
Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val Gln His 930 935 940			
His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp Ala Tyr 945 950 955 960			
Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu Ile Gly 965 970 975			
Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His Gly Arg 980 985 990			
Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe Asp Glu 995 1000 1005			
Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn Cys Arg 1010 1015 1020			
Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys Glu Asn 1025 1030 1035			
Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr Leu Pro 1040 1045 1050			
Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr Leu Leu 1055 1060 1065			
Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His Phe Ser Gly 1070 1075 1080			
His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met Ala Leu 1085 1090 1095			
Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met Leu Pro 1100 1105 1110			

-continued

Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly Glu His
 1115 1120 1125
 Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser Asn Lys
 1130 1135 1140
 Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg Asp Phe
 1145 1150 1155
 Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro Lys Leu
 1160 1165 1170
 Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser Thr Lys
 1175 1180 1185
 Glu Pro Phe Ser Trp Ile Lys Val Asp Leu Leu Ala Pro Met Ile
 1190 1195 1200
 Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe Ser Ser
 1205 1210 1215
 Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp Gly Lys
 1220 1225 1230
 Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu Met Val
 1235 1240 1245
 Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn Ile Phe
 1250 1255 1260
 Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro Thr His
 1265 1270 1275
 Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly Cys Asp
 1280 1285 1290
 Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys Ala Ile
 1295 1300 1305
 Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe
 1310 1315 1320
 Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg
 1325 1330 1335
 Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu
 1340 1345 1350
 Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr
 1355 1360 1365
 Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe
 1370 1375 1380
 Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe
 1385 1390 1395
 Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe
 1400 1405 1410
 Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr
 1415 1420 1425
 Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg
 1430 1435 1440
 Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
 1445 1450 1455

<210> SEQ ID NO 13

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

-continued

 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 13

```

atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc      60
accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg      120
ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac      180
acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt      240
gcaaaaccca gaccacctg gatgggactc ctgggaccca ccattcagge tgagggtgat      300
gacactgtgg tcatcacct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg      360
ggagtctcat actgaaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga      420
gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc      480
aaggagaatg gacctatggc atctgaccca ctctgctga catactccta cttttctcat      540
gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa      600
ggatccctgg ccaaggagaa aaccagaca ctgcacaagt tcattctcct gtttgtgtc      660
tttgatgagg gcaagtcttg gactctgaa acaaagaact ccctgatgca agacagggat      720
gtgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacagatca      780
ctgctgggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg      840
acaaccctg aagtgcactc cattttctg gagggacaca ctttctggt caggaaccac      900
agacaagcct ctctggagat ctctcccacc accttctca ctgcacagac actgctgatg      960
gaccttggac agttctctgct gttctgccc atctcttccc accagcatga tggcatggaa     1020
gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag     1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat     1140
gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca     1200
tgggtgcact acattgtctg tgaggaagag gactgggact atgcaccact ggtcctggcc     1260
cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccaaaaag aattggaaga     1320
aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc     1380
attcagcatg agtctggcat tctgggacca ctctctgatg gggaaagtggg agacaccctg     1440
ctcatcatct tcaagaacca ggctccagg ccctacaaca tctaccaca tggcatcact     1500
gatgtcagge ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc     1560
cccattctgc ctggagagat cttcaagtac aagtggactg tcaactgtga ggatggacca     1620
acaaagtctg accccagggt cctcaccaga tactactcct cttttgtgaa catggagaga     1680
gacctggcat ctggactgat tggacctg ctcatctgct acaaggagtc tgtggaccag     1740
agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag     1800
aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaacc tgctggggtg     1860
caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg     1920
tttgactctc tccagcttct tgtctgcctg catgaggtgg cctactggta cattctttct     1980
attggggcac aaactgactt cttttctgct ttcttctctg gatacacctt caagcacaag     2040
atggtgatag aggacacct gacactcttc ccattctctg gggaaactgt gttcatgagc     2100
atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga     2160

```

-continued

atgactgcac	tgctcaaagt	ctcctcctgt	gacaagaaca	ctggggacta	ctatgaggac	2220
tcttatgagg	acatctctgc	ctacctgctc	agcaagaaca	atgccattga	gcccagaagc	2280
ttctctcaga	atcccactgt	cctgaagaga	caccagagag	agatcaccag	gacaaccctc	2340
cagtctgacc	aggaagagat	tgactatgat	gacaccattt	ctgtggagat	gaagaaggag	2400
gactttgaca	tctatgatga	ggacgagAAC	cagtctccaa	gatcattcca	gaagaagaca	2460
agacactact	tcattgtctg	tgtggaaga	ctgtgggact	atggcatgct	ttcctctccc	2520
catgtcctca	ggaacagggc	acagtctggc	tctgtgccac	agttcaagaa	agtggctctc	2580
caggagtcca	ctgatggctc	attcaccocag	cccctgtaca	gaggggaact	gaatgagcac	2640
ctgggactcc	tgggaccata	catcagggct	gaggtggaag	acaacatcat	ggtgacattc	2700
agaaaccagg	cctccaggcc	ctacagcttc	tactcttccc	tcatcagcta	tgaggaagac	2760
cagagacaag	gggctgagcc	aagaaagaac	tttgtgaaac	ccaatgaaac	caagacctac	2820
ttctggaaag	tccagcacca	catggcaccc	accaaggatg	agtttgactg	caaggcctgg	2880
gcatacttct	ctgatgtgga	cctggagaaa	gatgtgcact	ctggcctgat	tggcccactc	2940
ctggtctgoc	acaccaacac	cctgaaccct	gcacatggaa	ggcaagtgac	tgtgcaggag	3000
tttgccctct	tcttcaccat	ctttgatgaa	accaagtcat	ggtacttcac	tgagaacatg	3060
gagagaaaact	gcagagcacc	atgcaacatt	cagatggaag	accccacctt	caaggagaac	3120
tacaggttcc	atgccatcaa	tggctacatc	atggacaccc	tgcctgggct	tgtcatggca	3180
caggaccaga	gaatcagatg	gtacctgctt	tctatgggat	ccaatgagaa	cattcactcc	3240
atccacttct	ctgggcatgt	cttcaactgtg	agaaagaagg	aggaatacaa	gatggcctg	3300
tacaacctct	accctggggg	ctttgagact	gtggagatgc	tgccctccaa	agctggcatc	3360
tggaggggtg	aatgcctcat	tggggagcac	ctgcatgctg	gcattgcaac	cctgttctctg	3420
gtctacagca	acaagtgcc	gacaccctg	ggaatggcct	ctggccacat	cagggacttc	3480
cagatcactg	cctctggcca	gtatggccag	tgggcaccca	aactggccag	gctccactac	3540
tctggctcca	tcaatgcatg	gtcaaccaag	gagccattct	cttggatcaa	ggtggacctg	3600
ctggcaccca	tgatcattca	tggcatcaag	acacaggggg	caagacagaa	attctcctct	3660
ctgtacatct	cacagttoat	catcatgtac	tctctggatg	gcaagaagtg	gcagacatac	3720
agaggcaact	ccactggcac	cctcatggtc	ttctttggca	atgtggacag	ctctggcatc	3780
aagcacaaca	tcttcaaccc	tcccatcatt	gccagataca	tcaggctgca	ccccaccac	3840
tactcaatca	gatcaaccct	caggatggaa	ctgatgggat	gtgacctgaa	ctcctgctca	3900
atgccctctg	gaatggagag	caaggccatt	tctgatgccc	agatcactgc	atcctcttac	3960
ttcaccaca	tgtttgccac	ctggctacca	tcaaaagcca	ggctgcacct	ccaggggaaga	4020
agcaatgcct	ggagacccca	ggtcaacaac	ccaaggaat	ggctgcaagt	ggacttccag	4080
aagacaatga	aagtcaactg	ggtgacaacc	cagggggctca	agtctctgct	cacctcaatg	4140
tatgtgaagg	agttcctgat	ctcttctctca	caggatggcc	accagtggac	actcttcttc	4200
cagaatggca	aagtcaagg	gttccagggc	aaccaggact	ctttcacacc	tgtgggtgaa	4260
tcaactggacc	ccccctctct	gacaagatac	ctgagaattc	accccagtc	ttgggtccac	4320
cagattgccc	tgagaatgga	agtcctggga	tgtgaggcac	aagacctgta	ctga	4374

-continued

```

<210> SEQ ID NO 14
<211> LENGTH: 4374
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

<400> SEQUENCE: 14
atgcagatcg aactgagcac ttgcttcttc ctgtgtctcc tgcgcttttg cttctccgcc      60
acaaggagat actatctcgg tgccgtggag ctcagctggg actacatgca gagcgacttg      120
ggtgaactgc ctgtggacgc cagggtttcca ccccgctgc ccaagagttt cccgttcaac      180
accagtgtcg tgtacaagaa aacctctctc gtggaattca cggaccacct gttcaacatc      240
gccaaaccgc gccctccctg gatggggctg ctcggcccga cgatccaggc tgaggctat      300
gacacggtgg tgattaccct caagaacatg gctagccacc cggtgagcct gcacgccgtg      360
ggcgtgtcct attggaagc gtccgagggt gcggagtacg atgaccagac ttcacagcgg      420
gagaaggaag acgacaaaagt gttccccggg ggttcccaca cctatgtctg gcaggctctg      480
aaggagaatg gtcctatggc ctccgacca ttgtgcctca cctactetta cctaagccat      540
gtggatctcg tcaaggacct gaactcgggg ctgatcggcg ccctgctcgt gtgccgggag      600
ggctcactgg ccaaggagaa gacccaaact ctgcacaagt tcatcctgct gttcgcggta      660
ttcgacgagg ggaagtctcg gcaactccgag accaagaaca gcctgatgca ggaccgcgac      720
gcagcctcgg cccgtgctg gccaaagatg cacaccgtga acggctacgt taacaggagc      780
ctacccggcc tgatcggctg ccaccgaaa tcggtctact ggcattgtgat cggaatgggc      840
acaacgcccc aggtccacag tatcttcttc gagggccaca ctttctcggg ccggaatcac      900
cgccaggcca gcctggagat cagccccata acctttctga cggcgcagac cttactcatg      960
gatctcggcc agttcctcct gttctgcacc atttctctcc accagcacga tgggatggaa      1020
gcatatgtga aagtggactc ctgccccgag gaaccccagc ttaggatgaa gaacaatgag      1080
gaggccgagg actacgacga tgaccttacc gattcagaaa tggacgtagt acgctttgac      1140
gacgacaact ctccatcctt catacagatt cgctccgtcg ccaagaagca ccctaagact      1200
tgggtgcact acatcgcggc cgaggaggag gactgggatt atgctcccct ggtgctggcc      1260
cccgacgacc gcagctacaa gagccagtac ctgaataacg ggccccagcg catcggccgg      1320
aagtacaaga aagtgcggtt catggcttac acggacgaga ccttcaagac ccgggaggct      1380
atccagcatg agagcggcat cttggggccc ctctgtacg gcgaagtgg agacacactg      1440
ctgatcatct tcaagaacca ggcgagcagg ccctacaaca tctaccccc cggcattacc      1500
gatgtccggc cgttgtacag ccgacggctg cccaagggcg tgaagcacct gaaggacttt      1560
ccgatcctgc cgggcgagat cttcaagtac aagtggactg tgaccgtgga ggatgggccg      1620
accaagagcg atccgcgctg cctgacccgt tactactcca gctttgtcaa tatggagcgc      1680
gacctcgeta gcgcttgat tggccctctg ctgatctgct acaaggagtc cgtggaccag      1740
agggggaatc agatcatgag tgacaagagg aacgtgatcc tgttctccgt gttcgcgaa      1800
aaccgcagct ggtatctcac cgagaatata cagcgtctcc tgccccaccc ggccggtgtg      1860
cagctggagg accccgagtt tcaggccagc aacatcatgc attctatcaa cggatatgtg      1920
tttgattccc tgcagctctc agtgtgtctg cacgaggtcg cctactggta tctctcagc      1980

```

-continued

attggggcac	agaccgactt	cctgagcgtg	ttcttctccg	ggtatacctt	caagcacaag	2040
atggtgtaag	aggataccct	gaccctgttc	ccctttagcg	gcgaaaccgt	gtttatgtct	2100
atggagaacc	ccgggctctg	gacccctggc	tgccataact	ccgacttccg	caaccgcgga	2160
atgaccgcgc	tcctgaaagt	gtcgagttgt	gacaagaaca	ccggcgacta	ttacgaggac	2220
agttacgagg	acatctctgc	gtacccctct	agcaagaata	acgccatcga	gccaagatcc	2280
ttcagccaga	acccccagct	gctgaagagg	catcagcggg	agatcacccg	cacgaccctg	2340
cagtcggatc	aggaggagat	tgattacgac	gacacgatca	gtgtggagat	gaagaaggag	2400
gacttcgaca	tctacgacga	agatgaaaac	cagtcccctc	ggctcctcca	aaagaagacc	2460
cggcactact	tcctgcgcgc	tgtggaacgc	ctgtgggact	atggaatgtc	ttctagccct	2520
cacgttttga	ggaaccgcgc	ccagtcgggc	agcgtgcccc	agttcaagaa	agtgggtgtc	2580
caggagttca	ccgacggctc	cttcacccag	ccactttacc	ggggcgagct	caatgaacat	2640
ctgggcctgc	tgggacccta	catcagggct	gaggtggagg	acaacatcat	ggtgacattc	2700
cggaatcagg	ccagcagacc	atacagtttc	tacagttcac	tcctctccta	cgaggaggac	2760
cagcgcagg	gggctgaacc	ccgtaagaac	ttcgtgaagc	caaacgaaac	aaagacctac	2820
ttctggaagg	tccagcacca	catggcacct	accaaggacg	agttcgattg	caaggcctgg	2880
gcctacttct	ccgacgtgga	cctggagaaa	gatgtgcaca	gcgccctgat	tggccctctg	2940
ctggtgtgtc	acacgaacac	actcaaccct	gcacacgggc	ggcaggtcac	tgtgcaggaa	3000
ttcgcctctg	tctttaccat	ctttgatgag	acgaagtctc	ggtatctcac	cgaaaacatg	3060
gagaggaact	gccgcgcacc	ctgcaacatc	cagatggaag	atccgacatt	caaggagaac	3120
taccggttcc	atgccatcaa	tggtctacatc	atggacaccc	tgctggcctc	cgtgatggcc	3180
caagaccagc	gtatccgctg	gtatctgctg	tcgatgggct	ccaacgagaa	catccatagt	3240
atccacttca	gcgggcatgt	cttcacgggt	aggaaaaagg	aggagtacaa	gatggcactg	3300
tacaacctct	atccccgctg	gttcgagacc	gtggagatgc	tgccctccaa	ggcggccatc	3360
tggagagtgg	aatgcctgat	cggcgagcac	ctccacgctg	ggatgtccac	gctgttctct	3420
gtttacagca	ataagtgcc	gacccctctg	ggcatggcga	gcgccacat	ccgcgacttc	3480
cagattacag	ccagcggcca	gtacggtcag	tgggctccaa	agctggcccc	tctgcactac	3540
tccggatcca	tcaacgcctg	gtccaccaag	gaaccgttct	cctggatcaa	agtagacctg	3600
ctagccccc	tgatcattca	cggcatcaag	acacaaggcg	cccacagaa	gttctcgagc	3660
ctctatatct	cccagttcat	catcatgtat	agcctggacg	gaaagaagtg	gcagacttac	3720
cgcggaaaact	cgacagggac	cctgatggta	ttcttcggta	acgtggacag	ctccggaatc	3780
aagcacaaca	tcttcaacct	accattatc	gcccgtaca	tccgctgca	ccccactcac	3840
tatagcatta	ggccaccctc	gcgaatggag	ctcatgggct	gtgacctgaa	cagctgtagc	3900
atgcccctcg	gcatggagtc	taaggcgatc	tccgacgcac	agataacggc	atcatctac	3960
tttaccaca	tgttcgctac	ctgggtcccc	tccaaggccc	gactccacct	gcaagggaga	4020
tccaacgcct	ggcggccaca	ggtcaacaat	cccaaggagt	ggctgcaagt	ggactttcag	4080
aaaactatga	aagtcaccgg	agtgaccaca	caggagtgta	agtctctgct	gaccagcatg	4140
tacgtgaagg	agttcctcat	ctccagttcg	caggatggcc	accagtgac	ggtgttcttc	4200
caaaacggta	aagtcaaaagt	cttccaaggg	aaccaggaca	gctttacacc	cgctgtgaac	4260

-continued

```
tccctggacc ccccgttct cactagatac ctccgcattc accctcagag ctgggtgcac 4320
cagattgccc tgcgcatgga ggttctgggg tgtgaagccc aggacctgta ctaa 4374
```

```
<210> SEQ ID NO 15
<211> LENGTH: 4374
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide
```

```
<400> SEQUENCE: 15
```

```
atgcagattg agctctccac ctgcttcttt ctctgcttc ttcgctctg cttttctgcc 60
acacgcaggt actatttggg agcagtgga ctgagctggg attacatgca gactgacctt 120
ggtgaacttc ctgtggagcg tcttttcca cctagagttc ccaagtcctt cccctcaac 180
acctcagtg tctacaagaa aacgctggtt gtggagtcca ctgacctctt cttcaacatt 240
gccccaaaca gaccccttg gatgggattg ctgggaccca caatacaagc agaagtctac 300
gacacggtgg tgattacct gaagaacatg gcgtcacacc ctgtttcact tcacgctggt 360
ggggtcagtt attgaaaagc ctacaggggt gcggaatacg atgatcaaac cagccagagg 420
gagaaggaag atgacaaggt ctttctggg ggtagccata cctatgtttg gcaggtgctg 480
aaagagaatg ggctatggc ctctgacctc ttgtgctca catactctta cctgagtcac 540
gtcgcacctg tgaagacct gaatagcggg ctgattggtg cactgcttgt ttgtagagag 600
gggagtttgg ccaaggagaa aactcagact ctccacaagt ttatctctct gtttgctgtg 660
ttcgacgagg gcaagtcttg gcactctgaa acaagaact ccctgatgca ggacagagat 720
gctgcatctg caagggcttg gcaaaaaatg cacacagtga acggtatgt gaatcgatca 780
ctgccaggac tgataggctg tcactgcaag tcagtgtatt ggcacgttat cgggatggga 840
acaactccag aagtgcacag catcttctct gagggccaca ctttctctgt tcggaatcat 900
agacaggcca gccttgagat cagcccaatc acctttctga ctgccccaac cttgctgatg 960
gatctgggac agttctctct gttttgtcac atctctctcc accaacatga cgggatggag 1020
gcttatgtga aggtcgatag ctgtccggag gaaccacaac tgaggatgaa gaacaacgaa 1080
gaggcagagg actatgacga cgactctgact gacagtgaaa tggacgtggt tcggttcgac 1140
gatgacaatt ctcttctatt tatccagatc cgttccgtgg ccaagaagca ccccaagact 1200
tgggttcatt acatcgctgc tgaggaggag gattgggact acgcgccctt ggtgttgccc 1260
ccagacgacg cctcatacaa gagccagtac cttaacaatg gtccacaaag gatcggccgg 1320
aagtacaaga aggttagatt tatggcttat accgacgaga cttttaaaac tagggaagca 1380
attcagcatg aaagtggcat tcttgacccc ctgctgtatg gcgaggttgg cgacacctg 1440
ctgattatct ttaagaacca ggcaagccgg cctacaaca tctaccgca cggcataacc 1500
gatgtacgac ccctgtacag tcgcagactt cctaagggg tgaaacacct gaaggacttc 1560
ccaattctgc ccggggagat cttcaagtat aaatggaccg tgacggttga ggatggtccc 1620
acaaagtccg atccgagatg ccttaccgca tattattcca gcttcgtgaa catggaaagg 1680
gacctggcca gcgggctgat tggcccactg ctgatttgtt acaaggagtc tgtcgatcaa 1740
agaggaaaac aaataatgag cgacaaacgt aacgtcatcc tgttcagcgt ctttgatgag 1800
aatagaagct ggtacctcac agaaaatatt cagcggtttc tgcctaacc cgcagcgtc 1860
```

-continued

cagctggaag atccccaggt ccaagcctca aacatcatgc atagcatcaa cggatacgt 1920
ttcgatagcc tgcagctgtc cgtctgtctc catgaagtgg catattggta catcctgagt 1980
atcggggcgc agaccgactt cctgagcgtg ttcttttctg gatacacgtt caaacacaaa 2040
atggtctatg aagataccct gactctgttt ccattctcag gagagacagt ctttatgagt 2100
atggaaaatc ctggactgtg gatcctgggc tgtcacaatt ctgattttcg gaacagaggc 2160
atgacagccc tgcttaaagt gagctcatgc gacaagaaca ccggtgatta ctacgaagat 2220
agctatgagg acatcagtg cgtattgtctc tccaagaaca acgctatcga gccacggtct 2280
ttcagtcaga atcctcccg tctgaagcgg catcagcgcg aaataacacg cacaccctt 2340
cagtcagacc aagaggaaat cgactacgat gatactatct ctgtggagat gaagaaggag 2400
gatttcgaca tttacgacga ggacgagaat cagtcccca ggagctttca gaagaaaaca 2460
agacactatt tcattgcccgc cgtggagcga ctgtgggact acggcatgtc tagctctccg 2520
catgtactta gaaatagggc acaaacggga tccgtgcctc agtttaagaa agttgtcttt 2580
caggagtta cagatggctc cttcaccag cccttgatc gcggggaact caatgaacac 2640
ctgggcctcc tgggtcctta tattagggcc gaagtcgagg acaatcatcat ggtgacctt 2700
aggaaccagg catctagacc ttactcttctc tactcctccc tgataccta tgaggaggac 2760
cagcggcaag gcgctgagcc tcggaagaac tttgtgaagc caaatgaaac caaacatac 2820
ttttgaaag ttcagcacca catggctccc acgaaggacg aatttgactg taaagcctgg 2880
gcctacttct cagatgtaga tctcgagaaa gacgtgcaact cagggtcctc tggtcccctc 2940
ctggtctgtc atactaatac cctcaatcca gcacacggac gtcaggtaac cgtccaggaa 3000
tttgccctgt tctttaccat tttcgatgag actaaatcct ggtactttac cgaaaacatg 3060
gagaggaatt gcagagcccc atgcaacatc cagatggagg accctacctt caaagagaac 3120
tatcgcttcc atgccattaa cggttacatt atggatactc tcccaggact tgtgatggca 3180
caggatcagc ggataagatg gtatctgttg agcatgggct ccaacgagaa tattcacagc 3240
atccatttct ccggtcactg gtttacagtg agaaagaaag aagagtacaa gatggctctg 3300
tataatctct atccaggcgt attcgaaacg gtggagatgt tgcctagcaa ggccggcatt 3360
tggcgagttag aatgccttat cggggaacat ctgcatgccg gaatgagcac gctcttctg 3420
gtgtatagta acaagtgcc aactccgctg ggcattggcat ctggccatat acgggacttt 3480
cagattacgg ctgacgggca gtatgggagc tgggcaccca aacttgccgcg actgcactat 3540
tcaggctcta tcaatgcatg gtccaccaag gaacccttct cttggattaa ggtggacctt 3600
tggcgccca tgataatcca tgggatcaaa acccagggcg ctgctcagaa attctcatca 3660
ctctacatct ctcagttcat aataatgat tcaactggatg ggaagaaatg gcagacttac 3720
agaggaaaca gcaccgggac gctgatggtg ttctttggca acgtggacag cagcggcatc 3780
aaacacaaca tcttcaatcc tcccattatt gcccggtata ttagactgca tcccactcac 3840
tactctatag gcagcacact taggatggag ctcatgggat gcgacctgaa cagttgtagt 3900
atgcccttgg ggatggagtc caaagctata agcgcgcac aaattacagc tagctcttac 3960
tttacgaata tgttgccac gtggagccca agcaaagccc ggetgcattt gcagggtcgg 4020
agtaatgctt ggcgcccaca ggtgaataac cctaaggaat ggttgaagt agatttcag 4080
aaaactatga aggtaaccgg cgtcactaca caggagatca agtcctctt gacctctatg 4140

-continued

```

tacgtcaagg agttcctgat tagcagcagt caggatgggc accaatggac actgttcttc 4200
cagaatggga aagttaaagt atttcagggt aaccaggact cctttacacc tgtggtgaat 4260
agcctcgacc caccctgct gacacgatac ctccgcatcc accctcagtc ttgggtgcat 4320
caaattgccc tgcgaaatgga ggtgttggga tgccaagctc aggacctcta ctga 4374

```

```

<210> SEQ ID NO 16
<211> LENGTH: 4374
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
                        polynucleotide

```

```

<400> SEQUENCE: 16

```

```

atgcagatcg aactctctac ttgcttcttc ctgtgccttc tgaggttctg cttctctgcc 60
actcgccgat attacctcgg ggcctgaggag ttgagttggg actacatgca atcagatctg 120
ggcgaactcc ctgtggatgc ccgattocca ccgcccgtgc ccaagtcttt cccatttaat 180
acttctgtgg tgtacaagaa gacattgttt gtggagtta ccgatcacct gttcaacatc 240
gccccaaaccg ggcctccatg gatgggtctg cttgggcccc ccattcaagc ggaggtctat 300
gatacagtgg tgataacgct taagaacatg gcgagccacc cagtgtctct gcatgccgtt 360
ggtgtatcat attggaaggc cagcgaagga gcggagtacg atgaccagac ctctcagaga 420
gagaaggaag acgataaggt ttttctcggc ggaagtcata catatgatg gcaggtcctg 480
aaagagaatg gcccgatggc ttctgacccc ctttgtctta cctatagtta tctgagccac 540
gtggacctgg tcaaggacct caacagtggg ctgattgggg ctctgcttgt ttgtagagag 600
ggtagccttg ctaaggagaa aacccaaaca ctccataagt tcattttgct gttcgcggtg 660
ttcgacgagg gaaagagttg gcacagcgaa acaagaatt cactgatgca agacagggac 720
gccccttccg caagggcttg gcctaagatg catacgggta atgggtatgt gaaccggagc 780
ctcccggggc tgatcgggtg ccategcaag tctgtttact ggcacgtcat tggaatgggg 840
acaacgccag aggtacatag tatatttctt gaaggccaca cgttctcgt acggaaccac 900
cgacaggctt ccctggagat aagccccatt acctttctga ccgctcagac tctgctgatg 960
gaccttggcc agtttctcct gttctgccat attagcagcc accagcacga cggtatggaa 1020
gcatacgtga aagtcgatag ctgtcctgag gaccctcagc tcagaatgaa gaacaacgag 1080
gaggccgaag actatgacga tgaccttaca gattccgaga tggacgtggg gcgctttgac 1140
gacgataaca gtctagttt cattcaaatc agatccgtag ccaaaaagca tccaaagaca 1200
tgggtgcatt acattgcagc cgaagaggag gattgggatt atgcgcccct tgttctggct 1260
ccagatgaca ggagctataa gtcccagtac ttgaacaacg ggccacagcg aatcggtaga 1320
aatataaga aggtaagatt catggcctac actgacgaaa catttaaac cagggaagct 1380
atccaacaag aatctggaat tctcggccct ctgctctacg gtgaggtggg ggacaccttg 1440
ctgatcattt tcaaaaatca ggcatccagg ccttacaaca tataccccca tggcatcacc 1500
gatgtccgcc cgctgtattc cagaagactc cccaagggag tgaacatct gaaagatctt 1560
cccctcctgc cgggcgagat ctttaatac aaatggactg tgactgtaga ggacggccct 1620
acaaaatcag acccacggtg cctgacaagg tattacagta gcttcgtcaa catggaacgc 1680

```

-continued

gacctcgcca gcgactcat tggcccactg ttgatctggt acaaagagtc agtggatcag	1740
aggggaaatc agatcatgag cgataagaga aacgttatcc tgtttagtgt cttcgacgag	1800
aaccggctctt ggtaccttac tgagaacatc cagaggttcc tgccgaatcc ggctggcgtt	1860
cagctcgagg acccagagtt ccaggccagt aatataatgc actcaatcaa cggttatgtg	1920
ttcgatagcc tgcagctgag cgtctgcctc cacgaggtag cctattggta catattgtcc	1980
atcggggctc agaccgattt tctgtccgtg ttctttagcg ggtatacctt taaacataaa	2040
atggtctatg aagacacct gacctgttc ccattctccg gtgagactgt gttcatgtcc	2100
atggagaacc cagggtctgt gatcctgggg tgtcacaata gtgactttag gaatcgggga	2160
atgacggcac tgetgaaggt gagtctctgc gataaaaata caggagatta ctatgaggat	2220
agttacgagg atatcagtgc ctatctgctt tcaaaaaaca acgcaattga gccccggtct	2280
ttctcacaaa accccccggt gctgaagcgc caccagcgcg aaattaccgg gacaaccttg	2340
cagtccgacc aggaggaaat cgattatgac gatactatca gtgtagaaat gaaaaaggag	2400
gattttgata tttacgacga agacgagaac cagtctccgc gaagttttca gaagaaaacg	2460
cgacactact ttatagctgc cgtggaacga ctctgggatt atggcatgtc ctccagccct	2520
catgtcctta ggaatcgagc gcagagtggc tctgtgcctc agttcaaaaa ggttgtgttc	2580
caggaattca ccgacggctc atttaccag ccgctgtaca gaggcgaact caacgaacac	2640
cttgggctgc ttggggcata tattcgagca gaggtggaag ataatatcat ggtaaccttt	2700
agaaaccagg cgtcaagacc ctattccttc tacagttctc tgatcagcta cgaggaggac	2760
caaagacagg gagctgaacc caggaagaac tttgtgaaac ctaatgagac caagacctac	2820
ttctggaagg tccagcacca tatggcccca actaaagatg aattcgattg caaggcctgg	2880
gcttatttca gcgacgtgga tctcgaaaag gatgtgcaca gcgggttgat cggaccgctt	2940
ttggtgtgcc acacaaatac cctcaatcct gcccaagggc ggcaggtcac agttcaagag	3000
tttgactct tctttacaat atttgacgag acaaagtcat ggtattttac agagaatatg	3060
gagagaaatt gtcgcgacc ttgcaacatt cagatggagg accccacatt taaggagaat	3120
tacagatttc atgctatcaa tgggtacatt atggatactc tgcttggctt ggtcatggcc	3180
caggatcagc gcataagggt gtacttctgt agcatgggat ctaatgagaa tatacacagc	3240
attcacttca gtggccacgt ttttactgtt agaaagaagg aggagtacaa aatggcgctc	3300
tacaaccttt acccgggtgt gtttgagaca gtggagatgc tgccaagcaa ggcaggcatc	3360
tggagggttg agtgtcttat tggggagcat ctgcatgctg gaatgtccac cctctttctt	3420
gtgtacagca ataagtgcc gacaccgctt ggcattggcca gcgccacat tagggacttt	3480
cagataactg ccagtggaca gtacggccag tgggctccca agcttgcaag actccactac	3540
tccggaagca taaacgatg gagcaccaag gaaccttct cttggattaa ggtggacctg	3600
ctggcgccaa tgatcattca cggcataaaa acccaagggg cagcagagaa attttcatct	3660
ttgtatatta gtcagtttat catcatgtac agcttggatg gaaagaagtg gcagacgtac	3720
aggggcaatt ctacaggaac acttatggtg ttttttggga atgtcgattc cagcgggatc	3780
aaacataaca tcttcaatcc tcctattatc gcccgatata tccgctgca ccctacgcat	3840
tactccatca ggtccacatt gagaatggaa ctgatggggg gcgacctgaa tagttgtagt	3900
atgccactgg gcattggatc taaagccatc agcgatgcac agatcaactgc cagctcttac	3960

-continued

ttcaccaaca tgtttgcaac ttggtccccc tctaaagctc gcctgcatct gcagggacgc	4020
tcaaatgcat ggcgaccaca ggtgaacaat ccaaagagt ggctccaggt cgactttcag	4080
aagacaatga aggtaacagg agtgacaacc caggggtgaa aaagcctcct tacgagtatg	4140
tacgttaagg agtttctgat ttctagctcc caggacggac accagtggac tctgttcttc	4200
cagaacggca aagtgaaggt atttcaggga aaccaggatt cttttacccc ggtagtgaat	4260
agcctggatc caccgttctg gaccgctat ctgagaattc atccacaatc ctgggtgcat	4320
cagattgccc tccggatgga agtgcctggc tgtgaagctc aggatctgta ttag	4374

<210> SEQ ID NO 17

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 17

atgcaaatag agctctccac ctgcttcttt ctgtgccttt tgcgattctg ctttagtgcc	60
accagaagat actacctggg tgcagtgaa ctgtcatggg actatatgca aagtgatctc	120
ggtgagctgc ctgtggacgc aagatttctt cctagagtgc caaaatcttt tccattcaac	180
acctcagtcg tgtacaaaa gactctggtt gtagaattca cggatcacct tttcaacatc	240
gctaagccaa ggccaccctg gatgggtctg ctaggctcta ccatccaggc tgaggtttat	300
gatacagtg tcatctact taagaacatg gcttccatc ctgtcagctc tcatgctgtt	360
ggtgtatcct actggaaagc ttctgaggga gctgaatatg atgatcagac cagtcaaagg	420
gagaagaag atgataaagt cttccctggg ggaagccata catatgtctg gcaggctctg	480
aaagagaatg gtccaatggc ctctgacca ctgtgcctta cctactcata tctttctcat	540
gtggacctgg taaaagactt gaattcaggc ctcatggag ccctactagt atgtagagaa	600
gggagtctgg ccaaggaaaa gacacagacc ttgcacaaat ttatactact ttttgctgta	660
tttgatgaag ggaaggttg gcactcagaa acaaagaact ccttgatgca ggatagggat	720
gctgcatctg ctccggcctg gcttaaaatg cacacagtca atggttatgt aaacaggtct	780
ctgccaggtc tgattggatg ccacaggaaa tcagtctatt ggcagtgtat tggaatgggc	840
accactcctg aagtgcactc aatattcttc gaaggtcaca catttcttgt gaggaacat	900
cgccaggcgt ccttggaat ctcccaata actttcctta ctgctcaaac actcttgatg	960
gaccttgac agtttctact gttttgtcat atctcttccc accaacatga tggcatggaa	1020
gcttatgtca aagtagacag ctgtccagag gaacccaac tacgaatgaa aaataatgaa	1080
gaagcggaa actatgatga tgatcttact gattctgaaa tggatgtggc caggtttgat	1140
gatgacaact ctcttctct tatccaaatt cgctcagttg ccaagaagca tcctaaaact	1200
tgggtacatt acattgctgc tgaagaggag gactgggact atgctccctt agtccctgcc	1260
cccgatgaca gaagtataa aagtcaatat ttgaacaatg gccctcagcg gattggtagg	1320
aagtacaaaa aagtccgatt tatggcatac acagatgaaa cctttaagac tcgtgaagct	1380
attcagcatg aatcaggaat cttgggacct ttactttatg gggaagttgg agacacactg	1440
ttgattatat ttaagaatca agcaagcaga ccatataaca tctaccctca cggaatcact	1500
gatgtccgtc ctttgattc aaggagatta ccaaaggtg taaaacattt gaaggatttt	1560

-continued

ccaattctgc	caggagaaat	attcaaatat	aatggacag	tgactgtaga	agatgggcca	1620
actaaatcag	atcctcggtg	cctgacccgc	tattactcta	gtttcgttaa	tatggagaga	1680
gatctagctt	caggactcat	tggccctctc	ctcatctgct	acaagaatc	tgtagatcaa	1740
agaggaaaacc	agataatgtc	agacaagagg	aatgtcatcc	tgttttctgt	atttgatgag	1800
aaccgaagct	ggtacctcac	agagaatata	caacgctttc	tccccaatcc	agctggagtg	1860
cagcttgagg	atccagagtt	ccaagcctcc	aacatcatgc	acagcatcaa	tggctatggt	1920
tttgatagtt	tgcagttgtc	agtttgttg	catgaggtgg	catactggta	cattctaage	1980
attggagcac	agactgactt	cctttctgtc	ttcttctctg	gatatacctt	caaacacaaa	2040
atggtctatg	aagacacact	caccctattc	ccattctcag	gagaaactgt	cttcatgtcg	2100
atggaaaacc	caggtctatg	gattctgggg	tgccacaact	cagactttcg	gaacagagggc	2160
atgaccgcct	tactgaaggt	ttctagttgt	gacaagaaca	ctggtgatta	ttacgaggac	2220
agttatgaag	atatttcagc	atacttgctg	agtaaaaaa	atgccattga	accaagaagc	2280
ttctcccaga	atccaccagt	cttgaaacgc	catcaacggg	aaataactcg	tactactctt	2340
cagtcagatc	aagaggaaat	tgactatgat	gataccatat	cagttgaaat	gaagaaggaa	2400
gattttgaca	tttatgatga	ggatgaaaat	cagagcccc	gcagctttca	aaagaaaaca	2460
cgacactatt	ttattgctgc	agtggagagg	ctctgggatt	atgggatgag	tagctcccca	2520
catgttctaa	gaaacagggc	tcagagtggc	agtgtccctc	agttcaagaa	agttgttttc	2580
caggaattta	ctgatggctc	ctttactcag	cccttatacc	gtggagaact	aatgaacat	2640
ttgggactcc	tggggccata	tataagagca	gaagttgaag	ataatatcat	ggtaactttc	2700
agaaatcagg	cctctcgtcc	ctattccttc	tattctagcc	ttatttctta	tgaggaagat	2760
cagaggcaag	gagcagaacc	tagaaaaaac	ttgtcaagc	ctaatgaaac	caaaacttac	2820
ttttgaaag	tgcaacatca	tatggcacc	actaaagatg	agtttgactg	caaagcctgg	2880
gcttatttct	ctgatgttga	cctggaaaaa	gatgtgcact	caggcctgat	tggaccctt	2940
ctggtctgcc	acactaacac	actgaaccct	gctcatggga	gacaagtgac	agtacaggaa	3000
tttgcctctg	ttttcaccat	ctttgatgag	acccaaaagct	ggtacttcac	tgaaaatatg	3060
gaaagaaact	gcagggctcc	ctgcaatata	cagatggaag	atcccacttt	taaagagaat	3120
tatcgcttcc	atgcaatcaa	tggctacata	atggatacac	tacctggctt	agtaatggct	3180
caggatcaaa	ggattcagatg	gtatctgctc	agcatgggca	gcaatgaaaa	catccattct	3240
attcatttca	gtggacatgt	gttcaactgta	cgaaaaaaag	aggagtataa	aatggcactg	3300
tacaatctct	atccaggtgt	ttttgagaca	gtggaaatgt	taccatccaa	agctggaatt	3360
tggcgggtgg	aatgccttat	tggcgagcat	ctacatgctg	ggatgagcac	actttttctg	3420
gtgtacagca	ataagtgta	gactccctg	ggaatggctt	ctggacacat	tagagatttt	3480
cagattacag	cttcaggaca	atatggacag	tgggccccaa	agctggccag	acttcattat	3540
tccggatcaa	tcaatgcctg	gagcaccaag	gagccctttt	cttggatcaa	ggtggatctg	3600
ttggcaccaa	tgattattca	cggcatcaag	accaggggtg	cccgtcagaa	gttctccagc	3660
ctctacatct	ctcagtttat	catcatgat	agtcttgatg	ggaagaagtg	gcagacttat	3720
cgaggaaatt	ccactggaac	cttaatggtc	ttctttggca	atgtggattc	atctgggata	3780
aaacacaata	tttttaacc	tccaattatt	gctcgataca	tccgtttgca	cccaactcat	3840

-continued

tatagcattc gcagcactct tcgcatggag ttgatgggct gtgatttaa tagttgcagc	3900
atgccattgg gaatggagag taaagcaata tcagatgcac agattactgc ttcacctac	3960
tttaccaata tgtttgccac ctggctcct tcaaaagctc gacttcacct ccaagggag	4020
agtaatgcct ggagacctca ggtgaataat ccaaagagt ggctgcaagt ggacttcag	4080
aagacaatga aagtcacagg agtaactact cagggagtaa aatctctgct taccagcatg	4140
tatgtgaagg agttcctcat ctccagcagt caagatggcc atcagtgga cctcttttt	4200
cagaatggca aagtaaagg ttttcaggga aatcaagact ccttcacacc tgtggtgaa	4260
tctctagacc caccgttact gactcgtac ctctgaattc acccccagag ttgggtgcac	4320
cagattgccc tgaggatgga ggttctgggc tgcgaggcac aggacctcta ctga	4374

<210> SEQ ID NO 18

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 18

atgcagatcg agctgtccac atgcttttt ctgtgcctgc tgcggttctg ctccagcgcc	60
acccggcggg actacctggg cgccgtggag ctgtcctggg actacatgca gagcgacctg	120
ggcgagctgc ccgtggacgc ccggttcccc cccagagtgc ccaagagctt ccccttcaac	180
accagcgtgg tgtacaagaa aaccctgttc gtggagtcca ccgaccacct gttcaacatc	240
gccaagccca ggccccctg gatgggcctg ctgggcccc ccatccaggc cgaggtgtac	300
gacacctggt tgatcacct gaagaacatg gccagccacc ccgtgagcct gcacgacctg	360
ggcgtgagct actggaaggc ctccgagggc gccaggtacg acgaccagac cagccagcgg	420
gagaaaagg acgacaaagt ctttctggc ggcagccaca cctacgtgtg gcaggtcctg	480
aaagaaaacg gccccatggc ctccgacccc ctgtgctga cctacagcta cctgagccac	540
gtggacctgg tgaaggacct gaacagcggg ctgattgggg ccctgctggt ctgccgggag	600
ggcagcctgg ccaaaagaa aaccagacc ctgcacaagt tcctcctgct gttcgcctg	660
ttcgacgagg gcaagagctg gcacagcag accaagaaca gcctgatgca ggaccgggac	720
gccgcctctg ccagagcctg gcccaagatg cacaccgtga acggtactgt gaacagaagc	780
ctgcccgccc tgattggctg ccaccggaag agcgtgtact ggcacgtgat cggcatgggc	840
accacaccgg aggtgcacag catctttctg gaagggcaca cctttctggt gcggaaccac	900
cggcaggcca gcctggaat cagccctatc accttctga ccgccagac actgctgatg	960
gacctgggcc agttcctgct gttttgccac atcagctctc accagcacga cggcatggaa	1020
gcctacgtga aggtggactc ctgccccgag gaaccccagc tgcggatgaa gaacaacgag	1080
gaagccgagg actacgacga cgacctgacc gacagcgaga tggacgtggt gcggttcgac	1140
gacgacaaca gccccagctt catccagatc agaagcgtgg ccaagaagca ccccaagacc	1200
tgggtgcaact acatcgccgc cgaggaagag gactgggact acgccccct ggtgctggcc	1260
cccagcaca gaagctacaa gagccagtac ctgaacaatg gccccagcg gatcgccgg	1320
aagtacaaga aagtgcgggt catggcctac accgacgaga ccttcaagac ccgggaggcc	1380

-continued

atccagcaag agagcggcat cctgggcccc ctgctgtacg gcgaagtggg cgacacactg	1440
ctgatcatct tcaagaacca ggccagcccg cctacaaca tctaccccca cggcatcacc	1500
gacgtgcggc ccctgtacag caggcggctg cccaagggcg tgaagcacct gaaggacttc	1560
cccatcctgc cggcggagat cttcaagtae aagtggaccg tgaccgtgga ggacggcccc	1620
accaagagcg accccagatg cctgaccccg tactacagca gcttcgtgaa catggaacgg	1680
gacctggcct cggggtgat cggacctctg ctgatctgct acaaagaaag cgtggaccag	1740
cggggcaacc agatcatgag cgacaagcgg aacgtgatcc tgttcagcgt gttcgatgag	1800
aaccggctct ggtatctgac cgagaacatc cagcggtttc tgcccaaccc tgccggggtg	1860
cagctggaag atccccagtt ccaggccagc aacatcatgc actccatcaa tggctacgtg	1920
ttcgacagcc tgcagctgtc cgtgtgtctg caccaggtgg cctactggta catcctgagc	1980
atcggcgcgc agaccgactt cctgagcgtg ttcttcagcg gctacacctt caagcacaag	2040
atggtgtacg aggacaccct gacctgttc cctttcagcg gcgagaccgt gttcatgagc	2100
atgaaaaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttccg gaaccggggc	2160
atgaccgccc tgctgaaggt gtcacagctc gacaagaaca ccggcgacta ctacaggac	2220
agctacgagg atatcagcgc ctacctgtg tccaagaaca acgcatcga gcccagaagc	2280
ttcagccaga acccccctgt gctgaagcgg caccagagag agatcaccgg gaccacctg	2340
cagtccgacc aggaagagat cgattacgac gacaccatca gcgtggagat gaaaaaagaa	2400
gatttcgaca tctacgacga ggacgagAAC cagagcccc ggtccttcca gaagaaaacc	2460
cggcactact ttatcgcgcg cgtggagcgg ctgtgggact acggcatgag cagcagcccc	2520
cacgtgctgc ggaaccgggc ccagagcggc agcgtgcccc agttcaagaa agtgggtgtc	2580
caggaaatca ccgacggcag cttcaccocg cccctgtacc ggggcgagct gaacgagcac	2640
ctggggctgc tggggcccta catcagggcc gaagtggagg acaacatcat ggtgacctc	2700
cggaatcagg ccagcagacc ctactcctc tacagcagcc tgatcagcta cgaagaggac	2760
cagcggcagg gcgctgaacc ccggaagaac ttcgtgaagc ccaatgagac caagacctac	2820
ttctgaaaag tgcagcacca catggcccc accaaggacg agttcgactg caaggcctgg	2880
gcctacttca gcgacgtgga tctggaaaag gacgtgcact ctggactgat tggccctctg	2940
ctggtgtgcc acaccaacac cctgaacccc gccacggcc ggcaggtgac cgtgcaggaa	3000
ttcgccctgt tcttcaccat cttcgacgag accaagtccct ggtacttcac cgagaatatg	3060
gaacggaaact gcagagcccc ctgcaacatc cagatggaag atcctacctt caaagagaac	3120
taccggttcc acgcatcaa cggctacatc atggaccccc tgctggcct ggtgatggcc	3180
caggaccaga ggatccgggtg gtatctgctg tccatgggca gcaacgagaa tatccacagc	3240
atccacttca cggcccacgt gttcacctg aggaagaaag aagagtacaa gatggccctg	3300
tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggcggcatc	3360
tggcgggtgg agtgtctgat cggcagacac ctgcatgccg ggatgagcac cctgtttctg	3420
gtgtacagca acaagtgcc aacccccctg ggcattggcca gcggccacat ccgggacttc	3480
cagatcaccc cctccggcca gtacggccag tgggccccca agctggcccg gctgcactac	3540
agcggcagca tcaacgcctg gtccacaaa gagcccttca gctggatcaa ggtggacctg	3600
ctggccccta tgatcatoca cggcattaag acccagggcg ccaggcagaa gttcagcagc	3660

-continued

```

ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720
cggggcaaca gcaccggcac cctgatgggtg ttcttcggca acgtggacag cagcggcatc 3780
aagcacaaca tcttcaaccc ccccatcadc gcccggtaca tccgggtgca ccccaccac 3840
tacagcatca gatccacctc gcggatggaa ctgatgggct gcgacctgaa ctctgcagc 3900
atgcctctgg gcattgaaag caaggccatc agcgacgccc agatcacagc cagcagctac 3960
ttcaccaaca tgttcgccac ctgggtcccc tccaaggcca ggctgcacct gcagggccgg 4020
tccaacgcct ggcggcctca ggtgaacaac cccaaagaat ggctgcaggt ggactttcag 4080
aaaaccatga aggtgaccgg cgtgaccacc cagggcgtga aaagcctgct gaccagcatg 4140
tacgtgaaag agtttctgat cagcagcagc caggacggcc accagtggac cctgttcttt 4200
cagaacggca aggtgaaagt gttccagggc aaccaggact ccttcacccc cgtggtgaac 4260
tccctggacc cccccctgct gaccgcctac ctgctggatcc acccccagtc ttgggtgcac 4320
cagatcgccc tgaggatgga agtgcctggga tgtgaggccc aggatctgta ctga 4374

```

<210> SEQ ID NO 19

<211> LENGTH: 2351

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 19

```

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1           5           10           15
Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20          25          30
Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
35          40          45
Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
50          55          60
Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65          70          75          80
Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
85          90          95
Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
100         105         110
His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
115         120         125
Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
130         135         140
Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145         150         155         160
Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
165         170         175
Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
180         185         190
Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
195         200         205
Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
210         215         220
Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225         230         235         240

```


-continued

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
245 250 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
260 265 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
275 280 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
290 295 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305 310 315 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
325 330 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
340 345 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
355 360 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
370 375 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385 390 395 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
405 410 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
420 425 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
435 440 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
450 455 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465 470 475 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
485 490 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
500 505 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
515 520 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
530 535 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545 550 555 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
565 570 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
580 585 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
595 600 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
610 615 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625 630 635 640

-continued

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
645 650 655
Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
660 665 670
Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
675 680 685
Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
690 695 700
Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705 710 715 720
Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735
Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
740 745 750
Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro
755 760 765
Ser Thr Arg Gln Lys Gln Phe Asn Ala Thr Thr Ile Pro Glu Asn Asp
770 775 780
Ile Glu Lys Thr Asp Pro Trp Phe Ala His Arg Thr Pro Met Pro Lys
785 790 795 800
Ile Gln Asn Val Ser Ser Ser Asp Leu Leu Met Leu Leu Arg Gln Ser
805 810 815
Pro Thr Pro His Gly Leu Ser Leu Ser Asp Leu Gln Glu Ala Lys Tyr
820 825 830
Glu Thr Phe Ser Asp Asp Pro Ser Pro Gly Ala Ile Asp Ser Asn Asn
835 840 845
Ser Leu Ser Glu Met Thr His Phe Arg Pro Gln Leu His His Ser Gly
850 855 860
Asp Met Val Phe Thr Pro Glu Ser Gly Leu Gln Leu Arg Leu Asn Glu
865 870 875 880
Lys Leu Gly Thr Thr Ala Ala Thr Glu Leu Lys Lys Leu Asp Phe Lys
885 890 895
Val Ser Ser Thr Ser Asn Asn Leu Ile Ser Thr Ile Pro Ser Asp Asn
900 905 910
Leu Ala Ala Gly Thr Asp Asn Thr Ser Ser Leu Gly Pro Pro Ser Met
915 920 925
Pro Val His Tyr Asp Ser Gln Leu Asp Thr Thr Leu Phe Gly Lys Lys
930 935 940
Ser Ser Pro Leu Thr Glu Ser Gly Gly Pro Leu Ser Leu Ser Glu Glu
945 950 955 960
Asn Asn Asp Ser Lys Leu Leu Glu Ser Gly Leu Met Asn Ser Gln Glu
965 970 975
Ser Ser Trp Gly Lys Asn Val Ser Ser Thr Glu Ser Gly Arg Leu Phe
980 985 990
Lys Gly Lys Arg Ala His Gly Pro Ala Leu Leu Thr Lys Asp Asn Ala
995 1000 1005
Leu Phe Lys Val Ser Ile Ser Leu Leu Lys Thr Asn Lys Thr Ser
1010 1015 1020
Asn Asn Ser Ala Thr Asn Arg Lys Thr His Ile Asp Gly Pro Ser
1025 1030 1035
Leu Leu Ile Glu Asn Ser Pro Ser Val Trp Gln Asn Ile Leu Glu

-continued

1040	1045	1050
Ser Asp Thr Glu Phe Lys Lys	Val Thr Pro Leu Ile His Asp Arg	1065
1055	1060	
Met Leu Met Asp Lys Asn Ala	Thr Ala Leu Arg Leu Asn His Met	1080
1070	1075	
Ser Asn Lys Thr Thr Ser Ser	Lys Asn Met Glu Met Val Gln Gln	1095
1085	1090	
Lys Lys Glu Gly Pro Ile Pro	Pro Asp Ala Gln Asn Pro Asp Met	1110
1100	1105	
Ser Phe Phe Lys Met Leu Phe	Leu Pro Glu Ser Ala Arg Trp Ile	1125
1115	1120	
Gln Arg Thr His Gly Lys Asn	Ser Leu Asn Ser Gly Gln Gly Pro	1140
1130	1135	
Ser Pro Lys Gln Leu Val Ser	Leu Gly Pro Glu Lys Ser Val Glu	1155
1145	1150	
Gly Gln Asn Phe Leu Ser Glu	Lys Asn Lys Val Val Val Gly Lys	1170
1160	1165	
Gly Glu Phe Thr Lys Asp Val	Gly Leu Lys Glu Met Val Phe Pro	1185
1175	1180	
Ser Ser Arg Asn Leu Phe Leu	Thr Asn Leu Asp Asn Leu His Glu	1200
1190	1195	
Asn Asn Thr His Asn Gln Glu	Lys Lys Ile Gln Glu Glu Ile Glu	1215
1205	1210	
Lys Lys Glu Thr Leu Ile Gln	Glu Asn Val Val Leu Pro Gln Ile	1230
1220	1225	
His Thr Val Thr Gly Thr Lys	Asn Phe Met Lys Asn Leu Phe Leu	1245
1235	1240	
Leu Ser Thr Arg Gln Asn Val	Glu Gly Ser Tyr Asp Gly Ala Tyr	1260
1250	1255	
Ala Pro Val Leu Gln Asp Phe	Arg Ser Leu Asn Asp Ser Thr Asn	1275
1265	1270	
Arg Thr Lys Lys His Thr Ala	His Phe Ser Lys Lys Gly Glu Glu	1290
1280	1285	
Glu Asn Leu Glu Gly Leu Gly	Asn Gln Thr Lys Gln Ile Val Glu	1305
1295	1300	
Lys Tyr Ala Cys Thr Thr Arg	Ile Ser Pro Asn Thr Ser Gln Gln	1320
1310	1315	
Asn Phe Val Thr Gln Arg Ser	Lys Arg Ala Leu Lys Gln Phe Arg	1335
1325	1330	
Leu Pro Leu Glu Glu Thr Glu	Leu Glu Lys Arg Ile Ile Val Asp	1350
1340	1345	
Asp Thr Ser Thr Gln Trp Ser	Lys Asn Met Lys His Leu Thr Pro	1365
1355	1360	
Ser Thr Leu Thr Gln Ile Asp	Tyr Asn Glu Lys Glu Lys Gly Ala	1380
1370	1375	
Ile Thr Gln Ser Pro Leu Ser	Asp Cys Leu Thr Arg Ser His Ser	1395
1385	1390	
Ile Pro Gln Ala Asn Arg Ser	Pro Leu Pro Ile Ala Lys Val Ser	1410
1400	1405	
Ser Phe Pro Ser Ile Arg Pro	Ile Tyr Leu Thr Arg Val Leu Phe	1425
1415	1420	

-continued

Gln Asp	Asn Ser Ser His Leu	Pro Ala Ala Ser Tyr	Arg Lys Lys
1430	1435	1440	
Asp Ser	Gly Val Gln Glu Ser	Ser His Phe Leu Gln	Gly Ala Lys
1445	1450	1455	
Lys Asn	Asn Leu Ser Leu Ala	Ile Leu Thr Leu Glu	Met Thr Gly
1460	1465	1470	
Asp Gln	Arg Glu Val Gly Ser	Leu Gly Thr Ser Ala	Thr Asn Ser
1475	1480	1485	
Val Thr	Tyr Lys Lys Val Glu	Asn Thr Val Leu Pro	Lys Pro Asp
1490	1495	1500	
Leu Pro	Lys Thr Ser Gly Lys	Val Glu Leu Leu Pro	Lys Val His
1505	1510	1515	
Ile Tyr	Gln Lys Asp Leu Phe	Pro Thr Glu Thr Ser	Asn Gly Ser
1520	1525	1530	
Pro Gly	His Leu Asp Leu Val	Glu Gly Ser Leu Leu	Gln Gly Thr
1535	1540	1545	
Glu Gly	Ala Ile Lys Trp Asn	Glu Ala Asn Arg Pro	Gly Lys Val
1550	1555	1560	
Pro Phe	Leu Arg Val Ala Thr	Glu Ser Ser Ala Lys	Thr Pro Ser
1565	1570	1575	
Lys Leu	Leu Asp Pro Leu Ala	Trp Asp Asn His Tyr	Gly Thr Gln
1580	1585	1590	
Ile Pro	Lys Glu Glu Trp Lys	Ser Gln Glu Lys Ser	Pro Glu Lys
1595	1600	1605	
Thr Ala	Phe Lys Lys Lys Asp	Thr Ile Leu Ser Leu	Asn Ala Cys
1610	1615	1620	
Glu Ser	Asn His Ala Ile Ala	Ala Ile Asn Glu Gly	Gln Asn Lys
1625	1630	1635	
Pro Glu	Ile Glu Val Thr Trp	Ala Lys Gln Gly Arg	Thr Glu Arg
1640	1645	1650	
Leu Cys	Ser Gln Asn Pro Pro	Val Leu Lys Arg His	Gln Arg Glu
1655	1660	1665	
Ile Thr	Arg Thr Thr Leu Gln	Ser Asp Gln Glu Glu	Ile Asp Tyr
1670	1675	1680	
Asp Asp	Thr Ile Ser Val Glu	Met Lys Lys Glu Asp	Phe Asp Ile
1685	1690	1695	
Tyr Asp	Glu Asp Glu Asn Gln	Ser Pro Arg Ser Phe	Gln Lys Lys
1700	1705	1710	
Thr Arg	His Tyr Phe Ile Ala	Ala Val Glu Arg Leu	Trp Asp Tyr
1715	1720	1725	
Gly Met	Ser Ser Ser Pro His	Val Leu Arg Asn Arg	Ala Gln Ser
1730	1735	1740	
Gly Ser	Val Pro Gln Phe Lys	Lys Val Val Phe Gln	Glu Phe Thr
1745	1750	1755	
Asp Gly	Ser Phe Thr Gln Pro	Leu Tyr Arg Gly Glu	Leu Asn Glu
1760	1765	1770	
His Leu	Gly Leu Leu Gly Pro	Tyr Ile Arg Ala Glu	Val Glu Asp
1775	1780	1785	
Asn Ile	Met Val Thr Phe Arg	Asn Gln Ala Ser Arg	Pro Tyr Ser
1790	1795	1800	

-continued

Phe	Tyr	Ser	Ser	Leu	Ile	Ser	Tyr	Glu	Glu	Asp	Gln	Arg	Gln	Gly
1805						1810					1815			
Ala	Glu	Pro	Arg	Lys	Asn	Phe	Val	Lys	Pro	Asn	Glu	Thr	Lys	Thr
1820						1825					1830			
Tyr	Phe	Trp	Lys	Val	Gln	His	His	Met	Ala	Pro	Thr	Lys	Asp	Glu
1835						1840					1845			
Phe	Asp	Cys	Lys	Ala	Trp	Ala	Tyr	Phe	Ser	Asp	Val	Asp	Leu	Glu
1850						1855					1860			
Lys	Asp	Val	His	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Val	Cys	His
1865						1870					1875			
Thr	Asn	Thr	Leu	Asn	Pro	Ala	His	Gly	Arg	Gln	Val	Thr	Val	Gln
1880						1885					1890			
Glu	Phe	Ala	Leu	Phe	Phe	Thr	Ile	Phe	Asp	Glu	Thr	Lys	Ser	Trp
1895						1900					1905			
Tyr	Phe	Thr	Glu	Asn	Met	Glu	Arg	Asn	Cys	Arg	Ala	Pro	Cys	Asn
1910						1915					1920			
Ile	Gln	Met	Glu	Asp	Pro	Thr	Phe	Lys	Glu	Asn	Tyr	Arg	Phe	His
1925						1930					1935			
Ala	Ile	Asn	Gly	Tyr	Ile	Met	Asp	Thr	Leu	Pro	Gly	Leu	Val	Met
1940						1945					1950			
Ala	Gln	Asp	Gln	Arg	Ile	Arg	Trp	Tyr	Leu	Leu	Ser	Met	Gly	Ser
1955						1960					1965			
Asn	Glu	Asn	Ile	His	Ser	Ile	His	Phe	Ser	Gly	His	Val	Phe	Thr
1970						1975					1980			
Val	Arg	Lys	Lys	Glu	Glu	Tyr	Lys	Met	Ala	Leu	Tyr	Asn	Leu	Tyr
1985						1990					1995			
Pro	Gly	Val	Phe	Glu	Thr	Val	Glu	Met	Leu	Pro	Ser	Lys	Ala	Gly
2000						2005					2010			
Ile	Trp	Arg	Val	Glu	Cys	Leu	Ile	Gly	Glu	His	Leu	His	Ala	Gly
2015						2020					2025			
Met	Ser	Thr	Leu	Phe	Leu	Val	Tyr	Ser	Asn	Lys	Cys	Gln	Thr	Pro
2030						2035					2040			
Leu	Gly	Met	Ala	Ser	Gly	His	Ile	Arg	Asp	Phe	Gln	Ile	Thr	Ala
2045						2050					2055			
Ser	Gly	Gln	Tyr	Gly	Gln	Trp	Ala	Pro	Lys	Leu	Ala	Arg	Leu	His
2060						2065					2070			
Tyr	Ser	Gly	Ser	Ile	Asn	Ala	Trp	Ser	Thr	Lys	Glu	Pro	Phe	Ser
2075						2080					2085			
Trp	Ile	Lys	Val	Asp	Leu	Leu	Ala	Pro	Met	Ile	Ile	His	Gly	Ile
2090						2095					2100			
Lys	Thr	Gln	Gly	Ala	Arg	Gln	Lys	Phe	Ser	Ser	Leu	Tyr	Ile	Ser
2105						2110					2115			
Gln	Phe	Ile	Ile	Met	Tyr	Ser	Leu	Asp	Gly	Lys	Lys	Trp	Gln	Thr
2120						2125					2130			
Tyr	Arg	Gly	Asn	Ser	Thr	Gly	Thr	Leu	Met	Val	Phe	Phe	Gly	Asn
2135						2140					2145			
Val	Asp	Ser	Ser	Gly	Ile	Lys	His	Asn	Ile	Phe	Asn	Pro	Pro	Ile
2150						2155					2160			
Ile	Ala	Arg	Tyr	Ile	Arg	Leu	His	Pro	Thr	His	Tyr	Ser	Ile	Arg
2165						2170					2175			
Ser	Thr	Leu	Arg	Met	Glu	Leu	Met	Gly	Cys	Asp	Leu	Asn	Ser	Cys

-continued

2180	2185	2190
Ser Met Pro Leu Gly Met Glu Ser Lys Ala Ile Ser Asp Ala Gln 2195 2200 2205		
Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe Ala Thr Trp Ser 2210 2215 2220		
Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg Ser Asn Ala Trp 2225 2230 2235		
Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu Gln Val Asp Phe 2240 2245 2250		
Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr Gln Gly Val Lys 2255 2260 2265		
Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe Leu Ile Ser Ser 2270 2275 2280		
Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe Gln Asn Gly Lys 2285 2290 2295		
Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe Thr Pro Val Val 2300 2305 2310		
Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile His 2315 2320 2325		
Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg Met Glu Val Leu 2330 2335 2340		
Gly Cys Glu Ala Gln Asp Leu Tyr 2345 2350		

<210> SEQ ID NO 20

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 20

```

atgcagattg agctgagcac ctgcttcttc ctgtgctgc tgaggttctg cttctctgcc      60
accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg      120
ggcgagctgc ctgtggacgc cagggtcccc cccagagtgc ccaagagctt ccccttcaac      180
acctcagtgg tgtacaagaa gacctgttc gtggagtcca cggaccacct gttcaacate      240
gccaagccca ggccccctg gatgggctg ctgggccccca ccatccaggg cgaggtgtac      300
gacaccgtgg tgatcaccct gaagaacatg gccagcccacc ccgtgagcct gcacgcccgtg      360
ggcgtgagct actggaaggg ctctgagggc gccgagtatg acgaccagac cagccagagg      420
gagaaggagg acgacaaggt gttccccggc gccagccaca cctacgtgtg gcaggtgctg      480
aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacageta cctgagccac      540
gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggg gtgcaggggag      600
ggcagcctgg ccaaggagaa gacctcagacc ctgcacaagt tcatcctgct gttcgccgtg      660
ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat      720
gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc      780
ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc      840
accacccccg aggtgcacag catcttctctg gagggccaca ccttctctgt gaggaaccac      900

```

-continued

aggcaggcca	gcctggagat	cagccccatc	accttcctga	ccgcccagac	cctgctgatg	960
gacctgggcc	agttcctgct	gtttctgccac	atcagcagcc	accagcacga	cggcatggag	1020
gcctacgtga	agggtggacag	ctgccccgag	gagccccagc	tgaggatgaa	gaacaacgag	1080
gaggccgagg	actatgatga	tgacctgacc	gactctgaga	tggacgtggt	gaggtttgat	1140
gatgacaaca	gccccagctt	catccagatc	aggctctgtg	ccaagaagca	ccccaaagacc	1200
tgggtgcact	acatcgccgc	cgaggaggag	gactgggact	acgccccct	ggtgctggcc	1260
cccagcgaca	ggagctaaa	gagccagtac	ctgaacaacg	gccccagag	gatcggcagg	1320
aagtacaaga	aggtcagatt	catggcctac	accgacgaga	ccttcaagac	cagggaggcc	1380
atccagcacg	agtctggcat	cctgggcccc	ctgctgtacg	gcgagggtgg	cgacaccctg	1440
ctgatcatct	tcaagaacca	ggccagcag	ccctacaaca	tctaccccc	cggcatacc	1500
gatgtgaggc	ccctgtacag	caggaggctg	cccaagggcg	tgaagcacct	gaaggacttc	1560
cccatacctgc	ccggcgagat	cttcaagtac	aagtggacgc	tgaccgtgga	ggatggcccc	1620
accaagtctg	accccaggtg	cctgaccagg	tactacagca	gcttcgtgaa	catggagagg	1680
gacctggcct	ctggcctgat	cggccccctg	ctgatctgct	acaaggagag	cgtggaccag	1740
aggggcaacc	agatcatgtc	tgacaagagg	aacgtgatcc	tgttctctgt	gttcgatgag	1800
aacaggagct	ggtatctgac	cgagaacatc	cagaggttcc	tgccccacc	cgccggcgtg	1860
cagctggagg	accccaggtt	ccaggccagc	aacatcatgc	acagcatcaa	cggtacgtg	1920
ttcgacagcc	tgcagctgtc	tgtgtgctc	cacgaggtg	cctactggta	catcctgagc	1980
atcggcgccc	agaccgactt	cctgtctgtg	ttcttctctg	gctacacctt	caagcacaag	2040
atggtgtacg	aggacaccct	gacctgttc	cccttcagcg	gcgagaccgt	gttcgatgagc	2100
atggagaacc	ccggcctgtg	gatcctgggc	tgccacaaca	gcgacttcag	gaacaggggc	2160
atgaccgccc	tgctgaaagt	cagcagctgc	gacaagaaca	ccggcgacta	ctacgaggac	2220
agctacgagg	acatcagcgc	ctacctgtg	agcaagaaca	acgccatcga	gcccaggagc	2280
ttcagccaga	acccccctg	gctgaagagg	caccagagg	agatcaccag	gaccaccctg	2340
cagagcgacc	aggaggagat	cgactatgat	gacaccatca	gcgtggagat	gaagaaggag	2400
gacttcgaca	tctacgacga	ggacgagaac	cagagcccca	ggagcttcca	gaagaagacc	2460
aggcactact	tcatcgccgc	cgtggagagg	ctgtgggact	atggcatgag	cagcagcccc	2520
cacgtgctga	ggaacagggc	ccagagcggc	agcgtgcccc	agttcaagaa	ggtggtgttc	2580
caggagttea	ccgacggcag	cttcacccag	cccctgtaca	gaggcgagct	gaacgagcac	2640
ctgggcctgc	tgggccccta	catcagggcc	gaggtggagg	acaacatcat	ggtgaccttc	2700
aggaaccagg	ccagcaggcc	ctacagcttc	tacagcagcc	tgatcagcta	cgaggaggac	2760
cagaggcagg	gcgccgagcc	caggaagaac	ttcgtgaagc	ccaacgagac	caagacctac	2820
ttctggaagg	tgcagcacca	catggcccc	accaaggacg	agttcgactg	caaggcctgg	2880
gcctactctt	ctgatgtgga	cctggagaag	gacgtgcaca	gcgccctgat	cggccccctg	2940
ctggtgtgcc	acaccaaac	cctgaacccc	gcccacggca	ggcaggtgac	cgtgcaggag	3000
ttcgccctgt	tcttaccat	cttcgacgag	accaagagct	ggtacttcac	cgagaacatg	3060
gagaggaact	gcagggcccc	ctgcaacatc	cagatggagg	acccacctt	caaggagaac	3120
tacaggttcc	acgccatcaa	cggtacatc	atggacaccc	tgccccct	ggtgatggcc	3180

-continued

```

caggaccaga ggatcaggtg gtatctgctg agcatgggca gcaacgagaa catccacagc 3240
atccacttca gcgccacgt gttcaccgtg aggaagaagg aggagtacaa gatggccctg 3300
tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccggcatc 3360
tggagggtgg agtgcctgat cggcgagcac ctgcacgccc gcatgagcac cctgttctctg 3420
gtgtacagca acaagtgcc aacccccctg ggcatggcca gcgccacat cagggacttc 3480
cagatcacgg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac 3540
agcggcagca tcaacgcctg gaccaccaag gagcccttca gctggatcaa ggtggactctg 3600
ctggccccca tgatcatcca cggcatcaag acccagggcg ccaggcagaa gttcagcagc 3660
ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720
aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc 3780
aagcacaaca tcttcaaccc ccccatcatc gccaggtaca tcaggctgca cccccaccac 3840
tacagcatca ggagcacctc gcggtggaa ctgatgggct gcgacctgaa cagctgcagc 3900
atgccctctg gcatggagag caaggccatc tctgacgccc agatcacgcg cagcagctac 3960
ttcaccaaca tgttcgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg 4020
agcaacgcct ggaggcccc ggtgaacaac cccaaggagt ggctgcaggt ggacttcag 4080
aagaccatga aggtgaccgg cgtgaccacc cagggcgtga agagcctgct gaccagcatg 4140
tacgtgaagg agttcctgat cagcagcagc caggacggcc accagtgagc cctgttcttc 4200
cagaacggca aagtgaaggt gttccagggc aaccaggaca gcttcacccc cgtggtgaac 4260
agcctggacc cccccctgct gaccaggtat ctgaggatcc acccccagag ctgggtgcac 4320
cagatcgccc tgagaatgga agtgcctggga tgcgaggccc aggacctgta ctga 4374

```

<210> SEQ ID NO 21

<211> LENGTH: 1457

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 21

```

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1             5             10            15
Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20            25            30
Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
35            40            45
Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
50            55            60
Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65            70            75            80
Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
85            90            95
Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
100           105           110
His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
115           120           125
Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp

```


-continued

130			135			140									
Asp	Lys	Val	Phe	Pro	Gly	Gly	Ser	His	Thr	Tyr	Val	Trp	Gln	Val	Leu
145					150					155					160
Lys	Glu	Asn	Gly	Pro	Met	Ala	Ser	Asp	Pro	Leu	Cys	Leu	Thr	Tyr	Ser
				165					170						175
Tyr	Leu	Ser	His	Val	Asp	Leu	Val	Lys	Asp	Leu	Asn	Ser	Gly	Leu	Ile
			180					185					190		
Gly	Ala	Leu	Leu	Val	Cys	Arg	Glu	Gly	Ser	Leu	Ala	Lys	Glu	Lys	Thr
		195					200					205			
Gln	Thr	Leu	His	Lys	Phe	Ile	Leu	Leu	Phe	Ala	Val	Phe	Asp	Glu	Gly
210						215					220				
Lys	Ser	Trp	His	Ser	Glu	Thr	Lys	Asn	Ser	Leu	Met	Gln	Asp	Arg	Asp
225					230						235				240
Ala	Ala	Ser	Ala	Arg	Ala	Trp	Pro	Lys	Met	His	Thr	Val	Asn	Gly	Tyr
				245					250						255
Val	Asn	Arg	Ser	Leu	Pro	Gly	Leu	Ile	Gly	Cys	His	Arg	Lys	Ser	Val
			260					265					270		
Tyr	Trp	His	Val	Ile	Gly	Met	Gly	Thr	Thr	Pro	Glu	Val	His	Ser	Ile
		275					280					285			
Phe	Leu	Glu	Gly	His	Thr	Phe	Leu	Val	Arg	Asn	His	Arg	Gln	Ala	Ser
290						295					300				
Leu	Glu	Ile	Ser	Pro	Ile	Thr	Phe	Leu	Thr	Ala	Gln	Thr	Leu	Leu	Met
305					310					315					320
Asp	Leu	Gly	Gln	Phe	Leu	Leu	Phe	Cys	His	Ile	Ser	Ser	His	Gln	His
				325					330						335
Asp	Gly	Met	Glu	Ala	Tyr	Val	Lys	Val	Asp	Ser	Cys	Pro	Glu	Glu	Pro
			340						345				350		
Gln	Leu	Arg	Met	Lys	Asn	Asn	Glu	Glu	Ala	Glu	Asp	Tyr	Asp	Asp	Asp
		355					360					365			
Leu	Thr	Asp	Ser	Glu	Met	Asp	Val	Val	Arg	Phe	Asp	Asp	Asp	Asn	Ser
370						375					380				
Pro	Ser	Phe	Ile	Gln	Ile	Arg	Ser	Val	Ala	Lys	Lys	His	Pro	Lys	Thr
385					390					395					400
Trp	Val	His	Tyr	Ile	Ala	Ala	Glu	Glu	Glu	Asp	Trp	Asp	Tyr	Ala	Pro
				405						410					415
Leu	Val	Leu	Ala	Pro	Asp	Asp	Arg	Ser	Tyr	Lys	Ser	Gln	Tyr	Leu	Asn
			420						425				430		
Asn	Gly	Pro	Gln	Arg	Ile	Gly	Arg	Lys	Tyr	Lys	Lys	Val	Arg	Phe	Met
		435					440					445			
Ala	Tyr	Thr	Asp	Glu	Thr	Phe	Lys	Thr	Arg	Glu	Ala	Ile	Gln	His	Glu
450						455					460				
Ser	Gly	Ile	Leu	Gly	Pro	Leu	Leu	Tyr	Gly	Glu	Val	Gly	Asp	Thr	Leu
465					470					475					480
Leu	Ile	Ile	Phe	Lys	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Asn	Ile	Tyr	Pro
				485					490						495
His	Gly	Ile	Thr	Asp	Val	Arg	Pro	Leu	Tyr	Ser	Arg	Arg	Leu	Pro	Lys
			500						505					510	
Gly	Val	Lys	His	Leu	Lys	Asp	Phe	Pro	Ile	Leu	Pro	Gly	Glu	Ile	Phe
			515					520					525		
Lys	Tyr	Lys	Trp	Thr	Val	Thr	Val	Glu	Asp	Gly	Pro	Thr	Lys	Ser	Asp
530						535					540				

-continued

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
 545 550 555 560
 Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
 565 570 575
 Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
 580 585 590
 Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
 595 600 605
 Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
 610 615 620
 Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
 625 630 635 640
 Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
 645 650 655
 Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
 660 665 670
 Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
 675 680 685
 Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
 690 695 700
 Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
 705 710 715 720
 Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
 725 730 735
 Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
 740 745 750
 Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu
 755 760 765
 Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln
 770 775 780
 Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu
 785 790 795 800
 Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe
 805 810 815
 Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp
 820 825 830
 Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln
 835 840 845
 Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr
 850 855 860
 Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His
 865 870 875 880
 Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile
 885 890 895
 Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser
 900 905 910
 Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
 915 920 925
 Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val
 930 935 940

-continued

Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp
 945 950 955 960

Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu
 965 970 975

Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His
 980 985 990

Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe
 995 1000 1005

Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn
 1010 1015 1020

Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys
 1025 1030 1035

Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr
 1040 1045 1050

Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr
 1055 1060 1065

Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His Phe
 1070 1075 1080

Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met
 1085 1090 1095

Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met
 1100 1105 1110

Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly
 1115 1120 1125

Glu His Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser
 1130 1135 1140

Asn Lys Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg
 1145 1150 1155

Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro
 1160 1165 1170

Lys Leu Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser
 1175 1180 1185

Thr Lys Glu Pro Phe Ser Trp Ile Lys Val Asp Leu Leu Ala Pro
 1190 1195 1200

Met Ile Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe
 1205 1210 1215

Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp
 1220 1225 1230

Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu
 1235 1240 1245

Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn
 1250 1255 1260

Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro
 1265 1270 1275

Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly
 1280 1285 1290

Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys
 1295 1300 1305

Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn
 1310 1315 1320

Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln

-continued

1325	1330	1335
Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu		
1340	1345	1350
Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val		
1355	1360	1365
Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys		
1370	1375	1380
Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu		
1385	1390	1395
Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp		
1400	1405	1410
Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr		
1415	1420	1425
Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile Ala		
1430	1435	1440
Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr		
1445	1450	1455

<210> SEQ ID NO 22
 <211> LENGTH: 2220
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 22

```

gccaccagga gatactaact gggcgccgtg gagctgagct gggactacat gcagctctgac    60
ctgggcgagc tgcctgtgga cgccagggtc cccccagag tgcccaagag cttccccttc    120
aacacctcag tgggtgtacaa gaagaccctg ttcgtggagt tcaccgacca cctgttcaac    180
atcgccaagc ccaggccccc ctggatgggc ctgctgggccc ccaccatcca ggccgaggtg    240
tacgacaccg tggatgatcac cctgaagaac atggccagcc acccctgag cctgcaagccc    300
gtgggctgga gctactggaa ggcctctgag ggcgcccagt atgacgacca gaccagccag    360
agggagaagg aggacgacaa ggtgttcccc ggcggcagcc acacctacgt gtggcaggtg    420
ctgaaggaga acgcccccat ggcacagcacc cccctgtgcc tgacctacag ctacctgagc    480
cacgtggacc tgggtgaagga cctgaactct ggcctgatcg gcgcccctgct ggtgtgcagg    540
gagggcagcc tggccaagga gaagaccacc agcctgcaca agttcctcct gctgttccgc    600
gtgttcgatg agggcaagag ctggcacagc gagaccaaga acagcctgat gcaggacagg    660
gatgccgccc ctgccagggc ctggcccaga atgcacaccg tgaacggcta cgtgaacagg    720
agcctgcccc gcctgatcgg ctgccacagg aagtctgtgt actggcacgt gatcggcctg    780
ggcaccaccc ccgaggtgca cagcatcttc ctggagggccc acaccttccc ggtgaggaac    840
cacaggcagg ccagcctgga gatcagcccc atcaccttcc tgaccgcccc gaccctgctg    900
atggacctgg gccagttccc gctgttctgc cacatcagca gccaccagca cgacggcctg    960
gaggcctacg tgaaggtgga cagctgcccc gaggagcccc agctgaggat gaagaacaac    1020
gaggaggccg aggactatga tgatgacctg accgactctg agatggacgt ggtgaggttt    1080
gatgatgaca acagccccag cttcatccag atcaggctctg tggccaagaa gcacccccaa    1140
acctgggtgc actacatcgc cgccgaggag gaggactggg actacgcccc cctggtgctg    1200
    
```

-continued

```

gccccgacg acaggagcta caagagccag tacctgaaca acggcccca gaggatcggc 1260
aggaagtaca agaaggtcag attcatggcc tacaccgacg agaccttcaa gaccagggag 1320
gcatccagc acgagtctgg catcctgggc cccctgctgt acggcgagggt gggcgacacc 1380
ctgctgatca tcttcaagaa ccaggccagc aggcctaca acatctaccc ccacggcatc 1440
accgatgtga ggcccctgta cagcaggagg ctgcccagg gcgtgaagca cctgaaggac 1500
tccccatcc tgcccggcga gatcttcaag tacaagtggg ccgtgaccgt ggaggatggc 1560
cccaccaagt ctgacccag gtgcctgacc aggtactaca gcagcttctg gaacatggag 1620
agggacctgg cctctggcct gatcggcccc ctgctgatct gctacaagga gagcgtggac 1680
cagaggggca accagatcat gtctgacaag aggaacgtga tcctgttctc tgtgttcgat 1740
gagaacagga gctggatctc gaccgagaac atccagaggt tcctgcccc ccccgccggc 1800
gtgcagctgg aggacccga gttccaggcc agcaacatca tgcacagcat caacggctac 1860
gtgttcgaca gcctcgagct gtctgtgtgc ctgcacgagg tggcctactg gtacatcctg 1920
agcatcggcg cccagaccga cttcctgtct gtgttcttct ctggctacac cttcaagcac 1980
aagatggtgt acgaggacac cctgacctg tccccctca gcggcgagac cgtgttcatg 2040
agcatggaga accccggcct gtggatcctg ggctgccaca acagcgactt caggaacagg 2100
ggcatgaccg cctcgtgtaa agtcagcagc tgcgacaaga acaccggcga ctactacgag 2160
gacagctacg aggacatcag cgcctacctg ctgagcaaga acaacgccat cgagcccagg 2220

```

<210> SEQ ID NO 23

<211> LENGTH: 2052

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 23

```

gagatcacca ggaccacct gcagagcgac caggaggaga tcgactatga tgacaccate 60
agcgtggaga tgaagaagga ggacttcgac atctacgacg aggacgagaa ccagagcccc 120
aggagcttcc agaagaagac caggcactac ttcctcgcgc ccgtggagag gctgtgggac 180
tatggcatga gcagcagccc ccacgtgctg aggaacaggg cccagagcgg cagcgtgccc 240
cagttcaaga aggtggtggt ccaggagttc accgacggca gcttcaccca gccctgtac 300
agaggcgagc tgaacgagca cctgggcctg ctgggcccc acatcagggc cgagggtggag 360
gacaacatca tggtgacctt caggaaccag gccagcaggc cctacagctt ctacagcagc 420
ctgatcagct acgaggagga ccagaggcag ggcgcccagc ccaggaagaa cttcgtgaag 480
cccaacgaga ccaagaccta cttctggaag gtgcagcacc acatggcccc caccaaggac 540
gagttcgact gcaaggcctg ggcctacttc tctgatgtgg acctggagaa ggacgtgcac 600
agcggcctga tcggccccct gctggtgtgc cacaccaaca ccctgaaccc cgcccacggc 660
aggcaggtga ccgtgcagga gttcgcctg ttcttcacca tcttcgacga gaccaagagc 720
tggtaactta ccgagaacat ggagaggaac tgcagggccc cctgcaacat ccagatggag 780
gacccacact tcaaggagaa ctacaggttc cacgccatca acggctacat catggacacc 840
ctgcccggcc tggatgatggc ccaggaccag aggatcaggt ggtatctgct gagcatgggc 900

```

-continued

```

agcaacgaga acatccacag catccacttc agcggccacg tgttcaccgt gaggaagaag 960
gaggagtaca agatggccct gtacaacctg taccocggcg tgttcogagac cgtggagatg 1020
ctgcccagca aggccggcat ctggaggggtg gagtgccctga tcggcgagca cctgcacgcc 1080
ggcatgagca ccctgttctt ggtgtacagc aacaagtgcc agacccccct gggcatggcc 1140
agcggccaca tcagggactt ccagatcacc gcctctggcc agtacggcca gtgggcccc 1200
aagctggcca ggctgcaacta cagcggcagc atcaacgcct ggagacacaa ggagcccttc 1260
agctggatca aggtggacct gctggcccc atgatcatcc acggcatcaa gaccagggc 1320
gccaggcaga agttcagcag cctgtacatc agccagttca tcatcatgta cagcctggac 1380
ggcaagaagt ggcagaccta caggggcaac agcaccggca ccctgatggt gttcttcggc 1440
aacgtggaca gcagcggcat caagcacaac atcttcaacc ccccatcat cgccaggta 1500
atcaggctgc accccacca ctacagcatc aggagcacc tcgggatgga actgatgggc 1560
tgcgacctga acagctgcag catgccctg ggcattgaga gcaaggccat ctctgacgcc 1620
cagatcacgg ccagcagcta cttaccaac atgttcgcca cctggagccc cagcaaggcc 1680
aggctgcacc tgcaggcag gagcaacgcc tggaggcccc aggtgaacaa cccaaggag 1740
tggctgcagg tggacttoca gaagaccatg aaggtgacgg gcgtgaccac ccaggcgtg 1800
aagagcctgc tgaccagcat gtacgtgaag gagttcctga tcagcagcag ccaggacggc 1860
caccagtgga ccctgttctt ccagaacggc aaagtgaagg tgttcagggg caaccaggac 1920
agcttcaccc ccgtggtgaa cagcctggac cccccctgc tgaccaggta tctgaggatc 1980
caccaccaga gctgggtgca ccagatcgcc ctgagaatgg aagtgctggg atgagaggcc 2040
caggacctgt ac 2052

```

<210> SEQ ID NO 24

<211> LENGTH: 2220

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 24

```

gccaccagga gatactacct gggggctgtg gaactttctt gggactacat gcagtctgac 60
ctgggagagc tgctgtgga tgccagggtc ccaccagag tgcccaagtc cttccattc 120
aacacctctg tggcttacia gaagacactc tttgtggaat tcaactgacca cctgttcaac 180
attgcaaac ccagaccacc ctggatggga ctccctggac ccaccattca ggctgaggtg 240
tatgacactg tggatcatc cctcaagaac atggcatccc accctgtgtc tctgcatgct 300
gtgggagtct catactggaa agcctctgaa ggggctgagt atgatgacca gacatcccag 360
agagagaaag aggatgacaa ggtgttcctt ggggatctc acacctatgt gtggcaagtc 420
ctcaaggaga atggaccat gccatctgac ccactctgcc tgacatactc ctacctttct 480
catgtggacc tggtaagga cctcaactct ggactgattg gggcactgct ggtgtgcagg 540
gaaggatccc tggccaagga gaaaaccag acaactgaca agttcattct cctgtttgct 600
gtctttgatg agggcaagtc ttggcactct gaaacaaaga actccctgat gcaagacagg 660
gatgctgct ctgccaggc atggcccaag atgcacactg tgaatggcta tgtgaacaga 720
tcaactgctg gactcattgg ctgccacag aaatctgtct actggcatgt gattggcatg 780

```

-continued

```

gggacaaccc ctgaagtgca ctccattttc ctggaggac acaccttctt ggtcaggaac 840
cacagacaag cctctctgga gatctctccc atcaccttcc tcaactgcaca gacactgctg 900
atggaccttg gacagtctct gctgttctgc cacatctctt cccaccagca tgatggcatg 960
gaagcctatg tcaaggtgga ctcatgccct gaggaaccac agctcaggat gaagaacaat 1020
gaggaggctg aggactatga tgatgaactg actgactctg agatggatgt ggtcagattt 1080
gatgatgaca actctccatc cttcattcag atcaggctcg tggcaaagaa acacccaag 1140
acatgggtgc actacattgc tctgaggaa gaggactggg actatgcacc actggctctg 1200
gcccctgatg acaggagcta caagtctcag tacctcaaca atggcccaca aagaattgga 1260
agaaagtaca agaaagttag attcatggcc tcaactgatg aaaccttcaa gacaagagaa 1320
gccattcagc atgagtctgg cattctggga ccaactctgt atggggaagt gggagacacc 1380
ctgctcatca tcttcaagaa ccaggcctcc aggccctaca acatctaccc acatggcatc 1440
actgatgtca gggccctgta cagcaggaga ctgccaaaag gggtgaaaca cctcaaggac 1500
ttccccattc tgcttgaga gatcttcaag tacaagtgga ctgtcactgt ggaggatgga 1560
ccaacaaagt ctgaccccag gtgcctcacc agatactact cctcttttgt gaacatggag 1620
agagacctgg catctggact gattggacca ctgctcatct gctacaagga gtctgtggac 1680
cagagaggca accagatcat gcttgacaag agaaatgtga ttctgttctc tgtctttgat 1740
gagaacagat catggtacct gactgagaac attcagagat tcctgcccac cctgtctggg 1800
gtgcaactgg aagacctga gttccaggca agcaacatca tgcactccat caatggctat 1860
gtgtttgact ctctccagct ttctgtctgc ctgcatgagg tggcctactg gtacattctt 1920
tctattgggg cacaaactga cttcctttct gtcttcttct ctggatacac cttcaagcac 1980
aagatggtgt atgaggacac cctgacactc ttcccattct ctggggaaac tgtgttcatg 2040
agcatggaga acctggact gtggattctg ggatgccaca actctgactt cagaaacagg 2100
ggaatgactg cactgctcaa agtctcctcc tgtgacaaga aactgggga ctactatgag 2160
gactcttatg aggacatctc tgccctactg ctcagcaaga acaatgccat tgagcccaga 2220

```

<210> SEQ ID NO 25

<211> LENGTH: 2052

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 25

```

gagatcacca ggacaaccct ccagtctgac caggaagaga ttgactatga tgacaccatt 60
tctgtggaga tgaagaagga ggactttgac atctatgatg aggacgagaa ccagtctcca 120
agatcattcc agaagaagac aagacactac ttcattgctg ctgtggaaaag actgtggggc 180
tatggcatgt cttcctctcc ccatgtctc aggaacaggg cacagtctgg ctctgtgcca 240
cagttcaaga aagtgtctt ccaggagttc actgatggct cattcaccca gccctgtac 300
agaggggaac tgaatgagca cctgggactc ctgggacat acatcagggc tgaggtggaa 360
gacaacatca tgggtgacatt cagaaaccag gctccaggc cctacagctt ctactcttcc 420
ctcatcagct atgaggaaga ccagagacaa ggggctgagc caagaaagaa ctttgtgaaa 480

```

-continued

```

cccaatgaaa ccaagaccta cttctggaaa gtccagcacc acatggcacc caccaaggat 540
gagtttgact gcaaggcctg ggcatacttc tctgatgtgg acctggagaa agatgtgcac 600
tctggcctga ttgcccact cctggctctgc cacaccaaca ccctgaaccc tgcacatgga 660
aggcaagtga ctgtgcagga gtttgcoctc ttcttcacca tctttgatga aaccaagtca 720
tggtaactca ctgagaacat ggagagaaac tgcagagcac catgcaacat tcagatggaa 780
gaccccacct tcaaggagaa ctacaggttc catgccatca atggctacat catggacacc 840
ctgcctgggc ttgtcatggc acaggaccag agaatcagat ggtacctgct ttctatggga 900
tccaatgaga acattcactc catccacttc tctgggcatg tcttcaactgt gagaaagaag 960
gaggaataca agatggccct gtacaacctc taccctgggg tctttgagac tgtggagatg 1020
ctgcctcca aagctggcat ctggaggggtg gaatgcoctc ttggggagca cctgcatgct 1080
ggcatgtcaa ccctgttcoct ggtctacagc aacaagtgcc agacaccctt gggaatggcc 1140
tctggccaca tcagggaact ccagatcaact gcctctggcc agtatggcca gtgggcacc 1200
aaactggcca ggctccacta ctctggctcc atcaatgcat ggtcaaccaa ggagccattc 1260
tcttggatca agtgggacct gctggcacc atgatcattc atggcatcaa gacacagggg 1320
gcaagacaga aattctctc tctgtacatc tcacagtcca tcatcatgta ctctctggat 1380
ggcaagaagt ggagacata cagaggcaac tccactggca ccctcatggt cttctttggc 1440
aatgtggaca gctctggcat caagcacaac atcttcaacc ctcccatcat tgccagatac 1500
atcaggctgc accccacca ctactcaatc agatcaacc tcaggatgga actgatggga 1560
tgtgacctga actcctgctc aatgcccctg ggaatggaga gcaaggccat ttctgatgcc 1620
cagatcactg catcctctta cttcaccaac atgtttgcca cctggtcacc atcaaaagcc 1680
aggetgcacc tccaggaag aagcaatgcc tggagacccc aggtcaacaa cccaaaggaa 1740
tggctgcaag tggacttcca gaagacaatg aaagtcaactg gggtagaac ccagggggctc 1800
aagtctctgc tcacctcaat gtatgtgaag gagtctctga tctctctc acaggatggc 1860
caccagtgga cactctctc ccagaatggc aaagtcaagg tgttccaggg caaccaggac 1920
tctttcacac ctgtggtgaa ctactggac cccccctcc tgacaagata cctgagaatt 1980
caccctcagt cttgggtcca ccagattgcc ctgagaatgg aagtcctggg atgtgaggca 2040
caagacctgt ac 2052

```

<210> SEQ ID NO 26

<211> LENGTH: 4332

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 26

```

atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60
accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120
ggagagctgc ctgtggatgc cagggtccca cccagagtgc ccaagtcctt cccattcaac 180
acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240
gcaaaaccca gaccacctg gatgggactc ctgggaccca ccattcaggc tgagggtgat 300
gacactgtgg tcatcacct caagaacatg gcatccacc ctgtgtctct gcatgctgtg 360

```


-continued

ggagtctcat	actggaaagc	ctctgaaggg	gctgagtatg	atgaccagac	atcccagaga	420
gagaaagagg	atgacaaggt	gttccctggg	ggatctcaca	cctatgtgtg	gcaagtctctc	480
aaggagaatg	gacccatggc	atctgacceca	ctctgcctga	catactccta	cctttctcat	540
gtggacctgg	tcaaggacct	caactctgga	ctgattgggg	cactgctggt	gtgcagggaa	600
ggatccctgg	ccaaggagaa	aaccagaca	ctgcacaagt	tcattctcct	gtttctgtc	660
tttgatgagg	gcaagtcttg	gcactctgaa	acaaagaact	ccctgatgca	agacagggat	720
gctgcctctg	ccagggcctg	gcccagatg	cacactgtga	atggctatgt	gaacagatca	780
ctgcctggac	tcattggctg	ccacaggaaa	tctgtctact	ggcatgtgat	tggcatgggg	840
acaacccctg	aagtgcactc	cattttcctg	gagggacaca	ccttctctgt	caggaaccac	900
agacaagcct	ctctggagat	ctctcccctc	accttctctc	ctgcacagac	actgctgatg	960
gaccttggac	agttcctgct	gttctgccc	atctcttccc	accagcatga	tggcatggaa	1020
gcctatgtca	agggtggactc	atgccctgag	gaaccacagc	tcaggatgaa	gaacaatgag	1080
gaggctgagg	actatgatga	tgacctgact	gactctgaga	tggatgtggt	cagatttgat	1140
gatgacaact	ctccatcctt	cattcagatc	aggtctgtgg	caaagaaaca	ccccagaca	1200
tgggtgcact	acattgtctg	tgaggaagag	gactgggact	atgcaccact	ggtcctggcc	1260
cctgatgaca	ggagctacaa	gtctcagtac	ctcaacaatg	gcccacaaag	aattggaaga	1320
aagtacaaga	aagtcagatt	catggcctac	actgatgaaa	ccttcaagac	aagagaagcc	1380
attcagcatg	agtctggcat	tctgggacca	ctcctgtatg	gggaagtggg	agacaccctg	1440
ctcatcatct	tcaagaacca	ggcctccagg	ccctacaaca	tctaccaca	tggcatcact	1500
gatgtcaggc	ccctgtacag	caggagactg	ccaaaagggg	tgaaacacct	caaggacttc	1560
cccattctgc	ctggagagat	cttcaagtac	aagtggactg	tcactgtgga	ggatggacca	1620
acaaagtctg	accccagggtg	cctcaccaga	tactactcct	cttttgtgaa	catggagaga	1680
gacctggcat	ctggactgat	tggaccactg	ctcatctgct	acaaggagtc	tgtggaccag	1740
agaggcaacc	agatcatgtc	tgacaagaga	aatgtgattc	tgttctctgt	ctttgatgag	1800
aacagatcat	ggtagctgac	tgagaacatt	cagagattcc	tgcccacccc	tgctggggtg	1860
caactggaag	accctgagtt	ccaggcaagc	aacatcatgc	actccatcaa	tggctatgtg	1920
tttgactctc	tccagcttct	tgtctgcctg	catgaggtgg	cctactggta	cattctttct	1980
attggggcac	aaactgactt	cctttctgtc	ttcttctctg	gatacacctt	caagcacaag	2040
atggtgtatg	aggacaccct	gacactcttc	ccattctctg	gggaaactgt	gttcatgagc	2100
atggagaacc	ctggactgtg	gattctggga	tgccacaact	ctgacttcag	aaacagggga	2160
atgactgcac	tgetcaaaagt	ctcctcctgt	gacaagaaca	ctgggggacta	ctatgaggac	2220
tcttatgagg	acatctctgc	ctacctgtc	agcaagaaca	atgccattga	gcccagagag	2280
atcaccagga	caaccctcca	gtctgaccag	gaagagattg	actatgatga	caccatttct	2340
gtggagatga	agaaggagga	ctttgacatc	tatgatgagg	acgagaacca	gtctccaaga	2400
tcattccaga	agaagacaag	acactacttc	attgctgctg	tggaaagact	gtgggactat	2460
ggcatgtctt	cctctcccaca	tgtcctcagg	aacagggcac	agctgtgctc	tgtgccacag	2520
ttcaagaaag	tggctctcca	ggagttcact	gatggctcat	tcacccagcc	cctgtacaga	2580
ggggaactga	atgagcaact	gggactcctg	ggaccataca	tcagggtctga	ggtggaagac	2640

-continued

```

aacatcatgg tgacattcag aaaccaggcc tccaggccct acagcttcta ctcttcctc 2700
atcagctatg aggaagacca gagacaaggg gctgagccaa gaaagaactt tgtgaaacct 2760
aatgaaacca agacctactt ctggaaagtc cagcaccaca tggcaccac caaggatgag 2820
tttgactgca aggctgggc atacttctct gatgtggacc tggagaaaga tgtgactct 2880
ggcctgatg gccactcct ggtctgccac accaacacct tgaacctgc acatggaagg 2940
caagtgactg tgcaggagtt tgcctcttc ttcaccatct ttgatgaaac caagtcatg 3000
tacttcactg agaacatgga gagaaactgc agagcaccat gcaacattca gatggaagac 3060
cccacctca aggagaacta caggttccat gccatcaatg gctacatcat ggacacctg 3120
cctgggcttg tcatggcaca ggaccagaga atcagatggt acctgcttc tatgggatcc 3180
aatgagaaca ttcactccat ccacttctct gggcatgtct tcaactgtgag aaagaaggag 3240
gaatacaaga tggccctgta caacctctac cctggggtct ttgagactgt ggagatgctg 3300
ccctccaaag ctggcatctg gaggggtgaa tgcctcattg gggagcacct gcatgctggc 3360
atgtcaacct tgttctctgt ctacagcaac aagtgcaga caccctggg aatggcctct 3420
ggccacatca gggacttcca gatcactgcc tctggccagt atggccagtg ggcacccaaa 3480
ctggccaggc tccactactc tggctccatc aatgcatggt caaccaagga gccattctct 3540
tggatcaagg tggacctgct ggcacctatg atcattcatg gcatcaagac acagggggca 3600
agacagaaat tctcctctct gtacatctca cagttcatca tcatgtactc tctggatggc 3660
aagaagtggc agacatacag aggcaactcc actggcacc ccatggtctt ctttggcaat 3720
gtggacagct ctggcatcaa gcacaacatc tcaaccctc ccatcattgc cagatacatc 3780
aggctgcacc ccacctacta ctcaatcaga tcaaccctca ggatggaact gatgggatgt 3840
gacctgaact cctgctcaat gccctggga atggagagca aggccatttc tgatgccag 3900
atcactgcat cctcttactt caccaacatg tttgccacct ggtcaccatc aaaagccagg 3960
ctgcacctcc aggaagaag caatgcctgg agacccagg tcaacaacct aaaggaatgg 4020
ctgcaagtgg acttcagaa gacaatgaaa gtcactggg tgacaacca gggggtcaag 4080
tctctgctca cctcaatgta tgtgaaggag ttcctgatct ctctctcaca ggatggccac 4140
cagtggacac tcttcttcca gaatggcaaa gtcaaggtgt tccagggcaa ccaggactct 4200
ttcacacctg tggtgaactc actggacccc cccctcctga caagatacct gagaattcac 4260
ccccagtctt gggccaacca gattgcctg agaatggaag tcctgggatg tgaggcacia 4320
gacctgtact ga 4332

```

<210> SEQ ID NO 27

<211> LENGTH: 4368

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 27

```

atgcagattg agctgtccac ctgcttctt ctgtgctgc tgagattctg cttctctgcc 60
accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120
ggagagctgc ctgtggatgc caggttccca cccagatgc ccaagtctt cccattcaac 180

```

-continued

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt	240
gcaaaaccca gaccacccctg gatgggactc ctgggaccca ccattcaggc tgagggtgat	300
gacactgtgg tcatcacocct caagaacatg gcatccacc ctgtgtctct gcatgctgtg	360
ggagtctcat actggaaagc ctctgaaggg gctgagatg atgaccagac atcccagaga	420
gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtctc	480
aaggagaatg gaccatggc atctgacca ctctgctga catactccta cctttctcat	540
gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggg gtgcagggaa	600
ggatccctgg ccaaggagaa aaccagaca ctgcacaagt tcattctcct gtttgctgtc	660
tttgatgagg gcaagtcttg gcaactgaa acaaagaact ccctgatgca agacagggat	720
gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacagatca	780
ctgctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg	840
acaacccctg aagtgcactc cttttctctg gagggacaca ccttctctgt caggaaccac	900
agacaagcct ctctggagat ctctccatc accttctca ctgcacagac actgctgatg	960
gaccttgac agttcctgct gttctgccac atctcttccc accagcatga tggcatggaa	1020
gcctatgtca aggtggactc atgcccctgag gaaccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggg cagatttgat	1140
gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca cccaagaca	1200
tgggtgcaact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc	1260
cctgatgaca ggagctacaa gtctcagtag ctcaacaatg gccacaaaag aattggaaga	1320
aagtacaaga aagtacagatt catggcctac actgatgaaa ccttcaagac aagagaagcc	1380
attcagcatg agtctggcat tctgggacca ctccctgatg gggaaagtgg agacaccctg	1440
ctcatcatct tcaagaacca ggctccagc ccctacaaca tctaccaca tggcatcact	1500
gatgtcaggc ccctgtacag caggagactg ccaaagggg tgaaacacct caaggacttc	1560
cccattctgc ctggagagat cttcaagtag aagtggactg tcaactgtgga ggatggacca	1620
acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga	1680
gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag	1740
agaggcaacc agatcatgct tgacaagaga aatgtgattc tgttctctgt ctttgatgag	1800
aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaacc tgctggggtg	1860
caactggaag acctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg	1920
tttgactctc tccagcttct tgtctgctg catgaggtgg cctactggta cattctttct	1980
attggggcac aaactgactt ctttctgtc ttcttctctg gatacacctt caagcacaag	2040
atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc	2100
atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga	2160
atgactgca tgetcaaagt ctctctctgt gacaagaaca ctggggacta ctatgaggac	2220
tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagaagc	2280
ttctctcaga attccagaca cccagcacc agggagatca ccaggacaac cctccagtct	2340
gaccaggaag agattgacta tgatgacacc atttctgtgg agatgaagaa ggaggacttt	2400
gacatctatg atgaggacga gaaccagtct ccaagatcat tccagaagaa gacaagacac	2460

-continued

tacttcattg ctgctgtgga aagactgtgg gactatggca tgtcttctc tccccatgtc	2520
ctcaggaaca gggcacagtc tggctctgtg ccacagttca agaaagtggc cttccaggag	2580
ttcactgatg gctcattcac ccagcccctg tacagagggg aactgaatga gcacctggga	2640
ctcctgggac catacatcag ggctgagggt gaagacaaca tcatggtgac attcagaaac	2700
caggcctcca ggcctacag cttctactct tccctcatca gctatgagga agaccagaga	2760
caaggggctg agccaagaaa gaactttgtg aaaccaatg aaaccaagac ctactctgg	2820
aaagtccagc accacatggc accccaag gatgagtttg actgcaaggc ctgggcatac	2880
ttctctgatg tggacctgga gaaagatgtg cactctggcc tgattggccc actcctggtc	2940
tgccacacca acaccctgaa cctgcacat ggaaggcaag tgactgtgca ggagtttgc	3000
ctcttcttca ccattcttga tgaaccaag tcatggtact tcatgagaa catggagaga	3060
aactgcagag caccatgcaa cattcagatg gaagaccca ccttcaagga gaactacagg	3120
ttccatgcca tcaatggcta catcatggac accctgctg ggcttgcac ggcacaggac	3180
cagagaatca gatggtacct gctttctatg ggatccaatg agaacattca ctccatccac	3240
ttctctgggc atgtcttccac tgtgagaaaag aaggaggaat acaagatggc cctgtacaac	3300
ctctaccctg gggctcttga gactgtggag atgctgccct ccaaagetgg catctggagg	3360
gtggaatgcc tcattgggga gcacctgcat gctggcatgt caacctgtt cctggtctac	3420
agcaacaagt gccagacacc cctgggaatg gcctctggcc acatcagga cttccagatc	3480
actgcctctg gccagtatgg ccagtgggca cccaaactgg ccaggctcca ctactctggc	3540
tccatcaatg catggtcaac caaggagcca ttctcttga tcaagggtga cctgctggca	3600
cccatgatca ttcatggcat caagacacag ggggcaagac agaaattctc ctctctgtac	3660
atctcacagt tcacatcat gtactctctg gatggcaaga agtggcagac atacagagge	3720
aactccactg gcaccctcat ggtcttctt ggcaatgtgg acagctctgg catcaagcac	3780
aacatcttca accctcccat cattgccaga tacatcagge tgcacccac ccactactca	3840
atcagatcaa ccctcaggat ggaactgatg ggatgtgacc tgaactcctg ctcaatgccc	3900
ctgggaatgg agagcaaggc catttctgat gccagatca ctgcatcctc ttacttacc	3960
aacatgtttg ccacctggtc accatcaaaa gccaggctgc acctccaggg aagaagcaat	4020
gcctggagac ccaggtcaa caaccctaaag gaatggctgc aagtggactt ccagaagaca	4080
atgaaagtca ctggggtgac aaccagggg gtcaagtctc tgctcacctc aatgatgtg	4140
aaggagttcc tgatctcttc ctacacaggat ggccaccagt ggacactctt cttccagaat	4200
ggcaaagtca aggtgttcca gggcaaccag gactcttca cacctgtggg gaactcactg	4260
gacccccccc tctgacaag atacctgaga attcaccctc agtcttgggt ccaccagatt	4320
gccttgagaa tggaagtcc gggatgtgag gcacaagacc tgtactga	4368

<210> SEQ ID NO 28

<211> LENGTH: 4332

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 28

atgcagattg agctgagcac ctgcttctc ctgtgctgac tgaggttctg cttctctgcc 60

-continued

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg	120
ggcgagctgc ctgtggaocg caggttcccc cccagagtgc ccaagagctt ccccttcaac	180
acctcagtgg tgtacaagaa gaccctgttc gtggagtcca ccgaccacct gttcaacatc	240
gccaagccca ggccccctg gatgggctg ctgggccccca ccatccagge cgaggtgtac	300
gacaccgtgg tgatcacct gaagaacatg gccagccacc ccgtgagcct gcacgacctg	360
ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg	420
gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg	480
aaggagaacg gccccatggc cagcgacccc ctgtgctga cctacagcta cctgagccac	540
gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag	600
ggcagcctgg ccaaggagaa gaccagacc ctgcacaagt tcatcctgct gttcgccgtg	660
ttcgatgagg gcaagagctg gcacagcag accaagaaca gcctgatgca ggacagggat	720
gccgcctctg ccagggctg gcccaagatg cacaccgtga acggctacgt gaacaggagc	780
ctgccccggc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc	840
accacccccg aggtgcacag catcttctg gagggccaca ccttctggt gaggaaccac	900
aggcaggcca gcctggagat cagccccatc accttctga ccgcccagac cctgctgatg	960
gacctgggccc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag	1020
gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag	1080
gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat	1140
gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc	1200
tgggtgcaact acatcgccg cgaggaggag gactgggact acgccccct ggtgctggcc	1260
cccagcagca ggagctacaa gagccagtac ctgaacaacg gccccagag gatcggcagg	1320
aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac caggagggcc	1380
atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacacctg	1440
ctgatcatct tcaagaacca ggccagcagg cctacaaca tctaccccc cggcatcacc	1500
gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc	1560
cccatcctgc ccggcgagat cttcaagtae aagtggaccg tgaccgtgga ggatggcccc	1620
accaagtctg accccagggt cctgaccagg tactacagca gcttctgtaa catggagagg	1680
gacctggcct ctggcctgat cggccccctg ctgatctgct acaaggagag cgtggaccag	1740
aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcogatgag	1800
aacaggagct ggtatctgac cgagaacatc cagagggtcc tgccccaccc cgccggcgtg	1860
cagctggagg acccccagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg	1920
ttcgacagcc tgcagctgtc tgtgtgctg cacgaggtgg cctactggta catcctgagc	1980
atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacct caagcacaag	2040
atggtgtacg aggacacct gaccctgttc cccctcagcg gcgagaccgt gttcatgagc	2100
atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc	2160
atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacaggac	2220
agctacgagg acatcagcgc ctacctgtg agcaagaaca acgcatcga gcccagggag	2280
atcaccagga ccacctgca gagcgaccag gaggagatcg actatgatga caccatcagc	2340

-continued

gtggagatga agaaggagga cttcgacatc tacgacgagg acgagaacca gagccccagg	2400
agcttccaga agaagaccag gcactacttc atcgccgccc tggagaggct gtgggactat	2460
ggcatgagca gcagccccca cgtgctgagg aacagggccc agagcggcag cgtgccccag	2520
ttcaagaagg tgggttcca ggagttcacc gacggcagct tccccagcc cctgtacaga	2580
ggcgagctga acgagcaact gggcctgctg gggccctaca tcagggccga ggtggaggac	2640
aacatcatgg tgacctcag gaaccaggcc agcaggccct acagcttcta cagcagcctg	2700
atcagctacg aggaggacca gaggcagggc gccgagccca ggaagaactt cgtgaagccc	2760
aacgagacca agacctactt ctggaagggt cagcaccaca tggccccac caaggacgag	2820
ttcgactgca aggccctggc ctacttctct gatgtggacc tggagaagga cgtgcacagc	2880
ggcctgatcg gccccctgct ggtgtgccac accaacaccc tgaaccccgc ccacggcagg	2940
caggtgacgg tgcaggagtt cgccctgttc ttcaccatct tcgacgagac caagagctgg	3000
tacttcaccg agaacatgga gaggaactgc agggccccct gcaacatcca gatggaggac	3060
cccacctca aggagaacta caggttccac gccatcaacg gctacatcat ggacacctg	3120
cccggcctgg tgatggccca ggaccagagg atcaggtggt atctgctgag catgggcagc	3180
aacgagaaca tccacagcat ccacttcagc gggcacgtgt tcaccgtgag gaagaaggag	3240
gagtacaaga tggccctgta caacctgtac cccggcgtgt tcgagaccgt ggagatgctg	3300
cccagcaagg ccggcatctg gaggtggag tgcctgatcg gcgagcact gcacgcccgc	3360
atgagcacc tgttccctggt gtacagcaac aagtgccaga cccccctggg catggccagc	3420
ggccacatca gggacttcca gatcacggcc tctggccagt acggccagtg gggccccaa	3480
ctggccaggc tgcactacag cggcagcatc aacgcctgga gcaccaagga gcccttcagc	3540
tggatcaagg tggacctgct gggccccatg atcatccacg gcatacaagc ccagggcgcc	3600
aggcagaagt tcagcagcct gtacatcagc cagttcatca tcatgtacag cctggacggc	3660
aagaagtggc agacctacag gggcaacagc accggcacc cccctggtgt cttcggaac	3720
gtggacagca gcggcatcaa gcacaacatc tcaaccccc ccatcatcgc caggatcatc	3780
aggetgcacc ccaacctacta cagcatcagg agcaccctgc ggatggaact gatgggctgc	3840
gacctgaaca gctgcagcat gccctgggc atggagagca aggcoactc tgaogcccag	3900
atcacgccca gcagctactt caccaacatg ttcgccact ggagccccag caaggccagg	3960
ctgcacctgc agggcaggag caacgcctgg agggccccag tgaacaaccc caaggagtgg	4020
ctgcaggctg acttccagaa gaccatgaag gtgaccggcg tgaccacca gggcgtgaag	4080
agcctgctga ccagcatgta cgtgaaggag ttctgatca gcagcagcca ggacggccac	4140
cagtggaacc tgttcttcca gaacggcaaa gtgaagggtt tccagggcaa ccaggacagc	4200
ttccccccg tggatgaacag cctggacccc cccctgctga ccaggtatct gaggatccac	4260
cccagagct ggggtcacca gatcgccctg agaatggaag tgctgggatg cgaggcccag	4320
gacctgtact ga	4332

<210> SEQ ID NO 29

<211> LENGTH: 4368

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

-continued

polynucleotide

<400> SEQUENCE: 29

atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc	60
accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg	120
ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac	180
acctcagtgg tgtacaagaa gacctgttc gtggagtcca cggaccacct gttcaacatc	240
gccaaagcca ggccccctg gatgggctg ctgggccccca ccatccaggc cgagggtgac	300
gacacctggt tgatcacctt gaagaacatg gccagcccacc ccgtgagcct gcacgccgtg	360
ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg	420
gagaaggagg acgacaaggt gttccccggc gccagcccaca cctacgtgtg gcagggtgctg	480
aaggagaacg gccccatggc cagcagcccc ctgtgcctga cctacagcta cctgagccac	540
gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag	600
ggcagcctgg ccaaggagaa gacctagacc ctgcacaagt tcatcctgct gttcgccgtg	660
ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat	720
gccgcctctg ccagggcctg gcccagaatg cacaccgtga acggctacgt gaacaggagc	780
ctgccccggc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc	840
accacccccg aggtgcacag catcttctctg gagggccaca ccttctctgt gaggaaccac	900
aggcaggcca gcctggagat cagccccatc accttctga ccgcccagac cctgctgatg	960
gacctgggcc agttcctgct gttctgccc acacagcacc accagcacga cggcatggag	1020
gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag	1080
gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat	1140
gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca cccaagacc	1200
tgggtgcact acatcgccgc cgaggaggag gactgggact acgccccctt ggtgctggcc	1260
cccagcgaca ggaagctaaa gagccagtac ctgaacaacg gccccagag gatcggcagg	1320
aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc	1380
atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg	1440
ctgatcatct tcaagaacca ggccagcagg cctacaaca tctaccccc cggcatcacc	1500
gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc	1560
cccacctctg ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc	1620
accaagcttg accccagggt cctgaccagg tactacagca gcttctgtaa catggagagg	1680
gacctggcct ctggcctgat cgccccctg ctgatctgct acaaggagag cgtggaccag	1740
aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcogatgag	1800
aacaggagct ggtatctgac cgagaacatc cagaggttcc tgccccaccc cgcggcgtg	1860
cagctggagg acccccagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg	1920
ttcgacagcc tgcagctgtc tgtgtgctg caccagggtg cctactggta catcctgagc	1980
atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag	2040
atggtgtacg aggacacctt gacctgttc ccttcagcg gcgagaccgt gttcatgagc	2100
atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc	2160

-continued

atgaccgccc	tgctgaaagt	cagcagctgc	gacaagaaca	ccggcgacta	ctacgaggac	2220
agctacgagg	acatcagcgc	ctacctgctg	agcaagaaca	acgccatcga	gcccaggagc	2280
ttcagccaga	actccagaca	ccccagcacc	agggagatca	ccaggaccac	cctgcagagc	2340
gaccaggagg	agatcgacta	tgatgacacc	atcagcgtgg	agatgaagaa	ggaggacttc	2400
gacatctacg	acgaggacga	gaaccagagc	cccaggagct	tccagaagaa	gaccaggcac	2460
tacttcatcg	ccgccgtgga	gaggctgtgg	gactatggca	tgagcagcag	ccccacgtg	2520
ctgaggaaca	gggcccagag	cggcagcgtg	ccccagtcca	agaaggtggg	gttccaggag	2580
ttcaccgacg	gcagcttcac	ccagcccctg	tacagaggcg	agctgaacga	gcacctgggc	2640
ctgctggggc	cctacatcag	ggccgaggtg	gaggacaaca	tcatggtgac	cttcaggaac	2700
caggccagca	ggccctacag	cttctacagc	agcctgatca	gctacgagga	ggaccagagg	2760
cagggcgccg	agcccaggaa	gaacttcgtg	aagcccaacg	agaccaagac	ctacttctgg	2820
aaggtgcagc	accacatggc	ccccaccaag	gacgagttcg	actgcaaggg	ctgggcctac	2880
ttctctgatg	tggacctgga	gaaggacgtg	cacagcggcc	tgatcgggcc	cctgctggtg	2940
tgccacacca	acaccctgaa	ccccgcccac	ggcaggcagg	tgaccgtgca	ggagtctgcc	3000
ctgttcttca	ccatcttoga	cgagaccaag	agctgggtact	tcaccgagaa	catggagagg	3060
aactgcaggg	ccccctgcaa	catccagatg	gaggacccca	ccttcaagga	gaactacagg	3120
ttccacgcca	tcaacggcta	catcatggac	accctgcccc	gcctgggtgat	ggcccaggac	3180
cagaggatca	gggtggtatc	gctgagcatg	ggcagcaacg	agaacatcca	cagcatccac	3240
ttcagcggcc	acgtgttcac	cgtgaggaag	aaggaggagt	acaagatggc	cctgtacaac	3300
ctgtaccctg	gcgtgttoga	gaccgtggag	atgctgcccc	gcaaggccgg	catctggagg	3360
gtggagtgcc	tgatcggcga	gcacctgcac	gcccgcacga	gcacctggt	cctgggtgac	3420
agcaacaagt	gccagacccc	cctgggcatg	gccagcggcc	acatcaggga	cttccagatc	3480
accgcctctg	gccagtacgg	ccagtggggc	cccaagctgg	ccaggctgca	ctacagcggc	3540
agcatcaacg	cctggagcac	caaggagccc	ttcagctgga	tcaaggtgga	cctgctggcc	3600
cccatgatca	tccacggcat	caagaccag	ggcgccaggc	agaagttcag	cagcctgtac	3660
atcagccagt	tcatcatcat	gtacagcctg	gacggcaaga	agtggcagac	ctacaggggc	3720
aacagcaccg	gcaccctgat	ggtgttcttc	ggcaacgtgg	acagcagcgg	catcaagcac	3780
aacatcttca	acccccccat	catcgccagg	tacatcaggc	tgcaccccac	ccactacagc	3840
atcaggagca	ccctgcggat	ggaactgatg	ggctgcgacc	tgaacagctg	cagcatgccc	3900
ctgggcatgg	agagcaaggg	catctctgac	gcccagatca	ccgccagcag	ctacttcacc	3960
aacatgttgc	ccacctggag	ccccagcaag	gccaggctgc	acctgcaggg	caggagcaac	4020
gcctggaggc	cccaggtgaa	caaccccagg	gagtggtctg	aggtggactt	ccagaagacc	4080
atgaaggtga	ccggcgtgac	cacccagggc	gtgaagagcc	tgctgaccag	catgtacgtg	4140
aaggagtcc	tgatcagcag	cagccaggac	ggccaccagt	ggacctggt	cttccagaac	4200
ggcaaagtga	aggtgttcca	gggcaaccag	gacagcttca	ccccctggg	gaacagcctg	4260
gacccccccc	tgctgaccag	gtatctgagg	atccaccccc	agagctgggt	gcaccagatc	4320
gccctgagaa	tggaagtgct	gggatgcgag	gcccaggacc	tgtactga		4368

-continued

<211> LENGTH: 14
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 30

Ser Phe Ser Gln Asn Pro Pro Val Leu Lys Arg His Gln Arg
 1 5 10

<210> SEQ ID NO 31
 <211> LENGTH: 24
 <212> TYPE: PRT
 <213> ORGANISM: Sus sp.

<400> SEQUENCE: 31

Ser Phe Ala Gln Asn Ser Arg Pro Pro Ser Ala Ser Ala Pro Lys Pro
 1 5 10 15

Pro Val Leu Arg Arg His Gln Arg
 20

<210> SEQ ID NO 32
 <211> LENGTH: 16
 <212> TYPE: PRT
 <213> ORGANISM: Sus sp.

<400> SEQUENCE: 32

Ser Phe Ser Gln Asn Ser Arg His Gln Ala Tyr Arg Tyr Arg Arg Gly
 1 5 10 15

<210> SEQ ID NO 33
 <211> LENGTH: 8
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 33

Thr Thr Tyr Val Asn Arg Ser Leu
 1 5

<210> SEQ ID NO 34
 <211> LENGTH: 7
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 34

Pro Gln Leu Arg Met Lys Asn
 1 5

<210> SEQ ID NO 35
 <211> LENGTH: 7
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 35

Val Asp Gln Arg Gly Asn Gln
 1 5

-continued

<210> SEQ ID NO 36
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 36

agcttcagcc agaatgtgag caacaatgtg agcaacaatg ccaccaataa tgctaccaac 60
ccacctgtcc tgaaacgccca ccagagg 87

<210> SEQ ID NO 37
<211> LENGTH: 75
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 37

agcttcagcc agaatgtgag caacaatgcc accaacaatg tgagcaaccc acctgtcctg 60
aaacgccacc agagg 75

<210> SEQ ID NO 38
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 38

agcttcagcc agaatgtgag caataatgcc accaaccacac ctgtcctgaa acgccaccag 60
agg 63

<210> SEQ ID NO 39
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 39

agcttcagcc agaatgtgag caataatcca cctgtcctga aagccacca gagg 54

<210> SEQ ID NO 40
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 40

agcttcagcc agaataggag cctgccacct gtccctgaaac gccaccagag g 51

<210> SEQ ID NO 41
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 41

agcttcagcc agaatgccac taatgtgtct aacaactctg ctacctctgc tgactctgct 60
gtgagcccac ctgtctgaa acgccaccag agg 93

<210> SEQ ID NO 42
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 42

agcttcagcc agaatgccac caactatgtg aacaggagcc tgccacctgt cctgaaacgc 60
caccagagg 69

<210> SEQ ID NO 43
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<400> SEQUENCE: 43

agcttcagcc agaatgccac caactatgtg aacaggagcc tgtctgccac ctctgtgac 60
tctgtgtgga gccagaatcc acctgtctctg aaacgccacc agagg 105

<210> SEQ ID NO 44
<211> LENGTH: 84
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 44

agcttcagcc agaatgtgag caacaatgtg agcaatgctg tgtctgctgt gtctgtctca 60
cctgtctctga aacgccacca gagg 84

<210> SEQ ID NO 45
<211> LENGTH: 90
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 45

agcttcagcc agaatatcac tgtggcctct gccaccteta acatcaactgt ggctctgct 60
gaccacctg tcctgaaacg ccaccagagg 90

<210> SEQ ID NO 46
<211> LENGTH: 72
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 46

agcttcagcc agaatatcac tgtgaccaac atcactgtga ctgccccacc tgtcctgaaa 60

cgccaccaga gg 72

<210> SEQ ID NO 47

<211> LENGTH: 72

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 47

agcttcagcc agaatcagac tgtgaccaac atcactgtga ctgccccacc tgtcctgaaa 60

cgccaccaga gg 72

<210> SEQ ID NO 48

<211> LENGTH: 93

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 48

agcttcagcc agaatgccac taatgtgtct aacaacagca acaccagcaa tgacagcaat 60

gtgtctccac ctgtcctgaa acgccaccag agg 93

<210> SEQ ID NO 49

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<400> SEQUENCE: 49

atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccacctg gatgggactc ctgggaccca ccattcaggc tgaggtgat 300

gacactgtgg tcatcaccct caagaacatg gcatccccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacctatggc atctgaccca ctctgctga catactccta cttttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aaccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gtgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacagatca 780

-continued

ctgcctggac	tcattggctg	ccacaggaaa	tctgtctact	ggcatgtgat	tggcatgggg	840
acaacccctg	aagtgcactc	cattttcctg	gagggacaca	ccttctcgtt	caggaaccac	900
agacaagcct	ctctggagat	ctctccctc	accttctca	ctgcacagac	actgctgatg	960
gaccttggac	agttcctgct	gtcctgccac	atctcttccc	accagcatga	tggcatggaa	1020
gcctatgtca	agggtggactc	atgccctgag	gaaccacagc	tcaggatgaa	gaacaatgag	1080
gaggctgagg	actatgatga	tgacctgact	gactctgaga	tggatgtggt	cagatttgat	1140
gatgacaact	ctccatcctt	cattcagatc	aggctctgtg	caaagaaaca	ccccaaagaca	1200
tgggtgact	acattgtctg	tgaggaagag	gactgggact	atgcaccact	ggtcctggcc	1260
cctgatgaca	ggagctacaa	gtctcagtac	ctcaacaatg	gccacaaaag	aattggaaga	1320
aagtacaaga	aagtcagatt	catggcctac	actgatgaaa	ccttcaagac	aagagaagcc	1380
attcagcatg	agtctggcat	tctgggacca	ctcctgtatg	gggaagtggg	agacaccctg	1440
ctcatcatct	tcaagaacca	ggcctccagg	ccctacaaca	tctaccaca	tggcatcact	1500
gatgtcaggc	ccctgtacag	caggagactg	ccaaaagggg	tgaaacacct	caaggacttc	1560
cccattctgc	ctggagagat	cttcaagtac	aagtggactg	tactgtgga	ggatggacca	1620
acaaagtctg	accccagggt	cctcaccaga	tactactcct	ctttgtgaa	catggagaga	1680
gacctggcat	ctggactgat	tggaccactg	ctcatctgct	acaaggagtc	tgtggaccag	1740
agaggcaacc	agatcatgtc	tgacaagaga	aatgtgattc	tgttctctgt	ctttgatgag	1800
aacagatcat	ggtagctgac	tgagaacatt	cagagattcc	tgcccaaccc	tgctggggtg	1860
caactggaag	accctgagtt	ccaggcaagc	aacatcatgc	actccatcaa	tggctatgtg	1920
tttgactctc	tccagcttct	tgtctgcctg	catgaggtgg	cctactggta	cattctttct	1980
atggggcac	aaactgactt	cctttctgct	ttcttctctg	gatacacctt	caagcacaag	2040
atggtgtatg	aggacacctt	gacactcttc	ccattctctg	gggaaactgt	gttcatgagc	2100
atggagaacc	ctggactgtg	gattctggga	tgccacaact	ctgacttcag	aaacagggga	2160
atgactgcac	tgtctaaagt	ctcctcctgt	gacaagaaca	ctggggacta	ctatgaggac	2220
tcttatgagg	acatctctgc	ctacctgctc	agcaagaaca	atgccattga	gccagaagc	2280
ttctctcaga	atccacctgt	cctgaagaga	caccagagag	agatcaccag	gacaaccctc	2340
cagtctgacc	aggaagagat	tgactatgat	gacaccattt	ctgtggagat	gaagaaggag	2400
gactttgaca	tctatgatga	ggacgagaac	cagtctccaa	gatcattcca	gaagaagaca	2460
agacactact	tcattgtctg	tgtggaagaa	ctgtgggact	atggcatgtc	ttctctctcc	2520
catgtcctca	ggaacagggc	acagtctggc	tctgtgccac	agttcaagaa	agtggctctc	2580
caggagtcca	ctgatggctc	atccaccag	cccctgtaca	gaggggaact	gaatgagcac	2640
ctgggactcc	tgggaccata	catcagggct	gaggtggaag	acaacatcat	ggtgacattc	2700
agaaaccagg	cctccaggcc	ctacagcttc	tactctctcc	tcatacagcta	tgaggaagac	2760
cagagacaag	gggctgagcc	aagaaagaac	tttgtgaaac	ccaatgaaac	caagacctac	2820
ttctggaaag	tccagcacca	catggcacc	accaaggatg	agtttgactg	caaggcctgg	2880
gcatacttct	ctgatgtgga	cctggagaaa	gatgtgcact	ctggcctgat	tggccactc	2940
ctggtctgcc	acaccaacac	cctgaaccct	gcacatggaa	ggcaagtgac	tgtgcaggag	3000
tttgccctct	tcttcacat	ctttgatgaa	accaagtcac	ggtacttcac	tgagaacatg	3060

-continued

```

gagagaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagAAC 3120
tacagggtcc atgccatcaa tggctacatc atggaccccc tgccctgggt tgcctatggca 3180
caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cattcactcc 3240
atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggcctg 3300
tacaacctct accctgggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360
tggagggtgg aatgcctcat tggggagcac ctgcatgctg gcatgtcaac cctgttctctg 3420
gtctacagca acaagtgcc aacacctctg ggaatggcct ctggccacat cagggaactc 3480
cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac 3540
tctggctcca tcaatgcatg gtaaccaag gagccattct cttggatcaa ggtggacctg 3600
ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctcctct 3660
ctgtacatct cacagtccat catcatgtac tctctggatg gcaagaagtg gcagacatac 3720
agaggcaact ccaactggcac cctcatggtc ttctttggca atgtggacag ctctggcatc 3780
aagcacaaca tcttcaacc tccatcatt gccagataca tcaggctgca cccaccacc 3840
tactcaatca gatcaacct caggatggaa ctgatgggat gtgacctgaa ctctctctca 3900
atgccctctg gaatggagag caaggccatt tctgatgccc agatcactgc atctcttac 3960
ttcaccaca tgtttggccac ctggtcacca tcaaaagcca ggctgcacct ccaggaaga 4020
agcaatgctt ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttcag 4080
aagacaatga aagtcactgg ggtgacaacc caggggggtca agtctctgct cacctcaatg 4140
tatgtgaagg agttcctgat ctcttctca caggatggcc accagtggac actcttcttc 4200
cagaatggca aagtaagggt gttccagggc aaccaggact ctttcacacc tgtggtgAAC 4260
tcaactggacc cccccctct gacaagatac ctgagaattc acccccagtc ttgggtccac 4320
cagattgccc tgagaatgga agtcctggga tgtgaggcac aagacctgta ctga 4374

```

```

<210> SEQ ID NO 50
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(45)

```

```

<400> SEQUENCE: 50

```

```

gtg agc aac aat gtg agc aac aat gcc acc aat aat gct acc aac 45
Val Ser Asn Asn Val Ser Asn Asn Ala Thr Asn Asn Ala Thr Asn
1 5 10 15

```

```

<210> SEQ ID NO 51
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

```

```

<400> SEQUENCE: 51

```

```

Val Ser Asn Asn Val Ser Asn Asn Ala Thr Asn Asn Ala Thr Asn
1 5 10 15

```

-continued

<210> SEQ ID NO 52
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(33)

<400> SEQUENCE: 52

gtg agc aac aat gcc acc aac aat gtg agc aac 33
Val Ser Asn Asn Ala Thr Asn Asn Val Ser Asn
1 5 10

<210> SEQ ID NO 53
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 53

Val Ser Asn Asn Ala Thr Asn Asn Val Ser Asn
1 5 10

<210> SEQ ID NO 54
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(21)

<400> SEQUENCE: 54

gtg agc aat aat gcc acc aac 21
Val Ser Asn Asn Ala Thr Asn
1 5

<210> SEQ ID NO 55
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 55

Val Ser Asn Asn Ala Thr Asn
1 5

<210> SEQ ID NO 56
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(12)

-continued

<400> SEQUENCE: 56

gtg agc aat aat 12
 Val Ser Asn Asn
 1

<210> SEQ ID NO 57

<211> LENGTH: 4

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 57

Val Ser Asn Asn
 1

<210> SEQ ID NO 58

<211> LENGTH: 9

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (1)..(9)

<400> SEQUENCE: 58

agg agc ctg 9
 Arg Ser Leu
 1

<210> SEQ ID NO 59

<211> LENGTH: 3

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 59

Arg Ser Leu
 1

<210> SEQ ID NO 60

<211> LENGTH: 51

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (1)..(51)

<400> SEQUENCE: 60

gcc act aat gtg tct aac aac tct gct acc tct gct gac tct gct gtg 48
 Ala Thr Asn Val Ser Asn Asn Ser Ala Thr Ser Ala Asp Ser Ala Val
 1 5 10 15

agc 51
 Ser

<210> SEQ ID NO 61

<211> LENGTH: 17

<212> TYPE: PRT

-continued

```

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide

<400> SEQUENCE: 61

Ala Thr Asn Val Ser Asn Asn Ser Ala Thr Ser Ala Asp Ser Ala Val
1           5           10           15

Ser

<210> SEQ ID NO 62
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(27)

<400> SEQUENCE: 62

gcc acc aac tat gtg aac agg agc ctg                27
Ala Thr Asn Tyr Val Asn Arg Ser Leu
1           5

<210> SEQ ID NO 63
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide

<400> SEQUENCE: 63

Ala Thr Asn Tyr Val Asn Arg Ser Leu
1           5

<210> SEQ ID NO 64
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(63)

<400> SEQUENCE: 64

gcc acc aac tat gtg aac agg agc ctg tct gcc acc tct gct gac tct    48
Ala Thr Asn Tyr Val Asn Arg Ser Leu Ser Ala Thr Ser Ala Asp Ser
1           5           10           15

gct gtg agc cag aat                63
Ala Val Ser Gln Asn
                20

<210> SEQ ID NO 65
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide

<400> SEQUENCE: 65

```

-continued

Ala Thr Asn Tyr Val Asn Arg Ser Leu Ser Ala Thr Ser Ala Asp Ser
1 5 10 15

Ala Val Ser Gln Asn
20

<210> SEQ ID NO 66
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(42)

<400> SEQUENCE: 66

gtg agc aac aat gtg agc aat gct gtg tct gct gtg tct gct 42
Val Ser Asn Asn Val Ser Asn Ala Val Ser Ala Val Ser Ala
1 5 10

<210> SEQ ID NO 67
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 67

Val Ser Asn Asn Val Ser Asn Ala Val Ser Ala Val Ser Ala
1 5 10

<210> SEQ ID NO 68
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(48)

<400> SEQUENCE: 68

atc act gtg gcc tct gcc acc tct aac atc act gtg gcc tct gct gac 48
Ile Thr Val Ala Ser Ala Thr Ser Asn Ile Thr Val Ala Ser Ala Asp
1 5 10 15

<210> SEQ ID NO 69
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 69

Ile Thr Val Ala Ser Ala Thr Ser Asn Ile Thr Val Ala Ser Ala Asp
1 5 10 15

<210> SEQ ID NO 70
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

```

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(30)

<400> SEQUENCE: 70

atc act gtg acc aac atc act gtg act gcc          30
Ile Thr Val Thr Asn Ile Thr Val Thr Ala
1              5              10

<210> SEQ ID NO 71
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide

<400> SEQUENCE: 71

Ile Thr Val Thr Asn Ile Thr Val Thr Ala
1              5              10

<210> SEQ ID NO 72
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(30)

<400> SEQUENCE: 72

cag act gtg acc aac atc act gtg act gcc          30
Gln Thr Val Thr Asn Ile Thr Val Thr Ala
1              5              10

<210> SEQ ID NO 73
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      peptide

<400> SEQUENCE: 73

Gln Thr Val Thr Asn Ile Thr Val Thr Ala
1              5              10

<210> SEQ ID NO 74
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(51)

<400> SEQUENCE: 74

gcc act aat gtg tct aac aac agc aac acc agc aat gac agc aat gtg          48
Ala Thr Asn Val Ser Asn Asn Ser Asn Thr Ser Asn Asp Ser Asn Val
1              5              10              15

```

-continued

tct 51
Ser

<210> SEQ ID NO 75
<211> LENGTH: 17
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 75

Ala Thr Asn Val Ser Asn Asn Ser Asn Thr Ser Asn Asp Ser Asn Val
1 5 10 15

Ser

<210> SEQ ID NO 76
<211> LENGTH: 405
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 76

Met Pro Leu Leu Leu Tyr Thr Cys Leu Leu Trp Leu Pro Thr Ser Gly
1 5 10 15

Leu Trp Thr Val Gln Ala Met Asp Pro Asn Ala Ala Tyr Val Asn Met
20 25 30

Ser Asn His His Arg Gly Leu Ala Ser Ala Asn Val Asp Phe Ala Phe
35 40 45

Ser Leu Tyr Lys His Leu Val Ala Leu Ser Pro Lys Lys Asn Ile Phe
50 55 60

Ile Ser Pro Val Ser Ile Ser Met Ala Leu Ala Met Leu Ser Leu Gly
65 70 75 80

Thr Cys Gly His Thr Arg Ala Gln Leu Leu Gln Gly Leu Gly Phe Asn
85 90 95

Leu Thr Glu Arg Ser Glu Thr Glu Ile His Gln Gly Phe Gln His Leu
100 105 110

His Gln Leu Phe Ala Lys Ser Asp Thr Ser Leu Glu Met Thr Met Gly
115 120 125

Asn Ala Leu Phe Leu Asp Gly Ser Leu Glu Leu Leu Glu Ser Phe Ser
130 135 140

Ala Asp Ile Lys His Tyr Tyr Glu Ser Glu Val Leu Ala Met Asn Phe
145 150 155 160

Gln Asp Trp Ala Thr Ala Ser Arg Gln Ile Asn Ser Tyr Val Lys Asn
165 170 175

Lys Thr Gln Gly Lys Ile Val Asp Leu Phe Ser Gly Leu Asp Ser Pro
180 185 190

Ala Ile Leu Val Leu Val Asn Tyr Ile Phe Phe Lys Gly Thr Trp Thr
195 200 205

Gln Pro Phe Asp Leu Ala Ser Thr Arg Glu Glu Asn Phe Tyr Val Asp
210 215 220

Glu Thr Thr Val Val Lys Val Pro Met Met Leu Gln Ser Ser Thr Ile
225 230 235 240

Ser Tyr Leu His Asp Ser Glu Leu Pro Cys Gln Leu Val Gln Met Asn
245 250 255

Tyr Val Gly Asn Gly Thr Val Phe Phe Ile Leu Pro Asp Lys Gly Lys

-continued

Met	Asn	Thr	Val	Ile	Ala	Ala	Leu	Ser	Arg	Asp	Thr	Ile	Asn	Arg	Trp
	275						280					285			
Ser	Ala	Gly	Leu	Thr	Ser	Ser	Gln	Val	Asp	Leu	Tyr	Ile	Pro	Lys	Val
	290					295					300				
Thr	Ile	Ser	Gly	Val	Tyr	Asp	Leu	Gly	Asp	Val	Leu	Glu	Glu	Met	Gly
305					310					315					320
Ile	Ala	Asp	Leu	Phe	Thr	Asn	Gln	Ala	Asn	Phe	Ser	Arg	Ile	Thr	Gln
				325					330						335
Asp	Ala	Gln	Leu	Lys	Ser	Ser	Lys	Val	Val	His	Lys	Ala	Val	Leu	Gln
		340						345						350	
Leu	Asn	Glu	Glu	Gly	Val	Asp	Thr	Ala	Gly	Ser	Thr	Gly	Val	Thr	Leu
	355						360						365		
Asn	Leu	Thr	Ser	Lys	Pro	Ile	Ile	Leu	Arg	Phe	Asn	Gln	Pro	Phe	Ile
	370					375						380			
Ile	Met	Ile	Phe	Asp	His	Phe	Thr	Trp	Ser	Ser	Leu	Phe	Leu	Ala	Arg
385					390						395				400
Val	Met	Asn	Pro	Val											
				405											

<210> SEQ ID NO 77
 <211> LENGTH: 4
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 77

Asn Met Ser Asn
 1

<210> SEQ ID NO 78
 <211> LENGTH: 4
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 78

Asn Leu Thr Glu
 1

<210> SEQ ID NO 79
 <211> LENGTH: 4
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 79

Asn Lys Thr Gln
 1

<210> SEQ ID NO 80
 <211> LENGTH: 4
 <212> TYPE: PRT
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 80

Asn Gly Thr Val
 1

<210> SEQ ID NO 81
 <211> LENGTH: 4

-continued

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 81

Asn Phe Ser Arg

1

<210> SEQ ID NO 82

<211> LENGTH: 4

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 82

Asn Leu Thr Ser

1

<210> SEQ ID NO 83

<211> LENGTH: 41

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 83

Leu Ser Lys Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Ala
1 5 10 15

Thr Asn Val Ser Asn Asn Ser Asn Thr Ser Asn Asp Ser Asn Val Ser
20 25 30

Pro Pro Val Leu Lys Arg His Gln Arg
35 40

<210> SEQ ID NO 84

<211> LENGTH: 4

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 84

Asn Ala Thr Asn

1

<210> SEQ ID NO 85

<211> LENGTH: 4

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 85

Asn Val Ser Asn

1

<210> SEQ ID NO 86

<211> LENGTH: 4

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

-continued

<400> SEQUENCE: 86

Asn Asn Ser Asn
1

<210> SEQ ID NO 87

<211> LENGTH: 4

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 87

Asn Thr Ser Asn
1

<210> SEQ ID NO 88

<211> LENGTH: 4

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 88

Asn Asp Ser Asn
1

<210> SEQ ID NO 89

<211> LENGTH: 4

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 89

Asn Val Ser Pro
1

<210> SEQ ID NO 90

<211> LENGTH: 4377

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 90

atgcagattg agctgtccac ctgcttcttt ctgtgctgc tgagattctg cttctctgcc	60
accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg	120
ggagagctgc ctgtggatgc caggttocca cccagagtgc ccaagtcctt cccattcaac	180
acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt	240
gcaaaaacca gaccacctg gatgggactc ctgggaccca ccattcagc tgagggtgat	300
gacactgtgg tcatcacct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg	360
ggagtctcat actgaaaagc ctctgaagg gctgagtatg atgaccagac atcccagaga	420
gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc	480
aaggagaatg gacctatggc atctgacca ctctgctga catactecta cctttctcat	540

-continued

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa	600
ggatccctgg ccaaggagaa aaccagaca ctgcacaagt tcattctcct gtttgctgtc	660
tttgatgagg gcaagtcttg gcaactctgaa acaaagaact ccctgatgca agacagggat	720
gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacagatca	780
ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg	840
acaaccctg aagtgcactc cttttctctg gagggacaca ccttctctgt caggaaaccac	900
agacaagcct ctctggagat ctctccatc accttctca ctgcacagac actgctgatg	960
gaccttggac agttctctgt gtctctccac atctcttccc accagcatga tggcatggaa	1020
gcctatgtca aggtggactc atgcccctgag gaaccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat	1140
gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca cccaagaca	1200
tgggtgcaact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc	1260
cctgatgaca ggagctacaa gtctcagtag ctcaacaatg gccacaaaag aattggaaga	1320
aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc	1380
attcagcatg agtctggcat tctgggacca ctctctgatg gggaaagtgg agacaccctg	1440
ctcatcatct tcaagaacca ggctccagg ccctacaaca tctaccaca tggcatcact	1500
gatgtcaggc ccctgtacag caggagactg ccaaagggg tgaaacacct caaggacttc	1560
cccattctgc ctggagagat cttcaagtag aagtggaactg tcaactgtgga ggatggacca	1620
acaaagtctg accccagggtg cctcaccaga tactactcct cttttgtgaa catggagaga	1680
gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag	1740
agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag	1800
aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaacc tgctggggtg	1860
caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg	1920
tttgactctc tccagcttct tgtctgctg catgaggtgg cctactggta cattctttct	1980
attggggcac aaactgactt ctttctgtc ttcttctctg gatacacctt caagcacaag	2040
atggtgtatg aggacacct gacactcttc ccattctctg gggaaactgt gttcatgagc	2100
atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga	2160
atgactgcac tgctcaaagt ctctcctgt gacaagaaca ctgggacta ctatgaggac	2220
tcttatgagg acatctctgc ctacctgtc agcaagaaca ataccaccta cgtgaaccgc	2280
tccctgtctc agaatccacc tgtcctgaag agacaccaga gagagatcac caggacaacc	2340
ctccagtctg accaggaaga gattgactat gatgacacca tttctgtgga gatgaagaag	2400
gaggactttg acatctatga tgaggacgag aaccagtctc caagatcatt ccagaagaag	2460
acaagacact acttcattgc tgtctggaa agactgtggg actatggcat gtcttctct	2520
ccccatgtcc tcaggaacag ggcacagtct ggctctgtgc cacagttcaa gaaagtggtc	2580
ttccaggagt tcaactgatg ctcatcacc cagcccctgt acagagggga actgaatgag	2640
cacctgggac tcctgggacc atacatcagg gctgaggtgg aagacaacat catggtgaca	2700
ttcagaaacc aggcctccag gccctacagc ttctactctt ccctcatcag ctatgaggaa	2760
gaccagagac aaggggctga gccaaagaaag aactttgtga aacccaatga aaccaagacc	2820

-continued

tacttctgga aagtccagca ccacatggca cccaccaagg atgagtttga ctgcaaggcc	2880
tgggcatact tctctgatgt ggacctggag aaagatgtgc actctggcct gattggccca	2940
ctcctggtct gccacaccaa caccctgaac cctgcacatg gaaggcaagt gactgtgcag	3000
gagtttgccc tcttcttcac catctttgat gaaaccaagt catggtactt cactgagaac	3060
atggagagaa actgcagagc accatgcaac attcagatgg aagaccccac cttcaaggag	3120
aactacaggt tccatgccat caatggctac atcatggaca ccctgcctgg gcttgtcatg	3180
gcacaggacc agagaatcag atggtacctg ctttctatgg gatccaatga gaacattcac	3240
tccatccact tctctgggca tgtcttcact gtgagaaaga aggaggaata caagatggcc	3300
ctgtacaacc tctacctgg ggtctttgag actgtggaga tgetgccctc caaagctggc	3360
atctggaggg tggaatgctt cattggggag cacctgcctg ctggcatgtc aaccctgttc	3420
ctggtctaca gcaacaagtg ccagacaccc ctgggaatgg cctctggcca catcagggac	3480
ttccagatca ctgctctggt ccagtatggc cagtgggac ccaaactggc caggctccac	3540
tactctggct ccatcaatgc atggtcaacc aaggagccat tctcttggat caaggtggac	3600
ctgctggcac ccatgatcat tcatggcatc aagacacagg gggcaagaca gaaattctcc	3660
tctctgtaca tctcacagtt catcatcatg tactctctgg atggcaagaa gtggcagaca	3720
tacagaggca actccactgg caccctcatg gtcttctttg gcaatgtgga cagctctggc	3780
atcaagcaca acatcttcaa cctcccctc attgccagat acatcaggct gcaccccacc	3840
cactactcaa tcagatcaac cctcaggatg gaactgatgg gatgtgacct gaactcctgc	3900
tcaatgcccc tgggaatgga gagcaaggcc atttctgatg cccagatcac tgcacctct	3960
tacttcacca acatgtttgc cacctgggtc ccatcaaaag ccaggctgca cctccaggga	4020
agaagcaatg cctggagacc ccaggtcaac aacccaaagg aatggctgca agtggacttc	4080
cagaagacaa tgaagtcac tggggtgaca acccaggggg tcaagtctct gctcacctca	4140
atgtatgtga aggagttoct gatctcttcc tcacaggatg gccaccagtg gacactcttc	4200
ttccagaatg gcaaaagtc ggtgttccag ggcaaccagg actctttcac acctgtgggtg	4260
aactcactgg accccccct cctgacaaga tacctgagaa ttcaccccca gtcttgggtc	4320
caccagattg ccctgagaat ggaagtctct ggatgtgagg cacaagacct gtactga	4377

<210> SEQ ID NO 91

<211> LENGTH: 4377

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 91

atgcagattg agctgtccac ctgcttcttt ctgtgctctg tgagattctg cttctctgcc	60
accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg	120
ggagagctgc ctgtggatgc cagggttcca cccagatgct ccaagtcctt cccattcaac	180
acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gtteaacatt	240
gcaaaaccca gaccacctg gatgggactc ctgggaccca ccattcagge tgaggtgat	300
gacactgtgg tcgtcacctc caagaacatg gcatcccacc ctgtgtctct gcatgctgtg	360
ggagtctcat actggaatc ctctgaaggg gctgagtatg atgaccagac atcccagaga	420

-continued

gagaaagagg atgacaaggt gttccctggg aagtctcaca cctatgtgtg gcaagtoctc 480
aaggagaatg gaccactgc atctgaocca ccttgcoctga catactccta cctttctcat 540
gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcaggaa 600
ggatccctgg ccaaggagaa aaccagaca ctgcacaagt tcattctcct gtttgetgtc 660
tttgatgagg gcaagtcttg gcaactctgaa acaaagaact ccctgatgca agacagggat 720
gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780
ctgcoctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840
acaaccctg aagtgcactc cttttctctg gagggacaca ccttctcgtt caggaaccac 900
agacaagcct ctctggagat ctctccatc accttctca ctgcacagac actgctgatg 960
gacctggac agttctctgt gttctgccac atctcttccc accagcatga tggcatggaa 1020
gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140
gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200
tgggtgcact acattgtctg tgaggaagag gactgggact atgcaccact ggtcctggcc 1260
cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccaaaaag aattggaaga 1320
aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380
attcagcatg agtctggcat tctgggacca ctctctgatg gggaaagtggg agacaccctg 1440
ctcatcatct tcaagaacca ggctccagg ccttacaaca tctaccaca tggcatcact 1500
gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560
cccattctgc ctggagagat cttcaagtae aagtggactg tcaactgtgga ggatggacca 1620
acaaagtctg accccagggt cctcaccaga tactactcct cttttgtgaa catggagaga 1680
gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740
agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800
aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaacc tgctggggtg 1860
caactggaag acctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920
tttgactctc tccagcttct tgtctgctg catgaggtgg cctactggta cattcttct 1980
attggggcac aaactgactt ctttctgtc ttcttctctg gatacacctt caagcacaag 2040
atggtgtatg aggacacct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100
atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160
atgactgcac tgctcaaagt ctctctctgt gacaagaaca ctggggacta ctatgaggac 2220
tcttatgagg acatctctgc ctacctgctc agcaagaaca ataccaccta cgtgaaccgc 2280
tccctgtctc agaatccacc tgtcctgaag agacaccaga gagagatcac caggacaacc 2340
ctccagtctg accaggaaga gattgactat gatgacacca tttctgtgga gatgaagaag 2400
gaggactttg acatctatga tgaggacgag aaccagtctc caagatcatt ccagaagaag 2460
acaagacact acttcattgc tctgtgtgaa agactgtggg actatggcat gtcttctct 2520
ccccatgtcc tcaggaacag ggcacagtct ggctctgtgc cacagttcaa gaaagtgtc 2580
ttccaggagt tcaactgatg ctcatcacc cagcccctgt acagagggga actgaatgag 2640
cacctgggac tctgtggacc atacatcagg gctgaggtgg aagacaacat catggtgaca 2700

-continued

ttcagaaacc aggcctccag gccctacagc ttctactctt ccctcatcag ctatgaggaa 2760
gaccagagac aaggggctga gccaagaaag aactttgtga aaccaatga aaccaagacc 2820
tacttctgga aagtcagca ccacatggca cccaccaagg atgagtttga ctgcaaggcc 2880
tgggcatact tctctgatgt ggacctggag aaagatgtgc actctggcct gattggccca 2940
ctcctggtct gccacacca caccctgaac cctgcacatg gaaggcaagt gactgtgcag 3000
gagtttgccc tcttcttcac catctttgat gaaaccaagt catggtaact cactgagaac 3060
atggagagaa actgcagagc accatgcaac attcagatgg aagacccac cttcaaggag 3120
aactacaggt tccatgccoat caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180
gcacaggacc agagaatcag atggtaacctg ctttctatgg gatccaatga gaacattcac 3240
tccatccact tctctgggca tgtcttcaact gtgagaaaga aggaggaata caagatggcc 3300
ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgcctc caaagctggc 3360
atctggaggg tggaatgctt cattggggag cacctgcatg ctggcatgtc aaccctgttc 3420
ctggtctaca gcaacaagtg ccagacaccc ctgggaatgg cctctggcca catcagggac 3480
ttccagatca ctgcctctgg ccagtatggc cagtgggac ccaaactggc caggctccac 3540
tactctggct ccatcaatgc atggtaacc aaggagccat tctcttggat caagtggtgac 3600
ctgctggcac ccattgatcat tcatggcacc aagacacagg gggcaagaca gaaattctcc 3660
tctctgtaca tctcacagtt catcatcatg tactctctgg atggcaagaa gtggcagaca 3720
tacagaggca actccactgg caccctcatg gtcttctttg gcaatgtgga cagctctggc 3780
atcaagcaca acatcttcaa cctccaccac attgccagat acatcaggct gcacccacc 3840
cactactcaa tcagatcaac cctcaggatg gaactgatgg gatgtgacct gaactcctgc 3900
tcaatgcccc tgggaatgga gagcaaggcc atttctgatg ccagatcac tgcatctct 3960
tacttcacca acatgtttgc cacctggta ccatcaaaag ccagggtgca cctccaggga 4020
agaagcaatg cctggagacc ccaggtaac aaccxaaagg aatggctgca agtggaactc 4080
cagaagacaa tgaagtccac tggggtgaca acccaggggg tcaagtctct gctcaactca 4140
atgtatgtga aggagttcct gatctcttcc tcacaggatg gccaccagtg gacactcttc 4200
ttccagaatg gcaaagtcaa ggtgttcag ggcaaccagg actctttcac acctgtggtg 4260
aactcactgg accccccct cctgacaaga tacctgagaa ttcacccca gtcttgggtc 4320
caccagattg ccctgagaat ggaagtctg ggatgtgagg cacaagacct gtactga 4377

<210> SEQ ID NO 92

<211> LENGTH: 4377

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 92

atgcagattg agctgtccac ctgcttcttt ctgtgectgc tgagattctg cttctctgcc 60
accaggagat actacctggg ggctgtgaa ctttcttggg actacatgca gtctgacctg 120
ggagagctgc ctgtggatgc caggttocca cccagagtgc ccaagtcctt cccattcaac 180
acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

-continued

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcagggc tgagggtgat	300
gacactgtgg tcatcacccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg	360
ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga	420
gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcttc	480
aaggagaatg gacccatggc atctgaccca ctctgcoctga catactccta cctttctcat	540
gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggg gtgcagggaa	600
ggatccctgg ccaaggagaa aaccagaca ctgcacaagt tcattctcct gtttgetgtc	660
tttgatgagg gcaagtcttg gcaactctgaa acaaagaact ccctgatgca agacagggat	720
gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacagatca	780
ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg	840
acaacccctg aagtgcactc cttttctctg gagggacaca ccttctctgg caggaaccac	900
agacaagcct ctctggagat ctctcccctc accttctcct ctgcacagac actgctgatg	960
gaccttggac agttcctgct gttctgccc atctcttccc accagcatga tggcatggaa	1020
gcctatgtca aggtggactc atgcccctgag gaaccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggg cagatttgat	1140
gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca cccaagaca	1200
tgggtgcact acattgtctg tgaggaagag gactgggact atgcaccact ggtcctggcc	1260
cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gccacaaaag aattggaaga	1320
aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc	1380
attcagcatg agtctggcat tctgggacca ctctgtatg gggaaagtggg agacaccctg	1440
ctcatcatct tcaagaacca ggccctccagg ccttacaaca tctaccaca tggcatcact	1500
gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc	1560
cccattctgc ctggagagat cttcaagtac aagtggactg tcaactgtgga ggatggacca	1620
acaaagtctg accccagggt cctcaccaga tactactcct cttttgtgaa catggagaga	1680
gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag	1740
agaggcaacc agatcatgtc tgacaagaga aatgtgatc tgttctctgt ctttgatgag	1800
aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaacce tgcctgggtg	1860
caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg	1920
tttgactctc tccagcttct tgtctgctg catgaggtgg cctactggta cattctttct	1980
attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag	2040
atggtgtatg aggacacct gacactcttc ccattctctg gggaaactgt gttcatgagc	2100
atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga	2160
atgactgcac tgctcaaagt ctctctctgt gacaagaaca ctggggacta ctatgaggac	2220
tcttatgagg acatctctgc ctacctgctc agcaagaaca ataccaccta cgtgaaccgc	2280
tccctgtctc agaatccacc tgcctgaag agacaccaga gagagatcac caggacaacc	2340
ctccagtctg accaggaaga gattgactat gatgacacca tttctgtgga gatgaagaag	2400
gaggactttg acatctatga tgaggacgag aaccagtctc caagatcatt ccagaagaag	2460
acaagacact acttcattgc tgctgtggaa agactgtggg actatggcat gtcttctct	2520

-continued

```

ccccatgtcc tcaggaacag ggcacagtct ggctctgtgc cacagttcaa gaaagtggtc 2580
ttccaggagt tcaactgatgg ctcaattcacc cagcccctgt acagagggga actgaatgag 2640
cacctgggac tcctgggacc atacatcagg gctgaggtgg aagacaacat catggtgaca 2700
ttcagaaaacc aggcctccag gccctacagc ttctactctt ccctcatcag ctatgaggaa 2760
gaccagagac aaggggctga gcccaagaaag aactttgtga aacccaatga aaccaagacc 2820
tacttctgga aagtcagca ccacatggca cccaccaagg atgagtttga ctgcaaggcc 2880
tgggcatact tctctgatgt ggacctggag aaagatgtgc actctggcct gattggccca 2940
ctcctggtct gccacaccaa caccctgaac cctgcacatg gaaggcaagt gactgtgcag 3000
gagtttgccc tcttcttccac catctttgat gaaaccaagt catggtactt cactgagaac 3060
atggagagaa actgcagagc accatgcaac attcagatgg aagacccac cttcaaggag 3120
aactacaggt tccatgcat caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180
gcacaggacc agagaatcag atggtaacctg ctttctatgg gatccaatga gaacattcac 3240
tccatccact tctctgggca tgtcttcaact gtgagaaaga aggaggaata caagatggcc 3300
ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgcccctc caaagctggc 3360
atctggaggg tggaatgctt cattggggag cacctgcatg ctggcatgtc aaccctgttc 3420
ctggtctaca gcaacaagtg ccagacacc cctgggaatgg cctctggcca catcaggagc 3480
ttccagatca ctgctctgag ccagtatggc cagtgggcac ccaaactggc caggctccac 3540
tactctggct ccatcaatgc atggtaacc aaggagccat tctcttggat caaggtggac 3600
ctgctggcac ccatgatcat tcatggcatc aagacacagg gggcaagaca gaaattctcc 3660
tctctgtaca tctcacagtt catcatcatg tactctctgg atggcaagaa gtggcagaca 3720
tacagaggca actccactgg caccctcatg gtcttctttg gcaatgtgga cagctctggc 3780
atcaagcaca acatcttcaa ccctcccatc attgocagat acatcaggct gcaccccacc 3840
cactactcaa tcagatcaac cctcaggatg gaaactgatgg gatgtgacct gaactcctgc 3900
tcaatgcccc tgggaatgga gagcaaggcc atttctgatg cccagatcac tgcactctct 3960
tacttcacca acatgtttgc cacctggcca ccatcaaaag ccaggctgca cctccaggga 4020
agaagcaatg cctggagacc ccaggtcaac aacccaaagg aatggctgca agtggacttc 4080
cagaagacaa tgaagtccac tggggtgaca acccaggggg tcaagtctct gctcacctca 4140
atgtatgtga aggagtctct gatctcttcc tcacaggatg gccaccagtg gacactcttc 4200
ttccagaatg gcaaaagcaa ggtgttccag ggcaaccagg actctttcac acctgtggty 4260
aactcactgg accccccct cctgacaaga tacctgagaa ttcaccccca gtcttgggtc 4320
caccagattg ccctgagaat ggaagtctg ggatgtgagg cacaagacct gtactga 4377

```

<210> SEQ ID NO 93

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 93

```

atgcagattg agctgtccac ctgcttctt ctgtgctgc tgagattctg cttctctgcc 60
accaggagat actacctggg ggctgtgaa ctttcttggg actacatgca gtctgaactg 120

```

-continued

ggagagctgc	ctgtggatgc	caggttocca	cccagagtgc	ccaagtcctt	cccattcaac	180
acctctgtgg	tctacaagaa	gacactcttt	gtggaattca	ctgaccacct	gttcaacatt	240
gcaaaaccca	gaccaccctg	gatgggactc	ctgggaccca	ccattcaggc	tgagggtgat	300
gacactgtgg	tcgtcaccoct	caagaacatg	gcateccacc	ctgtgtctct	gcattgctgtg	360
ggagtctcat	actggaaatc	ctctgaaggg	gctgagtatg	atgaccagac	atcccagaga	420
gagaaagagg	atgacaaggt	gttccctggg	aagtctcaca	cctatgtgtg	gcaagtcctc	480
aaggagaatg	gaccactgc	atctgacca	cctgacctga	catactccta	cctttctcat	540
gtggacctgg	tcaaggacct	caactctgga	ctgattgggg	cactgctggg	gtgcagggaa	600
ggatccctgg	ccaaggagaa	aaccagaca	ctgcacaagt	tcattctcct	gtttgctgtc	660
tttgatgagg	gcaagtcttg	gcactctgaa	acaaagaact	ccctgatgca	agacagggat	720
gctgcctctg	ccagggcatg	gccaagatg	cacactgtga	atggctatgt	gaacagatca	780
ctgcctggac	tcattggctg	ccacaggaaa	tctgtctact	ggcatgtgat	tggcatgggg	840
acaaccctg	aagtgcactc	cattttctctg	gagggacaca	ccttctctgt	caggaaccac	900
agacaagcct	ctctggagat	ctctcccatc	accttctca	ctgcacagac	actgctgatg	960
gaccttgac	agttcctgct	gttctgccac	atctcttccc	accagcatga	tggcatggaa	1020
gcctatgtca	agggtgactc	atgccctgag	gaaccacagc	tcaggatgaa	gaacaatgag	1080
gaggctgagg	actatgatga	tgacctgact	gactctgaga	tggatgtggg	cagatttgat	1140
gatgacaact	ctccatcctt	cattcagatc	aggctctgtg	caaagaaaca	ccccaagaca	1200
tgggtgcact	acattgctgc	tgaggaagag	gactgggact	atgcaccact	ggtcctggcc	1260
cctgatgaca	ggagctacaa	gtctcagtac	ctcaacaatg	gcccacaaag	aattggaaga	1320
aagtacaaga	aagtcagatt	catggcctac	actgatgaaa	ccttcaagac	aagagaagcc	1380
attcagcatg	agtctggcat	tctgggacca	ctcctgtatg	gggaagtggg	agacaccctg	1440
ctcatcatct	tcaagaacca	ggcctccagg	ccctacaaca	tctaccaca	tggcatcact	1500
gatgtcaggc	ccctgtacag	caggagactg	ccaaaagggg	tgaaacacct	caaggacttc	1560
cccattctgc	ctggagagat	cttcaagtac	aagtgaggctg	tcaactgtgga	ggatggacca	1620
acaaagtctg	accccagggtg	cctcaccaga	tactactcct	cttttgtgaa	catggagaga	1680
gacctggcat	ctggactgat	tggaccactg	ctcatctgct	acaaggagtc	tgtggaccag	1740
agaggcaacc	agatcatgtc	tgacaagaga	aatgtgattc	tgttctctgt	ctttgatgag	1800
aacagatcat	ggtacctgac	tgagaacatt	cagagattcc	tgcccaacc	tgctgggggtg	1860
caactggaag	accctgagtt	ccaggcaagc	aacatcatgc	actccatcaa	tggctatgtg	1920
tttgactctc	tccagcttct	tgtctgctg	catgaggtgg	cctactggta	cattctttct	1980
attggggcac	aaactgactt	cctttctgtc	ttcttctctg	gatacacctt	caagcacaag	2040
atggtgtatg	aggacaccct	gacactcttc	ccattctctg	gggaaactgt	gttcatgagc	2100
atggagaacc	ctggactgtg	gattctggga	tgccacaact	ctgacttcag	aaacagggga	2160
atgactgcac	tgctcaaagt	ctcctcctgt	gacaagaaca	ctggggacta	ctatgaggac	2220
tcttatgagg	acatctctgc	ctacctgtc	agcaagaaca	atgccattga	gcccagaagc	2280
ttctctcaga	atccacctgt	cctgaagaga	caccagagag	agatcaccag	gacaaccctc	2340
cagtctgacc	aggaagagat	tgactatgat	gacaccattt	ctgtggagat	gaagaaggag	2400

-continued

gactttgaca tctatgatga ggacgagaac cagtctccaa gatcattcca gaagaagaca 2460
agacactact tcattgtctgc tgtggaaaga ctgtgggact atggcatgtc ttcctctccc 2520
catgtcctca ggaacagggc acagtctggc tctgtgccac agttcaagaa agtggctctc 2580
caggagtcca ctgatggctc attcaccocag cccctgtaca gaggggaact gaatgagcac 2640
ctgggactcc tgggaccata catcagggct gaggtggaag acaacatcat ggtgacattc 2700
agaaaccagg cctccaggcc ctacagcttc tactcttccc tcatcagcta tgaggaagac 2760
cagagacaag gggctgagcc aagaaagaac tttgtgaaac ccaatgaaac caagacctac 2820
ttctggaaaag tccagcacca catggcacc accaaggatg agtttgactg caaggcctgg 2880
gcatacttct ctgatgtgga cctggagaaa gatgtgcact ctggcctgat tggcccactc 2940
ctggtctgcc acaccaacac cctgaacctc gcacatggaa ggcaagtgac tgtgcaggag 3000
tttgccctct tcttcacat ctttgatgaa accaagtcac ggtacttcac tgagaacatg 3060
gagagaaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagaac 3120
tacagggtcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggca 3180
caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cattcactcc 3240
atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggccctg 3300
tacaacctct acctgggggt ctttgagact gtggagatgc tgcctccaa agctggcatc 3360
tggagggtgg aatgcctcat tggggagcac ctgcatgctg gcatgtcaac cctgttctctg 3420
gtctacagca acaagtgcc aacaccctc ggaatggcct ctggccacat cagggacttc 3480
cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac 3540
tctggctcca tcaatgcatg gtcaaccaag gagccattct cttggatcaa ggtggactctg 3600
ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctctctc 3660
ctgtacatct cacagttcat catcatgtac tctctggatg gcaagaagtg gcagacatac 3720
agaggcaact ccactggcac cctcatggtc ttctttggca atgtggacag ctctggcatc 3780
aagcacaaca tcttcaaccc tcccatcatt gccagataca tcaggctgca cccccaccac 3840
tactcaatca gatcaacct caggatggaa ctgatgggat gtgacctgaa ctctctctca 3900
atgcccctgg gaatggagag caaggccatt tctgatgccc agatcactgc atctcttac 3960
ttcacciaaca tgtttgccac ctggctacca tcaaaagcca ggctgcacct ccaggaaga 4020
agcaatgctt ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttcag 4080
aagacaatga aagtcactgg ggtgacaacc caggggtca agtctctgct cacctcaatg 4140
tatgtgaagg agttcctgat ctcttctca caggatggcc accagtgagc actcttcttc 4200
cagaatggca aagtcaaggt gttccagggc aaccaggact ctttcacacc tgtggtgaa 4260
tcaactggacc cccccctct gacaagatac ctgagaattc acccccagtc ttgggtccac 4320
cagattgccc tgagaatgga agtcctggga tgtgaggcac aagacctgta ctga 4374

<210> SEQ ID NO 94

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

-continued

<400> SEQUENCE: 94

atgcagattg agctgagcac ctgcttcttc ctgtgcoctg tgaggttctg cttctctgcc	60
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg	120
ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatecct cccattcaac	180
acctctgtgg tctacaagaa gacctcttt gtggagtcca ctgaccacct gttcaacatt	240
gccaaaccca ggccacctg gatgggactc ctgggaccca ccattcagge tgagggtgat	300
gacactgtgg tcgtcacocct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg	360
ggggctcagct actggaagtc ctctgagggg gctgagtatg atgaccagac ctcccagagg	420
gagaaggagg atgacaaagt gttccctggg aagagccaca cctatgtgtg gcaggtcctc	480
aaggagaatg gccccactgc ctctgaccca ccttgcoctga cctactccta cttttctcat	540
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcaggag	600
ggctccctgg ccaaaagaaa gaccagacc ctgcacaagt tcattctcct gtttgetgtc	660
tttgatgagg gcaagagctg gactctgaa accaagaact ccctgatgca ggacagggat	720
gtgcctctg ccagggctg gcccaagatg cacactgtga atggctatgt gaacaggagc	780
ctgcoctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg	840
acaaccctg aggtgcactc cttttctctg gagggccaca ccttctggt caggaaccac	900
agacaggcca gcctggagat cagccccatc accttctca ctgcccagac cctgctgatg	960
gacctgggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag	1020
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat	1140
gatgacaaca gccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc	1200
tgggtgcact acattgctgc tgaggaggag gactgggact atgcccact ggtcctggcc	1260
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gccacacagag gattggacgc	1320
aagtacaaga aagtcagggt catggcctac actgatgaaa ccttcaagac caggagggcc	1380
attcagcatg agtctggcat cctgggcca ctctgtatg gggaggtggg ggacacctg	1440
ctcatctct tcaagaacca ggctccagg ccctacaaca tctaccaca tggcatcact	1500
gatgtcagge ccctgtacag ccgagggctg ccaaaggggg tgaaacacct caaggacttc	1560
cccattctgc ctggggagat cttcaagtae aagtggactg tcaactgtga ggatggacca	1620
accaaactg accccagggt cctcaccaga tactactcca gctttgtgaa catggagagg	1680
gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag	1740
aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag	1800
aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccacccc tgctgggggtg	1860
cagctggagg acctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg	1920
tttgacagcc tccagcttct tgtctgcoct catgaggtgg cctactggta cattctttct	1980
attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag	2040
atggtgtatg aggacacct gacctcttc ccattctctg gggagactgt gttcatgagc	2100
atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc	2160
atgactgccc tgctcaaagt ctctcctgtg gacaagaaca ctggggacta ctataggac	2220

-continued

agctatgagg	acatctctgc	ctacctgctc	agcaagaaca	atgccattga	gcccaggagc	2280
ttcagccaga	atccacctgt	cctgaaacgc	caccagaggg	agatcaccag	gaccaccctc	2340
cagtctgacc	aggaggagat	tgactatgat	gacaccattt	ctgtggagat	gaagaaagag	2400
gactttgaca	tctatgacga	ggacgagAAC	cagagcccaa	ggagcttcca	gaagaagacc	2460
aggcactact	tcattgtctg	tgtggagcgc	ctgtgggact	atggcatgag	ctccagcccc	2520
catgtcctca	ggaacagggc	ccagtctggc	tctgtgccac	agttcaagaa	agtggctctc	2580
caagagttea	ctgatggcag	cttcacccag	cccctgtaca	gaggggagct	gaatgagcac	2640
ctgggactcc	tgggcccata	catcagggct	gaggtggagg	acaacatcat	ggtgaccttc	2700
cgcaaccagg	cctccaggcc	ctacagcttc	tacagctccc	tcatcagcta	tgaggaggac	2760
cagaggcagg	gggctgagcc	acgcaagaac	tttgtgaaac	ccaatgaaac	caagacctac	2820
ttctggaaag	tccagcacca	catggccccc	accaaggatg	agtttgactg	caaggcctgg	2880
gcctactctc	ctgatgtgga	cctggagaag	gatgtgcact	ctggcctgat	tggcccactc	2940
ctggtctgcc	acaccaacac	cctgaaccct	gcccattgaa	ggcaagtgac	tgtgcaggag	3000
tttgccctct	tcttcaccat	ctttgatgaa	accaagagct	ggtacttcac	tgagaacatg	3060
gagcgcaact	gcagggcccc	atgcaacatt	cagatggagg	acccccctt	caaagagaac	3120
taccgcttcc	atgccatcaa	tggtacatc	atggaccccc	tgctgggct	tgtcatggcc	3180
caggaccaga	ggatcagggt	gtacctgctt	tctatgggct	ccaatgagaa	cattcactcc	3240
atccactctc	ctgggcatgt	cttcaactgt	cgcaagaagg	aggagtacaa	gatggccctg	3300
tacaacctct	accctggggg	ctttgagact	gtggagatgc	tgccctccaa	agctggcatc	3360
tggaggggtg	agtgccctcat	tggggagcac	ctgcatgctg	gcatgagcac	cctgttctct	3420
gtctacagca	acaagtgcc	gacccccctg	ggaatggcct	ctggccacat	cagggacttc	3480
cagatcactg	cctctggcca	gtatggccag	tgggccccca	agctggccag	gctccactac	3540
tctggatcca	tcaatgcctg	gagcaccaag	gagccattca	gctggatcaa	agtggacctg	3600
ctggccccca	tgatcatcca	tggcatcaag	accagggggg	ccaggcagaa	gttctccage	3660
ctgtacatca	gccagttcat	catcatgtac	agcctggatg	gcaagaaatg	gcagacctac	3720
agaggcaact	ccactggaac	actcatggtc	ttctttgcca	atgtggacag	ctctggcatc	3780
aagcacaaca	tcttcaaccc	cccaatcctc	gccagataca	tcaggctgca	ccccaccacc	3840
tacagcatcc	gcagcacctc	caggatggag	ctgatgggct	gtgacctgaa	ctcctgcagc	3900
atgcccctgg	gcatggagag	caaggccatt	tctgatgccc	agatcactgc	ctccagctac	3960
ttcaccaca	tgtttgccac	ctggagccca	agcaaggcca	ggctgcacct	ccagggaggg	4020
agcaatgcct	ggaggcccc	ggtcaacaac	ccaaggaggt	ggctgcagggt	ggacttccag	4080
aagaccatga	aggctcactg	ggtgaccacc	caggggggtca	agagcctgct	caccagcatg	4140
tatgtgaagg	agttcctgat	cagctccagc	caggatggcc	accagtggac	cctcttcttc	4200
cagaatggca	aggtcaagggt	gttccagggc	aaccaggaca	gcttcacccc	tgtggtgaa	4260
agcctggacc	ccccctcctc	gaccagatac	ctgaggattc	acccccagag	ctgggtccac	4320
cagattgccc	tgaggatgga	ggtcctggga	tgtgagggcc	aggacctgta	ctga	4374

<210> SEQ ID NO 95

<211> LENGTH: 4377

<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 95

```

atgcagattg agctgagcac ctgtctcttc ctgtgctgc tgaggttctg cttctctgcc      60
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg      120
ggggagctgc ctgtggatgc cagggttccca cccagagtgc ccaaatcctt cccattcaac      180
acctctgtgg tctacaagaa gacctcttt gtggagtcca ctgaccacct gttcaacatt      240
gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgagggtgat      300
gacactgtgg tcatcacctc caagaacatg gcctcccacc ctgtgagcct gcatgctgtg      360
ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg      420
gagaaggagg atgacaaaagt gttccctggg ggcagccaca cctatgtgtg gcaggctctc      480
aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat      540
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggg gtgcaggagg      600
ggctccctgg ccaaagagaa gaccagacc ctgcacaagt tcattctcct gtttctgtc      660
tttgatgagg gcaagagctg gcaactctgaa accaagaact ccctgatgca ggacagggat      720
gctgcctctg ccagggcctg gcccagaatg cacactgtga atggctatgt gaacaggagc      780
ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcattgtgat tggcatgggg      840
acaacccctg aggtgcactc cttttctctg gagggccaca ccttctctgt caggaaccac      900
agacaggcca gcctggagat cagccccatc accttctcct ctgcccagac cctgctgatg      960
gacctcggac agttcctgct gttctgccc atcagctccc accagcatga tggcatggag      1020
gcctatgtca agtgggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag      1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggg ccgctttgat      1140
gatgacaaca gcccattcct cattcagatc aggtctgtgg ccaagaaaca cccaagacc      1200
tgggtgcact acattgtctg tgaggaggag gactgggact atgccccact ggtcctggcc      1260
cctgatgaca gtagctacaa gagccagtac ctcaacaatg gccacacagag gattggagcg      1320
aagtacaaga aagtcagggt catggcctac actgatgaaa ccttcaagac cagggaggcc      1380
attcagcatg agtctggcat cctgggccc actcctgatg gggagggtgg ggacaccctg      1440
ctcatcatct tcaagaacca ggctccagg ccctacaaca tctaccaca tggcatcact      1500
gatgtcaggc ccctgtacag ccgcaggctg ccaaggggg tgaaacacct caaggacttc      1560
cccattctgc ctggggagat cttcaagtac aagtggactg tcaactgtga ggatggacca      1620
accaaactctg accccagggt cctcaccaga tactactcca gctttgtgaa catggagagg      1680
gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag      1740
aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag      1800
aacaggagct ggtacctgac tgagaacatt cagcgtctcc tgcccacccc tgctgggggtg      1860
cagctggagg acctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg      1920
tttgacagcc tccagcttct tgtctgctg catgagggtg ctaactggta cattctttct      1980
attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag      2040
atggtgtatg aggacacct gacctcttc ccattctctg gggagactgt gttcatgagc      2100

```

-continued

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160
atgactgccc tgctcaaagt ctctcctgt gacaagaaca ctggggacta ctatgaggac 2220
agctatgagg acatctctgc ctacctgtc agcaagaaca ataccaccta cgtgaaccgc 2280
tccctgagcc agaateccacc tgtcctgaaa cgccaccaga gggagatcac caggaccacc 2340
ctccagtctg accaggagga gattgactat gatgacacca tttctgtgga gatgaagaaa 2400
gaggactttg acatctatga cgaggacgag aaccagagcc caaggagctt ccagaagaag 2460
accaggcaact acttcattgc tgetgtggag cgcctgtggg actatggcat gagctccage 2520
ccccatgtcc tcaggaacag ggcccagtct ggctctgtgc cacagttcaa gaaagtggtc 2580
ttccaagagt tcaactgatgg cagcttcacc cagcccctgt acagagggga gctgaatgag 2640
cacctgggac tcctgggccc atacatcagg gctgaggtgg aggacaacat catggtgacc 2700
ttccgcaacc aggcctccag gccctacagc ttctacagct ccctcatcag ctatgaggag 2760
gaccagaggc aggggggtga gccacgcaag aactttgtga aacccaatga aaccaagacc 2820
tacttctgga aagtcagca ccacatggcc cccaccaagg atgagtttga ctgcaaggcc 2880
tgggcctact tctctgatgt ggacctggag aaggatgtgc actctggcct gattggocca 2940
ctcctggtct gccacaccaa caccctgaac cctgcccag gaaggcaagt gactgtgcag 3000
gagtttgcct tcttcttacc catctttgat gaaaccaaga gctggactt cactgagaac 3060
atggagcgca actgcagggc cccatgcaac attcagatgg aggacccac cttcaaagag 3120
aactaccgct tccatgccc caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180
gcccaggacc agaggatcag gtggtaoctg ctttctatgg gctccaatga gaacattcac 3240
tccatccact tctctgggca tgtcttcaact gtgcgcaaga aggaggagta caagatggcc 3300
ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgcccctc caaagetggc 3360
atctggaggg tggagtgcct cattggggag cacctgcatg ctggcatgag caccctgttc 3420
ctggtctaca gcaacaagtg ccagaccccc ctgggaatgg cctctggcca catcagggac 3480
ttccagatca ctgcctctgg ccagtatggc cagtgggccc ccaagctggc caggctccac 3540
tactctggat ccatcaatgc ctggagcacc aaggagccat tcagctggat caaagtggac 3600
ctgctggccc ccatgatcat ccatggcatc aagacccagg gggccaggca gaagtctccc 3660
agcctgtaca tcagccagtt catcatcatg tacagcctgg atggcaagaa atggcagacc 3720
tacagaggca actccactgg aacactcatg gtcttctttg gcaatgtgga cagctctgga 3780
atcaagcaca acatcttcaa ccccccaatc atcgccagat acatcaggct gcaccccacc 3840
cactacagca tccgcagcac cctcaggatg gagctgatgg gctgtgacct gaactcctgc 3900
agcatgcccc tgggcatgga gagcaaggcc atttctgatg ccagatcac tgctccage 3960
tacttcacca acatgtttgc cacctggagc ccaagcaagg ccaggctgca cctccaggga 4020
aggagcaatg cctggaggcc ccaggtcaac aacccaagg agtggctgca ggtggacttc 4080
cagaagacca tgaaggtcac tggggtgacc acccaggggg tcaagagcct gctcaccagc 4140
atgtatgtga aggagtctct gatcagctcc agccaggatg gccaccagtg gaccctcttc 4200
ttccagaatg gcaaggtoaa ggtgttocag ggcaaccagg acagcttcaac cctgtggtg 4260
aacagcctgg accccccct cctgaccaga tacctgagga ttcaccccca gagctgggtc 4320
caccagatg ccctgaggat ggaggtcctg ggatgtgagg cccaggacct gtactga 4377

-continued

<210> SEQ ID NO 96
<211> LENGTH: 4377
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 96

atgcagattg agctgagcac ctgcttcttc ctgtgectgc tgaggttctg cttctctgcc 60
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120
ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180
acctctgtgg tctacaagaa gacctcttt gtggagtca ctgaccacct gttcaacatt 240
gccaaacca ggccacctg gatgggactc ctgggacca ccattcaggc tgaggtgat 300
gacctgtgg tegtaccct caagaacatg gctcccacc ctgtgagcct gcatgctgtg 360
ggggtcagct actggaagtc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420
gagaaggagg atgacaaagt gttccctggg aagagccaca cctatgtgtg gcaggtcctc 480
aaggagaatg gccccactgc ctctgacca cctgctga cctactccta cctttctcat 540
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcaggag 600
ggctccctgg ccaaaagaa gaccagacc ctgcacaagt tcattctcct gtttctgtc 660
tttgatgagg gcaagagctg gcaactgaa accaagaact ccctgatgca ggacagggat 720
gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780
ctgctggac tcattggctg ccacagaaa tctgtctact ggcatgtgat tggcatgggg 840
acaaccctg aggtgcactc cttttctctg gagggccaca ccttctggt caggaaccac 900
agacaggcca gcctggagat cagcccoact accttctca ctgcccagac cctgctgatg 960
gacctcggac agttctgct gttctgccac atcagctccc accagcatga tggcatggag 1020
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140
gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200
tgggtgcact acattgtctg tgaggaggag gactgggact atgcccact ggtcctggcc 1260
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gccacagag gattggacgc 1320
aagtacaaga aagtcagggt catggcctac actgatgaaa ccttcaagac caggagggcc 1380
attcagcatg agtctggcat cctgggcca ctccctgatg gggaggtggg ggacacctg 1440
ctcatcatct tcaagaacca ggctccagg ccctacaaca tctaccaca tggcatcact 1500
gatgtcaggc ccctgtacag ccgaggctg ccaaaggggg tgaaacacct caaggacttc 1560
cccattctgc ctggggagat cttcaagtac aagtgaggctg tcaactgtga ggatggacca 1620
accaaaatctg accccagggt cctcaccaga tactactcca gctttgtgaa catggagagg 1680
gacctggcct ctggcctgat tggccactg ctcatctgct acaaggagtc tgtggaccag 1740
aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt cttgatgag 1800
aacaggagct ggtacctgac tgagaacatt cagcgttcc tgcccaacc tgctggggtg 1860
cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

-continued

tttgacagcc	tccagcttcc	tgtctgctg	catgaggtgg	cctactggta	cattctttct	1980
attggggccc	agactgactt	cctttctgtc	ttcttctctg	gctacacctt	caaacacaag	2040
atggtgtatg	aggacacct	gacctcttc	ccattctctg	gggagactgt	gttcattgagc	2100
atggagaacc	ctggcctgtg	gattctggga	tgccacaact	ctgacttccg	caacaggggc	2160
atgactgccc	tgctcaaagt	ctcctctgt	gacaagaaca	ctggggacta	ctatgaggac	2220
agctatgagg	acatctctgc	ctacctgtc	agcaagaaca	ataccaccta	cgtgaaccgc	2280
tccctgagcc	agaatccacc	tgtcctgaaa	cgccaccaga	gggagatcac	caggaccacc	2340
ctccagtctg	accaggagga	gattgactat	gatgacacca	tttctgtgga	gatgaagaaa	2400
gaggactttg	acatctatga	cgaggacgag	aaccagagcc	caaggagctt	ccagaagaag	2460
accaggcact	acttcattgc	tgctgtggag	cgctgtggg	actatggcat	gagctccagc	2520
ccccatgtcc	tcaggaacag	ggcccagtct	ggctctgtgc	cacagttcaa	gaaagtggtc	2580
ttccaagagt	tactgatgg	cagcttcacc	cagccctgt	acagagggga	gctgaatgag	2640
cacctgggac	tcctgggccc	atacatcagg	gctgaggtgg	aggacaacat	catggtgacc	2700
ttccgcaacc	aggcctccag	gccctacagc	ttctacagct	ccctcatcag	ctatgaggag	2760
gaccagaggc	agggggctga	gccacgcaag	aactttgtga	aacccaatga	aaccaagacc	2820
tacttctgga	aagtccagca	ccacatggcc	cccaccaagg	atgagtttga	ctgcaaggcc	2880
tgggcctact	tctctgatgt	ggacctggag	aaggatgtgc	actctggcct	gattggccca	2940
ctcctggtct	gccacaccaa	cacctgaac	cctgcccattg	gaaggcaagt	gactgtgcag	3000
gagtttgccc	tcttcttcac	catctttgat	gaaaccaaga	gctggtactt	actgagaac	3060
atggagcgca	actgcagggc	cccatgcaac	atcagatgg	aggaccccac	cttcaaagag	3120
aactaccgct	tccatgccat	caatggctac	atcatggaca	ccctgcctgg	gcttgtcatg	3180
gcccaggacc	agaggatcag	gtggtaoctg	ctttctatgg	gctccaatga	gaacattcac	3240
tccatccact	tctctgggca	tgtcttcact	gtgcgcaaga	aggaggagta	caagatggcc	3300
ctgtacaacc	tctacctgg	ggtctttgag	actgtggaga	tgtgtccctc	caaagctggc	3360
atctggaggg	tggagtgcct	cattggggag	cacctgcatg	ctggcatgag	cacctgttc	3420
ctggtctaca	gcaacaagtg	ccagaccccc	ctgggaatgg	cctctggcca	catcagggac	3480
ttccagatca	ctgctctgg	ccagtatggc	cagtggggccc	ccaagctggc	caggctccac	3540
tactctggat	ccatcaatgc	ctggagcacc	aaggagccat	tcagctggat	caaagtggac	3600
ctgctggccc	ccatgatcat	ccatggcatc	aagaccagg	gggccaggca	gaagtctcc	3660
agcctgtaca	tcagccagtt	catcatcatg	tacagcctgg	atggcaagaa	atggcagacc	3720
tacagaggca	actccactgg	aacactcatg	gtcttctttg	gcaatgtgga	cagctctggc	3780
atcaagcaca	acatcttcaa	ccccccaatc	atgccagat	acatcaggct	gcaccccacc	3840
cactacagca	tccgcagcac	cctcaggatg	gagctgatgg	gctgtgacct	gaactcctgc	3900
agcatgcccc	tgggcatgga	gagcaaggcc	atctctgatg	cccagatcac	tgctccagc	3960
tacttcacca	acatgtttgc	cacctggagc	ccaagcaagg	ccaggctgca	cctccagggg	4020
aggagcaatg	cctggaggcc	ccaggtcaac	aaccacaagg	agtggctgca	ggtggacttc	4080
cagaagacca	tgaaggtcac	tggggtgacc	accaggggg	tcaagagcct	gctcaccagc	4140
atgtatgtga	aggagttoct	gatcagctcc	agccaggatg	gccaccagtg	gacctcttc	4200

-continued

ttccagaatg gcaaggtcaa ggtgttccag ggcaaccagg acagettcac cctgtggtg	4260
aacagcctgg accccccct cctgaccaga tacctgagga ttcacccca gagctgggtc	4320
caccagattg ccctgaggat ggaggtcctg ggatgtgagg cccaggacct gtactga	4377

<210> SEQ ID NO 97

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 97

atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc	60
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg	120
ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaactcctt cccattcaac	180
acctctgtgg tctacaagaa gacctcttt gtggagtcca ctgaccacct gttcaacatt	240
gccccaaacca ggccaccctg gatgggactc ctgggaccca ccattcagge tgagggtgat	300
gacactgtgg tcatcacct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg	360
ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg	420
gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggctctc	480
aaggagaatg gccccatggc ctctgaccca ctctgctga cctactccta ctttctcat	540
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcaggag	600
ggctccctgg ccaaaagaaa gaccagacc ctgcacaagt tcattctctt gtttctgtc	660
tttgatgagg gcaagagctg gacctctgaa accaagaact ccctgatgca ggacagggat	720
gctgcctctg ccagggcctg gccaagatg cacactgtga atggctatgt gaacaggagc	780
ctgctgggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg	840
acaaccctg aggtgcactc catttctctg gagggccaca ctttctggt caggaaccac	900
agacaggcca gcctggagat cagccccatc accttctca ctgcccagac cctgctgatg	960
gacctcggac agttcctgct gtctctccac atcagctccc accagcatga tggcatggag	1020
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat	1140
gatgacaaca gcccacctt cattcagatc aggtctgtgg ccaagaaaaca ccccaagacc	1200
tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc	1260
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc	1320
aagtacaaga aagtcagggt catggcctac actgatgaaa ccttcaagac caggagggcc	1380
attcagcatg agtctggcat cctgggcca ctctctgatg gggagggtgg ggacacctg	1440
ctcatcatct tcaagaacca ggctccagg ccctacaaca tctaccaca tggcatcact	1500
gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc	1560
cccattctgc ctggggagat cttcaagtac aagtggactg tcaactgtgga ggatggacca	1620
accaaactg accccagggt cctcaccaga tactactcca gctttgtgaa catggagagg	1680
gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag	1740
aggggaaacc agatcatgct tgacaaggg aatgtgattc tgttctctgt ctttgatgag	1800

-continued

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860
cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggetatgtg 1920
tttgacagcc tccagcttcc tgtctgctg catgagggtg cctactggta cattctttct 1980
attggggccc agactgaact cctttctgtc ttcttctctg gctacacctt caaacacaag 2040
atggtgtatg aggacacct gacctcttc ccattctctg gggagactgt gttcatgagc 2100
atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160
atgactgccc tgctcaaagt ctctctctgt gacaagaaca ctggggacta ctatgaggac 2220
agctatgagg acatctctgc ctacctgtc agcaagaaca atgccattga gcccaggagc 2280
ttcagccaga atcccactgt cctgaaacgc caccagaggg agatcaccag gaccacctc 2340
cagtctgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag 2400
gactttgaca tctatgacga ggacgagaac cagagcccaa ggagcttcca gaagaagacc 2460
aggcactact tcattgtctg tgtggagcgc ctgtgggact atggcatgag ctccagcccc 2520
catgtctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtggctctc 2580
caagagttca ctgatggcag cttcaccag cccctgtaca gaggggagct gaatgagcac 2640
ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgacctc 2700
cgcaaccagg cctccaggcc ctacagcttc tacagctccc tcatcagcta tgaggaggac 2760
cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac 2820
ttctggaaaag tccagcacca catggcccc accaaggatg agtttgactg caaggcctgg 2880
gcctactct ctgatgtgga cctggagaag gatgtgcaact ctggcctgat tggcccactc 2940
ctggtctgcc acaccaaac cctgaacct gcccatggaa ggcaagtgac tgtgcaggag 3000
tttgccctct tctcaccat ctttgatgaa accaagagct ggtacttcaac tgagaacatg 3060
gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaac 3120
taccgcttcc atgccatcaa tggtacatc atggaccccc tgctgggct tgtcatggcc 3180
caggaccaga ggatcagggt gtacctgtt tctatgggt ccaatgagaa cattcactcc 3240
atccactct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggccctg 3300
tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360
tggaggggtg agtgccctcat tggggagcac ctgcatgctg gcatgagcac cctgttctctg 3420
gtctacagca acaagtgcc aacccccctg ggaatggcct ctggccacat cagggacttc 3480
cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac 3540
tctggatcca tcaatgcctg gagcaccaag gagccattca gctggatcaa agtggacctg 3600
ctggccccca tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc 3660
ctgtacatca gccagttcat catcatgtac agcctggatg gcaagaaatg gcagacctac 3720
agaggcaact ccaactggaac actcatggtc ttctttggca atgtggacag ctctggcatc 3780
aagcacaaca tcttcaaccc cccaatcatc gccagataca tcaggctgca cccccaccac 3840
tacagcatcc gcagcaccct caggatggag ctgatgggt gtgacctgaa ctctgcagc 3900
atgccctgg gcatggagag caaggccatt tctgatgcc agatcactgc ctccagctac 3960
ttcaccaca tgtttgccac ctggagccca agcaaggcca ggctgcacct ccaggaagg 4020
agcaatgctt ggaggcccc ggtcaacaac ccaaaggagt ggctgaggt ggacttccag 4080

-continued

aagaccatga aggtcactgg ggtgaccacc caggggggtca agagcctgct caccagcatg	4140
tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctctctctc	4200
cagaatggca aggtcaaggt gttccagggc aaccaggaca gcttcacccc tgtggtgaac	4260
agcctggacc cccccctcct gaccagatac ctgaggattc acccccagag ctgggtccac	4320
cagattgccc tgaggatgga ggtcctggga tgtgaggccc aggacctgta ctga	4374

<210> SEQ ID NO 98

<211> LENGTH: 4377

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 98

atgcagattg agctgagcac ctgcttcttc ctgtgctgc tgaggttctg cttctctgcc	60
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg	120
ggggagctgc ctgtggatgc cagggtccca cccagagtgc ccaaatcctt cccattcaac	180
acctctgtgg tctacaagaa gacctcttt gtggagtcca ctgaccacct gttcaacatt	240
gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgagggtgat	300
gacactgtgg tcatcacctc caagaacatg gcctcccacc ctgtgagcct gcatgctgtg	360
ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg	420
gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggctctc	480
aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat	540
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcaggag	600
ggctccctgg ccaaagagaa gaccagacc ctgcacaagt tcattctcct gtttgctgtc	660
tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat	720
gctgcctctg ccagggcctg gcccagatg cacactgtga atggctatgt gaacaggagc	780
ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcattgtgat tggcatgggg	840
acaacccctg aggtgcactc cttttctctg gagggccaca ccttctctgt caggaaccac	900
agacaggcca gcctggagat cagccccatc accttctcct ctgcccagac cctgctgatg	960
gacctcggac agttcctgct gtcctgccc atcagctccc accagcatga tggcatggag	1020
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat	1140
gatgacaaca gcccacctt cattcagatc aggtctgtgg ccaagaaaca cccaagacc	1200
tgggtgcact acattgtctg tgaggaggag gactgggact atgccccact ggtcctggcc	1260
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc	1320
aagtacaaga aagtccaggt catggcctac actgatgaaa ccttcaagac cagggaggcc	1380
attcagcatg agtctggcat cctgggccc ctctctgatg gggagggtggg ggacaccctg	1440
ctcatcatct tcaagaacca ggctccagc ccctacaaca tctaccaca tggcatcact	1500
gatgtcaggc ccctgtacag ccgaggctg ccaaaggggg tgaaacacct caaggacttc	1560
cccattctgc ctggggagat cttcaagtac aagtggactg tcaactgtgga ggatggacca	1620

-continued

accaaactcg	accccagggtg	cctcaccaga	tactactcca	gctttgtgaa	catggagagg	1680
gacctggcct	ctggcctgat	tggcccactg	ctcatctgct	acaaggagtc	tgtggaccag	1740
aggggaaacc	agatcatgtc	tgacaagagg	aatgtgattc	tgttctctgt	ctttgatgag	1800
aacaggagct	ggtacctgac	tgagaacatt	cagcgcttcc	tgcccaacct	tgctggggtg	1860
cagctggagg	accctgagtt	ccaggccagc	aacatcatgc	actccatcaa	tggtatgtg	1920
tttgacagcc	tccagcttcc	tgtctgctg	catgagggtg	cctactggta	cattctttct	1980
attggggccc	agactgaatt	cctttctgtc	ttcttctctg	gctacacctt	caaacacaag	2040
atggtgtatg	aggacacctt	gacctcttcc	ccattctctg	gggagactgt	gttcatgagc	2100
atggagaacc	ctggcctgtg	gattctggga	tgccacaact	ctgacttccg	caacaggggc	2160
atgactgccc	tgctcaaagt	ctcctcctgt	gacaagaaca	ctggggacta	ctatgaggac	2220
agctatgagg	acatctctgc	ctacctgtcc	agcaagaaca	ataccaccta	cgtgaaccgc	2280
tccctgagcc	agaatccacc	tgtcctgaaa	cgccaccaga	gggagatcac	caggaccacc	2340
ctccagtctg	accaggagga	gattgactat	gatgacacca	tttctgtgga	gatgaagaaa	2400
gaggactttg	acatctatga	cgaggacgag	aaccagagcc	caaggagctt	ccagaagaag	2460
accaggcact	acttcattgc	tgtctgtggg	cgctctgtgg	actatggcat	gagctccagc	2520
ccccatgtcc	tcaggaacag	ggcccagtct	ggctctgtgc	cacagttcaa	gaaagtggtc	2580
ttccaagagt	tactgatgag	cagcttccac	cagcccctgt	acagagggga	gctgaatgag	2640
cacctgggac	tcctggggccc	atacatcagg	gctgagggtg	aggacaacat	catggtgacc	2700
ttccgcaacc	aggcctccag	gccctacagc	ttctacagct	ccctcatcag	ctatgaggag	2760
gaccagaggg	agggggctga	gccacgcaag	aactttgtga	aacccaatga	aaccaagacc	2820
tacttctgga	aagtcagca	ccacatggcc	cccaccaagg	atgagtttga	ctgcaaggcc	2880
tgggcctaact	tctctgatgt	ggacctggag	aaggatgtgc	actctggcct	gattggccca	2940
ctcctggtct	gccacaccaa	cacctggaac	cctgcccctg	gaaggcaagt	gactgtgcag	3000
gagtttgccc	tcttcttcc	catctttgat	gaaaccaaga	gctgggtactt	cactgagAAC	3060
atggagcgca	actgcagggc	cccatgcaac	attcagatgg	aggacccacc	cttcaaagag	3120
aactaccgct	tccatgccat	caatggctac	atcatggaca	ccctgcctgg	gcttgtccatg	3180
gcccaggacc	agaggatcag	gtggtaacctg	ctttctatgg	gctccaatga	gaacattcac	3240
tccatccact	tctctgggca	tgtcttccact	gtgcgcaaga	aggaggagta	caagatggcc	3300
ctgtacaacc	tctacctggtg	ggtctttgag	actgtggaga	tgctgcccctc	caaagctggc	3360
atctggaggg	tggagtgcct	cattgggggag	cacctgcatg	ctggcatgag	cacctgttcc	3420
ctggtctaca	gcaacaagtg	ccagaccccc	ctgggaatgg	cctctggcca	catcagggac	3480
ttccagatca	ctgctctctg	ccagtatggc	cagtggggccc	ccaagctggc	caggctccac	3540
tactctggat	ccatcaatgc	ctggagcacc	aaggagccat	tcagctggat	caaagtggac	3600
ctgctggccc	ccatgatcat	ccatggcacc	aagaccaccg	gggcccaggca	gaagtctctcc	3660
agcctgtaca	tcagccagtt	catcatcatg	tacagcctgg	atggcaagaa	atggcagacc	3720
tacagaggca	actccactgg	aacctcctatg	gtcttctttg	gcaatgtgga	cagctctggc	3780
atcaagcaca	acatcttcaa	cccccaatc	atcgccagat	acatcaggct	gcacccacc	3840
cactacagca	tccgagcacc	cctcaggatg	gagctgatgg	gctgtgacct	gaactcctgc	3900

-continued

```

agcatgcccc tgggcatgga gagcaaggcc atttctgatg cccagatcac tgcctccage 3960
tacttcacca acatgtttgc cacctggagc ccaagcaagg ccaggctgca cctccaggga 4020
aggagcaatg cctggaggcc ccaggtaaac aaccctaaagg agtggtgca ggtggacttc 4080
cagaagacca tgaaggtcac tggggtgacc acccaggggg tcaagagcct gctcaccagc 4140
atgtatgtga aggagtctct gatcagctcc agccaggatg gccaccagtg gaccctcttc 4200
ttccagaatg gcaaggtaaa ggtgttccag ggcaaccagg acagcttcac cctgtggtg 4260
aacagcctgg accccccct cctgaccaga tacctgagga ttcacccca gagctgggtc 4320
caccagattg ccctgaggat ggaggtcctg ggatgtgagg cccaggacct gtactga 4377

```

<210> SEQ ID NO 99

<211> LENGTH: 4377

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 99

```

atgcagattg agctgagcac ctgcttcttc ctgtgctgc tgaggttctg cttctctgcc 60
accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120
ggcgagctgc ctgtggagcg caggttcccc cccagagtgc ccaagagctt ccccttaaac 180
acctcagtgg tgtacaagaa gaccctgttc gtggagtcca ccgaccacct gttcaacatc 240
gccaagccca ggccccctg gatgggctct ctgggccccca ccatccaggc cgaggtgtac 300
gacacctggt tgatcacct gaagaacatg gccagccacc ccgtgagcct gcacgctgtg 360
ggcgtgagct actggaaggc ctctgagggc gccgagatg acgaccagac cagccagagg 420
gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480
aaggagaaag gccccatggc cagcgacccc ctgtgctga cctacagcta cctgagccac 540
gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcaggag 600
ggcagcctgg ccaaggagaa gaccagacc ctgcacaagt tcctcctgct gttcgcctgtg 660
ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720
gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggtactgt gaacaggagc 780
ctgccccggc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840
accacccccg aggtgcacag catcttctct gagggccaca ccttctctgt gaggaaccac 900
aggcaggcca gcctggagat cagccccatc accttctga ccgcccagac cctgctgatg 960
gacctgggcc agttcctgct gtctctccac atcagcagcc accagcacga cggcatggag 1020
gcctactgta aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080
gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140
gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200
tgggtgcact acatcgccgc cgaggaggag gactgggact acgccccctt ggtgctggcc 1260
cccagcgaca ggagctacaa gagccagtac ctgaacaacg gccccagag gatcggcagg 1320
aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac caggagggcc 1380
atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacacctg 1440
ctgatcatct tcaagaacca ggccagcagg cctacaaca tctacccca cggcatcacc 1500

```

-continued

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560
cccatcctgc cggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620
accaagtctg accccagggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680
gacctggcct ctggcctgat cggccccctg ctgatctgct acaaggagag cgtggaccag 1740
aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgtatgag 1800
aacaggagct ggtatctgac cgagaacatc cagagggtcc tgcccaaccc cgccggcgtg 1860
cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggtacgtg 1920
ttcgacagcc tgcagctgtc tgtgtgctg cagcagggtg cctactggta catcctgagc 1980
atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040
atggtgtacg aggacacct gacctgttc ccttcagcg gcgagaccgt gttcatgagc 2100
atggagaacc cggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160
atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacagggac 2220
agctacgagg acatcagcgc ctacctgctg agcaagaaca acaccaccta cgtgaaccgc 2280
tccctgagcc agaaccccc cgtgctgaag aggcaccaga gggagatcac caggaccacc 2340
ctgcagagcg accaggagga gatcgactat gatgacacca tcagcgtgga gatgaagaag 2400
gaggacttcg acatctacga cgaggacgag aaccagagcc ccaggagctt ccagaagaag 2460
accaggcact acttcatcgc cgccgtggag aggctgtggg actatggcat gagcagcagc 2520
ccccactgtc tgaggaacag ggcccagagc ggcagcgtgc ccagttcaa gaaggtggtg 2580
ttccaggagt tcaccgacgg cagcttcacc cagccctgt acagaggcga gctgaacgag 2640
cacctgggcc tgctgggccc ctacatcagg gccgaggtgg aggacaacat catggtgacc 2700
tcagggaacc aggccagcag gcctacagc ttctacagca gcctgatcag ctacagggag 2760
gaccagaggg agggcgccga gccccaggaag aacttcgtga agcccaacga gaccaagacc 2820
tacttctgga aggtgcagca ccacatggcc cccaccaagg acgagttcga ctgcaaggcc 2880
tgggcctact tctctgatgt ggacctggag aaggacgtgc acagcgccct gatcgcccc 2940
ctgctggtgt gccacaccaa caccctgaac cccgcccacg gcaggcaggt gaccgtgacg 3000
gagttcggcc tgttcttcac catcttcgac gagaccaaga gctggtactt caccgagaac 3060
atggagagga actgcagggc cccctgcaac atccagatgg aggacccac cttcaaggag 3120
aactacaggt tccacgccat caacggctac atcatggaca ccctgcccgg cctggtgatg 3180
gcccaggacc agaggatcag gtggtatctg ctgagcatgg gcagcaacga gaacatccac 3240
agcatccact tcagcggcca cgtgttcacc gtgaggaaga aggaggagta caagatggcc 3300
ctgtacaacc tgtaccccg cgtgttcgag accgtggaga tgctgcccag caaggccggc 3360
atctggaggg tggagtgcct gatcggcgag cacctgcacg ccggcatgag caccctgttc 3420
ctggtgtaca gcaacaagtg ccagaccccc ctgggcatgg ccagcggcca catcagggac 3480
ttccagatca ccgctctgg ccagtacggc cagtgggccc ccaagctggc caggctgcac 3540
tacagcggca gcatcaacgc ctggagcacc aaggagccct tcagctggat caaggtggac 3600
ctgctggccc ccatgatcat ccacggcatc aagaccagg gcgccaggca gaagttcagc 3660
agcctgtaca tcagccagtt catcatcatg tacagcctgg acggcaagaa gtggcagacc 3720
tacaggggca acagcacgg caccctgatg gtgttcttcg gcaacgtgga cagcagcggc 3780

-continued

```

atcaagcaca acatcttcaa cccccccatc atcgccagggt acatcagggt gcacccccacc 3840
cactacagca tcaggagcac cctgcggtatg gaactgatgg gctgogacct gaacagctgc 3900
agcatgcccc tgggcatgga gagcaaggcc atctctgacg cccagatcac cgccagcagc 3960
tacttcacca acatgttcgc cacctggagc cccagcaagg ccaggtgca cctgcagggc 4020
aggagcaacg cctggaggcc ccaggtgaac aacccaagg agtggctgca ggtggacttc 4080
cagaagacca tgaaggtgac cggcgtgacc acccaggcgc tgaagagcct gctgaccagc 4140
atgtacgtga aggagttoct gatcagcagc agccaggacg gccaccagtg gaccctgttc 4200
ttccagaacg gcaaagtga ggtgttcag ggcaaccagg acagcttcac ccccggtgtg 4260
aacagcctgg accccccct gctgaccagg tatctgagga tccaccccc gagctgggtg 4320
caccagatcg ccctgagaat ggaagtgtg ggatgaggg cccaggacct gtactga 4377

```

<210> SEQ ID NO 100

<211> LENGTH: 4377

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 100

```

atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60
accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120
ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180
acctcagtgg tgtacaagaa gaccctgttc gtggagtcca ccgaccacct gttcaacatc 240
gccaagcccc ggccccctg gatgggctc ctgggcccc ccatccaggc cgaggtgtac 300
gacaccgtgg tgatcacctt gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360
ggcgtgagct actggaaggc ctctgagggc gccagatag acgaccagac cagccagagg 420
gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480
aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacagcta cctgagccac 540
gtggacctgg tgaaggacct gaactctggc ctgatcgccg ccctgctggt gtgcagggag 600
ggcagcctgg ccaaggagaa gaccagacc ctgcacaagt tcatcctgct gttcgccgtg 660
ttcgatgagg gcaagagctg gcacagcag accaagaaca gcctgatgca ggacagggat 720
gccgcctctg ccagggcctg gccaagatg cacaccgtga acggctacgt gaacaggagc 780
ctgccccggc tgatcgctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840
accacccccg aggtgcacag catcttctct gagggccaca ccttctggt gaggaaccac 900
aggcaggcca gcctggagat cagccccatc accttctgga ccgcccagac cctgctgatg 960
gacctgggcc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag 1020
gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080
gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140
gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca cccaagacc 1200
tgggtgcact acatcgccgc cgaggaggag gactgggact acgccccct ggtgctggcc 1260
cccagcagca ggagctacaa gagccagtac ctgaacaacg gccccagag gatcggcagg 1320

```

-continued

aagtacaaga	aggtcagatt	catggcctac	accgacgaga	ccttcaagac	cagggaggcc	1380
atccagcaag	agtctggcat	cctgggcccc	ctgctgtacg	gcgaggtggg	cgacaccctg	1440
ctgatcatct	tcaagaacca	ggccagcagg	ccctacaaca	tctaccccc	cgcatcacc	1500
gatgtgaggc	ccctgtacag	caggaggctg	cccaagggcg	tgaagcacct	gaaggacttc	1560
cccatoctgc	ccggcgagat	cttcaagtac	aagtggaccg	tgaccgtgga	ggatggcccc	1620
accaagtctg	accccagggtg	cctgaccagg	tactacagca	gcttcgtgaa	catggagagg	1680
gacctggcct	ctggcctgat	cgccccctg	ctgatctgct	acaaggagag	cgtggaccag	1740
aggggcaacc	agatcatgtc	tgacaagagg	aacgtgatcc	tgttctctgt	gttcgatgag	1800
aacaggagct	ggatctgtac	cgagaacatc	cagaggttcc	tgccccaccc	cgccggcgtg	1860
cagctggagg	accccaggtt	ccaggccagc	aacatcatgc	acagcatcaa	cggtacgtg	1920
ttcgacagcc	tgacagctgc	tgtgtgcctg	cacgaggtgg	cctactggta	catcctgagc	1980
atcggcgccc	agaccgactt	cctgtctgtg	ttcttctctg	gctacacctt	caagcacaag	2040
atggtgtaag	aggacacctt	gacctgttc	cccttcagcg	gcgagaccgt	gttcgatgagc	2100
atggagaacc	ccggcctgtg	gatcctgggc	tgccacaaca	gcgacttcag	gaacaggggc	2160
atgaccgccc	tgctgaaagt	cagcagctgc	gacaagaaca	ccggcgacta	ctacgaggac	2220
agctacgagg	acatcagcgc	ctacctgctg	agcaagaaca	acaccacctt	cgtgaaccgc	2280
tccctgagcc	agaaccccc	cgtgctgaag	aggcaccaga	gggagatcac	caggaccacc	2340
ctgcagagcg	accaggagga	gatcgactat	gatgacacca	tcagcgtgga	gatgaagaag	2400
gaggacttcg	acatctacga	cgaggacgag	aaccagagcc	ccaggagctt	ccagaagaag	2460
accaggcact	acttcatcgc	cgccgtggag	aggctgtggg	actatggcat	gagcagcagc	2520
ccccacgtgc	tgaggaacag	ggcccagagc	ggcagcgtgc	cccagttcaa	gaaggtggtg	2580
ttccaggagt	tcaccgacgg	cagcttcacc	cagccoctgt	acagaggcga	gctgaacgag	2640
cacctgggcc	tgctgggccc	ctacatcagg	ggcaggtggg	aggacaacat	catggtgacc	2700
ttcaggaacc	aggccagcag	gcccacagc	ttctacagca	gcctgatcag	ctacgaggag	2760
gaccagaggc	agggcgccga	gcccaggaag	aacttcgtga	agcccaacga	gaccaagacc	2820
tacttctgga	aggtgcagca	ccacatggcc	cccaccaagg	acgagttcga	ctgcaaggcc	2880
tgggcctact	tctctgatgt	ggacctggag	aaggacgtgc	acagcggcct	gatcggcccc	2940
ctgctggtgt	gccacaccaa	cacctgaa	cccgccacg	gcaggcaggt	gaccgtgcag	3000
gagttcgcct	tgttcttcac	catcttcgac	gagaccaaga	gctggtactt	caccgagaac	3060
atggagagga	actgcagggc	cccctgcaac	atccagatgg	aggaccccac	cttcaaggag	3120
aactacaggt	tccacgcat	caacggctac	atcatggaca	ccctgccccg	cctggtgatg	3180
gcccaggacc	agaggatcag	gtggatctg	ctgagcatgg	gcagcaacga	gaacatccac	3240
agcatccact	tcagcgcca	cgtgttcacc	gtgaggaaga	aggaggagta	caagatggcc	3300
ctgtacaacc	tgtacccccg	cgtgttcgag	accgtggaga	tgctgcccag	caaggccggc	3360
atctggaggg	tggagtgcct	gatcggcgag	cacctgcacg	ccggcatgag	cacctgttc	3420
ctggtgtaca	gcaacaagtg	ccagaccccc	ctgggcatgg	ccagcggcca	catcagggac	3480
ttccagatca	ccgctctg	ccagtaggc	cagtgggccc	ccaagctggc	caggctgcac	3540
tacagcgcca	gcatcaacgc	ctggagcacc	aaggagccct	tcagctggat	caaggtggac	3600

-continued

ctgctggccc	ccatgatcat	ccacggcacc	aagacccagg	gcgccaggca	gaagttcagc	3660
agcctgtaca	tcagccagtt	catcatcatg	tacagcctgg	acggcaagaa	gtggcagacc	3720
tacaggggca	acagcaccgg	caccctgatg	gtgttcttcg	gcaacgtgga	cagcagcggc	3780
atcaagcaca	acatcttcaa	ccccccatc	atcgccaggt	acatcaggct	gcaccccacc	3840
cactacagca	tcaggagcac	cctgcggtatg	gaactgatgg	gctgcgacct	gaacagctgc	3900
agcatgcccc	tgggcatgga	gagcaaggcc	atctctgacg	cccagatcac	cgccagcagc	3960
tacttcacca	acatgttcgc	cacctggagc	cccagcaagg	ccaggctgca	cctgcagggc	4020
aggagcaacg	cctggaggcc	ccaggtgaac	aacccaagg	agtggctgca	ggtggacttc	4080
cagaagacca	tgaaggtgac	cgcgctgacc	accaggggcg	tgaagagcct	gctgaccagc	4140
atgtacgtga	aggagttoct	gatcagcagc	agccaggacg	gccaccagtg	gaccctgttc	4200
ttccagaacg	gcaagtga	ggtgttcocag	ggcaaccagg	acagcttcac	ccccgtggtg	4260
aacagcctgg	accccccoct	gctgaccagg	tatctgagga	tccaccccca	gagctgggtg	4320
caccagatcg	ccctgagaat	ggaagtgtg	ggatgagagg	cccaggacct	gtactga	4377

<210> SEQ ID NO 101

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 101

atgcagattg	agctgagcac	ctgcttcttc	ctgtgctgc	tgaggttctg	cttctctgcc	60
accaggagat	actacctggg	cgccgtggag	ctgagctggg	actacatgca	gtctgacctg	120
ggcgagctgc	ctgtggacgc	cagggtcccc	cccagagtgc	ccaagagctt	ccccttcaac	180
acctcagtgg	tgtacaagaa	gacctgttc	gtggagtcca	ccgaccacct	gttcaacatc	240
gccaagccca	ggccccctg	gatgggctg	ctgggcccc	ccatccaggc	cgaggtgtac	300
gacaccgtgg	tggtcacct	gaagaacatg	gccagccc	ccgtgagcct	gcacgccgtg	360
ggcgtgagct	actggaagtc	ctctgagggc	gccgagtatg	acgaccagac	cagccagagg	420
gagaaggagg	acgacaaggt	gttccccggc	aagagccaca	cctacgtgtg	gcaggtgctg	480
aaggagaacg	gccccactgc	cagcgacccc	ccctgcctga	cctacagcta	cctgagccac	540
gtggacctgg	tgaaggacct	gaactctggc	ctgatcggcg	ccctgctggt	gtgcagggag	600
ggcagcctgg	ccaaggagaa	gacctcagacc	ctgcacaagt	tcatcctgct	gttcgocgtg	660
ttcgatgagg	gcaagagctg	gcacagcag	accaagaaca	gcctgatgca	ggacagggat	720
gccgcctctg	ccagggcctg	gcccgaagt	cacaccgtga	acggctacgt	gaacagggagc	780
ctgccccggc	tgatcggctg	ccacaggaag	tctgtgtact	ggcacgtgat	cgccatgggc	840
accacccccg	aggtgcacag	catcttctctg	gagggccaca	ccttctctggt	gaggaaccac	900
aggcaggcca	gcctggagat	cagccccatc	accttctga	ccgcccagac	cctgctgatg	960
gacctgggccc	agttcctgct	gttctgccac	atcagcagcc	accagcacga	cggcatggag	1020
gcctacgtga	aggtggacag	ctgccccgag	gagccccagc	tgaggatgaa	gaacaacgag	1080
gaggccgagg	actatgatga	tgacctgacc	gactctgaga	tggacgtggt	gaggtttgat	1140
gatgacaaca	gccccagctt	catccagatc	aggtctgtgg	ccaagaagca	ccccaaagacc	1200

-continued

tgggtgcaact acatcgccgc cgaggaggag gactgggact acgccccct ggtgctggcc 1260
cccgacgaca ggagctacaa gagccagtac ctgaacaacg gccccagag gatcggcagg 1320
aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380
atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgagggtgg cgacaccctg 1440
ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccc cggeatcacc 1500
gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560
cccatcctgc cggcgagat cttcaagtac aagtggacgg tgaccgtgga ggatggcccc 1620
accaagtctg accccagggt cctgaccagg tactacagca gcttcgtgaa catggagagg 1680
gacctggcct ctggcctgat cggccccctg ctgatctgct acaaggagag cgtggaccag 1740
aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800
aacaggagct ggtatctgac cgagaacatc cagaggttcc tgccccaccc cgccggcgtg 1860
cagctggagg accccaggtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920
ttcgacagcc tgcagctgtc tgtgtgctg cagcagggtg cctactggta catcctgagc 1980
atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040
atggtgtacg aggacaccct gaccctgttc cccctcagcg gcgagaccgt gttcatgagc 2100
atggagaacc cggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160
atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacaggagc 2220
agctacgagg acatcagcgc ctacctgtg agcaagaaca acgccatcga gcccaggagc 2280
ttcagccaga accccccctg gctgaagagg caccagaggg agatcaccag gaccaccctg 2340
cagagcgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag 2400
gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc 2460
aggcactact tcatcgccgc cgtggagagg ctgtgggact atggcatgag cagcagcccc 2520
cacgtgctga ggaacagggc ccagagcggc agcgtgcccc agttcaagaa ggtggtgttc 2580
caggagttea ccgacggcag cttcaccag cccctgtaca gaggcgagct gaacgagcac 2640
ctgggcctgc tgggccccca catcagggcc gaggtggagg acaacatcat ggtgaccttc 2700
aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta caggaggagc 2760
cagaggcagg gcgccgagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac 2820
ttctggaagg tgcagcacca catggcccc accaaggacg agttcgactg caaggcctgg 2880
gcctacttct ctgatgtgga cctggagaag gacgtgcaca gcggcctgat cgccccctg 2940
ctggtgtgcc acaccaacac cctgaacccc gccacggca ggcagggtgac cgtgcaggag 3000
ttcgccctgt tcttcaccat cttcgacgag accaagagct ggtacttcac cgagaacatg 3060
gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac 3120
tacaggttcc acgccatcaa cggctacatc atggaccccc tgcccggcct ggtgatggcc 3180
caggaccaga ggatcagggt gtatctgctg agcatgggca gcaacgagaa catccacagc 3240
atccacttca cggcccacgt gttcaccgtg aggaagaagg aggagtacaa gatggccctg 3300
tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccggcacc 3360
tggaggggtg agtgccctgat cggcgagcac ctgcacgccc gcatgagcac cctgttctctg 3420
gtgtacagca acaagtgcc aacccccctg ggcattggcca gcggccacat cagggacttc 3480

-continued

cagatcacgg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac 3540
agcggcagca tcaacgcctg gaccaccaag gagcccttca gctggatcaa ggtggactg 3600
ctggccccca tgatcatcca cggcatcaag acccagggcg ccaggcagaa gttcagcagc 3660
ctgtacatca gccagtteat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720
aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc 3780
aagcacaaca tcttcaaccc ccccatcacc gccaggtaga tcaggctgca cccccaccac 3840
tacagcatca ggagcacctc gcggatggaa ctgatgggct gcgacctgaa cagctgcagc 3900
atgccctctg gcctggagag caaggccacc tctgacgccc agatcacccg cagcagctac 3960
ttcaccaca tgttcgcacc ctggagcccc agcaaggcca ggctgcacct gcagggcagg 4020
agcaacgcct ggaggccccca ggtgaacaac cccaaggagt ggctgcagggt ggacttcag 4080
aagaccatga aggtgaccgg cgtgaccacc cagggcgtga agagcctgct gaccagcatg 4140
tacgtgaagg agttcctgat cagcagcagc caggacggcc accagtggac cctgttcttc 4200
cagaacggca aagtgaaggt gttccagggc aaccaggaca gcttcacccc cgtggtgaa 4260
agcctggacc cccccctgct gaccaggat ctgaggatcc acccccagag ctgggtgcac 4320
cagatcgccc tgagaatgga agtgcctggga tgcgaggccc aggacctgta ctga 4374

<210> SEQ ID NO 102

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 102

atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60
accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120
ggcgagctgc ctgtggaagc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180
acctcagtgg tgtacaagaa gacctgttc gtggagtcca ccgaccacct gttcaacatc 240
gccaagcccc ggccccctg gatgggctct ctgggccccca ccattccaggc cgaggtgtac 300
gacacctggt tgatcacctc gaagaacatg gccagccacc ccgtgagcct gcacgacctg 360
ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg 420
gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480
aaggagaacg gccccatggc cagcgaacccc ctgtgcctga cctacagcta cctgagccac 540
gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcaggag 600
ggcagcctgg ccaaggagaa gaccagacc ctgcacaagt tcattcctgct gttcgcctg 660
ttcgtatgag gcaagagctg gcacagcag accaagaaca gcctgatgca ggacagggat 720
gcccctctg ccagggctg gcccagatg cacacctga acggctacgt gaacaggagc 780
ctgcccgccc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840
accacccccg aggtgcacag catcttctct gagggccaca ccttctggt gaggaaccac 900
aggcaggcca gcctggagat cagccccacc accttctgta ccgcccagac cctgctgatg 960
gacctgggcc agttcctgct gtccctgccac atcagcagcc accagcacga cggcatggag 1020

-continued

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag	1080
gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat	1140
gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc	1200
tgggtgcaact acatcgccgc cgaggaggag gactgggact acgccccctt ggtgctggcc	1260
cccgacgaca ggagctacaa gagccagtac ctgaacaacg gccccagag gatcggcag	1320
aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac caggaggacc	1380
atccagcaag agtctggcat cctgggcccc ctgctgtacg gcgagggtgg cgacacctg	1440
ctgatcatct tcaagaacca ggccagcagg cctacaaca tctaccccc cgcatcacc	1500
gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc	1560
cccatcctgc ccggcgagat cttcaagtae aagtggaccg tgaccgtgga ggatggcccc	1620
accaagctctg accccagggtg cctgaccagg tactacagca gcttcgtgaa catggagagg	1680
gacctggcct ctggcctgat cggccccctg ctgatctgct acaaggagag cgtggaccag	1740
aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcagatgag	1800
aacaggagct ggtatctgac cgagaacatc cagagggtcc tgcccaaccc cgcggcgtg	1860
cagctggagg acccccagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg	1920
ttcgacagcc tgcagctgtc tgtgtgctg cacgagggtg cctactggta catcctgagc	1980
atcggcgcce agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag	2040
atggtgtacg aggacacctt gacctgttc ccttcagcg gcgagaccgt gttcatgagc	2100
atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc	2160
atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacaggac	2220
agctacgagg acatcagcgc ctacctgtc agcaagaaca acgcatcga gccaggagc	2280
ttcagccaga accccccctg gctgaagagg caccagaggg agatcaccag gaccacctg	2340
cagagcgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag	2400
gacttcgaca tctacgaca ggacgagaac cagagcccca ggagcttcca gaagaagacc	2460
aggcactact tcatcgccgc cgtggagagg ctgtgggact atggcatgag cagcagcccc	2520
cacgtgctga ggaacagggc ccagagcggc agcgtgcccc agttcaagaa ggtggtgttc	2580
caggagttca ccgacggcag cttcaccag cccctgtaca gaggcgagct gaacgagcac	2640
ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc	2700
aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac	2760
cagaggcagg gcgcccagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac	2820
ttctggaagg tgcagcacca catggcccc accaaggacg agttcagctg caaggcctgg	2880
gcctacttct ctgatgtgga cctggagaag gacgtgcaca gcggcctgat cggccccctg	2940
ctggtgtgcc acaccaaac cctgaacccc gccacggca ggcagggtgac cgtgcaggag	3000
ttcgccctgt tcttcacat cttcgacgag accaagagct ggtacttcac cgagaacatg	3060
gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac	3120
tacaggttcc acgcatcaa cggctacatc atggacaccc tgcccggcct ggtgatggcc	3180
caggaccaga ggatcagggtg gtatctgctg agcatgggca gcaacgagaa catccacagc	3240
atccacttca gcggccacgt gttcacctg aggaagaagg aggagtaca gatggccctg	3300

-continued

tacaacctgt	acccccgggt	gttcgagacc	gtggagatgc	tgcccagcaa	ggccggcate	3360
tggaggggtg	agtgccctgat	cgcgagcac	ctgcacgccg	gcatgagcac	cctgttctctg	3420
gtgtacagca	acaagtgcc	gacccccctg	ggcatggcca	gcgccacat	cagggacttc	3480
cagatcaccg	cctctggcca	gtacggccag	tgggccccca	agctggccag	gctgcactac	3540
agcggcagca	tcaacgcctg	gagcaccaag	gagcccttca	gctggatcaa	ggtggacctg	3600
ctggccccca	tgatcatcca	cgcatcaag	accagggcg	ccaggcagaa	gttcagcagc	3660
ctgtacatca	gccagtctcat	catcatgtac	agcctggacg	gcaagaagtg	gcagacctac	3720
aggggcaaca	gcaccggcac	cctgatggtg	ttcttcggca	acgtggacag	cagcggcatc	3780
aagcacaaca	tcttcaaccc	ccccatcctc	gccaggtaca	tcaggctgca	ccccaccac	3840
tacagcatca	ggagcacctc	gcgatggaa	ctgatgggct	gcgacctgaa	cagctgcagc	3900
atgcccctgg	gcatggagag	caaggccatc	tctgacgccc	agatcacccg	cagcagctac	3960
ttcaccaca	tgttcgccac	ctggagcccc	agcaaggcca	ggctgcacct	gcagggcagg	4020
agcaacgcct	ggaggccccca	ggtgaacaac	cccaaggagt	ggctgcaggt	ggacttccag	4080
aagaccatga	aggtgaccgg	cgtgaccacc	cagggcgtga	agagcctgct	gaccagcatg	4140
tacgtgaagg	agttcctgat	cagcagcagc	caggacggcc	accagtggac	cctgttcttc	4200
cagaacggca	aagtgaaggt	gttccagggc	aaccaggaca	gcttcacccc	cgtggtgaac	4260
agcctggacc	ccccctgct	gaccaggtat	ctgaggatcc	acccccagag	ctgggtgcac	4320
cagatcgccc	tgagaatgga	agtgtctggga	tgcgaggccc	aggacctgta	ctga	4374

<210> SEQ ID NO 103

<211> LENGTH: 4377

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 103

atgcagattg	agctgagcac	ctgcttcttc	ctgtgcctgc	tgaggttctg	cttctctgcc	60
accaggagat	actacctggg	cgccgtggag	ctgagctggg	actacatgca	gtctgacctg	120
ggcgagctgc	ctgtggaagc	caggttcccc	cccagagtgc	ccaagagctt	cccttcaac	180
acctcagtgg	tgtacaagaa	gaccctgttc	gtggagtcca	ccgaccacct	gttcaacatc	240
gccaagcccc	ggccccctg	gatgggctctg	ctgggccccca	ccatccaggc	cgaggtgtac	300
gacaccgtgg	tggtcacctc	gaagaacatg	gccagccacc	ccgtgagcct	gcacgccgtg	360
ggcgtgagct	actggaagtc	ctctgagggc	gccagatgatg	acgaccagac	cagccagagg	420
gagaaggagg	acgacaaggt	gttccccggc	aagagccaca	cctacgtgtg	gcaggtgctg	480
aaggagaacg	gccccactgc	cagcgacccc	ccctgctctga	cctacagcta	cctgagccac	540
gtggacctgg	tgaaggacct	gaactctggc	ctgatcggcg	ccctgctggt	gtgcaggag	600
ggcagcctgg	ccaaggagaa	gaccagacc	ctgcacaagt	tcctcctgct	gttcgcccgtg	660
ttcgtatgag	gcaagagctg	gcacagcgag	accaagaaca	gcctgatgca	ggacagggat	720
gcccctctg	ccagggcctg	gccccagatg	cacaccgtga	acggctacgt	gaacaggagc	780
ctgccccggc	tgatcggctg	ccacaggaag	tctgtgtact	ggcacgtgat	cggcatgggc	840
accacccccg	aggtgcacag	catcttctctg	gagggccaca	ccttctctggt	gaggaaccac	900

-continued

aggcaggcca gcctggagat cagcccacac accttctga cgcaccagac cctgctgatg	960
gacctgggccc agttcctgct gttctgccac atcagcagcc accagcacga cggeatggag	1020
gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag	1080
gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat	1140
gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca cccaagacc	1200
tgggtgcact acatcgccgc cgaggaggag gactgggact acgccccct ggtgctggcc	1260
cccgcagaca ggagctacaa gagccagtac ctgaacaacg gccccagag gatcggcagg	1320
aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac caggagggcc	1380
atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg	1440
ctgatcatct tcaagaacca ggccagcagg cctacaaca tctaccccc cgcatcacc	1500
gatgtgaggc ccctgtacag caggaggctg cccaaggcgg tgaagcacct gaaggacttc	1560
cccctcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc	1620
accaagtctg accccagggt cctgaccagg tactacagca gcttcgtgaa catggagagg	1680
gacctggcct ctggcctgat cggccccctg ctgatctgct acaaggagag cgtggaccag	1740
aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcogatgag	1800
aacaggagct ggtatctgac cgagaacac cagaggttcc tgccccaccc cgccggcgtg	1860
cagctggagg accccagatt ccaggccagc aacatcatgc acagatcaa cggctacgtg	1920
ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc	1980
atcgccgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag	2040
atggtgtaag aggcaccctt gaccctgttc cccttcagcg gcgagaccgt gttcatgagc	2100
atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc	2160
atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacaggagc	2220
agctacgagg acatcagcgc ctacctgctg agcaagaaca acaccaccta cgtgaaccgc	2280
tccctgagcc agaaccccc cgtgctgaag aggcaccaga gggagatcac caggaccacc	2340
ctgcagagcg accaggagga gatcgactat gatgacacca tcagcgtgga gatgaagaag	2400
gaggacttcc acatctacga cgaggacgag aaccagagcc ccaggagctt ccagaagaag	2460
accaggcact acttcctcgc cgccgtggag aggtgtggg actatggcat gagcagcagc	2520
ccccacgtgc tgaggaacag ggcccagagc ggcagcgtgc ccagttcaa gaaggtggtg	2580
ttccaggagt tcaccgacgg cagcttcacc cagcccctgt acagaggcga gctgaacgag	2640
cacctgggcc tgctgggccc ctacatcagg gccgaggtgg aggacaacat catggtgacc	2700
ttcaggaacc agggccagcag gccctacagc ttctacagca gcctgatcag ctacaggag	2760
gaccagaggc agggcgccga gcccaggaag aacttcgtga agccccacga gaccaagacc	2820
tacttctgga aggtgcagca ccacatggcc cccaccaagg acgagttcga ctgcaaggcc	2880
tgggcctact tctctgatgt ggacctggag aaggacgtgc acagcggcct gatcgcccc	2940
ctgctggtgt gccacaccaa caccctgaac cccgccacg gcaggcaggt gaccgtgcag	3000
gagttcgcgc tgttcttcac catcttcgac gagaccaaga gctggctactt caccgagaac	3060
atggagagga actgcagggc cccctgcaac atccagatgg aggacccac cttcaaggag	3120
aactacaggt tccacgcoat caacggctac atcatggaca ccctgccccg cctggtgatg	3180

-continued

```

gcccaggacc agaggatcag gtggtatctg ctgagcatgg gcagcaacga gaacatccac 3240
agcatccaact tcagcggcca cgtgttcacc gtgaggaaga aggaggagta caagatggcc 3300
ctgtacaacc tgtaccccggt cgtgttcgag accgtggaga tgctgcccag caaggccggc 3360
atctggagggg tggagtgcct gatcggcgag cacctgcacg ccggcatgag caccctgttc 3420
ctggtgtaca gcaacaagtg ccagaccccc ctgggcatgg ccagcggcca catcagggac 3480
ttccagatca ccgctctgg ccagtcaggc cagtggggccc ccaagctggc caggctgcac 3540
tacagcggca gcatcaacgc ctggagcacc aaggagccct tcagctggat caaggtggac 3600
ctgctggccc ccatgatcat ccacggcatc aagacccagg gcgccaggca gaagttcagc 3660
agcctgtaca tcagccagtt catcatcatg tacagcctgg acggcaagaa gtggcagacc 3720
tacaggggca acagcaccgg caccctgatg gtgttcttcg gcaacgtgga cagcagcggc 3780
atcaagcaca acatcttcaa cccccccatc atcgccaggt acatcaggt gcaccccacc 3840
cactacagca tcaggagcac cctgcggatg gaactgatgg gctgcgacct gaacagctgc 3900
agcatgcccc tgggcatgga gagcaaggcc atctctgacg cccagatcac cgccagcagc 3960
tacttcacca acatgttcgc cacctggagc cccagcaagg ccaggctgca cctgcagggc 4020
aggagcaacg cctggaggcc ccaggtgaac aaccccaagg agtggctgca ggtggacttc 4080
cagaagacca tgaaggtgac cggcgtgacc acccagggcg tgaagagcct gctgaccagc 4140
atgtactgta aggagtctct gatcagcagc agccaggacg gccaccagtg gaccctgttc 4200
ttccagaacg gcaaaagtga ggtgttcocag ggcaaccagg acagcttcac ccccggtgtg 4260
aacagcctgg accccccctc gctgaccagg tatctgagga tccaccccca gagctgggtg 4320
caccagatcg ccctgagaat ggaagtgtg ggatgcgagg cccaggacct gtactga 4377

```

<210> SEQ ID NO 104

<211> LENGTH: 1458

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 104

```

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
 1             5             10             15
Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
 20             25             30
Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
 35             40             45
Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
 50             55             60
Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
 65             70             75             80
Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
 85             90             95
Ala Glu Val Tyr Asp Thr Val Val Val Thr Leu Lys Asn Met Ala Ser
 100            105            110
His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ser Ser
 115            120            125

```

-continued

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
 130 135 140
 Asp Lys Val Phe Pro Gly Lys Ser His Thr Tyr Val Trp Gln Val Leu
 145 150 155 160
 Lys Glu Asn Gly Pro Thr Ala Ser Asp Pro Pro Cys Leu Thr Tyr Ser
 165 170 175
 Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
 180 185 190
 Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
 195 200 205
 Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
 210 215 220
 Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
 225 230 235 240
 Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
 245 250 255
 Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
 260 265 270
 Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
 275 280 285
 Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
 290 295 300
 Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
 305 310 315 320
 Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
 325 330 335
 Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
 340 345 350
 Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
 355 360 365
 Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
 370 375 380
 Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
 385 390 395 400
 Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
 405 410 415
 Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
 420 425 430
 Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
 435 440 445
 Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
 450 455 460
 Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
 465 470 475 480
 Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
 485 490 495
 His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
 500 505 510
 Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
 515 520 525
 Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp

-continued

530					535					540					
Pro	Arg	Cys	Leu	Thr	Arg	Tyr	Tyr	Ser	Ser	Phe	Val	Asn	Met	Glu	Arg
545					550					555					560
Asp	Leu	Ala	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Ile	Cys	Tyr	Lys	Glu
				565					570					575	
Ser	Val	Asp	Gln	Arg	Gly	Asn	Gln	Ile	Met	Ser	Asp	Lys	Arg	Asn	Val
			580					585					590		
Ile	Leu	Phe	Ser	Val	Phe	Asp	Glu	Asn	Arg	Ser	Trp	Tyr	Leu	Thr	Glu
		595					600					605			
Asn	Ile	Gln	Arg	Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp
610					615					620					
Pro	Glu	Phe	Gln	Ala	Ser	Asn	Ile	Met	His	Ser	Ile	Asn	Gly	Tyr	Val
625					630					635					640
Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	Cys	Leu	His	Glu	Val	Ala	Tyr	Trp
				645					650					655	
Tyr	Ile	Leu	Ser	Ile	Gly	Ala	Gln	Thr	Asp	Phe	Leu	Ser	Val	Phe	Phe
		660						665					670		
Ser	Gly	Tyr	Thr	Phe	Lys	His	Lys	Met	Val	Tyr	Glu	Asp	Thr	Leu	Thr
		675					680					685			
Leu	Phe	Pro	Phe	Ser	Gly	Glu	Thr	Val	Phe	Met	Ser	Met	Glu	Asn	Pro
690					695					700					
Gly	Leu	Trp	Ile	Leu	Gly	Cys	His	Asn	Ser	Asp	Phe	Arg	Asn	Arg	Gly
705				710						715					720
Met	Thr	Ala	Leu	Leu	Lys	Val	Ser	Ser	Cys	Asp	Lys	Asn	Thr	Gly	Asp
				725					730					735	
Tyr	Tyr	Glu	Asp	Ser	Tyr	Glu	Asp	Ile	Ser	Ala	Tyr	Leu	Leu	Ser	Lys
		740						745				750			
Asn	Asn	Thr	Thr	Tyr	Val	Asn	Arg	Ser	Leu	Ser	Gln	Asn	Pro	Pro	Val
		755					760					765			
Leu	Lys	Arg	His	Gln	Arg	Glu	Ile	Thr	Arg	Thr	Thr	Leu	Gln	Ser	Asp
770				775						780					
Gln	Glu	Glu	Ile	Asp	Tyr	Asp	Asp	Thr	Ile	Ser	Val	Glu	Met	Lys	Lys
785				790						795					800
Glu	Asp	Phe	Asp	Ile	Tyr	Asp	Glu	Asp	Glu	Asn	Gln	Ser	Pro	Arg	Ser
				805					810					815	
Phe	Gln	Lys	Lys	Thr	Arg	His	Tyr	Phe	Ile	Ala	Ala	Val	Glu	Arg	Leu
				820				825					830		
Trp	Asp	Tyr	Gly	Met	Ser	Ser	Ser	Pro	His	Val	Leu	Arg	Asn	Arg	Ala
		835					840					845			
Gln	Ser	Gly	Ser	Val	Pro	Gln	Phe	Lys	Lys	Val	Val	Phe	Gln	Glu	Phe
850					855					860					
Thr	Asp	Gly	Ser	Phe	Thr	Gln	Pro	Leu	Tyr	Arg	Gly	Glu	Leu	Asn	Glu
865				870						875					880
His	Leu	Gly	Leu	Leu	Gly	Pro	Tyr	Ile	Arg	Ala	Glu	Val	Glu	Asp	Asn
				885					890					895	
Ile	Met	Val	Thr	Phe	Arg	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Ser	Phe	Tyr
			900					905					910		
Ser	Ser	Leu	Ile	Ser	Tyr	Glu	Glu	Asp	Gln	Arg	Gln	Gly	Ala	Glu	Pro
		915						920				925			
Arg	Lys	Asn	Phe	Val	Lys	Pro	Asn	Glu	Thr	Lys	Thr	Tyr	Phe	Trp	Lys
930					935					940					

-continued

Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala
 945 950 955 960
 Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly
 965 970 975
 Leu Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala
 980 985 990
 His Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile
 995 1000 1005
 Phe Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg
 1010 1015 1020
 Asn Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe
 1025 1030 1035
 Lys Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp
 1040 1045 1050
 Thr Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp
 1055 1060 1065
 Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His
 1070 1075 1080
 Phe Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys
 1085 1090 1095
 Met Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu
 1100 1105 1110
 Met Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile
 1115 1120 1125
 Gly Glu His Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr
 1130 1135 1140
 Ser Asn Lys Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile
 1145 1150 1155
 Arg Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala
 1160 1165 1170
 Pro Lys Leu Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp
 1175 1180 1185
 Ser Thr Lys Glu Pro Phe Ser Trp Ile Lys Val Asp Leu Leu Ala
 1190 1195 1200
 Pro Met Ile Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys
 1205 1210 1215
 Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu
 1220 1225 1230
 Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr
 1235 1240 1245
 Leu Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His
 1250 1255 1260
 Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His
 1265 1270 1275
 Pro Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met
 1280 1285 1290
 Gly Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser
 1295 1300 1305
 Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr
 1310 1315 1320

-continued

Asn Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu
 1325 1330 1335
 Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys
 1340 1345 1350
 Glu Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly
 1355 1360 1365
 Val Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val
 1370 1375 1380
 Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr
 1385 1390 1395
 Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln
 1400 1405 1410
 Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu
 1415 1420 1425
 Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile
 1430 1435 1440
 Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
 1445 1450 1455

<210> SEQ ID NO 105

<211> LENGTH: 1458

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 105

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
 1 5 10 15
 Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
 20 25 30
 Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
 35 40 45
 Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
 50 55 60
 Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
 65 70 75 80
 Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
 85 90 95
 Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
 100 105 110
 His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
 115 120 125
 Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
 130 135 140
 Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
 145 150 155 160
 Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
 165 170 175
 Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
 180 185 190
 Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
 195 200 205

-continued

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
 210 215 220
 Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
 225 230 235 240
 Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
 245 250 255
 Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
 260 265 270
 Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
 275 280 285
 Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
 290 295 300
 Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
 305 310 315 320
 Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
 325 330 335
 Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
 340 345 350
 Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
 355 360 365
 Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
 370 375 380
 Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
 385 390 395 400
 Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
 405 410 415
 Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
 420 425 430
 Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
 435 440 445
 Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
 450 455 460
 Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
 465 470 475 480
 Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
 485 490 495
 His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
 500 505 510
 Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
 515 520 525
 Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
 530 535 540
 Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
 545 550 555 560
 Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
 565 570 575
 Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
 580 585 590
 Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
 595 600 605

-continued

Asn	Ile	Gln	Arg	Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp	610	615	620	
Pro	Glu	Phe	Gln	Ala	Ser	Asn	Ile	Met	His	Ser	Ile	Asn	Gly	Tyr	Val	625	630	635	640
Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	Cys	Leu	His	Glu	Val	Ala	Tyr	Trp	645	650	655	
Tyr	Ile	Leu	Ser	Ile	Gly	Ala	Gln	Thr	Asp	Phe	Leu	Ser	Val	Phe	Phe	660	665	670	
Ser	Gly	Tyr	Thr	Phe	Lys	His	Lys	Met	Val	Tyr	Glu	Asp	Thr	Leu	Thr	675	680	685	
Leu	Phe	Pro	Phe	Ser	Gly	Glu	Thr	Val	Phe	Met	Ser	Met	Glu	Asn	Pro	690	695	700	
Gly	Leu	Trp	Ile	Leu	Gly	Cys	His	Asn	Ser	Asp	Phe	Arg	Asn	Arg	Gly	705	710	715	720
Met	Thr	Ala	Leu	Leu	Lys	Val	Ser	Ser	Cys	Asp	Lys	Asn	Thr	Gly	Asp	725	730	735	
Tyr	Tyr	Glu	Asp	Ser	Tyr	Glu	Asp	Ile	Ser	Ala	Tyr	Leu	Leu	Ser	Lys	740	745	750	
Asn	Asn	Thr	Thr	Tyr	Val	Asn	Arg	Ser	Leu	Ser	Gln	Asn	Pro	Pro	Val	755	760	765	
Leu	Lys	Arg	His	Gln	Arg	Glu	Ile	Thr	Arg	Thr	Thr	Leu	Gln	Ser	Asp	770	775	780	
Gln	Glu	Glu	Ile	Asp	Tyr	Asp	Asp	Thr	Ile	Ser	Val	Glu	Met	Lys	Lys	785	790	795	800
Glu	Asp	Phe	Asp	Ile	Tyr	Asp	Glu	Asp	Glu	Asn	Gln	Ser	Pro	Arg	Ser	805	810	815	
Phe	Gln	Lys	Lys	Thr	Arg	His	Tyr	Phe	Ile	Ala	Ala	Val	Glu	Arg	Leu	820	825	830	
Trp	Asp	Tyr	Gly	Met	Ser	Ser	Ser	Pro	His	Val	Leu	Arg	Asn	Arg	Ala	835	840	845	
Gln	Ser	Gly	Ser	Val	Pro	Gln	Phe	Lys	Lys	Val	Val	Phe	Gln	Glu	Phe	850	855	860	
Thr	Asp	Gly	Ser	Phe	Thr	Gln	Pro	Leu	Tyr	Arg	Gly	Glu	Leu	Asn	Glu	865	870	875	880
His	Leu	Gly	Leu	Leu	Gly	Pro	Tyr	Ile	Arg	Ala	Glu	Val	Glu	Asp	Asn	885	890	895	
Ile	Met	Val	Thr	Phe	Arg	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Ser	Phe	Tyr	900	905	910	
Ser	Ser	Leu	Ile	Ser	Tyr	Glu	Glu	Asp	Gln	Arg	Gln	Gly	Ala	Glu	Pro	915	920	925	
Arg	Lys	Asn	Phe	Val	Lys	Pro	Asn	Glu	Thr	Lys	Thr	Tyr	Phe	Trp	Lys	930	935	940	
Val	Gln	His	His	Met	Ala	Pro	Thr	Lys	Asp	Glu	Phe	Asp	Cys	Lys	Ala	945	950	955	960
Trp	Ala	Tyr	Phe	Ser	Asp	Val	Asp	Leu	Glu	Lys	Asp	Val	His	Ser	Gly	965	970	975	
Leu	Ile	Gly	Pro	Leu	Leu	Val	Cys	His	Thr	Asn	Thr	Leu	Asn	Pro	Ala	980	985	990	
His	Gly	Arg	Gln	Val	Thr	Val	Gln	Glu	Phe	Ala	Leu	Phe	Phe	Thr	Ile	995	1000	1005	
Phe	Asp	Glu	Thr	Lys	Ser	Trp	Tyr	Phe	Thr	Glu	Asn	Met	Glu	Arg					

-continued

1010	1015	1020
Asn Cys Arg Ala Pro Cys	Asn Ile Gln Met Glu Asp	Pro Thr Phe
1025	1030	1035
Lys Glu Asn Tyr Arg Phe His	Ala Ile Asn Gly Tyr	Ile Met Asp
1040	1045	1050
Thr Leu Pro Gly Leu Val Met	Ala Gln Asp Gln Arg	Ile Arg Trp
1055	1060	1065
Tyr Leu Leu Ser Met Gly Ser	Asn Glu Asn Ile His	Ser Ile His
1070	1075	1080
Phe Ser Gly His Val Phe Thr	Val Arg Lys Lys Glu	Glu Tyr Lys
1085	1090	1095
Met Ala Leu Tyr Asn Leu Tyr	Pro Gly Val Phe Glu	Thr Val Glu
1100	1105	1110
Met Leu Pro Ser Lys Ala Gly	Ile Trp Arg Val Glu	Cys Leu Ile
1115	1120	1125
Gly Glu His Leu His Ala Gly	Met Ser Thr Leu Phe	Leu Val Tyr
1130	1135	1140
Ser Asn Lys Cys Gln Thr Pro	Leu Gly Met Ala Ser	Gly His Ile
1145	1150	1155
Arg Asp Phe Gln Ile Thr Ala	Ser Gly Gln Tyr Gly	Gln Trp Ala
1160	1165	1170
Pro Lys Leu Ala Arg Leu His	Tyr Ser Gly Ser Ile	Asn Ala Trp
1175	1180	1185
Ser Thr Lys Glu Pro Phe Ser	Trp Ile Lys Val Asp	Leu Leu Ala
1190	1195	1200
Pro Met Ile Ile His Gly Ile	Lys Thr Gln Gly Ala	Arg Gln Lys
1205	1210	1215
Phe Ser Ser Leu Tyr Ile Ser	Gln Phe Ile Ile Met	Tyr Ser Leu
1220	1225	1230
Asp Gly Lys Lys Trp Gln Thr	Tyr Arg Gly Asn Ser	Thr Gly Thr
1235	1240	1245
Leu Met Val Phe Phe Gly Asn	Val Asp Ser Ser Gly	Ile Lys His
1250	1255	1260
Asn Ile Phe Asn Pro Pro Ile	Ile Ala Arg Tyr Ile	Arg Leu His
1265	1270	1275
Pro Thr His Tyr Ser Ile Arg	Ser Thr Leu Arg Met	Glu Leu Met
1280	1285	1290
Gly Cys Asp Leu Asn Ser Cys	Ser Met Pro Leu Gly	Met Glu Ser
1295	1300	1305
Lys Ala Ile Ser Asp Ala Gln	Ile Thr Ala Ser Ser	Tyr Phe Thr
1310	1315	1320
Asn Met Phe Ala Thr Trp Ser	Pro Ser Lys Ala Arg	Leu His Leu
1325	1330	1335
Gln Gly Arg Ser Asn Ala Trp	Arg Pro Gln Val Asn	Asn Pro Lys
1340	1345	1350
Glu Trp Leu Gln Val Asp Phe	Gln Lys Thr Met Lys	Val Thr Gly
1355	1360	1365
Val Thr Thr Gln Gly Val Lys	Ser Leu Leu Thr Ser	Met Tyr Val
1370	1375	1380
Lys Glu Phe Leu Ile Ser Ser	Ser Gln Asp Gly His	Gln Trp Thr
1385	1390	1395

-continued

Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln
 1400 1405 1410

Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu
 1415 1420 1425

Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile
 1430 1435 1440

Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
 1445 1450 1455

<210> SEQ ID NO 106

<211> LENGTH: 1457

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 106

Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
 1 5 10 15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
 20 25 30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
 35 40 45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
 50 55 60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
 65 70 75 80

Ala Lys Pro Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln
 85 90 95

Ala Glu Val Tyr Asp Thr Val Val Val Thr Leu Lys Asn Met Ala Ser
 100 105 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ser Ser
 115 120 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
 130 135 140

Asp Lys Val Phe Pro Gly Lys Ser His Thr Tyr Val Trp Gln Val Leu
 145 150 155 160

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

Tyr Leu Ser His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile
 180 185 190

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

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
 210 215 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
 225 230 235 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
 245 250 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
 260 265 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile

-continued

275					280					285					
Phe	Leu	Glu	Gly	His	Thr	Phe	Leu	Val	Arg	Asn	His	Arg	Gln	Ala	Ser
290						295					300				
Leu	Glu	Ile	Ser	Pro	Ile	Thr	Phe	Leu	Thr	Ala	Gln	Thr	Leu	Leu	Met
305					310					315					320
Asp	Leu	Gly	Gln	Phe	Leu	Leu	Ser	Cys	His	Ile	Ser	Ser	His	Gln	His
				325					330					335	
Asp	Gly	Met	Glu	Ala	Tyr	Val	Lys	Val	Asp	Ser	Cys	Pro	Glu	Glu	Pro
			340					345					350		
Gln	Leu	Arg	Met	Lys	Asn	Asn	Glu	Glu	Ala	Glu	Asp	Tyr	Asp	Asp	Asp
		355				360						365			
Leu	Thr	Asp	Ser	Glu	Met	Asp	Val	Val	Arg	Phe	Asp	Asp	Asp	Asn	Ser
370						375					380				
Pro	Ser	Phe	Ile	Gln	Ile	Arg	Ser	Val	Ala	Lys	Lys	His	Pro	Lys	Thr
385					390					395					400
Trp	Val	His	Tyr	Ile	Ala	Ala	Glu	Glu	Glu	Asp	Trp	Asp	Tyr	Ala	Pro
				405					410					415	
Leu	Val	Leu	Ala	Pro	Asp	Asp	Arg	Ser	Tyr	Lys	Ser	Gln	Tyr	Leu	Asn
			420					425					430		
Asn	Gly	Pro	Gln	Arg	Ile	Gly	Arg	Lys	Tyr	Lys	Lys	Val	Arg	Phe	Met
		435				440						445			
Ala	Tyr	Thr	Asp	Glu	Thr	Phe	Lys	Thr	Arg	Glu	Ala	Ile	Gln	His	Glu
450						455					460				
Ser	Gly	Ile	Leu	Gly	Pro	Leu	Leu	Tyr	Gly	Glu	Val	Gly	Asp	Thr	Leu
465					470					475					480
Leu	Ile	Ile	Phe	Lys	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Asn	Ile	Tyr	Pro
				485					490					495	
His	Gly	Ile	Thr	Asp	Val	Arg	Pro	Leu	Tyr	Ser	Arg	Arg	Leu	Pro	Lys
			500					505					510		
Gly	Val	Lys	His	Leu	Lys	Asp	Phe	Pro	Ile	Leu	Pro	Gly	Glu	Ile	Phe
		515					520					525			
Lys	Tyr	Lys	Trp	Thr	Val	Thr	Val	Glu	Asp	Gly	Pro	Thr	Lys	Ser	Asp
530						535					540				
Pro	Arg	Cys	Leu	Thr	Arg	Tyr	Tyr	Ser	Ser	Phe	Val	Asn	Met	Glu	Arg
545					550					555					560
Asp	Leu	Ala	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Ile	Cys	Tyr	Lys	Glu
				565				570						575	
Ser	Val	Asp	Gln	Arg	Gly	Asn	Gln	Ile	Met	Ser	Asp	Lys	Arg	Asn	Val
			580					585					590		
Ile	Leu	Phe	Ser	Val	Phe	Asp	Glu	Asn	Arg	Ser	Trp	Tyr	Leu	Thr	Glu
		595					600					605			
Asn	Ile	Gln	Arg	Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp
610						615					620				
Pro	Glu	Phe	Gln	Ala	Ser	Asn	Ile	Met	His	Ser	Ile	Asn	Gly	Tyr	Val
625					630					635					640
Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	Cys	Leu	His	Glu	Val	Ala	Tyr	Trp
				645				650						655	
Tyr	Ile	Leu	Ser	Ile	Gly	Ala	Gln	Thr	Asp	Phe	Leu	Ser	Val	Phe	Phe
			660					665					670		
Ser	Gly	Tyr	Thr	Phe	Lys	His	Lys	Met	Val	Tyr	Glu	Asp	Thr	Leu	Thr
		675					680					685			

-continued

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
 690 695 700
 Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
 705 710 715 720
 Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
 725 730 735
 Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
 740 745 750
 Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu
 755 760 765
 Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln
 770 775 780
 Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu
 785 790 795 800
 Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe
 805 810 815
 Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp
 820 825 830
 Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln
 835 840 845
 Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr
 850 855 860
 Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His
 865 870 875 880
 Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile
 885 890 895
 Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser
 900 905 910
 Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
 915 920 925
 Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val
 930 935 940
 Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp
 945 950 955 960
 Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu
 965 970 975
 Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His
 980 985 990
 Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe
 995 1000 1005
 Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn
 1010 1015 1020
 Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys
 1025 1030 1035
 Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr
 1040 1045 1050
 Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr
 1055 1060 1065
 Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His Phe
 1070 1075 1080

-continued

Ser Gly 1085	His Val Phe Thr	Val 1090	Arg Lys Lys Glu	Glu Tyr Lys Met 1095
Ala Leu 1100	Tyr Asn Leu Tyr	Pro 1105	Gly Val Phe Glu Thr	Val Glu Met 1110
Leu Pro 1115	Ser Lys Ala Gly	Ile 1120	Trp Arg Val Glu Cys	Leu Ile Gly 1125
Glu His 1130	Leu His Ala Gly	Met 1135	Ser Thr Leu Phe Leu	Val Tyr Ser 1140
Asn Lys 1145	Cys Gln Thr Pro	Leu 1150	Gly Met Ala Ser Gly	His Ile Arg 1155
Asp Phe 1160	Gln Ile Thr Ala	Ser 1165	Gly Gln Tyr Gly Gln	Trp Ala Pro 1170
Lys Leu 1175	Ala Arg Leu His	Tyr 1180	Ser Gly Ser Ile Asn	Ala Trp Ser 1185
Thr Lys 1190	Glu Pro Phe Ser	Trp 1195	Ile Lys Val Asp Leu	Leu Ala Pro 1200
Met Ile 1205	Ile His Gly Ile	Lys 1210	Thr Gln Gly Ala Arg	Gln Lys Phe 1215
Ser Ser 1220	Leu Tyr Ile Ser	Gln 1225	Phe Ile Ile Met Tyr	Ser Leu Asp 1230
Gly Lys 1235	Lys Trp Gln Thr	Tyr 1240	Arg Gly Asn Ser Thr	Gly Thr Leu 1245
Met Val 1250	Phe Phe Gly Asn	Val 1255	Asp Ser Ser Gly Ile	Lys His Asn 1260
Ile Phe 1265	Asn Pro Pro Ile	Ile 1270	Ala Arg Tyr Ile Arg	Leu His Pro 1275
Thr His 1280	Tyr Ser Ile Arg	Ser 1285	Thr Leu Arg Met Glu	Leu Met Gly 1290
Cys Asp 1295	Leu Asn Ser Cys	Ser 1300	Met Pro Leu Gly Met	Glu Ser Lys 1305
Ala Ile 1310	Ser Asp Ala Gln	Ile 1315	Thr Ala Ser Ser Tyr	Phe Thr Asn 1320
Met Phe 1325	Ala Thr Trp Ser	Pro 1330	Ser Lys Ala Arg Leu	His Leu Gln 1335
Gly Arg 1340	Ser Asn Ala Trp	Arg 1345	Pro Gln Val Asn Asn	Pro Lys Glu 1350
Trp Leu 1355	Gln Val Asp Phe	Gln 1360	Lys Thr Met Lys Val	Thr Gly Val 1365
Thr Thr 1370	Gln Gly Val Lys	Ser 1375	Leu Leu Thr Ser Met	Tyr Val Lys 1380
Glu Phe 1385	Leu Ile Ser Ser	Ser 1390	Gln Asp Gly His Gln	Trp Thr Leu 1395
Phe Phe 1400	Gln Asn Gly Lys	Val 1405	Lys Val Phe Gln Gly	Asn Gln Asp 1410
Ser Phe 1415	Thr Pro Val Val	Asn 1420	Ser Leu Asp Pro Pro	Leu Leu Thr 1425
Arg Tyr 1430	Leu Arg Ile His	Pro 1435	Gln Ser Trp Val His	Gln Ile Ala 1440
Leu Arg 1445	Met Glu Val Leu	Gly 1450	Cys Glu Ala Gln Asp	Leu Tyr 1455

-continued

```

<210> SEQ ID NO 107
<211> LENGTH: 1457
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polypeptide

<400> SEQUENCE: 107

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

```


-continued

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
 355 360 365
 Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
 370 375 380
 Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
 385 390 400
 Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
 405 410 415
 Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
 420 425 430
 Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
 435 440 445
 Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
 450 455 460
 Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
 465 470 475 480
 Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
 485 490 495
 His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
 500 505 510
 Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
 515 520 525
 Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
 530 535 540
 Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
 545 550 555 560
 Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
 565 570 575
 Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
 580 585 590
 Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
 595 600 605
 Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
 610 615 620
 Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
 625 630 635 640
 Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
 645 650 655
 Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
 660 665 670
 Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
 675 680 685
 Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
 690 695 700
 Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
 705 710 715 720
 Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
 725 730 735
 Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
 740 745 750
 Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu

-continued

755					760					765					
Lys	Arg	His	Gln	Arg	Glu	Ile	Thr	Arg	Thr	Thr	Leu	Gln	Ser	Asp	Gln
770					775					780					
Glu	Glu	Ile	Asp	Tyr	Asp	Asp	Thr	Ile	Ser	Val	Glu	Met	Lys	Lys	Glu
785					790					795					800
Asp	Phe	Asp	Ile	Tyr	Asp	Glu	Asp	Glu	Asn	Gln	Ser	Pro	Arg	Ser	Phe
				805					810					815	
Gln	Lys	Lys	Thr	Arg	His	Tyr	Phe	Ile	Ala	Ala	Val	Glu	Arg	Leu	Trp
				820					825					830	
Asp	Tyr	Gly	Met	Ser	Ser	Ser	Pro	His	Val	Leu	Arg	Asn	Arg	Ala	Gln
				835					840					845	
Ser	Gly	Ser	Val	Pro	Gln	Phe	Lys	Lys	Val	Val	Phe	Gln	Glu	Phe	Thr
				850					855					860	
Asp	Gly	Ser	Phe	Thr	Gln	Pro	Leu	Tyr	Arg	Gly	Glu	Leu	Asn	Glu	His
				865					870					875	
Leu	Gly	Leu	Leu	Gly	Pro	Tyr	Ile	Arg	Ala	Glu	Val	Glu	Asp	Asn	Ile
				885					890					895	
Met	Val	Thr	Phe	Arg	Asn	Gln	Ala	Ser	Arg	Pro	Tyr	Ser	Phe	Tyr	Ser
				900					905					910	
Ser	Leu	Ile	Ser	Tyr	Glu	Glu	Asp	Gln	Arg	Gln	Gly	Ala	Glu	Pro	Arg
				915					920					925	
Lys	Asn	Phe	Val	Lys	Pro	Asn	Glu	Thr	Lys	Thr	Tyr	Phe	Trp	Lys	Val
				930					935					940	
Gln	His	His	Met	Ala	Pro	Thr	Lys	Asp	Glu	Phe	Asp	Cys	Lys	Ala	Trp
				945					950					955	
Ala	Tyr	Phe	Ser	Asp	Val	Asp	Leu	Glu	Lys	Asp	Val	His	Ser	Gly	Leu
				965					970					975	
Ile	Gly	Pro	Leu	Leu	Val	Cys	His	Thr	Asn	Thr	Leu	Asn	Pro	Ala	His
				980					985					990	
Gly	Arg	Gln	Val	Thr	Val	Gln	Glu	Phe	Ala	Leu	Phe	Phe	Thr	Ile	Phe
				995					1000					1005	
Asp	Glu	Thr	Lys	Ser	Trp	Tyr	Phe	Thr	Glu	Asn	Met	Glu	Arg	Asn	
				1010					1015					1020	
Cys	Arg	Ala	Pro	Cys	Asn	Ile	Gln	Met	Glu	Asp	Pro	Thr	Phe	Lys	
				1025					1030					1035	
Glu	Asn	Tyr	Arg	Phe	His	Ala	Ile	Asn	Gly	Tyr	Ile	Met	Asp	Thr	
				1040					1045					1050	
Leu	Pro	Gly	Leu	Val	Met	Ala	Gln	Asp	Gln	Arg	Ile	Arg	Trp	Tyr	
				1055					1060					1065	
Leu	Leu	Ser	Met	Gly	Ser	Asn	Glu	Asn	Ile	His	Ser	Ile	His	Phe	
				1070					1075					1080	
Ser	Gly	His	Val	Phe	Thr	Val	Arg	Lys	Lys	Glu	Glu	Tyr	Lys	Met	
				1085					1090					1095	
Ala	Leu	Tyr	Asn	Leu	Tyr	Pro	Gly	Val	Phe	Glu	Thr	Val	Glu	Met	
				1100					1105					1110	
Leu	Pro	Ser	Lys	Ala	Gly	Ile	Trp	Arg	Val	Glu	Cys	Leu	Ile	Gly	
				1115					1120					1125	
Glu	His	Leu	His	Ala	Gly	Met	Ser	Thr	Leu	Phe	Leu	Val	Tyr	Ser	
				1130					1135					1140	
Asn	Lys	Cys	Gln	Thr	Pro	Leu	Gly	Met	Ala	Ser	Gly	His	Ile	Arg	
				1145					1150					1155	

-continued

Asp	Phe	Gln	Ile	Thr	Ala	Ser	Gly	Gln	Tyr	Gly	Gln	Trp	Ala	Pro
1160						1165					1170			
Lys	Leu	Ala	Arg	Leu	His	Tyr	Ser	Gly	Ser	Ile	Asn	Ala	Trp	Ser
1175						1180					1185			
Thr	Lys	Glu	Pro	Phe	Ser	Trp	Ile	Lys	Val	Asp	Leu	Leu	Ala	Pro
1190						1195					1200			
Met	Ile	Ile	His	Gly	Ile	Lys	Thr	Gln	Gly	Ala	Arg	Gln	Lys	Phe
1205						1210					1215			
Ser	Ser	Leu	Tyr	Ile	Ser	Gln	Phe	Ile	Ile	Met	Tyr	Ser	Leu	Asp
1220						1225					1230			
Gly	Lys	Lys	Trp	Gln	Thr	Tyr	Arg	Gly	Asn	Ser	Thr	Gly	Thr	Leu
1235						1240					1245			
Met	Val	Phe	Phe	Gly	Asn	Val	Asp	Ser	Ser	Gly	Ile	Lys	His	Asn
1250						1255					1260			
Ile	Phe	Asn	Pro	Pro	Ile	Ile	Ala	Arg	Tyr	Ile	Arg	Leu	His	Pro
1265						1270					1275			
Thr	His	Tyr	Ser	Ile	Arg	Ser	Thr	Leu	Arg	Met	Glu	Leu	Met	Gly
1280						1285					1290			
Cys	Asp	Leu	Asn	Ser	Cys	Ser	Met	Pro	Leu	Gly	Met	Glu	Ser	Lys
1295						1300					1305			
Ala	Ile	Ser	Asp	Ala	Gln	Ile	Thr	Ala	Ser	Ser	Tyr	Phe	Thr	Asn
1310						1315					1320			
Met	Phe	Ala	Thr	Trp	Ser	Pro	Ser	Lys	Ala	Arg	Leu	His	Leu	Gln
1325						1330					1335			
Gly	Arg	Ser	Asn	Ala	Trp	Arg	Pro	Gln	Val	Asn	Asn	Pro	Lys	Glu
1340						1345					1350			
Trp	Leu	Gln	Val	Asp	Phe	Gln	Lys	Thr	Met	Lys	Val	Thr	Gly	Val
1355						1360					1365			
Thr	Thr	Gln	Gly	Val	Lys	Ser	Leu	Leu	Thr	Ser	Met	Tyr	Val	Lys
1370						1375					1380			
Glu	Phe	Leu	Ile	Ser	Ser	Ser	Gln	Asp	Gly	His	Gln	Trp	Thr	Leu
1385						1390					1395			
Phe	Phe	Gln	Asn	Gly	Lys	Val	Lys	Val	Phe	Gln	Gly	Asn	Gln	Asp
1400						1405					1410			
Ser	Phe	Thr	Pro	Val	Val	Asn	Ser	Leu	Asp	Pro	Pro	Leu	Leu	Thr
1415						1420					1425			
Arg	Tyr	Leu	Arg	Ile	His	Pro	Gln	Ser	Trp	Val	His	Gln	Ile	Ala
1430						1435					1440			
Leu	Arg	Met	Glu	Val	Leu	Gly	Cys	Glu	Ala	Gln	Asp	Leu	Tyr	
1445						1450					1455			

<210> SEQ ID NO 108
 <211> LENGTH: 4374
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 108

atgcagattg agctgtccac ctgcttcttt ctgtgctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa cttctctggg actacatgca gtctgacctg 120

-continued

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac	180
acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt	240
gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgagggtgat	300
gacactgtgg tgcgtcacct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg	360
ggagtctcat actggaatc ctctgaagg gctgagtatg atgaccagac atcccagaga	420
gagaaagagg atgacaaggt gttccctggg aagtctcaca cctatgtgtg gcaagtcctc	480
aaggagaatg gaccactgc atctgaocca cctgctctga catactccta cctttctcat	540
gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa	600
ggatccctgg ccaaggagaa aaccagaca ctgcacaagt tcattctcct gtttctgtc	660
tttgatgagg gcaagtcttg gactctgaa acaagaact ccctgatgca agacagggat	720
gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca	780
ctgctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg	840
acaaccctg aagtgcactc cttttctctg gagggacaca ccttctctgt caggaaccac	900
agacaagcct ctctggagat ctctcccatc accttctca ctgcacagac actgctgatg	960
gacctggac agttctctgt gtctctccac atctcttccc accagcatga tggcatggaa	1020
gcctatgtca aggtggactc atgcctgag gaaccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat	1140
gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca	1200
tgggtgcaact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc	1260
cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccaaaaag aattggaaga	1320
aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc	1380
attcagcatg agtctggcat tctgggacca ctctctgatg gggaaagtggg agacaccctg	1440
ctcatcatct tcaagaacca ggctccagg ccctacaaca tctaccaca tggcatcact	1500
gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc	1560
cccattctgc ctggagagat cttcaagtac aagtggactg tcaactgtga ggatggacca	1620
acaaagtctg accccagggt cctcaccaga tactactcct cttttgtgaa catggagaga	1680
gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag	1740
agaggcaacc agatcatgct tgacaagaga aatgtgattc tgttctctgt ctttgatgag	1800
aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaacct tgctggggtg	1860
caactggaag acctgaggt ccaggcaagc aacatcatgc actccatcaa tggctatgtg	1920
tttgactctc tccagcttct tgtctgctg catgaggtgg cctactggta cattcttct	1980
attggggcac aaactgactt ctttctgtc ttcttctctg gatacacctt caagcacaag	2040
atggtgtatg aggacacct gacactcttc ccattctctg gggaaactgt gttcatgagc	2100
atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga	2160
atgactgcac tgctcaaagt ctctctctgt gacaagaaca ctggggacta ctatgaggac	2220
tcttatgagg acatctctgc ctacctgctc agcaagaaca atgocattga gcccagaagc	2280
ttctctcaga atccacctgt cctgaagaga caccagagag agatcaccag gacaaccctc	2340
cagtctgacc aggaagagat tgactatgat gacaccattt ctgtggagat gaagaaggag	2400

-continued

gactttgaca tctatgatga ggacgagaac cagtctccaa gatcattcca gaagaagaca	2460
agacactact tcattgtgtgc tgtggaaaga ctgtgggact atggcatgtc ttcctctccc	2520
catgtcctca ggaacagggc acagtctggc tctgtgccac agttcaagaa agtggctctc	2580
caggagttea ctgatggctc attcaccag cccctgtaca gaggggaact gaatgagcac	2640
ctgggactcc tgggaccata catcagggct gaggtggaag acaacatcat ggtgacattc	2700
agaaaccagg cctccaggcc ctacagcttc tactcttccc tcatcagcta tgaggaagac	2760
cagagacaag gggctgagcc aagaaagaac tttgtgaaac ccaatgaaac caagacctac	2820
ttctggaaag tccagcacca catggcacc accaaggatg agtttgactg caaggcctgg	2880
gcatacttct ctgatgtgga cctggagaaa gatgtgcaact ctggcctgat tggcccactc	2940
ctggtctgoc acaccaacac cctgaaccct gcacatggaa ggcaagtgac tgtgcaggag	3000
tttgccctct tcttcacat ctttgatgaa accaagtcat ggtacttcac tgagaacatg	3060
gagagaaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagaa	3120
tacaggttcc atgccatcaa tggctacatc atggacaccc tgctgggct tgtcatggca	3180
caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cattcactcc	3240
atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggcctg	3300
tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc	3360
tggaggggtg aatgcctcat tggggagcac ctgcatgctg gcatgtcaac cctgttctctg	3420
gtctacagca acaagtgcc aacaccctg ggaatggcct ctggccacat cagggacttc	3480
cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac	3540
tctggctcca tcaatgcatg gtcaaccaag gagccattct cttggatcaa ggtggacctg	3600
ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctctctt	3660
ctgtacatct cacagttoat catcatgtac tctctggatg gcaagaagtg gcagacatac	3720
agaggcaact ccactggcac cctcatggtc ttctttggca atgtggacag ctctggcatc	3780
aagcacaaca tcttcaaccc tccatcatt gccagataca tcaggctgca cccccccac	3840
tactcaatca gatcaacct caggatgga ctgatgggat gtgacctgaa ctctgtctca	3900
atgcccctgg gaatggagag caaggccatt tctgatgccc agatcactgc atctcttac	3960
ttcaccaca tgtttgccac ctggtcacca tcaaaagcca ggctgcacct ccaggaaga	4020
agcaatgcct ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttcag	4080
aagacaatga aagtcaactg ggtgacaacc cagggggta agtctctgct cacctcaatg	4140
tatgtgaagg agttctgat ctcttctca caggatggcc accagtggac actcttcttc	4200
cagaatggca aagtcaaggt gttccagggc aaccaggact ctttcacacc tgtggtgaac	4260
tcactggacc cccccctct gacaagatac ctgagaattc acccccagtc ttgggtccac	4320
cagattgccc tgagaatgga agtccctggga tgtgaggcac aagacctgta ctga	4374

<210> SEQ ID NO 109

<211> LENGTH: 4374

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 109

-continued

atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc	60
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg	120
ggggagctgc ctgtggatgc cagggttccca cccagagtgc ccaaatcctt cccattcaac	180
acctctgtgg tctacaagaa gacctcttt gtggagtcca ctgaccacct gttcaacatt	240
gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgagggtgat	300
gacactgtgg tcgtcacccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg	360
ggggctcagct actggaagtc ctctgagggg gctgagtatg atgaccagac ctcccagagg	420
gagaaggagg atgacaaagt gttccctggg aagagccaca cctatgtgtg gcaggctctc	480
aaggagaatg gccccactgc ctctgaccca cctgcctga cctactccta cctttctcat	540
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggg gtgcagggag	600
ggctccctgg ccaagagaa gaccagacc ctgcacaagt tcattctcct gtttctgtc	660
tttgatgagg gcaagagctg gcaactctgaa accaagaact ccctgatgca ggacagggat	720
gctgcctctg ccagggctg gccaagatg cacactgtga atggctatgt gaacaggagc	780
ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcattgtgat tggcatgggg	840
acaacccctg aggtgcactc cttttctctg gagggccaca ccttctctgt caggaaccac	900
agacaggcca gcctggagat cagccccatc accttctca ctgcccagac cctgctgatg	960
gacctcggac agttcctgct gtcctgccac atcagctccc accagcatga tggcatggag	1020
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag	1080
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggg ccgctttgat	1140
gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca cccaagacc	1200
tgggtgcact acattgtctg tgaggaggag gactgggact atgccccact ggtcctggcc	1260
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gccacagag gattggagcg	1320
aagtacaaga aagtcagggt catggcctac actgatgaaa ccttcaagac caggagggcc	1380
attcagcatg agtctggcat cctgggcca ctctctgatg gggagggtgg ggacaccctg	1440
ctcatcatct tcaagaacca ggctccagg cctacaaca tctaccaca tggcatcact	1500
gatgtcaggc ccctgtacag ccgcaggctg ccaaggggg tgaaacacct caaggacttc	1560
cccattctgc ctggggagat cttcaagtac aagtggactg tcaactgtga ggatggacca	1620
accaaactcg accccagggt cctcaccaga tactactcca gctttgtgaa catggagagg	1680
gacctggcct ctggcctgat tgcccactg ctcatctgct acaaggagtc tgtggaccag	1740
aggggaaacc agatcatgtc tgacaagagg aatgtgatc tgttctctgt ctttgatgag	1800
aacaggagct ggtacctgac tgagaacatt cagcgtctcc tgccccaccc tgctggggtg	1860
cagctggagg acctgaggt ccaggccagc aacatcatgc actccatcaa tggetatgtg	1920
tttgacagcc tccagcttct tgtctgctg catgaggtgg cctactggta cattcttct	1980
attggggccc agactgactt cctttctgtc ttcttctctg gctacacct caaacacaag	2040
atggtgtatg aggacaccct gacctcttc ccattctctg gggagactgt gttcatgagc	2100
atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc	2160
atgactgccc tgctcaaagt ctctcctgt gacaagaaca ctggggacta ctatgaggac	2220
agctatgagg acatctctgc ctacctgtc agcaagaaca atgccattga gcccaggagc	2280

-continued

ttcagccaga atccacctgt cctgaaacgc caccagaggg agatcaccag gaccaccctc 2340
cagctcgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag 2400
gactttgaca tctatgacga ggacgagaa cagagcccaa ggagcttcca gaagaagacc 2460
aggcactact tcattgtctg tgtggagcgc ctgtgggact atggcatgag ctccagcccc 2520
catgtcctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtggctctc 2580
caagagtcca ctgatggcag cttcaccag cccctgtaca gaggggagct gaatgagcac 2640
ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgacctc 2700
cgcaaccagg cctccaggcc ctacagcttc tacagctccc tcatcagcta tgaggaggac 2760
cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac 2820
ttctggaaaag tccagcacca catggcccc accaaggatg agtttgactg caaggcctgg 2880
gcctacttct ctgatgtgga cctggagaag gatgtgcact ctggcctgat tggcccactc 2940
ctggtctgcc acaccaacac cctgaaccct gcccatggaa ggcaagtgac tgtgcaggag 3000
tttgccctct tcttcacat ctttgatgaa accaagagct ggtacttcac tgagaacatg 3060
gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaa 3120
taccgcttcc atgccatcaa tggtacatc atggaccccc tgctgggct tgtcatggcc 3180
caggaccaga ggatcaggtg gtacctgctt tctatgggct ccaatgagaa cattcactcc 3240
atccacttct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggccctg 3300
tacaacctct accctggggt ctttgagact gtggagatgc tgccctcaa agctggcatc 3360
tggaggggtg agtgcctcat tggggagcac ctgcatgctg gcatgagcac cctgttctctg 3420
gtctacagca acaagtgcc accccctctg ggaatggcct ctggccacat cagggacttc 3480
cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac 3540
tctggatcca tcaatgctg gagcaccaag gagccattca gctggatcaa agtggactg 3600
ctggccccca tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc 3660
ctgtacatca gccagtcoat catcatgtac agcctggatg gcaagaaatg gcagacctac 3720
agaggcaact ccaactggaac actcatggtc ttctttgca atgtggacag ctctggcatc 3780
aagcacaaca tcttcaaccc cccaatcacc gccagataca tcaggctgca cccccccac 3840
tacagcatcc gcagcaccct caggatggag ctgatgggct gtgacctgaa ctctgcagc 3900
atgcccctgg gcatggagag caaggccatt tctgatgccc agatcactgc ctccagctac 3960
ttcaccaca tgtttgccac ctggagcccc agcaaggcca ggctgcacct ccaggaagag 4020
agcaatgcct ggaggcccc ggtcaacaac ccaaaggagt ggctgcagggt ggacttccag 4080
aagaccatga aggtcactgg ggtgaccacc caggggggtca agagcctgct caccagcatg 4140
tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctcttcttc 4200
cagaatggca aggtcaagggt gttccagggc aaccaggaca gcttcaacccc tgtgggtgaa 4260
agcctggacc cccccctct gaccagatac ctgaggatcc acccccagag ctgggtccac 4320
cagattgccc tgaggatgga ggtcctggga tgtgagggcc aggacctgta ctga 4374

1-74. (canceled)

75. A polynucleotide comprising a nucleotide sequence encoding a Factor VIII polypeptide, the Factor VIII polypeptide comprising a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain,

wherein the heavy chain of the Factor VIII polypeptide comprises a first polypeptide sequence having at least 95% identity to amino acids 20 to 759 of SEQ ID NO: 2;

wherein the light chain of the Factor FVIII polypeptide comprises a second polypeptide sequence having at least 95% identity to nucleic acids 774 to 1457 of SEQ ID NO: 2; and

wherein the polypeptide linker comprises a furin cleavage site and a glycosylation peptide having an amino acid sequence of SEQ ID NO: 55.

76. The polynucleotide of claim **75**, wherein the encoded Factor VIII polypeptide comprises I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to SEQ ID NO: 19.

77. The polynucleotide of claim **75**, wherein the polypeptide linker comprises an amino acid sequence of SEQ ID NO:113.

78. The polynucleotide of claim **75**, wherein the polypeptide linker comprises an amino acid sequence of SEQ ID NO:113.

79. The polynucleotide of claim **75**,

wherein the heavy chain of the Factor VIII polypeptide comprises a first polypeptide sequence having at least 99% identity to amino acids 20 to 759 of SEQ ID NO: 2;

wherein the light chain of the Factor FVIII polypeptide comprises a second polypeptide sequence having at least 99% identity to nucleic acids 774 to 1457 of SEQ ID NO: 2.

80. The polynucleotide of claim **79**, wherein the encoded Factor VIII polypeptide comprises I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to SEQ ID NO: 19.

81. The polynucleotide of claim **79**, wherein the polypeptide linker comprises an amino acid sequence of SEQ ID NO:113.

82. The polynucleotide of claim **80**, wherein the polypeptide linker comprises an amino acid sequence of SEQ ID NO:113.

83. The polynucleotide of claim **75**, wherein the encoded Factor VIII polypeptide comprises a sequence having at least 99% sequence identity to SEQ ID NO: 2.

84. The polynucleotide of claim **83**, wherein the encoded Factor VIII polypeptide comprises I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to SEQ ID NO: 19.

85. The polynucleotide of claim **83**, wherein the polypeptide linker comprises an amino acid sequence of SEQ ID NO:113.

86. The polynucleotide of claim **84**, wherein the polypeptide linker comprises an amino acid sequence of SEQ ID NO:113.

87. The polynucleotide of claim **75**, further comprising a promoter element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

88. The polynucleotide of claim **87**, wherein the promoter element is a liver-specific promoter sequence upstream of the nucleotide sequence encoding the Factor VIII polypeptide.

89. An adeno-associated virus (AAV) vector comprising a polynucleotide according to claim **75**.

90. An adeno-associated virus (AAV) particle comprising a polynucleotide according to claim **75**.

91. The adeno-associated virus (AAV) particle according to claim **90**, wherein the AAV particle is an AAV-8 particle.

92. A host cell infected with an adeno-associated virus (AAV) particle comprising a polynucleotide according to claim **75**.

93. A method for producing an adeno-associated virus (AAV) particle comprising introducing a polynucleotide according to claim **75** into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.

94. A method for treating hemophilia A comprising administering, to a patient in need thereof, an adeno-associated virus (AAV) particle according to claim **90**.

* * * * *