(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0158531
(43) 공개일자 2023년11월20일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) *A61K 48/00* (2006.01)
C12N 15/10 (2017.01) *C12N 15/113* (2010.01)
C12N 15/86 (2006.01) *C12N 9/78* (2006.01)
(52) CPC특허분류
C12N 9/22 (2013.01)
A61K 48/00 (2013.01)
(21) 출원번호 10-2023-7034367
(22) 출원일자(국제) 2022년03월23일
심사청구일자 없음
(85) 번역문제출일자 2023년10월06일
(86) 국제출원번호 PCT/US2022/021523
(87) 국제공개번호 WO 2022/204268
국제공개일자 2022년09월29일
(30) 우선권주장
63/164,798 2021년03월23일 미국(US)

(71) 출원인
빔 테라퓨틱스, 인크.
미국 02142 매사추세츠주 캠브리지 메인 스트리트 238 나인쓰 플로어
(72) 발명자
제츠, 베른트
미국 02139 매사추세츠 케임브리지 랜즈다운 스트리트 26 2 플로어 빔 테라퓨틱스, 인크. (내)
바레라, 루이스
미국 02139 매사추세츠 케임브리지 랜즈다운 스트리트 26 2 플로어 빔 테라퓨틱스, 인크. (내)
(뒷면에 계속)
(74) 대리인
특허법인 남앤남

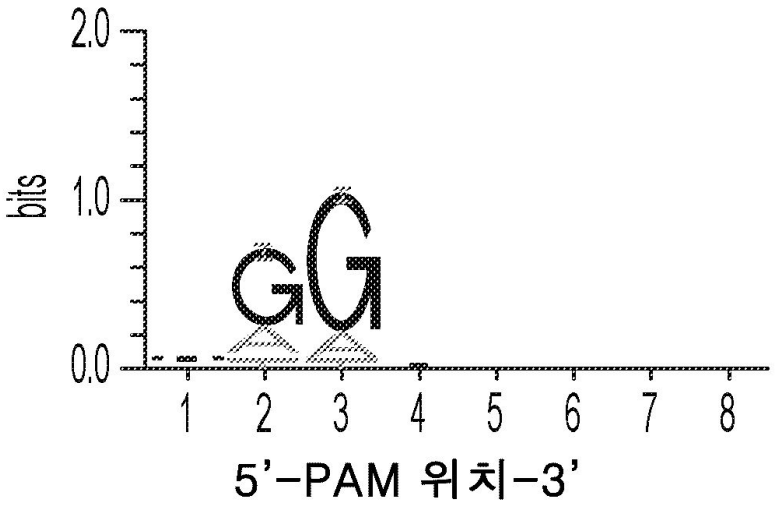
전체 청구항 수 : 총 86 항

(54) 발명의 명칭 **신규 CRISPR 효소, 방법, 시스템 및 그 용도**

(57) 요 약

본 발명은 핵산 표적화 및 조작을 위해 인간 세포에 최적화된 재조합적으로 조작된 신규 Cas9 효소를 제조하고 사용하기 위한 신규 시스템, 방법 및 조성물을 제공한다. 본 발명은 인간 세포에 사용하기 위해 코돈-최적화되고 재조합적으로 생성된 *스트렙토코커스 콘스텔라투스(Streptococcus constellatus)*, *샤르펜(Sharpen) spp. 단리물 RUG017*, *베일로넬라 파르불라(Veillonella parvula)*, *에자키엘라 페루엔시스(Ezakiella peruensis)*, *락토바실러스 퍼멘텀 균주(Lactobacillus fermentum strain) AF15-40LB 균주* 및 *펩토니필러스 sp. 마르세유(Peptoniphilus sp. Marseille)-P3761* 박테리아로부터 신규 Cas9 효소의 발견에 기반한다. 일부 구현예에서, 신규 Cas9 효소는 염기 편집을 위해 사용될 수 있다. 일부 구현예에서, 신규 조작된 Cas9 효소는 인간 질병을 치료하기 위해 사용된다.

대 표 도



(52) CPC특허분류
C12N 15/102 (2013.01)
C12N 15/113 (2013.01)
C12N 15/86 (2013.01)
C12N 9/78 (2013.01)
C12Y 305/04001 (2013.01)
C12Y 305/04004 (2013.01)
C12Y 305/04005 (2013.01)
C12N 2310/20 (2017.05)
C12N 2800/22 (2013.01)
(72) 발명자
본, 데이비드, 에이.
미국 02139 매사추세츠 케임브리지 랜즈다운 스트리트 26 2 플로어 빔 테라퓨틱스, 인크. (내)
쑨, 밍
미국 02139 매사추세츠 케임브리지 랜즈다운 스트리트 26 2 플로어 빔 테라퓨틱스, 인크. (내)

명 세 서

청구범위

청구항 1

스트렙토코커스 콘스텔라투스(*Streptococcus constellatus*) Cas9, *샤르페아*(*Sharpea*) Cas9, *베일로넬라 파르불라*(*Veillonella parvula*) Cas9, *에자키엘라 페루엔시스*(*Ezakiella peruensis*) Cas9, *락토바실러스 퍼멘텀 균주*(*Lactobacillus fermentum strain*) *AF15-40LB* Cas9 또는 *펩토니필러스 sp. 마르세유*(*Peptoniphilus sp. Marseille*)-*P3761* Cas9로부터 변형되는 조작된, 비천연 발생 Cas9 단백질.

청구항 2

제1항에 있어서, 상기 *스트렙토코커스 콘스텔라투스* Cas9는

MGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLD
DSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEI
TVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGIL
TVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNG
SIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPK
HSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKA
NEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQI
IDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQ
NGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKR
QLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRS
RKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVA
YSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTTLLYHSYRI
EKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIH
QSITGLYETRIDLSKLGED (서열번호 1)와 적어도 80% 서열 동일성을 갖는, Cas9 단백질.

청구항 3

제1항에 있어서, 상기 *샤르페아 Cas9*는

MAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEE
DKQKNLGNDYKDNYNLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFAKVYEDIEEKLDNTLKKFMSLNDLDNLF
VDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYIEFIEMLHSI
NSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEK
FMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWNYDQIVDIDATAEGFIERMQN
TGTYFPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIF
GEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIEES
NKKDIEGPFRYDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTD
ALYLYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFSLIRTDFNEKDK
KRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYKKDVQEHNN
YGFILNSMNHIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKAKIPVNKYRKDVNKYGGFSGDAPIMLAIEADKG
KKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKLESIDDANFNEVI
QELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKL
(서열번호 4)와 적어도 80% 서열 동일성을 갖는, Cas9 단백질.

청구항 4

제1항에 있어서, 상기 *베일로넬라 파르불라 Cas9*는

MSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELF
ADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDV

IKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKL
QITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEET
SCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYYVGPLNTH
HNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTK
NRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQK
LLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVT
RSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSF
DNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDF
RHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTG
ALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISIIYGKL
CINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMK
GNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIV (서열번호 8)와 적
어도 80% 서열 동일성을 갖는, Cas9 단백질.

청구항 5

제1항에 있어서, 상기 *에자키엘라 페루엔시스 Cas9*는
MTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKD
QKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINI
LTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEAD
KSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADL
ELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKD
VVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFK
QGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREY
NLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFI
SEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCY
AYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGN
VYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKS
SIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSAL
QLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLK
SIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKL (서열번호 14)와 적어도 80% 서열 동일성을 갖는, Cas9
단백질.

청구항 6

제1항에 있어서, 상기 *락토바실러스 퍼멘텀 균주 AF15-40LB Cas9*는

MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQ
AFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAY
EELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFRE
ELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEID
ELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEI
TPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVE
GLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWN
TDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFE
DQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAKMKPYWNQL
LKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLN
RRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQN
KKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFI
LSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFG
SDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL (서열번호 84)와 적어도 80% 서열 동일성을 갖는,
Cas9 단백질.

청구항 7

제1항에 있어서, 상기 *펩토니필러스 sp. 마르세유-P3761* Cas9는

MEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFR
LNESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIE
LRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYN
AVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIE
KEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLI
GRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKV
PGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQ
LINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKE
NLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDN
LTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALY
PYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKD
LDPSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIA
MEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAI
SRRRYNSMTGLFDGQVIYQSLTGLYETRMRFED (서열번호 86)와 적어도 80% 서열 동일성을 갖는, Cas9 단백질.

청구항 8

제2항 내지 제7항 중 어느 한 항에 있어서, 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열을 포함하는 Cas9 단백질.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함하는 Cas9 단백질.

청구항 10

제9항에 있어서, 상기 *스트렙토코커스 콘스텔라투스 Cas9*는

MPKKKRKVGGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVD
ENFFQRLDDSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTI
EESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYD
AILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLR
KQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYL
PEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILD
KSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFK
EEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQND
RLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEE
DKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEY
PKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYG
GFDSPVVAYSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTT
LLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTE
CLNATLIHQSITGLYETRIDLSKLGEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 2)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

[청구항 10b]

제9항에 있어서, 상기 *스트렙토코커스 콘스텔라투스 Cas9*는

MPKKKRKVGMGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKV
DENFFQRLDDSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKT

IEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLY
DAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFL
RKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLY
LPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKIL
DKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSF
KEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQN
DRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTE
EDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGE
YPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKY
GGFMQPVVAYSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVIPQRFT
TLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVT
ECLNATLIHQSITGLYETRIDLSKLGEDGEGADKRTADGSEFESPKKKRKV (서열번호 95)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

[청구항 10c]

제9항에 있어서, 상기 *스트렙토코커스 콘스텔라투스 Cas9*는

MPKKKRKVGMGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKV
DENFFQRLDDSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKT
IEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLY
DAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFL
RKQRTFDNGIIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLY
LPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKIL
DKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERLHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSF
KEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQN
DRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTE
EDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGE
YPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKY
GGFMQPVVAYSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVIPQRFT
TLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVT
ECLNATLIHQSITGLYETRIDLSKLGEDGEGADKRTADGSEFESPKKKRKV (서열번호 96)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

청구항 11

제9항에 있어서, 상기 *샤르페아 Cas9*가

MPKKKRKVGAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLL
NVSFLDEEDKQKNLGNDYKDNYNLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALAHHIVKYRGNFLKEGQSFAKVYEDIEEKLDNTLKKFMS
LNDLDNLFVDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYIE
FIEMLHSINSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREI
LERIDLEKFMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWNYDQIVDIDATAE
GFIERMQNTGTYFPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAP
WIDFTKIFGEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLG
FKQIIEESNKKDIEGPFRYDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKED
AKKKIDTDALYLYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFSLIR
TDFNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYK
KDVQEHNNYGFILNSMNHIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKAKIPVNKYRKDVNKYGGFSGDAPIM
LAIEADKGKKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKLESID
DANFNEVIQELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTG
LFSKKYKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 5)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

청구항 12

제9항에 있어서, 상기 *베일로넬라 파르불라* Cas9는

MPKKKRKVGSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLR
LKLLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDS
NASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEE
LEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVVDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYI
KQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPY
YVGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFK
QDHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYR
DWGRLSQKLLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALP
ARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPR
SLTKDDSFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFV
KAENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVT
KRLLEQTGALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKD
ISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLK
MPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAG
QAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 9)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas 단백질.

청구항 13

제9항에 있어서, 상기 *에자키엘라 페루엔시스* Cas9는

MPKKKRKVGTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNS
DLYMEDKDQKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEF
DNENLININLTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSIL
ENLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINK
VNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSE
ITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILL
SGTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGS
IIHFMREYNLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFY
KDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPD
EILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDA
YINIIVGNVVYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRP
LKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDK
TLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRA
SGVRQDLKSIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서
열번호 15) (D12A 돌연변이체는 굵은 글씨)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

청구항 14

제9항에 있어서, 상기 *락토바실러스 퍼멘텀 균주 AF15-40LB* Cas9는

MPKKKRKVGKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIH
PEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKS
FNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSE
DEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSK
KENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWA
KRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTR
GVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQ
RIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRR
TQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAK
MKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLT

TFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESS
SLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFL
VDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIK
ILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDV
PDYA (서열번호 85)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

청구항 15

제9항에 있어서, 상기 *펩토니필러스 sp. 마르세유-P3761* Cas9가

MPKKKRKVGEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDE
VDENFFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNA
LFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLF
DHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGS
EKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSA
EKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYS
TYVELAKVPGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQ
HLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAA
LGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGL
MTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAI
AINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKG
SRIPIKKDLDPSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVE
LQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPAD
FIFMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 87)
와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서, 상기 Cas9 단백질의 아미노산 서열은 서열번호 1, 4, 8, 14, 84 또는 86에서 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 또는 적어도 10개의 돌연변이를 포함하는, Cas9 단백질.

청구항 17

제16항에 있어서, 상기 돌연변이는 아미노산 치환인, Cas9 단백질.

청구항 18

제1항 내지 제17항 중 어느 한 항에 있어서, 상기 Cas9 단백질이 닉카제 활성을 갖는, Cas9 단백질.

[청구항 18b]

제18항에 있어서, 아미노산 위치에서의 닉카제 돌연변이가 야생형 SpCas9의 1개 이상의 아미노산 10, 12, 17, 762, 840, 854, 863, 982, 983, 984, 986, 987에 상응하는, Cas9 단백질.

청구항 19

제16항에 있어서, 적어도 1개의 돌연변이가 불활성 Cas9(dCas9)를 초래하는, Cas9 단백질.

청구항 20

제1항 내지 제19항 중 어느 한 항에 있어서, 상기 Cas9 단백질은 PAM 상호작용, HNH 및/또는 RuvC 도메인에 적어도 1개의 아미노산 돌연변이를 포함하는, Cas9 단백질.

[청구항 20b]

제20항에 있어서, 아미노산 위치에서의 돌연변이는 SirCas9의 RuvC 도메인에서 아미노산 14에 상응하는, Cas9 단백질.

[청구항 20c]

제20항에 있어서, 아미노산 위치에서의 돌연변이는 EpeCas9의 RuvC 도메인에서 아미노산 12에 상응하는, Cas9 단백질.

[청구항 20d]

제20항에 있어서, 아미노산 위치에서의 돌연변이는 LfeCas9의 RuvC 도메인에서 아미노산 9에 상응하는, Cas9 단백질.

[청구항 20e]

제20항에 있어서, 아미노산 위치에서의 돌연변이는 PmaCas9의 RuvC 도메인에서 아미노산 12에 상응하는, Cas9 단백질.

[청구항 20f]

제20항에 있어서, 상기 Cas9 단백질은 초정밀 Cas9인, Cas9 단백질.

[청구항 20g]

제20항에 있어서, 상기 Cas9 단백질은 SpyCas9(서열번호 173)에 관하여 N692A, M694A, Q695A 및/또는 H698A에 상응하는 돌연변이를 포함하는, Cas9 단백질.

[청구항 20h]

제20항에 있어서, 상기 Cas9 단백질은 고충실 Cas9인, Cas9 단백질.

[청구항 20i]

제20항에 있어서, 상기 Cas9 단백질은 SpyCas9(서열번호 173)에 관하여 N467A, R661A, Q695A 및/또는 Q926A에 상응하는 돌연변이를 포함하는, Cas9 단백질.

[청구항 20j]

제20항에 있어서, 상기 Cas9 단백질은 SuperFi-Cas9인, Cas9 단백질.

[청구항 20k]

제20항에 있어서, SpyCas9에 상응하는 Y1016, R1019, Y1010, Y1013, K1031, Q1027 및/또는 VI018 잔기가 아스파르트산으로 돌연변이되는, Cas9 단백질.

청구항 21

서열번호 1, 4, 8, 14, 84 또는 86과 적어도 80% 동일성을 갖는 Cas9 단백질을 포함하고, 상기 Cas9 단백질은 히스톤 데메틸라제, 전사 활성화인자 또는 데아미나제에 융합되는, 조작된 비천연 발생 Cas9 융합 단백질.

[청구항 21b]

제21항에 있어서, 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함하는 조작된 비천연 발생 Cas9 융합 단백질.

[청구항 21c]

제22항에 있어서, 서열번호 2, 5, 9, 15, 85, 87, 95 또는 96과 적어도 80% 동일성을 갖는 조작된 비-천연 발생 Cas9 융합 단백질.

청구항 22

제21항에 있어서, 상기 Cas9 단백질은 시토신 데아미나제 또는 아데노신 데아미나제에 융합되는, Cas9 단백질.

청구항 23

제22항에 있어서, 상기 Cas9 단백질은 아데노신 데아미나제에 융합되고

(a)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GGKPYSIGLAIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLD
DSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEI
TVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGIL
TVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNG
SIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPK
HSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKA
NEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQI
IDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQ
NGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVKEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKR
QLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRS
RKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVA
YSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTTLLYHSYRI
EKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIH
QSITGLYETRIDLSKLGEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 20);

(b)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GAKNKDIRYSIGLAIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEE
DKQKNLGNDYKDNYNLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFAKVYEDIEEKLDNTLKKFMSLNDLDNLF
VDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYIEFIEMLHSI
NSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEK
FMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWNYDQIVDIDATAEGFIERMQN
TGTYFPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIF
GEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIEES
NKKDIEGPFRYDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTD
ALYLYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFSLIRTDFNEKDK
KRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYKKDVQEHNN
YGFILNSMNHIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKAKIPVNKYRKDVNKYGGFSGDAPIMLAIEADKG
KKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKLESIDDANFNEVI
QELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKL
GKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 6);

(c)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELF
ADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDV
IKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKL
QITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEET
SCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYYVGPLNTH
HNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTK
NRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQK
LLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVT
RSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSF
DNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDF
RHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTG
ALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISIIYGKL
CINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMK
GNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGS

YPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 10);

(d)

MPKKKRKVSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRL
KLLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSN
ASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEEL
EEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIK
QGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYY
VGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQ
DHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRD
WGRLSQKLLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPA
RIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRS
LTKDDSFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVK
AENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTK
RLLEQTGALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDI
SIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKM
PIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVKRPAATKKAGQA
KKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLY
VTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDV
PDYAYPYDVPDYAYPYDVPDYA (서열번호 11);

(e)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GTKVKDYYIGLAIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKD
QKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINI
LTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEAD
KSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADL
ELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKD
VVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFK
QGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREY
NLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFI
SEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCY
AYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGN
VYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKS
SIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSAL
QLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLK
SIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호
16);

(f)

MPKKKRKVTKVKDYYIGLAIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSD
LYMEDKDQKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFD
NENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILE
NLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKV
NGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEI
TPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLS
GTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSI
IHFMREYNLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYK
DGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDE
ILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAY
INIIVGNVYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPL

KDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKT
LELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRAS
GVRQDLKSIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 17);

(g)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQ
AFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAY
EELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFRE
ELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEID
ELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEI
TPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVE
GLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWN
TDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFE
DQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAKMKPYWNQL
LKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLN
RRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQN
KKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFI
LSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFG
SDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서
열번호 88);

(h)

MPKKKRKVGKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIH
PEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKS
FNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSE
DEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSK
KENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWA
KRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTR
GVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQ
RIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRR
TQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAK
MKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLT
TFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESS
SLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFL
VDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIK
ILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번
호 89); 또는

(i)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSGKPYS
IGLAIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVD
EDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALS
ILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNS
TKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQI
HLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYE

AFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIE
DIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDD
IANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMY
TGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETR
QITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATE
KFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVF
ADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVIPQRFTTLLYHSYRIEKDYEP
EHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITGL
YETRIDLSKLGEDGEGADKRTADGSEFESPKKKRKV (서열번호 98);

(j)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRV
VFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSGKPYS
IGLAIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVD
EDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALS
ILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNS
TKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGIIPHQI
HLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYE
AFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIE
DIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERLHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDD
IANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMY
TGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETR
QITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATE
KFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVF
ADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVIPQRFTTLLYHSYRIEKDYEP
EHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITGL
YETRIDLSKLGEDGEGADKRTADGSEFESPKKKRKV (서열번호 99)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

청구항 24

제22항에 있어서, 상기 Cas9 단백질은 시토신 데아미나제에 융합되고

(a)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGGKPYSIGLAIGTNSVGWAV
VTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIFGNI
AAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLEN
LIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYE
EHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGE
HYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYV
NEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDREM
IRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAI
KKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYD
IDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERF
NTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKK
EDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKLRKV
QDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFK
ELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSITGLYETRIDLSKLGEDG
KRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML

YPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 21);

(b)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGSIINFQRRGLMETQASNQL
ISSHLKGYPIKDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELFADAMAQVDPTFFMRLRESKY
HYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFDCNSAIS
SIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSD
EIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSD
SKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYYVGPLNTHHNVDNGGFAWAVRKASGRVT
PWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTKNRIEQFLKDNGYISETHKHE
ITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPETII
ILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDSRQKRLSDL
YAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSFDNLVLCKRTANAQKSDAYPI
SEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDFRHDNNFIKVRSLNHHHHAKD
AYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTGALADATIYKATVAGKAKDGA
YIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTN
SKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSL
EEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKE
TGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 12);

(c)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGTKVKDYYIGLAIGTSSVGW
AVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKDQKLKSKYTLFNDKDFKDKNF
HKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLNILTDPKLNKTAKKKELKSVIG
DTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEADKSKDGNKYISNAFVKKYNKH
GQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIK
SSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCT
YLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFKAIVGDKVDSD
DYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREYNLNLMELMSASFTFTEEIQK
LNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFISEIGEERYSYLLSEIEGEEE
NKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRL
TRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKDPMNFVKKQEKA
RSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKSSIEKYGGYTNINKACFAVLD
IKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSE
NDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLSSKTNNIGN
YQEFKIINQSITGLFENEVDLLKLGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDES
TDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLT
SDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 18);

(d)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGKEYHIGLAIGTSSIGWAVT
DSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLLKKNERGYP
TLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKI
GQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQNDILTEITSLFSQI
MLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVI

PHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIK
RMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKS
IVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQIDQ
LNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAP
DLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAKMKPYWNQLLKQGLITQRKFENLTMDVDQ
TIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKH
GSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQNKKSKLVVPLKNKLQPEYYGA
YTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDI
KLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFGSDLQNLKEIGFGTTPLGQFQ
VSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDIL
VHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDEST
DENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 90);

(e)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGEKKTNYTIGLAIGTDSVGW
AVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFSKNPIFG
TLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRT
VKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLS
AQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQV
HLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLIYQ
KYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMMDDPDNEDLMEEIV
KILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMID
PSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQN
GKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQR
QLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVYGEFRYKKKPQA
EKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSKYGGYIEEKMAYYAVI
SYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLVNFIYHLKNYDEISH
PESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSMTGLFDGQVIYQS
LTGLYETRMRFEDGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 93);

(f)

MPAAKRVKLDTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGG
STNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSPKKKRKVE
KKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLN
ESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELR
KEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYNAV
ELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKE
DFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGR
DTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPG
MKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLI
NDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENL
PGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDNLT
KAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPY
MADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLD
PSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAME
QDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISR
RRYNSMTGLFDGQVIYQSLTGLYETRMRFEDKRPAATKKAGQAKKKKGSSGGSSGGSSGSETPGTSESATPESSGGSSGGSTSEKGPSTGDPTLRRRIESWE

FDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMD QRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLAT GLIHPSVTWRYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 94)와 적어도 80% 동일한 아미노산 서열을 갖는, Cas9 단백질.

청구항 25

제2항에 있어서, 상기 Cas9 단백질은 5'-NGG-3'을 포함하는 PAM 서열을 인식하는, Cas9 단백질.

청구항 26

제3항에 있어서, 상기 Cas9 단백질은 5'-NAGHC-3'을 포함하는 PAM 서열을 인식하고, H는 아데닌, 시토신 또는 티민인, Cas9 단백질.

청구항 27

제4항에 있어서, 상기 Cas9 단백질은 5'-NRHRRH-3'을 포함하는 PAM 서열을 인식하고, H는 아데닌, 시토신 또는 티민이고, R은 아데닌 또는 구아닌인, Cas9 단백질.

청구항 28

제5항 또는 제7항에 있어서, 상기 Cas9 단백질은 5'-NGG-3'을 포함하는 PAM 서열을 인식하는, Cas9 단백질.

청구항 29

제6항에 있어서, 상기 Cas9 단백질은 5'-NNAAA-3'을 포함하는 PAM 서열을 인식하는, Cas9 단백질.

[청구항 29b]

제10b항 또는 제10c항에 있어서, 상기 Cas9 단백질은 5'-NGG-3'을 포함하는 PAM 서열을 인식하는, Cas9 단백질.

청구항 30

제1항 내지 제29항 중 어느 한 항의 Cas9 단백질을 암호화하는 핵산.

청구항 31

제30항에 있어서, 상기 핵산은 포유동물 세포에서 발현되도록 코돈-최적화되는, 핵산.

청구항 32

제31항에 있어서, 상기 핵산은 인간 세포에서 발현되도록 코돈-최적화되는, 핵산.

청구항 33

제29항의 Cas9 단백질을 포함하는 진핵 세포.

청구항 34

제33항에 있어서, 상기 세포는 인간 세포인, 진핵 세포.

청구항 35

진핵 세포에서 표적 핵산을 절단하는 방법으로,

상기 세포를 제1항 내지 제29항 중 어느 한 항의 Cas9, 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하고, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고,

상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열의 파손을 유발할 수 있는, 방법.

청구항 36

진핵 세포에서 표적 핵산의 발현을 변경(altering)하는 방법으로,

상기 세포를 제1항 내지 제29항 중 어느 한 항의 Cas9, 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하고, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고,

상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열의 파손을 유발할 수 있는, 방법.

청구항 37

진핵 세포에서 표적 핵산의 발현을 변경하는 방법으로,

상기 세포를 제1항 내지 제29항 중 어느 한 항의 Cas9, 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하고, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고,

상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있는, 방법.

청구항 38

진핵 세포에서 표적 핵산을 변형하는 방법으로,

상기 세포를 제1항 내지 제29항 중 어느 한 항의 Cas9, 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하고, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고,

상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있는, 방법.

청구항 39

제37항 또는 제38항에 있어서, 상기 Cas9 단백질은 불활성 Cas9(dCas9)인, 방법.

청구항 40

제39항에 있어서, 상기 dCas9는 데아미나제에 융합되는, 방법.

청구항 41

제35항 내지 제40항 중 어느 한 항에 있어서, 상기 RNA 가이드는 crRNA 및 tracrRNA를 포함하는, 방법.

청구항 42

제35항 내지 제39항 중 어느 한 항에 있어서, 상기 RNA 가이드는 sgRNA를 포함하는, 방법.

청구항 43

제42항에 있어서, *스트렙토코커스 콘스텔라투스* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUUAGAGCUGUGCUGUUUAAACAACACAGCAAGUUAAAAUAAGGCUUUGUCCGUACUCAAGCUUGCAAAAGCGUGCACCGAUUCGGUGCU-3' (서열번호 3)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 방법.

청구항 44

제42항에 있어서, *샤르페아* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUUAGAGUUGUGUUAUUGAAAAAUAACACAACGAGUUAAAAUAAAGCUUAUGCUUAAAUGCCAGCUUUGCUGGUGUCAUUUAGAUGACUUUACUAAGGUUGCUUCGGCAACCUUUUU-3' (서열번호 7)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 방법.

청구항 45

제42항에 있어서, *베일로넬라 파르불라* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUGAGAGUAGUGUGAAAACAUUACGAGUUCAAAUACAAAUUAAUUUACAAUGCCUUCGGGCUGCCCGACGUAGGGCACCUACUCUCAAUUCUUCGGAAUUGAGUU-3' (서열번호 13)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 방법.

청구항 46

제42항에 있어서, *에자키엘라 페루엔시스* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUGAGAGUUAUGUAAUUGAAAAAUUACAUGACGAGUUCAAAUAAAAAUUUAUUCAAACCGCCUAUUUAUAGGCCGCAGAUGUUCUGCAUUAUGCUUGCUAUUGCAAGCUU-3' (서열번호 19)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 방법.

청구항 47

제42항에 있어서, *락토바실러스 퍼멘텀 균주 AF15-40LB* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUCUUGGAUGAGUGUGAAAACACUCAUAGUCAAGAUCAAACGAGUGGUUUUCCACGAGUUAUUACUUUUGAGGUCUUAUAUGGCCCAUACAUAAAAAGGAGUCGGAAUUUCCGGCUCCUUUUCUU-3' (서열번호 95)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 방법.

청구항 48

제42항에 있어서, *펩토니필러스 sp. 마르세유-P3761* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUUAGAGCCAUGUAGAAAUACAUUGCAAGUUAAAAUAAGGCUUUGUCCGUAAUCAACUUGAAAAAGUGGCGCUGUUUCGGCGCUUU-3' (서열번호 96)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 방법.

청구항 49

제41항에 있어서, 상기 crRNA는 약 16 내지 26개의 뉴클레오티드 길이의 가이드 서열을 포함하는, 방법.

청구항 50

제49항에 있어서, 상기 crRNA는 18 내지 24개의 뉴클레오티드 길이의 가이드 서열을 포함하는, 방법.

청구항 51

제35항 또는 제36항에 있어서, 상기 표적 핵산의 파손은 단일 가닥 또는 이중 가닥의 파손인, 방법.

청구항 52

제51항에 있어서, 상기 표적 핵산의 파손은 단일 가닥의 파손인, 방법.

청구항 53

제34항 또는 제35항에 있어서, 상기 Cas9 단백질은 상기 표적 핵산 서열의 두 가닥 모두를 절단하는 뉴클레아제이거나, 상기 표적 핵산 서열의 한 가닥을 절단하는 닉카제인, 방법.

청구항 54

제34항 내지 제53항 중 어느 한 항에 있어서, 상기 표적 핵산은 5' 내지 프로토스페이서 인접 모티프(PAM) 서열인, 방법.

청구항 55

제34항 내지 제54항 중 어느 한 항에 있어서, 상기 Cas9가 진핵 세포에서 발현되도록 프로모터 서열에 작동 가능하게 연결되고, 상기 가이드 RNA가 진핵 세포에서 발현되도록 프로모터 서열에 작동 가능하게 연결되는, 방법.

청구항 56

제55항에 있어서, 상기 진핵 세포는 인간 세포인, 방법.

청구항 57

제55항에 있어서, 상기 프로모터 서열은 진핵 또는 바이러스의 프로모터인, 방법.

청구항 58

조작된 비천연 발생 CRISPR-Cas 시스템으로,

RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산으로, 상기 RNA 가이드는 직접 반복 서열, 및 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하는 것인, 상기 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산; 및

코돈-최적화된 CRISPR-연관(Cas) 단백질로, 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 80% 서열 동일성을 갖고, 상기 Cas 단백질은 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열에서 파손을 유발할 수 있는 것인, 코돈-최적화된 CRISPR-연관(Cas) 단백질

을 포함하는, 조작된 비천연 발생 CRISPR-Cas 시스템.

[청구항 58b]

제58항에 있어서, 상기 코돈-최적화된 CRISPR-연관(Cas) 단백질은 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함하는, 조작된 비천연 발생 CRISPR-Cas 시스템.

[청구항 58c]

제59항에 있어서, 상기 코돈-최적화된 CRISPR-연관(Cas) 단백질은 서열번호 2, 5, 9, 15, 85, 87, 95 또는 96과 적어도 80% 서열 동일성을 갖고, 상기 Cas 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열의 파손을 유발할 수 있는, 조작된 비천연 발생 CRISPR-Cas 시스템.

청구항 59

조작된 비천연 발생 CRISPR-Cas 시스템으로,

RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산으로, 상기 RNA 가이드는 직접 반복 서열, 및 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하는 것인, 상기 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산; 및

서열번호 1, 4, 8, 14, 84 또는 86과 적어도 80% 서열 동일성을 갖는 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함하고,

상기 Cas 단백질은 데아미나제에 융합되고, Cas 단백질 융합은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있는, 조작된 비천연 발생 CRISPR-Cas 시스템.

[청구항 59b]

제59항에 있어서, 상기 코돈-최적화된 CRISPR-연관(Cas) 단백질은 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함하는, 조작된 비천연 발생 CRISPR-Cas 시스템.

[청구항 59c]

제59b항에 있어서, 상기 코돈-최적화된 CRISPR-연관(Cas) 단백질은 서열번호 2, 5, 9, 15, 85, 87, 95 또는 96과 적어도 80% 서열 동일성을 갖고, 상기 Cas 단백질은 데아미나제에 융합되고, Cas 단백질 융합은 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있는, 조작된 비천연 발생 CRISPR-Cas 시스템.

청구항 60

제59항에 있어서, 상기 Cas9 단백질은 불활성 Cas9(dCas9)인, 시스템.

청구항 61

제58항 내지 제60항 중 어느 한 항에 있어서, 상기 RNA 가이드는 crRNA 및 tracrRNA를 포함하는, 시스템.

청구항 62

제58항 내지 제60항 중 어느 한 항에 있어서, 상기 RNA 가이드는 sgRNA를 포함하는, 시스템.

청구항 63

제62항에 있어서, *스트렙토코커스 콘스텔라투스* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUUAGAGCUGUGCUGUUUAAACAACACAGCAAGUUAAAAUAAGGCUUUGUCCGUACUCAAGCUUGCAAAAGCGUGCACCGAUUCGGUGCU-3' (서열번호 3)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 시스템.

청구항 64

제62항에 있어서, *사르페아* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUUAGAGUUGUGUUAUUGAAAAAUAACACAACGAGUUAAAAUAAAGCUUAUGCUUAAAUGCCAGCUUUGCUGGUGUCAUUUAGAUGACUUUACUAAGGUUGCUUCGGCAACCUUUUU-3' (서열번호 7)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 시스템.

청구항 65

제62항에 있어서, *베일로넬라 파르불라* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUGAGAGUAGUGUGAAAACAUUACGAGUUCAAAUACAAAUUAAUUUACAAUGCCUUCGGGCUGCCCGACGUAGGGCACCUACUCUCAAUUCUUCGGAAUUGAGUU-3' (서열번호 13)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 시스템.

청구항 66

제62항에 있어서, *에자키엘라 페루엔시스* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUGAGAGUUAUGUAAUUGAAAAAUUACAUGACGAGUUCAAAUAAAAAUUUAUUCAAACCGCCUAUUUAUAGGCCGCAGAUGUUCUGCAUUAUGCUUGCUAUUGCAAGCUU-3' (서열번호 19)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 시스템.

청구항 67

제62항에 있어서, *락토바실러스 퍼멘텀 균주 AF15-40LB* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUCUUGGAUGAGUGUGAAAACACUCAUAGUCAAGAUCAAACGAGUGGUUUUCCACGAGUUAUUACUUUUGAGGUCUUAUAUGGCCCAUACAUAAAAAGGAGUCGGAAUUUCCGGCUCCUUUUCUU-3' (서열번호 95)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 시스템.

청구항 68

제62항에 있어서, *펩토니필러스 sp. 마르세유-P3761* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUUAGAGCCAUGUAGAAAUACAUUGCAAGUUAAAAUAAGGCUUUGUCCGUAAUCAACUUGAAAAAGUGGCGCUGUUUCGGCGCUUU-3' (서열번호 96)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함하는, 시스템.

청구항 69

제58항 내지 제68항 중 어느 한 항에 있어서, 상기 Cas 단백질은 진핵 세포에서 발현되도록 프로모터 서열에 작동 가능하게 연결되고, 상기 가이드 RNA는 진핵 세포에서 발현되도록 프로모터 서열에 작동 가능하게 연결되는, 시스템.

청구항 70

제69항에 있어서, 상기 진핵 세포는 인간 세포인, 시스템.

청구항 71

제70항에 있어서, 상기 프로모터 서열은 진핵 프로모터 서열인, 시스템.

청구항 72

제58항 내지 제71항 중 어느 한 항의 시스템을 암호화하는 핵산.

청구항 73

제58항 내지 제72항 중 어느 한 항의 시스템을 포함하는 벡터.

청구항 74

제73항에 있어서, 상기 벡터는 플라스미드 벡터 또는 바이러스 벡터인, 벡터.

청구항 75

제74항에 있어서, 상기 바이러스 벡터는 아데노 연관 바이러스(AAV) 벡터 또는 렌티바이러스 벡터인, 벡터.

청구항 76

제75항에 있어서, 상기 바이러스 벡터는 AAV 벡터인, 벡터.

청구항 77

제76항에 있어서, 1개 초과의 AAV 벡터는 제59항 내지 제71항 중 어느 한 항의 시스템을 패키징하기 위해 사용되는, 벡터.

청구항 78

장애 또는 질병 치료를 필요로 하는 대상체에서 장애 또는 질병을 치료하는 방법으로, 상기 방법은 제58항 내지 제71항 중 어느 한 항의 시스템을 상기 대상체에게 투여하는 것을 포함하고,

상기 가이드 RNA는 병태 또는 질병과 연관된 표적 핵산의 적어도 10개 뉴클레오티드에 상보적이고;

상기 Cas 단백질은 상기 가이드 RNA와 연관되고;

상기 가이드 RNA는 상기 표적 핵산에 결합하고;

상기 Cas 단백질은 표적 핵산의 파손을 유발하고, 선택적으로 상기 Cas9는 데아미나제에 융합된 불활성 Cas9(dCas9)이고 상기 표적 핵산에서 1개 이상의 염기 편집을 초래하여 상기 장애 또는 질병을 치료하는, 방법.

청구항 79

제78항에 있어서, 상기 가이드 RNA는 약 18 내지 24개의 뉴클레오티드에 상보적인, 방법.

청구항 80

제79항에 있어서, 상기 가이드 RNA는 20개의 뉴클레오티드에 상보적인, 방법.

청구항 81

제16항 내지 제19항 중 어느 한 항의 융합 단백질을 포함하는 염기 편집기.

청구항 82

제81항에 있어서, 아데노신 데아미나제 도메인 또는 시티딘 데아미나제 도메인을 포함하는 염기 편집기.

[청구항 82b]

제81항에 있어서, 아데노신 데아미나제 도메인 및 시티딘 데아미나제 도메인을 포함하는 염기 편집기.

청구항 83

폴리뉴클레오티드의 핵염기를 편집하는 방법으로, 상기 방법은 1개 이상의 가이드 RNA와의 복합체에서 제81항의 염기 편집기와 상기 폴리뉴클레오티드를 접촉시키는 것을 포함하고, 상기 염기 편집기는 아데노신 데아미나제 도메인을 포함하고, 상기 1개 이상의 가이드 RNA는 상기 폴리뉴클레오티드에서 A·T에서 G·C로 변경을 이루도록 염기 편집기를 표적으로 하는, 방법.

청구항 84

폴리뉴클레오티드의 핵염기를 편집하는 방법으로, 상기 방법은 1개 이상의 가이드 RNA와의 복합체에서 제81항의 염기 편집기와 상기 폴리뉴클레오티드를 접촉시키는 것을 포함하고, 상기 염기 편집기는 시티딘 데아미나제 도

메인을 포함하고, 상기 1개 이상의 가이드 RNA는 상기 폴리뉴클레오티드에서 G·C에서 T·A로 변경을 이루도록 염기 편집기를 표적으로 하는, 방법.

청구항 85

제83항 또는 제84항에 있어서, 상기 편집이 표적 폴리뉴클레오티드 서열에서 50% 미만의 삽입결실 형성을 초래하는, 방법.

청구항 86

제83항 내지 제85항 중 어느 한 항에 있어서, 상기 편집이 점 돌연변이를 발생시키는, 방법.

발명의 설명

기 술 분 야

[0001] **관련 출원에 대한 상호 참조**

[0002] 본 출원은 모든 목적을 위해 전체적으로 본원에 참고로 통합되는 2021년 3월 23일에 출원된 미국 가특허 출원 일련번호 제63/164,798호의 우선권을 주장한다.

배 경 기 술

[0003] 원핵생물의 일정한 간격으로 놓인 짧은 회문 반복부 집합(CRISPR) 및 CRISPR-연관 단백질(CRISPR-Cas) 시스템 유래의 효소가 진핵생물에서 사용하기 위해 재프로그래밍 가능한 매우 특이적인 게놈 편집 도구(tool)로서 활용되었다. 게놈 편집 및 절단 이외에, CRISPR-Cas9는 게놈 상의 특이적 부위에 이펙터 분자를 위치시키기 위해 사용될 수 있어, 다양한 메커니즘을 통한 유전적 및 후생적 조절과 전사적 조절을 가능하게 한다.

[0004] 그러나, 다양한 게놈과 게놈 표적은 효과적인 유전자 조작을 위한 다양한 도구를 필요로 하고, 다양한 서열을 인식하고 표적으로 할 수 있는 신규 Cas 단백질을 발견하고 조작하여 CRISPR 도구상자를 확장해야 할 필요성이 여전히 남아 있다.

[0005] 유전자를 녹아웃시키거나 유전자 발현을 변형시키기 위해 CRISPR-Cas9 시스템을 사용할 수 있지만, 특정 종류의 유전자 편집에는 유전자 내 단일 염기 편집과 같이 표적 유전자에 대한 정확한 변형이 필요하다. 이러한 정확한 변형은 여전히 기술적 과제로 남아 있고, 광범위한 표적 유전자에서 정확한 게놈 변형을 수행하기 위해 다양한 유전자 편집 도구키트를 필요로 한다.

발명의 내용

[0006] **발명의 요약**

[0007] 고유의 프로토스페이서 인접 모티프(PAM)에 대하여 특이성을 갖는 신규 Cas9 효소를 식별하여, 유전자 편집을 위해 이용가능한 도구를 확장할 수 있다. 본 발명은 특히 *스트렙토코커스 콘스텔라투스*(*Streptococcus constellatus*), *샤르페아*(*Sharpea*) *spp. 단리물 RUG017*, *베일로넬라 파르불라*(*Veillonella parvula*), *에자키엘라 페루엔시스*(*Ezakiella peruensis*), *락토바실러스 퍼멘텀 균주*(*Lactobacillus fermentum strain*) *AF15-40LB* 또는 *펩토니필러스 sp. 마르세유*(*Peptoniphilus sp. Marseille*)*-P3761* 박테리아로부터 단리된 조작된 비천연 발생한 신규 Cas9 효소를 제공한다. 본 발명은 특이적 PAM 서열을 인식하는 상이한 박테리아로부터 발견된 신규 Cas9 효소가 진핵 세포(예: 인간, 식물 등)에서 발현되도록 조작될 수 있다는 놀라운 발견에 부분적으로 기반한다. 따라서, 기술된 Cas9 효소 및 그 변이체는 진핵생물에서 기능한다. 여기에 제공된 예는 다양한 게놈 부위를 표적으로 하기 위해 다양한 PAM 인식 서열을 사용하여 인간 세포에서 조작된 비천연 Cas9 효소의 사용을 보여준다. 예를 들어, *스트렙토코커스 콘스텔라투스*, *에자키엘라 페루엔시스* 및 *펩토니필러스 sp. 마르세유-P3761*로부터 조작된 Cas9는 공통 PAM 서열 5'-NGG-3'을 인식한다. *샤르페아 spp. 단리물 RUG017*로부터 단리된 Cas9에 의해 인식되는 공통 PAM 서열은 5'-NAGHC-3'이다. *베일로넬라 파르불라*로부터 단리된 Cas9에 의해 인식되는 공통 PAM 서열은 5'-NRHRRH-3'으로 식별되었다. *락토바실러스 퍼멘텀* 균주 AF15-40LB로부터 단리된 Cas9에 의해 인식되는 공통 PAM 서열은 5'-NNAAA-3'으로 식별되었다.(H=A, C 또는 T; R=A 또는 G).

[0008] 일 측면에서, *스트렙토코커스 콘스텔라투스* Cas9, *샤르페아* Cas9, *베일로넬라 파르불라* Cas9, *에자키엘라 페루*

엔시스 Cas9, *락토바실러스 퍼멘텀* 균주 AF15-40LB Cas9 또는 *펩토니필러스 sp. 마르세유-P3761* Cas9로부터 변형되는 조작된, 비천연 발생 Cas9 단백질이 본원에 제공된다.

[0009] 일부 구현예에서, 상기 *스트렙토코커스 콘스텔라투스* Cas9 단백질은

[0010] MGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLD
DSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEI
TVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGIL
TVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNG
SIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPK
HSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKA
NEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQI
IDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQ
NGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKR
QLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRS
RKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVA
YSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTTLLYHSYRI
EKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIH
QSITGLYETRIDLSKLGED (서열번호 1)와 적어도 80% 서열 동일성을 가진다.

[0011] 일부 구현예에서, 상기 *샤르페아 Cas9* 단백질은

[0012] MAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEE
DKQKNLGNDYKDNYNLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFAKVYEDIEEKLDNTLKKFMSLNDLDNLF
VDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYIEFIEMLHSI
NSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEK
FMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWNYDQIVDIDATAEGFIERMQN
TGTYFPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIF
GEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIEES
NKKDIEGPFRYDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTD
ALYLYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFSLIRTDFNEKDK
KRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYKKDVQEHNN
YGFILNSMNHIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKAKIPVNKYRKDVNKYGGFSGDAPIMLAIEADKG
KKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKLESIDDANFNEVI
QELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKL
(서열번호 4)와 적어도 80% 서열 동일성을 가진다.

[0013] 일부 구현예에서, 상기 *베일로넬라 파르불라* Cas9 단백질은
MSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELF
ADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDV
IKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKL
QITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEET
SCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYYVGPLNTH
HNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTK
NRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQK
LLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVT
RSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSF
DNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDF
RHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTG
ALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISIIYGKL
CINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMK
GNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIV (서열번호 8)와 적어도 80% 서열 동일성을 가진다.

[0014] 일부 구현예에서, 상기 *에자키엘라 페루엔시스* Cas9 단백질은 MTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKD QKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINI LTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEAD KSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADL ELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKD VVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFK QGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREY NLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFI SEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCY AYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGN VYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKS SIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSAL QLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLK SIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKL (서열번호 14)와 적어도 80% 서열 동일성을 가진다.

[0015] 일부 구현예에서, 상기 *락토바실러스 퍼멘텀* Cas9 단백질은

[0016] MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQ AFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAY EELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFRE ELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEID ELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEI TPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVE GLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWN TDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFE DQAHELVKDTSLTEELEKAPDLSDRYYFYFTQQGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAKMKPYWNQL LKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLN RRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQN KKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFI LSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFG SDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL (서열번호 84)와 적어도 80% 서열 동일성을 가진다.

[0017] 일부 구현예에서, 상기 *쾀토니필러스 sp. 마르세유-P3761* Cas9 단백질은

[0018] MEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFR LNESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIE LRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYN AVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIE KEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLI GRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKV PGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQ LINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKE NLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDN LTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALY PYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKD LDPSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIA MEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAI SRRRYNSMTGLFDGQVIYQSLTGLYETRMRFED (서열번호 86)와 적어도 80% 서열 동일성을 가진다.

[0019] 일부 구현예에서, Cas9 단백질은 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 85%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일한 아미노산을 포함한다.

[0020] 일부 구현예에서, Cas9 단백질은 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함한다.

[0021] 일부 구현예에서, 상기 *스트렙토코커스 콘스텔라투스* *Cas9*는 MPKKKRKVGGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVD ENFFQRLDDSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTI EESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYD AILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLR KQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYL PEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILD KSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFK EEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQND RLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEE DKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEY PKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYG GFDSPVVAYSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTT LLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTE CLNATLIHQSITGLYETRIDLSKLGEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 2)와 적어도 80% 서열 동일한 아미노산 서열을 가진다.

[0022] 일부 구현예에서, 상기 *샤르페아 Cas9*는

[0023] MPKKKRKVGAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLL NVSFLDEEDKQKNLGNDYKDNYNLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFAKVYEDIEEKLDNTLKKFMS LNDLDNLFVDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYIE FIEMLHSINSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREI LERIDLEKFMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWNYDQIVDIDATAE GFIERMQNTGTYFPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAP WIDFTKIFGEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLG FKQIIEESNKKDIEGPFRYDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKED AKKKIDTDALYLYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFSLIR TDFNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYK KDVQEHNNYGFILNSMNHIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKAKIPVNKYRKDVNKYGGFSGDAPIM LAIEADKGKKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKLESID DANFNEVIQELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTG LFSKKYKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 5)와 적어도 80% 동일한 아미노산 서열을 가진다.

[0024] 일부 구현예에서, 상기 *베일로벨라 파르불라* Cas9는

[0025] MPKKKRKVGSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLR LKLLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDS NASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEE LEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYI KQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPY YVGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFK QDHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYR DWGRLSQKLLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALP ARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPR SLTKDDSFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFV KAENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVT KRLLEQTGALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKD ISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLK MPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAG QAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 9)와 적어도 80% 동일한 아미노산 서열을 가진다.

[0026] 일부 구현예에서, 상기 *에자키엘라 페루엔시스* Cas9는

[0027] MPKKKRKVGTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNS
DLYMEDKDQKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEF
DNENLININILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSIL
ENLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINK
VNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSE
ITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILL
SGTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGS
IIHFMREYNLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFY
KDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPD
EILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDA
YINIIVGNVVYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRP
LKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDK
TLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRA
SGVRQDLKSIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서
열번호 15)와 적어도 80% 동일한 아미노산 서열을 가진다.

[0028] 일부 구현예에서, 상기 *락토바실러스 퍼멘텀 균주 AF15-40LB* Cas9는

[0029] MPKKKRKVGKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIH
PEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKS
FNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSE
DEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSK
KENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWA
KRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTR
GVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQ
RIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRR
TQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAK
MKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLT
TFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESS
SLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFL
VDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIK
ILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDV
PDYA (서열번호 85)와 적어도 80% 동일한 아미노산 서열을 가진다.

[0030] 일부 구현예에서, 상기 *펩토니필러스 sp.* 마르세유-P3761 Cas9가

[0031] MPKKKRKVGEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDE
VDENFFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNA
LFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLF
DHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGS
EKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSA
EKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYS
TYVELAKVPGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQ
HLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAA
LGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGL
MTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAI
AINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKG
SRIPIKKDLDPSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVE
LQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPAD
FIFMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 87)
와 적어도 80% 동일한 아미노산 서열을 가진다.

[0032] 일부 구현예에서, 상기 Cas9 단백질의 아미노산 서열은 서열번호 1, 4, 8, 14, 84 또는 86에서 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개 또는 적어도 10개의 돌연변이를 포함한다.

[0033] 일부 구현예에서, 상기 돌연변이는 아미노산 치환이다.

[0034] 일부 구현예에서, 상기 Cas9 단백질은 닉카제 활성을 가진다.

[0035] 일부 구현예에서, 야생형 SpCas9의 1개 이상의 아미노산 10, 12, 17, 762, 840, 854, 863, 982, 983, 984, 986, 987에 상응하는 아미노산 위치에서 닉카제 돌연변이를 포함하는 Cas9 단백질이 본원에 제공된다.

[0036] 일부 구현예에서, 적어도 1개의 돌연변이가 불활성 Cas9(dCas9)를 초래한다.

[0037] 일부 구현예에서, 상기 Cas9 단백질은 PAM 상호작용, HNH 및/또는 RuvC 도메인에 적어도 1개의 아미노산 돌연변이를 포함한다.

[0038] 일부 구현예에서, 아미노산 위치에서 돌연변이는 SirCas9의 RuvC 도메인에서의 아미노산 14에 상응하는 Cas9 단백질이 본원에 제공된다.

[0039] 일부 구현예에서, 아미노산 위치에서 돌연변이는 EpeCas9의 RuvC 도메인에서의 아미노산 12에 상응하는 Cas9 단백질이 본원에 제공된다.

[0040] 일부 구현예에서, 아미노산 위치에서 돌연변이는 LfeCas9의 RuvC 도메인에서의 아미노산 9에 상응하는 Cas9 단백질이 본원에 제공된다.

[0041] 일부 구현예에서, 아미노산 위치에서 돌연변이는 PmaCas9의 RuvC 도메인에서의 아미노산 12에 상응하는 Cas9 단백질이 본원에 제공된다.

[0042] 일부 구현예에서, 상기 Cas9 단백질은 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함한다.

[0043] 일 측면에서, 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 80% 동일성을 갖는 Cas9 단백질을 포함하는 조작된 비천연 발생 Cas9 융합 단백질이 본원에 제공되고 상기 Cas9 단백질은 히스톤 데메틸라제, 전사 활성화인자 또는 데아미나제에 융합된다.

[0044] 일부 구현예에서, 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함하는 조작된 비천연 발생 Cas9 융합 단백질이 본원에 제공된다.

[0045] 일부 구현예에서, 서열번호 2, 5, 9, 15, 85, 87, 95 또는 96과 적어도 80% 동일성을 갖는 조작된 비천연 발생 Cas9 융합 단백질이 본원에 제공된다.

[0046] 일부 구현예에서, 상기 Cas9 단백질은 시토신 데아미나제 또는 아데노신 데아미나제에 융합된다.

[0047] 일부 구현예에서, 상기 Cas9 단백질은 아데노신 데아미나제에 융합되고

[0048] (a)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GGKPYSIGLAIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLD
DSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEI
TVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGIL
TVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNG
SIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPK
HSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKA
NEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQI
IDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQ
NGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKR
QLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRS
RKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVA
YSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTTLLYHSYRI
EKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIH

QSITGLYETRIDLSKLGEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 20);

[0049] (b)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GAKNKDIRYSIGLAIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEE
DKQKNLGNDYKDNYNLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFAKVYEDIEEKLDNTLKKFMSLNDLDNLF
VDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYIEFIEMLHSI
NSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEK
FMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYVFGPLNTHSEFAWIKKFEDKQKERILPWNYDQIVDIDATAEGFIERMQN
TGTYFPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIF
GEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIEES
NKKDIEGPFRYDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTD
ALYLYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFSLIRTDFNEKDK
KRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYKKDVQEHNN
YGFILNSMNHIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKAKIPVNKYRKDVNKYGGFSGDAPIMLAIEADKG
KKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKLESIDDANFNEVI
QELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKL
GKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 6);

[0050] (c)

[0051] MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELF
ADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDV
IKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKL
QITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEET
SCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYYVGPLNTH
HNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTK
NRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQK
LLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVT
RSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSF
DNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDF
RHDNNFIKVRSLNHHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTG
ALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISIIYGKL
CINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMK
GNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGS
YPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 10);

[0052] (d)

MPKKKRKVSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRL
KLLEELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSN
ASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEEL
EEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIK
QGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYY
VGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQ
DHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRD
WGRLSQKLLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPA
RIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRS
LTKDDSFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVK
AENVTDFRHDNNFIKVRSLNHHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTK
RLLEQTGALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDI

SIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKM
PIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVKRPAATKKAGQA
KKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLY
VTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDV
PDYAYPYDVPDYAYPYDVPDYA (서열번호 11);

[0053] (e)

[0054] MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GTKVKDYYIGLAIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKD
QKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINI
LTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEAD
KSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADL
ELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKD
VVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFK
QGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREY
NLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFI
SEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCY
AYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGN
VYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKS
SIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSAL
QLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLK
SIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호
16);

[0055] (f)

MPKKKRKVTKVKDYYIGLAIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSD
LYMEDKDQKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFD
NENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILE
NLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKV
NGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEI
TPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLS
GTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSI
IHFMREYNLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYK
DGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDE
ILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAY
INIIVGNVYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPL
KDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKT
LELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRAS
GVRQDLKSIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTL
AKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 17);

[0056] (g)

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQ
AFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAY
EELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFRE
ELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEID
ELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEI
TPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVE

GLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWN
TDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFE
DQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAKMKPYWNQL
LKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLN
RRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQN
KKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFI
LSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFG
SDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서
열번호 88);

[0057] (h)

MPKKKRKVGKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIH
PEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKS
FNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSE
DEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSK
KENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWA
KRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTR
GVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQ
RIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRR
TQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAK
MKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLT
TFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESS
SLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFL
VDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIK
ILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHE
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKT
GAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번
호 89); 또는

[0058] (i)

[0059] MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAG
AMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPESSGPKKKRKV
GEKKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFR
LNESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIE
LRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYN
AVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIE
KEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLI
GRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKV
PGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQ
LINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKE
NLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDN
LTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALY
PYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKD
LDPSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIA
MEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAI
SRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDGKRPAATKKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 91);

[0060] (j)

MPKKKRKVEKKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEV
DENFFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNAL
FSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFD
HAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSE

KFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAE
KFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYST
YVELAKVPGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQH
LNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAAL
GSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLM
TQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIA
INLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGS
RIPIKKDLDPSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVEL
QKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADF
IFMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDKRPAATKKAGQAKKKKSGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVP
VGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVE
ITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDPAAKRVKLDGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 92)와 적어도 80% 동일한 아미노산 서열을 갖는다.

[0061] 일부 구현예에서, Cas9 단백질은 시토신 데아미나제에 융합되고

[0062] (a)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGGKPYSIGLAIGTNSVGWAV
VTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIFGNI
AAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLEN
LIAHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYE
EHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGE
HYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYV
NEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDREM
IRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAI
KKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYD
IDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERF
NTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKK
EDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKLRKV
QDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFK
ELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSITGLYETRIDLSKLGEDG
KRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDS
NGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML
YPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 21);

[0063] (b)

[0064] MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGSIINFQRRGLMETQASNQL
ISSHLKGYPIKDYFVGLAIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELFADAMAQVDPTFFMRLRESKY
HYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFDCNSAIS
SIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSD
EIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSD
SKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYYVGPLNTHHNVDNGGFAWAVRKASGRVT
PWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTKNRIEQFLKDNGYISETHKHE
ITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPETII
ILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDSRQKRLSDL
YAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSFDNLVLCKRTANAQKSDAYPI
SEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDFRHDNNFIKVRSLNHHHHAKD
AYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTGALADATIYKATVAGKAKDGA

YIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTN
SKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSL
EEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYSNEVTIVGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKE
TGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQE
SILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 12);

[0065] (c)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGTKVKDYYIGLAIGTSSVGW
AVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKDQKLKSKYTLFNDKDFKDKNF
HKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIG
DTKFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEADKSKDGNKYISNAFVKKYNKH
GQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIK
SSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCT
YLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFKAIVGDKVDSD
DYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREYNLNLMELMSASFTFTEEIQK
LNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFISEIGEERYSYLLSEIEGEEE
NKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRL
TRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKDPMNFVKKQEKA
RSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKSSIEKYGGYTNINKACFAVLD
IKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSE
NDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLSSKTNNIGN
YQEFKIINQSITGLFENEVDLLKLGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDES
TDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLT
SDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 18);

[0066] (d)

MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGKEYHIGLAIGTSSIGWAVT
DSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLLKKNERGYP
TLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKI
GQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQNDILTEITSLFSQI
MLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVI
PHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIK
RMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKS
IVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQIDQ
LNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAP
DLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAKMKPYWNQLLKQGLITQRKFENLTMDVDQ
TIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKH
GSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQNKKSKLVVPLKNKLQPEYYGA
YTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDI
KLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFGSDLQNLKEIGFGTTPLGQFQ
VSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNLGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDIL
VHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDEST
DENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 90);

[0067] (e)

[0068] MPAAKRVKLDTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPC
WECSQAIREFLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKI
SRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSPKKKRKVGEKKTNYTIGLAIGTDSVGW

AVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFSKNPIFG
TLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRT
VKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLS
AQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQV
HLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLIYQ
KYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMMDDPDNEDLMEEIV
KILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMID
PSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQN
GKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQR
QLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVYGEFRYKKKPQA
EKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSKYGGYIEEKMAYYAVI
SYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLVNFIYHLKNYDEISH
PESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSMTGLFDGQVIYQS
LTGLYETRMRFEDGKRPAATKKAGQAKKKKGSSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWAL
VIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 93); 또는

[0069] (f)

[0070] MPAAKRVKLDTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGG
STNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSPKKKRKVE
KKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLN
ESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELR
KEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYNAV
ELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKE
DFLRKQRSFYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGR
DTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPG
MKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLI
NDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENL
PGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDNLT
KAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPY
MADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLD
PSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAME
QDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISR
RRYNSMTGLFDGQVIYQSLTGLYETRMRFEDKRPAATKKAGQAKKKKGSSGGSSGGSSGSETPGTSESATPESSGGSSGGSTSEKGPSTGDPTLRRRIESWE
FDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARLFWHMD
QRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLAT
GLIHPSVTWRYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 94)와 적어도 80% 동일한 아미노산 서열을 가진다.

[0071] 일부 구현예에서, *스트렙토코커스 콘스텔라투스* Cas9 단백질은 5'-NGG-3'을 포함하는 PAM 서열을 인식한다.

[0072] 일부 구현예에서, *스트렙토코커스 콘스텔라투스* Cas9 단백질은 5'-NGC-3'을 포함하는 PAM 서열을 인식한다.

[0073] 일부 구현예에서, 본원에 개시된 Cas9 단백질(예: SirCas9, VapCas9, EpeCas9, LfeCas9 또는 PmaCas9)은 5'-NGC-3'을 포함하는 PAM 서열을 인식한다.

[0074] 일부 구현예에서, *샤르페아* Cas9 단백질은 5'-NAGHC-3'을 포함하는 PAM 서열을 인식하고, H는 A, C 또는 T이다.

[0075] 일부 구현예에서, *베일로넬라 파르불라* Cas9 단백질은 5'-NRHRRH-3'을 포함하는 PAM 서열을 인식하고, H는 아데닌, 시토신 또는 티민이고, R은 아데닌 또는 구아닌이다.

[0076] 일부 구현예에서, *에자키엘라 페루엔시스* Cas9 단백질은 5'-NGG-3'을 포함하는 PAM 서열을 인식한다.

[0077] 일부 구현예에서, *락토바실러스 퍼멘텀* 균주 AF15-40LB Cas9 단백질은 5'-NGG-3'을 포함하는 PAM 서열을 인식한다.

[0078] 일부 구현예에서, *펩토니필러스 sp. 마르세유-P3761* Cas9 단백질은 5'-NNAAA-3'을 포함하는 PAM 서열을 인식한

다.

[0079] 일부 구현예에서, 상기 Cas9 단백질을 암호화하는 핵산이 제공된다.

[0080] 일부 구현예에서, 상기 핵산은 포유동물 세포에서 발현되도록 코돈-최적화된다.

[0081] 일부 구현예에서, 상기 핵산은 인간 세포에서 발현되도록 코돈-최적화된다.

[0082] 일부 구현예에서, 상기 Cas9 단백질을 포함하는 진핵 세포가 제공된다.

[0083] 일부 구현예에서, 상기 세포는 인간 세포이다. 일부 구현예에서, 상기 세포는 식물 세포이다.

[0084] 일 측면에서, 세포를 본원에 기술된 바와 같은 Cas9, 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하고, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열의 파손을 유발할 수 있는, 진핵 세포에서 표적 핵산을 절단하는 방법이 제공된다.

[0085] 일 측면에서, 세포를 본원에 기술된 바와 같은 Cas9, 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하고, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열의 파손을 유발할 수 있는, 진핵 세포에서 표적 핵산의 발현을 변경하는 방법이 제공된다.

[0086] 일 측면에서, 세포를 본원에 기술된 바와 같은 Cas9, 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하고, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있는, 진핵 세포에서 표적 핵산의 발현을 변경하는 방법이 제공된다.

[0087] 일 측면에서, 세포를 본원에 기술된 바와 같은 Cas9, 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하고, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있는, 진핵 세포에서 표적 핵산을 변형하는 방법이 제공된다.

[0088] 일부 구현예에서, 상기 Cas9 단백질은 불활성 Cas9(dCas9)이다.

[0089] 일부 구현예에서, 상기 dCas9는 데아미나제에 융합된다.

[0090] 일부 구현예에서, 상기 RNA 가이드는 crRNA 및 tracrRNA를 포함한다.

[0091] 일부 구현예에서, 상기 RNA 가이드는 sgRNA를 포함한다.

[0092] 일부 구현예에서, *스트렙토코커스 콘스텔라투스* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUUAGAGCUGUGCUGUUUAAACAACACAGCAAGUUAAAAUAAGGCUUUGUCCGUACUCAAGCUUGCAAAAGCGUGCACCGAUUCGGUGCU-3' (서열번호 3)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함한다.

[0093] 일부 구현예에서, *샤르페아* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUUAGAGUUGUGUUAUUGAAAAAUAACACAACGAGUUAAAAUAAAGCUUAUGCUUAAAUGCCAGCUUUGCUGGUGUCAUUUAGAUGACUUUACUAAGGUUGCUUCGGCAACCUUUUU-3' (서열번호 7)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함한다.

[0094] 일부 구현예에서, *베일로넬라 파르불라* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUGAGAGUAGUGUGAAAACAUUACGAGUUCAAAUACAAAUUAAUUUACAAUGCCUUCGGGCUGCCCGACGUAGGGCACCUACUCUCAAUUCUUCGGAAUUGAGUU-3' (서열번호 13)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함한다.

[0095] 일부 구현예에서, *에자키엘라 페루엔시스* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUUUGAGAGUUAUGUAAUUGAAAAAUUACAUGACGAGUUCAAAUAAAAAUUUAUUCAAACCGCCUAUUUAUAGGCCGCAGAUGUUCUGCAUUAUGCUUGCUAUUGCAAGCUU-3' (서열번호 19)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함한다.

[0096] 일부 구현예에서, *락토바실러스 퍼멘텀 균주 AF15-40LB* Cas9와 함께 사용하기 위한 sgRNA는 5'-GUCUUGGAUGAGUGUGAAAACACUCAUAGUCAAGAUCAAACGAGUGGUUUUCCACGAGUUAUUACUUUUGAGGUCUUAUAUGGCCCAUACAUAAAAAGGAGUCGGAAUUUCCGGCUCCUUUUCUU-3' (서열번호 95)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함한다.

[0097] 일부 구현예에서, *펩토니필러스 sp. 마르세유-P3761* Cas9와 함께 사용하기 위한 sgRNA는 5'-

GUUUUAGAGCCAUGUAGAAAUACAUUGCAAGUUAAAAUAAGGCUUUGUCCGUAAUCAACUUGAAAAAGUGGCGCUGUUUCGGCGCUUU-3' (서열번호 96)와 적어도 약 80% 동일성을 갖는 서열을 포함하는 스캐폴드를 포함한다.

[0098] 일부 구현예에서, 상기 crRNA는 약 16 내지 26개의 뉴클레오티드 길이의 가이드 서열을 포함한다.

[0099] 일부 구현예에서, 상기 crRNA는 18 내지 24개의 뉴클레오티드 길이의 가이드 서열을 포함한다.

[0100] 일부 구현예에서, 표적 핵산에서의 파손은 단일 가닥 또는 이중 가닥의 파손이다.

[0101] 일부 구현예에서, 표적 핵산에서의 파손은 단일 가닥의 파손이다.

[0102] 일부 구현예에서, Cas9 단백질은 상기 표적 핵산 서열의 두 가닥 모두를 절단하는 뉴클레아제이다. 일부 구현예에서, Cas9는 상기 표적 핵산 서열의 한 가닥을 절단하는 닉카제이다.

[0103] 일부 구현예에서, 상기 표적 핵산은 5' 내지 프로토스페이서 인접 모티프(PAM)이다.

[0104] 일부 구현예에서, 상기 Cas9는 진핵 세포에서 발현되도록 프로모터 서열에 작동 가능하게 연결되고, 상기 가이드 RNA는 진핵 세포에서 발현되도록 프로모터 서열에 작동 가능하게 연결된다.

[0105] 일부 구현예에서, 상기 진핵 세포는 인간 세포이다.

[0106] 일부 구현예에서, 상기 프로모터 서열은 진핵 또는 바이러스의 프로모터이다.

[0107] 일 측면에서, RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산; 및 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 80% 서열 동일성을 갖는 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함하는 조작된 비천연 발생 CRISPR-Cas 시스템이 본원에 제공되고, 상기 RNA 가이드는 직접 반복 서열, 및 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas 단백질은 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열에서의 파손을 유발할 수 있다.

[0108] 일부 구현예에서, 서열번호 2, 5, 9, 15, 85, 87, 95 또는 96과 적어도 80% 서열 동일성을 갖는 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함하는 조작된 비천연 발생 CRISPR-Cas 시스템이 본원에 제공되고, 상기 Cas 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열에서의 파손을 유발할 수 있다.

[0109] 일 측면에서, RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산; 및 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 80% 서열 동일성을 갖는 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함하는 조작된 비천연 발생 CRISPR-Cas 시스템이 본원에 제공되고, 상기 RNA 가이드는 직접 반복 서열, 및 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas 단백질은 데아미나제에 융합되고, Cas 단백질 융합은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있다.

[0110] 일부 구현예에서, 조작된 비천연 발생 CRISPR-Cas 시스템은 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함하는 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함한다.

[0111] 일부 구현예에서, 조작된 비천연 발생 CRISPR-Cas 시스템은 서열번호 2, 5, 9, 15, 85, 87, 95 또는 96과 적어도 80% 서열 동일성을 갖는 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함하고, 상기 Cas 단백질은 데아미나제에 융합되고, 상기 Cas 단백질 융합은 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있다.

[0112] 일 구현예에서, 상기 Cas9 단백질은 불활성 Cas9(dCas9)이다.

[0113] 일 구현예에서, 상기 RNA 가이드는 crRNA 및 tracrRNA를 포함한다.

[0114] 일 구현예에서, 상기 RNA 가이드는 sgRNA를 포함한다.

[0115] 일 구현예에서, 상기 Cas 단백질은 진핵 세포에서 발현되도록 프로모터 서열에 작동 가능하게 연결되고, 상기 가이드 RNA는 진핵 세포에서 발현되도록 프로모터 서열에 작동 가능하게 연결된다.

[0116] 일 구현예에서, 상기 진핵 세포는 인간 세포이다.

[0117] 일 구현예에서, 상기 프로모터 서열은 진핵 프로모터 서열이다.

[0118] 일 구현예에서, 본원에 기술된 시스템을 암호화하는 핵산이 제공된다.

[0119] 일 구현예에서, 본원에 기술된 시스템을 포함하는 벡터가 제공된다.

[0120] 일 구현예에서, 상기 벡터는 플라스미드 벡터 또는 바이러스 벡터이다.

[0121] 일 구현예에서, 상기 바이러스 벡터는 아데노 연관 바이러스(AAV) 벡터 또는 렌티바이러스 벡터이다.

[0122] 일 구현예에서, 상기 바이러스 벡터는 AAV 벡터이다.

[0123] 일 구현예에서, 1개 초과의 AAV 벡터는 상기 시스템을 패키징하기 위해 사용된다.

[0124] 일 구현예에서, 장애 또는 질병 치료를 필요로 하는 대상체에서 장애 또는 질병을 치료하는 방법은 본원에 기술된 시스템을 상기 대상체에게 투여하는 것을 포함하고, 상기 가이드 RNA는 병태 또는 질병과 연관된 표적 핵산의 적어도 10개 뉴클레오티드에 상보적이고; 상기 Cas 단백질은 상기 가이드 RNA와 연관되고; 상기 가이드 RNA는 상기 표적 핵산에 결합하고; 상기 Cas 단백질은 표적 핵산에서의 파손을 유발하고, 선택적으로 상기 Cas9는 데아미나제에 융합된 불활성 Cas9(dCas9)이고 상기 표적 핵산에서 하나 이상의 염기 편집을 초래하여, 상기 장애 또는 질병을 치료한다.

[0125] 일부 구현예에서, 상기 가이드 RNA는 약 18 내지 24개의 뉴클레오티드에 상보적이다.

[0126] 일부 구현예에서, 상기 가이드 RNA는 20개의 뉴클레오티드에 상보적이다.

[0127] 일부 구현예에서, 상기 염기 편집기는 융합 단백질을 포함한다.

[0128] 일부 구현예에서, 상기 염기 편집기는 아데노신 데아미나제 도메인 또는 시티딘 데아미나제 도메인을 포함한다.

[0129] 일부 구현예에서, 폴리뉴클레오티드의 핵염기를 편집하는 방법이 본원에 제공되고, 상기 방법은 1개 이상의 가이드 RNA와의 복합체에서 염기와 폴리뉴클레오티드를 접촉시키는 것을 포함하고, 상기 염기 편집기는 아데노신 데아미나제 도메인을 포함하고, 상기 1개 이상의 가이드 RNA는 폴리뉴클레오티드에서 A·T에서 G·C로 변경을 이루도록 염기 편집기를 표적으로 한다.

[0130] 일부 구현예에서, 폴리뉴클레오티드의 핵염기를 편집하는 방법이 본원에 제공되고, 상기 방법은 1개 이상의 가이드 RNA와의 복합체에서 염기 편집기와 폴리뉴클레오티드를 접촉시키는 것을 포함하고, 상기 염기 편집기는 시티딘 데아미나제 도메인을 포함하고, 상기 1개 이상의 가이드 RNA는 폴리뉴클레오티드에서 G·C에서 T·A로 변경을 이루도록 염기 편집기를 표적으로 한다.

[0131] 일부 구현예에서, 상기 편집이 표적 폴리뉴클레오티드 서열에서 50% 미만의 삽입결실 형성을 초래한다.

[0132] 일부 구현예에서, 상기 편집이 점 돌연변이를 발생시킨다.

[0133] **정의**

[0134] 본 발명이 보다 쉽게 이해되도록, 일부 용어가 다음에서 먼저 정의된다. 다음 용어 및 다른 용어에 대한 추가 정의가 명세서 전반에 명시되어 있다.

[0135] *부정관사(A 또는 An)*: 부정관사는 하나 또는 1개 초과의(즉, 적어도 하나의) 문법상 물품 객체를 지칭하기 위해 본원에서 사용된다. 예를 들어, "하나의 요소"는 1개 요소 또는 1개보다 많은 요소를 의미한다.

[0136] *대략 또는 약*: 본원에서 사용될 때 용어 "대략" 또는 "약"은 1개 이상의 관심 값에 적용되는 경우에 명시된 참조 값과 유사한 값을 의미한다. 일부 구현예에서, 용어 "대략" 또는 "약"은 문맥에서 달리 명시하거나 달리 명백하지 않는 한, 명시된 참조 값의 한 방향으로(초과 또는 미만으로) 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1% 이하 이내에 속하는 값의 범위를 의미한다(해당 숫자가 가능한 값의 100%를 초과하는 경우는 예외).

[0137] *연관된*: 상기 용어가 본원에서 사용될 때 하나의 존재, 수치 및/또는 형태가 다른 것과 상관관계를 가지는 경우 2개의 사건 또는 단위체가 서로에 "연관된"다. 예를 들어, 특정 단위체(예: 폴리펩티드)는 그 존재, 수치 및/또는 형태가 질병, 장애 또는 병태의 발생률 및/또는 감수성과 상관관계를 가지는 경우(예: 관련 모집단 전체에 걸쳐) 특정 질병, 장애 또는 병태와 연관된 것으로 간주한다. 일부 구현예에서, 2개 이상의 단위체가 직접적으로 또는 간접적으로 상호작용하여 서로에 물리적으로 근접하게 유지되는 경우 서로에 물리적으로 "연관된"다. 일부 구현예에서, 서로에 물리적으로 연관된 2개 이상의 단위체는 서로 공유적으로 연결되며; 일부 구현예에서, 서로에 물리적으로 연관된 2개 이상의 단위체는 서로 공유적으로 연결되지 않지만 예를 들어 수소 결합, 반 데르 발스 상호작용, 소수성 상호작용, 자성 및 이들의 조합에 의해 비공유적으로 연관된다.

[0138] *염기 편집기*: "염기 편집기(BE)" 또는 "핵염기 편집기(NBE)"라 함은 폴리뉴클레오티드에 결합하고 핵염기 변형 활성을 갖는 작용제(agent)를 의미한다. 다양한 구현예에서, 염기 편집기는 핵염기 변형 폴리펩티드(예를 들어, 데아미나제) 및 가이드 폴리뉴클레오티드(예를 들어, 가이드 RNA)와 함께 폴리뉴클레오티드 프로그래밍 가능한 뉴클레오티드 결합 도메인을 포함한다. 다양한 구현예에서, 작용제는 염기 편집 활성을 갖는 단백질 도메인, 즉 핵산 분자(예를 들어, DNA) 내에 염기(예를 들어, A, T, C, G 또는 U)를 변형시킬 수 있는 도메인을 포함하는 생체분자 복합체이다. 일부 구현예에서, 폴리뉴클레오티드 프로그래밍 가능한 DNA 결합 도메인은 데아미나제 도메인에 융합되거나 연결된다. 일 구현예에서, 작용제는 염기 편집 활성을 갖는 1개 이상의 도메인을 포함하는 융합 단백질이다. 또 다른 구현예에서, 염기 편집 활성을 갖는 단백질 도메인은 가이드 RNA에 연결된다(예를 들어, 가이드 RNA 상의 RNA 결합 모티프 및 데아미나제에 융합된 RNA 결합 도메인을 통함). 일부 구현예에서, 염기 편집 활성을 갖는 도메인은 핵산 분자 내의 염기를 탈아미노화할 수 있다. 일부 구현예에서, 염기 편집기는 DNA 분자 내의 1개 이상의 염기를 탈아미노화할 수 있다. 일부 구현예에서, 염기 편집기는 DNA 내의 시토신(C) 또는 아데노신(A)을 탈아미노화할 수 있다. 일부 구현예에서, 염기 편집기는 DNA 내의 시토신(C)과 아데노신(A)을 탈아미노화할 수 있다. 일부 구현예에서, 염기 편집기는 시티딘 염기 편집기(CBE)이다. 일부 구현예에서, 염기 편집기는 아데노신 염기 편집기(ABE)이다. 일부 구현예에서, 염기 편집기는 아데노신 염기 편집기(ABE) 및 시티딘 염기 편집기(CBE)이다. 일부 구현예에서, 염기 편집기는 아데노신 데아미나제에 융합된 뉴클레아제-불활성 Cas9(dCas9)이다. 일부 구현예에서, 염기 편집기는 염기 절제 복구 억제제, 예를 들어 UGI 도메인, 또는 dISN 도메인에 융합된다. 일부 구현예에서, 융합 단백질은 데아미나제 및 염기 절제 복구 억제제, 예컨대 UGI 또는 dISN 도메인에 융합된 Cas9 닉카제를 포함한다. 다른 구현예에서, 염기 편집기는 무염기 염기 편집기이다. 염기 편집기에 대한 상세한 설명은 국제 PCT 출원 PCT/2017/045381호(WO2018/027078호) 및 PCT/US2016/058344호(WO2017/070632호)에 기술되어 있으며, 이들 각각은 전체적으로 본원에 참고로 통합된다. 또한 Komor, A.C., 등, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., 등, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); Komor, A.C., 등, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017), 및 Rees, H.A., 등, "Base editing: precision chemistry on the genome and transcriptome of living cells." Nat Rev Genet. 2018 Dec;19(12):770-788. doi: 10.1038/s41576-018-0059-1을 참조하며, 이들 전체 내용은 본원에 참고로 통합된다.

[0139] *염기 편집 활성*: "염기 편집 활성"이라 함은 폴리뉴클레오티드 내의 염기를 화학적으로 변경시키는 작용을 의미한다. 일 구현예에서, 제1 염기는 제2 염기로 전환된다. 일 구현예에서, 염기 편집 활성은 시티딘 데아미나제 활성으로, 예를 들어 표적 C·G를 T·A로 전환시킨다. 또 다른 구현예에서, 염기 편집 활성은 아데노신 또는 아데닌 데아미나제 활성으로, 예를 들어 A·T를 G·C로 전환시킨다. 또 다른 구현예에서, 염기 편집 활성은 시토신 또는 시티딘 데아미나제 활성으로, 예를 들어 표적 C·G를 T·A로 전환시키고 아데노신 또는 아데닌 데아미나제 활성으로, 예를 들어 A·T를 G·C로 전환시킨다.

[0140] *염기 편집기 시스템*: 용어 "염기 편집기 시스템"은 표적 뉴클레오티드 서열의 핵염기를 편집하기 위한 시스템을 의미한다. 다양한 구현예에서, 염기 편집기(BE) 시스템은 (1) 표적 뉴클레오티드 서열에서 핵염기를 탈아미노화하기 위한 폴리뉴클레오티드 프로그래밍 가능한 뉴클레오티드 결합 도메인(예를 들어, Cas9), 데아미나제 도메인 및 시티딘 데아미나제 도메인; 및 (2) 폴리뉴클레오티드 프로그래밍 가능한 뉴클레오티드 결합 도메인과 함께 1개 이상의 가이드 폴리뉴클레오티드(예를 들어, 가이드 RNA)를 포함한다. 다양한 구현예에서, 염기 편집기(BE) 시스템은 아데노신 데아미나제 또는 시티딘 데아미나제로부터 선택된 핵염기 편집기 도메인, 및 핵산 서열 특이적 결합 활성을 갖는 도메인을 포함한다. 일부 구현예에서, 염기 편집기 시스템은 (1) 표적 뉴클레오티드 서열에서 1개 이상의 핵염기를 탈아미노화하기 위한 폴리뉴클레오티드 프로그래밍 가능한 DNA 결합 도메인 및 데아미나제 도메인을 포함하는 염기 편집기(BE); 및 (2) 폴리뉴클레오티드 프로그래밍 가능한 DNA 결합 도메인과 함께 1개 이상의 가이드 RNA를 포함한다. 일부 구현예에서, 폴리뉴클레오티드 프로그래밍 가능한 뉴클레오티드 결합 도메인은 폴리뉴클레오티드 프로그래밍 가능한 DNA 결합 도메인이다. 일부 구현예에서, 염기 편집기는 시티딘 염기 편집기(CBE)이다. 일부 구현예에서, 염기 편집기는 아데닌 또는 아데노신 염기 편집기(ABE)이다. 일부 구현예에서, 염기 편집기는 아데닌 또는 아데노신 염기 편집기(ABE) 또는 시티딘 염기 편집기(CBE)이다.

[0141] 일부 구현예에서, 폴리뉴클레오티드 프로그래밍 가능한 뉴클레오티드 결합 도메인은 데아미나제 도메인과 비공유적으로 상호작용하거나 연관됨으로써 데아미나제 도메인을 표적 뉴클레오티드 서열에 대해 표적으로 할 수 있다. 예를 들어, 일부 구현예에서, 핵염기 편집 구성요소, 예를 들어 데아미나제 구성요소는 폴리뉴클레오티드

프로그래밍 가능한 뉴클레오타이드 결합 도메인의 일부인 추가적인 이종 부분 또는 도메인과 상호작용할 수 있거나 연관될 수 있거나 복합체를 형성할 수 있는 추가적인 이종 부분 또는 도메인을 포함할 수 있다. 일부 구현예에서, 추가적인 이종 부분은 폴리펩티드와 결합하거나, 상호작용하거나, 연관되거나, 복합체를 형성할 수 있다. 일부 구현예에서, 추가적인 이종 부분은 폴리뉴클레오티드와 결합하거나, 상호작용하거나, 연관되거나, 복합체를 형성할 수 있다. 일부 구현예에서, 추가적인 이종 부분은 가이드 폴리뉴클레오티드와 결합할 수 있다. 일부 구현예에서, 추가적인 이종 부분은 폴리펩티드 링커와 결합할 수 있다. 일부 구현예에서, 추가적인 이종 부분은 폴리뉴클레오티드 링커와 결합할 수 있다. 추가적인 이종 부분은 단백질 도메인일 수 있다. 일부 구현예에서, 추가적인 이종 부분은 K 상동체(KH) 도메인, MS2 코트 단백질 도메인, PP7 코트 단백질 도메인, SfMu Com 코트 단백질 도메인, 스테릴 알파 모티프, 텔로머라제 Ku 결합 모티프 및 Ku 단백질, 텔로머라제 Sm7 결합 모티프 및 Sm7 단백질, 또는 RNA 인식 모티프일 수 있다.

[0142] *생물학적 활성*: 본원에서 사용될 때 "생물학적 활성" 문구는 생물학적 시스템에서, 및 특히 유기체에서 활성을 갖는 임의의 작용제의 특징을 의미한다. 예를 들어, 유기체에 투여될 때 해당 유기체에 생물학적 효과를 갖는 작용제는 생물학적 활성인 것으로 간주된다. 특정 구현예에서, 펩티드가 생물학적 활성인 경우, 펩티드의 적어도 하나의 생물학적 활성을 공유하는 펩티드 부분은 일반적으로 "생물학적 활성" 부분으로 지칭된다.

[0143] *절단*: 본원에서 사용될 때 절단은 본원에 기술된 CRISPR 시스템의 뉴클레아제에 의해 형성된 표적 핵산에서의 파손을 의미한다. 일부 구현예에서, 절단 사건은 이중 가닥 DNA의 파손이다. 일부 구현예에서, 절단 사건은 단일 가닥 DNA의 파손이다. 일부 구현예에서, 절단 사건은 단일 가닥 RNA의 파손이다. 일부 구현예에서, 절단 사건은 이중 가닥 RNA의 파손이다.

[0144] *상보적*: 본원에서 사용될 때 상보적은 왓슨-크릭 염기 쌍을 형성하여 A 염기가 T와 쌍을 이루고 C 염기가 G와 쌍을 이루거나 두 번째 핵산 가닥 상에서 염기와 비전형적인 염기 쌍을 형성하도록 하는 핵산 가닥을 의미한다. 즉, 이는 적절한 조건 하에서 서로 혼성화되는 핵산을 의미한다.

[0145] *일정한 간격으로 놓인 짧은 회문 반복부 집합(CRISPR)-연관(Cas) 시스템*: 본원에서 사용될 때 CRISPR-Cas9 시스템은 CRISPR 이펙터를 암호화하는 서열을 포함하는 CRISPR-이펙터, RNA 가이드, 및 CRISPR 유전자좌의 기타 서열 및 전사체의 발현에 수반되거나 그 활성을 지시하는 핵산 및/또는 단백질을 의미한다. 일부 구현예에서, CRISPR 시스템은 조작된 비천연 발생 CRISPR 시스템이다. 일부 구현예에서, CRISPR 시스템의 구성요소는 시스템의 1개 이상의 구성요소를 암호화하는 핵산(들)(예를 들어, 벡터), 단백질 형태의 구성요소(들) 또는 이들의 조합을 포함할 수 있다.

[0146] *CRISPR 어레이*: 본원에서 사용될 때 용어 "CRISPR 어레이"는 CRISPR 반복부, 및 첫 번째 CRISPR 반복부의 첫 번째 뉴클레오티드에서 시작하고 마지막(말단) CRISPR 반복부의 마지막 뉴클레오티드에서 끝나는 스페이서를 포함하는 핵산(예를 들어, DNA) 세그먼트를 의미한다. 일반적으로, CRISPR 어레이의 각 스페이서는 2개 반복부 사이에 위치한다. 본원에서 사용될 때 용어 "CRISPR 반복부" 또는 "CRISPR 직접 반복부" 또는 "직접 반복부"는 다수의 짧은 직접 반복 서열을 의미하며, 이는 CRISPR 어레이 내 서열 변이를 거의 또는 전혀 나타내지 않는다.

[0147] *CRISPR-연관 단백질(Cas)*: 본원에서 사용될 때 용어 "CRISPR-연관 단백질", "CRISPR 이펙터", "이펙터" 또는 "CRISPR 효소"는 효소 활성을 수행하거나 RNA 가이드에 의해 지정된 핵산 상의 표적 부위에 결합하는 단백질을 의미한다. 다른 구현예에서, CRISPR 이펙터는 엔도뉴클레아제 활성, 닉카제 활성, 엑소뉴클레아제 활성, 트랜스포사제 활성 및/또는 절제 활성을 가진다. 일부 구현예에서, Cas는 고정밀 Cas이다. 일부 구현예에서, Cas는 고충실 Cas이다. 일부 구현예에서, Cas는 SuperFi-Cas이다. 일부 구현예에서, 고정밀, 고충실 및 SuperFi-Cas는 Bravo, J. 등 Structural basis for mismatch surveillance by CRISPR-Cas9 *Nature*, 603, March 2022에 기술된 바와 같다.

[0148] *crRNA*: 본원에서 사용될 때 용어 "CRISPR RNA" 또는 "crRNA"는 특정 핵산 서열을 표적으로 하기 위해 CRISPR 이펙터에 의해 사용되는 가이드 서열을 포함하는 RNA 분자를 의미한다. 일반적으로, crRNA는 표적 인식을 매개하는 서열과 tracrRNA와 이중구조를 형성하는 서열을 함유한다. 일부 구현예에서, crRNA:tracrRNA 이중구조는 CRISPR 이펙터에 결합한다.

[0149] *생체외*: 본원에서 사용될 때 용어 "생체외"는 다세포 유기체 내부가 아닌 외부에서 성장한 세포 또는 조직에서 발생한 사건을 의미한다.

[0150] *기능적 등가물 또는 유사체*: 본원에서 사용될 때 용어 "기능적 등가물" 또는 "기능적 유사체"는 아미노산 서열의 기능적 유도체와 관련하여 원래 서열과 실질적으로 유사한 생물학적 활성(기능적 또는 구조적)을 유지하는

분자를 지칭한다. 기능적 유도체 또는 등가물은 천연 유도체일 수 있거나 합성으로 제조된다. 예시적인 기능적 유도체는 단백질의 생물학적 활성이 보존된다는 전제 하에 하나 이상의 아미노산의 치환, 결실 또는 추가가 있는 아미노산 서열을 포함한다. 아미노산 치환은 바람직하게는 치환된 아미노산과 유사한 화학적-물리적 특성을 가진다. 바람직한 유사한 화학적-물리적 특성에는 전하, 부피, 소수성, 친수성 등의 유사성이 포함된다.

[0151] *반감기*: 본원에서 사용될 때 용어 "반감기"는 단백질 농도 또는 활성과 같은 양이 기간 시작 시에 측정된 값의 절반으로 떨어지는 데 필요한 시간이다.

[0152] *개선하다, 증가하다 또는 낮추다*: 본원에서 사용될 때 용어 "개선하다", "증가하다" 또는 "낮추다" 또는 문법적 동등어는 본원에 기술된 치료를 시작하기 전의 동일 개체에서 측정치 또는 본원에 기술된 치료를 받지 않은 대조 대상체(또는 다수의 대조 대상체)에서 측정치와 같은 기준선 측정치와 비교한 값을 나타낸다. "대조 대상체"는 치료받는 대상체와 동일 형태의 질병을 갖고 치료받는 대상체와 거의 동일한 연령을 갖는 대상체이다.

[0153] *억제*: 본원에서 사용될 때 용어 "억제", "억제하다" 및 "억제함"은 관심 단백질 또는 유전자의 활성 및/또는 발현을 감소시키거나 낮추는 프로세스 또는 방법을 의미한다. 전형적으로, 단백질 또는 유전자를 억제함은 본원에 기술되거나 당업계에서 인식된 하나 이상의 방법에 의해 측정될 때 단백질 또는 유전자의 발현 또는 관련 활성을 적어도 10% 이상, 예를 들어 20%, 30%, 40% 또는 50%, 60%, 70%, 80%, 90% 이상 낮추거나 발현 또는 관련 활성을 1배, 2배, 3배, 4배, 5배, 10배, 50배, 100배 초과하여 감소시킴을 의미한다.

[0154] *혼성화*: 본원에서 사용될 때 용어 "혼성화"는 둘 이상의 핵산이 두 핵산의 염기 사이에 왓슨-크릭 쌍에 의한 수소 결합, 후그스타인(Hoogstein) 결합 또는 다른 서열-특이적 결합을 통해 서로 결합하는 반응을 의미한다. 다른 서열과 혼성화할 수 있는 서열은 서열의 "상보"라고 하며, "상보적"이라고 하거나 "상보성"을 보인다고 한다.

[0155] *삽입결실*: 본원에서 사용될 때 용어 "삽입결실"은 핵산 서열에서 염기의 삽입 또는 결실을 의미한다. 이는 일반적으로 돌연변이를 초래하고, 유전적 변이의 일반적인 형태이다.

[0156] *시험관내*: 본원에서 사용될 때 용어 "시험관내"는 다세포 유기체 내에서가 아니라 인위적 환경, 예를 들어 시험관이나 반응 용기 내에서, 세포 배양물 등에서 발생하는 사건을 의미한다.

[0157] *생체내*: 본원에서 사용될 때 용어 "생체내"는 인간 및 비인간 동물과 같은 다세포 유기체 내에서 발생하는 사건을 의미한다. 세포 기반 시스템의 맥락에서, 상기 용어는 살아있는 세포 내(예를 들어 시험관내 시스템과 반대임)에서 발생하는 사건을 지칭하기 위해 사용될 수 있다.

[0158] *링커*: 용어 "링커"는 둘 이상의 단위체를 연결하기 위해 사용되는 임의의 수단, 단위체 또는 모이어티를 의미한다. 일부 구현예에서, 링커는 공유 링커이다. 일부 구현예에서, 링커는 비공유 링커이다. 공유 링커의 예에는 공유 결합, 또는 연결될 단백질 또는 도메인의 하나 이상에 공유적으로 부착된 링커 모이어티가 포함된다. 일부 구현예에서, 링커는 비공유 결합, 예를 들어 백금 원자와 같은 금속 중심을 통한 유기금속 결합이다. 연결은 영구적 또는 가역적일 수 있다. 공유 연결의 경우 카본산 유도체를 포함한 아미드 기, 에테르, 유기 및 무기 에스테르를 포함한 에스테르, 아미노, 우레탄, 요소 등과 같은 다양한 작용기가 사용될 수 있다. 연결을 제공하기 위해 도메인은 산화, 수산화, 치환, 환원 등에 의해 변형되어 커플링을 위한 부위를 제공할 수 있다. 접합 방법은 당업자에게 잘 알려져 있고, 본 발명에서 사용을 위해 포함된다. 링커 모이어티는 화학적 링커 모이어티, 또는 예를 들어 펩티드 링커 모이어티(링커 서열)를 포함하지만 이에 국한되지는 않는다. RNA-결합 도메인 및 이펙터 도메인의 기능을 유의하게 감소시키지 않는 변형이 바람직하다는 것이 이해될 것이다.

[0159] *돌연변이*: 본원에서 사용될 때 용어 "돌연변이"는 당업계에서 일반적인 의미를 가지며, 예를 들어 점 돌연변이, 치환, 삽입, 결실, 역위 및 결실을 포함한다.

[0160] *올리고뉴클레오티드*: 본원에서 사용될 때 용어 "올리고뉴클레오티드"는 일반적으로 단일 또는 이중 가닥 DNA의 약 5 내지 약 100개 뉴클레오티드의 폴리뉴클레오티드를 의미한다. 올리고뉴클레오티드는 "올리고머" 또는 "올리고"라고도 알려져 있고, 유전자로부터 단리될 수 있거나 화학적으로 합성될 수 있다.

[0161] *PAM*: 용어 'PAM' 또는 '프로토스페이서 인접 모티프'는 CRISPR-Cas9와 같이 CRISPR 시스템에 의해 절단하도록 표적이 되는 핵산 영역이 잇따르는 짧은 핵산 서열(일반적으로 2 내지 6개 염기쌍 길이)을 의미한다. PAM은 Cas 뉴클레아제가 절단하는 데 필요하고, 일반적으로 절단 부위의 3 내지 4개 뉴클레오티드 다운스트림에서 발견된다.

[0162] *폴리펩티드*; 본원에서 사용될 때 용어 "폴리펩티드"는 펩티드 결합을 통해 함께 연결된 일련의 아미노산 사슬을

의미한다. 이 용어는 임의의 길이의 아미노산 사슬을 지칭하기 위해 사용되지만, 당업자는 이 용어가 긴 사슬에 제한되지 않고 펩티드 결합을 통해 연결된 2개 아미노산을 포함하는 최소 사슬을 지칭할 수 있음을 이해할 것이다. 당업자에게 알려진 바와 같이, 폴리펩티드는 처리 및/또는 변형될 수 있다. 본원에서 사용될 때 용어 "폴리펩티드" 및 "펩티드"는 상호교환적으로 사용된다.

[0163] *예방하다*: 본원에서 사용될 때 용어 "예방하다" 또는 "예방"은 질병, 장애 및/또는 병태의 발생과 관련하여 사용되는 경우 질병, 장애 및/또는 병태가 발생할 위험을 낮추는 것을 의미한다.

[0164] *단백질*: 본원에서 사용될 때 용어 "단백질"은 개별 단위체로 기능하는 1개 이상의 폴리펩티드를 의미한다. 단일 폴리펩티드가 개별 기능 단위체이고, 개별 기능 단위체를 형성하기 위해 다른 폴리펩티드와의 영구적인 또는 일시적인 물리적 연관을 필요로 하지 않는 경우, 용어 "폴리펩티드"와 "단백질"은 상호교환적으로 사용될 수 있다. 개별 기능 단위체가 서로 물리적으로 연관되는 1개보다 많은 폴리펩티드로 구성되는 경우, 용어 "단백질"은 물리적으로 커플링되어 개별 단위체로 함께 기능하는 다수의 폴리펩티드를 의미한다.

[0165] *참조*: "참조" 단위체, 시스템, 양, 조건 설정 등은 테스트 단위체, 시스템, 양, 조건 설정 등이 본원에 기술된 바와 같이 비교되는 대상이다. 예를 들어, 일부 구현예에서, "참조" 항체는 본원에 기술된 바와 같이 조작되지 않은 대조 항체이다.

[0166] *RNA 가이드*: 용어 RNA 가이드는 표적 핵산에 대해 본원에 기술된 단백질을 표적으로 하는 것을 촉진하는 RNA 분자를 의미한다. 예시적인 "RNA 가이드" 또는 "가이드 RNA"는 crRNA, 또는 동족 tracrRNA와 조합된 crRNA를 포함하지만 이에 국한되지는 않는다. 후자는 독립적인 RNA이거나 링커를 사용하여 단일 RNA로 융합될 수 있다(sgRNA). 일부 구현예에서, RNA 가이드는 화학적 또는 생화학적 변형을 포함하도록 조작되고, 일부 구현예에서, RNA 가이드는 1개 이상의 뉴클레오티드를 포함할 수 있다.

[0167] *대상체*: 본원에서 사용될 때 용어 "대상체"는 진단, 예후 또는 요법이 요구되는 임의의 대상체를 의미한다. 예를 들어, 대상체는 포유동물, 예를 들어 인간 또는 비인간 영장류(예컨대 유인원, 원숭이, 오랑우탄 또는 침팬지), 개, 고양이, 기니피그, 토끼, 랫트, 마우스, 말, 소 또는 암소일 수 있다.

[0168] *sgRNA*: 용어 "sgRNA" 또는 "단일 가이드 RNA"는 (i) 가이드 서열(crRNA 서열) 및 (ii) Cas9 뉴클레아제-모집 서열(tracrRNA)을 함유하는 단일 가이드 RNA를 의미한다.

[0169] *실질적 동일*: "실질적 동일" 문구는 아미노산 또는 핵산 서열 간의 비교를 지칭하기 위해 본원에서 사용된다. 당업자들이 인식하는 바와 같이 2개 서열이 상응하는 위치에 동일 잔기를 함유하는 경우 일반적으로 "실질적 동일"로 간주된다. 당업계에 잘 알려진 바와 같이, 아미노산 또는 핵산 서열은 뉴클레오티드 서열에 대한 BLASTN, 및 아미노산 서열에 대한 BLASTP, 갭 BLAST 및 PSI-BLAST와 같은 시판 컴퓨터 프로그램에서 이용 가능한 알고리즘을 포함하여 임의의 다양한 알고리즘을 사용하여 비교될 수 있다. 예시적인 이러한 프로그램은 Altschul, 등, Basic local alignment search tool, *J. Mol. Biol.*, 215(3): 403-410, 1990; Altschul, 등, *Methods in Enzymology*; Altschul 등, *Nucleic Acids Res.* 25:3389-3402, 1997; Baxevanis 등, *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins, Wiley, 1998;* 및 Misener, 등, *(eds.), Bioinformatics Methods and Protocols (Methods in Molecular Biology, Vol. 132), Humana Press, 1999*에 기술되어 있다. 동일 서열을 식별하는 것 외에, 상기에서 언급한 프로그램은 일반적으로 동일성 정도에 대한 지표를 제공한다. 일부 구현예에서, 두 서열이 상응하는 잔기의 적어도 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 이상이 잔기의 관련 구간에 걸쳐 동일한 경우 실질적 동일로 간주된다. 일부 구현예에서, 관련 구간은 완전 서열이다. 일부 구현예에서, 관련 구간은 적어도 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475 또는 500개 이상의 잔기이다.

[0170] *표적 핵산*: 본원에서 사용될 때 용어 "표적 핵산"은 CRISPR-Cas9 시스템이 결합되는 임의의 길이의 뉴클레오티드(올리고뉴클레오티드 또는 폴리뉴클레오티드)인 데옥시리보뉴클레오티드, 리보뉴클레오티드 또는 이들의 유사체를 의미한다. 표적 핵산은 암호화 또는 비암호화 영역을 포함하여 3차원 구조를 가질 수 있고, 엑손, 인트론, mRNA, tRNA, rRNA, siRNA, shRNA, miRNA, 리보자임, cDNA, 플라스미드, 벡터, 외인성 서열, 내인성 서열을 포함할 수 있다. 표적 핵산은 변형된 뉴클레오티드를 포함할 수 있고, 메틸화 뉴클레오티드 또는 뉴클레오티드 유사체를 포함할 수 있다. 표적 핵산은 비핵산 구성요소와 함께 배치될 수 있다. 표적 핵산은 단일-, 이중- 또는 다중-가닥의 DNA 또는 RNA, 게놈 DNA, cDNA, DNA-RNA 하이브리드, 또는 퓨린 및 피리미딘 염기를 포함하는 중합체, 또는 기타 천연, 화학적 또는 생화학적 변형된, 비천연 또는 유도체화된 뉴클레오티드 염기에 제한되지 않

는다.

[0171] *치료적 유효량*: 본원에서 사용될 때 용어 "치료적 유효량"은 임의의 의학적 치료에 적용될 수 있는 합리적인 이익/위험 비로 치료 효과를 치료 대상체에게 부여하는 치료 분자(예를 들어 본원에 기술된 조작된 항체)의 양을 의미한다. 치료 효과는 객관적(즉, 일부 테스트 또는 마커에 의해 측정 가능함) 또는 주관적(즉, 대상체가 효과의 조짐 또는 느낌을 제공함)일 수 있다. 특히, "치료적 유효량"은 특정 질병 또는 병태를 치료, 개선 또는 예방하는 데 효과적이거나 예컨대 질병과 연관된 증상을 개선시키고/거나 질병 발병을 예방 또는 지연시키고/거나 질병 증상의 중증도 또는 빈도를 저감시킴으로써 감지 가능한 치료 또는 예방 효과를 나타내는 데 효과적인 치료 분자 또는 조성물의 양을 의미한다. 치료적 유효량은 다수의 단위 용량을 포함할 수 있는 용량 섭생법으로 투여될 수 있다. 임의의 특정 치료 분자의 경우, 치료적 유효량(및/또는 유효한 용량 섭생법 내의 적절한 단위 용량)은 예를 들어 투여 경로, 다른 약제와의 조합에 따라 달라질 수 있다. 또한, 임의의 특정 대상체를 위한 구체적인 치료적 유효량(및/또는 단위 용량)은 치료되는 장애 및 장애 중증도; 사용되는 구체적인 약제의 활성; 사용되는 구체적인 조성; 대상체의 연령, 체중, 전반적인 건강, 성별 및 식이요법; 사용되는 구체적인 치료 분자의 투여 시간, 투여 경로 및/또는 배설률 또는 대사율; 치료 기간; 및 의학 분야에 잘 알려진 유사 요인을 포함한 다양한 요인에 따라 달라질 수 있다.

[0172] *tracrRNA*: 본원에서 사용될 때 용어 "tracrRNA" 또는 "트랜스-활성화 crRNA"는 CRISPR-연관 단백질이 명시된 표적 핵산에 결합하는 데 필요한 구조를 형성하는 서열을 포함하는 RNA를 의미한다.

[0173] *치료*: 본원에서 사용될 때 용어 "치료"(또한 "치료하다" 또는 "치료함")는 특정 질병, 장애 및/또는 병태의 하나 이상의 증상 또는 특징의 발병을 부분적으로 또는 완전히 완화, 개선, 경감, 억제, 지연시키고/거나 중증도를 낮추고/거나 발생률을 낮추는 치료 분자(예를 들어, 본원에 기술된 CRISPR-Cas 치료 단백질 또는 시스템)의 임의의 투여를 의미한다. 이러한 치료는 관련 질병, 장애 및/또는 병태의 징후를 나타내지 않는 대상체 및/또는 질병, 장애 및/또는 병태의 초기 징후만을 나타내는 대상체에 대한 것일 수 있다. 대안적으로 또는 추가적으로, 이러한 치료는 관련 질병, 장애 및/또는 병태의 하나 이상의 규명된 징후를 나타내는 대상체에 대한 것일 수 있다.

도면의 간단한 설명

[0174] 도면은 설명 목적으로만 제공되고 제한을 위한 것이 아니다.

도 1a는 인간 코돈-최적화된 *스트렙토코커스 콘스텔라투스* Cas 9에 의해 인식되는 공통 PAM 모티프를 보여주는 그래프이다. 도 1b는 인간 코돈-최적화된 *샤르페아 spp. 단리물 RUG017* Cas9에 의해 인식되는 공통 PAM 모티프를 보여주는 그래프이다. 도 1c는 인간 코돈-최적화된 *베일로넬라 파르불라* Cas9에 의해 인식되는 공통 PAM 모티프를 보여주는 그래프이다. 도 1d는 인간 코돈-최적화된 *에자키엘라 페루엔시스*에 의해 인식되는 공통 PAM 모티프를 보여주는 그래프이다. 도 1e는 인간 코돈-최적화된 *락토바실러스 퍼멘텀 균주 AF15-40LB*에 의해 인식되는 공통 PAM 모티프를 보여주는 그래프이다. 도 1f는 인간 코돈-최적화된 *펩토니필러스 sp. 마르세유-P3761*에 의해 인식되는 공통 PAM 모티프를 보여주는 그래프이다.

도 2a는 Geneious 소프트웨어를 사용하여 인간 코돈-최적화된 *스트렙토코커스 콘스텔라투스* ScoCas9에 대한 sgRNA의 예측된 RNA 접힘 구조를 보여주는 도면이다. 도 2a는 서열번호 3을 포함하는 sgRNA을 나타낸다. 도 2b는 Geneious 소프트웨어를 사용하여 인간 코돈-최적화된 *샤르페아 spp. 단리물 RUG017* SirCas9에 대한 sgRNA의 예측된 RNA 접힘 구조를 보여주는 도면이다. 도 2b는 서열번호 7을 포함하는 sgRNA을 나타낸다. 도 2c는 Geneious 소프트웨어를 사용하여 인간 코돈-최적화된 *베일로넬라 파르불라* VapCas9에 대한 sgRNA의 예측된 RNA 접힘 구조를 보여주는 도면이다. 도 2c는 서열번호 13을 포함하는 sgRNA을 나타낸다. 도 2d는 Geneious 소프트웨어를 사용하여 인간 코돈-최적화된 *에자키엘라 페루엔시스* EpeCas9에 대한 sgRNA의 예측된 RNA 접힘 구조를 보여주는 도면이다. 도 2d는 서열번호 19를 포함하는 sgRNA을 나타낸다. 도 2e는 Geneious 소프트웨어를 사용하여 인간 코돈-최적화된 *락토바실러스 퍼멘텀 균주 AF15-40LB* LfeCas9에 대한 sgRNA의 예측된 RNA 접힘 구조를 보여주는 도면이다. 도 2d는 서열번호 95을 포함하는 sgRNA을 나타낸다. 도 2f는 Geneious 소프트웨어를 사용하여 인간 코돈-최적화된 *펩토니필러스 sp. 마르세유-P3761* PmaCas9에 대한 sgRNA의 예측된 RNA 접힘 구조를 보여주는 도면이다. 도 2d는 서열번호 96을 포함하는 sgRNA을 나타낸다.

도 3은 HEK293T 세포에서 인간 코돈-최적화된 ScoCas9의 생체외 절단 활성의 예시적인 결과를 보여주는 그래프이다. 그래프의 y축은 PAM 공통 모티프에 상응하는 서열에 인접한 A-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA를 사용하여 얻은 삽입결실 빈도를 보여준다(도 1a 참조).

도 4a는 아데닌 염기 편집기(ABE) 또는 시토신 염기 편집기(CBE)에 N-말단에서 융합된 ScoCas9 D10A 돌연변이체의 작제물을 보여주는 도면이다. 도 4b는 ScoCas9 D10A 돌연변이체의 N-말단에 융합된 ABE를 포함하는 염기 편집기를 사용하여 달성된 삽입결실 빈도 및 아데닌에서 구아닌 염기로(A에서 G로) 전환율의 결과를 보여주는 그래프이다. A에서 G로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 A-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 8)에 대해 도시되어 있다(도 1a 참조). 도 4c는 ScoCas9 D10A 돌연변이체의 N-말단에 융합된 ABE를 포함하는 염기 편집기를 사용하여 달성된 삽입결실 빈도 및 시토신에서 티민 염기로(C에서 T로) 전환율의 결과를 보여주는 그래프이다. C에서 T로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 C-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 8)에 대해 도시되어 있다(도 1a 참조).

도 5a는 WT SirCas9 및 아데닌 염기 편집기(ABE)에 N-말단에서 융합된 SirCas9 D14A 돌연변이체의 작제물을 보여주는 도면이다. 도 5b는 SirCas9 D14A 돌연변이체의 N-말단에 융합된 ABE를 포함하는 염기 편집기를 사용하여 달성된 삽입결실 빈도 및 A에서 G로 전환의 결과를 보여주는 그래프이다. A에서 G로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 A-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 9)에 대해 도시되어 있다(도 1b 참조).

도 6a는 WT VapCas9 및 아데닌 염기 편집기(ABE) 또는 시토신 염기 편집기(CBE)에 N-말단에서 융합된 VapCas9 D38A 돌연변이체를 보여주는 작제물의 도면이다. 도 6b는 염기결실 빈도, VapCas9 D38A 돌연변이체의 N-말단에 융합된 ABE를 포함하는 염기 편집기를 사용하여 달성된 A에서 G로 전환 및 VapCas9 D38A의 N-말단에 융합된 CBE를 포함하는 염기 편집기를 사용하여 달성된 C에서 T로 전환을 보여주는 그래프이다. A에서 G로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 A-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 10)에 대해 도시되어 있다(도 1c 참조). C에서 T로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 C-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 10)에 대해 도시되어 있다(도 1c 참조).

도 7a는 VapCas9의 N-말단 또는 VapCas9의 C 말단에 융합된 ABE를 보여주는 작제물의 도면이다. 도 7b는 VapCas9의 N-말단에 융합된 ABE 및 C-말단에 융합된 ABE를 포함하는 염기 편집기를 사용하여 달성된 A에서 G로 전환의 비교를 보여주는 그래프이다. A에서 G로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 A-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 11)에 대해 도시되어 있다(도 1c 참조).

도 8a는 WT EpeCas9 및 ABE와 CBE에 N-말단에서 융합된 EpeCas9 D38A 돌연변이체를 보여주는 작제물의 도면이다. 도 8b는 삽입결실 빈도, EpeCas9 D38A 돌연변이체의 N-말단에 융합된 ABE를 포함하는 염기 편집기를 사용하여 달성된 A에서 G로 전환 및 EpeCas9 D38A의 N-말단에 융합된 CBE를 포함하는 염기 편집기를 사용하여 달성된 C에서 T로 전환의 결과를 보여주는 그래프이다. A에서 G로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 A-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 12)에 대해 도시되어 있다(도 1d 참조). C에서 T로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 C-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 12)에 대해 도시되어 있다(도 1d 참조).

도 9a는 WT LfeCas9 및 ABE와 CBE에 N-말단에서 융합된 LfeCas9 D9A 돌연변이체를 보여주는 도면이다. 도 9b는 LfeCas9를 사용한 삽입결실 빈도의 결과를 보여주는 그래프이다. 도 9c는 LfeCas9 D9A 돌연변이체의 N-말단에 융합된 ABE를 포함하는 염기 편집기를 사용하여 달성된 A에서 G로 전환의 결과를 보여주는 그래프이다. A에서 G로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 A-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 13)에 대해 도시되어 있다(도 1e 참조). 도 9d는 LfeCas9 D9A 돌연변이체의 N-말단에 융합된 CBE를 포함하는 염기 편집기를 사용하여 달성된 C에서 T로 전환의 결과를 보여주는 그래프이다. C에서 T로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 C-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 13)에 대해 도시되어 있다(도 1e 참조).

도 10a는 WT PmaCas9 및 N-말단과 C-말단에서 ABE와 CBE에 융합된 PmaCas9 D12A 돌연변이체를 보여주는 도면이다. 도 10b는 PmaCas9 D12A 돌연변이체의 N-말단 또는 C-말단에 융합된 ABE 또는 CBE를 포함하는 염기 편집기를 사용하여 달성된 A에서 G로 또는 C에서 T로 전환의 결과를 보여주는 그래프이다. A에서 G로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 A-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 14)에 대해 도시되어 있다(도 1f 참조). C에서 T로 전환율(y축)은 PAM 공통 모티프에 상응하는 서열에 인접한 C-풍부 게놈 테스트 부위를 표적으로 하는 다양한 가이드 RNA(x축; 표 14)에 대해 도시되어 있다(도 1f 참조).

도 11a는 2개의 ScoCas9-NGC 변이체인 ScoCas9-NGC-v1 및 ScoCas9-NGC-v2(x축)로 세포를 형질감염시킴으로써

측정한 삽입결실 빈도(y축; 삽입결실 빈도%)의 예시적인 결과를 보여주는 그래프이다. 미형질감염 세포 대조군도 표시된다.

도 11b는 NGC PAM 모티프를 인식하도록 조작된 ScoCas9-NGC 변이체인 ScoCas9-NGC-v1 및 ScoCas9-NGC-v2를 포함하는 A에서 G로 염기 편집기(ABE)로 형질감염된 HEK293T 세포(x축)에서 예시적인 A에서 G로 전환(y축; A에서 G로 전환%)을 보여주는 그래프이다. NGC를 인식하지 않는 ScoCas9-NGG 변이체는 A에서 G로 전환을 보여주지 않았다. SpyCas9-NGC 대조 벡터는 A에서 G로 편집을 보여주었다. 미형질감염 세포 대조군도 표시된다.

발명을 실시하기 위한 구체적인 내용

[0175] 일정한 간격으로 놓인 짧은 회문 반복부 서열(CRISPR)은 박테리아와 고세균에서 적응형 면역 시스템으로 처음 발견된 후 살아있는 세포와 유기체에서 표적이 된 DNA 파손을 발생시키도록 조작되었다. 세포 DNA 복구 프로세스동안 다양한 DNA 변화가 도입될 수 있다. 다양한 확장성 CRISPR 도구상자를 사용하여 프로그래밍 가능한 게놈 편집, 후성유전체 편집 및 전사체 조절을 할 수 있다.

[0176] CRISPR-Cas 시스템은 Cas 유전자 조직과 구성요소 단백질의 서열 및 구조에 기반한 3가지 주요 유형(I, II 및 III)을 포함한다. 3가지 CRISPR 시스템 각각은 다음의 고유 Cas 유전자를 특징으로 한다: 유형 I에서 표적-열화성 뉴클레아제/헬리카제인 Cas3; 유형 II에서 RNA-결합 및 표적-열화성 뉴클레아제인 Cas9; 유형 III에서 다중 기능을 위한 대형 단백질인 Cas 10. 또한 3가지 CRISPR 유형은 연관된 이펙터 복합체에서 상이하다. 유형 I Cas 시스템은 캐스케이드 이펙터 복합체와 연관되고, 유형 II 이펙터 복합체는 단일 Cas9와 1개 이상의 RNA 분자로 이루어지고, 유형 III 간섭 복합체는 유형 III-A(DNA를 표적으로 하는 Csm 복합체)와 유형 III-B(RNA를 표적으로 하는 Cmr 복합체)로 더 나누어진다. Cas 단백질은 모든 CRISPR-Cas 시스템에서 이펙터 복합체의 중요 구성요소이다.

[0177] 현재 게놈 편집 기술은 DNA 절단을 위한 단일-단백질 이펙터 뉴클레아제를 함유하는 클래스 II CRISPR-Cas 시스템에 초점을 두고 있으며, 구체적으로는 CRISPR RNA(crRNA)와 tracrRNA 둘 모두를 필요로 하고 HNH 및 RuvC 뉴클레아제 도메인 둘 모두를 함유하는 이중-RNA 가이드되는 뉴클레아제인 Cas9와, crRNA만을 필요로 하고 단일 RuvC 도메인을 함유하는 단일-RNA 가이드되는 뉴클레아제인 Cas 12a에 초점을 두고 있다.

[0178] 본 발명의 다양한 측면은 다음 항에서 자세히 기술된다. 이러한 항의 사용은 본 발명을 제한하려는 것이 아니다. 각 항은 본 발명의 임의의 측면에 적용될 수 있다. 본 출원에서 "또는"의 사용은 달리 명시되지 않는 한 "및/또는"을 의미한다.

[0179] **조작된 비천연 발생 Cas9 단백질**

[0180] *스트렙토코커스 콘스텔라투스*(ScoCas9), *샤르페아 spp. 단리물 RUG017*(SirCas9), *베일로넬라 파르불라*(상호 교환적으로 사용되는 VapCas9 또는 VpaCas9), *에자키엘라 페루엔시스*(EpeCas9), *락토바실러스 퍼멘텀*(LfeCas9) 및 *펩토니필러스 sp. 마르세유-P3761*(PmaCas9) 박테리아에서 얻은 WT Cas9로부터 변형된 조작된 비천연 발생 Cas9 단백질이 본원에 기술되어 있다.

[0181] 일부 구현예에서, 본원에 기술된 조작된 비천연 발생 Cas9 단백질은 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 60%(예를 들어, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 이상) 동일한 아미노산 서열을 포함한다. 일부 구현예에서, Cas9 단백질은 서열번호 1, 4, 8, 14, 84 또는 86과 80% 동일하다. 일부 구현예에서, Cas9 단백질의 아미노산 서열은 서열번호 1, 4, 8, 14, 84 또는 86과 동일하다. 예시적인 Cas9 아미노산 서열은 아래 표 1에 제공된다.

[0182] 표 1. 예시적인 Cas9 아미노산 서열

야생형 *스트렙토코커스 콘스텔라투스* Cas9
MGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGETAEA TRLKRTARRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIFGNI AAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQA LFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIAL SLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGI LTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESG VKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQG EHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSA EKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFD HVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDD KANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGI RNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSP AIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDN PVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNR VLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGF IKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVRE INDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRF FKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKL IPRKTKNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKN PVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQRFTTLLYH SYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSS LAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSITGLYETRIDLSKL GED (서열번호 1).
핵 위치 신호 (NLS) 및 링커를 갖는 *스트렙토코커스 콘스텔라투스* Cas9
MPKKKRKVGGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLF DSGETAEATRLKRTARRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGE HHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLK AENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNT

[0183]

LFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLY
DAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKG
YASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEM
HAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDD
ILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDT
NMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKI
LDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRL
SYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDDIANV
VHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQ
DSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFI
KDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGL
TEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKE
FELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFL
FYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESIL
PKGESDKLIPRKTKNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKLRKVQDMVGITIM
EKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQ
RFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSK
NKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSITGLYE
TRIDLSKLGEDG**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDYAYPYDVPDYA*

(서열번호 2).

야생형 *사르페아* Cas9

MAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIR
RRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGNDYKDNYNLFIEK
DFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFAKVYEDI
EEKLDNTLKKFMSLNDLDNLFVDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYK
EFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYIEFIEMLHS
INSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNY
LGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEKFMLKQNSRTNGSIPYQMQK
DEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILP
WNYDQIVDIDATAEGFIERMQNTGTYFPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPV
ETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTK
IFGEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEK

[0184]

[0185]

LLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIEESNKKDIEGPFRYDEVKKLA GSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQ TKEDQLVYESLKKEDAKKKIDTDALYLYYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLI KDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFSLIRTDFNE KDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYKNRDLNDLH HAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYKKDVQEHNNYGFILNSMNHIHFNDDN SVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKAKIPVNKYR KDVNKYGGFSGDAPIMLAIEADKGKKHVRQVIAFPLRLKNYNDEERIKFIEKEKNLKNVKI LTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKLESIDDANF NEVIQELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMSAKATQGYI KPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKL (서열번호 4).
핵 위치 신호 (NLS) 및 링커를 갖는 *사르페아* Cas9
MPKKKRKVGAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAAT RRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGNDYKD NYNLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQS FAKVYEDIEEKLDNTLKKFMSLNDLDNLFVDNDINSMITVLSKIYQRSKKADDLLKIMNPT KEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYI EFIEMLHSINSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFRE DNPKLHNYLGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEKFMLKQNSRTNG SIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFED KQKERILPWNYDQIVDIDATAEGFIERMQNTGTYFPDKPVMAKNSLTVSKFEVLNELNKIR INGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSL APWIDFTKIFGEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYT GWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIEESNKKDIEGPFR YDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQK IYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYYLQMGKSMYSGKPLDIDKLSTYHID HILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFS LIRTDFNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYK NRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYKKDVQEHNNYGFILNSMN HIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKA KIPVNKYRKDVNKYGGFSGDAPIMLAIEADKGKKHVRQVIAFPLRLKNYNDEERIKFIEKE KNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKL

ESIDDANFNEVIQELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMS
AKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKLG**KRPAATK**
KAGQAKKKKGS*YPYDVPDYAYPYDVPDYAYPYDVPDYA* (서열번호 5).

야생형 *베일로넬라 파르불라* Cas9

MSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRSH
KMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELFADAMAQVDPTFFMRLRESKY
HYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKY
RGNFLYEGATFDSNASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAI
EHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKK
LQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKD
DLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLS
DSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSV
AEKLIKMLEFRIPYYVGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIK
NLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKM
TKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDIT
IFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPET
IIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQA
LRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEI
FGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKD
DSFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTAD
DLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDFRHDNNFIKVRSLNHHHH
AKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTS
MDTVKKMMDSNDVRVTKRLLEQTGALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKY
GGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIPQAKDISII
YGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDL
KANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFK
KMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGLYS
NEVTIV (서열번호 8).

핵 위치 신호 (NLS) 및 링커를 갖는 *베일로넬라 파르불라* Cas9

MPKKKRKVGSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAY
ELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELFADAMAQVDPTFF

[0186]

[0187]

MRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFL
AVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKT
KSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVE
ELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKV
KAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYT
KKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLN
EVADGFSVAEKLIKMLEFRIPYYVGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDRE
KSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAV
FKQDHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMA
EEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDK
AGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPA
VKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVL
LNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHI
YPRSLTKDDSFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLT
RITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDFRHDNNFIKV
RSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDG
KAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTGALADATIYKATVAGKAKDGAYIGMKTKSS
VFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIP
QAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNN
MRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELS
EKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVIN
ESITGLYSNEVTIVG**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDYAYPYDVPDYA*

(서열번호 9).

야생형 *에자키엘라 페루엔시스* Cas9

MTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRGARRR
LDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKDQKLKSKYTLFNDKDFKDKNF
HKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELK
VHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLV
DLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEA
DKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSI
SNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKI
KSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNP

[0188]

NSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYM
VLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGD
FKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSE
IKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREYNLNLMELMSASFTFTEEI
QKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEV
KARKESRKNQLLKFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRC
MYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNK
NCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICK
NSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKDPMNFVKKQ
EKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESG
ALFNATLQRKNKGSRPLKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVERE
IYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLEL
RSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFKTSVFKKSF
LNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRLSSKTN
NIGNYQEFKIINQSITGLFENEVDLLKL (서열번호 14).

핵 위치 신호 (NLS) 및 링커를 갖는 *에자키엘라 페루엔시스* Cas9

MPKKKRKVGTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERR
GQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKDQKLKSKYTLFND
KDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSF
ENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIM
IGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSI
LENLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNY
VSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDME
FKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIP
YYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKA
SLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEI
LLSGTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGY
GKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREYNLNLMELMSA
SFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIE
MARGKEEVKARKESRKNQLLKFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYL
YYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYP
IPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETA

[0189]

NLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKD
PMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFT
RMNYIESGALFNATLQRKNKGSRPLKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIER
KLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITG
GSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFK
TSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLKSIDLKSDY
GRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLG**KRPAATKKAGQAKKKK**GS*YPYDVP*
DYAYPYDVPDYAYPYDVPDYA(서열번호 15).

야생형 *락토바실러스 파렌텀* 균주 *AF15-40LB* Cas9

MKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKR
RKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLLKKNERGYP
TLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASV
DKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAK
LLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFR
EELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPAS
ARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGV
IPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTP
EVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANS
LLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPS
VEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVE
WLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQI
DQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEF
ARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMY
TGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAKMKPYW
NQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTD
IIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMK
FKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNE
TYEETGKYFDASHESSSLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDK
KGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGE
EKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYY

[0190]

KLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFGSDLQNLKEIGFGTTPL

GQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL (서열번호 84).

핵 위치 신호 (NLS) 및 링커를 갖는 *락토바실러스 파렌팀 균주AF15-40LB* Cas9

MPKKKRKVGKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFR
TTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLL
KKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRG
HFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKL
DRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETS
EDEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKE
YLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPK
QRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPY
YVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLN
EDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAK
TRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIF
ADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMQIVN
QPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNA
PKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYF
TQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLY
AAKMKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGS
MYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSF
FVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINL
KVMLISNETYEETGKYFDASHESSSLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYM
VILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLK
GTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNI
LRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFGSDLQNLKE
IGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL

G**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDYAYPYDVPDYA* (서열번호 85).

야생형 *캠토니필러스 sp. 마르세유-P3761* Cas9

MEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFESGQTA
KDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFSKNPIFG
TLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAI

NENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQS
PMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVY
NAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQ
KKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQ
VHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIK
PWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQK
RMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELA
KVPGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGW
GRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQM
IDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKA
LSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQY
DEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRK
FDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVL
LKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVYGEFRYKKK
PQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEER
TGMLFKETINGKTNKGSRIPIKKDLDPSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGIS
IMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTKGKVELQKA
NQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFLDAETNINR
LRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSMTGLFDGQV
IYQSLTGLYETRMRFED (서열번호 86).

핵 위치 신호 (NLS) 및 링커를 갖는 *캠토니필러스 sp. 마르세유-P3761* Cas9

MPKKKRKVGEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSL
LFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKA
FSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGK
LDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKD
IGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADL
FDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMT
FVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSF
YNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLE
RKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKI
AYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNAD
YSTYVELAKVPGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKEL

[0191]

AKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFK
KTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMAREN
QTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPL
DFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKD
AGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNK
IRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVY
GEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVN
VVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSKYGGYIEEKMAYYAVISYEDKKKKP
GKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTK
GKVELQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFL
DAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSM
TGLFDGQVIYQSLTGLYETRMRFEDG**KRPAATKKAGQAKKKK***GSYPYDVPDYAYPYDVPDY*
AYPYDVPDYA (서열번호 87).

[0192]

[0193] NLS(굵은 글씨)는 다른 NLS로 대체될 수 있다.

[0194] 링커(밑줄)는 제거되거나 연장될 수 있다.

[0195] 3xHA 태그(이탤릭체)는 다른 태그로 대체될 수 있다.

[0196] 일부 구현예에서, Cas9 단백질은 서열번호 1, 4, 8, 14, 84 또는 86에 대하여 1개 이상의 돌연변이를 포함한다. 예를 들어, Cas9 단백질의 아미노산 서열은 서열번호 1, 4, 8, 14, 84 또는 86에 적어도 1개, 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 적어도 6개, 적어도 7개, 적어도 8개, 적어도 9개, 적어도 10개의 돌연변이를 포함한다. 다양한 돌연변이가 당업계에 알려져 있으며, 예를 들어 아미노산 치환을 포함한다.

[0197] 일부 구현예에서, Cas9의 2개 이상의 촉매 도메인(RuvC1, RuvCII, RuvCIII)이 돌연변이되어, 불활성 또는 핵산 절단 활성이 없는 "죽은" Cas9(dCas9)를 생성한다. 일부 구현예에서, 1개 이상의 돌연변이는 PAM 상호작용, HNH 및 또는 RuvC 도메인에 있다. 일부 구현예에서, Cas9는 비돌연변이 형태에 비해 DNA 절단 활성을 약 25%, 15%,

10%, 5%, 1%, 0.1%, 0.01% 이하로 낮추도록 돌연변이된다.

[0198] 일부 구현예에서 Cas9의 닉카제-돌연변이체 버전이 제공된다. 일부 구현예에서, 닉카제 돌연변이체는 RuvC 및/또는 HNH 도메인에 1개 이상의 아미노산 치환을 가진다. 다양한 닉카제 돌연변이가 SpCas9(*스트렙토코커스 피오게네스*)와 관련하여 알려져 있으며, 예를 들어 야생형 SpCas9의 아미노산 위치 10, 12, 17, 762, 840, 854, 863, 982, 983, 984, 986, 987 중 하나 이상에 돌연변이를 포함한다. 예를 들어, *SpCas9*의 D10A에 상응하는 아스파르트산에서 알라닌으로 치환은 닉카제 형성을 초래한다. 일부 구현예에서, 본원에 기술된 Cas9는 닉카제 형성을 초래하는 1개 이상의 돌연변이를 가진다. 일부 구현예에서, 본원에 기술된 Cas9는 SpCas9의 아미노산 10, 12, 17, 762, 840, 854, 863, 982, 983, 984, 986, 987 중 하나 이상에 상응하는 아미노산 위치에 1개 이상의 돌연변이를 가진다.

[0199] 일부 구현예에서, 돌연변이는 ScoCas9의 RuvC 도메인에서 아스파르트산에서 알라닌으로 치환(D10A)이다. 일부 구현예에서, 돌연변이는 SirCas9의 RuvC 도메인에서 아스파르트산에서 알라닌으로 치환(D14A)이다. 일부 구현예에서, 돌연변이는 VapCas9의 RuvC 도메인에서 아스파르트산에서 알라닌으로 치환(D38A)(예를 들어 SpCas9의 D10A에 상응)이다. 일부 구현예에서, 돌연변이는 EpeCas9의 RuvC 도메인에서 아스파르트산에서 알라닌으로 치환(D12A)이다. 일부 구현예에서, 돌연변이는 LfeCas9의 RuvC 도메인에서 아스파르트산에서 알라닌으로 치환(D9A)이다. 일부 구현예에서, 돌연변이는 PmaCas9의 RuvC 도메인에서 아스파르트산에서 알라닌으로 치환(D12A)이다.

[0200] 일부 구현예에서, 돌연변이는 ScoCas9의 RuvC 도메인에서 아스파르트산에서 글리신으로 치환(D10G)이다. 일부 구현예에서, 돌연변이는 SirCas9의 RuvC 도메인에서 아스파르트산에서 글리신으로 치환(D14G)이다. 일부 구현예에서, 돌연변이는 VapCas9의 RuvC 도메인에서 아스파르트산에서 글리신으로 치환(D38G)이다(예를 들어 SpCas9의 D10G에 상응)이다. 일부 구현예에서, 돌연변이는 EpeCas9의 RuvC 도메인에서 아스파르트산에서 글리신으로 치환(D12G)이다. 일부 구현예에서, 돌연변이는 LfeCas9의 RuvC 도메인에서 아스파르트산에서 글리신으로 치환(D9A)이다. 일부 구현예에서, 돌연변이는 PmaCas9의 RuvC 도메인에서 아스파르트산에서 글리신으로 치환(D12G)이다.

[0201] 일부 구현예에서, 본원에 기술된 이러한 1개 이상의 돌연변이는 Cas9를 Cas9의 불활성 또는 "죽은" 버전(dCas9)으로 전환시킨다. 따라서, 일부 구현예에서, Cas9 단백질은 Cas9가 DNA 이중구조의 두 가닥을 모두 절단하는 능력을 억제하는 1개 이상의 돌연변이를 포함한다.

[0202] 일부 구현예에서, 가이드 RNA와 함께 공동발현될 때, 죽은 Cas9는 전사적 신장, RNA 폴리머라제 결합 또는 전사인자 결합을 특이적으로 방해할 수 있는 DNA 인식 복합체를 발생시킨다. 일부 구현예에서, 죽은 Cas9는 특이적 핵산 표적 부위에 다양한 기능의 이펙터 단백질을 특이적으로 표적으로 하기 위해 사용된다.

[0203] 일부 구현예에서, 고충실 Cas9 변이체는 증강된 특이성을 포함하여 표적외(off-target) 절단을 최소화한다. 일부 구현예에서, 조작된 변이체, 예를 들어 주로 REC3 도메인 내에 돌연변이를 포함하고 보다 높은 특이성과 충실성을 달성하는 '초정확 Cas9'(SpyCas9에 상응하는 N692A, M694A, Q695A 및/또는 H698A 돌연변이) 및/또는 '고충실 Cas9'(SpyCas9에 상응하는 N467A, R661A, Q695A 및/또는 Q926A 돌연변이)가 사용된다. 고충실 변이체는 Cas9가 미스매치를 안정화시키고 표적외 DNA 절단을 낮추는 능력을 낮춘다. 일부 구현예에서, 특이성 증가는 표적내(on-target) 절단 효율을 약 100배 감소시킴을 동반한다. 일부 구현예에서, 야생형 Cas9에 필적하는 표적내 절단율을 유지하는 고충실 변이체인 SuperFi-Cas9가 사용된다. 일부 구현예에서, SuperFi-Cas9는 RuvC 루프에 돌연변이를 포함한다. 일부 구현예에서, 돌연변이는 gRNA-TS 이중구조의 후속 절단을 촉진하는 꼬인 형태의 형성을 억제한다. 일부 구현예에서, SpyCas9에 상응하는 Y1016, R1019, Y1010, Y1013, K1031, Q1027 및/또는 V1018 잔기는 예를 들어, 아스파르트산으로 돌연변이된다.(Bravo, J. 등 Structural basis for mismatch surveillance by CRISPR-Cas9 *Nature*, 603, March 2022).

[0204] 조작된 비천연 발생 Cas9는 서열번호 2, 5, 9, 15, 85, 87, 95 또는 96에서 Cas9 아미노 서열과 적어도 80%(예를 들어, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 이상) 동일한 아미노산 서열을 가진다.

[0205] 일부 구현예에서, 조작된 비천연 발생 Cas9는 인간 세포에 대해 코돈-최적화된(예를 들어, 발현, 안정성 등에 대해 코돈 최적화된) 핵산 분자에서 암호화된다.

[0206] 핵 위치 신호(NLS) 및 링커가 있는 예시적인 Cas9 서열은 다음 표 2에 제공된다.

[0207] 표 2. NLS 및 링커가 있는 예시적인 Cas9 서열

핵 위치 신호 (NLS) 및 링커를 갖는 *Sco*Cas9 서열
MPKKKRKVGGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLF DSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGE HHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLK AENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNT LFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLY DAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKG YASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEM HAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDD ILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDT NMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKI LDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRL SYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDDIANV VHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQ DSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFI KDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGL

[0208]

TEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKE FELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFL FYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESIL PKGESDKLIPRKTKNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKLRKVQDMVGITIM EKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQKGNELVIPQ RFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSK NKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSITGLYE TRIDLSKLGEDG**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDYAYPYDVPDYA* (서열번호 2).
핵 위치 신호 (NLS) 및 링커를 갖는 *샤르페아*Cas9 서열
MPKKKRKVGAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAAT RRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGNDYKD NYNLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQS FAKVYEDIEEKLDNTLKKFMSLNDLDNLFVDNDINSMITVLSKIYQRSKKADDLLKIMNPT KEERAAYKEFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYI EFIEMLHSINSWVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFRE DNPKLHNYLGYIKYPKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEKFMLKQNSRTNG SIPYQMQKDEMIQIIDNQSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFED KQKERILPWNYDQIVDIDATAEGFIERMQNTGTYFPDKPVMAKNSLTVSKFEVLNELNKIR INGKLIPVETKKELLSDLFMKNKTITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSL APWIDFTKIFGEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYT GWSRLSEKLLTGIKSKNSKETILSILENSNMNLMEIINDESLGFKQIIEESNKKDIEGPFR YDEVKKLAGSPAIKRGIWQALLVVQEITKFMKHEPSHIYIEFAREEQEKVRTESRIAKLQK IYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYLYYLQMGKSMYSGKPLDIDKLSTYHID HILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIRNKMMGFWQKLHENGLMSNKKFFS LIRTDFNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTNVVTVRAELSHQFRERYKIYK NRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYKKDVQEHNNYGFILNSMN HIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTILPSDKNSENGVTKA KIPVNKYRKDVNKYGGFSGDAPIMLAIEADKGKKHVRQVIAFPLRLKNYNDEERIKFIEKE KNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLVDANKHNKL ESIDDANFNEVIQELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYLIAIMS

[0209]

AKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKLG**KRPAATK**

KAGQAKKKKGS*YPYDVPDYAYPYDVPDYAYPYDVPDYA* (서열번호 5).

핵 위치 신호 (NLS) 및 링커를 갖는 *베일로멀라 파르불라* Cas9 서열

MPKKKRKVGSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAY
ELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELFADAMAQVDPTFF
MRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFL
AVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKT
KSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVE
ELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKV
KAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYT
KKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLN
EVADGFSVAEKLIKMLEFRIPYYVGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDRE
KSAAAFIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAV
FKQDHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMA
EEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDK
AGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPA
VKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVL
LNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHI
YPRSLTKDDSFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLT
RITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDFRHDNNFIKV
RSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDG
KAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTGALADATIYKATVAGKAKDGAYIGMKTKSS
VFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVASIIP
QAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNN
MRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELS
EKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVIN
ESITGLYSNEVTIVG**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDYAYPYDVPDYA*
(서열번호 9).

핵 위치 신호 (NLS) 및 링커를 갖는 *여자키엘라 페루엔시스* 서열

[0210]

[0211]

MPKKKRKVGTKVKDYYIGLDIGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERR
GQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKDQKLKSKYTLFND
KDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSF
ENSLNELKVHLNDEYGLDLEFDNENLININLTDPKLNKTAKKKELKSVIGDTKFLKAVSAIM
IGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSI
LENLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNY
VSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDME
FKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIP
YYVGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKA
SLLYNEYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEI
LLSGTDGDFKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGY
GKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREYNLNLMELMSA
SFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIE
MARGKEEVKARKESRKNQLLKFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYL
YYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYP
IPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETA
NLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKD
PMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKELEGTNYRFT
RMNYIESGALFNATLQRKNKGSRPLKDKGPKSSIEKYGGYTNINKACFAVLDIKSKNKIER
KLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYFITG
GSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFK
TSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLKSIDLKSDY
GRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLG**KRPAATKKAGQAKKKK**GS*YPYDVP*

DYAYPYDVPDYAYPYDVPDYA (서열번호 15).

핵 위치 신호 (NLS) 및 링커를 갖는 *락토바실러스 퍼멘텀* 균주 *AF15-40LB* Cas9 서열

MPKKKRKVGKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFR
TTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLL
KKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRG
HFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKL
DRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETS
EDEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKE

YLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPK
QRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPY
YVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLN
EDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAK
TRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIF
ADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMQIVN
QPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNA
PKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYF
TQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLY
AAKMKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGS
MYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSF
FVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINL
KVMLISNETYEETGKYFDASHESSSLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYM
VILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLK
GTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNI
LRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFGSDLQNLKE
IGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL

G**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDYAYPYDVPDYA* (서열번호 85).

핵 위치 신호 (NLS) 및 링커를 갖는 *캠토니필러스 sp. 마르세유-P3761* Cas9 서열

MPKKKRKVGEKKTNYTIGLDIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSL
LFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKA
FSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGK
LDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKD
IGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADL
FDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMT
FVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSF
YNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLE
RKSDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKI
AYLDERQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNAD
YSTYVELAKVPGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKEL
AKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFK
KTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMAREN

[0212]

[0213]

QTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPL
DFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKD
AGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNK
IRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLLALYPYMADDFVY
GEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIKKVMGYRQVN
VVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSKYGGYIEEKMAYYAVISYEDKKKKP
GKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATKTTK
GKVELQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFL
DAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSM
TGLFDGQVIYQSLTGLYETRMRFEDG**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDY*
AYPYDVPDYA (서열번호 87).

핵 위치 신호 (NLS) 및 링커를 갖는 *Sco*Cas9 변이체 서열 (ScoCas9-NGC-v1)

MPKKKRKVGMGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALL
FDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRG
EHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDL
KAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKN
TLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNL
YDAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKK
GYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQE
MHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFD
DILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFD
TNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRK
ILDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGR
LSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDDIAN
VVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLL
QDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAF
IKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGG
LTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRK
EFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKF
LFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESI
LPKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVFADVEKGKSKKLRKVQDMVGITI
MEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVIP

QRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFS
KNKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITGLY
ETRIDLSKLGEDG**EGADKRTADGSEFESPKKKRKV** (서열번호 95)

핵 위치 신호 (NLS) 및 링커를 갖는 *Sco*Cas9 서열 (ScoCas9-NGC-v2)

MPKKKRKVGMGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALL
FDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRG
EHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDL
KAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKN
TLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNL
YDAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKK
GYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGIIPHQIHLQE
MHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFD
DILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFD
TNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRK
ILDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERLHYTGWGR
LSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDDIAN
VVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLL
QDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAF
IKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGG
LTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRK
EFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKF
LFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTGGFSKESI
LPKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVFADVEKGKSKKLRKVQDMVGITI
MEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVIP
QRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFS
KNKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITGLY
ETRIDLSKLGEDG**EGADKRTADGSEFESPKKKRKV** (서열번호 96)

[0214]

[0215] NLS(굵은 글씨)는 상이한 NLS로 대체될 수 있다.

[0216] 링커(밑줄)는 제거되거나 연장될 수 있다.

[0217] 일부 구현예에서, 조작된 비천연 발생 Cas9는 태그를 포함한다. 당업자에게 명백한 바와 같이, 목적에 따라 다양한 태그가 Cas9 변이체(예를 들어, 3xHA 태그)에 융합될 수 있다.

[0218] 다양한 종은 특정 유기체에서 해당 코돈에 대한 tRNA 종의 풍부함에 상응하는 메신저 RNA(mRNA) 내 코돈을 활용함으로써 mRNA의 번역 효율과 상관관계가 있는 코돈 편향(즉, 유기체에 의한 코돈 사용의 차이)을 나타낸다. 예를 들어 소프트웨어 사용에 의하는 것을 포함하여 당업계의 다양한 방법이 컴퓨터 최적화를 위해 사용될 수 있다. 일부 구현예에서, 코돈 최적화는 천연 아미노산 서열을 유지하면서 천연 서열의 적어도 1개의 코돈(예를 들어 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 이상의 코돈)을, 숙주 세포의 유전자에서 더 빈번하게 사용되거나 가장 빈번하게 사용되는 코돈으로 대체함으로써 관심 숙주 세포에서 증강되게 발현되도록 하기 위한 핵산 서열 변형을 의미한다.

[0219] 일부 구현예에서, 본원에 기술된 Cas9 단백질은 코돈 최적화된다. 이러한 유형의 최적화는 당업계에 알려져 있고, 동일 단백질을 암호화하면서 의도된 숙주 유기체 또는 세포의 코돈 선호를 모방하도록 외래-유래의 DNA 돌연변이를 수반한다. 따라서, 코돈은 변하지만, 암호화된 단백질은 변하지 않은 상태로 유지된다. 코돈 최적화는 가용성 단백질 수준을 향상시키고, 특정 종의 활성과 편집 효율을 증가시킨다. 또한 코돈 최적화는 번역 및 단백질 발현의 증가를 초래한다.

[0220] 일부 구현예에서, Cas9 단백질은 진핵 세포에서 발현되도록 코돈-최적화된다. 일부 구현예에서, Cas9 단백질은 인간 세포에서 발현되도록 코돈-최적화된다.

[0221] **프로토스페이서 인접 모티브(PAM)**

[0222] 각 Cas 엔도뉴클레아제는 비표적화된, 즉, 상보적 DNA 가닥 상에 프로토스페이서 인접 모티프(PAM)로 알려진 특정 서열이 있는 경우에만 표적 서열에 결합한다. 다양한 박테리아 종으로부터 단리된 Cas 뉴클레아제는 상이한 PAM 서열을 인식한다. 예를 들어 SpCas9 뉴클레아제(*스타필로코커스 피오게네스* 유래)는 PAM 서열 5'-NGG-3'(여

기서 "N"은 임의의 뉴클레오티드 염기일 수 있다)의 업스트림을 절단하고, SaCas9(스타필로코커스 아우레우스 유래)는 표적에서 PAM 서열 5'-NNGRR(N)-3'을 인식한다. 따라서, 상이한 Cas 단백질이 표적으로 할 수 있는 게놈 내 위치는 고유 PAM 서열의 위치에 의해 제한된다.

[0223] *스트렙토코커스 콘스텔라투스* 및 *에자키엘라 페루엔시스* 및 *펩토니필러스 sp. 마르세유-P3761* 종으로부터 조작된 본원에 개시된 Cas9 단백질이 공통 PAM 서열 5'-NGG-3'를 인식한다. *스트렙토코커스 콘스텔라투스* 및 *에자키엘라 페루엔시스* 및 *펩토니필러스 sp. 마르세유-P3761* 종으로부터 조작된 본원에 개시된 Cas9 단백질이 공통 PAM 시퀀스 5'-NGG-3'를 인식한다. 일부 구현예에서, 본원에 개시된 Cas9 단백질은 공통 PAM 서열 5'-NGC-3'을 인식하도록 조작된다. 예시적인 구현예가 아래에 기술되어 있고, 비제한적이어야 한다. 일부 구현예에서, *스트렙토코커스 콘스텔라투스* 유래의 Cas9 단백질은 공통 PAM 서열 5'-NGC-3'을 인식하도록 조작된다. 일부 구현예에서, NGC PAM 변이체는 ScoCas9(서열번호 1)와 관련하여 D1117M, S118Q, E1201F, A1299R, D1309A, R1312E 및 T1314R(총칭하여 "MQFRAER"라 함)로부터 선택되거나 이에 상응하는 1개 이상의 아미노산 치환을 포함한다. 일부 구현예에서, NGC PAM 변이체는 자연 발생 SpyCas9(서열번호 173)와 관련하여 D1135M, S1136Q, G1218K, E1219F, A1322R, D1332A, R1335E 및 T1337R(총칭하여 "MQKFRAER"라 함)로부터 선택되거나 이에 상응하는 1개 이상의 아미노산 치환을 포함한다. 일부 구현예에서, 유사하거나 상응하는 아미노산 치환이 SirCas9, VapCas9, EpeCas9, LfeCas9 또는 PmaCas9에 대해 이루어질 수 있다.

[0224] *스트렙토코커스 피오게네스* Cas9(*SpyCasP*; GenBank: QSG91308.1)

[0225] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE
ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL
RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS
GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIAR
KKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKY
FDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(서열번호 173).

[0226] 일부 구현예에서, 본원에 기술된 Cas9 단백질은 임의의 다른 PAM 서열과 결합하지 않거나 활성을 나타내지 않는다.

[0227] **RNA 가이드**

[0228] RNA 가이드는 표적 서열에 대해 상보성을 갖는 폴리뉴클레오티드 서열을 포함한다. RNA 가이드는 표적 핵산 서열과 혼성화하고, 표적 핵산에 대한 CRISPR 복합체의 서열-특이적 결합을 지시한다. 일부 구현예에서, RNA 가이드는 표적 핵산 서열에 대해 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 상보성을 가진다.

[0229] 일부 구현예에서, RNA 가이드는 약 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50 또는 75개 이상의 뉴클레오티드 길이를 가진다. 일부 구현예에서, RNA 가이드는 약 18 내지 24개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, RNA 가이드는 표적 핵산 서열에서 약 18 내지 24개의 뉴클레오티드에 상보적이다. 예를 들어, RNA 가이드는 표적 핵산 서열에서 약 18, 19, 20, 21, 22, 23 또는 24개의 뉴클레오티드에 상보적이다. 일부 구현예에서, RNA 가이드는 약 18 내지 22개의 뉴클레오티드에 상보적이다. 일부 구현예에서, RNA 가이드는 약 18 내지 21개의 뉴클레오티드에 상보적이다. 일부 구현예에서, RNA 가이드는 약 18 내지 20개의 뉴클레오티드에 상보적이다. 일부 구현예에서, RNA 가이드는 표적 핵산 서열에서 20개의 뉴클레오티드에 상보적이다.

[0230] RNA 가이드는 임의의 표적 서열을 표적으로 하기 위해 설계될 수 있다. 최적의 정렬은 Needleman-Wunsch 알고리즘, Smith-Waterman 알고리즘, Burrows-Wheeler 알고리즘, Clustlw, ClustlX, BLAST, Novoalign, SOAP, Maq

및 ELAND를 포함하여 임의의 서열 정렬 알고리즘을 사용하여 결정된다.

[0231] 일부 구현예에서, RNA 가이드는 세포 게놈 내의 고유 표적 서열에 대해 표적된다. 일부 구현예에서, RNA 가이드는 PAM 서열이 결여되도록 설계된다. 일부 구현예에서, RNA 가이드 서열은 mFold 또는 Geneious를 포함하는 접힘 알고리즘을 사용하여 최적의 이차 구조를 갖도록 설계된다. 일부 구현예에서, RNA 가이드의 발현은 유도성 프로모터, 예를 들어 호르몬 유도성, 테트라사이클린 또는 독시사이클린 유도성, 아라비노스 유도성 또는 광유도성 프로모터 하에서 이루어질 수 있다.

[0232] 일부 구현예에서 CRISPR 시스템은 1개 이상의 RNA 가이드, 예를 들어 crRNA, tracrRNA 및/또는 sgRNA를 포함한다. 따라서, 일부 구현예에서 RNA 가이드는 crRNA를 포함한다. 일부 구현예에서, RNA 가이드는 tracrRNA를 포함한다. 일부 구현예에서, RNA 가이드는 sgRNA를 포함한다. 일부 구현예에서, CRISPR 시스템은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 15개 이상의 RNA 가이드를 포함하는 다수 RNA 가이드를 포함한다.

[0233] 일부 구현예에서, RNA 가이드는 crRNA를 포함한다. 일부 구현예에서, CRISPR 시스템은 2 내지 15개의 crRNA를 포함하는 다수 crRNA를 포함한다. 일부 구현예에서, crRNA는 전구체 crRNA(pre-crRNA)이며, 이는 직접 반복 서열, 스페이서 서열 및 직접 반복 서열을 포함한다. 일부 구현예에서, crRNA는 첨두절단된 직접 반복 서열을 포함하는 처리된 또는 성숙 crRNA이다.

[0234] 일부 구현예에서, CRISPR 연관 단백질은 pre-crRNA를 절단하여, 처리된 또는 성숙 crRNA를 형성한다.

[0235] 일부 구현예에서, CRISPR 연관 단백질은 성숙 crRNA와 복합체를 형성하고, 스페이서 서열은 표적 핵산 내 상보적 서열에 대해 복합체를 표적으로 한다. 일부 구현예에서, RNA 가이드는 표적 핵산에 적합한 조건 하에서 혼성화할 수 있는 스페이서 서열 및 직접 반복 서열을 포함한다.

[0236] 일부 구현예에서, crRNA의 스페이서 길이는 약 15 내지 50개의 뉴클레오티드 범위일 수 있다. 일부 구현예에서, RNA 가이드의 스페이서 길이는 적어도 16개의 뉴클레오티드, 적어도 17개의 뉴클레오티드, 적어도 18개의 뉴클레오티드, 적어도 19개의 뉴클레오티드, 적어도 20개의 뉴클레오티드, 적어도 21개의 뉴클레오티드 또는 적어도 22개의 뉴클레오티드이다. 일부 구현예에서, 스페이서 길이는 15 내지 17개의 뉴클레오티드(예를 들어, 15, 16 또는 17개의 뉴클레오티드), 17 내지 20개의 뉴클레오티드(예를 들어, 17, 18, 19 또는 20개의 뉴클레오티드), 20 내지 24개의 뉴클레오티드(예를 들어, 20, 21, 22, 23 또는 24개의 뉴클레오티드), 23 내지 25개의 뉴클레오티드(예를 들어, 23, 24 또는 25개의 뉴클레오티드), 24 내지 27개의 뉴클레오티드, 27 내지 30개의 뉴클레오티드, 30 내지 45개의 뉴클레오티드(예를 들어, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44 또는 45개의 뉴클레오티드), 30 또는 35 내지 40개의 뉴클레오티드, 41 내지 45개의 뉴클레오티드, 45 내지 50개의 뉴클레오티드(예를 들어, 45, 46, 47, 48, 49 또는 50개의 뉴클레오티드) 이상이다.

[0237] 일부 구현예에서, RNA 가이드는 약 16 내지 26개의 뉴클레오티드 길이를 가지는 직접 반복(DR) 서열을 포함한다. 예를 들어, 일부 구현예에서, DR은 약 16개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 17개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 18개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 19개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 20개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 21개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 22개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 23개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 24개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 25개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, DR은 약 26개의 뉴클레오티드 길이를 가진다.

[0238] 일부 구현예에서, crRNA는 뉴클레오티드 가이드 서열 및 DR 서열을 포함한다. 뉴클레오티드 가이드 서열은 약 18 내지 24개의 뉴클레오티드 길이를 가질 수 있다. 따라서, 일부 구현예에서, 뉴클레오티드 가이드 서열은 약 18개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, 뉴클레오티드 가이드 서열은 약 19개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, 뉴클레오티드 가이드 서열은 약 20개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, 뉴클레오티드 가이드 서열은 약 21개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, 뉴클레오티드 가이드 서열은 약 22개의 뉴클레오티드 길이를 가진다. 일부 구현예에서, crRNA는 약 22개 뉴클레오티드 길이의 뉴클레오티드 가이드 서열 및 약 22개의 뉴클레오티드 길이의 직접 반복부를 포함한다.

[0239] 일부 구현예에서, crRNA 서열은, CRISPR-연관 단백질과 복합체를 형성할 수 있고 실질적인 뉴클레아제 활성 없이 특이적 표적에 결합할 수 있는 "죽은 crRNA", "죽은 가이드" 또는 "죽은 가이드 서열"로 변형될 수 있다.

[0240] 일부 구현예에서, crRNA는 당 인산염 골격(backbone) 또는 염기에서 화학적으로 변형될 수 있다. 일부 구현예에서, crRNA는 2'O-메틸, 2'-F 또는 잠긴 핵산을 사용하여 변형되어, 뉴클레아제 저항성 또는 염기쌍 짝짓기를 향

상시킬 수 있다. 일부 구현예에서, crRNA는 2-티오우리디엔 또는 N6-메틸아데노신과 같은 변형된 염기를 함유할 수 있다.

[0241] 일부 구현예에서, crRNA는 다른 올리고뉴클레오티드, 펩티드, 단백질, 태그, 염료 또는 폴리에틸렌 글리콜과 접합된다.

[0242] 일부 구현예에서, crRNA는 3차원 구조로 인해 특이적 표적 분자에 결합할 수 있는 앱타머 또는 리보스위치 서열을 포함할 수 있다.

[0243] 일부 구현예에서, 트랜스-활성화 RNA(tracrRNA)는 crRNA와 연관되어, Cas9 단백질과의 복합체 형성을 촉진한다. 일부 구현예에서, tracrRNA 서열은 약 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100개 이상 또는 이를 초과하는 뉴클레오티드 길이를 가진다. 일부 구현예에서, tracrRNA는 약 70개의 뉴클레오티드 길이를 가진다.

[0244] 일부 구현예에서, tracrRNA 및 crRNA는 단일 가이드 RNA(sgRNA)라고 불리는 단일 전사체에 함유된다. 일부 구현예에서, sgRNA는 tracrRNA와 sgRNA 사이에 루프를 포함한다.

[0245] 일부 구현예에서, 루프 형성 서열은 3, 4, 5개 이상의 뉴클레오티드 길이를 가진다. 일부 구현예에서, 루프는 서열 GAAA, AAAG, CAAA, AAAC, UUUU, UUAUAU, UUA, UUU 및/또는 AAUCA를 가진다. 일부 구현예에서, 루프는 서열 GAAA를 가진다. 일부 구현예에서, 루프는 서열 AAAG를 가진다. 일부 구현예에서, 루프는 서열 CAAA를 가진다. 일부 구현예에서, 루프는 서열 AAAC를 가진다. 일부 구현예에서, 루프는 서열 AAUCA를 가진다. 일부 구현예에서, 루프는 서열 UUUU를 가진다. 일부 구현예에서, 루프는 서열 UUAUAU를 가진다. 일부 구현예에서, 루프는 서열 UUA를 가진다. 일부 구현예에서, 루프는 서열 UUU를 가진다. 일부 구현예에서, 루프는 서열 AAUCA를 가진다.

[0246] 일부 구현예에서, tracrRNA 및 crRNA는 헤어핀 루프를 형성한다. 일부 구현예에서, sgRNA는 적어도 2개 이상의 헤어핀을 가진다. 일부 구현예에서, sgRNA는 2개, 3개, 4개 또는 5개의 헤어핀을 가진다.

[0247] 일부 구현예에서, sgRNA는 6개의 뉴클레오티드를 포함하는 폴리T 서열을 포함하는 전사 종결 서열을 포함한다.

[0248] 일부 구현예에서, sgRNA는 ScoCas9에 대해

[0249] 5'-GUUUUAGAGCUGUGCUGUUUAAACAACACAGCAAGUUAAAAUAAGGCUUUGUCCGUACUC (서열번호 3),

[0250] SirCas9에 대해 5'-

[0251] GUUUUAGAGUUGUGUUAUUGAAAAAUAACACAACGAGUUAAAAUAAAGCUUAUGCUUAAAUGCCAGCUUUGCUGGUGUCAUUUAGAUGACUUUACUAAGGUUGCUUCGGCAACCUUUUU-3' (서열번호 7),

[0252] VapCas9에 대해 5'-

[0253] GUUUGAGAGUAGUGUGAAAACAUUACGAGUUCAAAUACAAAUUAAUUUACAAUGCCUUCGGGCUGCCCGACGUAGGGCACCUACUCUCAAUUCUUCGGAAUUGAGUU-3' (서열번호 13),

[0254] EpeCas9에 대해 5'-

[0255] GUUUGAGAGUUAUGUAAUUGAAAAAUUACAUGACGAGUUCAAAUAAAAAUUUAUUCAAACCGCCUAUUUAUAGGCCGCAGAUGUUCUGCAUUAUGCUUGCUAUUGCAAGCUU-3' (서열번호 19),

[0256] LfeCas9에 대해 5'-

[0257] GUCUUGGAUGAGUGUGAAAACACUCAUAGUCAAGAUCAAACGAGUGGUUUUCCACGAGUUAUUACUUUUGAGGUCUUAUAUGGCCCAUACAUAAAAAGGAGUCGGAAUUUCCGGCUCCUUUUCUU-3' (서열번호 95) 및

[0258] PmaCas9에 대해 5'-

[0259] GUUUUAGAGCCAUGUAGAAAUACAUUGCAAGUUAAAAUAAGGCUUUGUCCGUAAUCAACUUGAAAAAGUGGCGCUGUUUCGGCGCUUU-3' (서열번호 96)에 적어도 80% 동일성을 갖는 서열을 포함한다.

[0260] 가이드 RNA는 Cas9의 5' 단부에 추가된다. 일부 구현예에서, sgRNA는 서열번호 3, 7, 13 19, 95 또는 96과 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99% 이상의 동일성을 갖는 서열을 포함한다. 일부 구현예에서, sgRNA는 서열번호 3, 7, 13, 19, 95 또는 96

과 동일한 서열을 포함한다.

[0261] 일부 구현예에서, tracrRNA는 동일한 전사체에 crRNA 서열과 함께 함유되지 않은 별도의 전사체이다.

[0262] **Cas9 융합 단백질**

[0263] 일부 구현예에서, Cas9 효소는 1개 이상의 이종 단백질 도메인에 융합된다. 일부 구현예에서, Cas9 효소는 약 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 이상의 단백질 도메인에 융합된다. 일부 구현예에서, 이종 단백질 도메인은 Cas9 효소의 C-말단에 융합된다. 일부 구현예에서, 이종 단백질 도메인은 Cas9 효소의 N-말단에 융합된다. 일부 구현예에서, 이종 단백질 도메인은 Cas9 효소의 C-말단과 N-말단 사이에 내부적으로 융합된다. 일부 구현예에서, 내부 융합은 Cas9 RuvCI, RuvC II, RuvCIII, HNH, REC I 또는 PAM 상호작용 도메인 내에서 이루어진다.

[0264] Cas9 단백질은 또 다른 단백질 도메인에 직접적으로 또는 간접적으로 연결될 수 있다. 일부 구현예에서, 적합한 CRISPR 시스템은 Cas9 단백질 및 이종 단백질을 연결하는 링커 또는 스페이서를 함유한다. 아미노산 링커 또는 스페이서는 일반적으로 2개 단백질 모이어티 사이에 알파 나선과 같은 구조를 개재시키도록 또는 유연하도록 설계된다. 링커 또는 스페이서는 상대적으로 짧을 수 있거나 길 수 있다. 일반적으로 링커 또는 스페이서는 예를 들어 1 내지 100개(예: 1 내지 100개, 5 내지 100개, 10 내지 100개, 20 내지 100개 30 내지 100개, 40 내지 100개, 50 내지 100개, 60 내지 100개, 70 내지 100개, 80 내지 100개, 90 내지 100개, 5 내지 55개, 10 내지 50개, 10 내지 45개, 10 내지 40개, 10 내지 35개, 10 내지 30개, 10 내지 25개, 10 내지 20개) 아미노산 길이를 함유한다. 일부 구현예에서, 링커 또는 스페이서는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 또는 100개 아미노산 길이와 동일하거나 이보다 길다. 일반적으로 링커가 길수록 입체 장애가 감소할 수 있다. 일부 구현예에서, 링커는 글리신과 세린 잔기의 혼합물을 포함한다. 일부 구현예에서, 링커는 트레오닌, 프롤린 및/또는 알라닌 잔기를 추가로 포함할 수 있다.

[0265] 일부 구현예에서 Cas9 단백질은 세포 위치 신호, 에피토프 태그, 리포터 유전자, 및 효소 활성, 후성적 변형 활성, RNA 절단 활성, 핵산 결합 활성, 전사 조절 활성을 갖는 단백질 도메인에 융합된다. 일부 구현예에서, Cas9 단백질은 핵 위치 서열(NLS), FLAG 태그, HIS 태그 및/또는 HA 태그에 융합된다.

[0266] 적합한 융합 파트너에는 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스파타제 활성, 유비퀴틴 리가제 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 데아데닐화 활성, SUMO화 활성, 탈SUMO화 활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성, 탈미리스토일화 활성, 인테그라제 활성, 트랜스포사제 활성, 재조합효소 활성, 폴리머라제 활성, 리가제 활성, 헬리카제 활성 또는 뉴클레아제 활성을 제공하는 폴리펩티드가 포함되지만 이에 국한되지는 않으며, 이들 중 임의의 것은 DNA 또는 DNA-연관 폴리펩티드(예: 히스톤 또는 DNA 결합 단백질)를 변형시킬 수 있다. 일부 구현예에서 Cas9 단백질은 히스톤 데메틸라제, 전사 활성화인자 또는 데아미나제에 융합된다.

[0267] 추가의 적합한 융합 파트너에는 경계 요소(예: CTCF), 주변 동원을 제공하는 단백질 및 그의 단편(예: 라민(Lamin) A, 라민 B 등) 및 단백질 도킹 요소(예: FKBP/FRB, Pill/Abyl 등)가 포함되지만 이에 국한되지는 않는다.

[0268] 특정 구현예에서, Cas9는 예를 들어 염기 편집에 사용하기 위해 시티딘 또는 아데노신 데아미나제 도메인에 융합된다. 일부 구현예에서, Cas9는 아데닌 및 시토신 염기 편집기(ACBE 또는 CASE)에 융합되고, ACBE 또는 CABE는 Cas9 닉카제(nCas9)의 N- 및 C-말단에 TadA 이종이량체 및 활성화-유도된 시티딘 데아미나제(AID)를 융합시킴으로써 발생된다. 일부 구현예에서, ACBE 또는 CABE는 동일 표적 부위에서 C에서 T로 및 A에서 G로 염기 편집을 동시에 유도한다. Xie, J 등 ACBE, a new base editor for simultaneous C-to-T and A-to-G substitutions in mammalian systems. BMC Biology (18: 131), 2020)

[0269] 일부 구현예에서, 용어 "시티딘 데아미나제" 및 "시토신 데아미나제"는 상호 교환하여 사용될 수 있다. 특정 구현예에서, 시티딘 데아미나제 도메인은 본원에 기술된 임의의 시티딘 데아미나제와 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99% 이상의 서열 동일성을 가질 수 있다. 일부 구현예에서, 시티딘 데아미나제 도메인은 시티딘 데아미나제 활성(예를 들어, C를 U로 전환)을 가진다. 특정 구현예에서, 아데노신 데아미나제 도메인은 본원에 기술된 임의의 아데노신 데아미나제와 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99% 이상의 서열 동일성을 가질 수 있다. 일부 구현예에서, 아데노신 데아미나제 도메인은 아데노신 데아미나제 활성(예를 들어, A를 I로 전환)을 가진다. 일부 구현예에서, 용어 "아데노신 데아미나제" 및 "아데닌 데아미나제"는 상호 교환하여 사용될 수 있다.

[0270] 일부 구현예에서, 시티딘 데아미나제는 아포지질단백질 B mRNA 편집 복합체(APOBEC) 계열 데아미나제의 전부 또는 일부를 포함할 수 있다. APOBEC은 진화적으로 보존된 시티딘 데아미나제 계열이다. 이 계열의 구성원은 C에서 U로 편집하는 효소이다. APOBEC 유사 단백질의 N-말단 도메인은 촉매 도메인인 반면, C-말단 도메인은 유사 촉매 도메인이다. 보다 구체적으로, 촉매 도메인은 아연 의존성 시티딘 데아미나제 도메인이고, 시티딘 탈아미노화에 중요하다. APOBEC 계열 구성원에는 APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D(현재 "APOBEC3E"는 이를 지칭함), APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4 및 활성화-유도된(시티딘 또는 시토신) 데아미나제를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC1 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC2 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3A 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3B 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3C 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3D 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3E 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3F 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3G 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC3H 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 APOBEC4 데아미나제의 전부 또는 일부를 포함한다. 일부 구현예에서, 융합 단백질에 혼입된 데아미나제는 활성화-유도된 데아미나제(AID)의 전부 또는 일부를 포함한다. 일부 구현예에서 융합 단백질에 혼입된 데아미나제는 시티딘 데아미나제 1(CDA1)의 전부 또는 일부를 포함한다. 융합 단백질은 임의의 적합한 유기체(예를 들어, 인간 또는 랫트)로부터의 데아미나제를 포함할 수 있음이 인식되어야 한다. 일부 구현예에서, 융합 단백질의 데아미나제 도메인은 인간, 침팬지, 고릴라, 원숭이, 소, 개, 랫트 또는 마우스에서 유래된다. 일부 구현예에서, 융합 단백질의 데아미나제 도메인은 랫트(예를 들어, 랫트 APOBEC1)로부터 유래된다. 일부 구현예에서, 데아미나제 도메인은 인간 APOBEC1이다. 일부 구현예에서, 데아미나제 도메인은 pmCDA1이다. 예시적인 시티딘 데아미나제의 서열이 아래에 제공된다.

[0271] pmCDA1 (*페트로마이존 마리너스(Petromyzon marinus)*)

[0272] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAE
KILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTK
SPAV (서열번호 22)

[0273] 인간 AID:

[0274] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNP
NLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKAPV (서열번호 23)

[0275] 인간 AID:

[0276] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNP
NLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL
(밑줄: 핵 위치 서열; 이중 밑줄; 핵외 수송 신호) (서열번호 24)

[0277] 마우스 AID:

[0278] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNP
NLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF
(밑줄: 핵 위치 서열; 이중 밑줄: 핵외 수송 신호) (서열번호 25)

[0279] 개 AID:

[0280] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYP
NLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL
(밑줄: 핵 위치 서열; 이중 밑줄: 핵외 수송 신호) (서열번호 26)

[0281] 소 AID:

[0282] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYP

NLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (밑줄: 핵 위치 서열; 이중 밑줄: 핵외 수송 신호) (서열번호 27)

[0283] 랫트 AID:

[0284] MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYL RNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWE GLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLGL (서열번호 28)

[0285] (밑줄: 핵 위치 서열; 이중 밑줄: 핵외 수송 신호)

[0286] clAID (*카니스 루푸스 파밀리아리스(Canis lupus familiaris)*):

[0287] MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYP NLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (서열 번호 29)

[0288] btAID (*보스 타우루스(Bos taurus)*):

[0289] MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYP NLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (서 열번호 30)

[0290] mAID (*무스 무스쿨루스(Mis musculus)*):

[0291] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNP NLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (서열 번호 31)

[0292] rAPOBEC-1 (*라투스 노르베기커스(Rattus Norvegicus)*):

[0293] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITE FLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT FFTIALQSCHYQRLPPHILWATGLK (서열번호 32)

[0294] maAPOBEC-1 (*메소크리세투스 아우라투스(Mesocricetus auratus)*):

[0295] MSSETGPVVVDPTLRRRIEPHEFDAFFDQGELRKETCLLYEIRWGGRHNIWRHTGQNTSRHVEINFIEKFTSERYFYPSTRCSIVWFLSWSPCGECSKAITE FLSGHPNVTLFIYAARLYHHTDQRNRQGLRDLISRGVTIRIMTEQEYCYCWRNFVNYPPSNEVYWPRYPNLWMRLYALELYCIHLGLPPCLKIKRRHQYPLT FFRLNLQSCHYQRIPPHILWATGFI (서열번호 33)

[0296] ppAPOBEC-1 (*퐁고 피그마에우스(Pongo pygmaeus)*):

[0297] MTSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIRE FLSQHPGVTLVIYVARLFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLA FFRLHLQNCHYQTIPPHILLATGLIHPSVTWR (서열번호 34)

[0298] ocAPOBEC1 (*오리코라그 쿠니쿨러스(Oryctolagus cuniculus)*):

[0299] MASEKGPSNKDYTLRRRIEPWEFEVFFDPQELRKEACLLYEIKWGASSKTWRSSGKNTTNHVEVNFLEKLTSEGRLGPSTCCSITWFLSWSPCWECSMAIRE FLSQHPGVTLIIFVARLFQHMDRRNRQGLKDLVTSGVTVRVMSVSEYCYCWENFVNYPPGKAAQWPRYPPRWMLMYALELYCIILGLPPCLKISRRHQKQLT FFSLTPQYCHYKMIPPYILLATGLLQPSVPWR (서열번호 35)

[0300] mdAPOBEC-1 (*모노델피스 도메스티카(Monodelphis domestica)*):

[0301] MNSKTGPSVGDATLRRRIKPWEFVAFFNPQELRKETCLLYEIKWGNQNIWRHSNQNTSQHAEINFMEKFTAERHFNSSVRCSITWFLSWSPCWECSKAIRKF LDHYPNVTLAIFISRLYWHMDQQHRQGLKELVHSGVTIQIMSYSEYHYCWRNFVDYPQGEEDYWPKYPYLWIMLYVLELHCIILGLPPCLKISGSHSNQLAL FSLDLQDCHYQKIPYNVLVATGLVQPFVTWR (서열번호 36)

[0302] ppAPOBEC-2 (*퐁고 피그마에우스*):

[0303] MAQKEEAAAATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEA FFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEELEIQDALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQP

WEDIQENFLYYEEKLADILK (서열번호 37)

[0304] btAPOBEC-2 (*보스 타우루스*):

[0305] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEA FFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEP WEDIQENFLYYEEKLADILK (서열번호 38)

[0306] mAPOBEC-3-(1) (*무스 무스쿨루스*):

[0307] MQPQRLGPRAGMGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPR EEFKITWYMSWSPCFECAEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKL QEILRPCYISVPSSSSSTLSNICLTKGLPETRFWVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLD KIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISR RTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS (서열번호 39)

[0308] 마우스 APOBEC-3-(2):

[0309] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSW SPCFEC*AEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPV PSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVT ITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKE SWGLQDLVNDFGNLQLGPPMS (이탤릭체: 핵산 편집 도메인) (서열번호 40)

[0310] 랫트 APOBEC-3:

[0311] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNRLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMS WSPCFEC*AEQVLRFLATHHNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIP VPSSSSSTLSNICLTKGLPETRFCVERRRVHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQV IITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIK ESWGLQDLVNDFGNLQLGPPMS (이탤릭체: 핵산 편집 도메인) (서열번호 41)

[0312] hAPOBEC-3A (*호모 사피엔스*(*Homo sapiens*)):

[0313] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCF SWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN (서 열번호 42)

[0314] hAPOBEC-3F (*호모 사피엔스*):

[0315] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAK LAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHF KNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARL YYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE (서열번호 43)

[0316] 레수스 마카쿠에 APOBEC-3G:

[0317] MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYHPEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLA KDPKVTLTIFVARLYYFWKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPW VSGQHETYLCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISNNEHVSLCIFAARI YDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (이탤릭체: 핵산 편집 도메인: 세포질 위 치 신호) (서열번호 44)

[0318] 침팬지 APOBEC-3G:

[0319] MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TR DVATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTS NFNNELWVRGRHETYLCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSC*AQEMAKFISNNKHVS LCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (서열번호 45) (이탤릭 체: 핵산 편집 도메인: 세포질 위치 신호)

[0320] 녹색 원숭이 APOBEC-3G:

[0321] MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKD*HPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALRILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSC*AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYSEFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI (서열번호 46) (이탤릭체: 핵산 편집 도메인: 세포질 위치 신호)

[0322] 인간 APOBEC-3G:

[0323] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKY*HPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSC*AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (서열번호 47) (이탤릭체: 핵산 편집 도메인: 세포질 위치 신호)

[0324] 인간 APOBEC-3F:

[0325] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEH*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETH*CHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE (서열번호 48)(이탤릭체: 핵산 편집 도메인)

[0326] 인간 APOBEC-3B:

[0327] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQY*HAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (서열번호 49) (이탤릭체: 핵산 편집 도메인)

[0328] 랫트 APOBEC-3B:

[0329] MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCALPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFYDCKLQEIFSRMNLLREDVFYLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVEILFLEKMRSMELSQVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFWRKKFQKGLCTLWRSGIHVDVMDLPQFADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL (서열번호 50)

[0330] 소 APOBEC-3B:

[0331] MDGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAPYYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAERFIDKINSLDLNPSQSYKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILTAPI (서열번호 51)

[0332] 침팬지 APOBEC-3B:

[0333] MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNNDPLVLRRHQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSFHSLTSCSIQPPCSSRIRETEGWASVSKEGRDLG (서열번호 52)

[0334] 인간 APOBEC-3C:

[0335] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETH*CHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ (서열번호 53)

[0336] (이탤릭체: 핵산 편집 도메인)

[0337] 고릴라 APOBEC-3C

[0338] MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETH*CHAERCFLSWECDDILSPNTNYQVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRFLKRRLQEILE (서열번호 54)

[0339] (이탤릭체: 핵산 편집 도메인)

[0340] 인간 APOBEC-3A:

[0341] MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGR*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN (서열번호 55)

[0342] (이탤릭체: 핵산 편집 도메인)

[0343] 레수스 마카쿠에 APOBEC-3A:

[0344] MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCGDYGC*HVELRFLCEVPSWQLDPAQTYRVTWFISWSPC*FRRGCAGQVRVFLQENKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN (서열번호 56)

[0345] (이탤릭체: 핵산 편집 도메인)

[0346] 소 APOBEC-3A:

[0347] MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPC*HAELYFLGKIHSWNLDRNQHYRLTCFISWSPC*YDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN (서열번호 57)

[0348] (이탤릭체: 핵산 편집 도메인)

[0349] 인간 APOBEC-3H:

[0350] MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV (서열번호 58)

[0351] (이탤릭체: 핵산 편집 도메인)

[0352] 레수스 마카쿠에 APOBEC-3H:

[0353] MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR (서열번호 59)

[0354] 인간 APOBEC-3D:

[0355] MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITW FVSWNPC*LPCVVKVTKFLAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ (서열번호 60)

[0356] (이탤릭체: 핵산 편집 도메인)

[0357] 인간 APOBEC-1:

[0358] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR (서열번호 61)

[0359] 마우스 APOBEC-1:

[0360] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLT

FFTITLQTCHYQRIPPHLLWATGLK (서열번호 62)

[0361] 랫트 APOBEC-1:

[0362] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITE
FLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT
FFTIALQSCHYQRLPPHILWATGLK (서열번호 63)

[0363] 인간 APOBEC-2:
MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEA
FFNTILPAFDPALRYNVTWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQP
WEDIQENFLYYEEKLADILK (서열번호 64)

[0364] 마우스 APOBEC-2:

[0365] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEA
FFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEP
WEDIQENFLYYEEKLADILK (서열번호 65)

[0366] 랫트 APOBEC-2:

[0367] MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTFLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEA
FFNTILPAFDPALKYNVTWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEP
WEDIQENFLYYEEKLADILK (서열번호 66)

[0368] 소 APOBEC-2:

[0369] MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEA
FFNSIMPTFDPALRYMVTWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEP
WEDIQENFLYYEEKLADILK (서열번호 67)

[0370] *페트로마이존 마리너스* CDA1 (pmCDA1):

[0371] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAE
KILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQ
LNENRWLEKTLKRAEKRRSELSFMIQVKILHTTKSPAV (서열번호 68)

[0372] 인간 APOBEC3G D316R D317R:

[0373] MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTR
DMATFLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKFNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHFMLGEILRHSMDPPTFTFN
FNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKKHVSLC
IFTARIYRRQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (서열번호 69)

[0374] 인간 APOBEC3G 사슬 A:

[0375] MDPPTFTFNFNNEPWWGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFI
SKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISFTYSEFKHCWDTFVDHQGCPFQPWDGLD EHSQDLSGRLRAILQ (서열번호 70)

[0376] 인간 APOBEC3G 사슬 A D120R D121R:

[0377] MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKF
ISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISFMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ (서열번호 71)

[0378] hAPOBEC-4 (*호모 사피엔스*):

[0379] MEPIYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTFPQTKHLTFYELKTSSGSLVQKGHASSCTGNYIHPESMLFEMN
GYLDSAIYNNDSIRHIILYSNNSPCNEANHCCISKMYNFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFISG
VSGSHVFQPILTGRALADRHNAYEINAITGVKPYFTDVLLQTKRNPNTKAQEALESYPLNNAFPGQFFQMPSGQLQPNLPPDLRAPVVFVLVPLRDLPPMHM
GQNPNKPRNIVRHLNMPQMSFQETKDLGRLPTGRSVEIVEITEQFASSKEADEKKKKKGKK (서열번호 72)

[0380] mAPOBEC-4 (*무스 무스쿨루스*):

[0381] MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNP NLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIGIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF (서열번호 73)

[0382] rAPOBEC-4 (*라투스 노르베기커스*):

[0383] MEPLYEEYLTHSGTIVKPYYWLSVSLNCTNCPYHIRTGEEARVPYTEFHQTFGFPWSTYPQTKHLTFYELRSSSGNLIQKGLASNCTGSHTHPESMLFERDG YLDSLIFHDSNIRHIILYSNNSPCDEANHCCISKMYNFLMNYPEVTLSVFFSQLYHTENQFPTSAWNREALRGLASLWPQVTLSAISGGIWQSILETFVSGI SEGLTAVRPFTAGRTLTDRYNAYEINCITEVKPYFTDALHSWQKENQDQKVWAASENQPLHNTTPAQWQPDMSQDCRTPAVFMLVPYRDLPPIHVNPSPQKP RTVVRHLNTLQLSASKVKALRKSPSGRPVKKEEARKGSTRSQEANETNKSKWKKQTLFIKSNICHLLEREQKKIGILSSWSV (서열번호 74)

[0384] mfAPOBEC-4 (*마카카 퍼시쿨라리스(Macaca fascicularis)*):

[0385] MEPTYEEYLANHGTIVKPYYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPYGTTYPQTKHLTFYELKTSSGSLVQKGHASSCTGNYIHPESMLFEMN GYLDSAIYNNDSIRHIILYCNNSPCNEANHCCISKVYNFLITYPGITLSIYFSQLYHTEMDFPASAWNREALRSLASLWPRVVLSPISGGIWHSVLHSFVSG VSGSHVFQPILTGRALTDRYNAYEINAITGVKPFFTDVLLHTKRNPNTKAQMALESYPLNNAFPGQSFQMTSGIPPDLRAPVVFVLLPLRDLPPMHMGQDPN KPRNIIRHLNMPQMSFQETKDLERLPTRRSVETVEITERFASSKQAEEKTKKKKGKK (서열번호 75)

[0386] pmCDA-1 (*페트로마이존 마리너스*):

[0387] MAGYECVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKL NPWLKNLLEEQGHTLTMHFSRIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGIP LHLFTLQTPLLSGRVVWWRV (서열번호 76)

[0388] pmCDA-2 (*페트로마이존 마리너스*):

[0389] MELREVVDCALASCVRHEPLSRVAFLRCFAAPSQKPRGTVILFYVEGAGRGVTGGHAVNYNKQGTSIHAEVLLLSAVRAALLRRRRCEDGEEATRGCTLHCY STYSPCRDCVEYIQEFGASTGVRVVIHCCRLYELDVNRRRSEAEGVLRSLSRLGRDFRLMGPRDAIALLLGGRLANTADGESGASGNAWVTETNVVEPLVDM TGFGDEDLHAQVQRNKQIREAYANYASAVSLMLGELHVDPDKFPFLAEFLAQTSVEPSGTPRETRGRPRGASSRGPEIGRQRPADFERALGAYGLFLHPRIV SREADREEIKRDLIVVMRKHNYQGP (서열번호 77)

[0390] pmCDA-5 (*페트로마이존 마리너스*):

[0391] MAGDENVRVSEKLDFDTFEFQFENLHYATERHRTYVIFDVKPQSAGGRSRRLWGYIINNPNVCHAELILMSMIDRHLESNPGVYAMTWYMSWSPCANCSSKL NPWLKNLLEEQGHTLMMHFSRIYDRDREGDHRGLRGLKHVSNSFRMGVVGRAEVKECLAEYVEASRRTLTWLDTTESMAAKMRRKLFCILVRCAGMRESGMP LHLFT (서열번호 78)

[0392] yCD (*사카로마이세스 세레비시아에*):

[0393] MVTGGMASKWDQKGMDIAYEEAALGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSATLHGEISTLENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYG IPRCVVGENVNFKSKGEKYLQTRGHEVVVVDDERCKKIMKQFIDERPQDWFEDIGE (서열번호 79)

[0394] rAPOBEC-1 (델타 177-186):

[0395] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITE FLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRGLPPCLNILRRKQPQLTFFTIALQSCHY QRLPPHILWATGLK (서열번호 80)

[0396] rAPOBEC-1 (델타 202-213):

[0397] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITE FLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQHY QRLPPHILWATGLK (서열번호 81)

[0398] 마우스 APOBEC-3:

[0399] MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSW SPCFEC*AEQIVRFLATHHNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPV PSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVT ITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKE SWGLQDLVNDFGNLQLGPPMS (서열번호 82)

[0400] (이탤릭체: 핵산 편집 도메인)

[0401] 일부 구현예에서, 아데노신 데아미나제는 아데노신 데아미나제 ADAR(예를 들어, ADAR1 또는 ADAR2)의 전부 또는 일부를 포함할 수 있다. 또 다른 구현예에서, 아데노신 데아미나제는 아데노신 데아미나제 ADAT의 전부 또는 일부를 포함할 수 있다. 일부 구현예에서, 아데노신 데아미나제는 다음 돌연변이 중 하나 이상을 포함하는 대장균(EcTadA) 유래의 ADAT의 전부 또는 일부를 포함할 수 있다: D108N, A106V, D147Y, E155V, L84F, H123Y, I157F, 또는 또 다른 아데노신 데아미나제에서의 상응하는 돌연변이. 아데노신 데아미나제는 임의의 적합한 유기체(예를 들어, 대장균)로부터 유래될 수 있다. 일부 구현예에서, 아데노신 데아미나제는 대장균, *스타필로코커스 아우레우스*, *살모넬라 타이피*, *쉬와넬라 푸트리파시엔스*, *헤모필루스 인플루엔자*, *카울로박터 크레센투스* 또는 *바실러스 서브틸리스*로부터 유래된다. 일부 구현예에서, 아데노신 데아미나제는 대장균으로부터 유래된다. 일부 구현예에서, 아데닌 데아미나제는 본원에 제공된 임의의 돌연변이(예를 들어, ecTadA의 돌연변이)에 상응하는 1개 이상의 돌연변이를 포함하는 자연-발생 아데노신 데아미나제이다. 임의의 상동성 단백질에서 상응하는 잔기는 예를 들어 서열 정렬 및 상동성 잔기 결정에 의해 식별될 수 있다. 이에 따라, 본원에 기술된 임의의 돌연변이(예를 들어, ecTadA에서 식별된 임의의 돌연변이)에 상응하는 임의의 자연-발생 아데노신 데아미나제(예를 들어, ecTadA와 상동성을 가짐)에서 돌연변이가 발생될 수 있다. 특정 구현예에서, TadA는 그 전체가 본원에 참고로 포함된 PCT/US2017/045381호(WO2018/027078호)에 기술된 TadA 중 임의의 하나이다. 돌연변이는 표 3에 나타낸 바와 같이 단일 가닥 DNA 상에서 바람직한 아데노신 데아미나제 활성을 갖는 진화 및 선택 횟수(예를 들어, TadA*7.10 = 7세대 진화로부터 변이체 10)를 통해 식별되었다.

[0402] 표 3.TadA 변이체의 유전자형

[0403]

TadA	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
0.1	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K
0.2	W	R	H	N	P		R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K

[0404]

TadA	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
1.1	W	R	H	N	P		R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K
1.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K
2.1	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.2	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.3	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.4	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.5	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.6	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.7	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.8	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.9	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.10	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.11	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
2.12	W	R	H	N	P		R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K
3.1	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.2	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.4	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.5	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.6	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.7	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
3.8	W	R	H	N	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K
4.1	W	R	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
4.2	W	G	H	N	P		R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K
4.3	W	R	H	N	P		R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
5.1	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.2	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
5.3	W	R	L	N	P		L	N	I	S	V	N	Y	G	A	C	Y	R	V	I	N	K
5.4	W	R	H	S	P		R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T
5.5	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.6	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.7	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.8	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.9	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K

TadA	23	26	36	37	48	49	51	72	84	87	105	108	123	125	142	145	147	152	155	156	157	16
5.10	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.11	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.12	W	R	L	N	P		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
5.13	W	R	H	N	P		L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K
5.14	W	R	H	N	S		L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K
6.1	W	R	H	N	S		L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K
6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K
6.3	W	R	L	N	S		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
6.4	W	R	L	N	S		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
6.5	W	R	L	N	I	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
7.1	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
7.2	W	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
7.3	I	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
7.4	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K
7.5	W	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K
7.6	W	R	L	N	A		L	N	I	S	V	N	Y	G	A	C	Y	P	V	I	N	K
7.7	L	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K
7.8	I	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K
7.9	L	R	L	N	A		L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K
7.10	R	R	L	N	A		L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K

[0405]

[0406] 일부 구현예에서, TadA는 단량체 또는 이량체(예를 들어, 야생형 대장균 TadA의 이종이량체 및 조작된 TadA 변이체)로 제공된다. 일부 구현예에서, 아데노신 데아미나제는 아래 표 4에 나타낸 바와 같은 8세대 TadA* 8 변이체이다.

[0407] 표 4: TadA8* 아데노신 데아미나제 변이체

아데노신 데아미나제	아데노신 데아미나제 설명
TadA*8.1	단량체_TadA*7.10 + Y147T
TadA*8.2	단량체_TadA*7.10 + Y147R
TadA*8.3	단량체_TadA*7.10 + Q154S
TadA*8.4	단량체_TadA*7.10 + Y123H
TadA*8.5	단량체_TadA*7.10 + V82S
TadA*8.6	단량체_TadA*7.10 + T166R
TadA*8.7	단량체_TadA*7.10 + Q154R
TadA*8.8	단량체_TadA*7.10 + Y147R_Q154R_Y123H
TadA*8.9	단량체_TadA*7.10 + Y147R_Q154R_I76Y
TadA*8.10	단량체_TadA*7.10 + Y147R_Q154R_T166R
TadA*8.11	단량체_TadA*7.10 + Y147T_Q154R
TadA*8.12	단량체_TadA*7.10 + Y147T_Q154S
TadA*8.13	단량체_TadA*7.10 + H123H_ Y147R_Q154R_I76Y
TadA*8.14	이종이합체_(WT) + (TadA*7.10 + Y147T)
TadA*8.15	이종이합체_ (WT) + (TadA*7.10 + Y147R)
TadA*8.16	이종이합체_(WT) + (TadA*7.10 + Q154S)
TadA*8.17	이종이합체_(WT) + (TadA*7.10 + Y123H)
TadA*8.18	이종이합체_(WT)+ (TadA*7.10 + V82S)
TadA*8.19	이종이합체_(WT) + (TadA*7.10 + T166R)
TadA*8.20	이종이합체_(WT)+ (TadA*7.10 + Q154R)
TadA*8.21	이종이합체_(WT) + (TadA*7.10 + Y147R_Q154R_Y123H)
TadA*8.22	이종이합체_(WT) + (TadA*7.10 + Y147R_Q154R_I76Y)
TadA*8.23	이종이합체_(WT) + (TadA*7.10 + Y147R_Q154R_T166R)
TadA*8.24	이종이합체_(WT) + (TadA*7.10 + Y147T_Q154R)
TadA*8.25	이종이합체_(WT) + (TadA*7.10 + Y147T_Q154S)
TadA*8.26	이종이합체_(WT) + (TadA*7.10 + H123H_Y147T_Q154R_I76Y)

[0408]

[0409] 일부 구현예에서, 아데노신 데아미나제는 다음으로부터 선택된 아미노산 위치에서의 변경을 함유하는 9세대 TadA*9 변이체이다: 아래 참조 서열에 표시된 TadA 변이체의 21, 23, 25, 38, 51, 54, 70, 71, 72, 72, 94, 124, 133, 138, 139, 146 및 158:

```
        10         20         30         40         50
MSEVEFSHEY WMRHALTLAK RARDEREVPV GAVLVLNNRV IGEGWNRAIG
        60         70         80         90        100
LHDPTAHAEI MALRQGGLVM QNYRLIDATL YVTFEPCVMC AGAMIHSRIG
       110        120        130        140        150
RVVFGVRNAK TGAAGSLMDV LHYPGMNHRV EITEGILADE CAALLCYFFR
       160
MPRQVFNAQK KAQSSTD (서열번호 83)
```

[0410]

[0411] 일 구현예에서, 아데노신 데아미나제 변이체는 다음으로부터 선택된 2개 이상의 아미노산 위치에서 변경을 함유한다: 상기 TadA 참조 서열의 21, 23, 25, 38, 51, 54, 70, 71, 72, 94, 124, 133, 138, 139, 146 및 158. 또 다른 구현예에서, 아데노신 데아미나제 변이체는 다음으로부터 선택된 1개 이상(예를 들어, 2, 3, 4개)의 변경을 함유한다: 서열번호 1의 R21N, R23H, E25F, N38G, L51W, P54C, M70V, Q71M, N72K, Y73S, M94V, P124W, T133K, D139L, D139M, C146R 및 A158K. 다른 구현예에서, 아데노신 데아미나제 변이체는 다음 변경 중 1개 이상을 추가로 함유한다: Y147T, Y147R, Q154S, Y123H 및 Q154R. 또 다른 구현예에서, 아데노신 데아미나제 변이체는 다음으로부터 선택된 상기 TadA 참조 서열에 대한 변경의 조합을 함유한다: E25F + V82S + Y123H, T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R +

Q154R; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; Q71M + V82S + Y123H + Y147R + Q154R; E25F + V82S + Y123H + T133K + Y147R + Q154R; E25F + V82S + Y123H + Y147R + Q154R; V82S + Y123H + P124W + Y147R + Q154R; L51W + V82S + Y123H + C146R + Y147R + Q154R; P54C + V82S + Y123H + Y147R + Q154R; Y73S + V82S + Y123H + Y147R + Q154R; N38G + V82T + Y123H + Y147R + Q154R; R23H + V82S + Y123H + Y147R + Q154R; R21N + V82S + Y123H + Y147R + Q154R; V82S + Y123H + Y147R + Q154R + A158K; N72K + V82S + Y123H + D139L + Y147R + Q154R; E25F + V82S + Y123H + D139M + Y147R + Q154R; M70V + V82S + M94V + Y123H + Y147R + Q154R; Q71M + V82S + Y123H + Y147R + Q154R; E25F + I76Y+ V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D138M + Y147R + Q154R; Y72S + I76Y + V82S + Y123H + Y147R + Q154R; E25F + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82T + Y123H + Y147R + Q154R; N38G + I76Y + V82S + Y123H + Y147R + Q154R; R23H + I76Y + V82S + Y123H + Y147R + Q154R; P54C + I76Y + V82S + Y123H + Y147R + Q154R; R21N + I76Y + V82S + Y123H + Y147R + Q154R; I76Y + V82S + Y123H + D138M + Y147R + Q154R; Y72S + I76Y + V82S + Y123H + Y147R + Q154R; 및 V82S + Q154R; N72K V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; M70V +Q71M +N72K +V82S + Y123H + Y147R + Q154R; N72K V82S + Y123H + Y147R + Q154R; Q71M_V82S + Y123H + Y147R + Q154R; M70V +V82S + M94V + Y123H + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R; V82S + Y123H + T133K + Y147R + Q154R + A158K; 및 M70V +Q71M +N72K +V82S + Y123H + Y147R + Q154R. 일부 구현예에서, 데아미나제 또는 다른 폴리펩티드 서열에는 예를 들어 융합 단백질의 구성요소로서 포함될 때 메티오닌이 결여되어 있다. 이는 위치 번호를 변경시킬 수 있다. 그러나, 당업자는 이러한 상응하는 돌연변이가 동일 돌연변이, 예를 들어 Y73S 및 Y72S 및 D139M 및 D138M을 의미함을 이해할 것이다.

[0412] 일부 구현예에서, Cas9는 SV40 대형 T 항원의 NLS, 뉴클레오플라스민, c-myc, hRNPA1 M9, 임포틴-알파의 IBB 도메인, 근종 T 단백질의 NLS, 인간 p53, c-abl IV, 인플루엔자 바이러스 NS1, 간염 바이러스 델타 항원, 마우스 Mx1, 인간 폴리(ADP-리보스) 폴리머라제, 스테로이드 호르몬 수용체(인간) 글루코코르티코이드를 포함하는 핵 위치 서열에 융합된다.

[0413] 일부 구현예에서, Cas9 단백질은 헤마글루티닌(HA) 태그, 히스티딘(His) 태그, FLAG 태그, Myc 태그, V5 태그, VSV-G 태그, SNAP 태그, 티오레독신(Trx) 태그를 포함하지만 이에 국한되지 않는 에피토프 태그에 융합된다.

[0414] 일부 구현예에서 Cas9는 글루타티온-S-트랜스퍼라제(GST), 서양고추냉이 퍼옥시다제(HRP), 클로람페니콜 트랜스퍼라제(CAT), HcRed, DsRed, 시안색 형광 단백질, 황색 형광 단백질 및 청색 형광 단백질, 증강된 버전 또는 초접힘 녹색 형광 단백질(GFP)를 비롯한 GFP뿐만 아니라 기타 변형된 리포터 유전자 버전을 포함하지만 이에 국한되지 않는 리포터 유전자에 융합된다.

[0415] 일부 구현예에서, 조작된 Cas9 단백질의 혈청 반감기는 이종 단백질, 예컨대 인간 혈청 알부민 단백질, 트랜스페린 단백질, 인간 IgG 및/또는 시알릴화 펩티드, 예컨대 카복시-말단 펩티드(융모성성선자극호르몬 β 사슬의 CTP)와의 융합에 의해 증가된다.

[0416] 일부 구현예에서, 조작된 Cas9 단백질의 혈청 반감기는 제미닌, 유비퀴틴, FKBP12-L106P 및/또는 디히드로폴레이트 환원효소를 포함하지만 이에 국한되지 않는 불안정화 도메인과의 융합에 의해 감소된다.

[0417] 증가된 또는 감소된 안정성을 제공하는 적합한 융합 파트너는 데그론 서열을 포함하지만 이에 국한되지는 않는다. 데그론은 이들이 일부에 해당하는 단백질의 안정성을 제어하는 아미노산 서열임을 당업자라면 쉽게 이해한다. 예를 들어, 데그론 서열을 포함하는 단백질의 안정성은 데그론 서열에 의해 적어도 부분적으로 제어된다. 일부 경우에, 적합한 데그론은 데그론이 실험 제어와 무관하게 단백질 안정성에 영향을 주도록 하는 구성요소이다(즉, 데그론은 약물 유도성, 온도 유도성 등이 아니다) 일부 경우에, 데그론은 변이체 Cas9 폴리펩티드가 원하는 조건에 따라 "켜짐"(즉, 안정함) 또는 "꺼짐(즉, 불안정함, 열화됨)으로 될 수 있도록 변이체 Cas9 폴리펩티드에 조절가능한 안정성을 제공한다. 예를 들어, 데그론이 온도 민감성 데그론인 경우, 변이체 Cas9 폴리펩티드는 역치 온도(예를 들어, 42℃, 41℃, 40℃, 39℃, 38℃, 37℃, 36℃, 35℃, 34℃, 33℃, 32℃, 31℃, 30℃ 등) 미만에서 기능적일 수 있지만(즉, "켜짐", 안정함), 임계 온도를 초과하는 온도에서는 비기능적일 수 있다(예: "꺼짐, 열화됨). 또 다른 예로서, 데그론이 약물 유도성 데그론인 경우, 약물의 존재 또는 부재는 단백질을 "꺼짐(즉, 불안정) 상태에서 "켜짐"(즉, 안정) 상태로 전환할 수 있거나 또는 그 반대일 수 있다. 예시적인

약물 유도성 데그론은 FKBP12 단백질로부터 유래된다. 데그론의 안정성은 데그론에 결합하는 소분자의 존재 또는 부재에 의해 조절된다.

[0418] 적합한 데그론의 예는 Shield-1, DHFR, 옥신 및/또는 온도에 의해 조절되는 데그론을 포함하지만 이에 국한되지는 않는다. 적합한 데그론의 비제한적인 예는 당업계에 알려져 있다(예를 들어, Dohmen 등, Science, 1994. 263(5151): p. 1273-1276: Heat-inducible degron: a method for constructing temperature-sensitive mutants; Schoeber 등, Am J Physiol Renal Physiol. 2009 Jan;296(1):F204-11: Conditional fast expression and function of multimeric TRPV5 channels using Shield-1; Chu 등, Bioorg Med Chem Lett. 2008 Nov 15;18(22):5941-4: Recent progress with FKBP-derived destabilizing domains; Kanemaki, Pflugers Arch. 2012 Dec 28: Frontiers of protein expression control with conditional degrons; Yang 등, Mol Cell. 2012 Nov 30;48(4):487-8: Titivated for destruction: the methyl degron; Barbour 등, Biosci Rep. 2013 Jan 18;33(1).: Characterization of the bipartite degron that regulates ubiquitin-independent degradation of thymidylate synthase; 및 Greussing 등, J Vis Exp. 2012 Nov 10;(69): Monitoring of ubiquitin-proteasome activity in living cells using a Degron (dgn)-destabilized green fluorescent protein (GFP)-based reporter protein; 이들 모두는 전체적으로 본원에 참고로 통합된다).

[0419] 예시적인 데그론 서열은 세포와 동물 모두에서 잘 특징규명되었고 테스트되었다. 따라서 죽은 Cas9를 데그론 서열에 융합시키면 "조정 가능"하고 "유도성"인 죽은 Cas9 폴리펩티드를 생성시킨다.

[0420] 본원에 기술된 임의의 융합 파트너는 임의의 바람직한 조합으로 사용될 수 있다. 이 점을 설명하기 위한 하나의 비제한적 예로서, Cas9 융합 단백질은 검출을 위한 YFP 서열, 안정성을 위한 데그론 서열 및 표적 DNA의 전사를 증가시키기 위한 전사 활성화인자 서열을 포함할 수 있다. 또한 dCas9 융합 단백질에서 사용될 수 있는 융합 파트너의 수는 제한되지 않는다. 일부 경우에, Cas9 융합 단백질은 1개 이상(예를 들어 2개 이상, 3개 이상, 4개 이상 또는 5개 이상)의 이종 서열을 포함한다.

[0421] **표적 핵산**

[0422] 표적 핵산은 단일 가닥, 이중 가닥 또는 다중 가닥의 DNA 또는 RNA인 DNA 분자, RNA 분자, 게놈 DNA, cDNA, DNA-RNA 혼성, 또는 퓨린 및 피리미딘 염기 또는 기타 천연, 화학적으로 또는 생화학적으로 변형된, 비천연 또는 유도체화된 뉴클레오티드 염기인 데옥시리보뉴클레오티드, 리보뉴클레오티드 또는 이들의 유사체를 포함하는 중합체이다. 표적 핵산은 3차원 구조를 가질 수 있고, 암호화 또는 비암호화 영역을 포함할 수 있으며, 엑손, 인트론, mRNA, tRNA, rRNA, siRNA, shRNA, miRNA, 리보자임, cDNA, 플라스미드, 벡터, 외인성 서열, 내인성 서열을 포함할 수 있다. 표적 핵산은 변형된 뉴클레오티드를 포함할 수 있고, 메틸화 뉴클레오티드 또는 뉴클레오티드 유사체를 포함할 수 있다. 일부 구현예에서, 표적 핵산은 비핵산 구성요소와 함께 산재될 수 있다.

[0423] 표적 핵산은 CRISPR-Cas9 시스템에 의해 인식되고, Cas9에 결합한다. 일부 구현예에서, 이는 Cas9의 결합으로 인해 변형되거나, 절단되거나, 변경 발현을 가진다. 표적 핵산은 특이적인 인식가능한 PAM 모티프, 예를 들어 5'-NGG-3', 5'-NGC-3', 5'-NAGHC-3', 5'-NRHRRH-3' 또는 5'-NNAAA-3'(H=A, C 또는 T; R=A 또는 G)를 함유한다.

[0424] **재조합 유전자 기술**

[0425] 본 개시내용에 따르면, 당업계 기술 내에 이용된 통상적인 분자 생물학, 미생물학 및 재조합 DNA 기술이 있을 수 있다. 이러한 기술은 문헌에 기술되어 있다(예를 들어 Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; DNA Cloning: A Practical Approach, Volumes I and II (D. N. Glover ed. 1985); Oligonucleotide Synthesis (M. J. Gait ed. 1984); Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. (1985)); Transcription And Translation (B. D. Hames & S. J. Higgins, eds. (1984)); Animal Cell Culture (R. I. Freshney, ed. (1986)); Immobilized Cells and Enzymes (IRL Press, (1986)); B. Perbal, A Practical Guide To Molecular Cloning (1984); F. M. Ausubel 등 (eds.), Current Protocols in Molecular Biology, John Wiley & Sons, Inc. (1994) 참조.

[0426] 본원에 기술된 조작된 Cas9 효소와 같이 폴리펩티드를 암호화하는 핵산과 같은 유전자의 재조합 발현은 폴리펩티드를 암호화하는 핵산을 함유하는 발현 벡터의 작제를 포함할 수 있다. 일단 폴리뉴클레오티드가 얻어지면, 폴리펩티드 생성을 위한 벡터는 당업계에 알려진 기술을 사용하여 재조합 DNA 기술에 의해 생성될 수 있다. 폴리펩티드 암호화 서열 및 적절한 전사 및 번역 조절 신호를 함유하는 발현 벡터를 작제하기 위해 공지 방법이 사용될 수 있다. 이러한 방법은 예를 들어 시험관내 재조합 DNA 기술, 합성 기술 및 생체내 유전자 재조합을 포

함한다.

[0427] 발현 벡터는 통상적인 기술에 의해 숙주 세포로 전달될 수 있고, 그 다음, 형질감염된 세포는 통상적인 기술에 의해 배양되어 폴리펩티드를 생성할 수 있다.

[0428] 일부 구현예에서, DNA-표적 RNA 및/또는 Cas9 단백질을 암호화하는 뉴클레오티드 서열은 조절 요소, 예를 들어 프로모터와 같은 전사 조절 요소에 작동가능하게 연결된다. 전사 조절 요소는 진핵 세포, 예를 들어 포유동물 세포; 또는 원핵 세포(예를 들어, 박테리아 또는 고세균 세포)에서 기능적일 수 있다. 일부 구현예에서, 진핵 세포는 인간 세포이다. 일부 구현예에서, DNA-표적 RNA 및/또는 신규 Cas9 단백질을 암호화하는 뉴클레오티드 서열은 원핵 세포 및 진핵 세포 모두에서 암호화된 뉴클레오티드 서열의 발현을 허용하는 다수의 조절 요소에 작동가능하게 연결된다.

[0429] 프로모터는 구성요소적으로 활성인 프로모터(즉, 활성/"켜짐" 상태에서 구성요소인 프로모터)일 수 있고, 유도성 프로모터(즉, 외부 자극, 예를 들어 특정 온도, 화합물 또는 단백질의 존재에 의해 조절되는 활성/"켜짐" 또는 불활성/"꺼짐" 상태를 갖는 프로모터)일 수 있고, 공간적으로 제한된 프로모터(즉, 전사 조절 요소, 인핸서 등)(예: 조직 특이적 프로모터, 세포 유형 특이적 프로모터 등)일 수 있고, 일시적으로 제한된 프로모터일 수 있다(즉, 프로모터는 배아 발달의 특정 단계 동안 또는 생물학적 프로세스의 특정 단계, 예를 들어 마우스의 모낭 주기 동안 "켜짐" 상태 또는 "꺼짐" 상태에 있다).

[0430] 적합한 프로모터는 바이러스로부터 유래될 수 있고 따라서 바이러스 프로모터로 지칭될 수 있거나, 원핵생물 또는 진핵생물 유기체를 포함하는 임의의 유기체로부터 유래될 수 있다. 임의의 RNA 폴리머라제(예를 들어, pol I, pol II, pol III)에 의한 발현을 구동시키기 위해 적합한 프로모터가 사용될 수 있다. 예시적인 프로모터는 SV40 초기 프로모터, 마우스 유방 종양 바이러스 긴 말단 반복(LTR) 프로모터; 아데노바이러스 주요 후기 프로모터(Ad MLP); 단순 포진 바이러스(HSV) 프로모터, 거대세포바이러스(CMV) 프로모터, 예컨대 CMV 즉시 초기(immediate early) 프로모터 영역(CMVIE), 육종 바이러스(RSV) 프로모터, 인간 U6 소형 핵 프로모터(U6)(Miyagishi 등, Nature Biotechnology 20, 497-500 (2002)), 증강된 U6 프로모터(예를 들어, Xia 등, Nucleic Acids Res. 2003 Sep 1,31(17)) 및/또는 인간 H1 프로모터(H1)를 포함하지만, 이에 국한되지 않는다.

[0431] 유도성 프로모터의 예는 T7 RNA 폴리머라제 프로모터, T3 RNA 폴리머라제 프로모터, 이소프로필-베타-D-티오갈락토피라노시드(IPTG)-조절된 프로모터, 락토스 유도된 프로모터, 열 충격 프로모터, 테트라사이클린-조절된 프로모터(예: Tet-켜짐, Tet-꺼짐 등), 스테로이드-조절된 프로모터, 금속-조절된 프로모터, 에스트로겐 수용체-조절된 프로모터 등을 포함하지만 이에 국한되지 않는다. 따라서 유도성 프로모터는 독시사이클린, RNA 폴리머라제, 예를 들어 T7 RNA 폴리머라제, 에스트로겐 수용체 및/또는 에스트로겐 수용체 융합을 포함하지만 이에 국한되지 않는 분자에 의해 조절될 수 있다.

[0432] 일부 구현예에서, 프로모터는 공간적으로 제한된 프로모터(즉, 세포 유형 특이적 프로모터, 조직 특이적 프로모터 등)이어서, 다세포 유기체에서 프로모터는 특정 세포의 하위세트에서 활성(즉, "켜짐")이다. 또한 공간적으로 제한된 프로모터는 인핸서, 전사 조절 요소, 조절 서열 등으로도 지칭될 수 있다. 임의의 편의적인 공간적으로 제한된 프로모터가 사용될 수 있고, 적합한 프로모터(예를 들어, 뇌 특이적 프로모터, 뉴런 하위세트에서 발현을 구동시키는 프로모터, 생식계열에서 발현을 구동시키는 프로모터, 폐에서 발현을 구동시키는 프로모터, 근육에서 발현을 구동시키는 프로모터, 췌장 섬 세포에서 발현을 구동시키는 프로모터 등)의 선택은 유기체에 따라 달라진다. 따라서 공간적으로 제한된 프로모터는 유기체에 따라 광범위한 상이한 조직 및 세포 유형에서 대상체 부위-특이적 폴리펩티드를 암호화하는 핵산의 발현을 조절하기 위해 사용될 수 있다. 또한 일부 공간적으로 제한된 프로모터는 배아 발달의 특정 단계 동안 또는 생물학적 프로세스의 특정 단계(예를 들어 모낭 주기) 동안 프로모터가 "켜짐" 상태 또는 "꺼짐" 상태에 있도록 일시적으로 제한된다.

[0433] 예시 목적으로, 공간적으로 제한된 프로모터의 예는 뉴런-특이적 프로모터, 지방세포-특이적 프로모터, 심근세포-특이적 프로모터, 평활근-특이적 프로모터, 광수용체-특이적 프로모터 등을 포함하지만 이제 국한되지는 않는다. 뉴런-특이적이고 공간적으로 제한된 프로모터는 뉴런-특이적 에놀라제(NSE) 프로모터, 방향족 아미노산 데카복실라제(AADC) 프로모터, 신경필라멘트 프로모터, 시냅신 프로모터, thy-1 프로모터, 세로토닌 수용체 프로모터, 티로신 하이드록실라제 프로모터(TH), GnRH 프로모터, L7 프로모터, DNMT 프로모터, 엔케팔린 프로모터, 미엘린 염기성 단백질(MBP) 프로모터, Ca^{2+}-칼모듈린-의존성 단백질 키나제 II-알파(CamKIIa) 프로모터 및/또는 CMV 인핸서/혈소판-유래된 성장 인자-β 프로모터를 포함하지만 이제 국한되지는 않는다.

[0434] 지방세포-특이적이고 공간적으로 제한된 프로모터는 aP2 유전자 프로모터/인핸서, 예를 들어 인간 aP2 유전자의

-5.4kb 내지 +21bp 영역, 글루코스 수송체-4(GLUT4) 프로모터, 지방산 트랜스로카제(FAT/CD36) 프로모터, 스테아로일-CoA 불포화효소-1(SCD1) 프로모터, 렙틴 프로모터 및 아디포넥틴 프로모터, 아디프신 프로모터 및/또는 레지스틴 프로모터를 포함하지만, 이에 국한되지는 않는다.

[0435] 심근세포-특이적이고 공간적으로 제한된 프로모터는 다음 유전자로부터 유래된 조절 서열을 포함하지만 이에 국한되지는 않는다: 미오신 경쇄-2, a-미오신 중쇄, AE3, 심장 트로포닌 C 및/또는 심장 액틴.

[0436] 평활근-특이적이고 공간적으로 제한된 프로모터는 SM22a 프로모터, 스무델린 프로모터 및/또는 a-평활근 액틴 프로모터를 포함하지만 이에 국한되지는 않는다.

[0437] 광수용체-특이적이고 공간적으로 제한된 프로모터는 로돕신 프로모터, 로돕신 키나제 프로모터, 베타 포스포디에스테라제 유전자 프로모터, 색소성 망막염 유전자 프로모터, 광수용기간 레티노이드-결합 단백질(IRBP) 유전자 인핸서 및/또는 IRBP 유전자 프로모터를 포함하지만, 이에 국한되지는 않는다.

[0438] **CRISPR-Cas9의 유전자 편집 사용**

[0439] 본원에 기술된 CRISPR-Cas9 시스템은 유전자 침묵 사건 또는 원하는 표적 유전자 발현의 발현 변경(예를 들면, 증가 또는 감소)을 초래할 수 있는 유전자 편집을 위해 사용될 수 있다. 따라서, 일부 구현예에서, 본원에 기술된 CRISPR-Cas9 시스템은 표적 핵산의 발현을 변경하는 방법에서 사용된다. 일부 구현예에서, 본원에 기술된 CRISPR-Cas9 시스템은 원하는 표적 세포에서 표적 핵산을 변형하는 방법에서 사용된다. 일부 구현예에서, 본 발명은 유전자 발현에서 원하는 변형을 유발하기 위한 진핵 세포 내 표적 핵산의 부위-특이적 변형 방법을 제공한다.

[0440] 일부 구현예에서, 본 발명은 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산; 및 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 80% 서열 동일성을 갖는 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함하는 조작된 비천연 발생 CRISPR-Cas 시스템을 제공하고, 상기 RNA 가이드는 직접 반복 서열, 및 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열에서 파손을 유발할 수 있다.

[0441] 일부 구현예에서, 본 발명은 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산; 및 서열번호 1, 4, 8, 14, 84 또는 86과 적어도 80% 서열 동일성을 갖는 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함하는 조작된 비천연 발생 CRISPR-Cas 시스템을 제공하고, 상기 RNA 가이드는 직접 반복 서열, 및 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas 단백질은 데아미나제에 융합하고, 상기 Cas 단백질 융합은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있다.

[0442] 일부 구현예에서, 코돈-최적화된 CRISPR-연관(Cas) 단백질을 포함하고 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함하는 조작된 비천연 발생 CRISPR-Cas 시스템이 본원에 제공된다.

[0443] 일부 구현예에서, 핵 위치 서열(NLS) 및/또는 FLAG, HIS 또는 HA 태그를 추가로 포함하는 조작된 비천연 발생 Cas9 융합 단백질이 본원에 제공된다.

[0444] 일부 구현예에서, 서열번호 2, 5, 9, 15, 85, 87, 95 또는 96과 적어도 80% 동일성을 갖는 조작된 비천연 발생 Cas9 융합 단백질이 본원에 제공된다.

[0445] 일부 구현예에서, 본 발명은 세포를, 본원에 기술된 Cas9 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하는, 진핵 세포 내 표적 핵산의 발현을 변경하는 방법으로, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열에서 파손을 유발할 수 있는 방법을 제공한다.

[0446] 일부 구현예에서, 본 발명은 세포를, 본원에 기술된 Cas9 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하는, 진핵 세포 내 표적 핵산의 발현을 변경하는 방법으로, 상기 RNA 가이드는 직접 반복 서열, 및 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas9 단백질은 상기 RNA 가이드에 결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있는 방법을 제공한다.

[0447] 일부 구현예에서, 본 발명은 세포를, 본원에 기술된 Cas9 및 RNA 가이드 또는 상기 RNA 가이드를 암호화하는 핵산과 접촉시키는 것을 포함하는, 진핵 세포 내 표적 핵산을 변형하는 방법으로, 상기 RNA 가이드는 직접 반복 서열, 및 상기 표적 핵산에 혼성화할 수 있는 스페이서 서열을 포함하고, 상기 Cas 단백질은 상기 RNA 가이드에

결합할 수 있고, 상기 RNA 가이드에 상보적인 표적 핵산 서열을 편집할 수 있는 방법을 제공한다.

[0448] 따라서, 일부 구현예에서, Cas 단백질은 서열번호 1, 4, 8, 14, 84 또는 86과 약 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 동일성을 가진다. 일부 구현예에서, Cas 단백질은 서열번호 1, 4, 8, 14, 84 또는 86과 동일하다.

[0449] CRISPR-Cas9 시스템 및 방법에서 사용하기에 적합한 가이드 RNA, Cas9 돌연변이 및 융합 단백질은 본 개시내용 전반에 걸쳐 기술된 바와 같다.

[0450] 일 측면에서, 방법은 CRISPR-Cas9를 표적 핵산에 결합시키고 표적 핵산 절단을 달성하는 것을 포함한다. 일부 구현예에서, CRISPR-Cas9 시스템은 이중 가닥의 파손을 도입함으로써 표적 DNA 또는 RNA 이중구조를 절단한다. 일부 구현예에서 CRISPR-Cas9 시스템은 단일 가닥의 파손 또는 틈(nick)을 도입함으로써 표적 DNA 또는 RNA를 절단한다.

[0451] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 부위-특이적 방식으로 표적 DNA를 변형하는 이펙터를 갖는 융합 단백질을 포함하며, 변형 활성에는 메틸트랜스퍼라제 활성, 데메틸라제 활성, 아세틸트랜스퍼라제 활성, 데아세틸라제 활성, 키나제 활성, 포스파타제 활성, 유비퀴틴 리가제 활성, 탈유비퀴틴화 활성, 아데닐화 활성, 탈아데닐화 활성, SUMO화 활성, 탈SUMO화 활성, 리보실화 활성, 탈리보실화 활성, 미리스토일화 활성, 탈미리스토일화 활성, 인테그라제 활성, 트랜스포사제 활성, 재조합효소 활성, 폴리머라제 활성, 리가제 활성, 헬리카제 활성 또는 뉴클레아제 활성이 포함되고, 이들 중 임의의 것은 DNA 또는 DNA-연관 폴리펩티드(예를 들어, 히스톤 또는 DNA 결합 단백질)를 변형시킬 수 있다.

[0452] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 아데노신 또는 시토신 염기를 변형시킬 수 있고 부위-특이적 염기 편집기로 기능할 수 있는 데아미나제 효소를 포함하여 뉴클레오티드 염기를 화학적으로 변형시킴으로써 DNA 서열을 편집할 수 있는 효소를 갖는 융합 단백질을 포함한다. 예를 들어, 일반적으로 RNA를 기질로 사용하는 APOBEC1 시티딘 데아미나제는 Cas9에 융합될 때 단일 가닥 및 이중 가닥의 DNA에 표적이 되어 시티딘을 유리딘으로 직접 전환시킬 수 있고, ADAR 효소는 아데노신을 이노신으로 탈아미노화한다. 따라서 데아미나제를 사용하는 '염기 편집'은 하나의 표적 DNA 염기를 다른 염기로 프로그래밍 가능한 전환을 가능하게 한다. 다양한 염기 편집기가 당업계에 알려져 있고, 본원에 기술된 방법 및 시스템에서 사용될 수 있다. 예시적인 염기 편집기는 예를 들어 Rees and Liu *Nature Review Genetics*, 2018, 19(12): 770-788에 기술되어 있고, 그 내용이 본원에 통합된다. 따라서, 일부 구현예에서, 본원에 기술된 Cas9 효소(ScoCas9, SirCas9, VapCas9, EpeCas9, LfeCas9, PmaCas9)는 핵염기 편집기의 구성요소이다. 일부 구현예에서, 염기 편집기는 아데닌 데아미나제 TadA8 또는 TadA9이다.

[0453] 일부 구현예에서, 염기 편집은 침묵 유전자에 정지 코돈 도입을 초래한다. 일부 구현예에서, 염기 편집은 아미노산 서열 변경에 의해 단백질 기능 변경을 초래한다.

[0454] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 히스톤과의 융합에 의한 표적 DNA의 후성적 변형을 포함한다. 일부 구현예에서, CRISPR-Cas9 시스템은 판독기, 기록기 또는 지우개(eraser) 단백질과 같은 후성적 변형 효소와의 융합에 의한 표적 DNA의 후성적 변형을 포함한다. 일부 구현예에서, CRISPR-Cas9 시스템은 표적 DNA의 선택된 영역에서 히스톤 변형 패턴을 변경시키는 히스톤 변형 효소와의 융합을 포함한다. 히스톤 변형은 메틸화, 아세틸화, 유비퀴틴화, 인산화를 포함하는 많은 상이한 방식으로 그리고 많은 상이한 조합으로 발생하여 DNA의 구조적 변화를 초래할 수 있다. 일부 구현예에서, 히스톤 변형은 전사 억제 또는 활성화를 초래한다.

[0455] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 전사 활성화인자 단백질 또는 전사 억제인자 단백질, 소분자/약물-반응성 전사 조절인자, 유도성 전사 조절인자와의 융합을 통해 전사를 증가시키거나 감소시킴으로써 표적 DNA의 전사를 조절한다. 일부 구현예에서, CRISPR-Cas9 시스템은 결합으로 유전자 발현의 증가 또는 감소를 초래하는 표적 암호화 mRNA(즉, 단백질 암호화 유전자)의 발현을 조절하기 위해 사용된다.

[0456] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 프로모터 또는 인핸서와 같은 유전자 조절 요소를 편집함으로써 유전자 조절을 조절하기 위해 사용된다.

[0457] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 tRNA, rRNA, snoRNA, siRNA, miRNA 및 긴 ncRNA를 포함하는 표적 비암호화 RNA의 발현을 조절하기 위해 사용된다.

[0458] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 염색질 루프 구조의 표적화된 조작을 위해 사용된다. 조절 게놈 영역 사이의 염색질 루프의 표적화된 조작은, 내인성 염색질 구조를 조종하고 새로운 인핸서-프로모터 연

결 형성이 유전적 결함을 극복하게 하거나 비정상적인 인핸서-프로모터 연결을 억제하도록 하는 수단을 제공한다.

[0459] 일부 구현예에서, CRISPR-Cas9는 살아있는 세포 영상화를 위해 사용된다. 형광 표지된 Cas9는 동원체 및 텔로미어와 같은 반복적인 게놈 영역을 표적으로 하여 세포 주기 전반에 걸쳐 천연 염색질 유전자좌를 추적하고, 3D 핵 공간에서 전사적 활성 및 불활성 영역의 차등적 위치를 결정한다.

[0460] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 유익한 임상 변이체 또는 억제 돌연변이의 삽입에 의한 병원성 돌연변이 교정을 위해 사용된다.

[0461] **핵염기 편집기**

[0462] Cas9를 포함하는 폴리뉴클레오티드의 표적 뉴클레오티드 서열을 편집, 변형 또는 변경하기 위한 신규 염기 편집기 또는 핵염기 편집기가 본원에 개시된다. 폴리뉴클레오티드 프로그래밍 가능한 뉴클레오티드 결합 도메인(예를 들어, Cas9) 및 핵염기 편집 도메인(예를 들어, 아데노신 데아미나제)을 포함하는 핵염기 편집기 또는 염기 편집기가 본원에 기술된다. 폴리뉴클레오티드 프로그래밍 가능한 뉴클레오티드 결합 도메인(예를 들어, Cas9)은 결합된 가이드 폴리뉴클레오티드(예를 들어, gRNA)와 함께 있을 때 표적 폴리뉴클레오티드 서열에 특이적으로 결합할 수 있고(즉, 결합된 가이드 핵산의 염기와 표적 폴리뉴클레오티드 서열의 염기 사이의 상보적 염기 쌍 형성을 통해), 그에 의해 편집하고자 하는 표적 핵산 서열에 염기 편집기를 위치시킬 수 있다. 일부 구현예에서, 표적 폴리뉴클레오티드 서열은 단일 가닥의 DNA 또는 이중 가닥의 DNA를 포함한다. 일부 구현예에서, 표적 폴리뉴클레오티드 서열은 RNA를 포함한다. 일부 구현예에서, 표적 폴리뉴클레오티드 서열은 DNA-RNA 혼성을 포함한다. 인간 질병과 연관하여 알려진 유전자 변이의 대부분은 점 돌연변이이므로, 정확한 점 돌연변이를 보다 효율적이고 깔끔하게 만들 수 있는 방법이 필요하다. 본원에 제공된 염기 편집 시스템은 이중 가닥 DNA 파손을 발생시키지 않고 공여체 DNA 주형을 요구하지 않고 과도한 확률론적 삽입 및 결실을 유도하지 않으면서 게놈 편집을 제공하는 신규 방법을 제공한다.

[0463] 본원에 제공된 염기 편집기는 유의한 비율의 삽입결실을 발생시키지 않으면서 특정 뉴클레오티드 염기를 변형시킬 수 있다. 본원에서 사용될 때 용어 "삽입결실(들)"은 핵산 내 뉴클레오티드 염기의 삽입 또는 결실을 의미한다. 이러한 삽입 또는 결실은 유전자의 암호화 영역 내의 프레임 이동 돌연변이를 이끌 수 있다. 일부 구현예에서, 표적 뉴클레오티드 서열에서 다수의 삽입 또는 결실(즉, 삽입결실)을 발생시키지 않으면서 핵산 내 특정 뉴클레오티드를 효율적으로 변형(예를 들어, 돌연변이 또는 탈아미노화)시키는 염기 편집기를 발생시키는 것이 바람직하다. 특정 구현예에서, 본원에 제공된 임의의 염기 편집기는 삽입결실 대비 의도한 변형(예를 들어 점 돌연변이 또는 탈아미노화) 비율을 더 크게 생성시킬 수 있다.

[0464] 일부 구현예에서, 본원에 제공된 임의의 염기 편집기 시스템은 표적 폴리뉴클레오티드 서열에서 50% 미만, 40% 미만, 30% 미만, 20% 미만, 19% 미만, 18% 미만, 17% 미만, 16% 미만, 15% 미만, 14% 미만, 13% 미만, 12% 미만, 11% 미만, 10% 미만, 9% 미만, 8% 미만, 7% 미만, 6% 미만, 5% 미만, 4% 미만, 3% 미만, 2% 미만, 1% 미만, 0.9% 미만, 0.8% 미만, 0.7% 미만, 0.6% 미만, 0.5% 미만, 0.4% 미만, 0.3% 미만, 0.2% 미만, 0.1% 미만, 0.09% 미만, 0.08% 미만, 0.07% 미만, 0.06% 미만, 0.05% 미만, 0.04% 미만, 0.03% 미만, 0.02% 미만 또는 0.01% 미만의 삽입결실 형성을 초래한다.

[0465] 본 개시내용의 일부 측면은 본원에 제공된 임의의 염기 편집기가 의도하지 않은 점 돌연변이와 같은 상당 수의 의도하지 않은 돌연변이를 발생시키는 것없이 핵산(예를 들어 대상체의 게놈 내 핵산)에서 점 돌연변이와 같은 의도된 돌연변이를 효율적으로 발생시킬 수 있다는 인식에 기반한다. 일부 구현예에서, 본원에 제공된 임의의 염기 편집기는 의도된 돌연변이를 적어도 0.01%(즉, 적어도 0.01% 염기 편집 효율) 발생시킬 수 있다. 일부 구현예에서, 본원에 제공된 임의의 염기 편집기는 의도된 돌연변이를 적어도 0.01%, 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 60%, 70%, 80%, 90%, 95% 또는 99%를 발생시킬 수 있다.

[0466] 일부 구현예에서, 본원에 제공된 염기 편집기는 1:1보다 큰, 의도된 점 돌연변이 대 삽입결실의 비율을 발생시킬 수 있다. 일부 구현예에서, 본원에 제공된 염기 편집기는 적어도 1.5:1, 적어도 2:1, 적어도 2.5:1, 적어도 3:1, 적어도 3.5:1, 적어도 4:1, 적어도 4.5:1, 적어도 5:1, 적어도 5.5:1, 적어도 6:1, 적어도 6.5:1, 적어도 7:1, 적어도 7.5:1, 적어도 8:1, 적어도 8.5:1, 적어도 9:1, 적어도 10:1, 적어도 11:1, 적어도 12:1, 적어도 13:1, 적어도 14:1, 적어도 15:1, 적어도 20:1, 적어도 25:1, 적어도 30:1, 적어도 40:1, 적어도 50:1, 적어도 100:1, 적어도 200:1, 적어도 300:1, 적어도 400:1, 적어도 500:1, 적어도 600:1, 적어도 700:1, 적어도 800:1, 적어도 900:1 또는 적어도 1000:1 이상의 의도된 점 돌연변이 대 삽입결실의 비율을 발생시킬 수 있다.

[0467] 의도된 돌연변이 및 삽입결실의 수는 예를 들어 국제 PCT 출원 제PCT/2017/045381호(WO2018/027078호) 및 PCT/US2016/058344호(WO2017/070632호); Komor, A.C., 등, "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage" Nature 533, 420-424 (2016); Gaudelli, N.M., 등, "Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage" Nature 551, 464-471 (2017); 및 Komor, A.C., 등, "Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity" Science Advances 3:eaao4774 (2017)에 기술된 바와 같은 임의의 적합한 방법을 사용하여 결정될 수 있으며, 이들 문헌의 전체 내용은 여기서 참고로 통합된다.

[0468] 일부 구현예에서, 삽입결실 빈도를 계산하기 위해, 삽입결실이 발생할 수 있는 양쪽 허용범위의 측면에 위치한 2개의 10-bp 서열의 정확한 매치에 대하여 서열분석 판독을 스캐닝한다. 정확한 매치가 없으면 이 판독은 분석에서 제외된다. 이 삽입결실 허용범위의 길이가 참조 서열과 정확히 매치되는 경우 판독은 삽입결실을 함유하지 않은 것으로 분류된다. 삽입결실 허용범위가 참조 서열보다 2개 초과 염기만큼 길거나 짧은 경우 이 염기서열 판독은 각각 삽입 또는 결실로 분류된다. 일부 구현예에서, 본원에 제공된 염기 편집기는 핵산 영역에서 삽입결실 형성을 제한할 수 있다. 일부 구현예에서, 상기 영역은 염기 편집기에 의해 표적화된 뉴클레오티드 또는 염기 편집기에 의해 표적화된 뉴클레오티드의 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 뉴클레오티드 내 영역이다.

[0469] 표적 뉴클레오티드 영역에서 형성된 삽입결실의 수는 핵산(예를 들어, 세포 게놈 내 핵산)이 염기 편집기에 노출되는 시간의 양에 따라 달라질 수 있다. 일부 구현예에서, 삽입결실의 수 또는 비율은 표적 뉴클레오티드 서열(예를 들어, 세포 게놈 내 핵산)을 염기 편집기에 노출시킨 후 적어도 1시간, 적어도 2시간, 적어도 6시간, 적어도 12시간, 적어도 24시간, 적어도 36시간, 적어도 48시간, 적어도 3일, 적어도 4일, 적어도 5일, 적어도 7일, 적어도 10일 또는 적어도 14일에 결정된다. 본원에 기술된 염기 편집기의 특징은 임의의 융합 단백질, 또는 본원에 제공된 융합 단백질을 사용하는 방법에 적용될 수 있음을 이해해야 한다.

[0470] **치료적 적용분야**

[0471] 본원에 기술된 CRISPR-Cas9 방법 또는 시스템은 다양한 치료적 적용분야를 가질 수 있다. 따라서, 일부 구현예에서, 장애 또는 질병 치료를 필요로 하는 대상체에서 장애 또는 질병을 치료하는 방법이 제공되고, 상기 방법은 본원에 기술된 바와 같은 Cas9를 포함하는 CRISPR-Cas9 시스템을 대상체에게 투여하는 것을 포함하고, 상기 가이드 RNA는 병태 또는 질병과 연관된 표적 핵산의 적어도 10개의 뉴클레오티드에 상보적이고; 상기 Cas 단백질은 상기 가이드 RNA와 연관되고; 상기 가이드 RNA는 표적 핵산에 결합하고; 상기 Cas 단백질은 표적 핵산의 파손을 유발하고, 선택적으로 상기 Cas9는 데아미나제에 융합된 불활성 Cas9(dCas9)이고 표적 핵산에서 1개 이상의 염기 편집을 초래하여 상기 장애 또는 질병을 치료한다.

[0472] 일부 구현예에서, CRISPR-Cas9 방법 또는 시스템은 다양한 질병 및 장애, 예를 들어 유전 장애(예를 들어 단일 유전 질병), 뉴클레아제 활성에 의해 치료될 수 있는 질병 및 다양한 암 등을 치료하기 위해 사용될 수 있다.

[0473] 일부 구현예에서, 본원에 기술된 CRISPR 방법 또는 시스템은 (예를 들어, 하나 이상의 핵산 잔기를 삽입, 결실 또는 돌연변이시킴으로써) 표적 핵산을 변형시키기 위해 표적 핵산을 편집하는 데 사용될 수 있다. 예를 들어, 일부 구현예에서 본원에 기술된 CRISPR 시스템은 바람직한 핵산 서열을 포함하는 외인성 공여체 주형 핵산(예를 들어, DNA 분자 또는 RNA 분자)을 포함한다. 본원에 기술된 CRISPR 시스템으로 유도된 절단 사건이 해결될 때 세포의 분자 기구는 절단 사건의 복구 및/또는 해결에서 외인성 공여체 주형 핵산을 활용한다. 대안적으로, 세포의 분자 기구는 절단 사건의 복구 및/또는 해결에서 내인성 주형을 활용할 수 있다. 일부 구현예에서, 본원에 기술된 CRISPR 시스템은 표적 핵산을 변경시키기 위해 사용되어 삽입, 결실 및/또는 점 돌연변이를 초래할 수 있다. 일부 구현예에서, 삽입은 무흔적 삽입(즉, 절단 사건이 해결될 때 표적 핵산 내로 의도된 핵산 서열 삽입으로 인해 추가의 의도하지 않은 핵산 서열이 없음)이다. 공여체 주형 핵산은 이중 가닥 또는 단일 가닥의 핵산 분자(예를 들어, DNA 또는 RNA)일 수 있다. 일부 구현예에서, 본원에 기술된 CRISPR 방법 또는 시스템은 핵염기 편집기를 포함한다. 예를 들어, 일부 구현예에서, 본원에 기술된 Cas9 단백질은 핵염기 편집 활성을 갖는 폴리펩티드에 융합된다.

[0474] 일 측면에서, 본원에 기술된 CRISPR 방법 또는 시스템은 RNA, 독성 RNA 및/또는 돌연변이된 RNA(예를 들어, 스플라이싱 결함 또는 첨두절단)의 과발현으로 유발된 질병을 치료하기 위해 사용될 수 있다.

[0475] 일부 구현예에서, 본원에 기술된 CRISPR 방법 또는 시스템은 또한 다양한 질병을 유발하는 RNA-의존적 기능에 영향을 주는 트랜스-작용(trans-acting) 돌연변이를 표적으로 할 수 있다.

[0476] 일부 구현예에서, 본원에 기술된 CRISPR 방법 또는 시스템은 또한 스플라이싱 결함 및 질병을 유발할 수 있는 시스-작용(cis-acting) 스플라이싱 코드를 파손시키는 돌연변이를 표적으로 하기 위해 사용될 수 있다.

[0477] 본원에 기술된 CRISPR 방법 또는 시스템은 특히 RNA 바이러스에 대항하는 항바이러스 활성을 위해 추가로 사용될 수 있다. CRISPR-연관 단백질은 바이러스 RNA 서열을 표적으로 하기 위해 선택된 적합한 RNA 가이드를 사용하여 바이러스 RNA를 표적으로 할 수 있다.

[0478] 본원에 기술된 CRISPR 방법 또는 시스템은 또한 대상체(예를 들어, 인간 대상체)에서 암을 치료하기 위해 사용될 수 있다. 예를 들어, 본원에 기술된 CRISPR-연관 단백질은 비정상적인(예를 들어, 점 돌연변이를 포함하거나 선택적 스플라이싱된) RNA 분자를 표적으로 하는 crRNA에 의해 프로그래밍될 수 있고, 암세포에서 발견되어 암세포에서 세포사를 유도할 수 있다(예를 들어, 아포톱시스를 통해).

[0479] 또한, 본원에 기술된 CRISPR 방법 또는 시스템은 대상체의 감염성 질병을 치료하기 위해 사용될 수 있다. 예를 들어, 본원에 기술된 CRISPR-연관 단백질은 감염원 세포에서 세포사를 표적으로 하고 유도하기 위해 감염원(예: 박테리아, 바이러스, 기생충 또는 원생동물)에 의해 발현되는 RNA 분자를 표적으로 하는 crRNA에 의해 프로그래밍될 수 있다. CRISPR 시스템은 세포내 감염원이 숙주 대상체의 세포를 감염시키는 질병을 치료하기 위해 사용될 수 있다. 감염원 유전자에 의해 암호화된 RNA 분자를 표적으로 하는 CRISPR-연관 단백질을 프로그래밍함으로써 감염원에 감염된 세포가 표적이 될 수 있고 세포사가 유도될 수 있다.

[0480] 또한, 시험관 내 RNA 감지 검정이 사용되어 특정 RNA 기질을 검출할 수 있다. CRISPR-연관 단백질은 살아있는 세포에서 RNA-기반 감지를 위해 사용될 수 있다. 적용분야의 예에는 예를 들어 질병-특이적 RNA 감지에 의한 진단이 있다.

[0481] 폴리뉴클레오티드 서열을 표적 DNA 서열 내로 삽입하는 것이 바람직한 적용분야에서, 삽입될 공여체 서열을 포함하는 폴리뉴클레오티드가 또한 세포에 제공된다. "공여체 서열" 또는 "공여체 폴리뉴클레오티드"라 함은 부위-특이적 변형 폴리펩티드에 의해 유도된 절단 부위에 삽입되는 핵산 서열을 의미한다. 공여체 폴리뉴클레오티드는 상기 공여체 폴리뉴클레오티드와 상기 공여체 폴리뉴클레오티드가 상동성을 구비한 대상인 게놈 서열과의 사이에 상동성-기반 복구를 뒷받침하기 위해 절단 부위에서 게놈 서열에 대해 충분한 상동성, 예컨대 절단 부위에서, 예컨대 절단 부위의 약 50개 이하의 염기 내에, 예컨대 약 30개 염기 내에, 약 15개 염기 내에, 약 10개 염기 내에, 약 5개 염기 내의 측면에 위치하거나 절단 부위의 바로 측면에 위치하는 뉴클레오티드 서열에 대해 예를 들어 70%, 80%, 85%, 90%, 95% 또는 100% 상동성을 함유한다. 공여체와 게놈 서열 사이의 서열 상동성의 대략 25, 50, 100 또는 200개의 뉴클레오티드, 또는 200개 초과의 뉴클레오티드(또는 10 내지 200개 이상의 뉴클레오티드의 임의의 적분 값)는 상동성-기반 복구를 뒷받침한다. 공여체 서열은 임의의 길이, 예컨대 10개 이상의 뉴클레오티드, 50개 이상의 뉴클레오티드, 100개 이상의 뉴클레오티드, 250개 이상의 뉴클레오티드, 500개 이상의 뉴클레오티드, 1000개 이상의 뉴클레오티드, 5000개 이상의 뉴클레오티드 등일 수 있다.

[0482] 공여체 서열은 일반적으로 이것이 대체하는 게놈 서열과 동일하지 않다. 오히려, 상동성-기반 복구를 뒷받침하기에 충분한 상동성이 존재하는 한, 공여체 서열은 게놈 서열과 관련하여 적어도 1개 이상의 단일 염기 변화, 삽입, 결실, 역위 또는 재배열을 함유할 수 있다. 일부 구현예에서, 공여체 서열은 2개의 상동성 영역이 측면에 위치하는 비-상동성 서열을 포함하여, 표적 DNA 영역과 2개의 측면에 위치하는 서열 사이의 상동성-기반 복구는 표적 영역에 비-상동성 서열의 삽입을 초래한다. 또한 공여체 서열은, 관심 DNA 영역에 상동성이 아니고 관심 DNA 영역 내로 삽입되도록 의도되지 않은 서열을 함유하는 벡터 골격을 포함할 수 있다. 일반적으로, 공여체 서열의 상동성 영역(들)은 재조합이 요구되는 게놈 서열에 대해 적어도 50% 서열 동일성을 가진다. 특정 구현예에서, 60%, 70%, 80%, 90%, 95%, 98%, 99% 또는 99.9% 서열 동일성이 존재한다. 공여체 폴리뉴클레오티드의 길이에 따라 1% 내지 100%의 임의의 값의 서열 동일성이 존재할 수 있다.

[0483] 공여체 서열은 게놈 서열과 비교할 때 특정 서열 차이, 예를 들어 제한 부위, 뉴클레오티드 다형성, 선택 가능한 마커(예: 약물 저항성 유전자, 형광 단백질, 효소 등) 등을 포함할 수 있으며, 이들은 예를 들어 절단 부위에 공여체 서열의 성공적 삽입을 평가하기 위해 사용될 수 있거나 일부 경우에 다른 목적(예: 표적으로 하는 게놈 유전자좌에서 발현을 보여줌)을 위해 사용될 수 있다. 일부 경우에, 암호화 영역에 위치하는 경우 이러한 뉴클레오티드 서열 차이는 아미노산 서열을 변화시키지 않거나 침묵 아미노산을 변화시킨다(즉, 단백질의 구조나 기능에 영향을 주지 않는 변화). 대안적으로, 이들 서열 차이는 마커 서열의 제거를 위해 나중에 활성화될 수 있는 FLP, loxP 서열 등과 같은 측면에 위치하는 재조합 서열을 포함할 수 있다.

[0484] 공여체 서열은 단일 가닥 DNA, 단일 가닥 RNA, 이중 가닥 DNA 또는 이중 가닥 RNA로서 세포에 제공될 수 있다.

이는 선형 또는 원형 형태로 세포에 도입될 수 있다. 선형 형태로 도입되는 경우, 공여체 서열의 말단은 당업계에 알려진 방법으로 (예를 들어, 핵산외 분해로부터) 보호될 수 있다. 예를 들어, 하나 이상의 디데옥시뉴클레오티드 잔기가 선형 분자의 3' 말단에 추가되고/되거나 자기-상보적 올리고뉴클레오티드가 한쪽 또는 양쪽 말단에 결찰된다. 외인성 폴리뉴클레오티드를 분해로부터 보호하기 위한 추가 방법은 말단 아미노기(들)의 추가 및 예를 들어 포스포로티오에이트, 포스포 아미데이트 및 O-메틸 리보스 또는 데옥시리보스 잔기와 같은 변형된 뉴클레오티드간 연결의 사용을 포함하지만 이에 국한되지는 않는다. 선형 공여체 서열의 말단을 보호하는 대안으로서, 재조합에 영향을 주지 않으면서 열화될 수 있는 상동성 영역 외부에 추가 길이의 서열이 포함될 수 있다. 공여체 서열은 예를 들어 복제 기원, 프로모터 및 항생제 저항성을 암호화하는 유전자와 같은 추가 서열을 갖는 벡터 분자의 일부로서 세포 내로 도입될 수 있다. 더욱이, 공여체 서열은 네이키드(naked) 핵산으로, 리포좀 또는 폴록사머와 같은 작용제와 복합화된 핵산으로 도입될 수 있거나, DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드를 암호화하는 핵산에 대해 전술한 바와같이 바이러스(예를 들어, 아데노바이러스, AAV)에 의해 전달될 수 있다.

[0485] 상기에 기술된 방법에 따라, 관심 DNA 영역은 생체외에서 절단되고 변형, 즉 "유전적으로 변형"될 수 있다. 일부 구현예에서, 선택 가능한 마커가 관심 DNA 영역 내로 삽입된 경우, 세포 집단은 유전적으로 변형된 세포를 나머지 집단으로부터 분리시킴으로써 유전 변형을 포함하는 세포로 풍부해질 수 있다. 풍부해지기 전, "유전적으로 변형된" 세포는 세포 집단의 단지 약 1% 이상(예를 들어, 2% 이상, 3% 이상, 4% 이상, 5% 이상, 6% 이상, 7% 이상, 8% 이상, 9% 이상, 10% 이상, 15% 이상 또는 20% 이상)을 구성할 수 있다. "유전적으로 변형된" 세포의 분리는 사용된 선택 가능한 마커에 적합한 임의의 편리한 분리 기술에 의해 달성될 수 있다. 예를 들어, 형광 마커가 삽입된 경우에 형광 활성화된 세포 분류를 통해 세포가 분리될 수 있는 반면, 세포 표면 마커가 삽입된 경우에는 친화도 분리 기술, 예를 들어 자기 분리, 친화도 크로마토그래피, 고체 매트릭스에 부착된 친화성 시약을 사용한 "패닝(panning)" 또는 기타 편리한 기술에 의해 세포가 이종 집단으로부터 분리될 수 있다. 정확한 분리를 제공하는 기술에는 다중 색상 채널, 예각 및 둔각 광산란 감지 채널, 임피던스 채널 등과 같이 다양한 수준의 정교함을 가질 수 있는 형광 활성화 세포 분류기가 포함된다. 죽은 세포와 연관된 염료(예를 들어, 요오드화 프로피듐)를 사용함으로써 세포를 죽은 세포에 대해 선택할 수 있다. 유전적으로 변형된 세포의 생존 가능성에 지나치게 해를 가하지 않는 임의의 기술이 사용될 수 있다. 변형된 DNA를 포함하는 세포가 매우 풍부한 세포 조성물은 이러한 방식으로 달성된다. "매우 풍부한"이라 함은 유전적으로 변형된 세포가 세포 조성물의 70% 이상, 75% 이상, 80% 이상, 85% 이상, 90% 이상, 예를 들어 세포 조성물의 약 95% 이상 또는 98% 이상임을 의미한다. 즉, 조성물은 유전적으로 변형된 세포의 실질적으로 순수한 조성물일 수 있다.

[0486] 본원에 기술된 방법에 의해 생산된 유전적으로 변형된 세포는 즉시 사용될 수 있다. 대안적으로, 세포는 액체 질소 온도에서 동결될 수 있고 장기간 보관되고 해동되고 재사용될 수 있다. 이러한 경우, 세포는 일반적으로 10% 디메틸술폭시드(DMSO), 50% 혈청, 40% 완충 배지, 또는 이러한 동결 온도에서 세포를 보존하기 위해 당업계에서 통상적으로 사용되는 일부 다른 용액에서 동결되고, 동결된 배양 세포를 해동하기 위해 당업계에 일반적으로 알려진 방식으로 해동된다.

[0487] 유전적으로 변형된 세포는 다양한 배양 조건 하에 시험관내에서 배양될 수 있다. 세포는 배양을 통해 확장될 수 있다, 즉, 증식을 촉진하는 조건 하에서 성장할 수 있다. 배양 배지는 액체 또는 반고체, 예를 들어 한천, 메틸셀룰로오스 등을 함유하는 것일 수 있다. 세포 집단은 송아지 태아 혈청(약 5 내지 10%), L-글루타민, 티올, 특히 2-머캅토에탄올 및 항생제, 예를 들어 페니실린과 스트렙토마이신으로 보충된 Iscove의 변형 DMEM 또는 RPMI 1640과 같은 적절한 영양 배지에 현탁될 수 있다. 배양물에는 조절 T 세포가 반응하는 성장 인자가 함유될 수 있다. 본원에 정의된 성장 인자는 막횡단 수용체에 대한 특정 효과를 통해 배양물 또는 비손상 조직에서 세포의 생존, 성장 및/또는 분화를 촉진할 수 있는 분자이다. 성장 인자는 폴리펩티드 및 비폴리펩티드 인자를 포함한다.

[0488] 이러한 방식으로 유전적으로 변형된 세포는 예를 들어 질병을 치료하기 위한 유전자 요법과 같은 목적을 위해 또는 항바이러스제, 항병원체 또는 항암 치료제로서, 농업에서 유전적으로 변형된 유기체의 생산을 위해 또는 생물학적 연구를 위해 대상체에게 이식될 수 있다. 대상체는 신생아, 청소년 또는 성인일 수 있다. 포유동물 대상체가 특히 흥미롭다. 본 방법으로 치료될 수 있는 포유동물 종은 개와 고양이; 말; 소; 양 등과 영장류, 특히 인간을 포함한다. 동물 모델, 특히 작은 포유동물(예를 들어, 마우스, 랫트, 기니피그, 햄스터, 토류목(예: 토끼) 등)은 실험적 연구를 위해 사용될 수 있다.

[0489] 세포는 단독으로 또는 적합한 기질 또는 매트릭스와 함께 대상체에게 제공되어, 예를 들어 세포가 이식되는 조

직에서의 성장 및/또는 조직화를 뒷받침할 수 있다. 일반적으로 적어도 $1x10^3$개 세포, 예를 들어 5x103개 세포, $1x10^4$개 세포, $5x10^4$개 세포, $1x10^5$개 세포, $1x10^6$개 세포 이상이 투여된다. 세포는 비경구, 피하, 정맥내, 두개내, 척수내, 안구내 또는 척수액 경로 중 임의의 경로를 통해 대상체에게 도입될 수 있다. 세포는 주사, 카테터 등에 의해 도입될 수 있다. 또한 세포는 유전자이식 동물(예: 유전자이식 마우스)을 발생시킬 목적으로 배아(예: 배반포) 내에 도입될 수 있다.

[0490] 대상체에 대한 치료 투여 횟수는 다를 수 있다. 유전적으로 변형된 세포를 대상체에게 도입하는 것은 일회성 사건일 수 있지만; 특정 상황에서 이러한 치료는 제한된 기간 동안 개선을 끌어낼 수 있으며 지속적인 반복 치료를 필요로 할 수 있다. 다른 상황에서, 효과가 관찰되기 전에 유전적으로 변형된 세포의 여러 번 투여가 필요할 수 있다. 정확한 프로토콜은 질병 또는 병태, 질병 단계 및 치료되는 개별 대상체의 매개변수에 따라 달라진다.

[0491] 본 발명의 다른 측면에서, DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드가 사용되어, 예를 들어 질병을 치료하기 위한 유전자 요법과 같은 목적으로 또는 항바이러스제, 항병원체 또는 항암 치료제로서, 농업에서 유전적으로 변형된 유기체의 생산을 위해 또는 생물학적 연구를 위해 생체 내 세포 DNA를 변형시킨다. 이들 생체내 구현예에서, DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드는 개체에 직접 투여된다. DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드는 펩티드, 소분자 및 핵산을 대상체에게 투여하기 위해 당업계에 잘 알려진 임의의 다수의 방법에 의해 투여될 수 있다. DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드는 다양한 제형 내에 통합될 수 있다. 보다 구체적으로, 본 발명의 DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드는 적절한 약제학적으로 허용 가능한 담체 또는 희석제와 조합하여 약제학적 조성물로 제형화될 수 있다.

[0492] 약제학적 제조물은 약제학적으로 허용 가능한 비히클에 존재하는 하나 이상의 DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드를 포함하는 조성물이다. "약제학적으로 허용 가능한 비히클"은 인간과 같은 포유동물에 대한 사용을 위해 연방 또는 주 정부의 규제 기관에 의해 승인되거나 미국 약전 또는 기타 일반적으로 인정되는 약전에 등재된 비히클일 수 있다. 용어 "비히클"은 본 발명의 화합물이 포유동물에 투여하기 위해 함께 제형화되는 희석제, 아쥬반트, 부형제 또는 담체를 의미한다. 이러한 약제학적 비히클은 지질, 예를 들어 리포솜, 예를 들어 리포솜 덴드리머; 액체, 예컨대 물 및, 석유, 동물성, 식물성 또는 합성 기원의 오일을 포함한 오일, 예컨대 낙화생유, 대두유, 광유, 참기름 등, 식염수; 아카시아검, 젤라틴, 전분 페이스트, 활석, 케라틴, 콜로이드성 실리카, 요소 등일 수 있다. 또한, 보조제, 안정제, 증점제, 윤활제 및 착색제가 사용될 수 있다. 약제학적 조성물은 정제, 캡슐, 산제, 과립, 연고, 용제, 좌제, 주사제, 흡입제, 겔, 소구체 및 에어로졸과 같은 고체, 반고체, 액체 또는 기체 형태의 제조물 내에 제형화될 수 있다. 이와 같이, DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드의 투여는 경구, 협측, 직장, 비경구, 복강내, 피내, 경피, 기관내, 안구내 등 투여를 포함하는 다양한 방식으로 달성될 수 있다. 활성 작용제는 투여 후 전신용일 수 있거나 국소 투여, 벽내 투여, 또는 이식 부위에서 활성 용량을 유지하는 작용을 하는 이식 장치의 사용에 의해 국소화될 수 있다. 활성 작용제는 즉각적인 활성을 위해 제형화될 수 있거나 서방용으로 제형화될 수 있다.

[0493] 일부 병태, 특히 중추신경계 병태의 경우, 혈액뇌장벽(BBB)을 통과하는 작용제를 제형화해야 할 수 있다. 혈액뇌장벽(BBB)을 통과하는 약물 전달의 한 가지 전략은 만니톨이나 류코트리엔과 같은 삼투압 수단에 의해 또는 브라디키닌과 같은 혈관활성 서브스턴스를 사용함으로써 생화학적으로 BBB 저해를 수반하는 것이다. 뇌 종양에 대해 특정 작용제를 표적으로 하기 위해 BBB 개구를 사용할 가능성 또한 선택안이다. BBB 저해제는 조성물이 혈관내 주사에 의해 투여될 때 본 발명의 치료 조성물과 함께 투여될 수 있다. BBB를 통과시키는 다른 전략은 Caveolin-1 매개 세포투과이동(trancytosis), 글루코스 및 아미노산 담체와 같은 담체-매개된 수송체, 인슐린 또는 트랜스페린에 대한 수용체-매개된 세포투과이동, 및 p-당단백질과 같은 활성 유출 수송체를 포함하는 내인성 수송체 시스템의 사용을 수반할 수 있다. 또한 활성 수송 모이어티는 혈관의 내피 벽을 가로지르는 수송을 촉진하기 위해 본 발명에서 사용하기 위한 치료 화합물에 접합될 수 있다.

[0494] 대안적으로, BBB 뒤에 치료제의 약물 전달은 국소 전달에 의해, 예를 들어 척수강내 전달에 의해 이루어질 수 있다.

[0495] 일반적으로, 유효량의 DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드가 제공된다. 생체외 방법과 관련하여 상기에서 검토된 바와 같이, 생체내 DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드의 유효량 또는 유효 용량은 음성 대조군, 예컨대 빈 벡터 또는 관

련 없는 폴리펩티드와 접촉된 세포와 비교하여 2개 상동성 서열 사이에서 관찰된 재조합 양을 2배 이상의 증가하도록 유도하는 양이다. 재조합 양은 예를 들어 상기에 기재되고 당업계에 알려진 임의의 편리한 방법으로 측정할 수 있다. 투여될 DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드의 유효량 또는 유효 용량의 계산은 당업계의 기술 범위 내에 있고 당업자에게는 일상적이다. 투여될 최종 양은 투여 경로 및 치료될 장애 또는 병태의 특성에 따라 달라진다.

[0496] 특정한 환자에게 제공되는 유효량은 다양한 요인에 따라 달라지며, 그 중 일부는 환자마다 다르다. 유능한 임상의는 필요에 따라 질병 병태를 중단시키거나 진행을 역전시키기 위해 환자에게 투여할 치료제의 유효량을 결정할 수 있다. LD50 동물 데이터 및 작용제에 대해 이용 가능한 다른 정보를 활용하여, 임상의는 투여 경로에 따라 개인에 대해 최대 안전 용량을 결정할 수 있다. 예를 들어, 치료 조성물이 투여되는 체액강이 큰 경우, 정맥내 투여 용량은 척수강내 투여 용량보다 많을 수 있다. 유사하게, 신체에서 빠르게 제거되는 조성물은 치료 농도를 유지하기 위해 더 높은 용량으로 또는 반복되는 용량으로 투여될 수 있다. 유능한 임상의는 통상의 기술을 활용하여 일상적인 임상 시험 과정에서 특정 치료제의 용량을 최적화할 수 있다.

[0497] 의약에 혼입시키기 위해, DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드를 적합한 상업적 공급원으로부터 얻을 수 있다. 일반적인 제안으로서, 용량당 비경구로 투여되는 DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드의 총 약제학적 유효량은 용량 반응 곡선으로 측정될 수 있는 범위 내에 있다.

[0498] DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드에 기반한 요법, 즉 치료적 투여를 위해 사용될 DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드의 제조물은 무균이어야 한다. 멸균 여과막(예: 0.2μm 막)을 통한 여과를 통해 무균이 쉽게 달성된다. 치료 조성물은 일반적으로 멸균 접근 포트를 갖는 용기, 예를 들어 정맥 주사액 백 또는 피하 주사 바늘이 관통할 수 있는 마개가 있는 바이알 내에 놓인다. DNA-표적 RNA 및/또는 부위-특이적 변형 폴리펩티드 및/또는 공여체 폴리뉴클레오티드에 기반한 요법은 단위 또는 다중-용량 용기, 예를 들어 밀봉 앰플 또는 바이알에 수용액으로 또는 재구성을 위한 동결건조 제형으로 보관될 수 있다. 동결건조 제형의 예로서, 10mL 바이알에 멸균-여과된 1%(w/v) 화합물 수용액 5ml를 채우고, 생성된 혼합물은 동결건조된다. 주입 용액은 정균 주사용수를 사용하여 동결건조된 화합물을 재구성함으로써 제조된다.

[0499] 약제학적 조성물은 동물 또는 인간 투여를 위한 약제학적 조성물을 제형화하기 위해 흔히 사용되는 비히클로 정의되는 약제학적으로 허용 가능한 무독성 희석제 담체를 원하는 제형에 따라 포함할 수 있다. 희석제는 조합의 생물학적 활성에 영향을 주지 않도록 선택된다. 이러한 희석제의 예는 증류수, 완충수, 생리식염수, PBS, 링거 용액, 덱스트로스 용액 및 행크 용액이다. 또한, 약제학적 조성물 또는 제형은 다른 담체, 아쥬반트 또는 비독성, 비치료적, 비면역원성 안정제, 부형제 등을 포함할 수 있다. 또한 조성물은 pH 조절제 및 완충제, 독성 조절제, 습윤제 및 청정제와 같은 생리학적 조건에 근접시키는 추가적인 서브스턴스를 포함할 수 있다.

[0500] 또한 조성물은 예를 들어 항산화제와 같은 다양한 안정제 중 임의의 것을 포함할 수 있다. 약제학적 조성물이 폴리펩티드를 포함할 때, 폴리펩티드는 폴리펩티드의 생체내 안정성을 증강시키거나 그렇지 않으면 약리학적 특성을 증강시키는(예를 들어, 폴리펩티드의 반감기를 증가시키고 그의 독성을 낮추고 용해도 또는 흡수를 증강시키는) 잘 알려진 다양한 화합물과 복합체를 형성할 수 있다. 이러한 변형제 또는 복합화 작용제의 예는 황산염, 글루콘산염, 시트르산염 및 인산염을 포함한다. 또한 조성물의 핵산 또는 폴리펩티드는 생체내 속성을 증강시키는 분자와 복합체를 형성할 수 있다. 이러한 분자는 예를 들어 탄수화물, 폴리아민, 아미노산, 기타 펩티드, 이온(예: 나트륨, 칼륨, 칼슘, 마그네슘, 망간) 및 지질을 포함한다.

[0501] 약제학적 조성물은 예방적 및/또는 요법적 치료를 위해 투여될 수 있다. 예를 들어 LD50(집단의 50%에서 치사가 이루어지는 용량) 및 ED50(집단의 50%에 치료적으로 유효한 용량) 결정을 포함하여 활성 성분의 독성 및 요법적 효능은 세포 배양물 및/또는 실험 동물에서 표준 약학 절차에 따라 결정될 수 있다. 독성 및 요법 효과 사이의 용량 비는 치료 계수이고, LD50/ED50 비로 표현될 수 있다. 큰 치료 계수를 나타내는 요법이 바람직하다.

[0502] 세포 배양물 및/또는 동물 연구에서 얻은 데이터는 인간을 위한 투여량 범위를 제형화하기 위해 사용될 수 있다. 활성 성분의 투여량은 일반적으로 낮은 독성과 함께 ED50을 포함하는 순환 농도 범위 내에 있다. 투여량은 이용되는 투여 형태 및 활용되는 투여 경로에 따라 이 범위 내에서 다를 수 있다.

[0503] 약제학적 조성물을 제형화하기 위해 사용되는 구성성분은 바람직하게는 고순도이고 잠재적으로 유해한 오염물질을 실질적으로 갖지 않는다(예를 들어, 적어도 국민 식품(NF) 등급, 일반적으로 최소한 분석 등급, 더 일반적으

로는 최소한 약제 등급). 더욱이, 생체내 사용으로 의도된 조성물은 일반적으로 무균이다. 주어진 화합물이 사용 전에 합성되어야 하는 경우, 생성된 생성물은 일반적으로 임의의 잠재적 독성 작용제, 특히 합성 또는 정제 프로세스 동안에 존재할 수 있는 임의의 내독소를 실질적으로 갖지 않는다. 또한 비경구 투여를 위한 조성물은 무균이고 실질적으로 등장성이며 GMP 조건 하에서 제조된다.

[0504] **전달 시스템**

[0505] 본원에 기술된 CRISPR 시스템 또는 이의 구성요소, 이의 핵산 분자, 및/또는 이의 구성요소를 암호화하거나 제공하는 핵산 분자, CRISPR-연관 단백질, 또는 RNA 가이드는 벡터, 예를 들어 플라스미드 및 전달 벡터와 같은 다양한 전달 시스템에 의해 전달될 수 있다. 예시적인 구현예가 아래에 기술된다. CRISPR 시스템(예를 들어, 본원에 기술된 핵염기 편집기를 포함하는 Cas9를 포함)은 바이러스 벡터에 함유된 핵산 상에서 암호화될 수 있다. 바이러스 벡터는 렌티바이러스, 아데노바이러스, 레트로바이러스 및 아데노-연관 바이러스(AAV)를 포함할 수 있다. 바이러스 벡터는 적용분야에 기반하여 선택될 수 있다. 예를 들어, AAV는 약한 면역원성으로 인해 생체내 유전자 전달을 위해 흔히 사용된다. 아데노바이러스는 이들이 유도하는 강력한 면역원성 반응 때문에 흔히 백신으로 사용된다. 바이러스 벡터의 패키징 용량은 벡터 내로 패키징될 수 있는 염기 편집기의 크기를 제한할 수 있다. 예를 들어, AAV의 패키징 용량은 2개의 145 염기 역위 말단 반복(ITR)을 포함하여 ~4.5kb이다.

[0506] AAV는 파보바이러스 계열에 속하는 작은 단일 가닥의 DNA 의존성 바이러스이다. 4.7kb 야생형(wt) AAV 게놈은 4개의 복제 단백질과 3개의 캡시드 단백질을 각각 암호화하는 2개의 유전자로 구성되고, 각 측면에는 145bp 역위 말단 반복(ITR)이 측면에 위치해 있다. 비리온은 차등 스플라이싱(Vp1) 및 선택적 번역 시작 부위(각각 Vp2 및 Vp3)를 제외하고 동일한 개방형 판독 프레임에서 1:1:10 비율로 생성된 3개의 캡시드 단백질인 Vp1, Vp2 및 Vp3로 구성된다. Vp3은 비리온에서 가장 풍부한 하위단위체이고 바이러스 향성을 정의하며 세포 표면에서 수용체 인식에 참여한다. 바이러스 감염성에서 기능하는 포스포리파제 도메인은 Vp1의 독특한 N 말단에서 식별되었다.

[0507] wt AAV와 유사하게, 재조합 AAV(rAAV)는 시스-작용 145-bp ITR을 활용하여 벡터 이식유전자 카세트의 측면에 위치하고 외래 DNA 패키징을 위해 최대 4.5kb를 제공한다. 감염에 이어, rAAV는 본 발명의 융합 단백질을 발현할 수 있고, 숙주 게놈 내로 통합되지 않으면서 머리에서 꼬리(head-to tail) 동일연쇄서열에 에피좀으로 존재함으로써 지속될 수 있다. 시험관내 및 생체내에서 이 시스템을 사용하는 rAAV 성공의 많은 예가 있지만, 유전자의 암호화 서열 길이가 크기가 wt AAV 게놈과 동일하거나 이보다 클 때 제한된 패키징 용량은 AAV-매개 유전자 전달의 사용을 제한하였다.

[0508] AAV 벡터의 작은 패키징 용량으로 인해 이 크기를 초과하는 다수의 유전자 전달 및/또는 큰 생리학적 조절 요소의 사용이 어려워진다. 이러한 어려움은 예를 들어 전달될 단백질(들)을 2개 이상의 단편으로 나눔으로써 해결될 수 있고, 상기 N-말단 단편은 분할된 인테인-N에 융합되고 C-말단 단편은 분할된 인테인-C에 융합된다. 그 다음, 이들 단편은 2개 이상의 AAV 벡터 내로 패키징된다. 본원에서 사용될 때 "인테인"은 측면에 위치한 N-말단 및 C-말단 엑스테인(예를 들어, 연결될 단편)을 결찰시키는 자가-스플라이싱 단백질 인트론(예를 들어, 펩티드)을 의미한다. 이종 단백질 단편을 연결하기 위한 특정 인테인의 사용이 예를 들어 Wood 등, J. Biol. Chem. 289(21); 14512-9 (2014)에 기술되어 있다. 예를 들어, 분리된 단백질 단편에 융합될 때 인테인 IntN과 IntC는 서로를 인식하고, 스스로 스플라이싱되고, 이들이 융합된 단백질 단편의 측면에 위치한 N- 및 C-말단 엑스테인을 동시에 결찰시켜, 2개 단백질 단편으로부터 전장 단백질을 재구성한다. 다른 적합한 인테인은 당업자에게 명백할 것이다.

[0509] 일부 구현예에서, 본 발명의 CRISPR 시스템은 다양한 길이를 가질 수 있다. 일부 구현예에서, 단백질 단편은 2개 아미노산 내지 약 1000개 아미노산 범위의 길이를 가진다. 일부 구현예에서, 단백질 단편은 약 5개 아미노산 내지 약 500개 아미노산 범위의 길이를 가진다. 일부 구현예에서, 단백질 단편은 약 20개 아미노산 내지 약 200개 아미노산 범위의 길이를 가진다. 일부 구현예에서, 단백질 단편은 약 10개 아미노산 내지 약 100개 아미노산 범위의 길이를 가진다. 다른 길이의 적합한 단백질 단편이 당업자에게 명백할 것이다.

[0510] 일부 구현예에서, 뉴클레아제(예를 들어, Cas9)의 일부 또는 단편은 인테인에 융합된다. 뉴클레아제는 인테인의 N-말단 또는 C-말단에 융합될 수 있다. 일부 구현예에서, 융합 단백질의 일부 또는 단편은 인테인에 융합되고 AAV 캡시드 단백질에 융합된다. 인테인, 뉴클레아제 및 캡시드 단백질은 임의의 배열(예를 들어, 뉴클레아제-인테인-캡시드, 인테인-뉴클레아제-캡시드, 캡시드-인테인-뉴클레아제 등)로 함께 융합될 수 있다. 일부 구현예에서, 인테인의 N-말단은 융합 단백질의 C-말단에 융합되고, 인테인의 C-말단은 AAV 캡시드 단백질의 N-말단에 융합된다.

[0511] 일 구현예에서, 이중 AAV 벡터는 큰 이식유전자 발현 카세트를 2개의 별도 반쪽(5' 및 3' 말단, 또는 머리와 꼬리)으로 분할함으로써 발생되고, 상기 카세트의 각 반쪽은 단일 AAV 벡터(< 5kb)에 패키징된다. 그 다음, 전장 이식유전자 발현 카세트의 재조립은 이중 AAV 벡터 둘 모두에 의한 동일 세포의 동시 감염 이후 다음에 의해 달성된다: (1) 5'와 3' 게놈 사이의 상동성 재조합(HR)(이중 AAV 중첩 벡터); (2) 5' 및 3' 게놈의 ITR-매개의 꼬리에서 머리로 연쇄동일서열화(이중 AAV 트랜스-스플라이싱 벡터); 또는 (3) 이들 두 메커니즘의 조합(이중 AAV 하이브리드 벡터). 생체내 이중 AAV 벡터의 사용으로 인해 전장 단백질이 발현된다. 이중 AAV 벡터 플랫폼의 사용은 >4.7kb 크기의 이식유전자에 대한 효율적이고 실행 가능한 유전자 전달 전략을 나타낸다.

[0512] 본원에 기술된 Cas9를 포함하는 CRISPR 시스템을 설계하기 위해 개시된 전략은 바이러스 벡터 내로 패키징될 수 있는 CRISPR 시스템의 발생에 유용할 수 있다. 염기 편집기 전달을 위해 RNA 또는 DNA 바이러스 기반 시스템을 사용하는 것은 배양물 또는 숙주에서 바이러스를 특정 세포에 대해 표적으로 하고 바이러스 페이로드를 핵 또는 숙주 세포 게놈에 전달하기 위해 고도로 진화된 프로세스를 활용한다. 바이러스 벡터는 배양물의 세포, 환자(생체 내)에게 직접 투여할 수 있거나 시험관내 세포를 치료하기 위해 사용될 수 있고, 변형된 세포는 선택적으로 환자(생체외)에게 투여될 수 있다. 종래의 바이러스 기반 시스템은 유전자 전달을 위한 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노 연관 및 단순 포진 바이러스 벡터를 포함할 수 있다. 레트로바이러스, 렌티바이러스 및 아데노 연관 바이러스 유전자 전달 방법을 사용하여 숙주 게놈 내 통합이 가능하여, 종종 삽입된 이식유전자의 장기 발현을 초래한다. 또한, 많은 상이한 세포 유형과 표적 조직에서 높은 형질도입 효율이 관찰되었다.

[0513] 레트로바이러스의 향성은 외래 외피 단백질을 혼입함으로써 변경될 수 있으며, 표적 세포의 잠재적인 표적 집단을 확장할 수 있다. 렌티바이러스 벡터는, 비분할 세포를 형질도입하거나 감염시킬 수 있고 일반적으로 높은 바이러스 역가를 생성할 수 있는 레트로바이러스 벡터이다. 따라서 레트로바이러스 유전자 전달 시스템의 선택은 표적 조직에 따라 달라진다. 레트로바이러스 벡터는 최대 6 내지 10kb의 외래 서열에 대한 패키징 용량을 갖는 시스-작용 긴 말단 반복으로 구성된다. 최소 시스-작용 LTR은 벡터의 복제 및 패키징에 충분하고, 이는 이후 영구적인 이식유전자 발현을 제공하기 위해 치료 유전자를 표적 세포 내로 통합하기 위해 사용된다. 널리 사용되는 레트로바이러스 벡터는 쥣과 백혈병 바이러스(MuLV), 긴팔원숭이 백혈병 바이러스(GaLV), 유인원 면역 결핍 바이러스(SIV), 인간 면역 결핍 바이러스(HIV) 및 이들의 조합에 기반한 것을 포함한다(예를 들어 Buchscher 등, J. Virol. 66:2731-2739 (1992); Johann 등, J. Virol. 66:1635-1640 (1992); Sommnerfelt 등, Virol. 176:58-59 (1990); Wilson 등, J. Virol. 63:2374-2378 (1989); Miller 등, J. Virol. 65:2220-2224 (1991); PCT/US94/05700 참조).

[0514] 레트로바이러스 벡터, 특히 렌티바이러스 벡터는 표적 세포 내로 효율적인 통합을 위해 주어진 길이보다 작은 폴리뉴클레오티드 서열을 필요로 할 수 있다. 예를 들어, 9kb보다 큰 길이의 레트로바이러스 벡터는 더 작은 길이의 레트로바이러스 벡터에 비해 낮은 바이러스 역가를 초래할 수 있다. 일부 측면에서, 본 개시내용의 CRISPR 시스템(예를 들어, 본원에 개시된 Cas9 포함)은 레트로바이러스 벡터를 통한 표적 세포 내로 효율적인 패키징 및 전달이 가능하도록 하는 충분한 크기를 가진다. 일부 경우에 Cas9는 가이드 핵산 및/또는 표적화 가능한 뉴클레아제 시스템의 다른 구성요소와 함께 발현될 때에도 효율적인 패키징 및 전달을 허용하도록 하는 크기를 가진다.

[0515] 일시적 발현이 바람직한 적용분야에서 아데노바이러스 기반 시스템이 사용될 수 있다. 아데노바이러스 기반 벡터는 많은 세포 유형에서 매우 높은 형질도입 효율을 가질 수 있고, 세포 분열을 필요로 하지 않는다. 이러한 벡터를 사용하면 높은 역가와 발현 수준이 얻어졌다. 이 벡터는 비교적 단순한 시스템에서 다량으로 생산될 수 있다. 또한 아데노 연관 바이러스("AAV") 벡터가 사용되어, 예를 들어 핵산 및 펩티드의 시험관내 생산을 위해 그리고 생체내 및 생체외 유전자 치료 절차를 위해 표적 핵산으로 세포를 형질도입할 수 있다(예를 들어, West 등, Virology 160:38-47 (1987); 미국 특허 제4,797,368호; WO 93/24641호; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994) 참조. 재조합 AAV 벡터의 작제는 미국 특허 제5,173,414호; Tratschin 등, Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, 등, Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); 및 Samulski 등, J. Virol. 63:03822-3828 (1989)을 포함하여 다수의 출판물에 기술되어 있다.

[0516] 따라서 본원에 기술된 CRISPR 시스템(예를 들어, 본원에 개시된 Cas9 포함)은 바이러스 벡터와 함께 전달될 수 있다. 염기 편집기 시스템의 하나 이상의 구성 요소는 하나 이상의 바이러스 벡터 상에서 암호화될 수 있다. 예를 들어, 염기 편집기와 가이드 핵산은 단일 바이러스 벡터 상에서 암호화될 수 있다. 다른 경우에 염기 편집기와 가이드 핵산이 상이한 바이러스 벡터 상에서 암호화된다. 어느 한 경우에, 염기 편집기 및 가이드 핵산은 각

각 프로모터 및 터미네이터에 작동 가능하게 연결될 수 있다.

[0517] 바이러스 벡터 상에서 암호화되는 구성요소의 조합은 선택된 바이러스 벡터의 운반대상(cargo) 크기 제약에 의해 결정될 수 있다.

[0518] ***염기 편집기의 비바이러스 전달***

[0519] CRISPR에 대한 비바이러스 전달 접근법도 사용할 수 있다. 비바이러스 핵산 벡터의 하나의 중요한 범주는 유기 또는 무기일 수 있는 나노입자이다. 나노입자는 당업계에 잘 알려져 있다. 임의의 적합한 나노입자 설계를 사용하여 게놈 편집 시스템 구성요소 또는 이러한 구성요소를 암호화하는 핵산을 전달할 수 있다. 예를 들어, 유기(예를 들어, 지질 및/또는 중합체) 나노입자는 본 개시내용의 특정 구현예에서 전달 비히클로서 사용하기에 적합할 수 있다. 나노입자 제형 및/또는 유전자 전달에 사용하기 위한 예시적인 지질은 표 5(아래)에 나와 있다.

[0520] **표 5**

유전자 전달에 사용되는 지질

지질	약어	특징
1,2-디올레오일-sn-글리세로-3-포스파티딜콜린	DOPC	헬퍼
1,2-디올레오일-sn-글리세로-3-포스파티딜에탄올아민	DOPE	헬퍼
콜레스테롤		헬퍼
N-[1-(2,3-디올레일옥시)프로필]N,N,N-트리메틸암모늄 클로라이드	DOTMA	양이온성
1,2-디올레오일옥시-3-트리메틸암모늄-프로판	DOTAP	양이온성
디옥타데실아미도글리실스퍼민	DOGS	양이온성
N-(3-아미노프로필)-N,N-디메틸-2,3-비스(도데실옥시)-1-프로파나미늄 브로마이드	GAP-DLRIE	양이온성
세틸트리메틸암모늄 브로마이드	CTAB	양이온성
6-라우르옥시헥실 오르니티네이트	LHON	양이온성
1-(2,3-디올레오일옥시프로필)-2,4,6-트리메틸피리디늄	2Oc	양이온성
2,3-디올레일옥시-N-[2(스퍼민카복스아미도-에틸]-N,N-디메틸-1-프로파나미늄 트리플루오로아세테이트	DOSPA	양이온성
1,2-디올레일-3-트리메틸암모늄-프로판	DOPA	양이온성
N-(2-하이드록시에틸)-N,N-디메틸-2,3-비스(테트라데실옥시)-1-프로파나미늄 브로마이드	MDRIE	양이온성
디미리스토옥시프로필 디메틸 하이드록시에틸 암모늄 브로마이드	DMRI	양이온성
3β-[N-(N',N'-디메틸아미노에탄)-카바모일]콜레스테롤	DC-Chol	양이온성
비스-구아니듐-트렌-콜레스테롤	BGTC	양이온성
1,3-디오데옥시-2-(6-카복시-스퍼밀)-프로필아미드	DOSPER	양이온성
디메틸옥타데실암모늄 브로마이드	DDAB	양이온성
디옥타데실아미도글리실스퍼미딘	DSL	양이온성
rac-[(2,3-디옥타데실옥시프로필)(2-하이드록시에틸)]-디메틸암모늄 클로라이드	CLIP-1	양이온성
rac-[2(2,3-디헥사데실옥시프로필-	CLIP-6	양이온성

[0521]

유전자 전달에 사용되는 지질		
지질	약어	특징
옥시메틸옥시)에틸]트리메틸암모늄 브로마이드		
에틸디미리스토일포스파티딜콜린	EDMPC	양이온성
1,2-디스테아릴옥시-N,N-디메틸-3-아미노프로판	DSDMA	양이온성
1,2-디미리스토일-트리메틸암모늄 프로판	DMTAP	양이온성
O,O'-디미리스틸-N-라이실 아스파르테이트	DMKE	양이온성
1,2-디스테로일-sn-글리세로-3-에틸포스포콜린	DSEPC	양이온성
N-팔미토일 D-에리트로-스핑고신 카바모일-스퍼민	CCS	양이온성
N-t-부틸-N0-테트라데실-3-테트라데실아미노프로피온아미딘	diC14-아미딘	양이온성
옥타데세놀리옥시[에틸-2-헵타데세닐-3 하이드록시에틸] 이미다졸리늄 클로라이드	DOTIM	양이온성
N1 -콜레스테릴옥시카보닐-3,7-디아자노난-1,9-디아민	CDAN	양이온성
2-(3-[비스(3-아미노-프로필)-아미노]프로필아미노)-N-디테트라데실카바모일메틸-아세트아미드	RPR209120	양이온성
1,2-디리놀레일옥시-3-디메틸아미노프로판	DLinDMA	양이온성
2,2-디리놀레일-4-디메틸아미노에틸-[1,3]-디옥솔란	DLin-KC2-DMA	양이온성
디리놀레일-메틸-4-디메틸아미노부티레이트	DLin-MC3-DMA	양이온성

[0522]

[0523] 표 6에는 유전자 전달 및/또는 나노입자 제형에 사용하기 위한 예시적인 중합체가 열거되어 있다.

[0524] **표 6**

유전자 전달에 사용되는 중합체	
중합체	약어
폴리(에틸렌)글리콜	PEG
폴리에틸렌이민	PEI
디티오비스 (숙신이미딜프로피오네이트)	DSP
디메틸-3,3'-디티오비스프로피온이미데이트	DTBP
폴리(에틸렌이민)비스카바메이트	PEIC
폴리(L-라이신)	PLL
히스티딘 개질된 PLL	
폴리(N-비닐피롤리돈)	PVP
폴리(프로필렌이민)	PPI
폴리(아미도아민)	PAMAM
폴리(아미도에틸렌이민)	SS-PAEI
트리에틸렌테트라민	TETA
폴리(β-아미노에스테르)	
폴리(4-하이드록시-L-프롤린 에스테르)	PHP
폴리(알릴아민)	
폴리(α-[4-아미노부틸]-L-글리콜산)	PAGA
폴리(D,L-락트-co-글리콜산)	PLGA
폴리(N-에틸-4-비닐피리디늄 브로마이드)	
폴리(포스파젠)	PPZ
폴리(포스포에스테르)	PPE
폴리(포스포아미데이트)	PPA
폴리(N-2-하이드록시프로필메타크릴아미드)	pHPMA
폴리 (2-(디메틸아미노)에틸 메타크릴레이트)	pDMAEMA
폴리(2-아미노에틸 프로필렌 포스페이트)	PPE-EA
키토산	
갈락토실화 키토산	

[0525]

유전자 전달에 사용되는 중합체	
중합체	약어
N-도데실화 키토산	
히스톤	
콜라겐	
덱스트란-스퍼민	D-SPM

[0526]

[0527] 표 7은 본원에 기술된 Cas9를 암호화하는 폴리뉴클레오티드에 대한 전달 방법을 요약한다.

[0528] **표 7**

전달	벡터/모드	비분화 세포 내로 전달	발현 기간	게놈 통합	전달되는 분자 유형
물리적	(예, 전기영동천공, 유전자총, 칼슘 포스페이트 형질감염	예	일시적	아니요	핵산 및 단백질
바이러스	레트로바이러스	아니요	안정적	예	RNA
	렌티바이러스	예	안정적	예/변형 시 아니요	RNA
	아데노바이러스	예	일시적	아니요	DNA
	아데노-연관 바이러스 (AAV)	예	안정적	아니요	DNA
	백시나 바이러스	예	매우 일시적	아니요	DNA
	단순 포진 바이러스	예	안정적	아니요	DNA
비바이러스	양이온성 리포좀	예	일시적	전달되는 물질에 따라 다름	핵산 및 단백질
	중합성 나노입자	예	일시적	전달되는 물질에 따라 다름	핵산 및 단백질
생물 비바이러스	약독화 박테리아	예	일시적	아니요	핵산
전달 비히클	조작된 박테리오파지	예	일시적	아니요	핵산
	포유동물의 바이러스 유사 입자	예	일시적	아니요	핵산
	생물 리포좀: 파열 적혈구 및 엑소좀	예	일시적	아니요	핵산

[0529]

[0530] 또 다른 측면에서, 게놈 편집 시스템 구성요소 또는 이러한 구성요소를 암호화하는 핵산, 예를 들어 생물학적 활성을 갖는 폴리펩티드(예를 들어 핵염기 편집기)에 선택적으로 융합되는 Cas9 또는 이의 변이체와 같은 핵산 결합 단백질, 및 관심 게놈 핵산 서열을 표적으로 하는 gRNA의 전달은 리보핵단백질(RNP)을 세포에 전달함으로써 달성될 수 있다. RNP는 표적으로 하는 gRNA와 복합체를 형성하는 핵산 결합 단백질, 예를 들어 Cas9를 포함한다. RNP는 공지 방법, 예컨대 전기영동천공, 핵감염 또는 예를 들어 Zuris, J.A. 등, 2015, *Nat. Biotechnology*, 33(1):73-8에 보고된 바와 같은 양이온성 지질-매개 방법을 사용하여 세포에 전달될 수 있다. RNP는 CRISPR 염기 편집 시스템에서 사용하기에, 특히 일차 세포와 같이 형질감염이 어려운 세포에 유리한다. 또한, RNP는 특히 CRISPR 플라스미드에 사용될 수 있는 진핵생물 프로모터, 예를들어 CMV 또는 EF1A가 잘 발현되지 않을 때 세포 내 단백질 발현에서 발생할 수 있는 어려움을 완화시킬 수도 있다. 유리하게, RNP의 사용은 외래 DNA를 세포 내로 전달할 것을 요구하지 않는다. 더욱이, 핵산 결합 단백질과 gRNA 복합체를 포함하는 RNP는 시간이 지남에 따라 열화되기 때문에 RNP 사용에는 표적외 효과를 제한할 가능성이 있다. 플라스미드 기반 기술과 유사한 방식으로 RNP를 사용하여 결합 단백질(예: Cas9 변이체)을 전달하고 상동성 기반 복구(HDR)를 지시할 수 있다.

[0531] CRISPR 시스템(예를 들어, 본원에 기술된 Cas9 포함)을 구동시키기 위해 사용되는 프로모터는 AAV ITR을 포함할 수 있다. 이는 벡터에서 공간을 차지할 수 있는 추가 프로모터 요소를 필요로 하지 않기 때문에 유리할 수

있다. 추가적인 여유 공간은 가이드 핵산 또는 선택 가능한 마커와 같은 추가적인 요소의 발현을 구동하기 위해 사용될 수 있다. ITR 활성은 상대적으로 약하므로, 이는 선택된 뉴클레아제의 과발현으로 인한 잠재적 독성을 낮추기 위해 사용될 수 있다.

[0532] 임의의 적합한 프로모터를 사용하여 Cas9 및 적절한 경우 가이드 핵산의 발현을 구동시킬 수 있다. 유비쿼터스 발현의 경우, 사용될 수 있는 프로모터는 CMV, CAG, CBh, PGK, SV40, 페리틴 중쇄 또는 경쇄 등을 포함한다. 뇌 또는 기타 CNS 세포 발현의 경우, 적합한 프로모터는 다음을 포함할 수 있다: 모든 뉴런의 경우 시냅신1, 흥분성 뉴런의 경우 CaMKII알파, GABA성 뉴런의 경우 GAD67 또는 GAD65 또는 VGAT 등을 포함할 수 있다. 간 세포 발현의 경우, 적합한 프로모터는 알부민 프로모터를 포함한다. 폐 세포 발현의 경우, 적합한 프로모터는 SP-B를 포함할 수 있다. 내피 세포의 경우, 적합한 프로모터는 ICAM을 포함할 수 있다. 조혈 세포의 경우, 적합한 프로모터는 IFN베타 또는 CD45를 포함할 수 있다. 조골세포의 경우, 적합한 프로모터는 OG-2를 포함할 수 있다.

[0533] 일부 경우에, 본 개시내용의 Cas9는 별개의 프로모터가 동일 핵산 분자 내에서 염기 편집기 및 상용가능한 가이드 핵산의 발현을 구동할 수 있도록 하는 충분히 작은 크기를 가진다. 예를 들어, 벡터 또는 바이러스 벡터는 염기 편집기를 암호화하는 핵산에 작동 가능하게 연결된 제1 프로모터 및 가이드 핵산에 작동 가능하게 연결된 제2 프로모터를 포함할 수 있다.

[0534] 가이드 핵산의 발현을 구동시키기 위해 사용되는 프로모터는 다음을 포함할 수 있다: U6와 같은 Pol III 프로모터 또는 Pol II 프로모터의 HI Use 및 gRNA 아데노 연관 바이러스(AAV)를 발현시키기 위한 인트론 카세트.

[0535] 1개 이상의 가이드 핵산이 있거나 없는 본원에 기술된 Cas9는 아데노 연관 바이러스(AAV), 렌티바이러스, 아데노바이러스 또는 기타 플라스미드 또는 바이러스 벡터 유형을 사용하여, 특히 예를 들어 미국 특허 제8,454,972호(아데노바이러스에 대한 제형, 용량), 미국 특허 제8,404,658호(AAV에 대한 제형, 용량) 및 미국 특허 제5,846,946호(DNA 플라스미드에 대한 제형, 용량) 및 렌티바이러스, AAV 및 아데노바이러스를 수반하는 임상 시험 및 상기 임상 시험에 관한 출판물의 제형 및 용량을 사용하여 전달될 수 있다. 예를 들어, AAV의 경우 투여 경로, 제형 및 용량은 미국 특허 제8,454,972호 및 AAV를 수반하는 임상 시험에서와 같을 수 있다. 아데노바이러스의 경우, 투여 경로, 제형 및 용량은 미국 특허 제8,404,658호 및 아데노바이러스를 수반하는 임상 시험에서와 같을 수 있다. 플라스미드 전달의 경우, 투여 경로, 제형 및 용량은 미국 특허 제5,846,946호 및 플라스미드를 수반하는 임상 시험에서와 같을 수 있다. 용량은 평균 70kg 개체(성인 남성 인간)을 기준으로 하거나 이에 따라 추정될 수 있고, 상이한 체중 및 종의 환자, 대상체, 포유동물에 대해 조정될 수 있다. 투여 빈도는 연령, 성별, 전반적인 건강 상태, 환자 또는 대상체의 기타 병태, 및 해결되는 구체적인 병태 또는 증상을 포함한 일반적인 요인에 따라 의료 또는 가축병 담당자(예: 의사, 수의사)의 범위 내에 있다. 바이러스 벡터는 관심 조직에 주사될 수 있다. 세포-유형 특이적 염기 편집의 경우, 염기 편집기 및 선택적 가이드 핵산의 발현은 세포-유형 특이적 프로모터에 의해 구동될 수 있다.

[0536] 생체내 전달의 경우, AAV는 다른 바이러스 벡터보다 유리할 수 있다. 일부 경우에 AAV가 낮은 독성을 허용하는데, 이는 면역 반응을 활성화시킬 수 있는 세포 입자의 초원심분리를 필요로 하지 않는 정제 방법 때문일 수 있다. 일부 경우에, AAV가 숙주 게놈 내로 통합되지 않기 때문에 삽입 돌연변이를 유발할 낮은 가능성을 가질 수 있다.

[0537] AAV는 4.5 또는 4.75Kb의 패키징 제한을 가진다. 4.5 또는 4.75Kb보다 큰 작제물은 유의하게 낮아진 바이러스 생성을 이끌 수 있다. 예를 들어 SpCas9는 상당히 크고 유전자 자체가 4.1Kb를 초과하므로 AAV 내로 채워지기가 어렵다. 따라서, 본 개시내용의 구현예는 종래의 Cas9보다 더 짧은 길이를 가지는 개시된 Cas9를 활용하는 것을 포함한다.

[0538] AAV는 AAV1, AAV2, AAV5 또는 이들의 임의의 조합일 수 있다. 표적으로 될 세포와 관련하여 AAV 유형을 선택할 수 있다: 예를 들어, 뇌 또는 신경 세포를 표적으로 하기 위해 AAV 혈청형 1, 2, 5 또는 혼성 캡시드 AAV1, AAV2, AAV5 또는 이들의 임의의 조합을 선택할 수 있고; 심장 조직을 표적으로 하기 위해 AAV4를 선택할 수 있다. AAV8은 간으로의 전달에 유용한다. 이들 세포에 관한 특정 AAV 혈청형의 도표는 Grimm, D. 등, J. Virol. 82: 5887-5911 (2008))에서 찾을 수 있다.

[0539] 렌티바이러스는 유사분열 및 유사분열 후 세포 모두에서 유전자를 감염시키고 발현시키는 능력을 가지는 복합 레트로바이러스이다. 가장 흔히 알려진 렌티바이러스는 인간 면역결핍 바이러스(HIV)로, 이는 다른 바이러스의 외피 당단백질을 사용하여 광범위한 세포 유형을 표적으로 한다.

[0540] 렌티바이러스는 다음과 같이 제조될 수 있다. pCasES10(렌티바이러스 전달 플라스미드 골격을 함유함)을 클로닝

한 후, 낮은 계대(p=5)의 HEK293FT를 T-75 플라스크에서 10% 소 태아 혈청이 있고 항생제가 없는 DMEM에 형질감염 전날 50% 컨플루언스까지 씨딩했다. 20시간 후, 배지를 OptiMEM(무혈청) 배지로 교체하고 4시간 후에 형질감염을 수행했다. 세포는 10㎍의 렌티바이러스 전달 플라스미드(pCasES10) 및 다음 패키징 플라스미드로 형질감염된다: 5㎍의 pMD2.G(VSV-g 슈도타입) 및 7.5㎍의 psPAX2(gag/pol/rev/tat). 형질감염은 양이온성 지질 전달제(50㎕의 리포펙타민 2000 및 100ul 플러스 시약)가 있는 4mL OptiMEM에서 수행할 수 있다. 6시간 후, 배지를 10% 소 태아 혈청이 있는 무항생제 DMEM으로 교체한다. 이들 방법은 세포 배양 동안에 혈청을 사용하지만, 무혈청 방법이 바람직하다.

[0541] 렌티바이러스는 다음과 같이 정제될 수 있다. 48시간 후에 바이러스 상등액이 수확된다. 먼저 상등액에서 잔해물을 제거하고 0.45㎛ 저단백질 결합(PVDF) 필터를 통해 여과한다. 그 다음, 이들은 초원심분리기에서 24,000rpm으로 2시간 동안 회전된다. 바이러스 펠릿을 4℃에서 하룻밤 동안 50㎕의 DMEM에 재현탁한다. 그 다음, 이를 분취하고 즉시 -80℃에서 냉동한다.

[0542] 또 다른 구현예에서, 말 감염성 빈혈 바이러스(EIAV)에 기반한 최소 비영장류 렌티바이러스 벡터도 고려된다. 또 다른 구현예에서, 망막하 주사를 통해 전달될 것으로 고려되는, 혈관신생억제 단백질 엔도스타틴 및 안지오스타틴을 발현하는 말 감염성 빈혈 바이러스 기반 렌티바이러스 유전자 요법 벡터인 RetinoStat®. 또 다른 구현예에서, 자가-불활성화 렌티바이러스 벡터의 사용이 고려된다.

[0543] 시스템의 임의의 RNA, 예를 들어 가이드 RNA 또는 Cas9-암호화 mRNA는 RNA 형태로 전달될 수 있다. Cas9 암호화 mRNA는 시험관내 전사를 사용하여 발생될 수 있다. 예를 들어, Cas9 mRNA는 다음 요소를 함유하는 PCR 카세트를 사용하여 합성될 수 있다: T7 프로모터, 선택적 코작(kozak) 서열(GCCACC), 뉴클레아제 서열 및 3' UTR, 예컨대 베타 글로빈-폴리A 꼬리의 3' UTR. 카세트는 T7 폴리머라제에 의한 전사를 위해 사용될 수 있다. 가이드 폴리뉴클레오티드(예를 들어, gRNA)는 또한 T7 프로모터, 이어서 서열 "GG" 및 가이드 폴리뉴클레오티드 서열을 함유하는 카세트로부터 시험관내 전사를 사용하여 전사될 수 있다.

[0544] 발현을 증강시키고 독성 가능성을 낮추기 위해 Cas9 서열 및/또는 가이드 핵산은 예를 들어 슈도-U 또는 5-메틸-C를 사용하여 하나 이상의 변형된 뉴클레오시드를 포함하도록 변형될 수 있다.

[0545] 일부 구현예의 개시내용은 세포 또는 유기체를 변형시키는 방법을 이해한다. 세포는 원핵 세포 또는 진핵 세포일 수 있다. 세포는 포유동물 세포일 수 있다. 포유동물 세포는 인간이 아닌 영장류, 소, 돼지, 설치류 또는 마우스 세포일 수 있다. 본 개시내용의 염기 편집기, 조성물 및 방법에 의해 세포에 도입되는 변형은 항체, 전분, 알코올 또는 기타 원하는 세포 배출물과 같은 생물학적 생성물의 향상된 생산을 위해 세포 및 세포 자손이 변경되도록 하는 것일 수 있다. 본 개시내용의 방법에 의해 세포에 도입되는 변형은 생산된 생물학적 생성물을 변화시키는 변경을 세포 및 세포 자손이 포함하도록 하는 것일 수 있다.

[0546] 시스템은 하나 이상의 상이한 벡터를 포함할 수 있다. 일 측면에서, Cas9는 원하는 세포 유형, 우선적으로는 진핵 세포, 바람직하게는 포유동물 세포 또는 인간 세포에서 발현되도록 코돈 최적화된다.

[0547] 일반적으로, 코돈 최적화는 천연 서열의 적어도 1개의 코돈(예를 들어 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 이상의 코돈)을, 천연 아미노산 서열을 유지하면서 해당 숙주 세포의 유전자에서 더 빈번하게 또는 가장 빈번하게 사용되는 코돈으로 대체함으로써 관심 숙주 세포에서의 증강된 발현을 위해 핵산 서열을 변형시키는 프로세스를 의미한다. 다양한 종은 특정 아미노산의 일부 코돈에 대해 특정 편향을 나타낸다. 코돈 편향(유기체 간의 코돈 용도의 차이)은 특히 번역되는 코돈의 특성 및 특정 전달 RNA(tRNA) 분자의 가용성에 따라 달라지는 것으로 생각되는 메신저 RNA(mRNA)의 번역 효율성과 종종 관련된다. 세포에서 선택된 tRNA의 우세는 일반적으로 펩티드 합성에서 가장 빈번하게 사용되는 코돈을 반영한 것이다. 따라서, 유전자는 코돈 최적화에 기반하여 주어진 유기체에서 최적의 유전자 발현을 위해 맞춤형으로 될 수 있다. 코돈 용도 표는 예를 들어 www.kazusa.orjp/codon/(2002년 7월 9일 방문)에서 입수 가능한 "코돈 용도 데이터베이스(Codon Usage Database)"에서 쉽게 입수 가능할 수 있으며, 이들 표는 다양한 방식으로 각색될 수 있다. Nakamura, Y., 등 "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000) 참조. 특정 숙주 세포에서 발현을 위해 특정 서열을 코돈 최적화하기 위한 컴퓨터 알고리즘, 예를 들어 Gene Forge(Aptagen; Jacobus, PA)도 입수 가능한다. 일부 구현예에서, 조작된 뉴클레아제를 암호화하는 서열에서의 1개 이상의 코돈(예를 들어 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 이상 또는 모든 코돈)은 특정 아미노산에 대해 가장 빈번하게 사용되는 코돈에 해당한다.

[0548] 패키징 세포는 일반적으로 숙주 세포를 감염시킬 수 있는 바이러스 입자를 형성하기 위해 사용된다. 이러한 세

포는 아데노바이러스를 패키징하는 293 세포, 및 레트로바이러스를 패키징하는 psi.2 세포 또는 PA317 세포를 포함한다. 유전자 요법에 사용되는 바이러스 벡터는 일반적으로 핵산 벡터를 바이러스 입자 내로 패키징하는 세포주를 생성시킴으로써 발생된다. 벡터는 일반적으로 패키징 및 후속적인 숙주 내로의 통합에 필요한 최소 바이러스 서열을 함유하며, 다른 바이러스 서열은 발현될 폴리뉴클레오티드(들)를 위한 발현 카세트로 대체된다. 누락된 바이러스 기능은 일반적으로 패키징 세포주에 의해 트랜스에서 공급된다. 예를 들어, 유전자 요법에서 사용되는 AAV 벡터는 일반적으로 패키징 및 숙주 게놈 내로의 통합에 필요한 AAV 게놈으로부터의 ITR 서열만을 보유한다. 바이러스 DNA는 다른 AAV 유전자를 암호화하는 헬퍼 플라스미드, 즉 rep 및 cap을 함유하지만 ITR 서열은 결여된 세포주에 패키징될 수 있다. 세포주는 헬퍼로서 아데노바이러스에 의해 감염될 수도 있다. 헬퍼 바이러스는 AAV 벡터의 복제와 헬퍼 플라스미드로부터 AAV 유전자 발현을 촉진할 수 있다. 일부 경우에서 헬퍼 플라스미드는 ITR 서열 결여로 인해 상당한 양으로 패키징되지 않는다. 아데노바이러스에 의한 오염은 예를 들어 열처리에 의해 낮아질 수 있으며, 아데노바이러스는 AAV보다 열처리에 더 민감하다.

[0549] **약제학적 조성물**

[0550] 본 개시내용의 다른 측면은 CRISPR 시스템(예를 들어, 본원에 개시된 Cas9 포함)을 포함하는 약제학적 조성물에 관한 것이다. 본원에서 사용될 때 용어 "약제학적 조성물"은 약제학적 용도로 제형화된 조성물을 의미한다. 일부 구현예에서, 약제학적 조성물은 약제학적으로 허용 가능한 담체를 추가로 포함한다. 일부 구현예에서, 약제학적 조성물은 추가적인 작용제(예를 들어, 특이적 전달용, 반감기 증가 또는 기타 요법적 화합물)를 포함한다.

[0551] 본원에서 사용될 때 용어 "약제학적으로 허용 가능한 담체"는 약제학적으로 허용가능한 물질, 조성물 또는 비히클, 예컨대 액체 또는 고체 충전제, 희석제, 부형제, 제조 보조제(예: 윤활제, 활석 마그네슘, 칼슘 또는 아연 스테아레이트 또는 스테아르산), 또는 신체의 한 부위(예: 전달 부위)에서 다른 부위(예: 신체의 기관, 조직 또는 일부)로 화합물을 운반하거나 수송하는 데 관여하는 용매 캡슐화 물질을 의미한다. 약제학적으로 허용 가능한 담체는 제형의 다른 성분과 상용성이 있고 대상체의 조직에 해를 가하지 않는다는 의미에서 "허용 가능"하다(예를 들어, 생리학적으로 상용성, 무균성, 생리학적 pH 등).

[0552] 약제학적으로 허용 가능한 담체로서 작용할 수 있는 물질의 일부 비제한적인 예는 다음을 포함한다: (1) 당, 예컨대 락토오스, 글루코스 및 수크로스; (2) 전분, 예컨대 옥수수 전분 및 감자 전분; (3) 셀룰로오스 및 그의 유도체, 예컨대 소듐 카복시메틸 셀룰로오스, 메틸셀룰로오스, 에틸 셀룰로오스, 미정질 셀룰로오스 및 셀룰로오스 아세테이트; (4) 분말 트라가칸트; (5) 맥아; (6) 젤라틴; (7) 윤활제, 예컨대 마그네슘 스테아레이트, 소듐 라우릴 설페이트 및 활석; (8) 부형제, 예컨대 코코아 버터 및 좌제 왁스; (9) 오일, 예컨대 낙화생유, 면실유, 홍화유, 참기름, 올리브유, 옥수수유 및 대두유; (10) 글리콜, 예컨대 프로필렌 글리콜; (11) 폴리올, 예컨대 글리세린, 소르비톨, 만니톨 및 폴리에틸렌 글리콜(PEG); (12) 에스테르, 예컨대 에틸 올레에이트 및 에틸 라우레이트; (13) 한천; (14) 완충제, 예컨대 마그네슘 하이드록사이드 및 알루미늄 하이드록사이드; (15) 알긴산; (16) 발열원이 없는 물; (17) 등장성 식염수; (18) 링거 용액; (19) 에틸 알코올; (20) pH 완충 용액; (21) 폴리에스테르, 폴리카보네이트 및/또는 폴리무수물; (22) 증량제, 예컨대 폴리펩티드 및 아미노산 (23) 혈청 알코올, 예컨대 에탄올; 및 (23) 약제학적 제형에 이용되는 기타 무독성 상용성 물질. 습윤제, 착색제, 이형제, 코팅제, 감미제, 향미제, 향료, 보존제 및 항산화제도 제형에 존재할 수 있다. "부형제", "담체", "약제학적으로 허용 가능한 담체", "비히클" 등과 같은 용어는 본원에서 상호 교환적으로 사용된다.

[0553] 약제학적 조성물은 약 5.0 내지 약 8.0 범위와 같은 생리학적 pH를 반영하는 사전 결정된 수준에서 제형의 pH를 유지하기 위해 하나 이상의 pH 완충 화합물을 포함할 수 있다. 수성 액체 제형에 사용되는 pH 완충 화합물은 아미노산 또는 아미노산 혼합물, 예컨대 히스티딘 또는 히스티딘과 글리신과 같은 아미노산의 혼합물일 수 있다. 대안적으로, pH 완충 화합물은 바람직하게는 제형의 pH를 사전 결정된 수준, 예컨대 약 5.0 내지 약 8.0 범위로 유지하고 칼슘 이온을 킬레이트화하지 않는 제형이다. 이러한 pH 완충 화합물의 예시적인 예는 이미다졸 및 아세테이트 이온을 포함하지만 이에 국한되지는 않는다. pH 완충 화합물은 제형의 pH를 사전 결정된 수준으로 유지하기 위해 적합한 임의의 양으로 존재할 수 있다.

[0554] 또한 약제학적 조성물은 하나 이상의 삼투성 조절제, 즉 제형의 삼투 특성(예를 들어 긴장성, 삼투질농도 및/또는 삼투압)을 수령자 개체의 혈류 및 혈액 세포에 허용 가능한 수준으로 조절하는 화합물을 함유할 수 있다. 삼투성 조절제는 칼슘 이온을 킬레이트화하지 않는 작용제일 수 있다. 삼투성 조절제는 제형의 삼투 특성을 조절하는, 당업자에게 알려져 있거나 이용가능한 임의의 화합물일 수 있다. 당업자는 본 발명의 제형에 사용하기 위한 주어진 삼투성 조절제의 적합성을 실증적으로 결정할 수 있다. 삼투성 조절제의 적합한 유형의 예시적인 예는 다음을 포함하지만 이에 국한되지는 않는다: 염, 예컨대 염화나트륨 및 아세트산나트륨; 당, 수크로스, 덱스

트로스 및 만니톨; 아미노산, 예컨대 글리신; 및 이들 작용제 및/또는 작용제 유형의 하나 이상의 혼합물. 삼투성 조절제(들)는 제형의 삼투 특성을 조절하기에 충분한 임의의 농도로 존재할 수 있다.

[0555] 일부 구현예에서, 약제학적 조성물은 예를 들어 유전자 편집을 위해 대상체에게 전달하기 위해 제형화된다. 본원에 기술된 약제학적 조성물을 투여하는 적합한 경로는 국소, 피하, 경피, 피내, 병변내, 관절내, 복강내, 방광내, 점막경유, 치은, 치아내, 와우내, 고막경유, 기관내, 경막외, 척수강내, 근육내, 정맥내, 혈관내, 골내, 눈주위, 종양내, 뇌내 및 뇌실내 투여를 포함한다.

[0556] 일부 구현예에서, 본원에 기술된 약제학적 조성물은 질병이 있는 부위에 국소 투여된다. 일부 구현예에서, 본원에 기술된 약제학적 조성물은 주사에 의해, 카테터에 의해, 좌제에 의해 또는 다공성, 비다공성 또는 젤라틴성 물질로 이루어지고 실라스틱 멤브레인과 같은 멤브레인 또는 섬유를 포함하는 임플란트에 의해 대상체에게 투여된다.

[0557] 다른 구현예에서, 본원에 기술된 약제학적 조성물은 방출 조절 시스템으로 전달된다. 일 구현예에서, 펌프가 사용될 수 있다(예를 들어, Langer, 1990, Science 249: 1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald 등, 1980, Surgery 88:507; Saudek 등, 1989, N. Engl. J. Med. 321:574 참조). 또 다른 구현예에서는 중합체 물질이 사용될 수 있다. (예를 들어, Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61 참조. 또한 Levy 등, 1985, Science 228: 190; During 등, 1989, Ann. Neurol. 25:351; Howard et ah, 1989, J. Neurosurg. 71: 105 참조.) 다른 방출 조절 시스템은 예를 들어 전술한 Langer에서 검토된다.

[0558] 일부 구현예에서, 약제학적 조성물은 대상체, 예를 들어 인간에게 정맥내 또는 피하 투여하도록 적합화된 조성물로서 통상적인 절차에 따라 제형화된다. 일부 구현예에서, 주사 투여를 위한 약제학적 조성물은 가용화제 및 주사 부위의 통증을 완화시키는 리그노카인과 같은 국소마취제로서 무균 등장성으로 사용되는 용액이다. 일반적으로, 성분은 개별적으로 공급되거나 단위 투여 형태로 혼합되어, 예를 들어 활성제의 양을 나타내는 앰플 또는 봉지와 같이 완전 밀봉 용기에 담긴 건조 동결건조 분말 또는 물이 없는 농축물로서 공급된다. 약제가 주입에 의해 투여되는 경우, 무균의 의약품 등급 물이나 식염수를 함유하는 주입병(infusion bottle)과 함께 제공될 수 있다. 약제학적 조성물이 주사로 투여되는 경우, 투여 전에 성분이 혼합될 수 있도록 무균 주사용수 또는 식염수의 앰플이 제공될 수 있다.

[0559] 전신 투여를 위한 약제학적 조성물은 액체, 예를 들어 무균 식염수, 젖산 링거액 또는 행크액일 수 있다. 또한, 약제학적 조성물은 고체 형태일 수 있으며, 사용 직전에 재용해되거나 현탁될 수 있다. 동결건조된 형태도 고려된다. 약제학적 조성물은 비경구 투여에도 적합한 리포솜 또는 미세결정과 같이 지질 입자 또는 소포 내에 함유될 수 있다. 조성물이 입자 내에 함유되어 있는 한, 입자는 단층 또는 다층과 같은 임의의 적합한 구조일 수 있다. 화합물은, 융합성 지질 디올레오일포스파티딜에탄올아민(DOPE), 낮은 수준(5 내지 10mol%)의 양이온성 지질을 함유하고 폴리에틸렌글리콜(PEG) 코팅에 의해 안정화되는 "안정화된 플라스미드-지질 입자"(SPLP)에 포획될 수 있다(Zhang Y. P. et ah, Gene Ther. 1999, 6: 1438-47). N-[1-(2,3-디올레오일옥시)프로필]-N,N,N-트리메틸암모늄메틸설페이트 또는 "DOTAP"와 같은 양으로 하전된 지질이 이러한 입자 및 소포에 특히 바람직하다. 이러한 지질 입자의 제조는 잘 알려져 있다. 예를 들어, 미국 특허 제4,880,635호; 제4,906,477호; 제4,911,928호; 제4,917,951호; 제4,920,016호; 및 제4,921,757호를 참고하고, 이들 각각은 본원에 참고로 통합된다.

[0560] 본원에 기술된 약제학적 조성물은 예를 들어 단위 용량으로 투여되거나 포장될 수 있다. 본 개시내용의 약제학적 조성물과 관련하여 사용될 때 용어 "단위 용량"은 대상체에 대한 단일 투여량으로 적합하게 물리적으로 구분된 단위를 의미하며, 각 단위는 필요한 희석제; 즉, 담체 또는 비히클과 연관되어 원하는 치료 효과를 생성시키도록 계산된 사전 결정된 양의 활성 물질을 함유한다.

[0561] 추가로, 약제학적 조성물은 (a) 동결건조 형태의 본 발명의 화합물을 함유하는 용기 및 (b) (본 발명의 동결건조 화합물의 재구성 또는 희석을 위해 사용되는 무균의) 약제학적으로 허용 가능한 희석제를 함유하는 제2 용기를 포함하는 약제학적 키트로 제공될 수 있다. 의약품 또는 생물학적 제품의 제조, 사용 또는 판매를 규제하는 정부 기관이 규정한 양식의 안내문이 이러한 용기(들)과 선택적으로 연관될 수 있고, 이 안내문은 인간 투여용의 제조, 사용 또는 판매에 대한 기관 승인을 반영한다.

[0562] 또 다른 측면에서, 상기에 기술된 질병 치료에 유용한 물질을 함유하는 제조 물품이 포함된다. 일부 구현예에서, 제조 물품은 용기와 라벨을 포함한다. 적합한 용기는 예를 들어 병, 바이알, 주사기 및 시험관을 포함한다. 용기는 유리 또는 플라스틱과 같은 다양한 물질로부터 형성될 수 있다. 일부 구현예에서, 용기는 본원에 기술된 질병을 치료하는 데 효과적인 조성물을 보유하고, 무균 접근 포트를 가질 수 있다. 예를 들어, 용기는 정맥 주사액 백 또는 피하 주사 바늘이 관통될 수 있는 마개가 있는 바이알일 수 있다. 조성물 중 활성제는 본 발명의 화합물이다. 일부 구현예에서, 용기 상에 있거나 용기와 연관된 라벨은 조성물이 선택된 질병을 치료하기 위해 사용됨을 나타낸다. 제조 물품은 인산염 완충 식염수, 링거액 또는 덱스트로스 용액과 같이 약제학적으로 허용 가능한 완충제를 포함하는 제2 용기를 추가로 포함할 수 있다. 이는 다른 완충제, 희석제, 필터, 바늘, 주사기 및 사용 설명서가 있는 포장 인서트를 포함하여 상업적 및 사용자 관점에서 바람직한 다른 물질을 추가로 포함할 수 있다.

[0563] 일부 구현예에서, CRISPR 시스템(예를 들어, 본원에 기술된 Cas9 포함)은 약제학적 조성물의 일부로서 제공된다. 일부 구현예에서, 약제학적 조성물은 본원에 제공된 임의의 융합 단백질(예를 들어, LubCas9를 포함하는 본원에 기술된 핵염기 편집기 포함)을 포함한다. 일부 구현예에서, 약제학적 조성물은 본원에 제공된 임의의 복합체를 포함한다. 일부 구현예에서, 약제학적 조성물은 gRNA 및 양이온성 지질과의 복합체를 형성하는 RNA-가이드 뉴클레아제(예를 들어, Cas9)를 포함하는 리보핵단백질 복합체를 포함한다. 일부 구현예에서 약제학적 조성물은 gRNA, 핵산 프로그래밍 가능한 DNA 결합 단백질, 양이온성 지질 및 약제학적으로 허용 가능한 부형제를 포함한다. 약제학적 조성물은 하나 이상의 추가적인 치료적 활성 서브스턴스를 선택적으로 포함할 수 있다.

[0564] **키트**

[0565] 일 측면에서, 본 발명은 상기 방법 및 조성물에 개시된 요소 중 임의의 하나 이상을 함유하는 키트를 제공한다. 일부 구현예에서, 키트는 벡터 시스템 및 키트 사용을 위한 설명서를 포함한다. 일부 구현예에서, 벡터 시스템은 가이드 서열을 삽입하기 위한 하나 이상의 삽입 부위를 포함하며, 상기 가이드 서열은 발현될 때 진핵 세포의 표적 서열에 대하여 CRISPR 복합체의 서열-특이적 결합을 지시하고, 상기 CRISPR 복합체는 (1) 표적 서열에 혼성화된 가이드 서열 및 (2) tracr 서열에 혼성화되는 서열과 복합체를 형성하는 CRISPR 효소; 및/또는 (b) 핵 위치 서열을 포함하는 상기 CRISPR 효소를 암호화하는 효소-암호화 서열에 작동 가능하게 연결된 제2 조절 요소를 포함한다. 요소는 개별적으로 또는 조합하여 제공될 수 있으며, 바이알, 병 또는 튜브와 같은 임의의 적합한 용기에서 제공될 수 있다. 일부 구현예에서, 키트는 하나 이상의 언어, 예를 들어 1개 초과의 언어로 작성된 설명서를 포함한다.

[0566] 일부 구현예에서, 키트는 핵염기 편집기를 포함한다. 예를 들어, 일부 구현예에서, 키트는 본원에 기술된 Cas9 효소(ScoCas9, SirCas9, VapCas9, EpeCas9, LfeCas9, PmaCas9)를 포함하는 핵염기 편집기를 포함한다.

[0567] 일부 구현예에서, 키트는 본원에 기술된 요소 중 하나 이상을 활용하는 프로세스에서 사용하기 위한 하나 이상의 시약을 포함한다. 시약은 임의의 적합한 용기에 제공될 수 있다. 예를 들어, 키트는 하나 이상의 반응 또는 저장 완충제를 제공할 수 있다. 시약은 특정 분석에 사용할 수 있는 형태로 또는 사용 전에 하나 이상의 다른 구성성분의 추가를 필요로 하는 형태(예: 농축물 또는 동결건조 형태)로 제공될 수 있다. 완충제는 탄산나트륨 완충제, 중탄산나트륨 완충제, 붕산염 완충제, 트리스 완충제, MOPS 완충제, HEPES 완충제 및 이들의 조합을 포함하지만 이에 국한되지 않는 임의의 완충제일 수 있다. 일부 구현예에서, 완충제는 알칼리성이다. 일부 구현예에서, 완충제는 약 7 내지 약 10의 pH를 가진다. 일부 구현예에서, 키트는 가이드 서열과 조절 요소를 작동 가능하게 연결하기 위해 벡터 내로 삽입하기 위한 가이드 서열에 상응하는 하나 이상의 올리고뉴클레오티드를 포함한다. 일부 구현예에서, 키트는 상동성 재조합 주형 폴리뉴클레오티드를 포함한다.

[0568] 본원에 언급된 모든 출판물, 특허 출원, 특허 및 기타 참고 문헌은 전체적으로 참고로 통합된다. 또한, 물질, 방법 및 실시예는 예시일 뿐이고 제한하려는 의도는 아니다. 달리 정의하지 않는 한, 본원에서 사용되는 모든 기술적 및 과학적 용어는 본 발명이 속하는 기술분야의 통상의 기술자가 일반적으로 이해하는 것과 동일한 의미를 가진다. 본원에 기술된 것과 유사하거나 동등한 방법 및 물질이 본 발명의 실시 또는 테스트에서 사용될 수 있으며, 적합한 방법 및 물질이 본원에 기술되어 있다.

[0569] **실시예**

[0570] 다음 실시예는 본 발명을 제조하고 실시하는 바람직한 방식 중 일부를 설명한다. 그러나, 이들 실시예는 단지 예시를 위한 것이며, 본 발명의 범위를 제한하려는 의도가 아닌 것으로 이해되어야 한다.

[0571] 실시예 1. *신규 Cas9 효소 스크리닝, 신규 Cas9 효소의 발견 및 최적화*

[0572] 이 실시예는 신규 Cas9 효소를 발견하기 위한 스크리닝을 기술한다. 본원에 기술된 바와 같이, 이 스크리닝을 사용하여 *스트렙토코커스 콘스텔라투스, 샤르페아 spp. 단리물 RUG017, 베일로넬라 파르불라, 에자키엘라 페루엔시스, 락토바실러스 퍼멘텀 균주 AF15-40LB* 및 *펩토니필러스 sp. 마르세유-P3761* 박테리아 유래의 신규 Cas9 효소를 단리하고 최적화하였다.

[0573] 신규 PAM 서열을 인식하는 신규 Cas9 효소를 발견하기 위한 검색에서 생물정보학 스크리닝을 사용하여 추가 효소를 검색하여서 CRISPR 표적 범위를 확장하였다. 스크리닝은 *S. 파오게네스, S. 아우레우스, S. 써모필러스* 및 *F. 노비시다*의 Cas9의 씨드 서열을 활용했다. BLAST 적중(hit)을 고려하기 위해 e-값 임계값이 1e-6인 BLAST의 tblastn 변형을 사용하여 생물정보학을 수행했다. 즉, 테스트를 위해 선택된 유전자좌는 다른 종의 Cas9 단백질의 존재 하에서 완전하게 유지되는 유전자좌이었다. CRISPR 어레이 내에 3개보다 많은 스페이서를 갖고 Cas9의 1kb 내인성 서열 5'보다 크고 CRISPR 어레이의 300nt 3'보다 큰 유전자좌를 선택하였다. 이 접근법을 사용하여, 신규 Cas 효소는 인간 세포에서 발현되도록 최적화된 코돈 및 상이한 박테리아 종으로부터 식별하였다. 그 다음, 신규 조작된 Cas9 효소를 재조합에 의해 생성하였고 테스트하였다.

[0574] *실시예 2. 스트렙토코커스 콘스텔라투스, 샤르페아 spp. 단리물 RUG017, 베일로넬라 파르불라, 에자키엘라 페루엔시스, 락토바실러스 퍼멘텀 균주 AF15-40LB 및 펩토니필러스 sp. 마르세유-P3761 박테리아 유래 신규 Cas9 효소에 대한 3'PAM 공통 모티프 식별*

[0575] 이 실시예는 *스트렙토코커스 콘스텔라투스, 샤르페아 spp. 단리물 RUG017, 베일로넬라 파르불라, 에자키엘라 페루엔시스, 락토바실러스 퍼멘텀 균주 AF15-40LB* 및 *펩토니필러스 sp. 마르세유-P3761* 균주로부터 최초 단리된 인간 코돈-최적화된 Cas9에 대한 프로토스페이서 인접 모티프(PAM) 서열의 식별을 예시한다.

[0576] 시험관내 PAM 식별 분석을 사용하여 PAM 서열 인식에 대해 인간 코돈-최적화된 Cas9를 테스트하였다. 무작위 PAM 서열을 구비한 플라스미드 라이브러리를 상이한 박테리아에서 단리된 Cas9와 함께 항온처리했다. 절단되지 않은 플라스미드를 정제하고 서열 분석하여, 절단된 특정 PAM 모티프를 식별했다. *스트렙토코커스 콘스텔라투스* Cas9에 의해 인식되는 공통 PAM 서열은 5'-NGG-3'으로 식별되었다(도 1a). *샤르페아 spp. 단리물 RUG017* Cas9에 의해 인식되는 공통 PAM 서열은 5'-NAGHC-3'으로 식별되었다(도 1b). *베일로넬라 파르불라* Cas9에 의해 인식되는 공통 PAM 서열은 5'-NRHRRH-3'(H=A, C 또는 T; R=A 또는 G)로 식별되었다(도 1c). *에자키엘라 페루엔시스* Cas9에 의해 인식되는 공통 PAM 서열은 5'-NGG-3'으로 식별되었다(도 1d). *락토바실러스 퍼멘텀 균주 AF15-40LB* Cas9에 의해 인식되는 공통 PAM 서열은 5'-NNAAA-3'으로 식별되었다(도 1e). *펩토니필러스 sp. 마르세유-P3761* Cas9에 의해 인식되는 공통 PAM 서열은 5'-NGG-3'으로 식별되었다(도 1f).

[0577] *실시예 3. 스트렙토코커스 콘스텔라투스, 샤르페아 spp. 단리물 RUG017, 베일로넬라 파르불라, 에자키엘라 페루엔시스, 락토바실러스 퍼멘텀 균주 AF15-40LB 및 펩토니필러스 sp. 마르세유-P3761 박테리아 유래 신규 Cas9 효소에 대한 sgRNA의 RNA 접힘 구조의 예측*

[0578] 이 실시예는 신규 Cas9 효소와 함께 사용하기 위한 crRNA 및 tracrRNA를 포함하는 예시적인 sgRNA의 예측된 RNA 접힘 구조를 입증한다.

[0579] Cas9 Crispr 유전자좌를 이종적으로 발현하는 대장균 균주로부터 유래된 RNA 상에서 소형 RNA 서열분석을 수행했다. 즉, 먼저 트리아졸에 대장균을 재현탁시킨 후 1분 주기로 3회 동안 균질화기에서 지르코니아/실리카 비드를 사용하여 박테리아를 균질화함으로써 정지상 박테리아로부터 RNA를 단리했다. 균질화된 샘플로부터 전체 RNA를 정제하고, DNAse를 처리하고, T4 폴리뉴클레오티드 키나제로 3' 탈인산화시키고, rRNA를 제거했다. rRNA가 고갈된 RNA로부터 RNA 라이브러리를 제조하고 소형 RNA에 대해 크기를 선택했다.

[0580] RNA 서열분석을 위해 전사체는 대장균 폴리(A) 폴리머라제 꼬리를 가진 폴리-A였으며, T4 RNA 리가제 1을 사용하여 5' RNA 어댑터와 결찰되었고, 역전사한 다음 바코드 프라이머로 cDNA를 PCR 증폭하고, MiSeq 상에서 서열분석했다. 각 샘플의 판독값은 연관된 바코드를 기반으로 식별했고, BWA를 사용하여 참조 서열에 대해 정렬시켰다. Picard 도구를 사용하여 전사체 서열을 추출하기 위해 쌍을 이룬 정렬을 사용했고, Geneious 소프트웨어를 사용하여 서열을 분석했다.

[0581] RNA 접힘은 Geneious 11.1.2 소프트웨어로부터의 예측을 기반으로 했다. 단일 sgRNA 전사체는 부위-특이적 Cas9 활성을 안내하기 위해 필요한 이중 RNA 구조를 모방하면서 crRNA를 tracrRNA에 융합시킨다. *스트렙토코커스 콘스텔라투스*로부터의 ScoCas9와 함께 사용하기 위한 키메라 sgRNA에 대해 예측된 RNA 접힘 구조가 도 2a에 도시

되어 있고, *샤르페아 spp. 단리물 RUG017*로부터의 SirCas9와 함께 사용하기 위한 sgRNA가 도 2b에 도시되어 있고, *베일로넬라 파르불라*로부터의 VasCas9와 함께 사용하기 위한 sgRNA가 도 2c에 도시되어 있고, *에자키엘라 페루엔시스*로부터의 EpeCas9와 함께 사용하기 위한 sgRNA가 도 2d에 도시되어 있고, *락토바실러스 퍼멘텀 균주 AF15-40LB*로부터의 LfeCas9와 함께 사용하기 위한 sgRNA가 도 2e에 도시되어 있고, *펩토니필러스 sp. 마르세유-P3761*로부터의 PmaCas9와 함께 사용하기 위한 sgRNA가 도 2f에 도시되어 있다.

[0582] 실시예 4. *HEK293T 세포에서 WT ScoCas9에 의한 생체외 절단 활성*

[0583] 이 실시예는 HEK293T 세포에서 *스트렙토코커스 콘스텔라투스*로부터 WT *ScoCas9*에 의한 생체외 핵산 절단 활성을 예시한다.

[0584] HEK293T 세포를 96웰 플레이트에 플레이팅했다. 플레이팅 후 24시간에 Cas9 및 가이드 RNA를 함유하는 발현 벡터(표 10)로 세포를 형질감염시켰다. 형질감염 후 72시간에 세포를 수확하고 전체 DNA를 추출했다.

[0585] HEK293T 세포에서 삽입결실 패턴을 특징규명하기 위해 심층 서열분석을 수행했다. 즉, Illumina 어댑터 및 표적 앰플리콘에 대한 고유 바코드를 추가하는 2회 PCR을 사용하여 예시적인 표적(표 8)을 증폭했다. PCR 산물을 2% 겔 상에서 실행하고 겔을 추출했다. 샘플을 모으고(pooled), 정량화하고, cDNA 라이브러리를 제조하고, MiSeq 상에서 서열분석했다. 삽입결실 빈도는 심층 서열분석에 의해 결정했다(도 3).

[0586] **표 8. 예시적인 가이드 RNA 서열 및 PAM 서열**

[0587]

종류(Sco/Pma)	5'->3' 가이드 서열	3' PAM
가이드 2	GAAACAATGATAACAAGACC (서열번호 97)	TGG
가이드 3	GTGGCCCCTGTGCCCAGCCC (서열번호 98)	TGG
가이드 4	GTCCCAAATATGTAGCTGTT (서열번호 99)	TGG
가이드 6	GCTCCCATCACATCAACCGG (서열번호 100)	TGG
가이드 7	GATGTCACCTCCAATGACTA (서열번호 101)	GGG
가이드 9	GTTGAAGATGAAGCCCAGAG (서열번호 102)	CGG
가이드 10	GCCAACACCAACCAGAACTT (서열번호 103)	GGG
가이드 11	TGCTGCACACAGCAGGCCTT (서열번호 104)	TGG

[0588] 데이터는 WT ScoCas9가 2 내지 32% 삽입결실 빈도를 달성했음을 보여주었다. 가이드 RNA 2와 9는 30%를 초과하는 삽입결실 돌연변이를 초래한 반면, 가이드 RNA 11은 약 2%의 삽입결실 돌연변이를 초래했다.

[0589] 실시예 5. *아데닌 염기 편집기(ABE) 또는 시티딘 염기 편집기(CBE)의 N-말단 융합이 있는 Cas9 효소에 의한 염기 편집*

[0590] 이 실시예는 아데닌 염기 편집기(ABE) 또는 시티딘 염기 편집기(CBE)에 융합된 Cas9 효소의 염기 전환 효율을 예시한다.

[0591] 즉, 96웰당 25,000개의 HEK293T 세포를 플레이팅했다. 100ng의 Cas9 발현 플라스미드 및 100ng의 가이드 발현 플라스미드를 플레이팅 후 24시간에 형질감염시켰다. 형질감염 후 5일에 세포를 수확하고 DNA를 추출했다.

[0592] HEK293T 세포에서 A에서 G로 전환 또는 C에서 T로 전환을 특징규명하기 위해 심층 서열분석을 수행했다. Illumina 어댑터 및 표적 앰플리콘에 대한 고유 바코드를 추가하는 2회 PCR 영역을 사용하여 예시적인 표적을 증폭했다. PCR 산물을 2% 겔 상에서 실행하고 겔을 추출했다. 샘플을 모으고, 정량화하고, cDNA 라이브러리를 제조하고, MiSeq 상에서 서열분석했다. N-말단 및 C-말단 TadA8 융합 작제물에 대한 심층 서열분석으로 A에서 G로 전환율을 결정했다. N-말단 및 C-말단 ppAPOBEC1 융합 작제물에 대한 심층 서열분석으로 C에서 T로의 전환율을 결정했다.

[0593] 도 4a는 ABE 또는 CBE에 N-말단에서 융합된 *ScoCas9* 작제물의 개요를 보여준다. 표 9는 *ScoCas9*와 함께 사용되는 가이드 RNA 서열을 보여준다. 도 4b는 인간 세포주(HEK293T)에서 게놈 부위에 관한 아데닌 염기 편집기(ABE)에 대한 *ScoCas9*의 N-말단 융합으로 달성된 삽입결실 돌연변이 및 표적화된 아데닌에서 구아닌으로 전환율의 그래프를 보여준다(도 4b). 도 4c는 인간 세포주(HEK293T)에서 게놈 부위에 관한 시티딘 염기 편집기에 대한 *ScoCas9*의 N-말단 융합으로 달성된 삽입결실 돌연변이 및 표적화된 시토신에서 티민으로 전환율의 그래프를 보

여준다(도 4c).

[0594] **표 9. ScoCas9와 함께 사용되는 가이드 RNA 서열 및 PAM 서열**

종류 (Sco)	5'->3' 가이드 서열	3' PAM
가이드 1	GAACACAAAGCATAGACTGC (서열번호 105)	GGG
가이드 2	GAAACAATGATAACAAGACC (서열번호 106)	TGG
가이드 3	GTGGCCCCTGTGCCCAGCCC (서열번호 107)	TGG
가이드 4	GTCCCAAATATGTAGCTGTT (서열번호 108)	TGG
가이드 5	AGAGGGACACACAGATCTAT (서열번호 109)	TGG
가이드 6	GCTCCCATCACATCAACCGG (서열번호 110)	TGG
가이드 7	GATGTCACCTCCAATGACTA (서열번호 111)	GGG
가이드 8	GGGCAACCACAAACCCACGA (서열번호 112)	GGG
가이드 9	GTTGAAGATGAAGCCCAGAG (서열번호 113)	CGG
가이드 10	GCCAACACCAACCAGAACTT (서열번호 114)	GGG
가이드 11	TGCTGCACACAGCAGGCCTT (서열번호 115)	TGG
가이드 12	GTGCCAGAAACAGGGGTGAC (서열번호 116)	GGG

[0595]

[0596] 도 5a는 WT *SirCas9*뿐만 아니라 ABE에 N-말단에서 융합된 *SirCas9*("D14A" 돌연변이체)의 작제물의 개략도를 보여준다. 표 10은 *SirCas9*와 함께 사용되는 예시적인 NAGMC 가이드 RNA 서열을 보여준다. 도 5b는 인간 세포주(HEK293T)에서 게놈 부위에 관한 아데닌 염기 편집기(ABE)에 대한 *SirCas9*의 N-말단 융합으로 달성된 삽입결실 돌연변이 및 표적화된 아데닌에서 구아닌으로 전환율의 그래프를 보여준다(도 5b).

[0597] **표 10. SirCa9와 함께 사용되는 가이드 RNA 서열 및 PAM 서열**

종류 (Sir)	5'->3' 서열	3' PAM
가이드 1	GCCTCAGCTGCTCACTT (서열번호 117)	GAGCC
가이드 2	CGGTCCCCAGAGGGTTC (서열번호 118)	TAGAC
가이드 3	ACCGGTTGATGTGATGG (서열번호 119)	GAGCC
가이드 4	TGGTCCCAGGCCTCAGC (서열번호 120)	CAGCC
가이드 5	GAAAATGAAACTTTCAA (서열번호 121)	AAGCC
가이드 6	AACCCAACTCCATCTAC (서열번호 122)	CAGCC
가이드 7	CCTTGAATTGCAGTATC (서열번호 123)	TAGCC
가이드 8	TAGACTGCGGGGCGGGC (서열번호 124)	CAGCC
가이드 9	AACTGGAACACAAAGCA (서열번호 125)	TAGAC
가이드 10	AGCATGTGGTAATTTTC (서열번호 126)	CAGCC
가이드 11	CCCGGAAACTCTGTCCA (서열번호 127)	GAGAC
가이드 12	ACCCCCACCAAGGTTCA (서열번호 128)	CAGCC

[0598]

[0599] 도 6a는 WT VapCas9뿐만 아니라 ABE 또는 CBE에 N-말단에서 융합된 VapCas9("D38A" 돌연변이체)를 보여주는 작제물의 개략도를 보여준다. 표 11은 *VapCas9*와 함께 사용되는 예시적인 NRHRRH[여기서 H는 아데닌, 시토신 또는 티민이고, R은 아데닌 또는 구아닌임] 가이드 RNA 서열을 보여준다. 도 6b는 인간 세포주(HEK293T)에서 게놈 부위에 관한 아데닌 염기 편집기(ABE)에 대한 *VapCas9*의 N-말단 융합으로 달성된 삽입결실 돌연변이 및 표적화된 아데닌에서 구아닌으로 전환율뿐만 아니라 시티딘 염기 편집기(CBE)에 대한 *VapCas9*의 N-말단 융합으로 달성된 표적화된 시토신에서 티민으로 전환율의 그래프를 보여준다(도 6b).

[0600] **표 11. VapCas9와 함께 사용하기 위한 가이드 RNA 서열 및 PAM 서열**

종류 (Vap)	5'->3' 서열	3' PAM
가이드 1	TGTTAACAGCTGACCCAATA (서열번호 129)	AGTGGC
가이드 2	GTTACTCGCCTGTCAAGTGG (서열번호 130)	CGTGAC
가이드 3	GGGCTCCCATCACATCAACC (서열번호 131)	GGTGGC
가이드 4	GCTTTGGGGAGGCCTGGAGT (서열번호 132)	CATGGC
가이드 5	TAGCTGCCAATGACTATAGC (서열번호 133)	AATAGC
가이드 6	TTAAAATAGGATCTACATCA (서열번호 134)	CGTAAC
가이드 7	GAATCCTGCCATACACTTTG (서열번호 135)	AATAGC
가이드 8	CTGCGGGGCGGGCCAGCCTG (서열번호 136)	AATAGC
가이드 9	ACATTGTCAGAGGGACACAC (서열번호 137)	TGTGGC
가이드 10	AGCAACTCCAGTCCCAAATA (서열번호 138)	TGTAGC
가이드 11	GTGGTGGCCGAGCGCCCCCT (서열번호 139)	AGTGAC
가이드 12	CATTCACCCAGCTTCCCTGT (서열번호 140)	GGTGGC

[0601]

[0602] 도 7a는 VapCas9에 대한 ABE의 N-말단 융합 및 ABE의 C-말단 융합을 보여주는 작제물의 개략도를 보여준다. 도 7b는 아데닌 염기 편집기(ABE)에 대한 N-말단 융합 및 C-말단 융합으로 달성된 표적화된 아데닌에서 구아닌으로 전환율의 그래프를 보여준다.

[0603] 도 8a는 EpeCas9에 대한 ABE 및 CBE의 N-말단 융합을 보여주는 작제물의 개략도를 보여준다. 표 12는 *EpeCas9*와 함께 사용되는 예시적인 가이드 RNA 서열을 보여준다. 도 8b는 삽입결실 돌연변이 그래프, ABE에 대한 N-말단 융합으로 달성된 표적화된 아데닌에서 티민으로 전환율 및 CBE에 대한 N-말단 융합으로 달성된 표적화된 시토신에서 티민으로 전환율의 그래프를 보여준다.

[0604] **표 12. EpeCas9와 함께 사용하기 위한 가이드 RNA 서열 및 PAM 서열**

종류 (Epe)	서열	PAM
가이드 1	GAACACAAAGCATAGACTGC (서열번호 141)	GGG
가이드 2	GAAACAATGATAACAAGACC (서열번호 142)	TGG
가이드 3	GTGGCCCCTGTGCCCAGCCC (서열번호 143)	TGG
가이드 4	GTCCCAAATATGTAGCTGTT (서열번호 144)	TGG
가이드 5	AGAGGGACACACAGATCTAT (서열번호 145)	TGG
가이드 6	GCTCCCATCACATCAACCGG (서열번호 146)	TGG
가이드 7	GATGTCACCTCCAATGACTA (서열번호 147)	GGG
가이드 8	GGGCAACCACAAACCCACGA (서열번호 148)	GGG
가이드 9	GTTGAAGATGAAGCCCAGAG (서열번호 149)	CGG
가이드 10	GCCAACACCAACCAGAACTT (서열번호 150)	GGG
가이드 11	TGCTGCACACAGCAGGCCTT (서열번호 151)	TGG
가이드 12	GTGCCAGAAACAGGGGTGAC (서열번호 152)	GGG

[0605]

[0606] 도 9a는 WT LfeCas9 및 ABE와 CBE에 N-말단에서 융합된 LfeCas9 D9A 돌연변이체를 보여주는 작제물의 개략도를 보여준다. 표 13은 *LfeCas9*와 함께 사용되는 예시적인 가이드 RNA 서열을 보여준다. 도 9b는 LfeCas9를 사용하여 달성된 삽입결실 돌연변이 빈도의 결과를 보여주는 그래프를 보여준다. 도 9c는 아데닌 염기 편집기에 대한 LfeCas9의 N-말단 융합으로 달성된 표적화된 아데닌에서 구아닌으로 전환 그래프를 보여준다. 도 9d는 LfeCas9 D9A 돌연변이체의 N-말단에 융합된 CBE를 포함하는 염기 편집기를 사용하여 달성된 표적화된 시토신에서 티민으로 전환 그래프를 보여준다.

[0607] **표 13. LfeCas9와 함께 사용하기 위한 가이드 RNA 서열 및 PAM 서열**

종류 (Lfe)	서열	PAM
가이드 1	TCACGGAGACTGAACACTCC (서열번호 153)	TCAAA
가이드 2	GTAACAGACATGGACCATCA (서열번호 154)	GGAAA
가이드 3	GGGAGGGAGGGGCACAGATG (서열번호 155)	AGAAA
가이드 4	TGTGGTTCCAGAACCGGAGG (서열번호 156)	ACAAA
가이드 5	AATGAGAGAAAATGAAACTT (서열번호 157)	TCAAA
가이드 6	GGCCATCAAGGATGCCCACG (서열번호 158)	AGAAA
가이드 7	AAATTGTCCAGCCCCATCTG (서열번호 159)	TCAAA
가이드 8	CCTGTAAAGGAAACTGGAAC (서열번호 160)	ACAAA
가이드 9	TACATGAAGCAACTCCAGTC (서열번호 161)	CCAAA
가이드 10	AAACTCCCCCCACCCCCTTT (서열번호 162)	CCAAA
가이드 11	GAGTTGGGTTTGGTGCTCAA (서열번호 163)	TGAAA
가이드 12	GCGGGCCAGCCTGAATAGCT (서열번호 164)	GCAAA

[0608]

[0609] 도 10a는 WT PmaCas9 및 ABE와 CBE에 N-말단과 C-말단에서 융합된 PmaCas9 D12A 돌연변이체를 보여주는 작제물의 개략도를 보여준다. 도 10b는 PmaCas9 D12A 돌연변이체의 N-말단 또는 C-말단에 융합된 ABE 또는 CBE를 포함하는 염기 편집기를 사용하여 달성된 A에서 G로 또는 C에서 T로 전환 결과를 보여주는 그래프를 보여준다.

[0610] **표 14. PmaCas9와 함께 사용하기 위한 가이드 RNA 서열 및 PAM 서열**

종류(Pma)	서열	PAM
가이드 2	GAAACAATGATAACAAGACC (서열번호 165)	TGG
가이드 3	GTGGCCCCTGTGCCCAGCCC (서열번호 166)	TGG
가이드 4	GTCCCAAATATGTAGCTGTT (서열번호 167)	TGG
가이드 6	GCTCCCATCACATCAACCGG (서열번호 168)	TGG
가이드 7	GATGTCACCTCCAATGACTA (서열번호 169)	GGG
가이드 9	GTTGAAGATGAAGCCCAGAG (서열번호 170)	CGG
가이드 10	GCCAACACCAACCAGAACTT (서열번호 171)	GGG
가이드 11	TGCTGCACACAGCAGGCCTT (서열번호 172)	TGG

[0611]

[0612] 표 15는 염기 편집 기능을 위한 예시적인 Cas9 아데노신 또는 아데닌 및 시토신 또는 시티딘 염기 편집기에 대한 서열을 개시한다.

[0613] 표 15. 예시적인 Cas9 아데노신 또는 아데닌 및 시토신 또는 시티딘 염기 편집기의 서열

서열번호. (설명)	DNA 절단 분석의 요소
아데닌 데아미나제, 닉카제 *Sco*Cas9 의 N-말단에 융합된 TadA8.13m-닉카제 (ABE-nScoCas9, D10A 돌연변이체)의 아미노산 서열	
MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP TAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESAT PESSGPKKKRKVGGKPYSIGLAIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALL FDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHH PIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTD VQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIAL SLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTV DDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEF YKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKEN QDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDL YLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLL NYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIEDIIQTLTLFE DREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANR NFMQLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNP ANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKD MYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSK LLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRR IRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEY PKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKK TEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVAYSILVFADVEKGKSKKLR KVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKELQK GNELVIPQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIE MLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYTSVTECLNATLIHQSITGL YETRIDLSKLGEDGKRPAATKKAGQAKKKKGS*YPYDVPDYAYPYDVPDYAYPYDVPDYA* (서열번호 20)	
시티딘 데아미나제, 닉카제 nScoCas9 의 N-말단에 융합된 ppAPOBEC1 (CBE-	

[0614]

nScoCas9, D10A 돌연변이체)의 아미노산 서열

M**PAAKRVKLD***TSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSG*
KNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARL
FWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELH
CIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSS
*GSETPGTSESATPESSGGSSGGS***PKKKRKV***GGKPYSIGLAIGTNSVGWAVVTDDYKVPAKKMKV*
LGNTDKQSIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFF
QRLDDSFLVDEDKRGEHHPIFGNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMI
KFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLI
AHYPTEKKNTLFGNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLF
ASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDK
NKKGYASYIESGVKQDEFYKYLKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQE
MHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDIL
DKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQE
IFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDD
KANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNK
ENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGI
LQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVEN
QQLQNDRLFLYYIQNGKDMYTGETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKS
DDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHV
AQILDERFNTEFDGNKRRIRNVKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGN
ALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKH
IKILRKVLSYPQVNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVA
YSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLF
ELENKRRRLLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYI
SVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENIDRKRYT
SVTECLNATLIHQSITGLYETRIDLSKLGEDG**KRPAATKKAGQAKKKK**GSSGGSGGSGGS***TNLS***
DIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA
LVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*YPYDVPDYAYPYDVPDYAYP*
YDVPDYA (서열번호 21)

아데닌 데아미나제, 닉카제 *Sir*Cas9 의 N-말단에 융합된 TadA8.13m (ABE-nSirCas9,

[0615]

D14A 돌연변이체)의 아미노산 서열.

M**PAAKRVKLD***GSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP*
TAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESAT
PESSG**PKKKRKV**GAKNKDIRYSIGL**A**IGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAA
TRRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGNDYKDNY
NLFIEKDFNDKTYYDKYPTIYHLRKELCENKEKADPRLIYLALHHIVKYRGNFLKEGQSFAKVY
EDIEEKLDNTLKKFMSLNDLDNLFVDNDINSMITVLSKIYQRSKKADDLLKIMNPTKEERAAYK
EFTKALVGLKFNVSKMILAQEVKKDDKDIELDFSNVDYDSTVDGLQAELGEYIEFIEMLHSINS
WVELQDILGNNSTISAAMVERYEEHKNDLRVLKKVIREELPDKYNEVFREDNPKLHNYLGYIKY
PKNTPVEEFYEYIKRLLAKVDTGEAREILERIDLEKFMLKQNSRTNGSIPYQMQKDEMIQIIDN
QSVYYPQLKENREKLISILEFRIPYYFGPLNTHSEFAWIKKFEDKQKERILPWNYDQIVDIDAT
AEGFIERMQNTGTYFPDKPVMAKNSLTVSKFEVLNELNKIRINGKLIPVETKKELLSDLFMKNK
TITDKKLKDWLVTHQYYDTNEELKIEGYQKDLQFSTSLAPWIDFTKIFGEINASNYQLIEKIIY
DISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGIKSKNSKETILSILENS
NMNLMEIINDESLGFKQIIEESNKKDIEGPFRYDEVKKLAGSPAIKRGIWQALLVVQEITKFMK
HEPSHIYIEFAREEQEKVRTESRIAKLQKIYKDLNLQTKEDQLVYESLKKEDAKKKIDTDALYL
YYLQMGKSMYSGKPLDIDKLSTYHIDHILPRSLIKDDSLDNRVLVLPKENEWKLDSETVPFEIR
NKMMGFWQKLHENGLMSNKKFFSLIRTDFNEKDKKRFINRQLVETRQIIKNVAVIINDHYTNTN
VVTVRAELSHQFRERYKIYKNRDLNDLHHAHDAYIACILGQFIHQNFGNMDVNMIYGQYKKNYK
KDVQEHNNYGFILNSMNHIHFNDDNSVIWDPSYIGKIKSCFCYKDVYVTKKLEQNDAKLFDLTI
LPSDKNSENGVTKAKIPVNKYRKDVNKYGGFSGDAPIMLAIEADKGKKHVRQVIAFPLRLKNYN
DEERIKFIEKEKNLKNVKILTEVKKNQLILINHQYFFITGTNELVNATQLKLSAKNTKNLFNLV
DANKHNKLESIDDANFNEVIQELICKLQEPIYSRYNSIGKEFEDSYEKINAVTKQDKLYIIEYL
IAIMSAKATQGYIKPELAREIGTNGKNKGRIKSFTIDLNKTTFISTSVTGLFSKKYKLG**KRPAA**

TKKAGQAKKKK*GSYPYDVPDYAYPYDVPDYAYPYDVPDYA* (서열번호 6)

아데닌 데아미나제, 닉카제 *Vap*Cas9 의 N-말단에 융합된 TadA8.13m (ABE-nVapCas9, D38A 돌연변이체)의 아미노산 서열

M**PAAKRVKLD***GSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP*
TAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESAT
PESSG**PKKKRKV**GSIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGL**A**IGTSSVGWAVTNKA

[0616]

YELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELFADAMAQVDPTFFMR
LRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHI
LKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAI
EHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKLQI
TGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILK
SLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYIL
SQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRI
PYYVGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVL
PKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTKNRIEQFLKDNGYISET
HKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCL
DDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFME
RIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNK
NEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGR
CAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSFDNLVLCKRTANAQKSDAYPISEEIQKTQKP
FWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFV
KAENVTDFRHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAK
MFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTGALADATIYKATVAGKAKD
GAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLI
NYVASIIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVL
EKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAE
LSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINE
SITGLYSNEVTIVG**KRPAATKKAGQAKKK***GSYPYDVPDYAYPYDVPDYAYPYDVPDYA*

(서열번호 10)

아데닌 데아미나제, 닉카제 *Vap*Cas9 의 C-말단에 융합된 TadA8.13m (nVapCas9-ABE8, D38A 돌연변이체)의 아미노산 서열

M**PKKKRKV**SIINFQRRGLMETQASNQLISSHLKGYPIKDYFVGL**A**IGTSSVGWAVTNKAYELLK
FRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEELFADAMAQVDPTFFMRLRESK
YHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRG
NFLYEGATFDSNASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVD
TYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLIDLFGSVEELEEDLKKLQITGDTY
DDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKS
DRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIEL

[0617]

[0618]

QILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEVADGFSVAEKLIKMLEFRIPYYVG
PLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDVLPKSSL
LYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTKNRIEQFLKDNGYISETHKHEI
TGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAI
KKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPETIIILMRNFSYNLMELLGDKFSFMERIQEI
NAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKK
KDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTG
EIIELSLLNTDNYDIDHIYPRSLTKDDSFDNLVLCKRTANAQKSDAYPISEEIQKTQKPFWTFL
KQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENV
TDFRHDNNFIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYD
VNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTGALADATIYKATVAGKAKDGAYIG
MKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINYVAS
IIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNN
MRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELSEKG
YGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVINESITGL
YSNEVTIV**KRPAATKKAGQAKKKK**SGSETPGTSESATPESSGSEVEFSHEYWMRHALTLAKRAR
DEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPC
VMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLCRFFRM
PRRVFNAQKKAQSSTD**PAAKRVKLD**GSYPYDVPDYAYPYDVPDYAYPYDVPDYA (서열번호 11)

시티딘 데아미나제, 닉카제 *Vap*Cas9 의 N-말단에 융합된 ppAPOBEC1 (CBE-nVapCas9, D38A 돌연변이체)의 아미노산 서열

M**PAAKRVKLD**TSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSG
KNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARL
FWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELH
CIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSS
GSETPGTSESATPESSGGSSGGS**PKKKRKV**GSIINFQRRGLMETQASNQLISSHLKGYPIKDYF
VGL**A**IGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLL
EELFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSEL
MKSGTDDIRKLFLAVHHILKYRGNFLYEGATFDSNASTLDDVIKQALENITFNCFDCNSAISSI
GQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLGLNASLI
DLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTIS
ESKVKAFNKHKDDLAILKSLLKSDRSIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKY

TKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCGPKFPFLNEV
ADGFSVAEKLIKMLEFRIPYYVGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAA
FIKNLTNKCTYLLGEDVLPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKM
TKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILGDGFDRSMAEEIITDITIFG
ESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPETIIILMR
NFSYNLMELLGDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAH
IKKALPARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIKKDDVLLNGLNNEIFGELKSSLAKYD
DAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSFDNLVLCKRTANA
QKSDAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVK
AATTLLRRLYPGVDVVFVKAENVTDFRHDNNFIKVRSLNHHHHHAKDAYLNIVVGNVYHERFTRN
FRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRVTKRLLEQTG
ALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEI
VPLPIYLTNRNTTDQDLINYVASIIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCID
NAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRIVEEKNIEFFDYLV
SKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASR
STVGSKISNQDEFKVINESITGLYSNEVTIVG**KRPAATKKAGQAKKKK**GSSGGSGGSGGS***TNLS***
DIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWA
LVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESD
ILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY

(서열번호 12)

아데닌 데아미나제, 닉카제 EpeCas9 의 N-말단에 융합된 TadA8.13m (ABE-nEpeCas9, D12A 돌연변이체)의 아미노산 서열

M**PAAKRVKLD***GSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP*
TAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM
DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESAT
PESSG**PKKKRKV**GTKVKDYYIGL**A**IGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEER
RGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKDQKLKSKYTLFNDKD
FKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLN
ELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLV
DLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEADKS
KDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRV
KANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIP

[0619]

[0620]

YQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKD
SEITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPI
TEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFKAIVGDK
VDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLT
GLEGANKITGERGSIIHFMREYNLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPS
VKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFISEIG
EERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIY
DDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDE
LVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDA
YINIIVGNVYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIK
RIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKSSIEKYGGYTNINKACFAV
LDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKID
GSYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDII
VNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLKSIDLKS
DYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLG**KRPAATKKAGQAKKKK**GS*YPYDVPD*
YAYPYDVPDYAYPYDVPDYA (서열번호 16)

아데닌 데아미나제, 닉카제 EpeCas9 의 C -말단에 융합된 TadA8.13m (nEpeCas9-ABE8, D12A 돌연변이체)의 아미노산 서열

M**PKKKRKV**TKVKDYYIGL**A**IGTSSVGWAVTDEAYNVLKFNSKKMWGVRLFDDAKTAEERRGQRG
ARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYMEDKDQKLKSKYTLFNDKDFKDKN
FHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHFIFEGQKFDTKSSFENSLNELKVH
LNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDTKFLKAVSAIMIGSSQKLVDLFEN
PEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILKEIYDSSILENLLKEADKSKDGNK
YISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKVNDNYVSYTKSSISNNKRVKANKF
TDQEAFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDMEFKNFMPKIKSSDNGVIPYQLKL
MELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYYVGPLNPNSKYAWIKKQKDSEITP
WNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYNEYMVLNELNNLKLNDLPITEEMK
KKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGDFKQGLNSYNDFKAIVGDKVDSDD
YRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKKMAGLNYKDWGRLSKKLLTGLEGA
NKITGERGSIIHFMREYNLNLMELMSASFTFTEEIQKLNPVDDRKLSYEMVDELYLSPSVKRML
WQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKNQLLKFYKDGKKAFISEIGEERYS
YLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISELSSKNIYDQDHIYPKSKIYDDSIE

[0621]

NRVLVKKDLNSKKGNSYPIPDEILNKNCYAYWKILYDKGLIGQKKYTRLTRRTGFTDDELVQFI
SRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQEFDIVKCRAVNDLHHMHDAYINII
VGNVYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGYTAWIADDEKGTVKNASIKRIRKE
LEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKSSIEKYGGYTNINKACFAVLDIKS
KNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIEDYRVVYPVVKMRTLLKIDGSYYF
ITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTIERIQDLTEELVYNTFDIIVNKFK
TSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVKAIRASGVRQDLKSIDLKSDYGRL
SSKTNNIGNYQEFKIINQSITGLFENEVDLLKL**KRPAATKKAGQAKKKK**SGSETPGTSESATPE
SSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMA
LRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGM
NHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD**PAAKRVKLD**GSYPYDVPDYAYPY
DVPDYAYPYDVPDYA (서열번호 17)

아데닌 데아미나제, 닉카제 EpeCas9 의 C-말단에 융합된 TadA8.13m (nEpeCas9-ABE8, D12A 돌연변이체)의 아미노산 서열

M**PAAKRVKLD**TSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSG
KNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARL
FWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELH
CIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSS
GSETPGTSESATPESSGGSSGGS**PKKKRKV**GTKVKDYYIGL**A**IGTSSVGWAVTDEAYNVLKFNS
KKMWGVRLFDDAKTAEERRGQRGARRRLDRKKERLSLLQDFFAEEVAKVDPNFFLRLDNSDLYM
EDKDQKLKSKYTLFNDKDFKDKNFHKKYPTIHHLLMDLIEDDSKKDIRLVYLACHYLLKNRGHF
IFEGQKFDTKSSFENSLNELKVHLNDEYGLDLEFDNENLINILTDPKLNKTAKKKELKSVIGDT
KFLKAVSAIMIGSSQKLVDLFENPEDFDDSAIKSVDFSTTSFDDKYSDYELALGDKIALVNILK
EIYDSSILENLLKEADKSKDGNKYISNAFVKKYNKHGQDLKEFKRLVRQYHKSAYFDIFRSEKV
NDNYVSYTKSSISNNKRVKANKFTDQEAFYKFAKKHLETIKYKINKVNGSKADLELIDGMLRDM
EFKNFMPKIKSSDNGVIPYQLKLMELNKILENQSKHHEFLNVSDEYGSVCDKIASIMEFRIPYY
VGPLNPNSKYAWIKKQKDSEITPWNFKDVVDLDSSREEFIDSLIGRCTYLKDEKVLPKASLLYN
EYMVLNELNNLKLNDLPITEEMKKKIFDQLFKTRKKVTLKAVANLLKKEFNINGEILLSGTDGD
FKQGLNSYNDFKAIVGDKVDSDDYRDKIEEIIKLIVLYGDDKSYLQKKIKAGYGKYFTDSEIKK
MAGLNYKDWGRLSKKLLTGLEGANKITGERGSIIHFMREYNLNLMELMSASFTFTEEIQKLNPV
DDRKLSYEMVDELYLSPSVKRMLWQSLRIVDEIKNIMGTDSKKIFIEMARGKEEVKARKESRKN
QLLKFYKDGKKAFISEIGEERYSYLLSEIEGEEENKFRWDNLYLYYTQLGRCMYSLEPIDISEL

[0622]

SSKNIYDQDHIYPKSKIYDDSIENRVLVKKDLNSKKGNSYPIPDEILNKNCYAYWKILYDKGLI
GQKKYTRLTRRTGFTDDELVQFISRQIVETRQATKETANLLKTICKNSEIVYSKAENASRFRQE
FDIVKCRAVNDLHHMHDAYINIIVGNVYNTKFTKDPMNFVKKQEKARSYNLENMFKYDVKRGGY
TAWIADDEKGTVKNASIKRIRKELEGTNYRFTRMNYIESGALFNATLQRKNKGSRPLKDKGPKS
SIEKYGGYTNINKACFAVLDIKSKNKIERKLMPVEREIYAKQKNDKKLSDEIFSKYLKDRFGIE
DYRVVYPVVKMRTLLKIDGSYYFITGGSDKTLELRSALQLILPKKNEWAIKQIDKSSENDYLTI
ERIQDLTEELVYNTFDIIVNKFKTSVFKKSFLNLFQDDKIENIDFKFKSMDFKEKCKTLLMLVK
AIRASGVRQDLKSIDLKSDYGRLSSKTNNIGNYQEFKIINQSITGLFENEVDLLKLGKRPAATK
KAGQAKKKKGSSGGSGGSGGS***TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVH***
TAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQ
LVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGEN
KIKMLYPYDVPDYAYPYDVPDYAY (서열번호 18)

아데닌 데아미나제, 닉카제 *Lfe*Cas9 의 N-말단에 융합된 TadA8.13m-닉카제 (ABE-nLfeCas9, D9A 돌연변이체)의 아미노산 서열

M**PAAKRVKLD**GSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPT
AHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDV
LHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESATPES
SG**PKKKRKV**GKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTR
RRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLLKKNERGY
PTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVDKF
KVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVA
DKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQNDIL
TEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYNKYIGQ
APKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAK
YYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEIT
PWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIK
QDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKSIVGDKVDDN
RYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQ
RIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQV
VRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAP
DLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDR
VPAKLYAAKMKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANIL

[0623]

GSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFV
YGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLIS
NETYEETGKYFDASHESSSLYLKNQNKKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKG
GFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFILS
SYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRA
KLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSN
TAFIIFKSPTGLFNRKLYLKNLG**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDYAYPYDVP*
DYA (서열번호 88)

아데닌 데아미나제, 닉카제 *Lfe*Cas9 의 C-말단에 융합된 TadA8.13m-닉카제 (nLfeCas9-ABE, D9A 돌연변이체)의 아미노산 서열

M**PKKKRKV**GKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTR
RRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDPAKNQAFIGKLLFPDLLKKNERG
YPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLNNASVD
KFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEV
KVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQ
NDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLAEVKEYLATQPASARKEFDQVYN
KYIGQAPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRI
IEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKR
KEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLPANSLLYQKYNVLNELNNVRVNG
RRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLKS
IVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKL
LTGIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYT
SPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQLQKLFEDQAHELVK
DTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRV
LVSRAENNKKSDRVPAKLYAAKMKPYWNQLLKQGLITQRKFENLTMDVDQTIKYRSLGFVKRQL
VETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDLPKVREVNDYHHAVDAYLTTFAG
QYLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEV
ADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQNKKSKLVVPLKNKLQPEYYGAYT
GITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYRNKIAEIISSDPRAPKSFEVVVP
RVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDNFKALKKLNVEMQTKKLIEIYDN
ILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEKVIIKILRGLHFGSDLQNLKEIG

FGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL**KRPAATKKAGQAKKKK**SGSETP
GTSESATPESSGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHD
PTAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSL
MDVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD**PAAKRVKLD**GSYPY
DVPDYAYPYDVPDYAYPYDVPDYA (서열번호 89)

시티딘 데아미나제, 닉카제 LfeCas9 의 N-말단에 융합된 ppAPOBEC1 (CBE-nLfeCas9, D9A 돌연변이체)의 아미노산 서열

M**PAAKRVKLD**TSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSG
KNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARL
FWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELH
CIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSS
GSETPGTSESATPESSGGSSGGS**PKKKRKV**GKEYHIGLAIGTSSIGWAVTDSQFKLMRIKGKTA
IGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDEIFAPHLQEVDENFLRRLKQSNIHPEDP
AKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLRE
VYLAVHHIVKYRGHFLNNASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQ
LLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLVLGYKADFATVAMANGNEWKI
DLSSETSEDEIEKFREELSDAQNDILTEITSLFSQIMLNEIVPNGMSISESMMDRYWTHERQLA
EVKEYLATQPASARKEFDQVYNKYIGQAPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLP
KQRTSANGVIPHQMHQQELDRIIEKQAKYYPWLATENPATGERDRHQAKYELDQLVSFRIPYYV
GPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLNEDVLP
ANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPS
VEGLSDPKKFNSNLATYLDLKSIVGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLT
DEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFMQIVNQPVFKEQIDQLNQK
AITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGE
ITRSRRTQLQKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEIST
KYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRVPAKLYAAKMKPYWNQLLKQGLITQRKFE
NLTMDVDQTIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFDL
PKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEG
DKSQGKVVDQQTGELITTRDEVADYFDWVINLKVMLISNETYEETGKYFDASHESSSLYLKNQN
KKSKLVVPLKNKLQPEYYGAYTGITQGYMVILKLLDKKGGFGVYRIPRYAADILNKCHDEVAYR
NKIAEIISSDPRAPKSFEVVVPRVLKGTFLVDGEEKFILSSYRYKVNATQLILPVSDIKLIQDN
FKALKKLNVEMQTKKLIEIYDNILRQVDKYYKLYDINKFRAKLHDGRSKFVELDDFGQDASKEK

[0624]

VIIKILRGLHFGSDLQNLKEIGFGTTPLGQFQVSEAGIRLSNTAFIIFKSPTGLFNRKLYLKNL

GKRPAATKKAGQAKKKKGSSGGSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNK

PESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDI

IEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALV

IQDSNGENKIKMLYPYDVPDYAYPYDVPDYAY (서열번호 90)

아데닌 데아미나제, 닉카제 *Pma*Cas9 의 N-말단에 융합된 TadA8.13m-닉카제 (ABE-PmaCas9, D12A 돌연변이체)의 아미노산 서열

MPAAKRVKLDGSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDP

TAHAEIMALRQGGLVMQNYRLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLM

DVLHHPGMNHRVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTDGSSGSETPGTSESAT

PESSGPKKKRKVGEKKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGS

LLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFS

KNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEH

IAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQS

PMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYNAV

ELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKG

YAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAI

LDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDR

SAEKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFET

LFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMMDDPDNEDLM

EEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTIL

DYLVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGI

LLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPG

NADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGN

KKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITK

NVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAI

AINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERD

IKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSKYGGYIEEKMAYYA

VISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYI

TGATKTTKGKVELQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHY

TRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYN

[0625]

SMTGLFDGQVIYQSLTGLYETRMRFEDG**KRPAATKKAGQAKKKK**GS*YPYDVPDYAYPYDVPDYA*

YPYDVPDYA (서열번호 91)

아데닌 데아미나제, 닉카제 *Pma*Cas9 의 C-말단에 융합된 TadA8.13m-닉카제 (nPmaCas9-ABE, D12A 돌연변이체)의 아미노산 서열

M**PKKKRKV**EKKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLLFES
GQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFSKNPIF
GTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIAINE
NLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPMMAF
ITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYNAVELAGI
LSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYAGYI
DGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILDRQE
NYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSAEKF
IEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLFKKR
SKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMMDDPDNEDLMEEIVK
ILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDYLVE
DDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILLGLK
IVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNADLS
QRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKKDDV
PSLEVVNRQLAYWRSLKDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNVAKL
LDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAINLL
ALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIKTIK
KVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSKYGGYIEEKMAYYAVISYE
DKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITGATK
TTKGKVELQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTRRFL
DAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSMTGL
FDGQVIYQSLTGLYETRMRFED**KRPAATKKAGQAKKKK**SGSETPGTSESATPESSG*SEVEFSHE*
YWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNY
RLYDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNHRVEITEGIL
*ADECAALLCRFFRMPRRVFNAQKKAQSSTD***PAAKRVKLD***GS**YPYDVPDYAYPYDVPDYAYPYDV*

PDYA (서열번호 92)

[0626]

시티딘 데아미나제, 닉카제 PmaCas9 의 N-말단에 융합된 ppAPOBEC1 (CBE-nPmaCas9, D12A 돌연변이체)의 아미노산 서열

M**PAAKRVKLD***TSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSSG*
KNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVARL
FWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELH
CIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRLKSGGSSGGSS
GSETPGTSESATPESSGGSSGGS**PKKKRKV**GEKKTNYTIGLAIGTDSVGWAVVKDDLELVKKRM
KVLGNTETNYIKKNLWGSLLFESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDEN
FFFRLNESFLVPEDKAFSKNPIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAH
MIKYRGHFLIEGKLDTEHIAINENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVK
EKRITSFLKDIGRANKQSPMMAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTI
GSDFADLFDHAQRVYNAVELAGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQ
YDMTFVAPKDAQKKDLKGYAGYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRS
FYNGVIPNQVHLAEMEAILDRQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERK
SDEKIKPWNFSDVVDLDRSAEKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDE
RQKRMNLSSVEKKEIFETLFKKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELA
KVPGMKSMMDDPDNEDLMEEIVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRL
SKKLLVGIRDKETQKTILDYLVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSAD
LYAQVQEIAGSPAIKKGILLGLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKI
KEGLAALGSSLLKENLPGNADLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTV
DNSLDNLVLTNSSQNRGNKKDDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDNLTKAMRGGLTDK
DRERFIQRQLVETRQITKNVAKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKV
RELNDYHHAHDAYLNAAIAINLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFG
EKQLFTPDHEDCWNKERDIKTIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKD
LDPSKYGGYIEEKMAYYAVISYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQI
ILKNYSLIAYPDGRRRYITGATKTTKGKVELQKANQIAMEQDLVNFIYHLKNYDEISHPESYAF
VQSHTDYFDRLFDSIEHYTRRFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAG
APADFIFMGEAISRRRYNSMTGLFDGQVIYQSLTGLYETRMRFEDG**KRPAATKKAGQAKKKK***GS*
*SGGSGGSGGS**TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENV***
MLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLP
EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*YPYDVP*
DYAYPYDVPDYAY (서열번호 93)

[0627]

시티딘 데아미나제, 닉카제 PmaCas9 의 C-말단에 융합된 ppAPOBEC1 (nPmaCas9-CBE, D12A 돌연변이체)의 아미노산 서열

MPAAKRVKLD*TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENV*
MLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLP
*EEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML*SGGSGG
SGGSPKKKRKVEKKTNYTIGLAIGTDSVGWAVVKDDLELVKKRMKVLGNTETNYIKKNLWGSLL
FESGQTAKDRRLKRVARRRYERRRNRLTELQKIFAPAIDEVDENFFFRLNESFLVPEDKAFSKN
PIFGTLGEDKTYYKTYPTIYHLRQHLADSEEKADVRLIYLALAHMIKYRGHFLIEGKLDTEHIA
INENLEQFFESYNALFSEEPIELRKEELIAIENILREKNSRTVKEKRITSFLKDIGRANKQSPM
MAFITLIVGKKAKFKAAFNLEEEISLNLTDDSYDENLEILLNTIGSDFADLFDHAQRVYNAVEL
AGILSGDVKNTHAKLSAQMVAMYERHKEQLKEYKSFIKANLPDQYDMTFVAPKDAQKKDLKGYA
GYIDGNMSQDSFYKFVKDQLKEVPGSEKFLDSIEKEDFLRKQRSFYNGVIPNQVHLAEMEAILD
RQENYYPWLKENREKIISLLTFRIPYYVGPLADGQSEFAWLERKSDEKIKPWNFSDVVDLDRSA
EKFIEQLIGRDTYLPDEYVLPKKSLIYQKYMVFNELTKIAYLDERQKRMNLSSVEKKEIFETLF
KKRSKVTEKQLVKFFENYLQIDNPTIFGIEDAFNADYSTYVELAKVPGMKSMMDDPDNEDLMEE
IVKILTVFEDRKMRRKQLEKYKERLSPEQIKELAKKHYTGWGRLSKKLLVGIRDKETQKTILDY
LVEDDNHSGGRQHLNRNLMQLINDDRLSFKKTIAELQMIDPSADLYAQVQEIAGSPAIKKGILL
GLKIVDEIIRVMGEKPENIVIEMARENQTTARGKALSKRREAKIKEGLAALGSSLLKENLPGNA
DLSQRKIYLYYTQNGKDIYLDEPLDFDRLSQYDEDHIIPQSFTVDNSLDNLVLTNSSQNRGNKK
DDVPSLEVVNRQLAYWRSLKDAGLMTQRKFDNLTKAMRGGLTDKDRERFIQRQLVETRQITKNV
AKLLDMRLNDKKDEAGNKIRETNIVLLKSAMASEFRKMFRLYKVRELNDYHHAHDAYLNAAIAI
NLLALYPYMADDFVYGEFRYKKKPQAEKATYEKLRQWNLIKRFGEKQLFTPDHEDCWNKERDIK
TIKKVMGYRQVNVVKKAEERTGMLFKETINGKTNKGSRIPIKKDLDPSKYGGYIEEKMAYYAVI
SYEDKKKKPGKTIVGISIMDKKEFEYDSISYLGKLGFSNPVVQIILKNYSLIAYPDGRRRYITG
ATKTTKGKVELQKANQIAMEQDLVNFIYHLKNYDEISHPESYAFVQSHTDYFDRLFDSIEHYTR
RFLDAETNINRLRRIYEEEKKKDPVDIEALVASFIELLKLTSAGAPADFIFMGEAISRRRYNSM
TGLFDGQVIYQSLTGLYETRMRFEDKRPAATKKAGQAKKKKGSSGGSSGGSSGSETPGTSESAT
*PESSGGSSGGS*TSEKGPSTGDPTLRRRIESWEFDVFYDPRELRKETCLLYEIKWGMSRKIWRSS
GKNTTNHVEVNFIKKFTSERRFHSSISCSITWFLSWSPCWECSQAIREFLSQHPGVTLVIYVAR
LFWHMDQRNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALEL
HCIILSLPPCLKISRRWQNHLAFFRLHLQNCHYQTIPPHILLATGLIHPSVTWRYPYDVPDYAY

PYDVPDYAYPYDVPDYA (서열번호 94)

[0628]

[0629] 링커(밑줄, 이탤릭체나 굵은 글씨 없음)

[0630] TadA8(ABE) 또는 ppABOBEC1(CBE)(이탤릭체 및 밑줄)

[0631] 닉카제 돌연변이: ScoCas9에서의 D10A 돌연변이, SirCas9에서의 D14A 돌연변이, VapCas9에서의 D38A, EpeCas9에서의 D12A, LfeCas9에서의 D9A, PmaCas9에서의 D12A(굵은 글씨 및 이탤릭체) 2xUGI(굵은 글씨, 이탤릭체 및 밑줄)

[0632] 3xHA 태그(이탤릭체)는 다른 태그로 대체 가능하다.

[0633] **실시예 6. 조작된 *스트렙토코커스 콘스텔라투스*(ScoCas9) NGC PAM 변이체**

[0634] 이 실시예는 NGC PAM 변이체를 인식하는 ScoCas9 변이체의 조작을 보여준다.

[0635] 즉, NGC PAM 인식에 대한 아미노산 치환을 함유하는 ScoCas9-NGC-v1 및 NGC PAM 인식에 대한 아미노산 치환과 SpyCas9 활성을 증강시키는 추가 아미노산 치환을 함유하는 ScoCas9-NGC-v2라는 2가지 변이체를 조작했다. *S. 피오게네스* SpyCas9와 *S. 콘스텔라투스* ScoCas9 간의 구조적 비교를 통해 아미노산 잔기를 식별했다. ScoCas9-NGC-v1의 아미노산 서열(서열번호 95)은 야생형 ScoCas9 서열로부터 다음 돌연변이를 포함했다: D1117M, S118Q, E1201F, A1299R, D1309A, R1312E, T1314R. ScoCas9-NGC-v2의 아미노산 서열(서열번호 96)은 야생형 ScoCas9 서열로부터 다음 돌연변이를 포함했다: S409I, R655L, D1117M, S118Q, E1201F, A1299R, D1309A, R1312E, T1314R.

스트렙토코커스 콘스텔라투스 (ScoCas9) 변이체 (ScoCas9-NGC-v1)의 아미노산 서열

MPKKKRKVMGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDS
GETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIF
GNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQA
LFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLD
LHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDN
STKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKY
LKGILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDK
IEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLP
EEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYL
NKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDRE
MIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFM
QLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANI
IIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYT
GETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVKEMKSFWSKLLS
VKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRN
VKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKY
NSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEE
QTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVFADVEKGKSKKLRKVQ
DMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNE
LVIPQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLF
SKNKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITGLYET
RIDLSKLGEDG***EGADKRTADGSEFESPKKKRKV*** (서열번호 95)

스트렙토코커스 콘스텔라투스 (ScoCas9) 변이체(ScoCas9-NGC-v2)의 아미노산 서열

MPKKKRKVMGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDS
GETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIF
GNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQA
LFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLD

[0636]

LHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDN
STKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKY
LKGILLKINGSGDFLDKIDREDFLRKQRTFDNGIIPHQIHLQEMHAILRRQGEHYPFLKENQDK
IEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLP
EEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYL
NKEFEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDRE
MIRQRLQKYSDIFTKAQLKKLERLHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFM
QLINDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANI
IIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYT
GETLDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLS
VKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRN
VKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKY
NSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEE
QTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVFADVEKGKSKKLRKVQ
DMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNE
LVIPQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLF
SKNKDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITGLYET
RIDLSKLGEDG***EGADKRTADGSEFESPKKKRKV*** (서열번호 96)

[0637]

[0638] NLS(이탤릭체 굵은 글씨)

[0639] 링커(밑줄있는 굵은 글씨)

[0640] 렌티바이러스 매개 삽입에 의해 HEK293T 세포의 게놈 내로 무작위 통합되고 뉴클레아제 및 염기 편집 활성에 대해 테스트된 게놈 유전자좌를 표적으로 하기 위해 ScoCas9-NGC 변이체를 사용했다.

[0641] 즉, HEK293T 세포를 96웰 플레이트에 플레이팅하였다. 플레이팅 후 24시간에 ScoCas9-NGC 변이체 및 가이드 RNA 서열 ATCGACAAGAAAGGGACTGA(서열번호 97)를 함유하는 발현 벡터로 세포를 형질감염시켰다. ScoCas9 변이체는 예

시적인 NGC 3' PAM 서열인 AGC를 인식했다. 형질감염 후 72시간에 세포를 수확하고 전체 DNA를 추출했다.

[0642] HEK293T 세포 내 삽입결실 패턴을 특징규명하기 위해 심층 서열분석을 수행했다. Illumina 어댑터 및 표적 앰플리콘에 대한 고유 바코드를 추가하는 2회 PCR을 사용하여 예시적인 표적을 증폭했다. PCR 산물을 2% 겔 상에서 실행하고, 겔을 추출했다. 샘플을 모으고, 정량화하고, cDNA 라이브러리를 제조하고, MiSeq 상에서 서열분석했다. 삽입결실 빈도는 형질감염 후 4일에 심층 서열분석으로 결정했다.

[0643] 결과는 ScoCas9-NGC 변이체 둘 모두의 뉴클레아제 활성을 보여주었다. ScoCas9-NGC-v1 및 ScoCas9-NGC-v2를 사용하여 약 20 내지 35%의 삽입결실 빈도를 달성했다(도 11a).

[0644] ABE 염기 편집기를 사용하여 ScoCas9-NGC 변이체의 융합이 작제되었다.

아데닌 염기 편집기에 융합된 ScoCas9 변이체의 아미노산 서열 (ABE-nScoCas9-NGC-v1)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
*RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSESATPESS**
GGSSGGSGKPYSIGLAIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGET
AEATRLKRTARRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIFGNI
AAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFK
DFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHP
NFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTK
APLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKG
ILLKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEK
ILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEK
VLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKE
FEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDREMIR
QRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLI
NDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIE
MARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGET
LDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKL
ITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKI
ITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSY
RSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTG
GFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVFADVEKGKSKKLRKVQDMV
GITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVI
PQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKN
KDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITGLYETRID
LSKLGEDG***EGADKRTADGSEFESPKKKRKV*** (서열번호 98)

아데닌 염기 편집기에 융합되는 ScoCas9 변이체의 아미노산 서열 (ABE-nScoCas9-NGC-v2)

[0645]

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALR
QGGLVMQNYRLYDATLYSTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHHPGMNH
*RVEITEGILADECAALLCRFFRMPRRVFNAQKKAQSSTD***SGGSSGGSSGSETPGTSESATPESS**
GGSSGGSGKPYSIGLAIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLLGALLFDSGET
AEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIFGNI
AAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFK
DFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTLFGNLIALSLDLHP
NFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTK
APLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKKGYASYIESGVKQDEFYKYLKG
ILLKINGSGDFLDKIDREDFLRKQRTFDNGIIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEK
ILTFRIPYYVGPLARKGSRFAWAEYKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEK
VLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKE
FEEFRIVNLTGLDKENKAFNSSLGTYHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDREMIR
QRLQKYSDIFTKAQLKKLERLHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLI
NDDALSFKEEIARAQIIDDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIE
MARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGET
LDINNLSQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKL
ITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNVKI
ITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSY
RSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQVNIVKKTEEQTG
GFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFMQPVVAYSILVFADVEKGKSKKLRKVQDMV
GITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELENKRRRLLASAKFLQKGNELVI
PQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKN
KDAEVSSLAKSFISLLTFTAFGAPRAFNFFGENIARKEYRSVTECLNATLIHQSITGLYETRID

LSKLGEDG***EGADKRTADGSEFESPKKKRKV*** (서열번호 99)

[0646]

[0647] 링커(밑줄있는 굵은 글씨)

[0648] TadA8(ABE)(이탤릭체)

[0649] NLS(굵은 이탤릭체)

[0650] HEK293T 세포에서 A에서 G로 전환을 특징규명하기 위해 심층 서열분석을 또한 수행했다(도 11b). 아데닌에서 구아닌으로(A에서 G로) 전환은 형질감염 후 4일에 NGS로 측정하였다. 결과는 ABE-nScoCas9-NGC 변이체 둘 모두에 의한 염기 편집 활동을 보여주었다. 변이체 둘 모두는 약 20 내지 30%의 A에서 G로 전환을 보여주었다. NGG를 인식하는 ScoCas9를 음성 대조군으로 사용했고 염기 편집을 나타내지 않았다. SpyCas9를 양성 대조군으로 사용했고 약 40%의 A에서 G로 전환을 보여주었다.

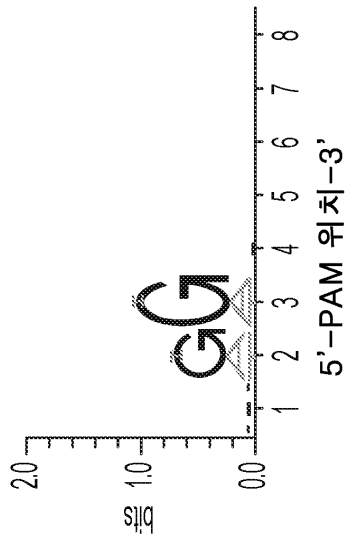
[0651] 전반적으로, 결과는 NGC PAM 서열을 인식하도록 조작된 ScoCas9 변이체가 뉴클레아제 및 염기 편집 활동을 수행할 수 있음을 보여주었다.

[0652] **균등물 및 범위**

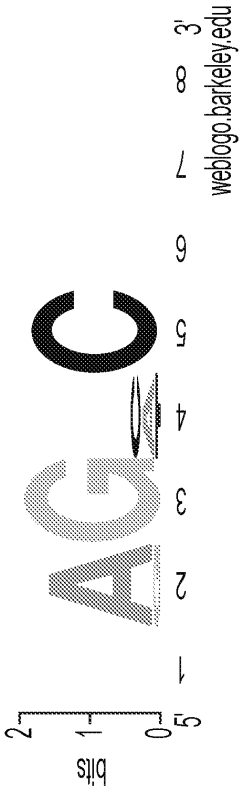
[0653] 당업자는 단지 통상 실험을 사용하여 본원에 기술된 본 발명의 특정 구현예에 대한 많은 균등물을 인식하거나 식별할 수 있을 것이다. 본 발명의 범위는 상기 설명에 국한되는 것이 아니며, 다음의 청구범위에 기재된 바와 같다.

도면

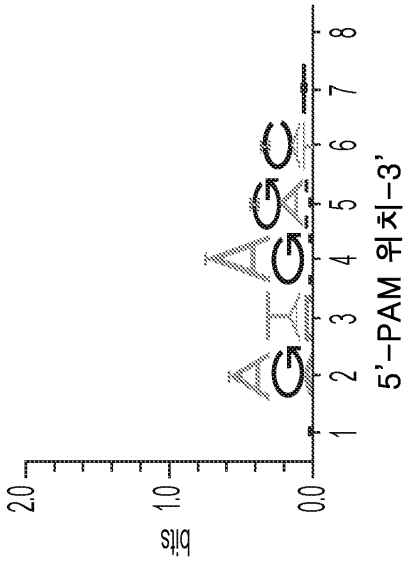
도면1a


bits
2.0
1.0
0.0
1 2 3 4 5 6 7 8
5'-PAM 위치-3'

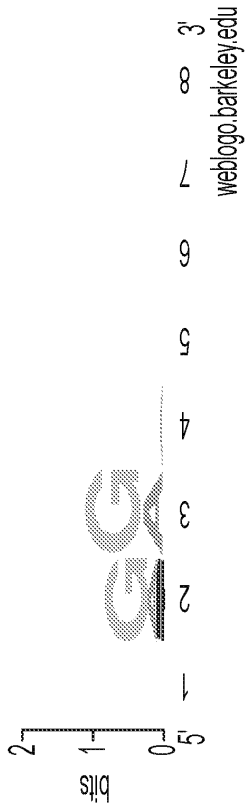
도면1b


bits
2
1
0
5'
1 2 3 4 5 6 7 8 3'
weblogo.berkeley.edu

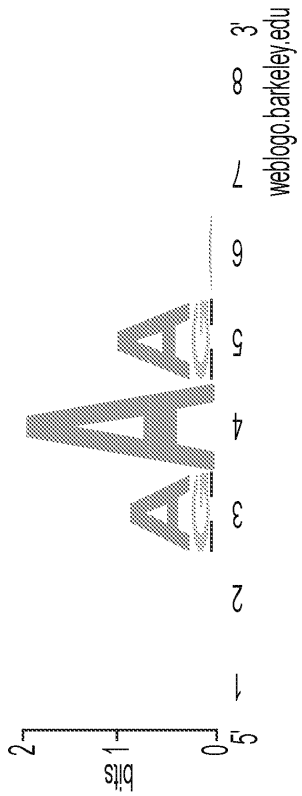
도면1c


bits
2.0
1.0
0.0
1 2 3 4 5 6 7 8
5'-PAM 위치-3'

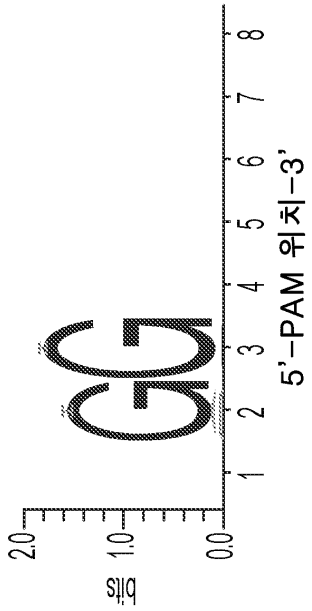
도면1d


bits
2
1
0
5'
1 2 3 4 5 6 7 8 3'
weblogo.berkeley.edu

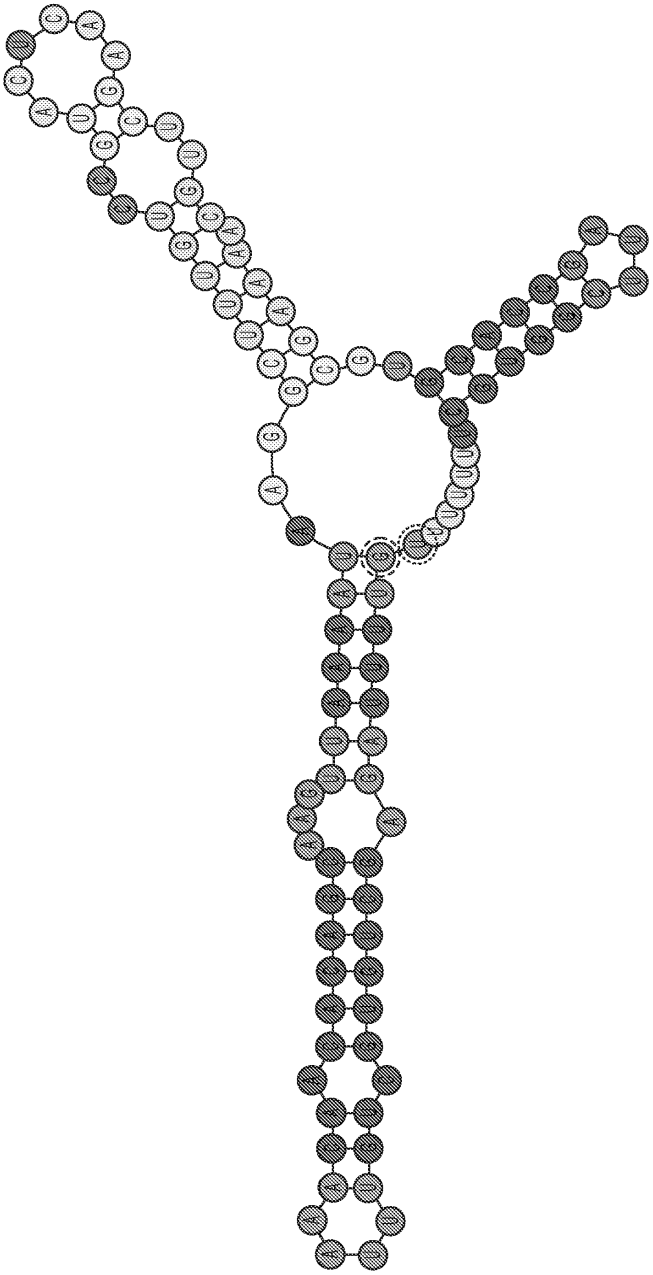
도면1e


bits
2
1
0
5'
1
2
3
4
5
6
7
8
3'
weblogo.barkeley.edu

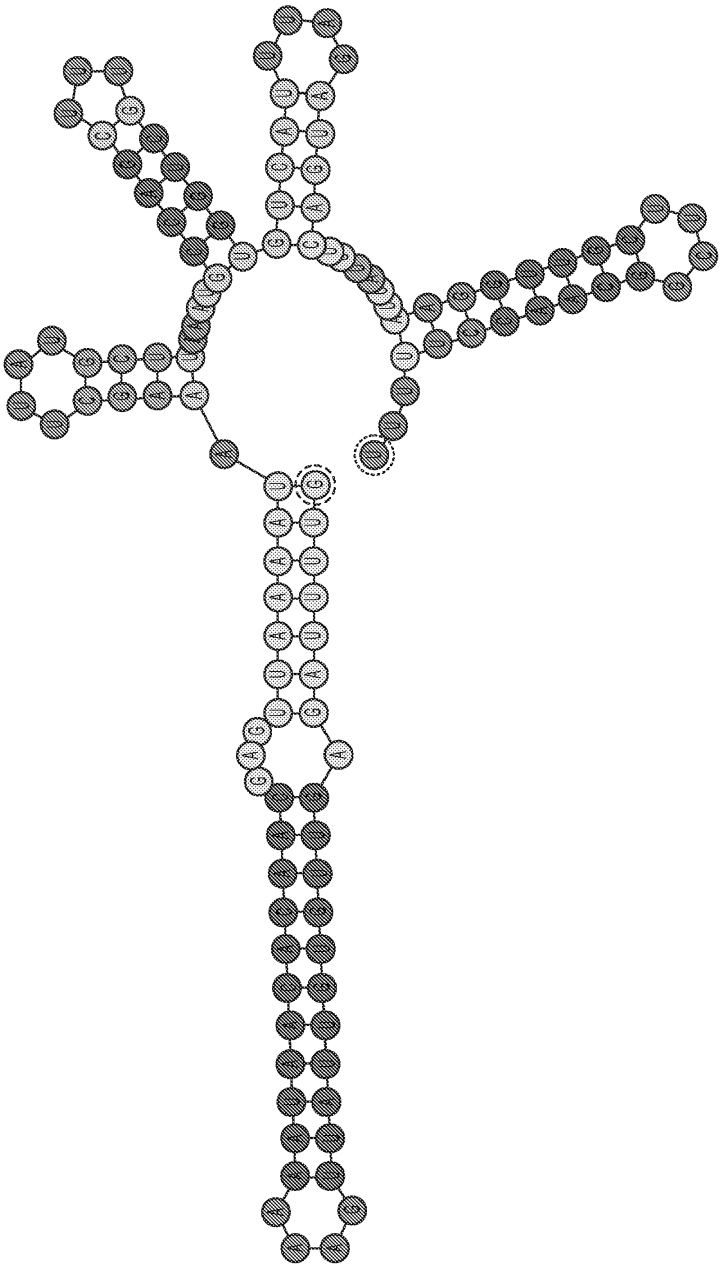
도면1f


bits
2.0
1.0
0.0
1
2
3
4
5
6
7
8
5'-PAM 위치-3'

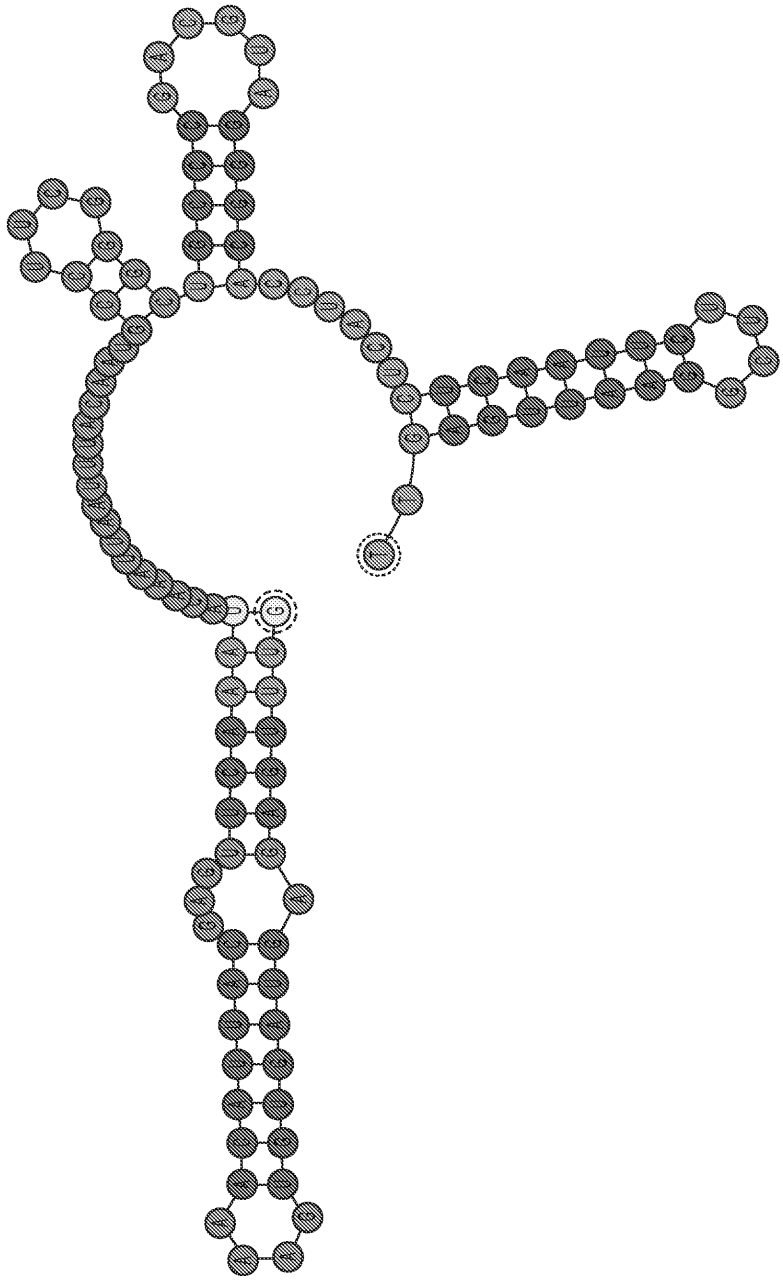
도면2a



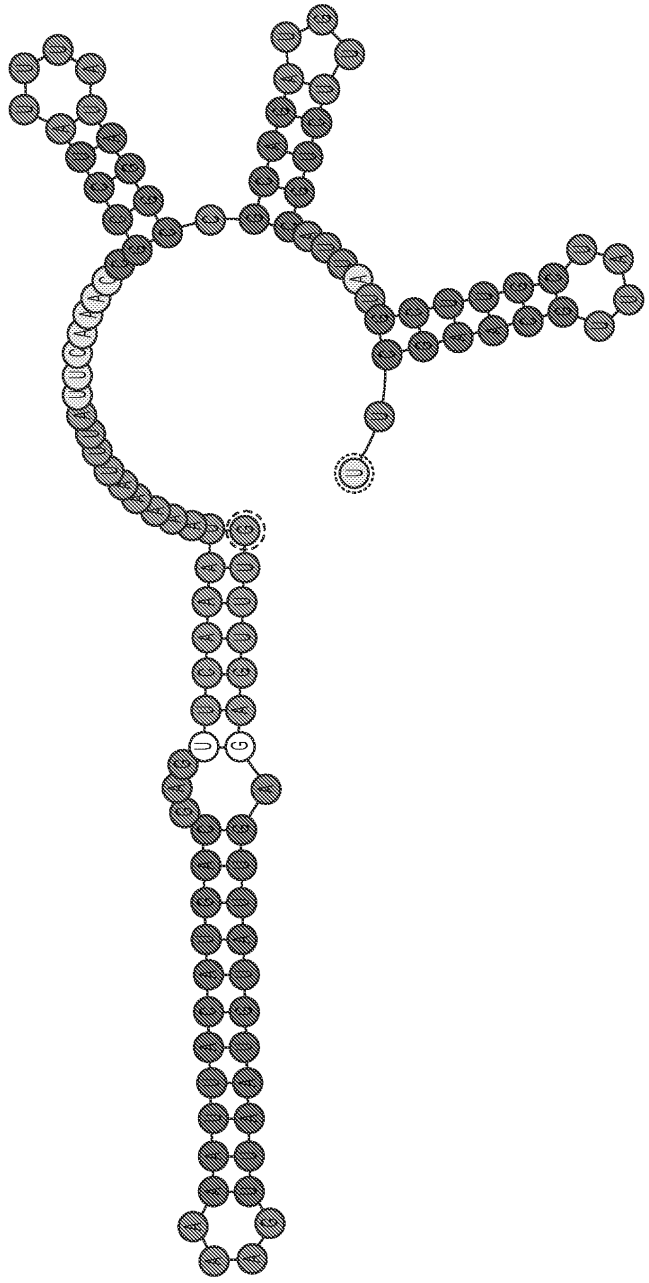
도면2b



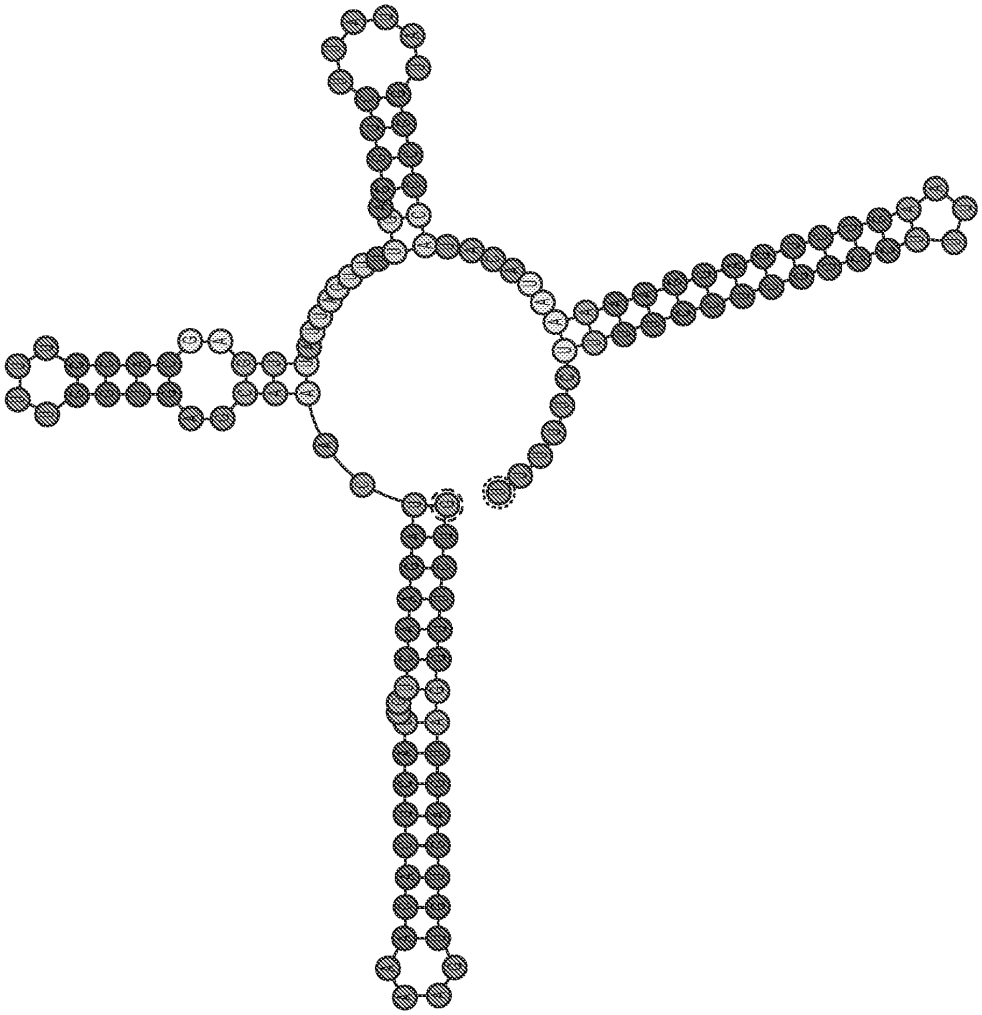
도면2c



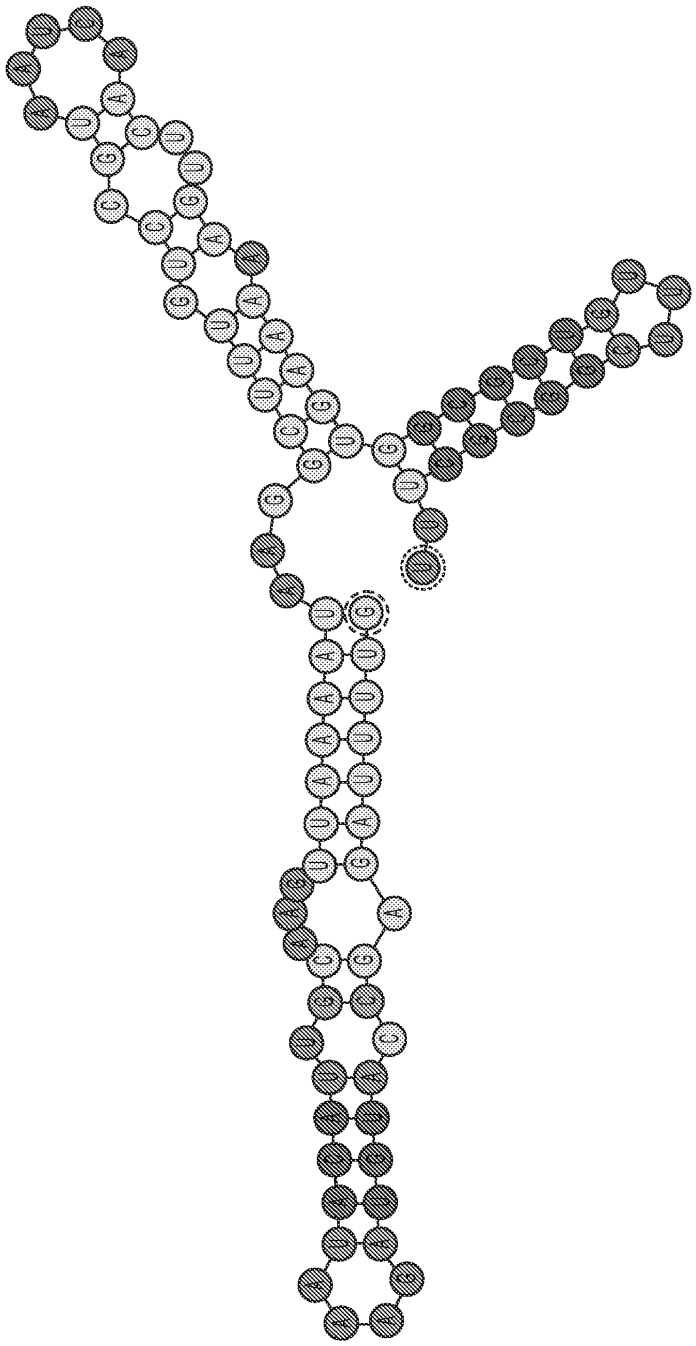
도면2d



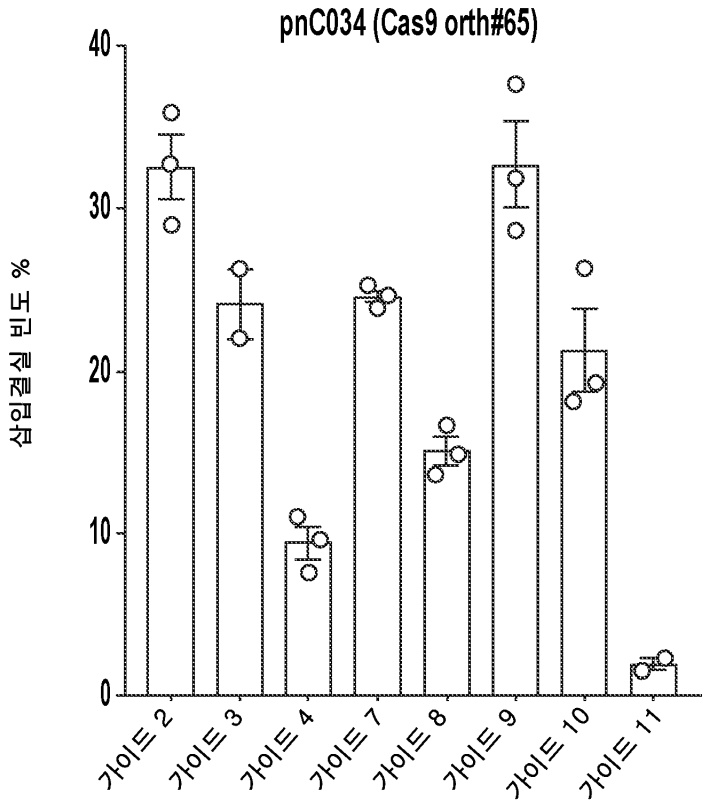
도면2e



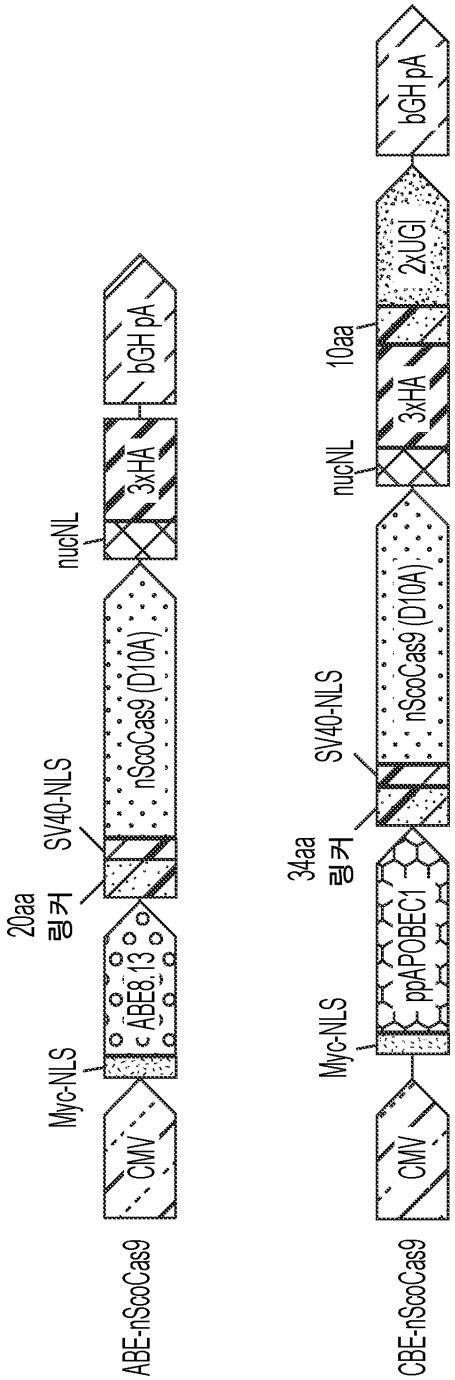
도면2f



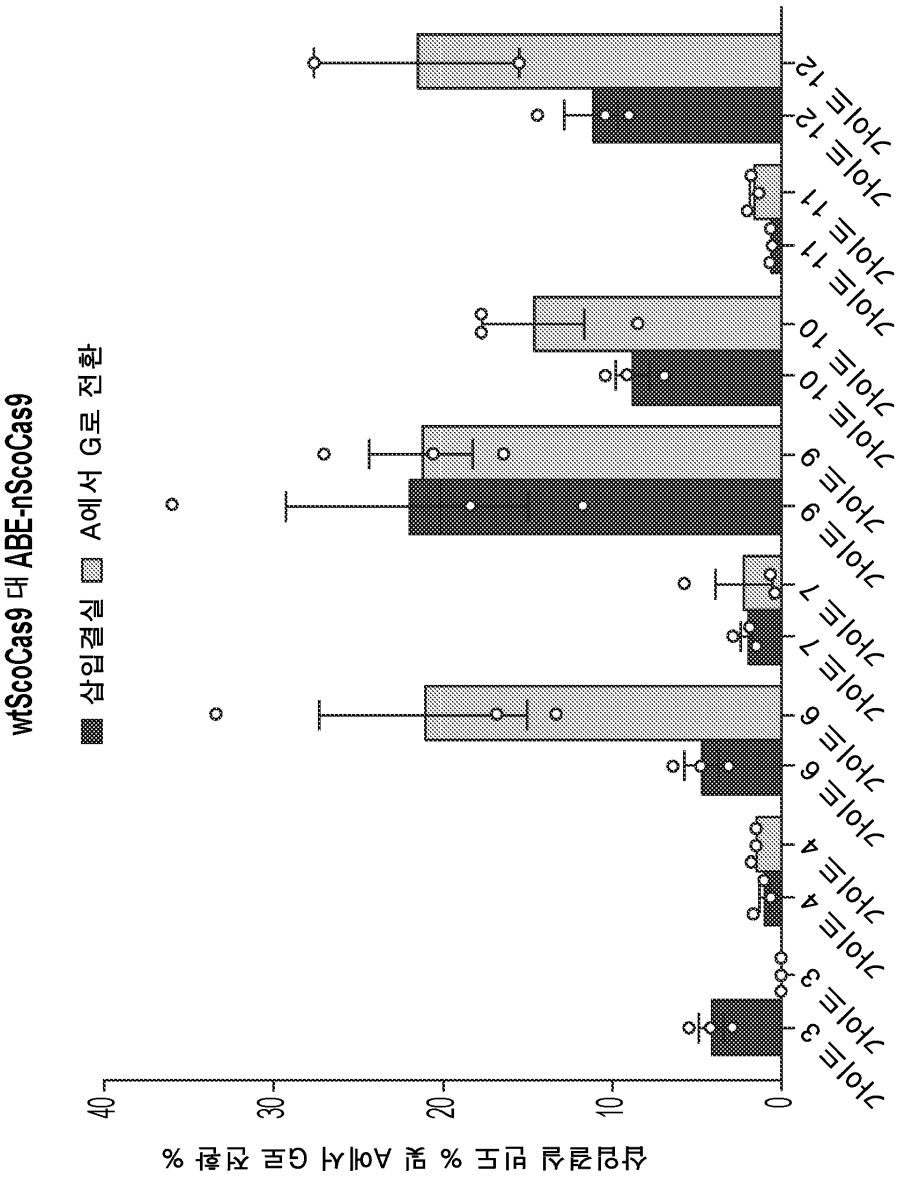
도면3


pnC034 (Cas9 orth#65)
삽입결실 빈도 %
40
30
20
10
0
가이드 2
가이드 3
가이드 4
가이드 7
가이드 8
가이드 9
가이드 10
가이드 11

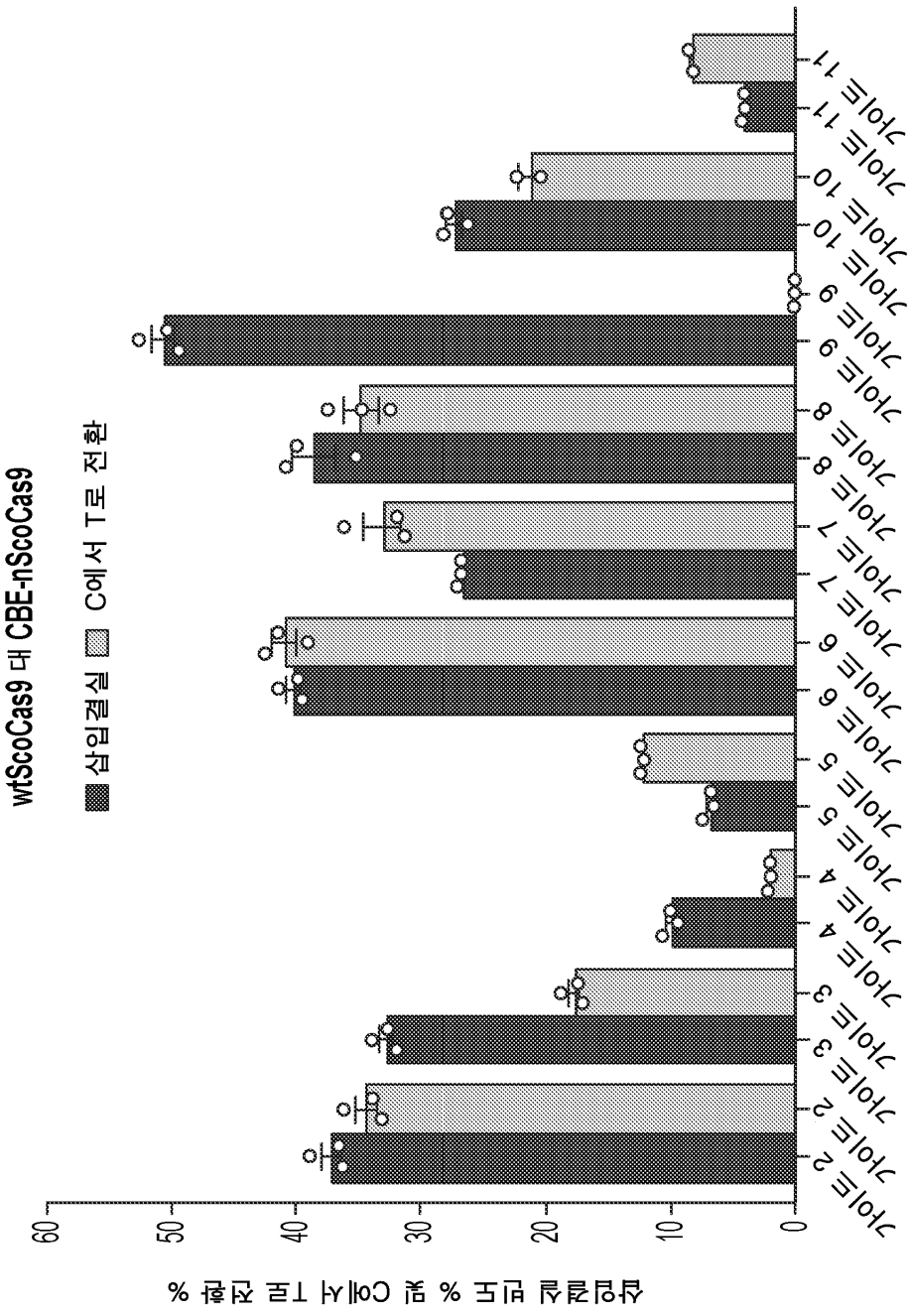
도면4a


ABE-nScoCas9
CMV
Myc-NLS
ABE8.13
20aa
링커
SV40-NLS
nScoCas9 (D10A)
nucNL
3xHA
bGH pA
CBE-nScoCas9
CMV
Myc-NLS
ppAPOBEC1
34aa
링커
SV40-NLS
nScoCas9 (D10A)
nucNL
10aa
3xHA
2xUGI
bGH pA

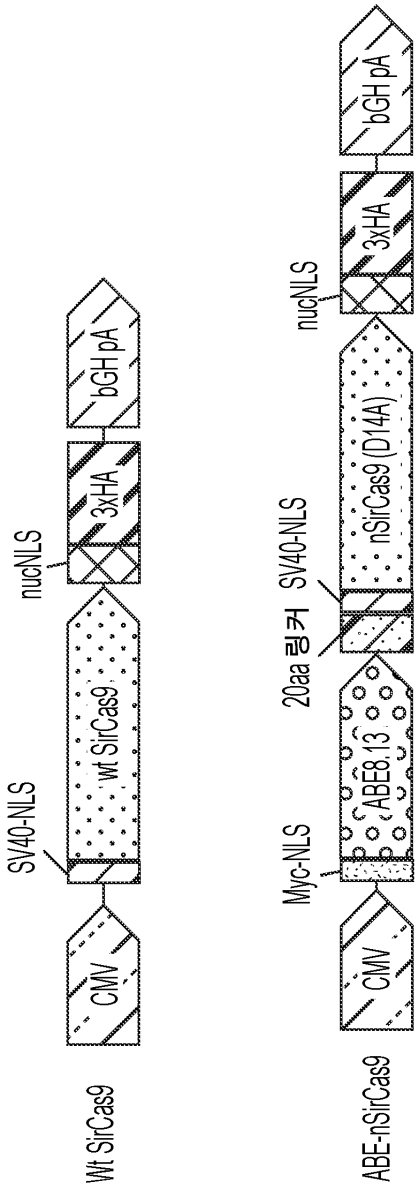
도면4b


wtScoCas9 대 ABE-nScoCas9
삽입결실
A에서 G로 전환
삽입결실 빈도 % 및 A에서 G로 전환 %
0
10
20
30
40
가이드 3
가이드 3
가이드 4
가이드 4
가이드 6
가이드 6
가이드 7
가이드 7
가이드 9
가이드 9
가이드 10
가이드 10
가이드 11
가이드 11
가이드 12
가이드 12

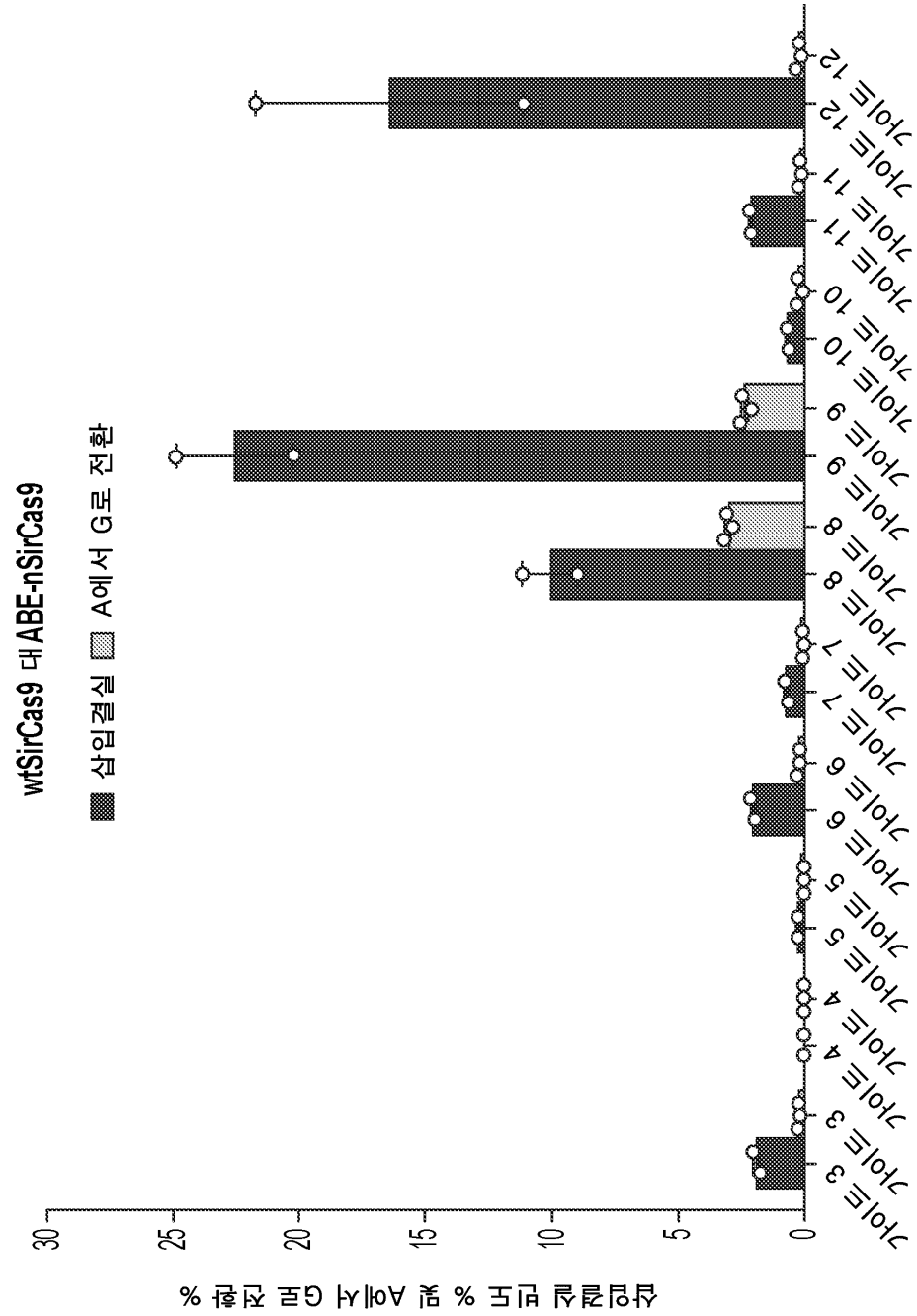
도면4c


wtScoCas9 대 CBE-nScoCas9
삽입결실 C에서 T로 전환
삽입결실 빈도 % 및 C에서 T로 전환 %
0
10
20
30
40
50
60
가이드 2
가이드 2
가이드 3
가이드 3
가이드 4
가이드 4
가이드 5
가이드 5
가이드 6
가이드 6
가이드 7
가이드 7
가이드 8
가이드 8
가이드 9
가이드 9
가이드 10
가이드 10
가이드 11
가이드 11

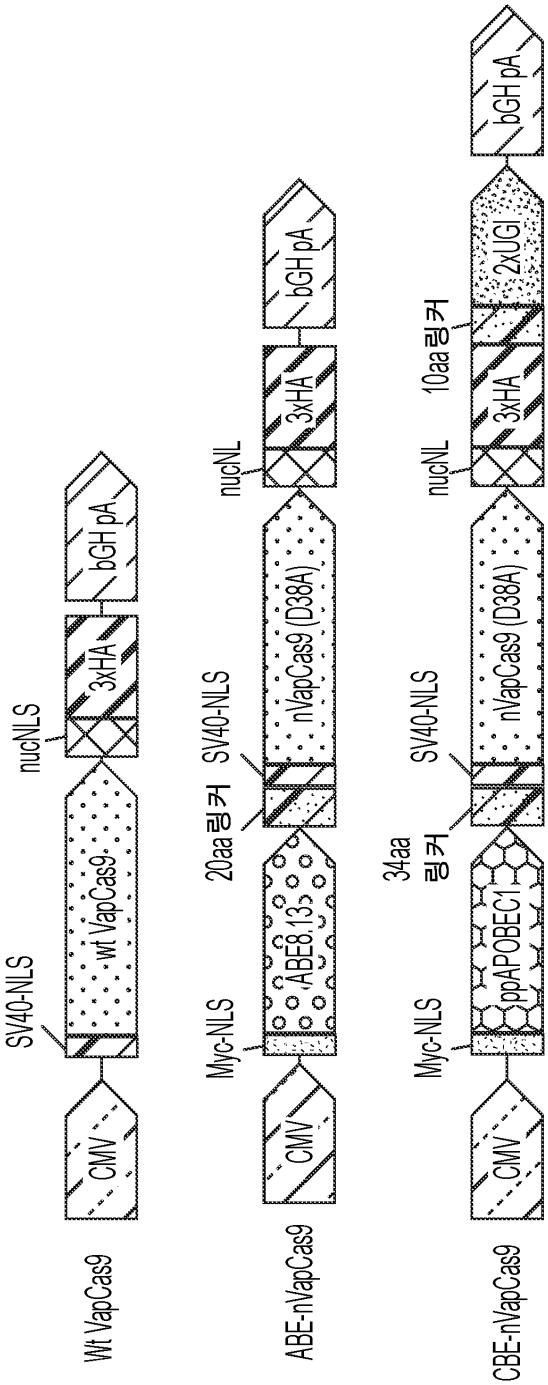
도면5a


Wt SirCas9
CMV
SV40-NLS
wt SirCas9
nucNLS
3xHA
bGH pA
ABE-nSirCas9
CMV
Myc-NLS
ABE8.13
20aa 링커
SV40-NLS
nSirCas9 (D14A)
nucNLS
3xHA
bGH pA

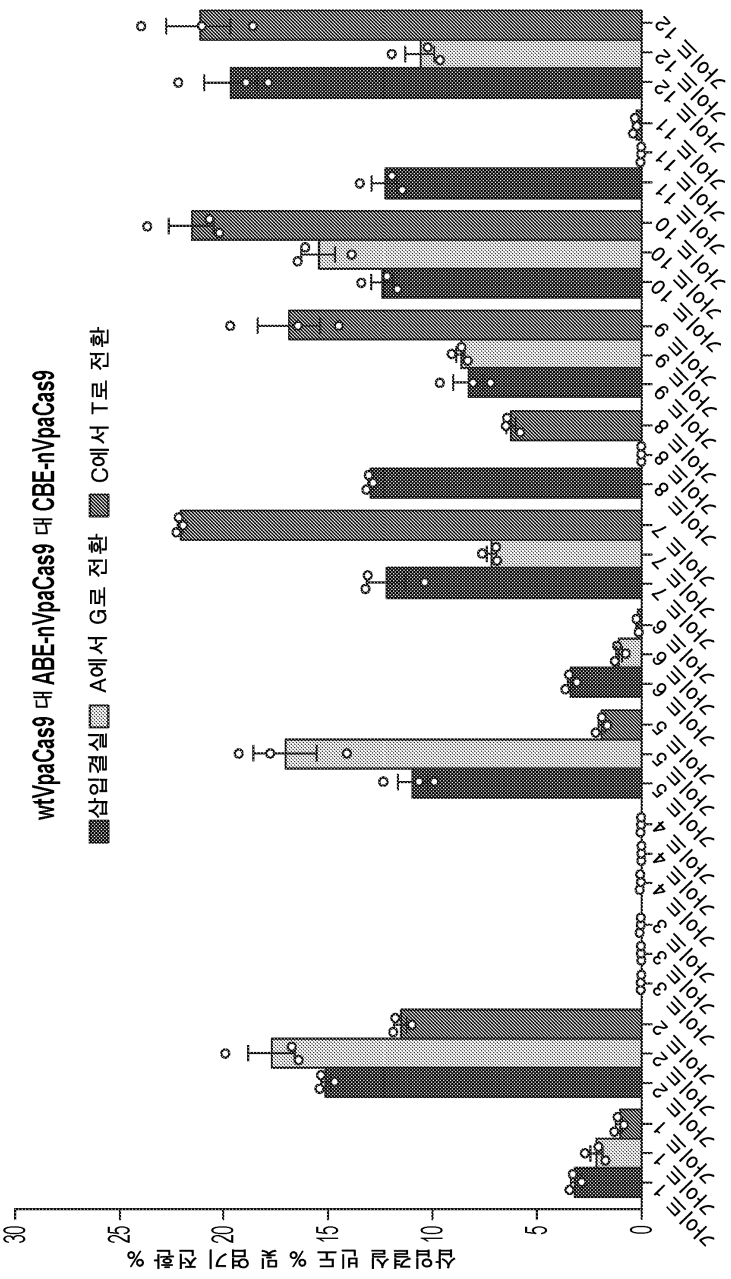
도면5b


wtSirCas9 대 ABE-nSirCas9
삽입결실
A에서 G로 전환
삽입결실 빈도 % 및 A에서 G로 전환 %
0
5
10
15
20
25
30
가이드 3
가이드 3
가이드 4
가이드 4
가이드 5
가이드 5
가이드 6
가이드 6
가이드 7
가이드 7
가이드 8
가이드 8
가이드 9
가이드 9
가이드 10
가이드 10
가이드 11
가이드 11
가이드 12
가이드 12

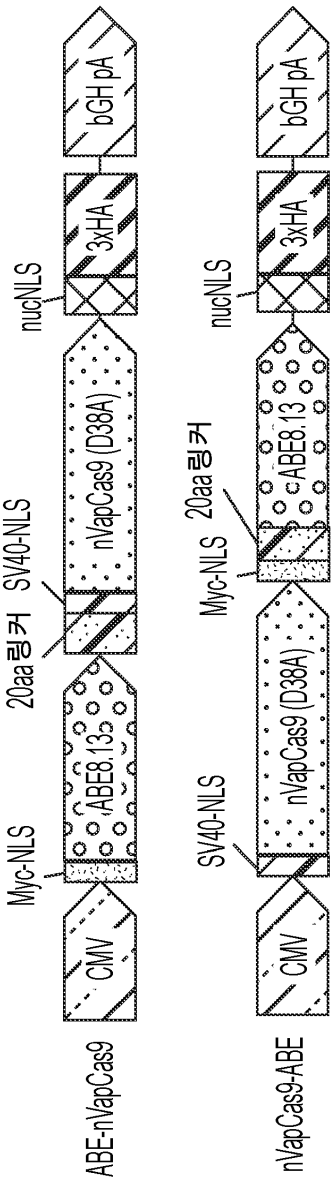
도면6a


Wt VapCas9
CMV
SV40-NLS
wt VapCas9
nucNLS
3xHA
bGH pA
ABE-nVapCas9
CMV
Myc-NLS
ABE8.13
20aa 링커
SV40-NLS
nVapCas9 (D38A)
nucNL
3xHA
bGH pA
CBE-nVapCas9
CMV
Myc-NLS
ppAPOBEC1
34aa 링커
SV40-NLS
nVapCas9 (D38A)
nucNL
10aa 링커
3xHA
2xUGI
bGH pA

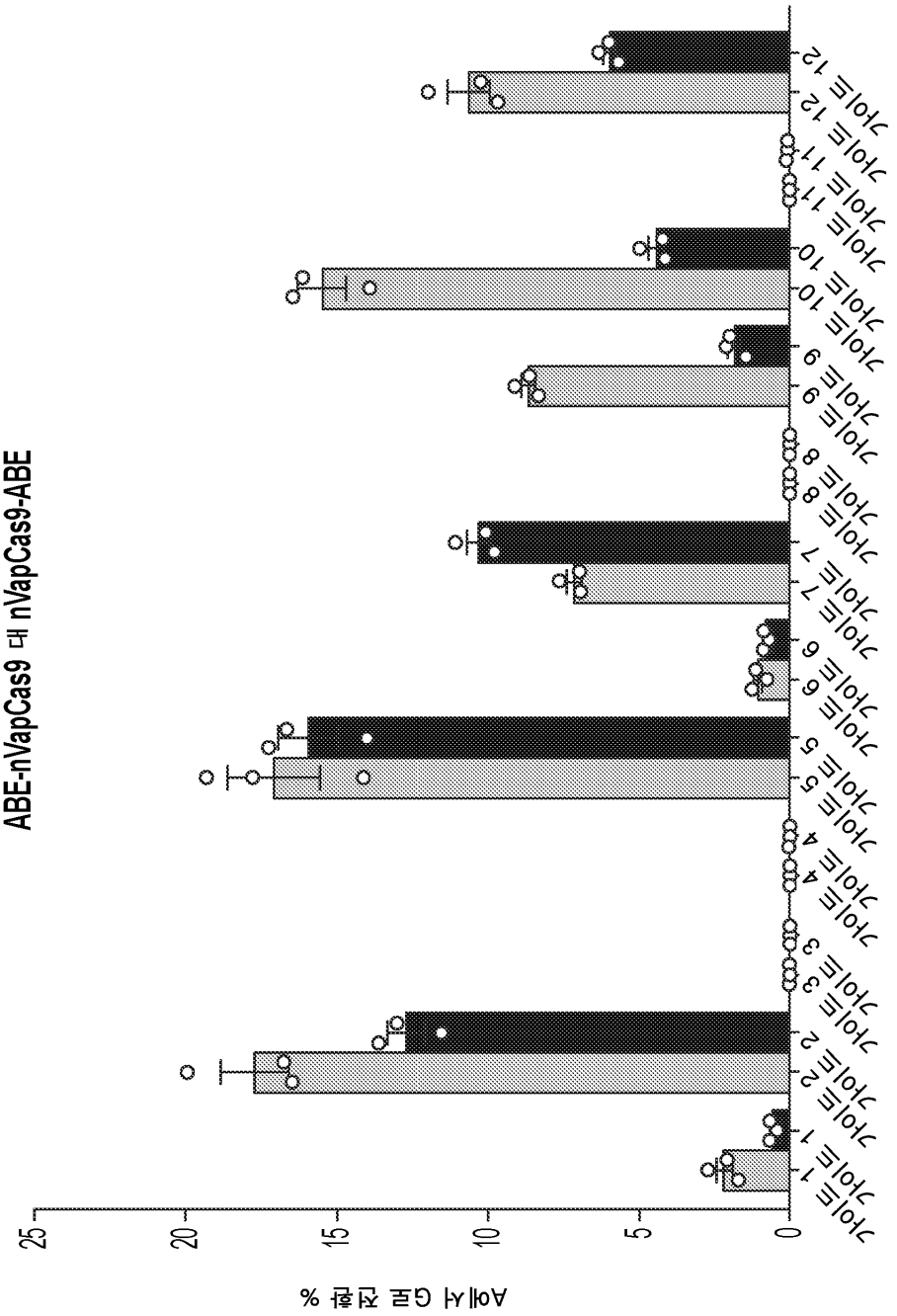
도면6b


wtVpaCas9 대 ABE-nVpaCas9 대 CBE-nVpaCas9
삽입결실 A에서 G로 전환 C에서 T로 전환
삽입결실 빈도 % 및 염기 전환 %
0
5
10
15
20
25
30
가이드 1
가이드 2
가이드 3
가이드 4
가이드 5
가이드 6
가이드 7
가이드 8
가이드 9
가이드 10
가이드 11
가이드 12

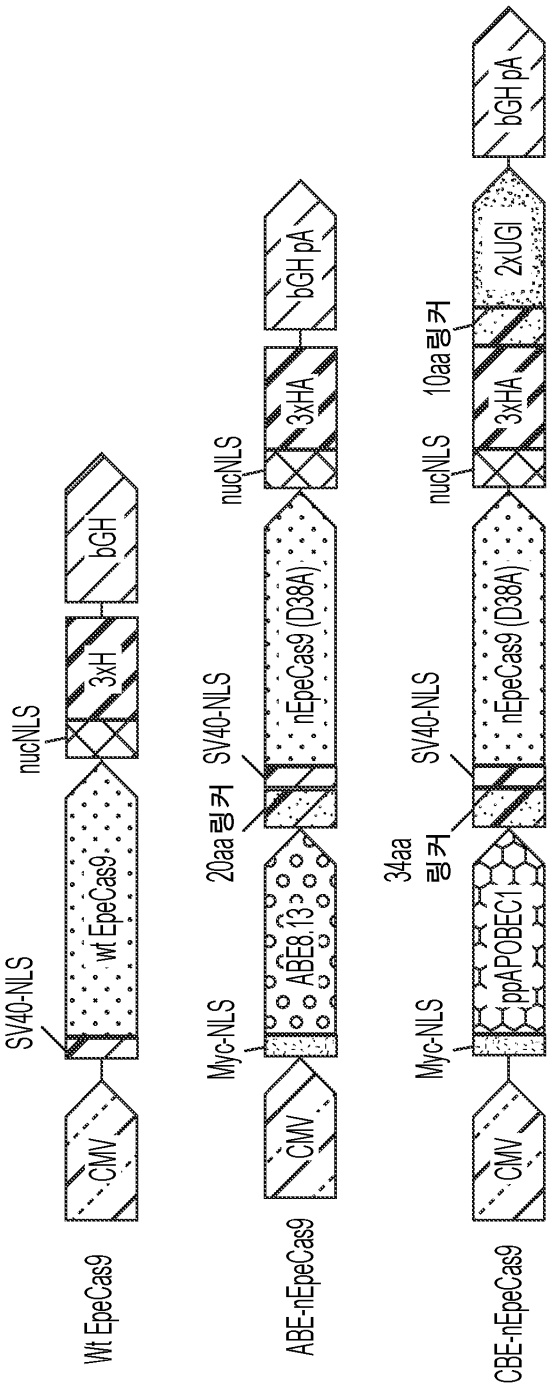
도면7a


ABE-nVapCas9
CMV
Myc-NLS
ABE8.13
20aa 링커
SV40-NLS
nVapCas9 (D38A)
nucNLS
3xHA
bGH pA
nVapCas9-ABE
CMV
SV40-NLS
nVapCas9 (D38A)
Myc-NLS
20aa 링커
cABE8.13
nucNLS
3xHA
bGH pA

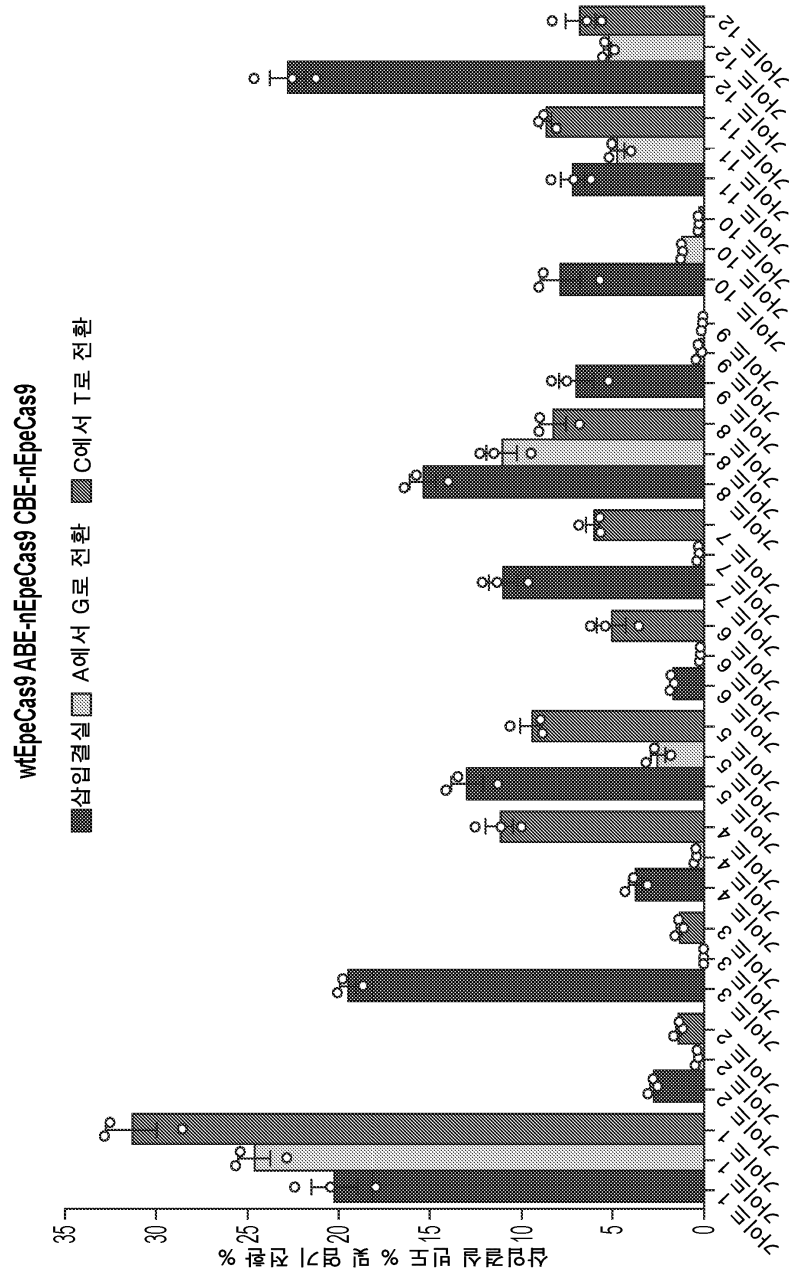
도면7b


ABE-nVapCas9 대 nVapCas9-ABE
A에서 G로 전환 %
0
5
10
15
20
25
가이드 1
가이드 1
가이드 2
가이드 2
가이드 3
가이드 3
가이드 4
가이드 4
가이드 5
가이드 5
가이드 6
가이드 6
가이드 7
가이드 7
가이드 8
가이드 8
가이드 9
가이드 9
가이드 10
가이드 10
가이드 11
가이드 11
가이드 12
가이드 12

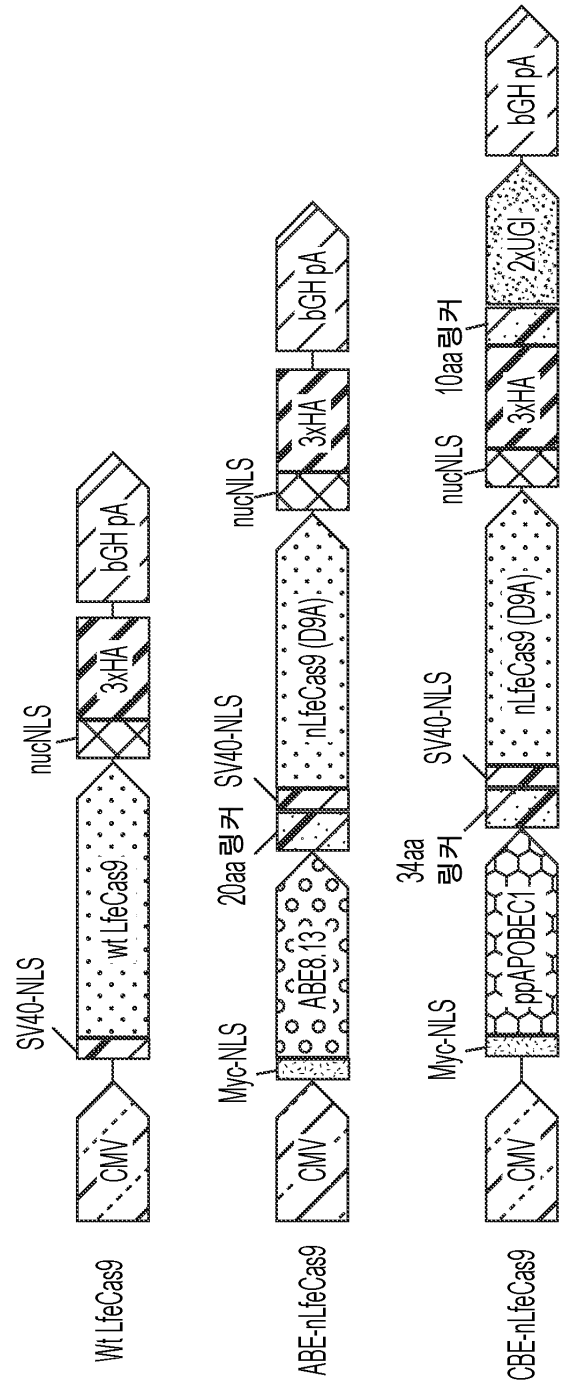
도면8a


Wt EpeCas9
CMV
SV40-NLS
wt EpeCas9
nucNLS
3xH
bGH
ABE-nEpeCas9
CMV
Myc-NLS
ABE8.13
20aa 링커
SV40-NLS
nEpeCas9 (D38A)
nucNLS
3xHA
bGH pA
CBE-nEpeCas9
CMV
Myc-NLS
ppAPOBEC1
34aa 링커
SV40-NLS
nEpeCas9 (D38A)
nucNLS
3xHA
10aa 링커
2xUGI
bGH pA

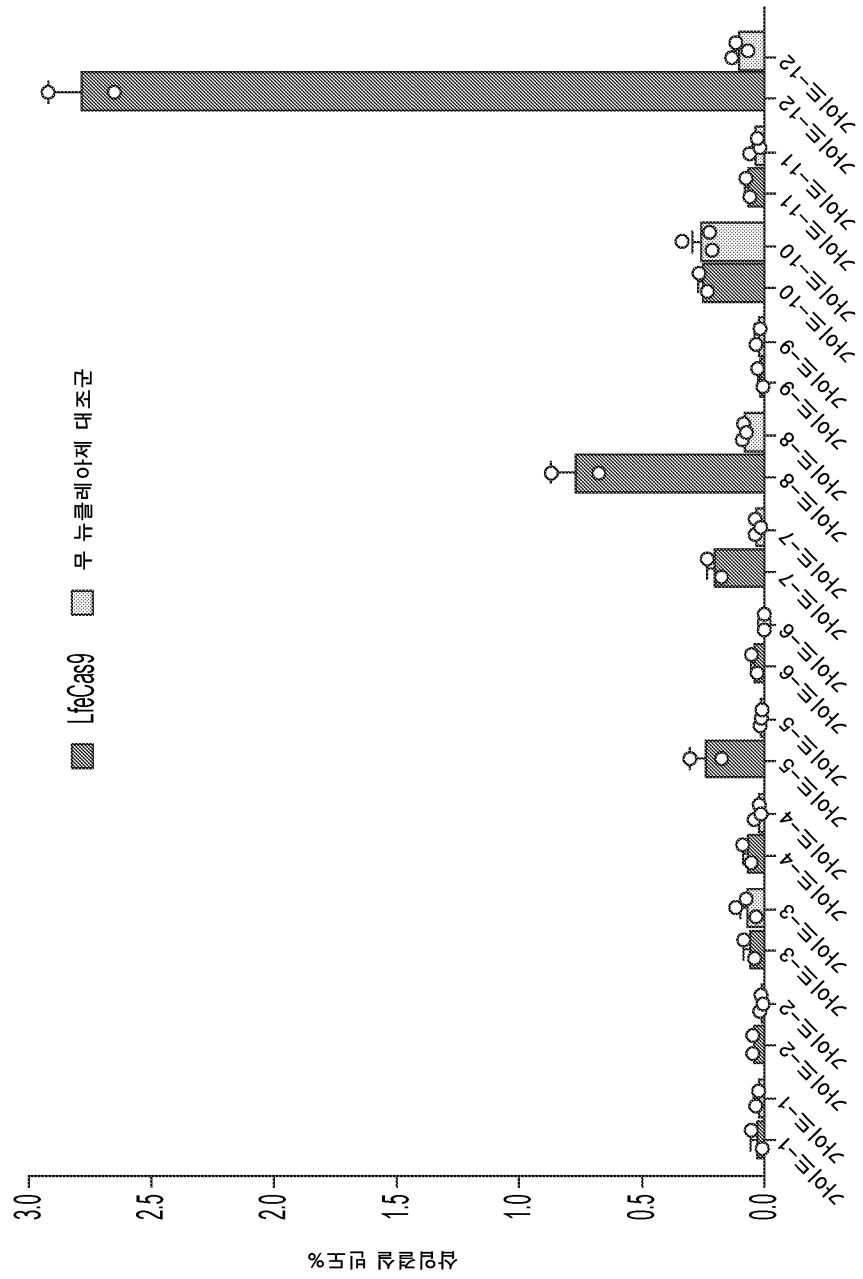
도면8b


wtEpeCas9 ABE-nEpeCas9 CBE-nEpeCas9
삽입결실 A에서 G로 전환 C에서 T로 전환
삽입결실 빈도 % 및 염기 전환 %
0
5
10
15
20
25
30
35
가이드 1
가이드 2
가이드 3
가이드 4
가이드 5
가이드 6
가이드 7
가이드 8
가이드 9
가이드 10
가이드 11
가이드 12

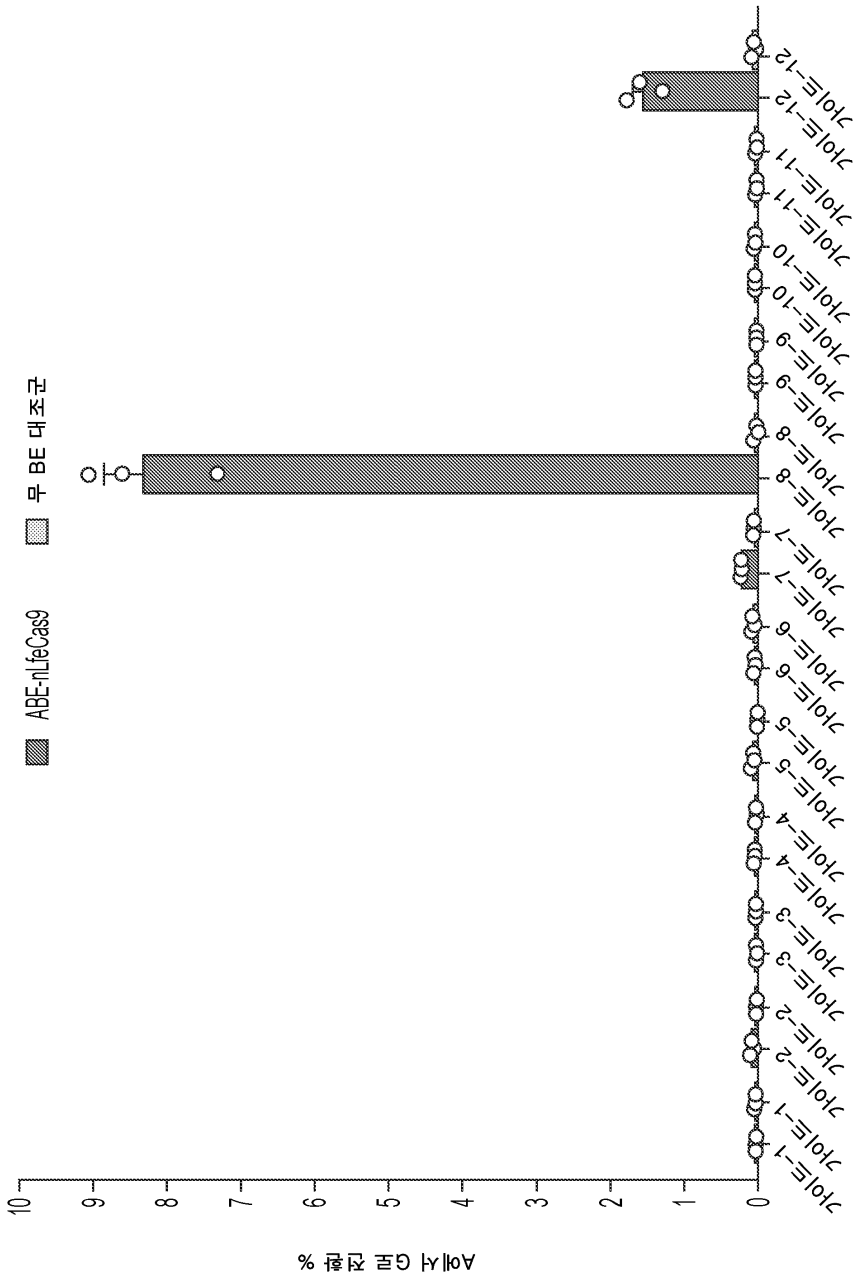
도면9a


Wt LfeCas9
CMV
SV40-NLS
wt LfeCas9
nucNLS
3xHA
bGH pA
ABE-nLfeCas9
CMV
Myc-NLS
ABE8.13
20aa 링커
SV40-NLS
nLfeCas9 (D9A)
nucNLS
3xHA
bGH pA
CBE-nLfeCas9
CMV
Myc-NLS
ppAPOBEC1
34aa 링커
SV40-NLS
nLfeCas9 (D9A)
nucNLS
10aa 링커
3xHA
2xUGI
bGH pA

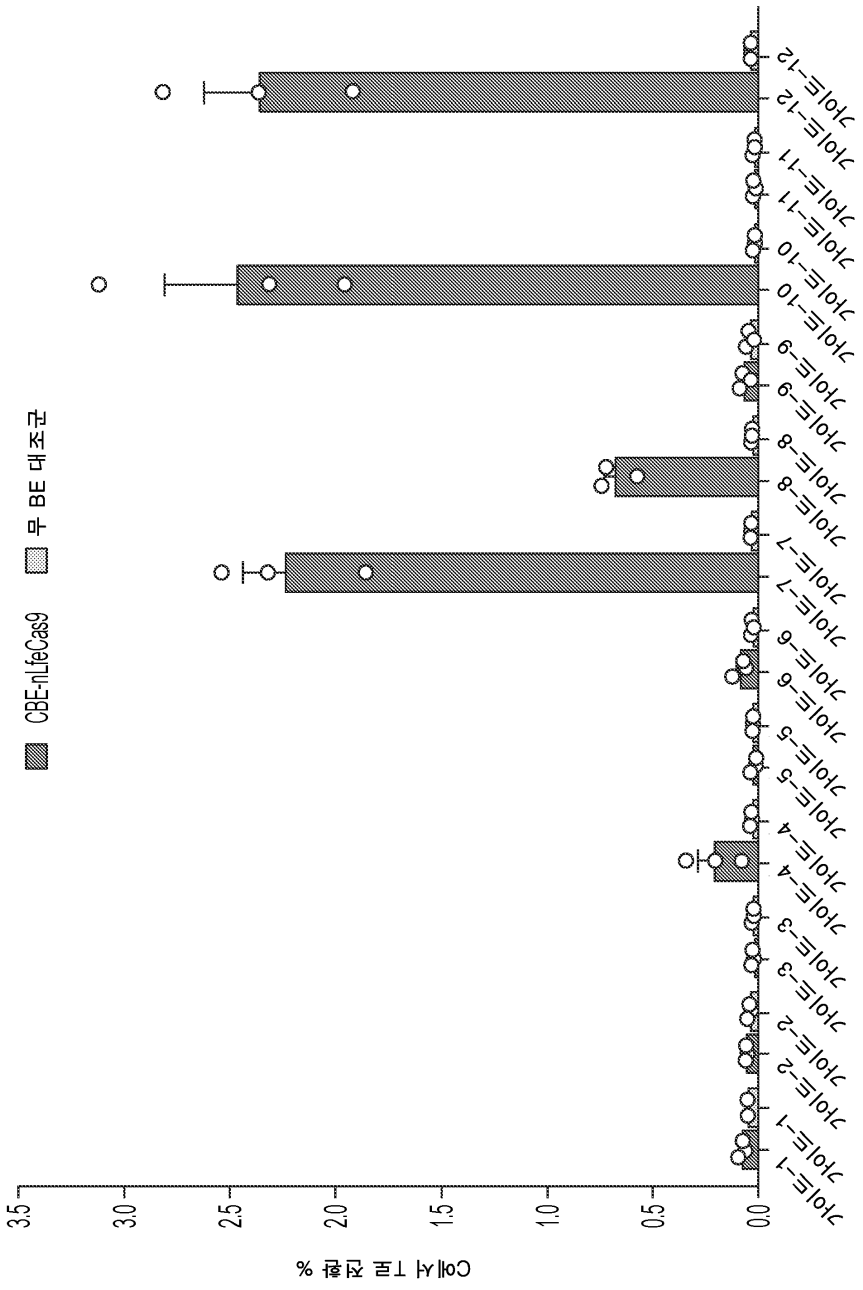
도면9b


LbCas9
무 뉴클레아제 대조군
삽입결실 빈도%
3.0
2.5
2.0
1.5
1.0
0.5
0.0
가이드-1
가이드-2
가이드-3
가이드-4
가이드-5
가이드-6
가이드-7
가이드-8
가이드-9
가이드-10
가이드-11
가이드-12

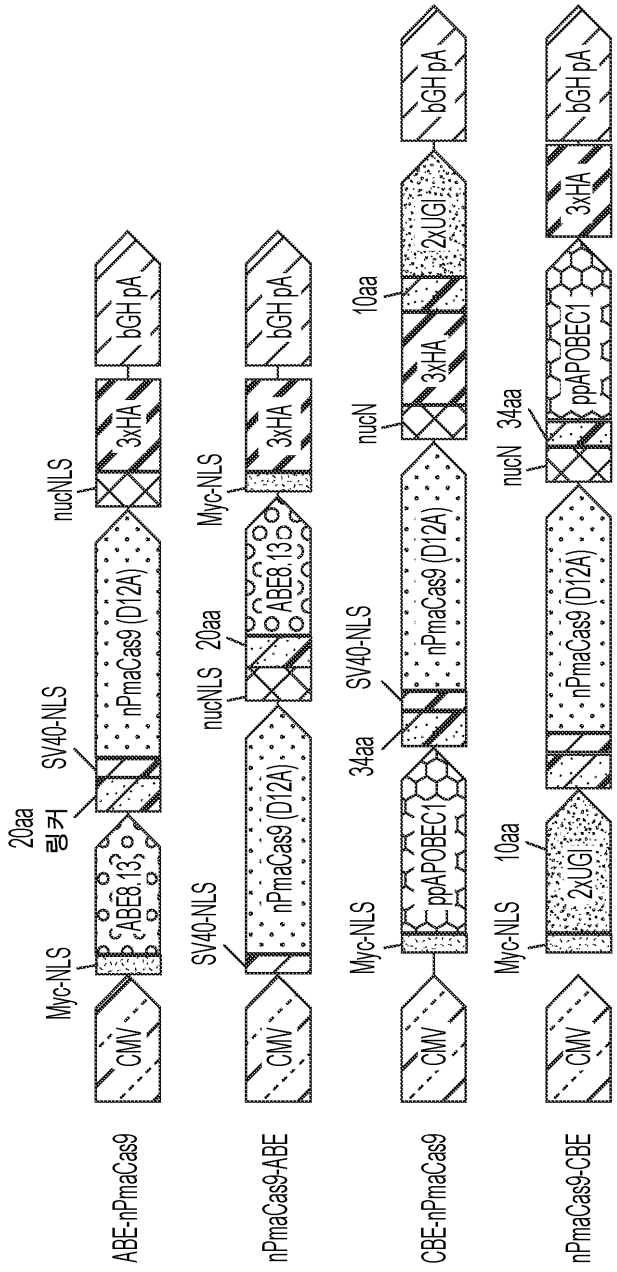
도면9c


ABE-nLfeCas9
무 BE 대조군
A에서 G로 전환 %
0
1
2
3
4
5
6
7
8
9
10
가이드-1
가이드-1
가이드-2
가이드-2
가이드-3
가이드-3
가이드-4
가이드-4
가이드-5
가이드-5
가이드-6
가이드-6
가이드-7
가이드-7
가이드-8
가이드-8
가이드-9
가이드-9
가이드-10
가이드-10
가이드-11
가이드-11
가이드-12
가이드-12

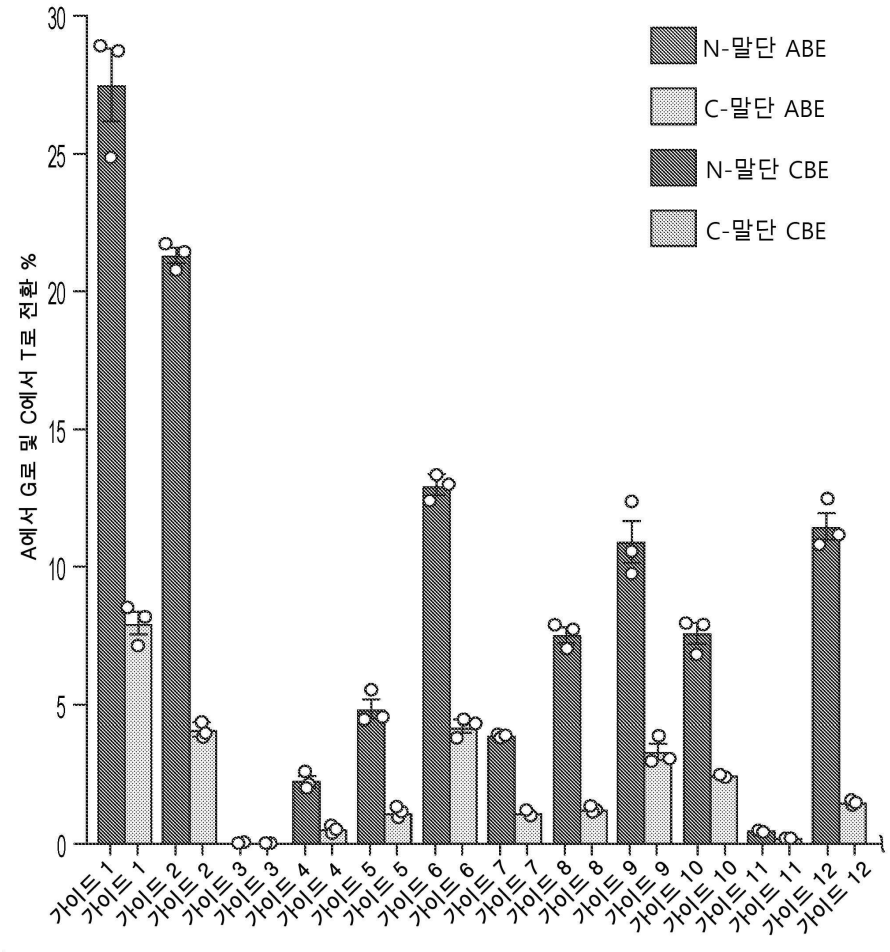
도면9d


CBE-nLiéCas9
무 BE 대조군
Cell 내 T로 전환 %
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
가이드-1
가이드-2
가이드-3
가이드-4
가이드-5
가이드-6
가이드-7
가이드-8
가이드-9
가이드-10
가이드-11
가이드-12

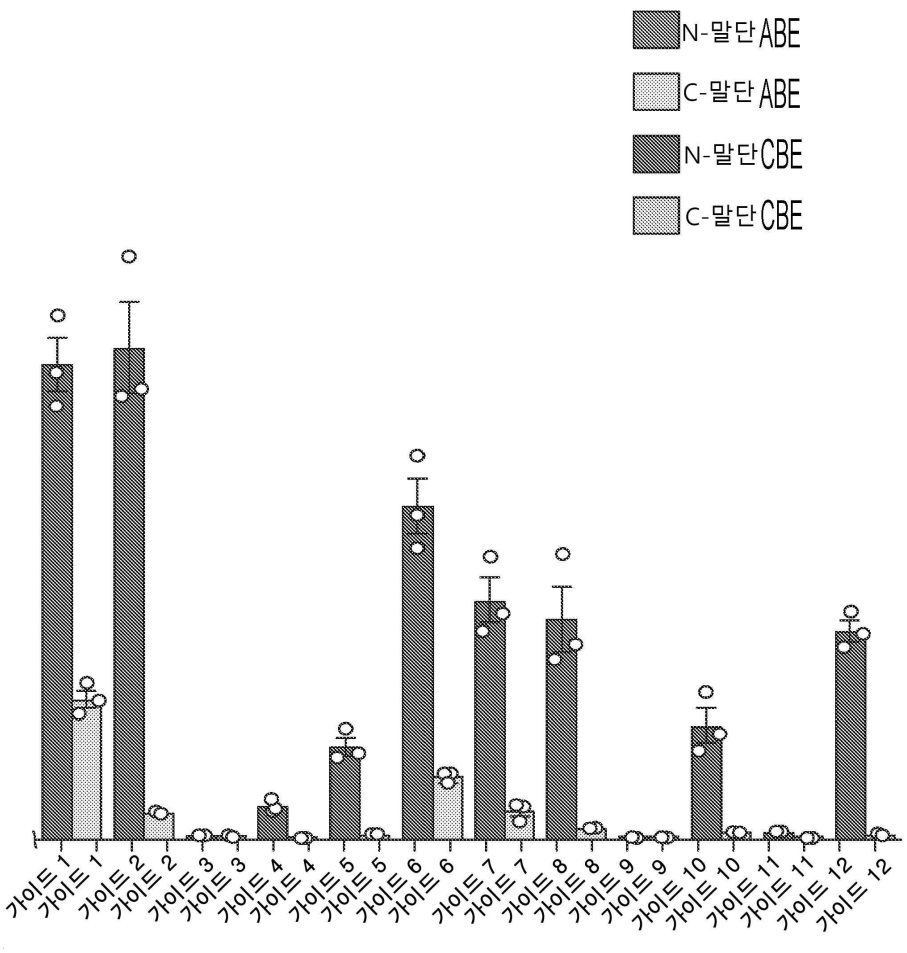
도면10a


ABE-nPmaCas9
CMV
Myc-NLS
ABE8.13
20aa
링커
SV40-NLS
nPmaCas9 (D12A)
nucNLS
3xHA
bGH pA
nPmaCas9-ABE
CMV
SV40-NLS
nPmaCas9 (D12A)
nucNLS
20aa
ABE8.13
Myc-NLS
3xHA
bGH pA
CBE-nPmaCas9
CMV
Myc-NLS
ppAPOBEC1
34aa
SV40-NLS
nPmaCas9 (D12A)
nucN
10aa
3xHA
2xUGI
bGH pA
nPmaCas9-CBE
CMV
Myc-NLS
10aa
2xUGI
nPmaCas9 (D12A)
nucN
34aa
ppAPOBEC1
3xHA
bGH pA

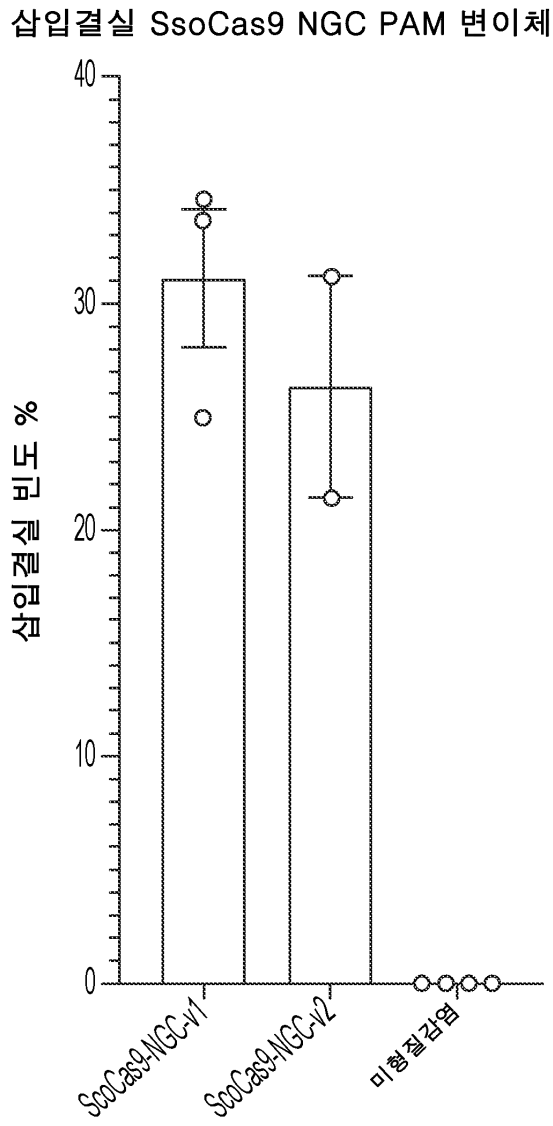
도면10ba


N-말단 ABE
C-말단 ABE
N-말단 CBE
C-말단 CBE
A에서 G로 및 C에서 T로 전환 %
30
25
20
15
10
5
0
가이드 1
가이드 1
가이드 2
가이드 2
가이드 3
가이드 3
가이드 4
가이드 4
가이드 5
가이드 5
가이드 6
가이드 6
가이드 7
가이드 7
가이드 8
가이드 8
가이드 9
가이드 9
가이드 10
가이드 10
가이드 11
가이드 11
가이드 12
가이드 12

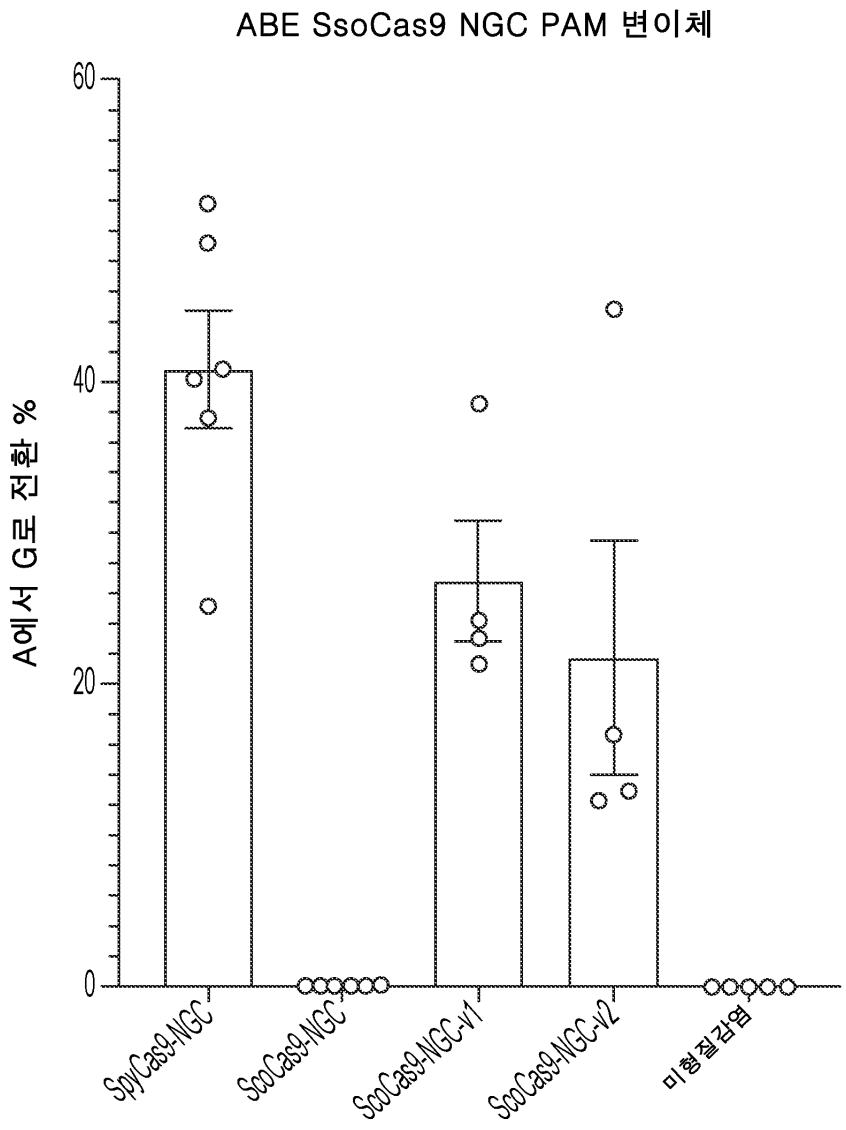
도면10bb


N-말단ABE
C-말단ABE
N-말단CBE
C-말단CBE
가이드 1
가이드 1
가이드 2
가이드 2
가이드 3
가이드 3
가이드 4
가이드 4
가이드 5
가이드 5
가이드 6
가이드 6
가이드 7
가이드 7
가이드 8
가이드 8
가이드 9
가이드 9
가이드 10
가이드 10
가이드 11
가이드 11
가이드 12
가이드 12

도면11a


삽입결실 SsoCas9 NGC PAM 변이체
삽입결실 빈도 %
40
30
20
10
0
SsoCas9-NGC-v1
SsoCas9-NGC-v2
미형질검염

도면11b



서 열 목 록

SEQUENCE LISTING

<110> BEAM THERAPEUTICS INC.

<120> NOVEL CRISPR ENZYMES, METHODS, SYSTEMS AND USES THEREOF

<130> BEM-004WO1

<140> PCT/US2022/021523

<141> 2022-03-23

<150> 63/164,798

<151> 2021-03-23

<160> 182

<170> PatentIn version 3.5

<210> 1

<211> 1345

<212> PRT

<213> Streptococcus constellatus

<400> 1

Met Gly Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Val Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys Met
20 25 30

Lys Val Leu Gly Asn Thr Asp Lys Gln Ser Ile Lys Lys Asn Leu Leu
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Leu Arg
65 70 75 80

Tyr Leu Gln Glu Ile Phe Thr Gly Glu Met Asn Lys Val Asp Glu Asn
85 90 95

Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val Asp Glu Asp Lys Arg
100 105 110

Gly Glu His His Pro Ile Phe Gly Asn Ile Ala Ala Glu Val Lys Tyr
115 120 125

His Asp Asp Phe Pro Thr Ile Tyr His Leu Arg Arg His Leu Ala Asp
130 135 140

Thr Ser Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Tyr Glu Gly Asp Leu Lys Ala
165 170 175

Glu Asn Thr Asp Val Gln Ala Leu Phe Lys Asp Phe Val Glu Glu Tyr
180 185 190

Asp Lys Thr Ile Glu Glu Ser His Leu Ser Glu Ile Thr Val Asp Ala
195 200 205

Leu Ser Ile Leu Thr Glu Lys Val Ser Lys Ser Ser Arg Leu Glu Asn

210 215 220

Leu Ile Ala His Tyr Pro Thr Glu Lys Lys Asn Thr Leu Phe Gly Asn

225 230 235 240

Leu Ile Ala Leu Ser Leu Asp Leu His Pro Asn Phe Lys Thr Asn Phe

245 250 255

Gln Leu Ser Glu Asp Ala Lys Leu Gln Phe Ser Lys Asp Thr Tyr Glu

260 265 270

Glu Asp Leu Glu Gly Phe Leu Gly Glu Val Gly Asp Glu Tyr Ala Asp

275 280 285

Leu Phe Ala Ser Ala Lys Asn Leu Tyr Asp Ala Ile Leu Leu Ser Gly

290 295 300

Ile Leu Thr Val Asp Asp Asn Ser Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320

Met Val Lys Arg Tyr Glu Glu His Gln Lys Asp Leu Lys Lys Leu Lys

325 330 335

Asp Phe Ile Lys Val Asn Ala Pro Asp Gln Tyr Asn Ala Ile Phe Lys

340 345 350

Asp Lys Asn Lys Lys Gly Tyr Ala Ser Tyr Ile Glu Ser Gly Val Lys

355 360 365

Gln Asp Glu Phe Tyr Lys Tyr Leu Lys Gly Ile Leu Leu Lys Ile Asn

370 375 380

Gly Ser Gly Asp Phe Leu Asp Lys Ile Asp Arg Glu Asp Phe Leu Arg

385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu

405 410 415

Gln Glu Met His Ala Ile Leu Arg Arg Gln Gly Glu His Tyr Pro Phe

420 425 430

Leu Lys Glu Asn Gln Asp Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile

435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Lys Gly Ser Arg Phe Ala Trp

450 455 460

Ala Glu Tyr Lys Ala Asp Glu Lys Ile Thr Pro Trp Asn Phe Asp Asp

465 470 475 480

Ile Leu Asp Lys Glu Lys Ser Ala Glu Lys Phe Ile Thr Arg Met Thr

485 490 495

Leu Asn Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser

500 505 510

Pro Leu Tyr Glu Ala Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys

515 520 525

Tyr Val Asn Glu Gln Gly Glu Ala Lys Phe Phe Asp Thr Asn Met Lys

530 535 540

Gln Glu Ile Phe Asp His Val Phe Lys Glu Asn Arg Lys Val Thr Lys

545 550 555 560

Asp Lys Leu Leu Asn Tyr Leu Asn Lys Glu Phe Glu Glu Phe Arg Ile

565 570 575

Val Asn Leu Thr Gly Leu Asp Lys Glu Asn Lys Ala Phe Asn Ser Ser

580 585 590

Leu Gly Thr Tyr His Asp Leu Arg Lys Ile Leu Asp Lys Ser Phe Leu

595 600 605

Asp Asp Lys Ala Asn Glu Lys Thr Ile Glu Asp Ile Ile Gln Thr Leu

610 615 620

Thr Leu Phe Glu Asp Arg Glu Met Ile Arg Gln Arg Leu Gln Lys Tyr

625 630 635 640

Ser Asp Ile Phe Thr Lys Ala Gln Leu Lys Lys Leu Glu Arg Arg His

645 650 655

Tyr Thr Gly Trp Gly Arg Leu Ser Tyr Lys Leu Ile Asn Gly Ile Arg

660 665 670

Asn Lys Glu Asn Lys Lys Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly

675 680 685

Tyr Ala Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Ala Leu Ser

690 695 700

Phe Lys Glu Glu Ile Ala Arg Ala Gln Ile Ile Asp Asp Val Asp Asp
705 710 715 720
Ile Ala Asn Val Val His Asp Leu Pro Gly Ser Pro Ala Ile Lys Lys

725 730 735
Gly Ile Leu Gln Ser Val Lys Ile Val Asp Glu Leu Val Lys Val Met
740 745 750
Gly His Asn Pro Ala Asn Ile Ile Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Asp Lys Gly Arg Arg Asn Ser Gln Gln Arg Leu Lys Leu Leu
770 775 780
Gln Asp Ser Leu Lys Asn Leu Asp Asn Pro Val Asn Ile Lys Asn Val

785 790 795 800
Glu Asn Gln Gln Leu Gln Asn Asp Arg Leu Phe Leu Tyr Tyr Ile Gln
805 810 815
Asn Gly Lys Asp Met Tyr Thr Gly Glu Thr Leu Asp Ile Asn Asn Leu
820 825 830
Ser Gln Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe Ile Lys Asp
835 840 845
Asn Ser Leu Asp Asn Arg Val Leu Thr Arg Ser Asp Lys Asn Arg Gly

850 855 860
Lys Ser Asp Asp Val Pro Ser Ile Glu Val Val His Glu Met Lys Ser
865 870 875 880
Phe Trp Ser Lys Leu Leu Ser Val Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Thr Glu Glu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925
His Val Ala Gln Ile Leu Asp Glu Arg Phe Asn Thr Glu Phe Asp Gly
930 935 940
Asn Lys Arg Arg Ile Arg Asn Val Lys Ile Ile Thr Leu Lys Ser Asn

945 950 955 960

Leu Val Ser Asn Phe Arg Lys Glu Phe Glu Leu Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

980 985 990

Gly Asn Ala Leu Leu Leu Lys Tyr Pro Gln Leu Glu Pro Glu Phe Val

995 1000 1005

Tyr Gly Glu Tyr Pro Lys Tyr Asn Ser Tyr Arg Ser Arg Lys Ser

1010 1015 1020

Ala Thr Glu Lys Phe Leu Phe Tyr Ser Asn Ile Leu Arg Phe Phe

1025 1030 1035

Lys Lys Glu Asp Ile Gln Thr Asn Glu Asp Gly Glu Ile Ala Trp

1040 1045 1050

Asn Lys Glu Lys His Ile Lys Ile Leu Arg Lys Val Leu Ser Tyr

1055 1060 1065

Pro Gln Val Asn Ile Val Lys Lys Thr Glu Glu Gln Thr Gly Gly

1070 1075 1080

Phe Ser Lys Glu Ser Ile Leu Pro Lys Gly Glu Ser Asp Lys Leu

1085 1090 1095

Ile Pro Arg Lys Thr Lys Asn Ser Tyr Trp Asp Pro Lys Lys Tyr

1100 1105 1110

Gly Gly Phe Asp Ser Pro Val Val Ala Tyr Ser Ile Leu Val Phe

1115 1120 1125

Ala Asp Val Glu Lys Gly Lys Ser Lys Lys Leu Arg Lys Val Gln

1130 1135 1140

Asp Met Val Gly Ile Thr Ile Met Glu Lys Lys Arg Phe Glu Lys

1145 1150 1155

Asn Pro Val Asp Phe Leu Glu Gln Arg Gly Tyr Arg Asn Val Arg

1160 1165 1170

Leu Glu Lys Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu

1175 1180 1185

Glu Asn Lys Arg Arg Arg Leu Leu Ala Ser Ala Lys Glu Leu Gln
1190 1195 1200
Lys Gly Asn Glu Leu Val Ile Pro Gln Arg Phe Thr Thr Leu Leu
1205 1210 1215
Tyr His Ser Tyr Arg Ile Glu Lys Asp Tyr Glu Pro Glu His Arg

1220 1225 1230
Glu Tyr Val Glu Lys His Lys Asp Glu Phe Lys Glu Leu Leu Glu
1235 1240 1245
Tyr Ile Ser Val Phe Ser Arg Lys Tyr Val Leu Ala Asp Asn Asn
1250 1255 1260
Leu Thr Lys Ile Glu Met Leu Phe Ser Lys Asn Lys Asp Ala Glu
1265 1270 1275
Val Ser Ser Leu Ala Lys Ser Phe Ile Ser Leu Leu Thr Phe Thr

1280 1285 1290
Ala Phe Gly Ala Pro Ala Ala Phe Asn Phe Phe Gly Glu Asn Ile
1295 1300 1305
Asp Arg Lys Arg Tyr Thr Ser Val Thr Glu Cys Leu Asn Ala Thr
1310 1315 1320
Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1325 1330 1335
Leu Ser Lys Leu Gly Glu Asp
1340 1345

<210> 2
<211> 1399
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 2
Met Pro Lys Lys Lys Arg Lys Val Gly Gly Lys Pro Tyr Ser Ile Gly
1 5 10 15
Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Val Thr Asp Asp

```
            20                  25                  30
Tyr Lys Val Pro Ala Lys Lys Met Lys Val Leu Gly Asn Thr Asp Lys
        35                  40                  45

Gln Ser Ile Lys Lys Asn Leu Leu Gly Ala Leu Leu Phe Asp Ser Gly
    50                  55                  60
Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr
65                  70                  75                  80
Thr Arg Arg Lys Asn Arg Leu Arg Tyr Leu Gln Glu Ile Phe Thr Gly
                85                  90                  95
Glu Met Asn Lys Val Asp Glu Asn Phe Phe Gln Arg Leu Asp Asp Ser
            100                 105                 110

Phe Leu Val Asp Glu Asp Lys Arg Gly Glu His His Pro Ile Phe Gly
        115                 120                 125
Asn Ile Ala Ala Glu Val Lys Tyr His Asp Asp Phe Pro Thr Ile Tyr
    130                 135                 140
His Leu Arg Arg His Leu Ala Asp Thr Ser Lys Lys Ala Asp Leu Arg
145                 150                 155                 160
Leu Val Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe
                165                 170                 175

Leu Tyr Glu Gly Asp Leu Lys Ala Glu Asn Thr Asp Val Gln Ala Leu
            180                 185                 190
Phe Lys Asp Phe Val Glu Glu Tyr Asp Lys Thr Ile Glu Glu Ser His
        195                 200                 205
Leu Ser Glu Ile Thr Val Asp Ala Leu Ser Ile Leu Thr Glu Lys Val
    210                 215                 220
Ser Lys Ser Ser Arg Leu Glu Asn Leu Ile Ala His Tyr Pro Thr Glu
225                 230                 235                 240

Lys Lys Asn Thr Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Asp Leu
                245                 250                 255
His Pro Asn Phe Lys Thr Asn Phe Gln Leu Ser Glu Asp Ala Lys Leu
            260                 265                 270
```

Gln Phe Ser Lys Asp Thr Tyr Glu Glu Asp Leu Glu Gly Phe Leu Gly
275 280 285
Glu Val Gly Asp Glu Tyr Ala Asp Leu Phe Ala Ser Ala Lys Asn Leu
290 295 300

Tyr Asp Ala Ile Leu Leu Ser Gly Ile Leu Thr Val Asp Asp Asn Ser
305 310 315 320
Thr Lys Ala Pro Leu Ser Ala Ser Met Val Lys Arg Tyr Glu Glu His
325 330 335
Gln Lys Asp Leu Lys Lys Leu Lys Asp Phe Ile Lys Val Asn Ala Pro
340 345 350
Asp Gln Tyr Asn Ala Ile Phe Lys Asp Lys Asn Lys Lys Gly Tyr Ala
355 360 365

Ser Tyr Ile Glu Ser Gly Val Lys Gln Asp Glu Phe Tyr Lys Tyr Leu
370 375 380
Lys Gly Ile Leu Leu Lys Ile Asn Gly Ser Gly Asp Phe Leu Asp Lys
385 390 395 400
Ile Asp Arg Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly
405 410 415
Ser Ile Pro His Gln Ile His Leu Gln Glu Met His Ala Ile Leu Arg
420 425 430

Arg Gln Gly Glu His Tyr Pro Phe Leu Lys Glu Asn Gln Asp Lys Ile
435 440 445
Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala
450 455 460
Arg Lys Gly Ser Arg Phe Ala Trp Ala Glu Tyr Lys Ala Asp Glu Lys
465 470 475 480
Ile Thr Pro Trp Asn Phe Asp Asp Ile Leu Asp Lys Glu Lys Ser Ala
485 490 495

Glu Lys Phe Ile Thr Arg Met Thr Leu Asn Asp Leu Tyr Leu Pro Glu
500 505 510
Glu Lys Val Leu Pro Lys His Ser Pro Leu Tyr Glu Ala Phe Thr Val

515 520 525

Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Asn Glu Gln Gly Glu Ala

530 535 540

Lys Phe Phe Asp Thr Asn Met Lys Gln Glu Ile Phe Asp His Val Phe

545 550 555 560

Lys Glu Asn Arg Lys Val Thr Lys Asp Lys Leu Leu Asn Tyr Leu Asn

565 570 575

Lys Glu Phe Glu Glu Phe Arg Ile Val Asn Leu Thr Gly Leu Asp Lys

580 585 590

Glu Asn Lys Ala Phe Asn Ser Ser Leu Gly Thr Tyr His Asp Leu Arg

595 600 605

Lys Ile Leu Asp Lys Ser Phe Leu Asp Asp Lys Ala Asn Glu Lys Thr

610 615 620

Ile Glu Asp Ile Ile Gln Thr Leu Thr Leu Phe Glu Asp Arg Glu Met

625 630 635 640

Ile Arg Gln Arg Leu Gln Lys Tyr Ser Asp Ile Phe Thr Lys Ala Gln

645 650 655

Leu Lys Lys Leu Glu Arg Arg His Tyr Thr Gly Trp Gly Arg Leu Ser

660 665 670

Tyr Lys Leu Ile Asn Gly Ile Arg Asn Lys Glu Asn Lys Lys Thr Ile

675 680 685

Leu Asp Tyr Leu Ile Asp Asp Gly Tyr Ala Asn Arg Asn Phe Met Gln

690 695 700

Leu Ile Asn Asp Asp Ala Leu Ser Phe Lys Glu Glu Ile Ala Arg Ala

705 710 715 720

Gln Ile Ile Asp Asp Val Asp Asp Ile Ala Asn Val Val His Asp Leu

725 730 735

Pro Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Val Lys Ile

740 745 750

Val Asp Glu Leu Val Lys Val Met Gly His Asn Pro Ala Asn Ile Ile

755 760 765

Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Asp Lys Gly Arg Arg Asn
770 775 780
Ser Gln Gln Arg Leu Lys Leu Leu Gln Asp Ser Leu Lys Asn Leu Asp
785 790 795 800
Asn Pro Val Asn Ile Lys Asn Val Glu Asn Gln Gln Leu Gln Asn Asp
805 810 815

Arg Leu Phe Leu Tyr Tyr Ile Gln Asn Gly Lys Asp Met Tyr Thr Gly
820 825 830
Glu Thr Leu Asp Ile Asn Asn Leu Ser Gln Tyr Asp Ile Asp His Ile
835 840 845
Ile Pro Gln Ala Phe Ile Lys Asp Asn Ser Leu Asp Asn Arg Val Leu
850 855 860
Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Ile
865 870 875 880

Glu Val Val His Glu Met Lys Ser Phe Trp Ser Lys Leu Leu Ser Val
885 890 895
Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
900 905 910
Gly Gly Leu Thr Glu Glu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu
915 920 925
Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Glu
930 935 940

Arg Phe Asn Thr Glu Phe Asp Gly Asn Lys Arg Arg Ile Arg Asn Val
945 950 955 960
Lys Ile Ile Thr Leu Lys Ser Asn Leu Val Ser Asn Phe Arg Lys Glu
965 970 975
Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Tyr His His Ala His
980 985 990
Asp Ala Tyr Leu Asn Ala Val Val Gly Asn Ala Leu Leu Leu Lys Tyr
995 1000 1005

Pro Gln Leu Glu Pro Glu Phe Val Tyr Gly Glu Tyr Pro Lys Tyr

1010 1015 1020

Asn Ser Tyr Arg Ser Arg Lys Ser Ala Thr Glu Lys Phe Leu Phe

1025 1030 1035

Tyr Ser Asn Ile Leu Arg Phe Phe Lys Lys Glu Asp Ile Gln Thr

1040 1045 1050

Asn Glu Asp Gly Glu Ile Ala Trp Asn Lys Glu Lys His Ile Lys

1055 1060 1065

Ile Leu Arg Lys Val Leu Ser Tyr Pro Gln Val Asn Ile Val Lys

1070 1075 1080

Lys Thr Glu Glu Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu

1085 1090 1095

Pro Lys Gly Glu Ser Asp Lys Leu Ile Pro Arg Lys Thr Lys Asn

1100 1105 1110

Ser Tyr Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Val

1115 1120 1125

Val Ala Tyr Ser Ile Leu Val Phe Ala Asp Val Glu Lys Gly Lys

1130 1135 1140

Ser Lys Lys Leu Arg Lys Val Gln Asp Met Val Gly Ile Thr Ile

1145 1150 1155

Met Glu Lys Lys Arg Phe Glu Lys Asn Pro Val Asp Phe Leu Glu

1160 1165 1170

Gln Arg Gly Tyr Arg Asn Val Arg Leu Glu Lys Ile Ile Lys Leu

1175 1180 1185

Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Lys Arg Arg Arg Leu

1190 1195 1200

Leu Ala Ser Ala Lys Glu Leu Gln Lys Gly Asn Glu Leu Val Ile

1205 1210 1215

Pro Gln Arg Phe Thr Thr Leu Leu Tyr His Ser Tyr Arg Ile Glu

1220 1225 1230

Lys Asp Tyr Glu Pro Glu His Arg Glu Tyr Val Glu Lys His Lys

1235 1240 1245

Asp Glu Phe Lys Glu Leu Leu Glu Tyr Ile Ser Val Phe Ser Arg
1250 1255 1260
Lys Tyr Val Leu Ala Asp Asn Asn Leu Thr Lys Ile Glu Met Leu
1265 1270 1275
Phe Ser Lys Asn Lys Asp Ala Glu Val Ser Ser Leu Ala Lys Ser
1280 1285 1290
Phe Ile Ser Leu Leu Thr Phe Thr Ala Phe Gly Ala Pro Ala Ala
1295 1300 1305

Phe Asn Phe Phe Gly Glu Asn Ile Asp Arg Lys Arg Tyr Thr Ser
1310 1315 1320
Val Thr Glu Cys Leu Asn Ala Thr Leu Ile His Gln Ser Ile Thr
1325 1330 1335
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Lys Leu Gly Glu Asp
1340 1345 1350
Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys
1355 1360 1365

Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro
1370 1375 1380
Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr
1385 1390 1395
Ala

<210> 3
<211> 92
<212> RNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 3
guuuuagagc ugugcuguuu aaacaacaca gcaaguuaaa auaaggcuuu guccguacuc 60
aagcuugcaa aagcgugcac cgauucggug cu 92

<210> 4

<211> 1326

<212> PRT

<213> Sharpea sp.

<400> 4

Met Ala Lys Asn Lys Asp Ile Arg Tyr Ser Ile Gly Leu Asp Ile Gly
1 5 10 15

Thr Asn Ser Val Gly Trp Ala Val Met Asp Glu His Tyr Glu Leu Leu
20 25 30

Lys Lys Gly Asn His His Met Trp Gly Ser Arg Leu Phe Asp Ala Ala
35 40 45

Glu Pro Ala Ala Thr Arg Arg Ala Ser Arg Ser Ile Arg Arg Arg Tyr
50 55 60

Asn Lys Arg Arg Glu Arg Ile Arg Leu Leu Arg Asp Leu Leu Gly Asp
65 70 75 80

Met Val Met Glu Val Asp Pro Thr Phe Phe Ile Arg Leu Leu Asn Val
85 90 95

Ser Phe Leu Asp Glu Glu Asp Lys Gln Lys Asn Leu Gly Asn Asp Tyr
100 105 110

Lys Asp Asn Tyr Asn Leu Phe Ile Glu Lys Asp Phe Asn Asp Lys Thr
115 120 125

Tyr Tyr Asp Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Glu Leu Cys
130 135 140

Glu Asn Lys Glu Lys Ala Asp Pro Arg Leu Ile Tyr Leu Ala Leu His
145 150 155 160

His Ile Val Lys Tyr Arg Gly Asn Phe Leu Lys Glu Gly Gln Ser Phe
165 170 175

Ala Lys Val Tyr Glu Asp Ile Glu Glu Lys Leu Asp Asn Thr Leu Lys
180 185 190

Lys Phe Met Ser Leu Asn Asp Leu Asp Asn Leu Phe Val Asp Asn Asp
195 200 205

Ile Asn Ser Met Ile Thr Val Leu Ser Lys Ile Tyr Gln Arg Ser Lys
210 215 220

Lys Ala Asp Asp Leu Leu Lys Ile Met Asn Pro Thr Lys Glu Glu Arg
225 230 235 240
Ala Ala Tyr Lys Glu Phe Thr Lys Ala Leu Val Gly Leu Lys Phe Asn

245 250 255
Val Ser Lys Met Ile Leu Ala Gln Glu Val Lys Lys Asp Asp Lys Asp
260 265 270
Ile Glu Leu Asp Phe Ser Asn Val Asp Tyr Asp Ser Thr Val Asp Gly
275 280 285
Leu Gln Ala Glu Leu Gly Glu Tyr Ile Glu Phe Ile Glu Met Leu His
290 295 300
Ser Ile Asn Ser Trp Val Glu Leu Gln Asp Ile Leu Gly Asn Asn Ser

305 310 315 320
Thr Ile Ser Ala Ala Met Val Glu Arg Tyr Glu Glu His Lys Asn Asp
325 330 335
Leu Arg Val Leu Lys Lys Val Ile Arg Glu Glu Leu Pro Asp Lys Tyr
340 345 350
Asn Glu Val Phe Arg Glu Asp Asn Pro Lys Leu His Asn Tyr Leu Gly
355 360 365
Tyr Ile Lys Tyr Pro Lys Asn Thr Pro Val Glu Glu Phe Tyr Glu Tyr

370 375 380
Ile Lys Arg Leu Leu Ala Lys Val Asp Thr Gly Glu Ala Arg Glu Ile
385 390 395 400
Leu Glu Arg Ile Asp Leu Glu Lys Phe Met Leu Lys Gln Asn Ser Arg
405 410 415
Thr Asn Gly Ser Ile Pro Tyr Gln Met Gln Lys Asp Glu Met Ile Gln
420 425 430
Ile Ile Asp Asn Gln Ser Val Tyr Tyr Pro Gln Leu Lys Glu Asn Arg

435 440 445
Glu Lys Leu Ile Ser Ile Leu Glu Phe Arg Ile Pro Tyr Tyr Phe Gly
450 455 460
Pro Leu Asn Thr His Ser Glu Phe Ala Trp Ile Lys Lys Phe Glu Asp

465 470 475 480

Lys Gln Lys Glu Arg Ile Leu Pro Trp Asn Tyr Asp Gln Ile Val Asp

485 490 495

Ile Asp Ala Thr Ala Glu Gly Phe Ile Glu Arg Met Gln Asn Thr Gly

500 505 510

Thr Tyr Phe Pro Asp Lys Pro Val Met Ala Lys Asn Ser Leu Thr Val

515 520 525

Ser Lys Phe Glu Val Leu Asn Glu Leu Asn Lys Ile Arg Ile Asn Gly

530 535 540

Lys Leu Ile Pro Val Glu Thr Lys Lys Glu Leu Leu Ser Asp Leu Phe

545 550 555 560

Met Lys Asn Lys Thr Ile Thr Asp Lys Lys Leu Lys Asp Trp Leu Val

565 570 575

Thr His Gln Tyr Tyr Asp Thr Asn Glu Glu Leu Lys Ile Glu Gly Tyr

580 585 590

Gln Lys Asp Leu Gln Phe Ser Thr Ser Leu Ala Pro Trp Ile Asp Phe

595 600 605

Thr Lys Ile Phe Gly Glu Ile Asn Ala Ser Asn Tyr Gln Leu Ile Glu

610 615 620

Lys Ile Ile Tyr Asp Ile Ser Ile Phe Glu Asp Lys Lys Ile Leu Lys

625 630 635 640

Arg Arg Leu Lys Lys Val Tyr Gln Leu Asp Asp Leu Leu Val Asp Lys

645 650 655

Ile Leu Lys Leu Asn Tyr Thr Gly Trp Ser Arg Leu Ser Glu Lys Leu

660 665 670

Leu Thr Gly Ile Lys Ser Lys Asn Ser Lys Glu Thr Ile Leu Ser Ile

675 680 685

Leu Glu Asn Ser Asn Met Asn Leu Met Glu Ile Ile Asn Asp Glu Ser

690 695 700

Leu Gly Phe Lys Gln Ile Ile Glu Glu Ser Asn Lys Lys Asp Ile Glu

705 710 715 720

Gly Pro Phe Arg Tyr Asp Glu Val Lys Lys Leu Ala Gly Ser Pro Ala
725 730 735
Ile Lys Arg Gly Ile Trp Gln Ala Leu Leu Val Val Gln Glu Ile Thr
740 745 750
Lys Phe Met Lys His Glu Pro Ser His Ile Tyr Ile Glu Phe Ala Arg

755 760 765
Glu Glu Gln Glu Lys Val Arg Thr Glu Ser Arg Ile Ala Lys Leu Gln
770 775 780
Lys Ile Tyr Lys Asp Leu Asn Leu Gln Thr Lys Glu Asp Gln Leu Val
785 790 795 800
Tyr Glu Ser Leu Lys Lys Glu Asp Ala Lys Lys Lys Ile Asp Thr Asp
805 810 815
Ala Leu Tyr Leu Tyr Tyr Leu Gln Met Gly Lys Ser Met Tyr Ser Gly

820 825 830
Lys Pro Leu Asp Ile Asp Lys Leu Ser Thr Tyr His Ile Asp His Ile
835 840 845
Leu Pro Arg Ser Leu Ile Lys Asp Asp Ser Leu Asp Asn Arg Val Leu
850 855 860
Val Leu Pro Lys Glu Asn Glu Trp Lys Leu Asp Ser Glu Thr Val Pro
865 870 875 880
Phe Glu Ile Arg Asn Lys Met Met Gly Phe Trp Gln Lys Leu His Glu

885 890 895
Asn Gly Leu Met Ser Asn Lys Lys Phe Phe Ser Leu Ile Arg Thr Asp
900 905 910
Phe Asn Glu Lys Asp Lys Lys Arg Phe Ile Asn Arg Gln Leu Val Glu
915 920 925
Thr Arg Gln Ile Ile Lys Asn Val Ala Val Ile Ile Asn Asp His Tyr
930 935 940
Thr Asn Thr Asn Val Val Thr Val Arg Ala Glu Leu Ser His Gln Phe

945 950 955 960
Arg Glu Arg Tyr Lys Ile Tyr Lys Asn Arg Asp Leu Asn Asp Leu His

965 970 975

His Ala His Asp Ala Tyr Ile Ala Cys Ile Leu Gly Gln Phe Ile His

980 985 990

Gln Asn Phe Gly Asn Met Asp Val Asn Met Ile Tyr Gly Gln Tyr Lys

995 1000 1005

Lys Asn Tyr Lys Lys Asp Val Gln Glu His Asn Asn Tyr Gly Phe

1010 1015 1020

Ile Leu Asn Ser Met Asn His Ile His Phe Asn Asp Asp Asn Ser

1025 1030 1035

Val Ile Trp Asp Pro Ser Tyr Ile Gly Lys Ile Lys Ser Cys Phe

1040 1045 1050

Cys Tyr Lys Asp Val Tyr Val Thr Lys Lys Leu Glu Gln Asn Asp

1055 1060 1065

Ala Lys Leu Phe Asp Leu Thr Ile Leu Pro Ser Asp Lys Asn Ser

1070 1075 1080

Glu Asn Gly Val Thr Lys Ala Lys Ile Pro Val Asn Lys Tyr Arg

1085 1090 1095

Lys Asp Val Asn Lys Tyr Gly Gly Phe Ser Gly Asp Ala Pro Ile

1100 1105 1110

Met Leu Ala Ile Glu Ala Asp Lys Gly Lys Lys His Val Arg Gln

1115 1120 1125

Val Ile Ala Phe Pro Leu Arg Leu Lys Asn Tyr Asn Asp Glu Glu

1130 1135 1140

Arg Ile Lys Phe Ile Glu Lys Glu Lys Asn Leu Lys Asn Val Lys

1145 1150 1155

Ile Leu Thr Glu Val Lys Lys Asn Gln Leu Ile Leu Ile Asn His

1160 1165 1170

Gln Tyr Phe Phe Ile Thr Gly Thr Asn Glu Leu Val Asn Ala Thr

1175 1180 1185

Gln Leu Lys Leu Ser Ala Lys Asn Thr Lys Asn Leu Phe Asn Leu

1190 1195 1200

Val Asp Ala Asn Lys His Asn Lys Leu Glu Ser Ile Asp Asp Ala
1205 1210 1215
Asn Phe Asn Glu Val Ile Gln Glu Leu Ile Cys Lys Leu Gln Glu
1220 1225 1230
Pro Ile Tyr Ser Arg Tyr Asn Ser Ile Gly Lys Glu Phe Glu Asp
1235 1240 1245
Ser Tyr Glu Lys Ile Asn Ala Val Thr Lys Gln Asp Lys Leu Tyr

1250 1255 1260
Ile Ile Glu Tyr Leu Ile Ala Ile Met Ser Ala Lys Ala Thr Gln
1265 1270 1275
Gly Tyr Ile Lys Pro Glu Leu Ala Arg Glu Ile Gly Thr Asn Gly
1280 1285 1290
Lys Asn Lys Gly Arg Ile Lys Ser Phe Thr Ile Asp Leu Asn Lys
1295 1300 1305
Thr Thr Phe Ile Ser Thr Ser Val Thr Gly Leu Phe Ser Lys Lys

1310 1315 1320
Tyr Lys Leu
1325
<210> 5
<211> 1380
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 5
Met Pro Lys Lys Lys Arg Lys Val Gly Ala Lys Asn Lys Asp Ile Arg
1 5 10 15
Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val
20 25 30

Met Asp Glu His Tyr Glu Leu Leu Lys Lys Gly Asn His His Met Trp
35 40 45
Gly Ser Arg Leu Phe Asp Ala Ala Glu Pro Ala Ala Thr Arg Arg Ala

50 55 60

Ser Arg Ser Ile Arg Arg Arg Tyr Asn Lys Arg Arg Glu Arg Ile Arg
65 70 75 80

Leu Leu Arg Asp Leu Leu Gly Asp Met Val Met Glu Val Asp Pro Thr
85 90 95

Phe Phe Ile Arg Leu Leu Asn Val Ser Phe Leu Asp Glu Glu Asp Lys
100 105 110

Gln Lys Asn Leu Gly Asn Asp Tyr Lys Asp Asn Tyr Asn Leu Phe Ile
115 120 125

Glu Lys Asp Phe Asn Asp Lys Thr Tyr Tyr Asp Lys Tyr Pro Thr Ile
130 135 140

Tyr His Leu Arg Lys Glu Leu Cys Glu Asn Lys Glu Lys Ala Asp Pro
145 150 155 160

Arg Leu Ile Tyr Leu Ala Leu His His Ile Val Lys Tyr Arg Gly Asn
165 170 175

Phe Leu Lys Glu Gly Gln Ser Phe Ala Lys Val Tyr Glu Asp Ile Glu
180 185 190

Glu Lys Leu Asp Asn Thr Leu Lys Lys Phe Met Ser Leu Asn Asp Leu
195 200 205

Asp Asn Leu Phe Val Asp Asn Asp Ile Asn Ser Met Ile Thr Val Leu
210 215 220

Ser Lys Ile Tyr Gln Arg Ser Lys Lys Ala Asp Asp Leu Leu Lys Ile
225 230 235 240

Met Asn Pro Thr Lys Glu Glu Arg Ala Ala Tyr Lys Glu Phe Thr Lys
245 250 255

Ala Leu Val Gly Leu Lys Phe Asn Val Ser Lys Met Ile Leu Ala Gln
260 265 270

Glu Val Lys Lys Asp Asp Lys Asp Ile Glu Leu Asp Phe Ser Asn Val
275 280 285

Asp Tyr Asp Ser Thr Val Asp Gly Leu Gln Ala Glu Leu Gly Glu Tyr
290 295 300

Ile Glu Phe Ile Glu Met Leu His Ser Ile Asn Ser Trp Val Glu Leu
305 310 315 320
Gln Asp Ile Leu Gly Asn Asn Ser Thr Ile Ser Ala Ala Met Val Glu
325 330 335
Arg Tyr Glu Glu His Lys Asn Asp Leu Arg Val Leu Lys Lys Val Ile
340 345 350

Arg Glu Glu Leu Pro Asp Lys Tyr Asn Glu Val Phe Arg Glu Asp Asn
355 360 365
Pro Lys Leu His Asn Tyr Leu Gly Tyr Ile Lys Tyr Pro Lys Asn Thr
370 375 380
Pro Val Glu Glu Phe Tyr Glu Tyr Ile Lys Arg Leu Leu Ala Lys Val
385 390 395 400
Asp Thr Gly Glu Ala Arg Glu Ile Leu Glu Arg Ile Asp Leu Glu Lys
405 410 415

Phe Met Leu Lys Gln Asn Ser Arg Thr Asn Gly Ser Ile Pro Tyr Gln
420 425 430
Met Gln Lys Asp Glu Met Ile Gln Ile Ile Asp Asn Gln Ser Val Tyr
435 440 445
Tyr Pro Gln Leu Lys Glu Asn Arg Glu Lys Leu Ile Ser Ile Leu Glu
450 455 460
Phe Arg Ile Pro Tyr Tyr Phe Gly Pro Leu Asn Thr His Ser Glu Phe
465 470 475 480

Ala Trp Ile Lys Lys Phe Glu Asp Lys Gln Lys Glu Arg Ile Leu Pro
485 490 495
Trp Asn Tyr Asp Gln Ile Val Asp Ile Asp Ala Thr Ala Glu Gly Phe
500 505 510
Ile Glu Arg Met Gln Asn Thr Gly Thr Tyr Phe Pro Asp Lys Pro Val
515 520 525
Met Ala Lys Asn Ser Leu Thr Val Ser Lys Phe Glu Val Leu Asn Glu
530 535 540

Leu Asn Lys Ile Arg Ile Asn Gly Lys Leu Ile Pro Val Glu Thr Lys

545 550 555 560

Lys Glu Leu Leu Ser Asp Leu Phe Met Lys Asn Lys Thr Ile Thr Asp

565 570 575

Lys Lys Leu Lys Asp Trp Leu Val Thr His Gln Tyr Tyr Asp Thr Asn

580 585 590

Glu Glu Leu Lys Ile Glu Gly Tyr Gln Lys Asp Leu Gln Phe Ser Thr

595 600 605

Ser Leu Ala Pro Trp Ile Asp Phe Thr Lys Ile Phe Gly Glu Ile Asn

610 615 620

Ala Ser Asn Tyr Gln Leu Ile Glu Lys Ile Ile Tyr Asp Ile Ser Ile

625 630 635 640

Phe Glu Asp Lys Lys Ile Leu Lys Arg Arg Leu Lys Lys Val Tyr Gln

645 650 655

Leu Asp Asp Leu Leu Val Asp Lys Ile Leu Lys Leu Asn Tyr Thr Gly

660 665 670

Trp Ser Arg Leu Ser Glu Lys Leu Leu Thr Gly Ile Lys Ser Lys Asn

675 680 685

Ser Lys Glu Thr Ile Leu Ser Ile Leu Glu Asn Ser Asn Met Asn Leu

690 695 700

Met Glu Ile Ile Asn Asp Glu Ser Leu Gly Phe Lys Gln Ile Ile Glu

705 710 715 720

Glu Ser Asn Lys Lys Asp Ile Glu Gly Pro Phe Arg Tyr Asp Glu Val

725 730 735

Lys Lys Leu Ala Gly Ser Pro Ala Ile Lys Arg Gly Ile Trp Gln Ala

740 745 750

Leu Leu Val Val Gln Glu Ile Thr Lys Phe Met Lys His Glu Pro Ser

755 760 765

His Ile Tyr Ile Glu Phe Ala Arg Glu Glu Gln Glu Lys Val Arg Thr

770 775 780

Glu Ser Arg Ile Ala Lys Leu Gln Lys Ile Tyr Lys Asp Leu Asn Leu

785 790 795 800

Gln Thr Lys Glu Asp Gln Leu Val Tyr Glu Ser Leu Lys Lys Glu Asp
805 810 815
Ala Lys Lys Lys Ile Asp Thr Asp Ala Leu Tyr Leu Tyr Tyr Leu Gln
820 825 830
Met Gly Lys Ser Met Tyr Ser Gly Lys Pro Leu Asp Ile Asp Lys Leu
835 840 845
Ser Thr Tyr His Ile Asp His Ile Leu Pro Arg Ser Leu Ile Lys Asp
850 855 860

Asp Ser Leu Asp Asn Arg Val Leu Val Leu Pro Lys Glu Asn Glu Trp
865 870 875 880
Lys Leu Asp Ser Glu Thr Val Pro Phe Glu Ile Arg Asn Lys Met Met
885 890 895
Gly Phe Trp Gln Lys Leu His Glu Asn Gly Leu Met Ser Asn Lys Lys
900 905 910
Phe Phe Ser Leu Ile Arg Thr Asp Phe Asn Glu Lys Asp Lys Lys Arg
915 920 925

Phe Ile Asn Arg Gln Leu Val Glu Thr Arg Gln Ile Ile Lys Asn Val
930 935 940
Ala Val Ile Ile Asn Asp His Tyr Thr Asn Thr Asn Val Val Thr Val
945 950 955 960
Arg Ala Glu Leu Ser His Gln Phe Arg Glu Arg Tyr Lys Ile Tyr Lys
965 970 975
Asn Arg Asp Leu Asn Asp Leu His His Ala His Asp Ala Tyr Ile Ala
980 985 990

Cys Ile Leu Gly Gln Phe Ile His Gln Asn Phe Gly Asn Met Asp Val
995 1000 1005
Asn Met Ile Tyr Gly Gln Tyr Lys Lys Asn Tyr Lys Lys Asp Val
1010 1015 1020
Gln Glu His Asn Asn Tyr Gly Phe Ile Leu Asn Ser Met Asn His
1025 1030 1035
Ile His Phe Asn Asp Asp Asn Ser Val Ile Trp Asp Pro Ser Tyr

1040 1045 1050

Ile Gly Lys Ile Lys Ser Cys Phe Cys Tyr Lys Asp Val Tyr Val
1055 1060 1065
Thr Lys Lys Leu Glu Gln Asn Asp Ala Lys Leu Phe Asp Leu Thr
1070 1075 1080
Ile Leu Pro Ser Asp Lys Asn Ser Glu Asn Gly Val Thr Lys Ala
1085 1090 1095
Lys Ile Pro Val Asn Lys Tyr Arg Lys Asp Val Asn Lys Tyr Gly
1100 1105 1110

Gly Phe Ser Gly Asp Ala Pro Ile Met Leu Ala Ile Glu Ala Asp
1115 1120 1125
Lys Gly Lys Lys His Val Arg Gln Val Ile Ala Phe Pro Leu Arg
1130 1135 1140
Leu Lys Asn Tyr Asn Asp Glu Glu Arg Ile Lys Phe Ile Glu Lys
1145 1150 1155
Glu Lys Asn Leu Lys Asn Val Lys Ile Leu Thr Glu Val Lys Lys
1160 1165 1170

Asn Gln Leu Ile Leu Ile Asn His Gln Tyr Phe Phe Ile Thr Gly
1175 1180 1185
Thr Asn Glu Leu Val Asn Ala Thr Gln Leu Lys Leu Ser Ala Lys
1190 1195 1200
Asn Thr Lys Asn Leu Phe Asn Leu Val Asp Ala Asn Lys His Asn
1205 1210 1215
Lys Leu Glu Ser Ile Asp Asp Ala Asn Phe Asn Glu Val Ile Gln
1220 1225 1230

Glu Leu Ile Cys Lys Leu Gln Glu Pro Ile Tyr Ser Arg Tyr Asn
1235 1240 1245
Ser Ile Gly Lys Glu Phe Glu Asp Ser Tyr Glu Lys Ile Asn Ala
1250 1255 1260
Val Thr Lys Gln Asp Lys Leu Tyr Ile Ile Glu Tyr Leu Ile Ala
1265 1270 1275

Ile Met Ser Ala Lys Ala Thr Gln Gly Tyr Ile Lys Pro Glu Leu
1280 1285 1290

Ala Arg Glu Ile Gly Thr Asn Gly Lys Asn Lys Gly Arg Ile Lys
1295 1300 1305
Ser Phe Thr Ile Asp Leu Asn Lys Thr Thr Phe Ile Ser Thr Ser
1310 1315 1320
Val Thr Gly Leu Phe Ser Lys Lys Tyr Lys Leu Gly Lys Arg Pro
1325 1330 1335
Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser
1340 1345 1350

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro
1355 1360 1365
Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1370 1375 1380

<210> 6
<211> 1576
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 6
Met Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Ser Glu Val Glu Phe
1 5 10 15

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala
20 25 30
Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn
35 40 45
Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro
50 55 60
Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met
65 70 75 80

Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro
85 90 95
Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
100 105 110
Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
115 120 125
Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
130 135 140

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg
145 150 155 160
Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr
165 170 175
Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr
180 185 190
Pro Glu Ser Ser Gly Pro Lys Lys Lys Arg Lys Val Gly Ala Lys Asn
195 200 205

Lys Asp Ile Arg Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
210 215 220
Gly Trp Ala Val Met Asp Glu His Tyr Glu Leu Leu Lys Lys Gly Asn
225 230 235 240
His His Met Trp Gly Ser Arg Leu Phe Asp Ala Ala Glu Pro Ala Ala
245 250 255
Thr Arg Arg Ala Ser Arg Ser Ile Arg Arg Arg Tyr Asn Lys Arg Arg
260 265 270

Glu Arg Ile Arg Leu Leu Arg Asp Leu Leu Gly Asp Met Val Met Glu
275 280 285
Val Asp Pro Thr Phe Phe Ile Arg Leu Leu Asn Val Ser Phe Leu Asp
290 295 300
Glu Glu Asp Lys Gln Lys Asn Leu Gly Asn Asp Tyr Lys Asp Asn Tyr
305 310 315 320
Asn Leu Phe Ile Glu Lys Asp Phe Asn Asp Lys Thr Tyr Tyr Asp Lys

325 330 335

Tyr Pro Thr Ile Tyr His Leu Arg Lys Glu Leu Cys Glu Asn Lys Glu
340 345 350
Lys Ala Asp Pro Arg Leu Ile Tyr Leu Ala Leu His His Ile Val Lys
355 360 365
Tyr Arg Gly Asn Phe Leu Lys Glu Gly Gln Ser Phe Ala Lys Val Tyr
370 375 380
Glu Asp Ile Glu Glu Lys Leu Asp Asn Thr Leu Lys Lys Phe Met Ser
385 390 395 400

Leu Asn Asp Leu Asp Asn Leu Phe Val Asp Asn Asp Ile Asn Ser Met
405 410 415
Ile Thr Val Leu Ser Lys Ile Tyr Gln Arg Ser Lys Lys Ala Asp Asp
420 425 430
Leu Leu Lys Ile Met Asn Pro Thr Lys Glu Glu Arg Ala Ala Tyr Lys
435 440 445
Glu Phe Thr Lys Ala Leu Val Gly Leu Lys Phe Asn Val Ser Lys Met
450 455 460

Ile Leu Ala Gln Glu Val Lys Lys Asp Asp Lys Asp Ile Glu Leu Asp
465 470 475 480
Phe Ser Asn Val Asp Tyr Asp Ser Thr Val Asp Gly Leu Gln Ala Glu
485 490 495
Leu Gly Glu Tyr Ile Glu Phe Ile Glu Met Leu His Ser Ile Asn Ser
500 505 510
Trp Val Glu Leu Gln Asp Ile Leu Gly Asn Asn Ser Thr Ile Ser Ala
515 520 525

Ala Met Val Glu Arg Tyr Glu Glu His Lys Asn Asp Leu Arg Val Leu
530 535 540
Lys Lys Val Ile Arg Glu Glu Leu Pro Asp Lys Tyr Asn Glu Val Phe
545 550 555 560
Arg Glu Asp Asn Pro Lys Leu His Asn Tyr Leu Gly Tyr Ile Lys Tyr
565 570 575

Pro Lys Asn Thr Pro Val Glu Glu Phe Tyr Glu Tyr Ile Lys Arg Leu
580 585 590

Leu Ala Lys Val Asp Thr Gly Glu Ala Arg Glu Ile Leu Glu Arg Ile
595 600 605

Asp Leu Glu Lys Phe Met Leu Lys Gln Asn Ser Arg Thr Asn Gly Ser
610 615 620

Ile Pro Tyr Gln Met Gln Lys Asp Glu Met Ile Gln Ile Ile Asp Asn
625 630 635 640

Gln Ser Val Tyr Tyr Pro Gln Leu Lys Glu Asn Arg Glu Lys Leu Ile
645 650 655

Ser Ile Leu Glu Phe Arg Ile Pro Tyr Tyr Phe Gly Pro Leu Asn Thr
660 665 670

His Ser Glu Phe Ala Trp Ile Lys Lys Phe Glu Asp Lys Gln Lys Glu
675 680 685

Arg Ile Leu Pro Trp Asn Tyr Asp Gln Ile Val Asp Ile Asp Ala Thr
690 695 700

Ala Glu Gly Phe Ile Glu Arg Met Gln Asn Thr Gly Thr Tyr Phe Pro
705 710 715 720

Asp Lys Pro Val Met Ala Lys Asn Ser Leu Thr Val Ser Lys Phe Glu
725 730 735

Val Leu Asn Glu Leu Asn Lys Ile Arg Ile Asn Gly Lys Leu Ile Pro
740 745 750

Val Glu Thr Lys Lys Glu Leu Leu Ser Asp Leu Phe Met Lys Asn Lys
755 760 765

Thr Ile Thr Asp Lys Lys Leu Lys Asp Trp Leu Val Thr His Gln Tyr
770 775 780

Tyr Asp Thr Asn Glu Glu Leu Lys Ile Glu Gly Tyr Gln Lys Asp Leu
785 790 795 800

Gln Phe Ser Thr Ser Leu Ala Pro Trp Ile Asp Phe Thr Lys Ile Phe
805 810 815

Gly Glu Ile Asn Ala Ser Asn Tyr Gln Leu Ile Glu Lys Ile Ile Tyr

820 825 830

Asp Ile Ser Ile Phe Glu Asp Lys Lys Ile Leu Lys Arg Arg Leu Lys

835 840 845

Lys Val Tyr Gln Leu Asp Asp Leu Leu Val Asp Lys Ile Leu Lys Leu

850 855 860

Asn Tyr Thr Gly Trp Ser Arg Leu Ser Glu Lys Leu Leu Thr Gly Ile

865 870 875 880

Lys Ser Lys Asn Ser Lys Glu Thr Ile Leu Ser Ile Leu Glu Asn Ser

885 890 895

Asn Met Asn Leu Met Glu Ile Ile Asn Asp Glu Ser Leu Gly Phe Lys

900 905 910

Gln Ile Ile Glu Glu Ser Asn Lys Lys Asp Ile Glu Gly Pro Phe Arg

915 920 925

Tyr Asp Glu Val Lys Lys Leu Ala Gly Ser Pro Ala Ile Lys Arg Gly

930 935 940

Ile Trp Gln Ala Leu Leu Val Val Gln Glu Ile Thr Lys Phe Met Lys

945 950 955 960

His Glu Pro Ser His Ile Tyr Ile Glu Phe Ala Arg Glu Glu Gln Glu

965 970 975

Lys Val Arg Thr Glu Ser Arg Ile Ala Lys Leu Gln Lys Ile Tyr Lys

980 985 990

Asp Leu Asn Leu Gln Thr Lys Glu Asp Gln Leu Val Tyr Glu Ser Leu

995 1000 1005

Lys Lys Glu Asp Ala Lys Lys Lys Ile Asp Thr Asp Ala Leu Tyr

1010 1015 1020

Leu Tyr Tyr Leu Gln Met Gly Lys Ser Met Tyr Ser Gly Lys Pro

1025 1030 1035

Leu Asp Ile Asp Lys Leu Ser Thr Tyr His Ile Asp His Ile Leu

1040 1045 1050

Pro Arg Ser Leu Ile Lys Asp Asp Ser Leu Asp Asn Arg Val Leu

1055 1060 1065

Val Leu Pro Lys Glu Asn Glu Trp Lys Leu Asp Ser Glu Thr Val
1070 1075 1080
Pro Phe Glu Ile Arg Asn Lys Met Met Gly Phe Trp Gln Lys Leu
1085 1090 1095

His Glu Asn Gly Leu Met Ser Asn Lys Lys Phe Phe Ser Leu Ile
1100 1105 1110
Arg Thr Asp Phe Asn Glu Lys Asp Lys Lys Arg Phe Ile Asn Arg
1115 1120 1125
Gln Leu Val Glu Thr Arg Gln Ile Ile Lys Asn Val Ala Val Ile
1130 1135 1140
Ile Asn Asp His Tyr Thr Asn Thr Asn Val Val Thr Val Arg Ala
1145 1150 1155

Glu Leu Ser His Gln Phe Arg Glu Arg Tyr Lys Ile Tyr Lys Asn
1160 1165 1170
Arg Asp Leu Asn Asp Leu His His Ala His Asp Ala Tyr Ile Ala
1175 1180 1185
Cys Ile Leu Gly Gln Phe Ile His Gln Asn Phe Gly Asn Met Asp
1190 1195 1200
Val Asn Met Ile Tyr Gly Gln Tyr Lys Lys Asn Tyr Lys Lys Asp
1205 1210 1215

Val Gln Glu His Asn Asn Tyr Gly Phe Ile Leu Asn Ser Met Asn
1220 1225 1230
His Ile His Phe Asn Asp Asp Asn Ser Val Ile Trp Asp Pro Ser
1235 1240 1245
Tyr Ile Gly Lys Ile Lys Ser Cys Phe Cys Tyr Lys Asp Val Tyr
1250 1255 1260
Val Thr Lys Lys Leu Glu Gln Asn Asp Ala Lys Leu Phe Asp Leu
1265 1270 1275

Thr Ile Leu Pro Ser Asp Lys Asn Ser Glu Asn Gly Val Thr Lys
1280 1285 1290
Ala Lys Ile Pro Val Asn Lys Tyr Arg Lys Asp Val Asn Lys Tyr

1295 1300 1305

Gly Gly Phe Ser Gly Asp Ala Pro Ile Met Leu Ala Ile Glu Ala

1310 1315 1320

Asp Lys Gly Lys Lys His Val Arg Gln Val Ile Ala Phe Pro Leu

1325 1330 1335

Arg Leu Lys Asn Tyr Asn Asp Glu Glu Arg Ile Lys Phe Ile Glu

1340 1345 1350

Lys Glu Lys Asn Leu Lys Asn Val Lys Ile Leu Thr Glu Val Lys

1355 1360 1365

Lys Asn Gln Leu Ile Leu Ile Asn His Gln Tyr Phe Phe Ile Thr

1370 1375 1380

Gly Thr Asn Glu Leu Val Asn Ala Thr Gln Leu Lys Leu Ser Ala

1385 1390 1395

Lys Asn Thr Lys Asn Leu Phe Asn Leu Val Asp Ala Asn Lys His

1400 1405 1410

Asn Lys Leu Glu Ser Ile Asp Asp Ala Asn Phe Asn Glu Val Ile

1415 1420 1425

Gln Glu Leu Ile Cys Lys Leu Gln Glu Pro Ile Tyr Ser Arg Tyr

1430 1435 1440

Asn Ser Ile Gly Lys Glu Phe Glu Asp Ser Tyr Glu Lys Ile Asn

1445 1450 1455

Ala Val Thr Lys Gln Asp Lys Leu Tyr Ile Ile Glu Tyr Leu Ile

1460 1465 1470

Ala Ile Met Ser Ala Lys Ala Thr Gln Gly Tyr Ile Lys Pro Glu

1475 1480 1485

Leu Ala Arg Glu Ile Gly Thr Asn Gly Lys Asn Lys Gly Arg Ile

1490 1495 1500

Lys Ser Phe Thr Ile Asp Leu Asn Lys Thr Thr Phe Ile Ser Thr

1505 1510 1515

Ser Val Thr Gly Leu Phe Ser Lys Lys Tyr Lys Leu Gly Lys Arg

1520 1525 1530

Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly
1535 1540 1545
Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val
1550 1555 1560
Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1565 1570 1575

<210>
7
<211> 119
<212> RNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polynucleotide
<400> 7
guuuuagagu uguguuauug aaaaauaaca caacgaguua aaauaaagcu uaugcuuaaa 60
ugccagcuuu gcugguguca uuuagaugac uuuacuaagg uugcuucggc aaccuuuuu 119

<210> 8
<211> 1409
<212> PRT
<213> Veillonella parvula
<400> 8
Met Ser Ile Ile Asn Phe Gln Arg Arg Gly Leu Met Glu Thr Gln Ala
1 5 10 15
Ser Asn Gln Leu Ile Ser Ser His Leu Lys Gly Tyr Pro Ile Lys Asp
20 25 30
Tyr Phe Val Gly Leu Asp Ile Gly Thr Ser Ser Val Gly Trp Ala Val
35 40 45
Thr Asn Lys Ala Tyr Glu Leu Leu Lys Phe Arg Ser His Lys Met Trp
50 55 60
Gly Ser Arg Leu Phe Asp Glu Gly Glu Ser Ala Val Ala Arg Arg Gly
65 70 75 80
Phe Arg Ser Met Arg Arg Arg Leu Glu Arg Arg Lys Leu Arg Leu Lys
85 90 95

Leu Leu Glu Glu Leu Phe Ala Asp Ala Met Ala Gln Val Asp Pro Thr
100 105 110
Phe Phe Met Arg Leu Arg Glu Ser Lys Tyr His Tyr Glu Asp Lys Thr
115 120 125
Thr Gly His Ser Ser Lys His Ile Leu Phe Ile Asp Lys Asn Tyr Asn
130 135 140
Asp Gln Asp Tyr Phe Lys Glu Tyr Pro Thr Val Tyr His Leu Arg Ser

145 150 155 160
Glu Leu Met Lys Ser Gly Thr Asp Asp Ile Arg Lys Leu Phe Leu Ala
165 170 175
Val His His Ile Leu Lys Tyr Arg Gly Asn Phe Leu Tyr Glu Gly Ala
180 185 190
Thr Phe Asp Ser Asn Ala Ser Thr Leu Asp Asp Val Ile Lys Gln Ala
195 200 205
Leu Glu Asn Ile Thr Phe Asn Cys Phe Asp Cys Asn Ser Ala Ile Ser

210 215 220
Ser Ile Gly Gln Ile Leu Met Glu Ala Gly Lys Thr Lys Ser Asp Lys
225 230 235 240
Ala Lys Ala Ile Glu His Leu Val Asp Thr Tyr Ile Ala Thr Asp Thr
245 250 255
Val Asp Thr Ser Ser Lys Thr Gln Lys Asp Gln Val Lys Glu Asp Lys
260 265 270
Lys Arg Leu Lys Ala Phe Ala Asn Leu Val Leu Gly Leu Asn Ala Ser

275 280 285
Leu Ile Asp Leu Phe Gly Ser Val Glu Glu Leu Glu Glu Asp Leu Lys
290 295 300
Lys Leu Gln Ile Thr Gly Asp Thr Tyr Asp Asp Lys Arg Asp Glu Leu
305 310 315 320
Ala Lys Ala Trp Ser Asp Glu Ile Tyr Ile Ile Asp Asp Cys Lys Ser
325 330 335
Val Tyr Asp Ala Ile Ile Leu Leu Ser Ile Lys Glu Pro Gly Leu Thr

340 345 350

Ile Ser Glu Ser Lys Val Lys Ala Phe Asn Lys His Lys Asp Asp Leu

355 360 365

Ala Ile Leu Lys Ser Leu Leu Lys Ser Asp Arg Ser Ile Tyr Asn Thr

370 375 380

Met Phe Lys Val Asp Glu Lys Gly Leu His Asn Tyr Val His Tyr Ile

385 390 395 400

Lys Gln Gly Arg Thr Glu Glu Thr Ser Cys Asn Arg Glu Asp Phe Tyr

405 410 415

Lys Tyr Thr Lys Lys Ile Val Glu Gly Leu Ser Asp Ser Lys Asp Lys

420 425 430

Glu Tyr Ile Leu Ser Gln Ile Glu Leu Gln Ile Leu Leu Pro Leu Gln

435 440 445

Arg Ile Lys Asp Asn Gly Val Ile Pro Tyr Gln Leu His Leu Glu Glu

450 455 460

Leu Lys Ala Ile Leu Ala Lys Cys Gly Pro Lys Phe Pro Phe Leu Asn

465 470 475 480

Glu Val Ala Asp Gly Phe Ser Val Ala Glu Lys Leu Ile Lys Met Leu

485 490 495

Glu Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Asn Thr His His Asn

500 505 510

Val Asp Asn Gly Gly Phe Ala Trp Ala Val Arg Lys Ala Ser Gly Arg

515 520 525

Val Thr Pro Trp Asn Phe Asp Asp Lys Ile Asp Arg Glu Lys Ser Ala

530 535 540

Ala Ala Phe Ile Lys Asn Leu Thr Asn Lys Cys Thr Tyr Leu Leu Gly

545 550 555 560

Glu Asp Val Leu Pro Lys Ser Ser Leu Leu Tyr Ser Glu Phe Met Leu

565 570 575

Leu Asn Glu Leu Asn Asn Val Arg Ile Asp Gly Lys Pro Leu Glu Lys

580 585 590

Val Val Lys Glu His Leu Ile Glu Ala Val Phe Lys Gln Asp His Lys

595 600 605

Lys Met Thr Lys Asn Arg Ile Glu Gln Phe Leu Lys Asp Asn Gly Tyr

610 615 620

Ile Ser Glu Thr His Lys His Glu Ile Thr Gly Leu Asp Gly Glu Ile

625 630 635 640

Lys Asn Asp Leu Ala Ser Tyr Arg Asp Met Val Arg Ile Leu Gly Asp

645 650 655

Gly Phe Asp Arg Ser Met Ala Glu Glu Ile Ile Thr Asp Ile Thr Ile

660 665 670

Phe Gly Glu Ser Lys Lys Met Leu Arg Glu Thr Leu Arg Lys Lys Phe

675 680 685

Ala Ser Cys Leu Asp Asp Glu Ala Ile Lys Lys Leu Thr Lys Leu Arg

690 695 700

Tyr Arg Asp Trp Gly Arg Leu Ser Gln Lys Leu Leu Asn Gly Ile Glu

705 710 715 720

Gly Cys Asp Lys Ala Gly Asp Gly Thr Pro Glu Thr Ile Ile Ile Leu

725 730 735

Met Arg Asn Phe Ser Tyr Asn Leu Met Glu Leu Leu Gly Asp Lys Phe

740 745 750

Ser Phe Met Glu Arg Ile Gln Glu Ile Asn Ala Lys Leu Thr Glu Gly

755 760 765

Gln Ile Val Asn Pro His Asp Ile Ile Asp Asp Leu Ala Leu Ser Pro

770 775 780

Ala Val Lys Arg Ala Val Trp Gln Ala Leu Arg Ile Val Asp Glu Val

785 790 795 800

Ala His Ile Lys Lys Ala Leu Pro Ala Arg Ile Phe Val Glu Val Thr

805 810 815

Arg Ser Asn Lys Asn Glu Lys Lys Lys Lys Asp Ser Arg Gln Lys Arg

820 825 830

Leu Ser Asp Leu Tyr Ala Ala Ile Lys Lys Asp Asp Val Leu Leu Asn

835 840 845

Gly Leu Asn Asn Glu Ile Phe Gly Glu Leu Lys Ser Ser Leu Ala Lys

850 855 860

Tyr Asp Asp Ala Ala Leu Arg Ser Lys Lys Leu Tyr Leu Tyr Tyr Thr

865 870 875 880

Gln Met Gly Arg Cys Ala Tyr Thr Gly Glu Ile Ile Glu Leu Ser Leu

885 890 895

Leu Asn Thr Asp Asn Tyr Asp Ile Asp His Ile Tyr Pro Arg Ser Leu

900 905 910

Thr Lys Asp Asp Ser Phe Asp Asn Leu Val Leu Cys Lys Arg Thr Ala

915 920 925

Asn Ala Gln Lys Ser Asp Ala Tyr Pro Ile Ser Glu Glu Ile Gln Lys

930 935 940

Thr Gln Lys Pro Phe Trp Thr Phe Leu Lys Gln Gln Gly Leu Ile Ser

945 950 955 960

Glu Arg Lys Tyr Glu Arg Leu Thr Arg Ile Thr Pro Leu Thr Ala Asp

965 970 975

Asp Leu Ser Gly Phe Ile Ala Arg Gln Leu Val Glu Thr Asn Gln Ser

980 985 990

Val Lys Ala Ala Thr Thr Leu Leu Arg Arg Leu Tyr Pro Gly Val Asp

995 1000 1005

Val Val Phe Val Lys Ala Glu Asn Val Thr Asp Phe Arg His Asp

1010 1015 1020

Asn Asn Phe Ile Lys Val Arg Ser Leu Asn His His His His Ala

1025 1030 1035

Lys Asp Ala Tyr Leu Asn Ile Val Val Gly Asn Val Tyr His Glu

1040 1045 1050

Arg Phe Thr Arg Asn Phe Arg Ala Phe Phe Lys Lys Asn Gly Ala

1055 1060 1065

Asn Arg Thr Tyr Asn Leu Ala Lys Met Phe Asn Tyr Asp Val Asn

1070 1075 1080

Cys Thr Asn Ala Lys Asp Gly Lys Ala Trp Asp Val Lys Thr Ser

1085 1090 1095

Met Asp Thr Val Lys Lys Met Met Asp Ser Asn Asp Val Arg Val

1100 1105 1110

Thr Lys Arg Leu Leu Glu Gln Thr Gly Ala Leu Ala Asp Ala Thr

1115 1120 1125

Ile Tyr Lys Ala Thr Val Ala Gly Lys Ala Lys Asp Gly Ala Tyr

1130 1135 1140

Ile Gly Met Lys Thr Lys Ser Ser Val Phe Ala Asp Val Ser Lys

1145 1150 1155

Tyr Gly Gly Met Thr Lys Ile Lys Asn Ala Tyr Ser Ile Ile Val

1160 1165 1170

Gln Tyr Thr Gly Lys Lys Gly Glu Val Ile Lys Glu Ile Val Pro

1175 1180 1185

Leu Pro Ile Tyr Leu Thr Asn Arg Asn Thr Thr Asp Gln Asp Leu

1190 1195 1200

Ile Asn Tyr Val Ala Ser Ile Ile Pro Gln Ala Lys Asp Ile Ser

1205 1210 1215

Ile Ile Tyr Gly Lys Leu Cys Ile Asn Gln Leu Val Lys Val Asn

1220 1225 1230

Gly Phe Tyr Tyr Tyr Leu Gly Gly Lys Thr Asn Ser Lys Phe Cys

1235 1240 1245

Ile Asp Asn Ala Ile Gln Val Ile Val Ser Asn Glu Trp Ile Pro

1250 1255 1260

Tyr Leu Lys Val Leu Glu Lys Phe Asn Asn Met Arg Lys Asp Asn

1265 1270 1275

Lys Asp Leu Lys Ala Asn Val Val Ser Thr Arg Ala Leu Asp Asn

1280 1285 1290

Lys His Thr Ile Glu Val Arg Ile Val Glu Glu Lys Asn Ile Glu

1295 1300 1305

Phe Phe Asp Tyr Leu Val Ser Lys Leu Lys Met Pro Ile Tyr Gln

1310 1315 1320

Lys Met Lys Gly Asn Lys Ala Ala Glu Leu Ser Glu Lys Gly Tyr

1325 1330 1335

Gly Leu Phe Lys Lys Met Ser Leu Glu Glu Gln Ser Ile His Leu

1340 1345 1350

Ile Glu Leu Leu Asn Leu Leu Thr Asn Gln Lys Thr Thr Phe Glu

1355 1360 1365

Val Lys Pro Leu Gly Ile Thr Ala Ser Arg Ser Thr Val Gly Ser

1370 1375 1380

Lys Ile Ser Asn Gln Asp Glu Phe Lys Val Ile Asn Glu Ser Ile

1385 1390 1395

Thr Gly Leu Tyr Ser Asn Glu Val Thr Ile Val

1400 1405

<210> 9

<211> 1463

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 9

Met Pro Lys Lys Lys Arg Lys Val Gly Ser Ile Ile Asn Phe Gln Arg

1 5 10 15

Arg Gly Leu Met Glu Thr Gln Ala Ser Asn Gln Leu Ile Ser Ser His

20 25 30

Leu Lys Gly Tyr Pro Ile Lys Asp Tyr Phe Val Gly Leu Asp Ile Gly

35 40 45

Thr Ser Ser Val Gly Trp Ala Val Thr Asn Lys Ala Tyr Glu Leu Leu

50 55 60

Lys Phe Arg Ser His Lys Met Trp Gly Ser Arg Leu Phe Asp Glu Gly

65 70 75 80

Glu Ser Ala Val Ala Arg Arg Gly Phe Arg Ser Met Arg Arg Arg Leu

85 90 95

Glu Arg Arg Lys Leu Arg Leu Lys Leu Leu Glu Glu Leu Phe Ala Asp
100 105 110

Ala Met Ala Gln Val Asp Pro Thr Phe Phe Met Arg Leu Arg Glu Ser
115 120 125
Lys Tyr His Tyr Glu Asp Lys Thr Thr Gly His Ser Ser Lys His Ile
130 135 140
Leu Phe Ile Asp Lys Asn Tyr Asn Asp Gln Asp Tyr Phe Lys Glu Tyr
145 150 155 160
Pro Thr Val Tyr His Leu Arg Ser Glu Leu Met Lys Ser Gly Thr Asp
165 170 175

Asp Ile Arg Lys Leu Phe Leu Ala Val His His Ile Leu Lys Tyr Arg
180 185 190
Gly Asn Phe Leu Tyr Glu Gly Ala Thr Phe Asp Ser Asn Ala Ser Thr
195 200 205
Leu Asp Asp Val Ile Lys Gln Ala Leu Glu Asn Ile Thr Phe Asn Cys
210 215 220
Phe Asp Cys Asn Ser Ala Ile Ser Ser Ile Gly Gln Ile Leu Met Glu
225 230 235 240

Ala Gly Lys Thr Lys Ser Asp Lys Ala Lys Ala Ile Glu His Leu Val
245 250 255
Asp Thr Tyr Ile Ala Thr Asp Thr Val Asp Thr Ser Ser Lys Thr Gln
260 265 270
Lys Asp Gln Val Lys Glu Asp Lys Lys Arg Leu Lys Ala Phe Ala Asn
275 280 285
Leu Val Leu Gly Leu Asn Ala Ser Leu Ile Asp Leu Phe Gly Ser Val
290 295 300

Glu Glu Leu Glu Glu Asp Leu Lys Lys Leu Gln Ile Thr Gly Asp Thr
305 310 315 320
Tyr Asp Asp Lys Arg Asp Glu Leu Ala Lys Ala Trp Ser Asp Glu Ile
325 330 335
Tyr Ile Ile Asp Asp Cys Lys Ser Val Tyr Asp Ala Ile Ile Leu Leu

340 345 350

Ser Ile Lys Glu Pro Gly Leu Thr Ile Ser Glu Ser Lys Val Lys Ala

355 360 365

Phe Asn Lys His Lys Asp Asp Leu Ala Ile Leu Lys Ser Leu Leu Lys

370 375 380

Ser Asp Arg Ser Ile Tyr Asn Thr Met Phe Lys Val Asp Glu Lys Gly

385 390 395 400

Leu His Asn Tyr Val His Tyr Ile Lys Gln Gly Arg Thr Glu Glu Thr

405 410 415

Ser Cys Asn Arg Glu Asp Phe Tyr Lys Tyr Thr Lys Lys Ile Val Glu

420 425 430

Gly Leu Ser Asp Ser Lys Asp Lys Glu Tyr Ile Leu Ser Gln Ile Glu

435 440 445

Leu Gln Ile Leu Leu Pro Leu Gln Arg Ile Lys Asp Asn Gly Val Ile

450 455 460

Pro Tyr Gln Leu His Leu Glu Glu Leu Lys Ala Ile Leu Ala Lys Cys

465 470 475 480

Gly Pro Lys Phe Pro Phe Leu Asn Glu Val Ala Asp Gly Phe Ser Val

485 490 495

Ala Glu Lys Leu Ile Lys Met Leu Glu Phe Arg Ile Pro Tyr Tyr Val

500 505 510

Gly Pro Leu Asn Thr His His Asn Val Asp Asn Gly Gly Phe Ala Trp

515 520 525

Ala Val Arg Lys Ala Ser Gly Arg Val Thr Pro Trp Asn Phe Asp Asp

530 535 540

Lys Ile Asp Arg Glu Lys Ser Ala Ala Ala Phe Ile Lys Asn Leu Thr

545 550 555 560

Asn Lys Cys Thr Tyr Leu Leu Gly Glu Asp Val Leu Pro Lys Ser Ser

565 570 575

Leu Leu Tyr Ser Glu Phe Met Leu Leu Asn Glu Leu Asn Asn Val Arg

580 585 590

Ile Asp Gly Lys Pro Leu Glu Lys Val Val Lys Glu His Leu Ile Glu
595 600 605
Ala Val Phe Lys Gln Asp His Lys Lys Met Thr Lys Asn Arg Ile Glu
610 615 620

Gln Phe Leu Lys Asp Asn Gly Tyr Ile Ser Glu Thr His Lys His Glu
625 630 635 640
Ile Thr Gly Leu Asp Gly Glu Ile Lys Asn Asp Leu Ala Ser Tyr Arg
645 650 655
Asp Met Val Arg Ile Leu Gly Asp Gly Phe Asp Arg Ser Met Ala Glu
660 665 670
Glu Ile Ile Thr Asp Ile Thr Ile Phe Gly Glu Ser Lys Lys Met Leu
675 680 685

Arg Glu Thr Leu Arg Lys Lys Phe Ala Ser Cys Leu Asp Asp Glu Ala
690 695 700
Ile Lys Lys Leu Thr Lys Leu Arg Tyr Arg Asp Trp Gly Arg Leu Ser
705 710 715 720
Gln Lys Leu Leu Asn Gly Ile Glu Gly Cys Asp Lys Ala Gly Asp Gly
725 730 735
Thr Pro Glu Thr Ile Ile Ile Leu Met Arg Asn Phe Ser Tyr Asn Leu
740 745 750

Met Glu Leu Leu Gly Asp Lys Phe Ser Phe Met Glu Arg Ile Gln Glu
755 760 765
Ile Asn Ala Lys Leu Thr Glu Gly Gln Ile Val Asn Pro His Asp Ile
770 775 780
Ile Asp Asp Leu Ala Leu Ser Pro Ala Val Lys Arg Ala Val Trp Gln
785 790 795 800
Ala Leu Arg Ile Val Asp Glu Val Ala His Ile Lys Lys Ala Leu Pro
805 810 815

Ala Arg Ile Phe Val Glu Val Thr Arg Ser Asn Lys Asn Glu Lys Lys
820 825 830
Lys Lys Asp Ser Arg Gln Lys Arg Leu Ser Asp Leu Tyr Ala Ala Ile

835 840 845

Lys Lys Asp Asp Val Leu Leu Asn Gly Leu Asn Asn Glu Ile Phe Gly

850 855 860

Glu Leu Lys Ser Ser Leu Ala Lys Tyr Asp Asp Ala Ala Leu Arg Ser

865 870 875 880

Lys Lys Leu Tyr Leu Tyr Tyr Thr Gln Met Gly Arg Cys Ala Tyr Thr

885 890 895

Gly Glu Ile Ile Glu Leu Ser Leu Leu Asn Thr Asp Asn Tyr Asp Ile

900 905 910

Asp His Ile Tyr Pro Arg Ser Leu Thr Lys Asp Asp Ser Phe Asp Asn

915 920 925

Leu Val Leu Cys Lys Arg Thr Ala Asn Ala Gln Lys Ser Asp Ala Tyr

930 935 940

Pro Ile Ser Glu Glu Ile Gln Lys Thr Gln Lys Pro Phe Trp Thr Phe

945 950 955 960

Leu Lys Gln Gln Gly Leu Ile Ser Glu Arg Lys Tyr Glu Arg Leu Thr

965 970 975

Arg Ile Thr Pro Leu Thr Ala Asp Asp Leu Ser Gly Phe Ile Ala Arg

980 985 990

Gln Leu Val Glu Thr Asn Gln Ser Val Lys Ala Ala Thr Thr Leu Leu

995 1000 1005

Arg Arg Leu Tyr Pro Gly Val Asp Val Val Phe Val Lys Ala Glu

1010 1015 1020

Asn Val Thr Asp Phe Arg His Asp Asn Asn Phe Ile Lys Val Arg

1025 1030 1035

Ser Leu Asn His His His His Ala Lys Asp Ala Tyr Leu Asn Ile

1040 1045 1050

Val Val Gly Asn Val Tyr His Glu Arg Phe Thr Arg Asn Phe Arg

1055 1060 1065

Ala Phe Phe Lys Lys Asn Gly Ala Asn Arg Thr Tyr Asn Leu Ala

1070 1075 1080

Lys Met Phe Asn Tyr Asp Val Asn Cys Thr Asn Ala Lys Asp Gly
1085 1090 1095
Lys Ala Trp Asp Val Lys Thr Ser Met Asp Thr Val Lys Lys Met
1100 1105 1110
Met Asp Ser Asn Asp Val Arg Val Thr Lys Arg Leu Leu Glu Gln
1115 1120 1125

Thr Gly Ala Leu Ala Asp Ala Thr Ile Tyr Lys Ala Thr Val Ala
1130 1135 1140
Gly Lys Ala Lys Asp Gly Ala Tyr Ile Gly Met Lys Thr Lys Ser
1145 1150 1155
Ser Val Phe Ala Asp Val Ser Lys Tyr Gly Gly Met Thr Lys Ile
1160 1165 1170
Lys Asn Ala Tyr Ser Ile Ile Val Gln Tyr Thr Gly Lys Lys Gly
1175 1180 1185

Glu Val Ile Lys Glu Ile Val Pro Leu Pro Ile Tyr Leu Thr Asn
1190 1195 1200
Arg Asn Thr Thr Asp Gln Asp Leu Ile Asn Tyr Val Ala Ser Ile
1205 1210 1215
Ile Pro Gln Ala Lys Asp Ile Ser Ile Ile Tyr Gly Lys Leu Cys
1220 1225 1230
Ile Asn Gln Leu Val Lys Val Asn Gly Phe Tyr Tyr Tyr Leu Gly
1235 1240 1245

Gly Lys Thr Asn Ser Lys Phe Cys Ile Asp Asn Ala Ile Gln Val
1250 1255 1260
Ile Val Ser Asn Glu Trp Ile Pro Tyr Leu Lys Val Leu Glu Lys
1265 1270 1275
Phe Asn Asn Met Arg Lys Asp Asn Lys Asp Leu Lys Ala Asn Val
1280 1285 1290
Val Ser Thr Arg Ala Leu Asp Asn Lys His Thr Ile Glu Val Arg
1295 1300 1305

Ile Val Glu Glu Lys Asn Ile Glu Phe Phe Asp Tyr Leu Val Ser

1310 1315 1320

Lys Leu Lys Met Pro Ile Tyr Gln Lys Met Lys Gly Asn Lys Ala

1325 1330 1335

Ala Glu Leu Ser Glu Lys Gly Tyr Gly Leu Phe Lys Lys Met Ser

1340 1345 1350

Leu Glu Glu Gln Ser Ile His Leu Ile Glu Leu Leu Asn Leu Leu

1355 1360 1365

Thr Asn Gln Lys Thr Thr Phe Glu Val Lys Pro Leu Gly Ile Thr

1370 1375 1380

Ala Ser Arg Ser Thr Val Gly Ser Lys Ile Ser Asn Gln Asp Glu

1385 1390 1395

Phe Lys Val Ile Asn Glu Ser Ile Thr Gly Leu Tyr Ser Asn Glu

1400 1405 1410

Val Thr Ile Val Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly

1415 1420 1425

Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp

1430 1435 1440

Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp

1445 1450 1455

Val Pro Asp Tyr Ala

1460

<210> 10

<211> 1659

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 10

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Ser Glu Val Glu Phe

1 5 10 15

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala

20 25 30

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn

35 40 45

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro

50 55 60

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

65 70 75 80

Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro

85 90 95

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val

100 105 110

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met

115 120 125

Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu

130 135 140

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg

145 150 155 160

Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr

165 170 175

Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr

180 185 190

Pro Glu Ser Ser Gly Pro Lys Lys Lys Arg Lys Val Gly Ser Ile Ile

195 200 205

Asn Phe Gln Arg Arg Gly Leu Met Glu Thr Gln Ala Ser Asn Gln Leu

210 215 220

Ile Ser Ser His Leu Lys Gly Tyr Pro Ile Lys Asp Tyr Phe Val Gly

225 230 235 240

Leu Ala Ile Gly Thr Ser Ser Val Gly Trp Ala Val Thr Asn Lys Ala

245 250 255

Tyr Glu Leu Leu Lys Phe Arg Ser His Lys Met Trp Gly Ser Arg Leu

260 265 270

Phe Asp Glu Gly Glu Ser Ala Val Ala Arg Arg Gly Phe Arg Ser Met

275 280 285

Arg Arg Arg Leu Glu Arg Arg Lys Leu Arg Leu Lys Leu Leu Glu Glu

290 295 300

Leu Phe Ala Asp Ala Met Ala Gln Val Asp Pro Thr Phe Phe Met Arg

305 310 315 320

Leu Arg Glu Ser Lys Tyr His Tyr Glu Asp Lys Thr Thr Gly His Ser

325 330 335

Ser Lys His Ile Leu Phe Ile Asp Lys Asn Tyr Asn Asp Gln Asp Tyr

340 345 350

Phe Lys Glu Tyr Pro Thr Val Tyr His Leu Arg Ser Glu Leu Met Lys

355 360 365

Ser Gly Thr Asp Asp Ile Arg Lys Leu Phe Leu Ala Val His His Ile

370 375 380

Leu Lys Tyr Arg Gly Asn Phe Leu Tyr Glu Gly Ala Thr Phe Asp Ser

385 390 395 400

Asn Ala Ser Thr Leu Asp Asp Val Ile Lys Gln Ala Leu Glu Asn Ile

405 410 415

Thr Phe Asn Cys Phe Asp Cys Asn Ser Ala Ile Ser Ser Ile Gly Gln

420 425 430

Ile Leu Met Glu Ala Gly Lys Thr Lys Ser Asp Lys Ala Lys Ala Ile

435 440 445

Glu His Leu Val Asp Thr Tyr Ile Ala Thr Asp Thr Val Asp Thr Ser

450 455 460

Ser Lys Thr Gln Lys Asp Gln Val Lys Glu Asp Lys Lys Arg Leu Lys

465 470 475 480

Ala Phe Ala Asn Leu Val Leu Gly Leu Asn Ala Ser Leu Ile Asp Leu

485 490 495

Phe Gly Ser Val Glu Glu Leu Glu Glu Asp Leu Lys Lys Leu Gln Ile

500 505 510

Thr Gly Asp Thr Tyr Asp Asp Lys Arg Asp Glu Leu Ala Lys Ala Trp

515 520 525

Ser Asp Glu Ile Tyr Ile Ile Asp Asp Cys Lys Ser Val Tyr Asp Ala
530 535 540
Ile Ile Leu Leu Ser Ile Lys Glu Pro Gly Leu Thr Ile Ser Glu Ser
545 550 555 560
Lys Val Lys Ala Phe Asn Lys His Lys Asp Asp Leu Ala Ile Leu Lys
565 570 575
Ser Leu Leu Lys Ser Asp Arg Ser Ile Tyr Asn Thr Met Phe Lys Val

580 585 590
Asp Glu Lys Gly Leu His Asn Tyr Val His Tyr Ile Lys Gln Gly Arg
595 600 605
Thr Glu Glu Thr Ser Cys Asn Arg Glu Asp Phe Tyr Lys Tyr Thr Lys
610 615 620
Lys Ile Val Glu Gly Leu Ser Asp Ser Lys Asp Lys Glu Tyr Ile Leu
625 630 635 640
Ser Gln Ile Glu Leu Gln Ile Leu Leu Pro Leu Gln Arg Ile Lys Asp

645 650 655
Asn Gly Val Ile Pro Tyr Gln Leu His Leu Glu Glu Leu Lys Ala Ile
660 665 670
Leu Ala Lys Cys Gly Pro Lys Phe Pro Phe Leu Asn Glu Val Ala Asp
675 680 685
Gly Phe Ser Val Ala Glu Lys Leu Ile Lys Met Leu Glu Phe Arg Ile
690 695 700
Pro Tyr Tyr Val Gly Pro Leu Asn Thr His His Asn Val Asp Asn Gly

705 710 715 720
Gly Phe Ala Trp Ala Val Arg Lys Ala Ser Gly Arg Val Thr Pro Trp
725 730 735
Asn Phe Asp Asp Lys Ile Asp Arg Glu Lys Ser Ala Ala Ala Phe Ile
740 745 750
Lys Asn Leu Thr Asn Lys Cys Thr Tyr Leu Leu Gly Glu Asp Val Leu
755 760 765
Pro Lys Ser Ser Leu Leu Tyr Ser Glu Phe Met Leu Leu Asn Glu Leu

770 775 780

Asn Asn Val Arg Ile Asp Gly Lys Pro Leu Glu Lys Val Val Lys Glu

785 790 795 800

His Leu Ile Glu Ala Val Phe Lys Gln Asp His Lys Lys Met Thr Lys

805 810 815

Asn Arg Ile Glu Gln Phe Leu Lys Asp Asn Gly Tyr Ile Ser Glu Thr

820 825 830

His Lys His Glu Ile Thr Gly Leu Asp Gly Glu Ile Lys Asn Asp Leu

835 840 845

Ala Ser Tyr Arg Asp Met Val Arg Ile Leu Gly Asp Gly Phe Asp Arg

850 855 860

Ser Met Ala Glu Glu Ile Ile Thr Asp Ile Thr Ile Phe Gly Glu Ser

865 870 875 880

Lys Lys Met Leu Arg Glu Thr Leu Arg Lys Lys Phe Ala Ser Cys Leu

885 890 895

Asp Asp Glu Ala Ile Lys Lys Leu Thr Lys Leu Arg Tyr Arg Asp Trp

900 905 910

Gly Arg Leu Ser Gln Lys Leu Leu Asn Gly Ile Glu Gly Cys Asp Lys

915 920 925

Ala Gly Asp Gly Thr Pro Glu Thr Ile Ile Ile Leu Met Arg Asn Phe

930 935 940

Ser Tyr Asn Leu Met Glu Leu Leu Gly Asp Lys Phe Ser Phe Met Glu

945 950 955 960

Arg Ile Gln Glu Ile Asn Ala Lys Leu Thr Glu Gly Gln Ile Val Asn

965 970 975

Pro His Asp Ile Ile Asp Asp Leu Ala Leu Ser Pro Ala Val Lys Arg

980 985 990

Ala Val Trp Gln Ala Leu Arg Ile Val Asp Glu Val Ala His Ile Lys

995 1000 1005

Lys Ala Leu Pro Ala Arg Ile Phe Val Glu Val Thr Arg Ser Asn

1010 1015 1020

Lys Asn Glu Lys Lys Lys Lys Asp Ser Arg Gln Lys Arg Leu Ser

1025 1030 1035

Asp Leu Tyr Ala Ala Ile Lys Lys Asp Asp Val Leu Leu Asn Gly

1040 1045 1050

Leu Asn Asn Glu Ile Phe Gly Glu Leu Lys Ser Ser Leu Ala Lys

1055 1060 1065

Tyr Asp Asp Ala Ala Leu Arg Ser Lys Lys Leu Tyr Leu Tyr Tyr

1070 1075 1080

Thr Gln Met Gly Arg Cys Ala Tyr Thr Gly Glu Ile Ile Glu Leu

1085 1090 1095

Ser Leu Leu Asn Thr Asp Asn Tyr Asp Ile Asp His Ile Tyr Pro

1100 1105 1110

Arg Ser Leu Thr Lys Asp Asp Ser Phe Asp Asn Leu Val Leu Cys

1115 1120 1125

Lys Arg Thr Ala Asn Ala Gln Lys Ser Asp Ala Tyr Pro Ile Ser

1130 1135 1140

Glu Glu Ile Gln Lys Thr Gln Lys Pro Phe Trp Thr Phe Leu Lys

1145 1150 1155

Gln Gln Gly Leu Ile Ser Glu Arg Lys Tyr Glu Arg Leu Thr Arg

1160 1165 1170

Ile Thr Pro Leu Thr Ala Asp Asp Leu Ser Gly Phe Ile Ala Arg

1175 1180 1185

Gln Leu Val Glu Thr Asn Gln Ser Val Lys Ala Ala Thr Thr Leu

1190 1195 1200

Leu Arg Arg Leu Tyr Pro Gly Val Asp Val Val Phe Val Lys Ala

1205 1210 1215

Glu Asn Val Thr Asp Phe Arg His Asp Asn Asn Phe Ile Lys Val

1220 1225 1230

Arg Ser Leu Asn His His His His Ala Lys Asp Ala Tyr Leu Asn

1235 1240 1245

Ile Val Val Gly Asn Val Tyr His Glu Arg Phe Thr Arg Asn Phe

1250 1255 1260

Arg Ala Phe Phe Lys Lys Asn Gly Ala Asn Arg Thr Tyr Asn Leu

1265 1270 1275

Ala Lys Met Phe Asn Tyr Asp Val Asn Cys Thr Asn Ala Lys Asp

1280 1285 1290

Gly Lys Ala Trp Asp Val Lys Thr Ser Met Asp Thr Val Lys Lys

1295 1300 1305

Met Met Asp Ser Asn Asp Val Arg Val Thr Lys Arg Leu Leu Glu

1310 1315 1320

Gln Thr Gly Ala Leu Ala Asp Ala Thr Ile Tyr Lys Ala Thr Val

1325 1330 1335

Ala Gly Lys Ala Lys Asp Gly Ala Tyr Ile Gly Met Lys Thr Lys

1340 1345 1350

Ser Ser Val Phe Ala Asp Val Ser Lys Tyr Gly Gly Met Thr Lys

1355 1360 1365

Ile Lys Asn Ala Tyr Ser Ile Ile Val Gln Tyr Thr Gly Lys Lys

1370 1375 1380

Gly Glu Val Ile Lys Glu Ile Val Pro Leu Pro Ile Tyr Leu Thr

1385 1390 1395

Asn Arg Asn Thr Thr Asp Gln Asp Leu Ile Asn Tyr Val Ala Ser

1400 1405 1410

Ile Ile Pro Gln Ala Lys Asp Ile Ser Ile Ile Tyr Gly Lys Leu

1415 1420 1425

Cys Ile Asn Gln Leu Val Lys Val Asn Gly Phe Tyr Tyr Tyr Leu

1430 1435 1440

Gly Gly Lys Thr Asn Ser Lys Phe Cys Ile Asp Asn Ala Ile Gln

1445 1450 1455

Val Ile Val Ser Asn Glu Trp Ile Pro Tyr Leu Lys Val Leu Glu

1460 1465 1470

Lys Phe Asn Asn Met Arg Lys Asp Asn Lys Asp Leu Lys Ala Asn

1475 1480 1485

Val Val Ser Thr Arg Ala Leu Asp Asn Lys His Thr Ile Glu Val
1490 1495 1500
Arg Ile Val Glu Glu Lys Asn Ile Glu Phe Phe Asp Tyr Leu Val

1505 1510 1515
Ser Lys Leu Lys Met Pro Ile Tyr Gln Lys Met Lys Gly Asn Lys
1520 1525 1530
Ala Ala Glu Leu Ser Glu Lys Gly Tyr Gly Leu Phe Lys Lys Met
1535 1540 1545
Ser Leu Glu Glu Gln Ser Ile His Leu Ile Glu Leu Leu Asn Leu
1550 1555 1560
Leu Thr Asn Gln Lys Thr Thr Phe Glu Val Lys Pro Leu Gly Ile

1565 1570 1575
Thr Ala Ser Arg Ser Thr Val Gly Ser Lys Ile Ser Asn Gln Asp
1580 1585 1590
Glu Phe Lys Val Ile Asn Glu Ser Ile Thr Gly Leu Tyr Ser Asn
1595 1600 1605
Glu Val Thr Ile Val Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala
1610 1615 1620
Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro

1625 1630 1635
Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr
1640 1645 1650
Asp Val Pro Asp Tyr Ala
1655
<210> 11
<211> 1654
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 11
Met Pro Lys Lys Lys Arg Lys Val Ser Ile Ile Asn Phe Gln Arg Arg

1 5 10 15

Gly Leu Met Glu Thr Gln Ala Ser Asn Gln Leu Ile Ser Ser His Leu
20 25 30
Lys Gly Tyr Pro Ile Lys Asp Tyr Phe Val Gly Leu Ala Ile Gly Thr
35 40 45
Ser Ser Val Gly Trp Ala Val Thr Asn Lys Ala Tyr Glu Leu Leu Lys
50 55 60
Phe Arg Ser His Lys Met Trp Gly Ser Arg Leu Phe Asp Glu Gly Glu
65 70 75 80

Ser Ala Val Ala Arg Arg Gly Phe Arg Ser Met Arg Arg Arg Leu Glu
85 90 95
Arg Arg Lys Leu Arg Leu Lys Leu Leu Glu Glu Leu Phe Ala Asp Ala
100 105 110
Met Ala Gln Val Asp Pro Thr Phe Phe Met Arg Leu Arg Glu Ser Lys
115 120 125
Tyr His Tyr Glu Asp Lys Thr Thr Gly His Ser Ser Lys His Ile Leu
130 135 140

Phe Ile Asp Lys Asn Tyr Asn Asp Gln Asp Tyr Phe Lys Glu Tyr Pro
145 150 155 160
Thr Val Tyr His Leu Arg Ser Glu Leu Met Lys Ser Gly Thr Asp Asp
165 170 175
Ile Arg Lys Leu Phe Leu Ala Val His His Ile Leu Lys Tyr Arg Gly
180 185 190
Asn Phe Leu Tyr Glu Gly Ala Thr Phe Asp Ser Asn Ala Ser Thr Leu
195 200 205

Asp Asp Val Ile Lys Gln Ala Leu Glu Asn Ile Thr Phe Asn Cys Phe
210 215 220
Asp Cys Asn Ser Ala Ile Ser Ser Ile Gly Gln Ile Leu Met Glu Ala
225 230 235 240
Gly Lys Thr Lys Ser Asp Lys Ala Lys Ala Ile Glu His Leu Val Asp
245 250 255

Thr Tyr Ile Ala Thr Asp Thr Val Asp Thr Ser Ser Lys Thr Gln Lys
260 265 270

Asp Gln Val Lys Glu Asp Lys Lys Arg Leu Lys Ala Phe Ala Asn Leu
275 280 285

Val Leu Gly Leu Asn Ala Ser Leu Ile Asp Leu Phe Gly Ser Val Glu
290 295 300

Glu Leu Glu Glu Asp Leu Lys Lys Leu Gln Ile Thr Gly Asp Thr Tyr
305 310 315 320

Asp Asp Lys Arg Asp Glu Leu Ala Lys Ala Trp Ser Asp Glu Ile Tyr
325 330 335

Ile Ile Asp Asp Cys Lys Ser Val Tyr Asp Ala Ile Ile Leu Leu Ser
340 345 350

Ile Lys Glu Pro Gly Leu Thr Ile Ser Glu Ser Lys Val Lys Ala Phe
355 360 365

Asn Lys His Lys Asp Asp Leu Ala Ile Leu Lys Ser Leu Leu Lys Ser
370 375 380

Asp Arg Ser Ile Tyr Asn Thr Met Phe Lys Val Asp Glu Lys Gly Leu
385 390 395 400

His Asn Tyr Val His Tyr Ile Lys Gln Gly Arg Thr Glu Glu Thr Ser
405 410 415

Cys Asn Arg Glu Asp Phe Tyr Lys Tyr Thr Lys Lys Ile Val Glu Gly
420 425 430

Leu Ser Asp Ser Lys Asp Lys Glu Tyr Ile Leu Ser Gln Ile Glu Leu
435 440 445

Gln Ile Leu Leu Pro Leu Gln Arg Ile Lys Asp Asn Gly Val Ile Pro
450 455 460

Tyr Gln Leu His Leu Glu Glu Leu Lys Ala Ile Leu Ala Lys Cys Gly
465 470 475 480

Pro Lys Phe Pro Phe Leu Asn Glu Val Ala Asp Gly Phe Ser Val Ala
485 490 495

Glu Lys Leu Ile Lys Met Leu Glu Phe Arg Ile Pro Tyr Tyr Val Gly

500 505 510

Pro Leu Asn Thr His His Asn Val Asp Asn Gly Gly Phe Ala Trp Ala

515 520 525

Val Arg Lys Ala Ser Gly Arg Val Thr Pro Trp Asn Phe Asp Asp Lys

530 535 540

Ile Asp Arg Glu Lys Ser Ala Ala Ala Phe Ile Lys Asn Leu Thr Asn

545 550 555 560

Lys Cys Thr Tyr Leu Leu Gly Glu Asp Val Leu Pro Lys Ser Ser Leu

565 570 575

Leu Tyr Ser Glu Phe Met Leu Leu Asn Glu Leu Asn Asn Val Arg Ile

580 585 590

Asp Gly Lys Pro Leu Glu Lys Val Val Lys Glu His Leu Ile Glu Ala

595 600 605

Val Phe Lys Gln Asp His Lys Lys Met Thr Lys Asn Arg Ile Glu Gln

610 615 620

Phe Leu Lys Asp Asn Gly Tyr Ile Ser Glu Thr His Lys His Glu Ile

625 630 635 640

Thr Gly Leu Asp Gly Glu Ile Lys Asn Asp Leu Ala Ser Tyr Arg Asp

645 650 655

Met Val Arg Ile Leu Gly Asp Gly Phe Asp Arg Ser Met Ala Glu Glu

660 665 670

Ile Ile Thr Asp Ile Thr Ile Phe Gly Glu Ser Lys Lys Met Leu Arg

675 680 685

Glu Thr Leu Arg Lys Lys Phe Ala Ser Cys Leu Asp Asp Glu Ala Ile

690 695 700

Lys Lys Leu Thr Lys Leu Arg Tyr Arg Asp Trp Gly Arg Leu Ser Gln

705 710 715 720

Lys Leu Leu Asn Gly Ile Glu Gly Cys Asp Lys Ala Gly Asp Gly Thr

725 730 735

Pro Glu Thr Ile Ile Ile Leu Met Arg Asn Phe Ser Tyr Asn Leu Met

740 745 750

Glu Leu Leu Gly Asp Lys Phe Ser Phe Met Glu Arg Ile Gln Glu Ile
755 760 765

Asn Ala Lys Leu Thr Glu Gly Gln Ile Val Asn Pro His Asp Ile Ile
770 775 780

Asp Asp Leu Ala Leu Ser Pro Ala Val Lys Arg Ala Val Trp Gln Ala
785 790 795 800

Leu Arg Ile Val Asp Glu Val Ala His Ile Lys Lys Ala Leu Pro Ala
805 810 815

Arg Ile Phe Val Glu Val Thr Arg Ser Asn Lys Asn Glu Lys Lys Lys
820 825 830

Lys Asp Ser Arg Gln Lys Arg Leu Ser Asp Leu Tyr Ala Ala Ile Lys
835 840 845

Lys Asp Asp Val Leu Leu Asn Gly Leu Asn Asn Glu Ile Phe Gly Glu
850 855 860

Leu Lys Ser Ser Leu Ala Lys Tyr Asp Asp Ala Ala Leu Arg Ser Lys
865 870 875 880

Lys Leu Tyr Leu Tyr Tyr Thr Gln Met Gly Arg Cys Ala Tyr Thr Gly
885 890 895

Glu Ile Ile Glu Leu Ser Leu Leu Asn Thr Asp Asn Tyr Asp Ile Asp
900 905 910

His Ile Tyr Pro Arg Ser Leu Thr Lys Asp Asp Ser Phe Asp Asn Leu
915 920 925

Val Leu Cys Lys Arg Thr Ala Asn Ala Gln Lys Ser Asp Ala Tyr Pro
930 935 940

Ile Ser Glu Glu Ile Gln Lys Thr Gln Lys Pro Phe Trp Thr Phe Leu
945 950 955 960

Lys Gln Gln Gly Leu Ile Ser Glu Arg Lys Tyr Glu Arg Leu Thr Arg
965 970 975

Ile Thr Pro Leu Thr Ala Asp Asp Leu Ser Gly Phe Ile Ala Arg Gln
980 985 990

Leu Val Glu Thr Asn Gln Ser Val Lys Ala Ala Thr Thr Leu Leu Arg

995 1000 1005

Arg Leu Tyr Pro Gly Val Asp Val Val Phe Val Lys Ala Glu Asn

1010 1015 1020

Val Thr Asp Phe Arg His Asp Asn Asn Phe Ile Lys Val Arg Ser

1025 1030 1035

Leu Asn His His His His Ala Lys Asp Ala Tyr Leu Asn Ile Val

1040 1045 1050

Val Gly Asn Val Tyr His Glu Arg Phe Thr Arg Asn Phe Arg Ala

1055 1060 1065

Phe Phe Lys Lys Asn Gly Ala Asn Arg Thr Tyr Asn Leu Ala Lys

1070 1075 1080

Met Phe Asn Tyr Asp Val Asn Cys Thr Asn Ala Lys Asp Gly Lys

1085 1090 1095

Ala Trp Asp Val Lys Thr Ser Met Asp Thr Val Lys Lys Met Met

1100 1105 1110

Asp Ser Asn Asp Val Arg Val Thr Lys Arg Leu Leu Glu Gln Thr

1115 1120 1125

Gly Ala Leu Ala Asp Ala Thr Ile Tyr Lys Ala Thr Val Ala Gly

1130 1135 1140

Lys Ala Lys Asp Gly Ala Tyr Ile Gly Met Lys Thr Lys Ser Ser

1145 1150 1155

Val Phe Ala Asp Val Ser Lys Tyr Gly Gly Met Thr Lys Ile Lys

1160 1165 1170

Asn Ala Tyr Ser Ile Ile Val Gln Tyr Thr Gly Lys Lys Gly Glu

1175 1180 1185

Val Ile Lys Glu Ile Val Pro Leu Pro Ile Tyr Leu Thr Asn Arg

1190 1195 1200

Asn Thr Thr Asp Gln Asp Leu Ile Asn Tyr Val Ala Ser Ile Ile

1205 1210 1215

Pro Gln Ala Lys Asp Ile Ser Ile Ile Tyr Gly Lys Leu Cys Ile

1220 1225 1230

Asn Gln Leu Val Lys Val Asn Gly Phe Tyr Tyr Tyr Leu Gly Gly
1235 1240 1245
Lys Thr Asn Ser Lys Phe Cys Ile Asp Asn Ala Ile Gln Val Ile
1250 1255 1260
Val Ser Asn Glu Trp Ile Pro Tyr Leu Lys Val Leu Glu Lys Phe
1265 1270 1275

Asn Asn Met Arg Lys Asp Asn Lys Asp Leu Lys Ala Asn Val Val
1280 1285 1290
Ser Thr Arg Ala Leu Asp Asn Lys His Thr Ile Glu Val Arg Ile
1295 1300 1305
Val Glu Glu Lys Asn Ile Glu Phe Phe Asp Tyr Leu Val Ser Lys
1310 1315 1320
Leu Lys Met Pro Ile Tyr Gln Lys Met Lys Gly Asn Lys Ala Ala
1325 1330 1335

Glu Leu Ser Glu Lys Gly Tyr Gly Leu Phe Lys Lys Met Ser Leu
1340 1345 1350
Glu Glu Gln Ser Ile His Leu Ile Glu Leu Leu Asn Leu Leu Thr
1355 1360 1365
Asn Gln Lys Thr Thr Phe Glu Val Lys Pro Leu Gly Ile Thr Ala
1370 1375 1380
Ser Arg Ser Thr Val Gly Ser Lys Ile Ser Asn Gln Asp Glu Phe
1385 1390 1395

Lys Val Ile Asn Glu Ser Ile Thr Gly Leu Tyr Ser Asn Glu Val
1400 1405 1410
Thr Ile Val Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala
1415 1420 1425
Lys Lys Lys Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
1430 1435 1440
Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu
1445 1450 1455

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp

1460 1465 1470

Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg

1475 1480 1485

Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro

1490 1495 1500

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val

1505 1510 1515

Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe

1520 1525 1530

Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile

1535 1540 1545

Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala

1550 1555 1560

Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg

1565 1570 1575

Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

1580 1585 1590

Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln

1595 1600 1605

Lys Lys Ala Gln Ser Ser Thr Asp Pro Ala Ala Lys Arg Val Lys

1610 1615 1620

Leu Asp Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro

1625 1630 1635

Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr

1640 1645 1650

Ala

<210> 12

<211> 1919

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 12

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Thr Ser Glu Lys Gly Pro
1 5 10 15

Ser Thr Gly Asp Pro Thr Leu Arg Arg Arg Ile Glu Ser Trp Glu Phe
20 25 30

Asp Val Phe Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu
35 40 45

Tyr Glu Ile Lys Trp Gly Met Ser Arg Lys Ile Trp Arg Ser Ser Gly
50 55 60

Lys Asn Thr Thr Asn His Val Glu Val Asn Phe Ile Lys Lys Phe Thr
65 70 75 80

Ser Glu Arg Arg Phe His Ser Ser Ile Ser Cys Ser Ile Thr Trp Phe
85 90 95

Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser Gln Ala Ile Arg Glu Phe
100 105 110

Leu Ser Gln His Pro Gly Val Thr Leu Val Ile Tyr Val Ala Arg Leu
115 120 125

Phe Trp His Met Asp Gln Arg Asn Arg Gln Gly Leu Arg Asp Leu Val
130 135 140

Asn Ser Gly Val Thr Ile Gln Ile Met Arg Ala Ser Glu Tyr Tyr His
145 150 155 160

Cys Trp Arg Asn Phe Val Asn Tyr Pro Pro Gly Asp Glu Ala His Trp
165 170 175

Pro Gln Tyr Pro Pro Leu Trp Met Met Leu Tyr Ala Leu Glu Leu His
180 185 190

Cys Ile Ile Leu Ser Leu Pro Pro Cys Leu Lys Ile Ser Arg Arg Trp
195 200 205

Gln Asn His Leu Ala Phe Phe Arg Leu His Leu Gln Asn Cys His Tyr
210 215 220

Gln Thr Ile Pro Pro His Ile Leu Leu Ala Thr Gly Leu Ile His Pro

225 230 235 240

Ser Val Thr Trp Arg Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser

245 250 255

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

260 265 270

Gly Gly Ser Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val Gly Ser

275 280 285

Ile Ile Asn Phe Gln Arg Arg Gly Leu Met Glu Thr Gln Ala Ser Asn

290 295 300

Gln Leu Ile Ser Ser His Leu Lys Gly Tyr Pro Ile Lys Asp Tyr Phe

305 310 315 320

Val Gly Leu Ala Ile Gly Thr Ser Ser Val Gly Trp Ala Val Thr Asn

325 330 335

Lys Ala Tyr Glu Leu Leu Lys Phe Arg Ser His Lys Met Trp Gly Ser

340 345 350

Arg Leu Phe Asp Glu Gly Glu Ser Ala Val Ala Arg Arg Gly Phe Arg

355 360 365

Ser Met Arg Arg Arg Leu Glu Arg Arg Lys Leu Arg Leu Lys Leu Leu

370 375 380

Glu Glu Leu Phe Ala Asp Ala Met Ala Gln Val Asp Pro Thr Phe Phe

385 390 395 400

Met Arg Leu Arg Glu Ser Lys Tyr His Tyr Glu Asp Lys Thr Thr Gly

405 410 415

His Ser Ser Lys His Ile Leu Phe Ile Asp Lys Asn Tyr Asn Asp Gln

420 425 430

Asp Tyr Phe Lys Glu Tyr Pro Thr Val Tyr His Leu Arg Ser Glu Leu

435 440 445

Met Lys Ser Gly Thr Asp Asp Ile Arg Lys Leu Phe Leu Ala Val His

450 455 460

His Ile Leu Lys Tyr Arg Gly Asn Phe Leu Tyr Glu Gly Ala Thr Phe

465 470 475 480

Asp Ser Asn Ala Ser Thr Leu Asp Asp Val Ile Lys Gln Ala Leu Glu
485 490 495
Asn Ile Thr Phe Asn Cys Phe Asp Cys Asn Ser Ala Ile Ser Ser Ile
500 505 510
Gly Gln Ile Leu Met Glu Ala Gly Lys Thr Lys Ser Asp Lys Ala Lys
515 520 525
Ala Ile Glu His Leu Val Asp Thr Tyr Ile Ala Thr Asp Thr Val Asp
530 535 540

Thr Ser Ser Lys Thr Gln Lys Asp Gln Val Lys Glu Asp Lys Lys Arg
545 550 555 560
Leu Lys Ala Phe Ala Asn Leu Val Leu Gly Leu Asn Ala Ser Leu Ile
565 570 575
Asp Leu Phe Gly Ser Val Glu Glu Leu Glu Glu Asp Leu Lys Lys Leu
580 585 590
Gln Ile Thr Gly Asp Thr Tyr Asp Asp Lys Arg Asp Glu Leu Ala Lys
595 600 605

Ala Trp Ser Asp Glu Ile Tyr Ile Ile Asp Asp Cys Lys Ser Val Tyr
610 615 620
Asp Ala Ile Ile Leu Leu Ser Ile Lys Glu Pro Gly Leu Thr Ile Ser
625 630 635 640
Glu Ser Lys Val Lys Ala Phe Asn Lys His Lys Asp Asp Leu Ala Ile
645 650 655
Leu Lys Ser Leu Leu Lys Ser Asp Arg Ser Ile Tyr Asn Thr Met Phe
660 665 670

Lys Val Asp Glu Lys Gly Leu His Asn Tyr Val His Tyr Ile Lys Gln
675 680 685
Gly Arg Thr Glu Glu Thr Ser Cys Asn Arg Glu Asp Phe Tyr Lys Tyr
690 695 700
Thr Lys Lys Ile Val Glu Gly Leu Ser Asp Ser Lys Asp Lys Glu Tyr
705 710 715 720
Ile Leu Ser Gln Ile Glu Leu Gln Ile Leu Leu Pro Leu Gln Arg Ile

725 730 735

Lys Asp Asn Gly Val Ile Pro Tyr Gln Leu His Leu Glu Glu Leu Lys
740 745 750

Ala Ile Leu Ala Lys Cys Gly Pro Lys Phe Pro Phe Leu Asn Glu Val
755 760 765

Ala Asp Gly Phe Ser Val Ala Glu Lys Leu Ile Lys Met Leu Glu Phe
770 775 780

Arg Ile Pro Tyr Tyr Val Gly Pro Leu Asn Thr His His Asn Val Asp
785 790 795 800

Asn Gly Gly Phe Ala Trp Ala Val Arg Lys Ala Ser Gly Arg Val Thr
805 810 815

Pro Trp Asn Phe Asp Asp Lys Ile Asp Arg Glu Lys Ser Ala Ala Ala
820 825 830

Phe Ile Lys Asn Leu Thr Asn Lys Cys Thr Tyr Leu Leu Gly Glu Asp
835 840 845

Val Leu Pro Lys Ser Ser Leu Leu Tyr Ser Glu Phe Met Leu Leu Asn
850 855 860

Glu Leu Asn Asn Val Arg Ile Asp Gly Lys Pro Leu Glu Lys Val Val
865 870 875 880

Lys Glu His Leu Ile Glu Ala Val Phe Lys Gln Asp His Lys Lys Met
885 890 895

Thr Lys Asn Arg Ile Glu Gln Phe Leu Lys Asp Asn Gly Tyr Ile Ser
900 905 910

Glu Thr His Lys His Glu Ile Thr Gly Leu Asp Gly Glu Ile Lys Asn
915 920 925

Asp Leu Ala Ser Tyr Arg Asp Met Val Arg Ile Leu Gly Asp Gly Phe
930 935 940

Asp Arg Ser Met Ala Glu Glu Ile Ile Thr Asp Ile Thr Ile Phe Gly
945 950 955 960

Glu Ser Lys Lys Met Leu Arg Glu Thr Leu Arg Lys Lys Phe Ala Ser
965 970 975

Cys Leu Asp Asp Glu Ala Ile Lys Lys Leu Thr Lys Leu Arg Tyr Arg
980 985 990

Asp Trp Gly Arg Leu Ser Gln Lys Leu Leu Asn Gly Ile Glu Gly Cys
995 1000 1005

Asp Lys Ala Gly Asp Gly Thr Pro Glu Thr Ile Ile Ile Leu Met
1010 1015 1020

Arg Asn Phe Ser Tyr Asn Leu Met Glu Leu Leu Gly Asp Lys Phe
1025 1030 1035

Ser Phe Met Glu Arg Ile Gln Glu Ile Asn Ala Lys Leu Thr Glu
1040 1045 1050

Gly Gln Ile Val Asn Pro His Asp Ile Ile Asp Asp Leu Ala Leu
1055 1060 1065

Ser Pro Ala Val Lys Arg Ala Val Trp Gln Ala Leu Arg Ile Val
1070 1075 1080

Asp Glu Val Ala His Ile Lys Lys Ala Leu Pro Ala Arg Ile Phe
1085 1090 1095

Val Glu Val Thr Arg Ser Asn Lys Asn Glu Lys Lys Lys Lys Asp
1100 1105 1110

Ser Arg Gln Lys Arg Leu Ser Asp Leu Tyr Ala Ala Ile Lys Lys
1115 1120 1125

Asp Asp Val Leu Leu Asn Gly Leu Asn Asn Glu Ile Phe Gly Glu
1130 1135 1140

Leu Lys Ser Ser Leu Ala Lys Tyr Asp Asp Ala Ala Leu Arg Ser
1145 1150 1155

Lys Lys Leu Tyr Leu Tyr Tyr Thr Gln Met Gly Arg Cys Ala Tyr
1160 1165 1170

Thr Gly Glu Ile Ile Glu Leu Ser Leu Leu Asn Thr Asp Asn Tyr
1175 1180 1185

Asp Ile Asp His Ile Tyr Pro Arg Ser Leu Thr Lys Asp Asp Ser
1190 1195 1200

Phe Asp Asn Leu Val Leu Cys Lys Arg Thr Ala Asn Ala Gln Lys

1205 1210 1215

Ser Asp Ala Tyr Pro Ile Ser Glu Glu Ile Gln Lys Thr Gln Lys

1220 1225 1230

Pro Phe Trp Thr Phe Leu Lys Gln Gln Gly Leu Ile Ser Glu Arg

1235 1240 1245

Lys Tyr Glu Arg Leu Thr Arg Ile Thr Pro Leu Thr Ala Asp Asp

1250 1255 1260

Leu Ser Gly Phe Ile Ala Arg Gln Leu Val Glu Thr Asn Gln Ser

1265 1270 1275

Val Lys Ala Ala Thr Thr Leu Leu Arg Arg Leu Tyr Pro Gly Val

1280 1285 1290

Asp Val Val Phe Val Lys Ala Glu Asn Val Thr Asp Phe Arg His

1295 1300 1305

Asp Asn Asn Phe Ile Lys Val Arg Ser Leu Asn His His His His

1310 1315 1320

Ala Lys Asp Ala Tyr Leu Asn Ile Val Val Gly Asn Val Tyr His

1325 1330 1335

Glu Arg Phe Thr Arg Asn Phe Arg Ala Phe Phe Lys Lys Asn Gly

1340 1345 1350

Ala Asn Arg Thr Tyr Asn Leu Ala Lys Met Phe Asn Tyr Asp Val

1355 1360 1365

Asn Cys Thr Asn Ala Lys Asp Gly Lys Ala Trp Asp Val Lys Thr

1370 1375 1380

Ser Met Asp Thr Val Lys Lys Met Met Asp Ser Asn Asp Val Arg

1385 1390 1395

Val Thr Lys Arg Leu Leu Glu Gln Thr Gly Ala Leu Ala Asp Ala

1400 1405 1410

Thr Ile Tyr Lys Ala Thr Val Ala Gly Lys Ala Lys Asp Gly Ala

1415 1420 1425

Tyr Ile Gly Met Lys Thr Lys Ser Ser Val Phe Ala Asp Val Ser

1430 1435 1440

Lys Tyr Gly Gly Met Thr Lys Ile Lys Asn Ala Tyr Ser Ile Ile
1445 1450 1455
Val Gln Tyr Thr Gly Lys Lys Gly Glu Val Ile Lys Glu Ile Val
1460 1465 1470

Pro Leu Pro Ile Tyr Leu Thr Asn Arg Asn Thr Thr Asp Gln Asp
1475 1480 1485
Leu Ile Asn Tyr Val Ala Ser Ile Ile Pro Gln Ala Lys Asp Ile
1490 1495 1500
Ser Ile Ile Tyr Gly Lys Leu Cys Ile Asn Gln Leu Val Lys Val
1505 1510 1515
Asn Gly Phe Tyr Tyr Tyr Leu Gly Gly Lys Thr Asn Ser Lys Phe
1520 1525 1530

Cys Ile Asp Asn Ala Ile Gln Val Ile Val Ser Asn Glu Trp Ile
1535 1540 1545
Pro Tyr Leu Lys Val Leu Glu Lys Phe Asn Asn Met Arg Lys Asp
1550 1555 1560
Asn Lys Asp Leu Lys Ala Asn Val Val Ser Thr Arg Ala Leu Asp
1565 1570 1575
Asn Lys His Thr Ile Glu Val Arg Ile Val Glu Glu Lys Asn Ile
1580 1585 1590

Glu Phe Phe Asp Tyr Leu Val Ser Lys Leu Lys Met Pro Ile Tyr
1595 1600 1605
Gln Lys Met Lys Gly Asn Lys Ala Ala Glu Leu Ser Glu Lys Gly
1610 1615 1620
Tyr Gly Leu Phe Lys Lys Met Ser Leu Glu Glu Gln Ser Ile His
1625 1630 1635
Leu Ile Glu Leu Leu Asn Leu Leu Thr Asn Gln Lys Thr Thr Phe
1640 1645 1650

Glu Val Lys Pro Leu Gly Ile Thr Ala Ser Arg Ser Thr Val Gly
1655 1660 1665
Ser Lys Ile Ser Asn Gln Asp Glu Phe Lys Val Ile Asn Glu Ser

1670 1675 1680

Ile Thr Gly Leu Tyr Ser Asn Glu Val Thr Ile Val Gly Lys Arg

1685 1690 1695

Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly

1700 1705 1710

Ser Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser

1715 1720 1725

Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu

1730 1735 1740

Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn

1745 1750 1755

Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser

1760 1765 1770

Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr

1775 1780 1785

Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys

1790 1795 1800

Ile Lys Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr

1805 1810 1815

Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val

1820 1825 1830

Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val

1835 1840 1845

Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr

1850 1855 1860

Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala

1865 1870 1875

Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly

1880 1885 1890

Glu Asn Lys Ile Lys Met Leu Tyr Pro Tyr Asp Val Pro Asp Tyr

1895 1900 1905

Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr

1910 1915

<210> 13

<211> 107

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 13

guuugagagu agugugaaaa cauuacgagu ucaaauacaa auuaauuuac aaugccuucg 60

ggcugcccga cguagggcac cuacucucaa uucuucggaa uugaguu 107

<210> 14

<211> 1370

<212> PRT

<213> Ezakiella peruensis

<400> 14

Met Thr Lys Val Lys Asp Tyr Tyr Ile Gly Leu Asp Ile Gly Thr Ser

1 5 10 15

Ser Val Gly Trp Ala Val Thr Asp Glu Ala Tyr Asn Val Leu Lys Phe

20 25 30

Asn Ser Lys Lys Met Trp Gly Val Arg Leu Phe Asp Asp Ala Lys Thr

35 40 45

Ala Glu Glu Arg Arg Gly Gln Arg Gly Ala Arg Arg Arg Leu Asp Arg

50 55 60

Lys Lys Glu Arg Leu Ser Leu Leu Gln Asp Phe Phe Ala Glu Glu Val

65 70 75 80

Ala Lys Val Asp Pro Asn Phe Phe Leu Arg Leu Asp Asn Ser Asp Leu

85 90 95

Tyr Met Glu Asp Lys Asp Gln Lys Leu Lys Ser Lys Tyr Thr Leu Phe

100 105 110

Asn Asp Lys Asp Phe Lys Asp Lys Asn Phe His Lys Lys Tyr Pro Thr

115 120 125

Ile His His Leu Leu Met Asp Leu Ile Glu Asp Asp Ser Lys Lys Asp
130 135 140
Ile Arg Leu Val Tyr Leu Ala Cys His Tyr Leu Leu Lys Asn Arg Gly
145 150 155 160
His Phe Ile Phe Glu Gly Gln Lys Phe Asp Thr Lys Ser Ser Phe Glu
165 170 175
Asn Ser Leu Asn Glu Leu Lys Val His Leu Asn Asp Glu Tyr Gly Leu

180 185 190
Asp Leu Glu Phe Asp Asn Glu Asn Leu Ile Asn Ile Leu Thr Asp Pro
195 200 205
Lys Leu Asn Lys Thr Ala Lys Lys Lys Glu Leu Lys Ser Val Ile Gly
210 215 220
Asp Thr Lys Phe Leu Lys Ala Val Ser Ala Ile Met Ile Gly Ser Ser
225 230 235 240
Gln Lys Leu Val Asp Leu Phe Glu Asn Pro Glu Asp Phe Asp Asp Ser

245 250 255
Ala Ile Lys Ser Val Asp Phe Ser Thr Thr Ser Phe Asp Asp Lys Tyr
260 265 270
Ser Asp Tyr Glu Leu Ala Leu Gly Asp Lys Ile Ala Leu Val Asn Ile
275 280 285
Leu Lys Glu Ile Tyr Asp Ser Ser Ile Leu Glu Asn Leu Leu Lys Glu
290 295 300
Ala Asp Lys Ser Lys Asp Gly Asn Lys Tyr Ile Ser Asn Ala Phe Val

305 310 315 320
Lys Lys Tyr Asn Lys His Gly Gln Asp Leu Lys Glu Phe Lys Arg Leu
325 330 335
Val Arg Gln Tyr His Lys Ser Ala Tyr Phe Asp Ile Phe Arg Ser Glu
340 345 350
Lys Val Asn Asp Asn Tyr Val Ser Tyr Thr Lys Ser Ser Ile Ser Asn
355 360 365
Asn Lys Arg Val Lys Ala Asn Lys Phe Thr Asp Gln Glu Ala Phe Tyr

370 375 380

Lys Phe Ala Lys Lys His Leu Glu Thr Ile Lys Tyr Lys Ile Asn Lys

385 390 395 400

Val Asn Gly Ser Lys Ala Asp Leu Glu Leu Ile Asp Gly Met Leu Arg

405 410 415

Asp Met Glu Phe Lys Asn Phe Met Pro Lys Ile Lys Ser Ser Asp Asn

420 425 430

Gly Val Ile Pro Tyr Gln Leu Lys Leu Met Glu Leu Asn Lys Ile Leu

435 440 445

Glu Asn Gln Ser Lys His His Glu Phe Leu Asn Val Ser Asp Glu Tyr

450 455 460

Gly Ser Val Cys Asp Lys Ile Ala Ser Ile Met Glu Phe Arg Ile Pro

465 470 475 480

Tyr Tyr Val Gly Pro Leu Asn Pro Asn Ser Lys Tyr Ala Trp Ile Lys

485 490 495

Lys Gln Lys Asp Ser Glu Ile Thr Pro Trp Asn Phe Lys Asp Val Val

500 505 510

Asp Leu Asp Ser Ser Arg Glu Glu Phe Ile Asp Ser Leu Ile Gly Arg

515 520 525

Cys Thr Tyr Leu Lys Asp Glu Lys Val Leu Pro Lys Ala Ser Leu Leu

530 535 540

Tyr Asn Glu Tyr Met Val Leu Asn Glu Leu Asn Asn Leu Lys Leu Asn

545 550 555 560

Asp Leu Pro Ile Thr Glu Glu Met Lys Lys Lys Ile Phe Asp Gln Leu

565 570 575

Phe Lys Thr Arg Lys Lys Val Thr Leu Lys Ala Val Ala Asn Leu Leu

580 585 590

Lys Lys Glu Phe Asn Ile Asn Gly Glu Ile Leu Leu Ser Gly Thr Asp

595 600 605

Gly Asp Phe Lys Gln Gly Leu Asn Ser Tyr Asn Asp Phe Lys Ala Ile

610 615 620

Val Gly Asp Lys Val Asp Ser Asp Asp Tyr Arg Asp Lys Ile Glu Glu

625 630 635 640

Ile Ile Lys Leu Ile Val Leu Tyr Gly Asp Asp Lys Ser Tyr Leu Gln

645 650 655

Lys Lys Ile Lys Ala Gly Tyr Gly Lys Tyr Phe Thr Asp Ser Glu Ile

660 665 670

Lys Lys Met Ala Gly Leu Asn Tyr Lys Asp Trp Gly Arg Leu Ser Lys

675 680 685

Lys Leu Leu Thr Gly Leu Glu Gly Ala Asn Lys Ile Thr Gly Glu Arg

690 695 700

Gly Ser Ile Ile His Phe Met Arg Glu Tyr Asn Leu Asn Leu Met Glu

705 710 715 720

Leu Met Ser Ala Ser Phe Thr Phe Thr Glu Glu Ile Gln Lys Leu Asn

725 730 735

Pro Val Asp Asp Arg Lys Leu Ser Tyr Glu Met Val Asp Glu Leu Tyr

740 745 750

Leu Ser Pro Ser Val Lys Arg Met Leu Trp Gln Ser Leu Arg Ile Val

755 760 765

Asp Glu Ile Lys Asn Ile Met Gly Thr Asp Ser Lys Lys Ile Phe Ile

770 775 780

Glu Met Ala Arg Gly Lys Glu Glu Val Lys Ala Arg Lys Glu Ser Arg

785 790 795 800

Lys Asn Gln Leu Leu Lys Phe Tyr Lys Asp Gly Lys Lys Ala Phe Ile

805 810 815

Ser Glu Ile Gly Glu Glu Arg Tyr Ser Tyr Leu Leu Ser Glu Ile Glu

820 825 830

Gly Glu Glu Glu Asn Lys Phe Arg Trp Asp Asn Leu Tyr Leu Tyr Tyr

835 840 845

Thr Gln Leu Gly Arg Cys Met Tyr Ser Leu Glu Pro Ile Asp Ile Ser

850 855 860

Glu Leu Ser Ser Lys Asn Ile Tyr Asp Gln Asp His Ile Tyr Pro Lys

865 870 875 880

Ser Lys Ile Tyr Asp Asp Ser Ile Glu Asn Arg Val Leu Val Lys Lys

885 890 895

Asp Leu Asn Ser Lys Lys Gly Asn Ser Tyr Pro Ile Pro Asp Glu Ile

900 905 910

Leu Asn Lys Asn Cys Tyr Ala Tyr Trp Lys Ile Leu Tyr Asp Lys Gly

915 920 925

Leu Ile Gly Gln Lys Lys Tyr Thr Arg Leu Thr Arg Arg Thr Gly Phe

930 935 940

Thr Asp Asp Glu Leu Val Gln Phe Ile Ser Arg Gln Ile Val Glu Thr

945 950 955 960

Arg Gln Ala Thr Lys Glu Thr Ala Asn Leu Leu Lys Thr Ile Cys Lys

965 970 975

Asn Ser Glu Ile Val Tyr Ser Lys Ala Glu Asn Ala Ser Arg Phe Arg

980 985 990

Gln Glu Phe Asp Ile Val Lys Cys Arg Ala Val Asn Asp Leu His His

995 1000 1005

Met His Asp Ala Tyr Ile Asn Ile Ile Val Gly Asn Val Tyr Asn

1010 1015 1020

Thr Lys Phe Thr Lys Asp Pro Met Asn Phe Val Lys Lys Gln Glu

1025 1030 1035

Lys Ala Arg Ser Tyr Asn Leu Glu Asn Met Phe Lys Tyr Asp Val

1040 1045 1050

Lys Arg Gly Gly Tyr Thr Ala Trp Ile Ala Asp Asp Glu Lys Gly

1055 1060 1065

Thr Val Lys Asn Ala Ser Ile Lys Arg Ile Arg Lys Glu Leu Glu

1070 1075 1080

Gly Thr Asn Tyr Arg Phe Thr Arg Met Asn Tyr Ile Glu Ser Gly

1085 1090 1095

Ala Leu Phe Asn Ala Thr Leu Gln Arg Lys Asn Lys Gly Ser Arg

1100 1105 1110

Pro Leu Lys Asp Lys Gly Pro Lys Ser Ser Ile Glu Lys Tyr Gly
1115 1120 1125
Gly Tyr Thr Asn Ile Asn Lys Ala Cys Phe Ala Val Leu Asp Ile

1130 1135 1140
Lys Ser Lys Asn Lys Ile Glu Arg Lys Leu Met Pro Val Glu Arg
1145 1150 1155
Glu Ile Tyr Ala Lys Gln Lys Asn Asp Lys Lys Leu Ser Asp Glu
1160 1165 1170
Ile Phe Ser Lys Tyr Leu Lys Asp Arg Phe Gly Ile Glu Asp Tyr
1175 1180 1185
Arg Val Val Tyr Pro Val Val Lys Met Arg Thr Leu Leu Lys Ile

1190 1195 1200
Asp Gly Ser Tyr Tyr Phe Ile Thr Gly Gly Ser Asp Lys Thr Leu
1205 1210 1215
Glu Leu Arg Ser Ala Leu Gln Leu Ile Leu Pro Lys Lys Asn Glu
1220 1225 1230
Trp Ala Ile Lys Gln Ile Asp Lys Ser Ser Glu Asn Asp Tyr Leu
1235 1240 1245
Thr Ile Glu Arg Ile Gln Asp Leu Thr Glu Glu Leu Val Tyr Asn

1250 1255 1260
Thr Phe Asp Ile Ile Val Asn Lys Phe Lys Thr Ser Val Phe Lys
1265 1270 1275
Lys Ser Phe Leu Asn Leu Phe Gln Asp Asp Lys Ile Glu Asn Ile
1280 1285 1290
Asp Phe Lys Phe Lys Ser Met Asp Phe Lys Glu Lys Cys Lys Thr
1295 1300 1305
Leu Leu Met Leu Val Lys Ala Ile Arg Ala Ser Gly Val Arg Gln

1310 1315 1320
Asp Leu Lys Ser Ile Asp Leu Lys Ser Asp Tyr Gly Arg Leu Ser
1325 1330 1335
Ser Lys Thr Asn Asn Ile Gly Asn Tyr Gln Glu Phe Lys Ile Ile

1340 1345 1350

Asn Gln Ser Ile Thr Gly Leu Phe Glu Asn Glu Val Asp Leu Leu

1355 1360 1365

Lys Leu

1370

<210> 15

<211> 1424

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 15

Met Pro Lys Lys Lys Arg Lys Val Gly Thr Lys Val Lys Asp Tyr Tyr

1 5 10 15

Ile Gly Leu Asp Ile Gly Thr Ser Ser Val Gly Trp Ala Val Thr Asp

20 25 30

Glu Ala Tyr Asn Val Leu Lys Phe Asn Ser Lys Lys Met Trp Gly Val

35 40 45

Arg Leu Phe Asp Asp Ala Lys Thr Ala Glu Glu Arg Arg Gly Gln Arg

50 55 60

Gly Ala Arg Arg Arg Leu Asp Arg Lys Lys Glu Arg Leu Ser Leu Leu

65 70 75 80

Gln Asp Phe Phe Ala Glu Glu Val Ala Lys Val Asp Pro Asn Phe Phe

85 90 95

Leu Arg Leu Asp Asn Ser Asp Leu Tyr Met Glu Asp Lys Asp Gln Lys

100 105 110

Leu Lys Ser Lys Tyr Thr Leu Phe Asn Asp Lys Asp Phe Lys Asp Lys

115 120 125

Asn Phe His Lys Lys Tyr Pro Thr Ile His His Leu Leu Met Asp Leu

130 135 140

Ile Glu Asp Asp Ser Lys Lys Asp Ile Arg Leu Val Tyr Leu Ala Cys

145 150 155 160

His Tyr Leu Leu Lys Asn Arg Gly His Phe Ile Phe Glu Gly Gln Lys
165 170 175

Phe Asp Thr Lys Ser Ser Phe Glu Asn Ser Leu Asn Glu Leu Lys Val
180 185 190
His Leu Asn Asp Glu Tyr Gly Leu Asp Leu Glu Phe Asp Asn Glu Asn
195 200 205
Leu Ile Asn Ile Leu Thr Asp Pro Lys Leu Asn Lys Thr Ala Lys Lys
210 215 220
Lys Glu Leu Lys Ser Val Ile Gly Asp Thr Lys Phe Leu Lys Ala Val
225 230 235 240

Ser Ala Ile Met Ile Gly Ser Ser Gln Lys Leu Val Asp Leu Phe Glu
245 250 255
Asn Pro Glu Asp Phe Asp Asp Ser Ala Ile Lys Ser Val Asp Phe Ser
260 265 270
Thr Thr Ser Phe Asp Asp Lys Tyr Ser Asp Tyr Glu Leu Ala Leu Gly
275 280 285
Asp Lys Ile Ala Leu Val Asn Ile Leu Lys Glu Ile Tyr Asp Ser Ser
290 295 300

Ile Leu Glu Asn Leu Leu Lys Glu Ala Asp Lys Ser Lys Asp Gly Asn
305 310 315 320
Lys Tyr Ile Ser Asn Ala Phe Val Lys Lys Tyr Asn Lys His Gly Gln
325 330 335
Asp Leu Lys Glu Phe Lys Arg Leu Val Arg Gln Tyr His Lys Ser Ala
340 345 350
Tyr Phe Asp Ile Phe Arg Ser Glu Lys Val Asn Asp Asn Tyr Val Ser
355 360 365

Tyr Thr Lys Ser Ser Ile Ser Asn Asn Lys Arg Val Lys Ala Asn Lys
370 375 380
Phe Thr Asp Gln Glu Ala Phe Tyr Lys Phe Ala Lys Lys His Leu Glu
385 390 395 400
Thr Ile Lys Tyr Lys Ile Asn Lys Val Asn Gly Ser Lys Ala Asp Leu

405 410 415

Glu Leu Ile Asp Gly Met Leu Arg Asp Met Glu Phe Lys Asn Phe Met

420 425 430

Pro Lys Ile Lys Ser Ser Asp Asn Gly Val Ile Pro Tyr Gln Leu Lys

435 440 445

Leu Met Glu Leu Asn Lys Ile Leu Glu Asn Gln Ser Lys His His Glu

450 455 460

Phe Leu Asn Val Ser Asp Glu Tyr Gly Ser Val Cys Asp Lys Ile Ala

465 470 475 480

Ser Ile Met Glu Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Asn Pro

485 490 495

Asn Ser Lys Tyr Ala Trp Ile Lys Lys Gln Lys Asp Ser Glu Ile Thr

500 505 510

Pro Trp Asn Phe Lys Asp Val Val Asp Leu Asp Ser Ser Arg Glu Glu

515 520 525

Phe Ile Asp Ser Leu Ile Gly Arg Cys Thr Tyr Leu Lys Asp Glu Lys

530 535 540

Val Leu Pro Lys Ala Ser Leu Leu Tyr Asn Glu Tyr Met Val Leu Asn

545 550 555 560

Glu Leu Asn Asn Leu Lys Leu Asn Asp Leu Pro Ile Thr Glu Glu Met

565 570 575

Lys Lys Lys Ile Phe Asp Gln Leu Phe Lys Thr Arg Lys Lys Val Thr

580 585 590

Leu Lys Ala Val Ala Asn Leu Leu Lys Lys Glu Phe Asn Ile Asn Gly

595 600 605

Glu Ile Leu Leu Ser Gly Thr Asp Gly Asp Phe Lys Gln Gly Leu Asn

610 615 620

Ser Tyr Asn Asp Phe Lys Ala Ile Val Gly Asp Lys Val Asp Ser Asp

625 630 635 640

Asp Tyr Arg Asp Lys Ile Glu Glu Ile Ile Lys Leu Ile Val Leu Tyr

645 650 655

Gly Asp Asp Lys Ser Tyr Leu Gln Lys Lys Ile Lys Ala Gly Tyr Gly
660 665 670
Lys Tyr Phe Thr Asp Ser Glu Ile Lys Lys Met Ala Gly Leu Asn Tyr
675 680 685

Lys Asp Trp Gly Arg Leu Ser Lys Lys Leu Leu Thr Gly Leu Glu Gly
690 695 700
Ala Asn Lys Ile Thr Gly Glu Arg Gly Ser Ile Ile His Phe Met Arg
705 710 715 720
Glu Tyr Asn Leu Asn Leu Met Glu Leu Met Ser Ala Ser Phe Thr Phe
725 730 735
Thr Glu Glu Ile Gln Lys Leu Asn Pro Val Asp Asp Arg Lys Leu Ser
740 745 750

Tyr Glu Met Val Asp Glu Leu Tyr Leu Ser Pro Ser Val Lys Arg Met
755 760 765
Leu Trp Gln Ser Leu Arg Ile Val Asp Glu Ile Lys Asn Ile Met Gly
770 775 780
Thr Asp Ser Lys Lys Ile Phe Ile Glu Met Ala Arg Gly Lys Glu Glu
785 790 795 800
Val Lys Ala Arg Lys Glu Ser Arg Lys Asn Gln Leu Leu Lys Phe Tyr
805 810 815

Lys Asp Gly Lys Lys Ala Phe Ile Ser Glu Ile Gly Glu Glu Arg Tyr
820 825 830
Ser Tyr Leu Leu Ser Glu Ile Glu Gly Glu Glu Glu Asn Lys Phe Arg
835 840 845
Trp Asp Asn Leu Tyr Leu Tyr Tyr Thr Gln Leu Gly Arg Cys Met Tyr
850 855 860
Ser Leu Glu Pro Ile Asp Ile Ser Glu Leu Ser Ser Lys Asn Ile Tyr
865 870 875 880

Asp Gln Asp His Ile Tyr Pro Lys Ser Lys Ile Tyr Asp Asp Ser Ile
885 890 895
Glu Asn Arg Val Leu Val Lys Lys Asp Leu Asn Ser Lys Lys Gly Asn

900 905 910

Ser Tyr Pro Ile Pro Asp Glu Ile Leu Asn Lys Asn Cys Tyr Ala Tyr

915 920 925

Trp Lys Ile Leu Tyr Asp Lys Gly Leu Ile Gly Gln Lys Lys Tyr Thr

930 935 940

Arg Leu Thr Arg Arg Thr Gly Phe Thr Asp Asp Glu Leu Val Gln Phe

945 950 955 960

Ile Ser Arg Gln Ile Val Glu Thr Arg Gln Ala Thr Lys Glu Thr Ala

965 970 975

Asn Leu Leu Lys Thr Ile Cys Lys Asn Ser Glu Ile Val Tyr Ser Lys

980 985 990

Ala Glu Asn Ala Ser Arg Phe Arg Gln Glu Phe Asp Ile Val Lys Cys

995 1000 1005

Arg Ala Val Asn Asp Leu His His Met His Asp Ala Tyr Ile Asn

1010 1015 1020

Ile Ile Val Gly Asn Val Tyr Asn Thr Lys Phe Thr Lys Asp Pro

1025 1030 1035

Met Asn Phe Val Lys Lys Gln Glu Lys Ala Arg Ser Tyr Asn Leu

1040 1045 1050

Glu Asn Met Phe Lys Tyr Asp Val Lys Arg Gly Gly Tyr Thr Ala

1055 1060 1065

Trp Ile Ala Asp Asp Glu Lys Gly Thr Val Lys Asn Ala Ser Ile

1070 1075 1080

Lys Arg Ile Arg Lys Glu Leu Glu Gly Thr Asn Tyr Arg Phe Thr

1085 1090 1095

Arg Met Asn Tyr Ile Glu Ser Gly Ala Leu Phe Asn Ala Thr Leu

1100 1105 1110

Gln Arg Lys Asn Lys Gly Ser Arg Pro Leu Lys Asp Lys Gly Pro

1115 1120 1125

Lys Ser Ser Ile Glu Lys Tyr Gly Gly Tyr Thr Asn Ile Asn Lys

1130 1135 1140

Ala Cys Phe Ala Val Leu Asp Ile Lys Ser Lys Asn Lys Ile Glu
1145 1150 1155
Arg Lys Leu Met Pro Val Glu Arg Glu Ile Tyr Ala Lys Gln Lys
1160 1165 1170
Asn Asp Lys Lys Leu Ser Asp Glu Ile Phe Ser Lys Tyr Leu Lys
1175 1180 1185

Asp Arg Phe Gly Ile Glu Asp Tyr Arg Val Val Tyr Pro Val Val
1190 1195 1200
Lys Met Arg Thr Leu Leu Lys Ile Asp Gly Ser Tyr Tyr Phe Ile
1205 1210 1215
Thr Gly Gly Ser Asp Lys Thr Leu Glu Leu Arg Ser Ala Leu Gln
1220 1225 1230
Leu Ile Leu Pro Lys Lys Asn Glu Trp Ala Ile Lys Gln Ile Asp
1235 1240 1245

Lys Ser Ser Glu Asn Asp Tyr Leu Thr Ile Glu Arg Ile Gln Asp
1250 1255 1260
Leu Thr Glu Glu Leu Val Tyr Asn Thr Phe Asp Ile Ile Val Asn
1265 1270 1275
Lys Phe Lys Thr Ser Val Phe Lys Lys Ser Phe Leu Asn Leu Phe
1280 1285 1290
Gln Asp Asp Lys Ile Glu Asn Ile Asp Phe Lys Phe Lys Ser Met
1295 1300 1305

Asp Phe Lys Glu Lys Cys Lys Thr Leu Leu Met Leu Val Lys Ala
1310 1315 1320
Ile Arg Ala Ser Gly Val Arg Gln Asp Leu Lys Ser Ile Asp Leu
1325 1330 1335
Lys Ser Asp Tyr Gly Arg Leu Ser Ser Lys Thr Asn Asn Ile Gly
1340 1345 1350
Asn Tyr Gln Glu Phe Lys Ile Ile Asn Gln Ser Ile Thr Gly Leu
1355 1360 1365

Phe Glu Asn Glu Val Asp Leu Leu Lys Leu Gly Lys Arg Pro Ala

1370 1375 1380

Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Tyr

1385 1390 1395

Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp

1400 1405 1410

Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1415 1420

<210> 16

<211> 1620

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 16

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Ser Glu Val Glu Phe

1 5 10 15

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala

20 25 30

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn

35 40 45

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro

50 55 60

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met

65 70 75 80

Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro

85 90 95

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val

100 105 110

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met

115 120 125

Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu

130 135 140

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg
145 150 155 160
Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr
165 170 175

Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr
180 185 190
Pro Glu Ser Ser Gly Pro Lys Lys Lys Arg Lys Val Gly Thr Lys Val
195 200 205
Lys Asp Tyr Tyr Ile Gly Leu Ala Ile Gly Thr Ser Ser Val Gly Trp
210 215 220
Ala Val Thr Asp Glu Ala Tyr Asn Val Leu Lys Phe Asn Ser Lys Lys
225 230 235 240

Met Trp Gly Val Arg Leu Phe Asp Asp Ala Lys Thr Ala Glu Glu Arg
245 250 255
Arg Gly Gln Arg Gly Ala Arg Arg Arg Leu Asp Arg Lys Lys Glu Arg
260 265 270
Leu Ser Leu Leu Gln Asp Phe Phe Ala Glu Glu Val Ala Lys Val Asp
275 280 285
Pro Asn Phe Phe Leu Arg Leu Asp Asn Ser Asp Leu Tyr Met Glu Asp
290 295 300

Lys Asp Gln Lys Leu Lys Ser Lys Tyr Thr Leu Phe Asn Asp Lys Asp
305 310 315 320
Phe Lys Asp Lys Asn Phe His Lys Lys Tyr Pro Thr Ile His His Leu
325 330 335
Leu Met Asp Leu Ile Glu Asp Asp Ser Lys Lys Asp Ile Arg Leu Val
340 345 350
Tyr Leu Ala Cys His Tyr Leu Leu Lys Asn Arg Gly His Phe Ile Phe
355 360 365

Glu Gly Gln Lys Phe Asp Thr Lys Ser Ser Phe Glu Asn Ser Leu Asn
370 375 380
Glu Leu Lys Val His Leu Asn Asp Glu Tyr Gly Leu Asp Leu Glu Phe

385 390 395 400

Asp Asn Glu Asn Leu Ile Asn Ile Leu Thr Asp Pro Lys Leu Asn Lys

405 410 415

Thr Ala Lys Lys Lys Glu Leu Lys Ser Val Ile Gly Asp Thr Lys Phe

420 425 430

Leu Lys Ala Val Ser Ala Ile Met Ile Gly Ser Ser Gln Lys Leu Val

435 440 445

Asp Leu Phe Glu Asn Pro Glu Asp Phe Asp Asp Ser Ala Ile Lys Ser

450 455 460

Val Asp Phe Ser Thr Thr Ser Phe Asp Asp Lys Tyr Ser Asp Tyr Glu

465 470 475 480

Leu Ala Leu Gly Asp Lys Ile Ala Leu Val Asn Ile Leu Lys Glu Ile

485 490 495

Tyr Asp Ser Ser Ile Leu Glu Asn Leu Leu Lys Glu Ala Asp Lys Ser

500 505 510

Lys Asp Gly Asn Lys Tyr Ile Ser Asn Ala Phe Val Lys Lys Tyr Asn

515 520 525

Lys His Gly Gln Asp Leu Lys Glu Phe Lys Arg Leu Val Arg Gln Tyr

530 535 540

His Lys Ser Ala Tyr Phe Asp Ile Phe Arg Ser Glu Lys Val Asn Asp

545 550 555 560

Asn Tyr Val Ser Tyr Thr Lys Ser Ser Ile Ser Asn Asn Lys Arg Val

565 570 575

Lys Ala Asn Lys Phe Thr Asp Gln Glu Ala Phe Tyr Lys Phe Ala Lys

580 585 590

Lys His Leu Glu Thr Ile Lys Tyr Lys Ile Asn Lys Val Asn Gly Ser

595 600 605

Lys Ala Asp Leu Glu Leu Ile Asp Gly Met Leu Arg Asp Met Glu Phe

610 615 620

Lys Asn Phe Met Pro Lys Ile Lys Ser Ser Asp Asn Gly Val Ile Pro

625 630 635 640

Tyr Gln Leu Lys Leu Met Glu Leu Asn Lys Ile Leu Glu Asn Gln Ser
645 650 655
Lys His His Glu Phe Leu Asn Val Ser Asp Glu Tyr Gly Ser Val Cys
660 665 670
Asp Lys Ile Ala Ser Ile Met Glu Phe Arg Ile Pro Tyr Tyr Val Gly
675 680 685

Pro Leu Asn Pro Asn Ser Lys Tyr Ala Trp Ile Lys Lys Gln Lys Asp
690 695 700
Ser Glu Ile Thr Pro Trp Asn Phe Lys Asp Val Val Asp Leu Asp Ser
705 710 715 720
Ser Arg Glu Glu Phe Ile Asp Ser Leu Ile Gly Arg Cys Thr Tyr Leu
725 730 735
Lys Asp Glu Lys Val Leu Pro Lys Ala Ser Leu Leu Tyr Asn Glu Tyr
740 745 750

Met Val Leu Asn Glu Leu Asn Asn Leu Lys Leu Asn Asp Leu Pro Ile
755 760 765
Thr Glu Glu Met Lys Lys Lys Ile Phe Asp Gln Leu Phe Lys Thr Arg
770 775 780
Lys Lys Val Thr Leu Lys Ala Val Ala Asn Leu Leu Lys Lys Glu Phe
785 790 795 800
Asn Ile Asn Gly Glu Ile Leu Leu Ser Gly Thr Asp Gly Asp Phe Lys
805 810 815

Gln Gly Leu Asn Ser Tyr Asn Asp Phe Lys Ala Ile Val Gly Asp Lys
820 825 830
Val Asp Ser Asp Asp Tyr Arg Asp Lys Ile Glu Glu Ile Ile Lys Leu
835 840 845
Ile Val Leu Tyr Gly Asp Asp Lys Ser Tyr Leu Gln Lys Lys Ile Lys
850 855 860
Ala Gly Tyr Gly Lys Tyr Phe Thr Asp Ser Glu Ile Lys Lys Met Ala
865 870 875 880

Gly Leu Asn Tyr Lys Asp Trp Gly Arg Leu Ser Lys Lys Leu Leu Thr

885 890 895

Gly Leu Glu Gly Ala Asn Lys Ile Thr Gly Glu Arg Gly Ser Ile Ile

900 905 910

His Phe Met Arg Glu Tyr Asn Leu Asn Leu Met Glu Leu Met Ser Ala

915 920 925

Ser Phe Thr Phe Thr Glu Glu Ile Gln Lys Leu Asn Pro Val Asp Asp

930 935 940

Arg Lys Leu Ser Tyr Glu Met Val Asp Glu Leu Tyr Leu Ser Pro Ser

945 950 955 960

Val Lys Arg Met Leu Trp Gln Ser Leu Arg Ile Val Asp Glu Ile Lys

965 970 975

Asn Ile Met Gly Thr Asp Ser Lys Lys Ile Phe Ile Glu Met Ala Arg

980 985 990

Gly Lys Glu Glu Val Lys Ala Arg Lys Glu Ser Arg Lys Asn Gln Leu

995 1000 1005

Leu Lys Phe Tyr Lys Asp Gly Lys Lys Ala Phe Ile Ser Glu Ile

1010 1015 1020

Gly Glu Glu Arg Tyr Ser Tyr Leu Leu Ser Glu Ile Glu Gly Glu

1025 1030 1035

Glu Glu Asn Lys Phe Arg Trp Asp Asn Leu Tyr Leu Tyr Tyr Thr

1040 1045 1050

Gln Leu Gly Arg Cys Met Tyr Ser Leu Glu Pro Ile Asp Ile Ser

1055 1060 1065

Glu Leu Ser Ser Lys Asn Ile Tyr Asp Gln Asp His Ile Tyr Pro

1070 1075 1080

Lys Ser Lys Ile Tyr Asp Asp Ser Ile Glu Asn Arg Val Leu Val

1085 1090 1095

Lys Lys Asp Leu Asn Ser Lys Lys Gly Asn Ser Tyr Pro Ile Pro

1100 1105 1110

Asp Glu Ile Leu Asn Lys Asn Cys Tyr Ala Tyr Trp Lys Ile Leu

1115 1120 1125

Tyr Asp Lys Gly Leu Ile Gly Gln Lys Lys Tyr Thr Arg Leu Thr
1130 1135 1140
Arg Arg Thr Gly Phe Thr Asp Asp Glu Leu Val Gln Phe Ile Ser
1145 1150 1155
Arg Gln Ile Val Glu Thr Arg Gln Ala Thr Lys Glu Thr Ala Asn
1160 1165 1170
Leu Leu Lys Thr Ile Cys Lys Asn Ser Glu Ile Val Tyr Ser Lys
1175 1180 1185

Ala Glu Asn Ala Ser Arg Phe Arg Gln Glu Phe Asp Ile Val Lys
1190 1195 1200
Cys Arg Ala Val Asn Asp Leu His His Met His Asp Ala Tyr Ile
1205 1210 1215
Asn Ile Ile Val Gly Asn Val Tyr Asn Thr Lys Phe Thr Lys Asp
1220 1225 1230
Pro Met Asn Phe Val Lys Lys Gln Glu Lys Ala Arg Ser Tyr Asn
1235 1240 1245

Leu Glu Asn Met Phe Lys Tyr Asp Val Lys Arg Gly Gly Tyr Thr
1250 1255 1260
Ala Trp Ile Ala Asp Asp Glu Lys Gly Thr Val Lys Asn Ala Ser
1265 1270 1275
Ile Lys Arg Ile Arg Lys Glu Leu Glu Gly Thr Asn Tyr Arg Phe
1280 1285 1290
Thr Arg Met Asn Tyr Ile Glu Ser Gly Ala Leu Phe Asn Ala Thr
1295 1300 1305

Leu Gln Arg Lys Asn Lys Gly Ser Arg Pro Leu Lys Asp Lys Gly
1310 1315 1320
Pro Lys Ser Ser Ile Glu Lys Tyr Gly Gly Tyr Thr Asn Ile Asn
1325 1330 1335
Lys Ala Cys Phe Ala Val Leu Asp Ile Lys Ser Lys Asn Lys Ile
1340 1345 1350
Glu Arg Lys Leu Met Pro Val Glu Arg Glu Ile Tyr Ala Lys Gln

1355 1360 1365

Lys Asn Asp Lys Lys Leu Ser Asp Glu Ile Phe Ser Lys Tyr Leu
1370 1375 1380
Lys Asp Arg Phe Gly Ile Glu Asp Tyr Arg Val Val Tyr Pro Val
1385 1390 1395
Val Lys Met Arg Thr Leu Leu Lys Ile Asp Gly Ser Tyr Tyr Phe
1400 1405 1410
Ile Thr Gly Gly Ser Asp Lys Thr Leu Glu Leu Arg Ser Ala Leu
1415 1420 1425

Gln Leu Ile Leu Pro Lys Lys Asn Glu Trp Ala Ile Lys Gln Ile
1430 1435 1440
Asp Lys Ser Ser Glu Asn Asp Tyr Leu Thr Ile Glu Arg Ile Gln
1445 1450 1455
Asp Leu Thr Glu Glu Leu Val Tyr Asn Thr Phe Asp Ile Ile Val
1460 1465 1470
Asn Lys Phe Lys Thr Ser Val Phe Lys Lys Ser Phe Leu Asn Leu
1475 1480 1485

Phe Gln Asp Asp Lys Ile Glu Asn Ile Asp Phe Lys Phe Lys Ser
1490 1495 1500
Met Asp Phe Lys Glu Lys Cys Lys Thr Leu Leu Met Leu Val Lys
1505 1510 1515
Ala Ile Arg Ala Ser Gly Val Arg Gln Asp Leu Lys Ser Ile Asp
1520 1525 1530
Leu Lys Ser Asp Tyr Gly Arg Leu Ser Ser Lys Thr Asn Asn Ile
1535 1540 1545

Gly Asn Tyr Gln Glu Phe Lys Ile Ile Asn Gln Ser Ile Thr Gly
1550 1555 1560
Leu Phe Glu Asn Glu Val Asp Leu Leu Lys Leu Gly Lys Arg Pro
1565 1570 1575
Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser
1580 1585 1590

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro
1595 1600 1605

Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1610 1615 1620

<210> 17

<211> 1615

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 17

Met Pro Lys Lys Lys Arg Lys Val Thr Lys Val Lys Asp Tyr Tyr Ile
1 5 10 15
Gly Leu Ala Ile Gly Thr Ser Ser Val Gly Trp Ala Val Thr Asp Glu
20 25 30

Ala Tyr Asn Val Leu Lys Phe Asn Ser Lys Lys Met Trp Gly Val Arg
35 40 45
Leu Phe Asp Asp Ala Lys Thr Ala Glu Glu Arg Arg Gly Gln Arg Gly
50 55 60
Ala Arg Arg Arg Leu Asp Arg Lys Lys Glu Arg Leu Ser Leu Leu Gln
65 70 75 80
Asp Phe Phe Ala Glu Glu Val Ala Lys Val Asp Pro Asn Phe Phe Leu
85 90 95

Arg Leu Asp Asn Ser Asp Leu Tyr Met Glu Asp Lys Asp Gln Lys Leu
100 105 110
Lys Ser Lys Tyr Thr Leu Phe Asn Asp Lys Asp Phe Lys Asp Lys Asn
115 120 125
Phe His Lys Lys Tyr Pro Thr Ile His His Leu Leu Met Asp Leu Ile
130 135 140
Glu Asp Asp Ser Lys Lys Asp Ile Arg Leu Val Tyr Leu Ala Cys His
145 150 155 160

Tyr Leu Leu Lys Asn Arg Gly His Phe Ile Phe Glu Gly Gln Lys Phe
165 170 175
Asp Thr Lys Ser Ser Phe Glu Asn Ser Leu Asn Glu Leu Lys Val His
180 185 190
Leu Asn Asp Glu Tyr Gly Leu Asp Leu Glu Phe Asp Asn Glu Asn Leu
195 200 205
Ile Asn Ile Leu Thr Asp Pro Lys Leu Asn Lys Thr Ala Lys Lys Lys
210 215 220

Glu Leu Lys Ser Val Ile Gly Asp Thr Lys Phe Leu Lys Ala Val Ser
225 230 235 240
Ala Ile Met Ile Gly Ser Ser Gln Lys Leu Val Asp Leu Phe Glu Asn
245 250 255
Pro Glu Asp Phe Asp Asp Ser Ala Ile Lys Ser Val Asp Phe Ser Thr
260 265 270
Thr Ser Phe Asp Asp Lys Tyr Ser Asp Tyr Glu Leu Ala Leu Gly Asp
275 280 285

Lys Ile Ala Leu Val Asn Ile Leu Lys Glu Ile Tyr Asp Ser Ser Ile
290 295 300
Leu Glu Asn Leu Leu Lys Glu Ala Asp Lys Ser Lys Asp Gly Asn Lys
305 310 315 320
Tyr Ile Ser Asn Ala Phe Val Lys Lys Tyr Asn Lys His Gly Gln Asp
325 330 335
Leu Lys Glu Phe Lys Arg Leu Val Arg Gln Tyr His Lys Ser Ala Tyr
340 345 350

Phe Asp Ile Phe Arg Ser Glu Lys Val Asn Asp Asn Tyr Val Ser Tyr
355 360 365
Thr Lys Ser Ser Ile Ser Asn Asn Lys Arg Val Lys Ala Asn Lys Phe
370 375 380
Thr Asp Gln Glu Ala Phe Tyr Lys Phe Ala Lys Lys His Leu Glu Thr
385 390 395 400
Ile Lys Tyr Lys Ile Asn Lys Val Asn Gly Ser Lys Ala Asp Leu Glu

405 410 415

Leu Ile Asp Gly Met Leu Arg Asp Met Glu Phe Lys Asn Phe Met Pro
420 425 430
Lys Ile Lys Ser Ser Asp Asn Gly Val Ile Pro Tyr Gln Leu Lys Leu
435 440 445
Met Glu Leu Asn Lys Ile Leu Glu Asn Gln Ser Lys His His Glu Phe
450 455 460
Leu Asn Val Ser Asp Glu Tyr Gly Ser Val Cys Asp Lys Ile Ala Ser
465 470 475 480

Ile Met Glu Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Asn Pro Asn
485 490 495
Ser Lys Tyr Ala Trp Ile Lys Lys Gln Lys Asp Ser Glu Ile Thr Pro
500 505 510
Trp Asn Phe Lys Asp Val Val Asp Leu Asp Ser Ser Arg Glu Glu Phe
515 520 525
Ile Asp Ser Leu Ile Gly Arg Cys Thr Tyr Leu Lys Asp Glu Lys Val
530 535 540

Leu Pro Lys Ala Ser Leu Leu Tyr Asn Glu Tyr Met Val Leu Asn Glu
545 550 555 560
Leu Asn Asn Leu Lys Leu Asn Asp Leu Pro Ile Thr Glu Glu Met Lys
565 570 575
Lys Lys Ile Phe Asp Gln Leu Phe Lys Thr Arg Lys Lys Val Thr Leu
580 585 590
Lys Ala Val Ala Asn Leu Leu Lys Lys Glu Phe Asn Ile Asn Gly Glu
595 600 605

Ile Leu Leu Ser Gly Thr Asp Gly Asp Phe Lys Gln Gly Leu Asn Ser
610 615 620
Tyr Asn Asp Phe Lys Ala Ile Val Gly Asp Lys Val Asp Ser Asp Asp
625 630 635 640
Tyr Arg Asp Lys Ile Glu Glu Ile Ile Lys Leu Ile Val Leu Tyr Gly
645 650 655

Asp Asp Lys Ser Tyr Leu Gln Lys Lys Ile Lys Ala Gly Tyr Gly Lys
660 665 670

Tyr Phe Thr Asp Ser Glu Ile Lys Lys Met Ala Gly Leu Asn Tyr Lys
675 680 685

Asp Trp Gly Arg Leu Ser Lys Lys Leu Leu Thr Gly Leu Glu Gly Ala
690 695 700

Asn Lys Ile Thr Gly Glu Arg Gly Ser Ile Ile His Phe Met Arg Glu
705 710 715 720

Tyr Asn Leu Asn Leu Met Glu Leu Met Ser Ala Ser Phe Thr Phe Thr
725 730 735

Glu Glu Ile Gln Lys Leu Asn Pro Val Asp Asp Arg Lys Leu Ser Tyr
740 745 750

Glu Met Val Asp Glu Leu Tyr Leu Ser Pro Ser Val Lys Arg Met Leu
755 760 765

Trp Gln Ser Leu Arg Ile Val Asp Glu Ile Lys Asn Ile Met Gly Thr
770 775 780

Asp Ser Lys Lys Ile Phe Ile Glu Met Ala Arg Gly Lys Glu Glu Val
785 790 795 800

Lys Ala Arg Lys Glu Ser Arg Lys Asn Gln Leu Leu Lys Phe Tyr Lys
805 810 815

Asp Gly Lys Lys Ala Phe Ile Ser Glu Ile Gly Glu Glu Arg Tyr Ser
820 825 830

Tyr Leu Leu Ser Glu Ile Glu Gly Glu Glu Glu Asn Lys Phe Arg Trp
835 840 845

Asp Asn Leu Tyr Leu Tyr Tyr Thr Gln Leu Gly Arg Cys Met Tyr Ser
850 855 860

Leu Glu Pro Ile Asp Ile Ser Glu Leu Ser Ser Lys Asn Ile Tyr Asp
865 870 875 880

Gln Asp His Ile Tyr Pro Lys Ser Lys Ile Tyr Asp Asp Ser Ile Glu
885 890 895

Asn Arg Val Leu Val Lys Lys Asp Leu Asn Ser Lys Lys Gly Asn Ser

900 905 910

Tyr Pro Ile Pro Asp Glu Ile Leu Asn Lys Asn Cys Tyr Ala Tyr Trp

915 920 925

Lys Ile Leu Tyr Asp Lys Gly Leu Ile Gly Gln Lys Lys Tyr Thr Arg

930 935 940

Leu Thr Arg Arg Thr Gly Phe Thr Asp Asp Glu Leu Val Gln Phe Ile

945 950 955 960

Ser Arg Gln Ile Val Glu Thr Arg Gln Ala Thr Lys Glu Thr Ala Asn

965 970 975

Leu Leu Lys Thr Ile Cys Lys Asn Ser Glu Ile Val Tyr Ser Lys Ala

980 985 990

Glu Asn Ala Ser Arg Phe Arg Gln Glu Phe Asp Ile Val Lys Cys Arg

995 1000 1005

Ala Val Asn Asp Leu His His Met His Asp Ala Tyr Ile Asn Ile

1010 1015 1020

Ile Val Gly Asn Val Tyr Asn Thr Lys Phe Thr Lys Asp Pro Met

1025 1030 1035

Asn Phe Val Lys Lys Gln Glu Lys Ala Arg Ser Tyr Asn Leu Glu

1040 1045 1050

Asn Met Phe Lys Tyr Asp Val Lys Arg Gly Gly Tyr Thr Ala Trp

1055 1060 1065

Ile Ala Asp Asp Glu Lys Gly Thr Val Lys Asn Ala Ser Ile Lys

1070 1075 1080

Arg Ile Arg Lys Glu Leu Glu Gly Thr Asn Tyr Arg Phe Thr Arg

1085 1090 1095

Met Asn Tyr Ile Glu Ser Gly Ala Leu Phe Asn Ala Thr Leu Gln

1100 1105 1110

Arg Lys Asn Lys Gly Ser Arg Pro Leu Lys Asp Lys Gly Pro Lys

1115 1120 1125

Ser Ser Ile Glu Lys Tyr Gly Gly Tyr Thr Asn Ile Asn Lys Ala

1130 1135 1140

Cys Phe Ala Val Leu Asp Ile Lys Ser Lys Asn Lys Ile Glu Arg
1145 1150 1155
Lys Leu Met Pro Val Glu Arg Glu Ile Tyr Ala Lys Gln Lys Asn
1160 1165 1170

Asp Lys Lys Leu Ser Asp Glu Ile Phe Ser Lys Tyr Leu Lys Asp
1175 1180 1185
Arg Phe Gly Ile Glu Asp Tyr Arg Val Val Tyr Pro Val Val Lys
1190 1195 1200
Met Arg Thr Leu Leu Lys Ile Asp Gly Ser Tyr Tyr Phe Ile Thr
1205 1210 1215
Gly Gly Ser Asp Lys Thr Leu Glu Leu Arg Ser Ala Leu Gln Leu
1220 1225 1230

Ile Leu Pro Lys Lys Asn Glu Trp Ala Ile Lys Gln Ile Asp Lys
1235 1240 1245
Ser Ser Glu Asn Asp Tyr Leu Thr Ile Glu Arg Ile Gln Asp Leu
1250 1255 1260
Thr Glu Glu Leu Val Tyr Asn Thr Phe Asp Ile Ile Val Asn Lys
1265 1270 1275
Phe Lys Thr Ser Val Phe Lys Lys Ser Phe Leu Asn Leu Phe Gln
1280 1285 1290

Asp Asp Lys Ile Glu Asn Ile Asp Phe Lys Phe Lys Ser Met Asp
1295 1300 1305
Phe Lys Glu Lys Cys Lys Thr Leu Leu Met Leu Val Lys Ala Ile
1310 1315 1320
Arg Ala Ser Gly Val Arg Gln Asp Leu Lys Ser Ile Asp Leu Lys
1325 1330 1335
Ser Asp Tyr Gly Arg Leu Ser Ser Lys Thr Asn Asn Ile Gly Asn
1340 1345 1350

Tyr Gln Glu Phe Lys Ile Ile Asn Gln Ser Ile Thr Gly Leu Phe
1355 1360 1365
Glu Asn Glu Val Asp Leu Leu Lys Leu Lys Arg Pro Ala Ala Thr

1370 1375 1380

Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Ser Gly Ser Glu Thr

1385 1390 1395

Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Ser Glu

1400 1405 1410

Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu

1415 1420 1425

Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val

1430 1435 1440

Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

1445 1450 1455

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu

1460 1465 1470

Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala

1475 1480 1485

Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala

1490 1495 1500

Met Ile His Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn

1505 1510 1515

Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp Val Leu His His

1520 1525 1530

Pro Gly Met Asn His Arg Val Glu Ile Thr Glu Gly Ile Leu Ala

1535 1540 1545

Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg Met Pro Arg

1550 1555 1560

Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr Asp Pro

1565 1570 1575

Ala Ala Lys Arg Val Lys Leu Asp Gly Ser Tyr Pro Tyr Asp Val

1580 1585 1590

Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro

1595 1600 1605

Tyr Asp Val Pro Asp Tyr Ala

1610 1615

<210> 18

<211> 1880

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 18

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Thr Ser Glu Lys Gly Pro
1 5 10 15
Ser Thr Gly Asp Pro Thr Leu Arg Arg Arg Ile Glu Ser Trp Glu Phe

20 25 30
Asp Val Phe Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu
35 40 45
Tyr Glu Ile Lys Trp Gly Met Ser Arg Lys Ile Trp Arg Ser Ser Gly
50 55 60
Lys Asn Thr Thr Asn His Val Glu Val Asn Phe Ile Lys Lys Phe Thr
65 70 75 80
Ser Glu Arg Arg Phe His Ser Ser Ile Ser Cys Ser Ile Thr Trp Phe

85 90 95
Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser Gln Ala Ile Arg Glu Phe
100 105 110
Leu Ser Gln His Pro Gly Val Thr Leu Val Ile Tyr Val Ala Arg Leu
115 120 125
Phe Trp His Met Asp Gln Arg Asn Arg Gln Gly Leu Arg Asp Leu Val
130 135 140
Asn Ser Gly Val Thr Ile Gln Ile Met Arg Ala Ser Glu Tyr Tyr His

145 150 155 160
Cys Trp Arg Asn Phe Val Asn Tyr Pro Pro Gly Asp Glu Ala His Trp
165 170 175
Pro Gln Tyr Pro Pro Leu Trp Met Met Leu Tyr Ala Leu Glu Leu His

180 185 190

Cys Ile Ile Leu Ser Leu Pro Pro Cys Leu Lys Ile Ser Arg Arg Trp

195 200 205

Gln Asn His Leu Ala Phe Phe Arg Leu His Leu Gln Asn Cys His Tyr

210 215 220

Gln Thr Ile Pro Pro His Ile Leu Leu Ala Thr Gly Leu Ile His Pro

225 230 235 240

Ser Val Thr Trp Arg Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser

245 250 255

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

260 265 270

Gly Gly Ser Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val Gly Thr

275 280 285

Lys Val Lys Asp Tyr Tyr Ile Gly Leu Ala Ile Gly Thr Ser Ser Val

290 295 300

Gly Trp Ala Val Thr Asp Glu Ala Tyr Asn Val Leu Lys Phe Asn Ser

305 310 315 320

Lys Lys Met Trp Gly Val Arg Leu Phe Asp Asp Ala Lys Thr Ala Glu

325 330 335

Glu Arg Arg Gly Gln Arg Gly Ala Arg Arg Arg Leu Asp Arg Lys Lys

340 345 350

Glu Arg Leu Ser Leu Leu Gln Asp Phe Phe Ala Glu Glu Val Ala Lys

355 360 365

Val Asp Pro Asn Phe Phe Leu Arg Leu Asp Asn Ser Asp Leu Tyr Met

370 375 380

Glu Asp Lys Asp Gln Lys Leu Lys Ser Lys Tyr Thr Leu Phe Asn Asp

385 390 395 400

Lys Asp Phe Lys Asp Lys Asn Phe His Lys Lys Tyr Pro Thr Ile His

405 410 415

His Leu Leu Met Asp Leu Ile Glu Asp Asp Ser Lys Lys Asp Ile Arg

420 425 430

Leu Val Tyr Leu Ala Cys His Tyr Leu Leu Lys Asn Arg Gly His Phe
435 440 445
Ile Phe Glu Gly Gln Lys Phe Asp Thr Lys Ser Ser Phe Glu Asn Ser
450 455 460
Leu Asn Glu Leu Lys Val His Leu Asn Asp Glu Tyr Gly Leu Asp Leu

465 470 475 480
Glu Phe Asp Asn Glu Asn Leu Ile Asn Ile Leu Thr Asp Pro Lys Leu
485 490 495
Asn Lys Thr Ala Lys Lys Lys Glu Leu Lys Ser Val Ile Gly Asp Thr
500 505 510
Lys Phe Leu Lys Ala Val Ser Ala Ile Met Ile Gly Ser Ser Gln Lys
515 520 525
Leu Val Asp Leu Phe Glu Asn Pro Glu Asp Phe Asp Asp Ser Ala Ile

530 535 540
Lys Ser Val Asp Phe Ser Thr Thr Ser Phe Asp Asp Lys Tyr Ser Asp
545 550 555 560
Tyr Glu Leu Ala Leu Gly Asp Lys Ile Ala Leu Val Asn Ile Leu Lys
565 570 575
Glu Ile Tyr Asp Ser Ser Ile Leu Glu Asn Leu Leu Lys Glu Ala Asp
580 585 590
Lys Ser Lys Asp Gly Asn Lys Tyr Ile Ser Asn Ala Phe Val Lys Lys

595 600 605
Tyr Asn Lys His Gly Gln Asp Leu Lys Glu Phe Lys Arg Leu Val Arg
610 615 620
Gln Tyr His Lys Ser Ala Tyr Phe Asp Ile Phe Arg Ser Glu Lys Val
625 630 635 640
Asn Asp Asn Tyr Val Ser Tyr Thr Lys Ser Ser Ile Ser Asn Asn Lys
645 650 655
Arg Val Lys Ala Asn Lys Phe Thr Asp Gln Glu Ala Phe Tyr Lys Phe

660 665 670
Ala Lys Lys His Leu Glu Thr Ile Lys Tyr Lys Ile Asn Lys Val Asn

675 680 685

Gly Ser Lys Ala Asp Leu Glu Leu Ile Asp Gly Met Leu Arg Asp Met

690 695 700

Glu Phe Lys Asn Phe Met Pro Lys Ile Lys Ser Ser Asp Asn Gly Val

705 710 715 720

Ile Pro Tyr Gln Leu Lys Leu Met Glu Leu Asn Lys Ile Leu Glu Asn

725 730 735

Gln Ser Lys His His Glu Phe Leu Asn Val Ser Asp Glu Tyr Gly Ser

740 745 750

Val Cys Asp Lys Ile Ala Ser Ile Met Glu Phe Arg Ile Pro Tyr Tyr

755 760 765

Val Gly Pro Leu Asn Pro Asn Ser Lys Tyr Ala Trp Ile Lys Lys Gln

770 775 780

Lys Asp Ser Glu Ile Thr Pro Trp Asn Phe Lys Asp Val Val Asp Leu

785 790 795 800

Asp Ser Ser Arg Glu Glu Phe Ile Asp Ser Leu Ile Gly Arg Cys Thr

805 810 815

Tyr Leu Lys Asp Glu Lys Val Leu Pro Lys Ala Ser Leu Leu Tyr Asn

820 825 830

Glu Tyr Met Val Leu Asn Glu Leu Asn Asn Leu Lys Leu Asn Asp Leu

835 840 845

Pro Ile Thr Glu Glu Met Lys Lys Lys Ile Phe Asp Gln Leu Phe Lys

850 855 860

Thr Arg Lys Lys Val Thr Leu Lys Ala Val Ala Asn Leu Leu Lys Lys

865 870 875 880

Glu Phe Asn Ile Asn Gly Glu Ile Leu Leu Ser Gly Thr Asp Gly Asp

885 890 895

Phe Lys Gln Gly Leu Asn Ser Tyr Asn Asp Phe Lys Ala Ile Val Gly

900 905 910

Asp Lys Val Asp Ser Asp Asp Tyr Arg Asp Lys Ile Glu Glu Ile Ile

915 920 925

Lys Leu Ile Val Leu Tyr Gly Asp Asp Lys Ser Tyr Leu Gln Lys Lys
930 935 940
Ile Lys Ala Gly Tyr Gly Lys Tyr Phe Thr Asp Ser Glu Ile Lys Lys
945 950 955 960
Met Ala Gly Leu Asn Tyr Lys Asp Trp Gly Arg Leu Ser Lys Lys Leu
965 970 975
Leu Thr Gly Leu Glu Gly Ala Asn Lys Ile Thr Gly Glu Arg Gly Ser

980 985 990
Ile Ile His Phe Met Arg Glu Tyr Asn Leu Asn Leu Met Glu Leu Met
995 1000 1005
Ser Ala Ser Phe Thr Phe Thr Glu Glu Ile Gln Lys Leu Asn Pro
1010 1015 1020
Val Asp Asp Arg Lys Leu Ser Tyr Glu Met Val Asp Glu Leu Tyr
1025 1030 1035
Leu Ser Pro Ser Val Lys Arg Met Leu Trp Gln Ser Leu Arg Ile

1040 1045 1050
Val Asp Glu Ile Lys Asn Ile Met Gly Thr Asp Ser Lys Lys Ile
1055 1060 1065
Phe Ile Glu Met Ala Arg Gly Lys Glu Glu Val Lys Ala Arg Lys
1070 1075 1080
Glu Ser Arg Lys Asn Gln Leu Leu Lys Phe Tyr Lys Asp Gly Lys
1085 1090 1095
Lys Ala Phe Ile Ser Glu Ile Gly Glu Glu Arg Tyr Ser Tyr Leu

1100 1105 1110
Leu Ser Glu Ile Glu Gly Glu Glu Glu Asn Lys Phe Arg Trp Asp
1115 1120 1125
Asn Leu Tyr Leu Tyr Tyr Thr Gln Leu Gly Arg Cys Met Tyr Ser
1130 1135 1140
Leu Glu Pro Ile Asp Ile Ser Glu Leu Ser Ser Lys Asn Ile Tyr
1145 1150 1155
Asp Gln Asp His Ile Tyr Pro Lys Ser Lys Ile Tyr Asp Asp Ser

1160 1165 1170

Ile Glu Asn Arg Val Leu Val Lys Lys Asp Leu Asn Ser Lys Lys

1175 1180 1185

Gly Asn Ser Tyr Pro Ile Pro Asp Glu Ile Leu Asn Lys Asn Cys

1190 1195 1200

Tyr Ala Tyr Trp Lys Ile Leu Tyr Asp Lys Gly Leu Ile Gly Gln

1205 1210 1215

Lys Lys Tyr Thr Arg Leu Thr Arg Arg Thr Gly Phe Thr Asp Asp

1220 1225 1230

Glu Leu Val Gln Phe Ile Ser Arg Gln Ile Val Glu Thr Arg Gln

1235 1240 1245

Ala Thr Lys Glu Thr Ala Asn Leu Leu Lys Thr Ile Cys Lys Asn

1250 1255 1260

Ser Glu Ile Val Tyr Ser Lys Ala Glu Asn Ala Ser Arg Phe Arg

1265 1270 1275

Gln Glu Phe Asp Ile Val Lys Cys Arg Ala Val Asn Asp Leu His

1280 1285 1290

His Met His Asp Ala Tyr Ile Asn Ile Ile Val Gly Asn Val Tyr

1295 1300 1305

Asn Thr Lys Phe Thr Lys Asp Pro Met Asn Phe Val Lys Lys Gln

1310 1315 1320

Glu Lys Ala Arg Ser Tyr Asn Leu Glu Asn Met Phe Lys Tyr Asp

1325 1330 1335

Val Lys Arg Gly Gly Tyr Thr Ala Trp Ile Ala Asp Asp Glu Lys

1340 1345 1350

Gly Thr Val Lys Asn Ala Ser Ile Lys Arg Ile Arg Lys Glu Leu

1355 1360 1365

Glu Gly Thr Asn Tyr Arg Phe Thr Arg Met Asn Tyr Ile Glu Ser

1370 1375 1380

Gly Ala Leu Phe Asn Ala Thr Leu Gln Arg Lys Asn Lys Gly Ser

1385 1390 1395

Arg Pro Leu Lys Asp Lys Gly Pro Lys Ser Ser Ile Glu Lys Tyr

1400 1405 1410

Gly Gly Tyr Thr Asn Ile Asn Lys Ala Cys Phe Ala Val Leu Asp

1415 1420 1425

Ile Lys Ser Lys Asn Lys Ile Glu Arg Lys Leu Met Pro Val Glu

1430 1435 1440

Arg Glu Ile Tyr Ala Lys Gln Lys Asn Asp Lys Lys Leu Ser Asp

1445 1450 1455

Glu Ile Phe Ser Lys Tyr Leu Lys Asp Arg Phe Gly Ile Glu Asp

1460 1465 1470

Tyr Arg Val Val Tyr Pro Val Val Lys Met Arg Thr Leu Leu Lys

1475 1480 1485

Ile Asp Gly Ser Tyr Tyr Phe Ile Thr Gly Gly Ser Asp Lys Thr

1490 1495 1500

Leu Glu Leu Arg Ser Ala Leu Gln Leu Ile Leu Pro Lys Lys Asn

1505 1510 1515

Glu Trp Ala Ile Lys Gln Ile Asp Lys Ser Ser Glu Asn Asp Tyr

1520 1525 1530

Leu Thr Ile Glu Arg Ile Gln Asp Leu Thr Glu Glu Leu Val Tyr

1535 1540 1545

Asn Thr Phe Asp Ile Ile Val Asn Lys Phe Lys Thr Ser Val Phe

1550 1555 1560

Lys Lys Ser Phe Leu Asn Leu Phe Gln Asp Asp Lys Ile Glu Asn

1565 1570 1575

Ile Asp Phe Lys Phe Lys Ser Met Asp Phe Lys Glu Lys Cys Lys

1580 1585 1590

Thr Leu Leu Met Leu Val Lys Ala Ile Arg Ala Ser Gly Val Arg

1595 1600 1605

Gln Asp Leu Lys Ser Ile Asp Leu Lys Ser Asp Tyr Gly Arg Leu

1610 1615 1620

Ser Ser Lys Thr Asn Asn Ile Gly Asn Tyr Gln Glu Phe Lys Ile

1625 1630 1635

Ile Asn Gln Ser Ile Thr Gly Leu Phe Glu Asn Glu Val Asp Leu

1640 1645 1650

Leu Lys Leu Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln

1655 1660 1665

Ala Lys Lys Lys Lys Gly Ser Ser Gly Gly Ser Gly Gly Ser Gly

1670 1675 1680

Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys

1685 1690 1695

Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val

1700 1705 1710

Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His

1715 1720 1725

Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr

1730 1735 1740

Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp

1745 1750 1755

Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Gly

1760 1765 1770

Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys Glu

1775 1780 1785

Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro

1790 1795 1800

Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile

1805 1810 1815

Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met

1820 1825 1830

Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val

1835 1840 1845

Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Tyr Pro

1850 1855 1860

Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr
1865 1870 1875

Ala Tyr
1880

<210> 19

<211> 112

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 19

guuugagagu uauguaauug aaaaauuaca ugacgaguuc aaauaaaaau uuauucaaac 60

cgccuauuua uaggccgcag auguucugca uuaugcuugc uauugcaagc uu 112

<210> 20

<211> 1595

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 20

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Ser Glu Val Glu Phe
1 5 10 15

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala
20 25 30

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn
35 40 45

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro
50 55 60

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met
65 70 75 80

Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro
85 90 95

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
100 105 110
Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
115 120 125
Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
130 135 140

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg
145 150 155 160
Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr
165 170 175
Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr
180 185 190
Pro Glu Ser Ser Gly Pro Lys Lys Lys Arg Lys Val Gly Gly Lys Pro
195 200 205

Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp Ala Val
210 215 220
Val Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys Met Lys Val Leu Gly
225 230 235 240
Asn Thr Asp Lys Gln Ser Ile Lys Lys Asn Leu Leu Gly Ala Leu Leu
245 250 255
Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala
260 265 270

Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Leu Arg Tyr Leu Gln Glu
275 280 285
Ile Phe Thr Gly Glu Met Asn Lys Val Asp Glu Asn Phe Phe Gln Arg
290 295 300
Leu Asp Asp Ser Phe Leu Val Asp Glu Asp Lys Arg Gly Glu His His
305 310 315 320
Pro Ile Phe Gly Asn Ile Ala Ala Glu Val Lys Tyr His Asp Asp Phe
325 330 335

Pro Thr Ile Tyr His Leu Arg Arg His Leu Ala Asp Thr Ser Lys Lys

340 345 350

Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His Met Ile Lys Phe

355 360 365

Arg Gly His Phe Leu Tyr Glu Gly Asp Leu Lys Ala Glu Asn Thr Asp

370 375 380

Val Gln Ala Leu Phe Lys Asp Phe Val Glu Glu Tyr Asp Lys Thr Ile

385 390 395 400

Glu Glu Ser His Leu Ser Glu Ile Thr Val Asp Ala Leu Ser Ile Leu

405 410 415

Thr Glu Lys Val Ser Lys Ser Ser Arg Leu Glu Asn Leu Ile Ala His

420 425 430

Tyr Pro Thr Glu Lys Lys Asn Thr Leu Phe Gly Asn Leu Ile Ala Leu

435 440 445

Ser Leu Asp Leu His Pro Asn Phe Lys Thr Asn Phe Gln Leu Ser Glu

450 455 460

Asp Ala Lys Leu Gln Phe Ser Lys Asp Thr Tyr Glu Glu Asp Leu Glu

465 470 475 480

Gly Phe Leu Gly Glu Val Gly Asp Glu Tyr Ala Asp Leu Phe Ala Ser

485 490 495

Ala Lys Asn Leu Tyr Asp Ala Ile Leu Leu Ser Gly Ile Leu Thr Val

500 505 510

Asp Asp Asn Ser Thr Lys Ala Pro Leu Ser Ala Ser Met Val Lys Arg

515 520 525

Tyr Glu Glu His Gln Lys Asp Leu Lys Lys Leu Lys Asp Phe Ile Lys

530 535 540

Val Asn Ala Pro Asp Gln Tyr Asn Ala Ile Phe Lys Asp Lys Asn Lys

545 550 555 560

Lys Gly Tyr Ala Ser Tyr Ile Glu Ser Gly Val Lys Gln Asp Glu Phe

565 570 575

Tyr Lys Tyr Leu Lys Gly Ile Leu Leu Lys Ile Asn Gly Ser Gly Asp

580 585 590

Phe Leu Asp Lys Ile Asp Arg Glu Asp Phe Leu Arg Lys Gln Arg Thr
595 600 605

Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gln Glu Met His
610 615 620

Ala Ile Leu Arg Arg Gln Gly Glu His Tyr Pro Phe Leu Lys Glu Asn
625 630 635 640

Gln Asp Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val
645 650 655

Gly Pro Leu Ala Arg Lys Gly Ser Arg Phe Ala Trp Ala Glu Tyr Lys
660 665 670

Ala Asp Glu Lys Ile Thr Pro Trp Asn Phe Asp Asp Ile Leu Asp Lys
675 680 685

Glu Lys Ser Ala Glu Lys Phe Ile Thr Arg Met Thr Leu Asn Asp Leu
690 695 700

Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser Pro Leu Tyr Glu
705 710 715 720

Ala Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Asn Glu
725 730 735

Gln Gly Glu Ala Lys Phe Phe Asp Thr Asn Met Lys Gln Glu Ile Phe
740 745 750

Asp His Val Phe Lys Glu Asn Arg Lys Val Thr Lys Asp Lys Leu Leu
755 760 765

Asn Tyr Leu Asn Lys Glu Phe Glu Glu Phe Arg Ile Val Asn Leu Thr
770 775 780

Gly Leu Asp Lys Glu Asn Lys Ala Phe Asn Ser Ser Leu Gly Thr Tyr
785 790 795 800

His Asp Leu Arg Lys Ile Leu Asp Lys Ser Phe Leu Asp Asp Lys Ala
805 810 815

Asn Glu Lys Thr Ile Glu Asp Ile Ile Gln Thr Leu Thr Leu Phe Glu
820 825 830

Asp Arg Glu Met Ile Arg Gln Arg Leu Gln Lys Tyr Ser Asp Ile Phe

835 840 845

Thr Lys Ala Gln Leu Lys Lys Leu Glu Arg Arg His Tyr Thr Gly Trp
850 855 860
Gly Arg Leu Ser Tyr Lys Leu Ile Asn Gly Ile Arg Asn Lys Glu Asn
865 870 875 880
Lys Lys Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Tyr Ala Asn Arg
885 890 895
Asn Phe Met Gln Leu Ile Asn Asp Asp Ala Leu Ser Phe Lys Glu Glu
900 905 910

Ile Ala Arg Ala Gln Ile Ile Asp Asp Val Asp Asp Ile Ala Asn Val
915 920 925
Val His Asp Leu Pro Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln
930 935 940
Ser Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly His Asn Pro
945 950 955 960
Ala Asn Ile Ile Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Asp Lys
965 970 975

Gly Arg Arg Asn Ser Gln Gln Arg Leu Lys Leu Leu Gln Asp Ser Leu
980 985 990
Lys Asn Leu Asp Asn Pro Val Asn Ile Lys Asn Val Glu Asn Gln Gln
995 1000 1005
Leu Gln Asn Asp Arg Leu Phe Leu Tyr Tyr Ile Gln Asn Gly Lys
1010 1015 1020
Asp Met Tyr Thr Gly Glu Thr Leu Asp Ile Asn Asn Leu Ser Gln
1025 1030 1035

Tyr Asp Ile Asp His Ile Ile Pro Gln Ala Phe Ile Lys Asp Asn
1040 1045 1050
Ser Leu Asp Asn Arg Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
1055 1060 1065
Lys Ser Asp Asp Val Pro Ser Ile Glu Val Val His Glu Met Lys
1070 1075 1080

Ser Phe Trp Ser Lys Leu Leu Ser Val Lys Leu Ile Thr Gln Arg
1085 1090 1095

Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Thr Glu
1100 1105 1110
Glu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
1115 1120 1125
Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Glu Arg Phe Asn
1130 1135 1140
Thr Glu Phe Asp Gly Asn Lys Arg Arg Ile Arg Asn Val Lys Ile
1145 1150 1155

Ile Thr Leu Lys Ser Asn Leu Val Ser Asn Phe Arg Lys Glu Phe
1160 1165 1170
Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Tyr His His Ala His
1175 1180 1185
Asp Ala Tyr Leu Asn Ala Val Val Gly Asn Ala Leu Leu Leu Lys
1190 1195 1200
Tyr Pro Gln Leu Glu Pro Glu Phe Val Tyr Gly Glu Tyr Pro Lys
1205 1210 1215

Tyr Asn Ser Tyr Arg Ser Arg Lys Ser Ala Thr Glu Lys Phe Leu
1220 1225 1230
Phe Tyr Ser Asn Ile Leu Arg Phe Phe Lys Lys Glu Asp Ile Gln
1235 1240 1245
Thr Asn Glu Asp Gly Glu Ile Ala Trp Asn Lys Glu Lys His Ile
1250 1255 1260
Lys Ile Leu Arg Lys Val Leu Ser Tyr Pro Gln Val Asn Ile Val
1265 1270 1275

Lys Lys Thr Glu Glu Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile
1280 1285 1290
Leu Pro Lys Gly Glu Ser Asp Lys Leu Ile Pro Arg Lys Thr Lys
1295 1300 1305
Asn Ser Tyr Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro

1310 1315 1320

Val Val Ala Tyr Ser Ile Leu Val Phe Ala Asp Val Glu Lys Gly

1325 1330 1335

Lys Ser Lys Lys Leu Arg Lys Val Gln Asp Met Val Gly Ile Thr

1340 1345 1350

Ile Met Glu Lys Lys Arg Phe Glu Lys Asn Pro Val Asp Phe Leu

1355 1360 1365

Glu Gln Arg Gly Tyr Arg Asn Val Arg Leu Glu Lys Ile Ile Lys

1370 1375 1380

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Lys Arg Arg Arg

1385 1390 1395

Leu Leu Ala Ser Ala Lys Glu Leu Gln Lys Gly Asn Glu Leu Val

1400 1405 1410

Ile Pro Gln Arg Phe Thr Thr Leu Leu Tyr His Ser Tyr Arg Ile

1415 1420 1425

Glu Lys Asp Tyr Glu Pro Glu His Arg Glu Tyr Val Glu Lys His

1430 1435 1440

Lys Asp Glu Phe Lys Glu Leu Leu Glu Tyr Ile Ser Val Phe Ser

1445 1450 1455

Arg Lys Tyr Val Leu Ala Asp Asn Asn Leu Thr Lys Ile Glu Met

1460 1465 1470

Leu Phe Ser Lys Asn Lys Asp Ala Glu Val Ser Ser Leu Ala Lys

1475 1480 1485

Ser Phe Ile Ser Leu Leu Thr Phe Thr Ala Phe Gly Ala Pro Ala

1490 1495 1500

Ala Phe Asn Phe Phe Gly Glu Asn Ile Asp Arg Lys Arg Tyr Thr

1505 1510 1515

Ser Val Thr Glu Cys Leu Asn Ala Thr Leu Ile His Gln Ser Ile

1520 1525 1530

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Lys Leu Gly Glu

1535 1540 1545

Asp Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys
1550 1555 1560
Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr
1565 1570 1575

Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp
1580 1585 1590
Tyr Ala
1595

<210> 21
<211> 1863
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 21

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Thr Ser Glu Lys Gly Pro
1 5 10 15
Ser Thr Gly Asp Pro Thr Leu Arg Arg Arg Ile Glu Ser Trp Glu Phe
20 25 30

Asp Val Phe Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu
35 40 45
Tyr Glu Ile Lys Trp Gly Met Ser Arg Lys Ile Trp Arg Ser Ser Gly
50 55 60
Lys Asn Thr Thr Asn His Val Glu Val Asn Phe Ile Lys Lys Phe Thr
65 70 75 80
Ser Glu Arg Arg Phe His Ser Ser Ile Ser Cys Ser Ile Thr Trp Phe
85 90 95

Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser Gln Ala Ile Arg Glu Phe
100 105 110
Leu Ser Gln His Pro Gly Val Thr Leu Val Ile Tyr Val Ala Arg Leu
115 120 125
Phe Trp His Met Asp Gln Arg Asn Arg Gln Gly Leu Arg Asp Leu Val

130 135 140

Asn Ser Gly Val Thr Ile Gln Ile Met Arg Ala Ser Glu Tyr Tyr His
145 150 155 160

Cys Trp Arg Asn Phe Val Asn Tyr Pro Pro Gly Asp Glu Ala His Trp
165 170 175

Pro Gln Tyr Pro Pro Leu Trp Met Met Leu Tyr Ala Leu Glu Leu His
180 185 190

Cys Ile Ile Leu Ser Leu Pro Pro Cys Leu Lys Ile Ser Arg Arg Trp
195 200 205

Gln Asn His Leu Ala Phe Phe Arg Leu His Leu Gln Asn Cys His Tyr
210 215 220

Gln Thr Ile Pro Pro His Ile Leu Leu Ala Thr Gly Leu Ile His Pro
225 230 235 240

Ser Val Thr Trp Arg Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser
245 250 255

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
260 265 270

Gly Gly Ser Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val Gly Gly
275 280 285

Lys Pro Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp
290 295 300

Ala Val Val Thr Asp Asp Tyr Lys Val Pro Ala Lys Lys Met Lys Val
305 310 315 320

Leu Gly Asn Thr Asp Lys Gln Ser Ile Lys Lys Asn Leu Leu Gly Ala
325 330 335

Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg
340 345 350

Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Leu Arg Tyr Leu
355 360 365

Gln Glu Ile Phe Thr Gly Glu Met Asn Lys Val Asp Glu Asn Phe Phe
370 375 380

Gln Arg Leu Asp Asp Ser Phe Leu Val Asp Glu Asp Lys Arg Gly Glu
385 390 395 400
His His Pro Ile Phe Gly Asn Ile Ala Ala Glu Val Lys Tyr His Asp
405 410 415

Asp Phe Pro Thr Ile Tyr His Leu Arg Arg His Leu Ala Asp Thr Ser
420 425 430
Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His Met Ile
435 440 445
Lys Phe Arg Gly His Phe Leu Tyr Glu Gly Asp Leu Lys Ala Glu Asn
450 455 460
Thr Asp Val Gln Ala Leu Phe Lys Asp Phe Val Glu Glu Tyr Asp Lys
465 470 475 480

Thr Ile Glu Glu Ser His Leu Ser Glu Ile Thr Val Asp Ala Leu Ser
485 490 495
Ile Leu Thr Glu Lys Val Ser Lys Ser Ser Arg Leu Glu Asn Leu Ile
500 505 510
Ala His Tyr Pro Thr Glu Lys Lys Asn Thr Leu Phe Gly Asn Leu Ile
515 520 525
Ala Leu Ser Leu Asp Leu His Pro Asn Phe Lys Thr Asn Phe Gln Leu
530 535 540

Ser Glu Asp Ala Lys Leu Gln Phe Ser Lys Asp Thr Tyr Glu Glu Asp
545 550 555 560
Leu Glu Gly Phe Leu Gly Glu Val Gly Asp Glu Tyr Ala Asp Leu Phe
565 570 575
Ala Ser Ala Lys Asn Leu Tyr Asp Ala Ile Leu Leu Ser Gly Ile Leu
580 585 590
Thr Val Asp Asp Asn Ser Thr Lys Ala Pro Leu Ser Ala Ser Met Val
595 600 605

Lys Arg Tyr Glu Glu His Gln Lys Asp Leu Lys Lys Leu Lys Asp Phe
610 615 620
Ile Lys Val Asn Ala Pro Asp Gln Tyr Asn Ala Ile Phe Lys Asp Lys

625 630 635 640

Asn Lys Lys Gly Tyr Ala Ser Tyr Ile Glu Ser Gly Val Lys Gln Asp

645 650 655

Glu Phe Tyr Lys Tyr Leu Lys Gly Ile Leu Leu Lys Ile Asn Gly Ser

660 665 670

Gly Asp Phe Leu Asp Lys Ile Asp Arg Glu Asp Phe Leu Arg Lys Gln

675 680 685

Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gln Glu

690 695 700

Met His Ala Ile Leu Arg Arg Gln Gly Glu His Tyr Pro Phe Leu Lys

705 710 715 720

Glu Asn Gln Asp Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr

725 730 735

Tyr Val Gly Pro Leu Ala Arg Lys Gly Ser Arg Phe Ala Trp Ala Glu

740 745 750

Tyr Lys Ala Asp Glu Lys Ile Thr Pro Trp Asn Phe Asp Asp Ile Leu

755 760 765

Asp Lys Glu Lys Ser Ala Glu Lys Phe Ile Thr Arg Met Thr Leu Asn

770 775 780

Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser Pro Leu

785 790 795 800

Tyr Glu Ala Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val

805 810 815

Asn Glu Gln Gly Glu Ala Lys Phe Phe Asp Thr Asn Met Lys Gln Glu

820 825 830

Ile Phe Asp His Val Phe Lys Glu Asn Arg Lys Val Thr Lys Asp Lys

835 840 845

Leu Leu Asn Tyr Leu Asn Lys Glu Phe Glu Glu Phe Arg Ile Val Asn

850 855 860

Leu Thr Gly Leu Asp Lys Glu Asn Lys Ala Phe Asn Ser Ser Leu Gly

865 870 875 880

Thr Tyr His Asp Leu Arg Lys Ile Leu Asp Lys Ser Phe Leu Asp Asp
885 890 895

Lys Ala Asn Glu Lys Thr Ile Glu Asp Ile Ile Gln Thr Leu Thr Leu
900 905 910

Phe Glu Asp Arg Glu Met Ile Arg Gln Arg Leu Gln Lys Tyr Ser Asp
915 920 925

Ile Phe Thr Lys Ala Gln Leu Lys Lys Leu Glu Arg Arg His Tyr Thr
930 935 940

Gly Trp Gly Arg Leu Ser Tyr Lys Leu Ile Asn Gly Ile Arg Asn Lys
945 950 955 960

Glu Asn Lys Lys Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Tyr Ala
965 970 975

Asn Arg Asn Phe Met Gln Leu Ile Asn Asp Asp Ala Leu Ser Phe Lys
980 985 990

Glu Glu Ile Ala Arg Ala Gln Ile Ile Asp Asp Val Asp Asp Ile Ala
995 1000 1005

Asn Val Val His Asp Leu Pro Gly Ser Pro Ala Ile Lys Lys Gly
1010 1015 1020

Ile Leu Gln Ser Val Lys Ile Val Asp Glu Leu Val Lys Val Met
1025 1030 1035

Gly His Asn Pro Ala Asn Ile Ile Ile Glu Met Ala Arg Glu Asn
1040 1045 1050

Gln Thr Thr Asp Lys Gly Arg Arg Asn Ser Gln Gln Arg Leu Lys
1055 1060 1065

Leu Leu Gln Asp Ser Leu Lys Asn Leu Asp Asn Pro Val Asn Ile
1070 1075 1080

Lys Asn Val Glu Asn Gln Gln Leu Gln Asn Asp Arg Leu Phe Leu
1085 1090 1095

Tyr Tyr Ile Gln Asn Gly Lys Asp Met Tyr Thr Gly Glu Thr Leu
1100 1105 1110

Asp Ile Asn Asn Leu Ser Gln Tyr Asp Ile Asp His Ile Ile Pro

1115 1120 1125

Gln Ala Phe Ile Lys Asp Asn Ser Leu Asp Asn Arg Val Leu Thr

1130 1135 1140

Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Ile

1145 1150 1155

Glu Val Val His Glu Met Lys Ser Phe Trp Ser Lys Leu Leu Ser

1160 1165 1170

Val Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala

1175 1180 1185

Glu Arg Gly Gly Leu Thr Glu Glu Asp Lys Ala Gly Phe Ile Lys

1190 1195 1200

Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln

1205 1210 1215

Ile Leu Asp Glu Arg Phe Asn Thr Glu Phe Asp Gly Asn Lys Arg

1220 1225 1230

Arg Ile Arg Asn Val Lys Ile Ile Thr Leu Lys Ser Asn Leu Val

1235 1240 1245

Ser Asn Phe Arg Lys Glu Phe Glu Leu Tyr Lys Val Arg Glu Ile

1250 1255 1260

Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

1265 1270 1275

Gly Asn Ala Leu Leu Leu Lys Tyr Pro Gln Leu Glu Pro Glu Phe

1280 1285 1290

Val Tyr Gly Glu Tyr Pro Lys Tyr Asn Ser Tyr Arg Ser Arg Lys

1295 1300 1305

Ser Ala Thr Glu Lys Phe Leu Phe Tyr Ser Asn Ile Leu Arg Phe

1310 1315 1320

Phe Lys Lys Glu Asp Ile Gln Thr Asn Glu Asp Gly Glu Ile Ala

1325 1330 1335

Trp Asn Lys Glu Lys His Ile Lys Ile Leu Arg Lys Val Leu Ser

1340 1345 1350

Tyr Pro Gln Val Asn Ile Val Lys Lys Thr Glu Glu Gln Thr Gly
1355 1360 1365

Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Gly Glu Ser Asp Lys
1370 1375 1380

Leu Ile Pro Arg Lys Thr Lys Asn Ser Tyr Trp Asp Pro Lys Lys
1385 1390 1395

Tyr Gly Gly Phe Asp Ser Pro Val Val Ala Tyr Ser Ile Leu Val
1400 1405 1410

Phe Ala Asp Val Glu Lys Gly Lys Ser Lys Lys Leu Arg Lys Val
1415 1420 1425

Gln Asp Met Val Gly Ile Thr Ile Met Glu Lys Lys Arg Phe Glu
1430 1435 1440

Lys Asn Pro Val Asp Phe Leu Glu Gln Arg Gly Tyr Arg Asn Val
1445 1450 1455

Arg Leu Glu Lys Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu
1460 1465 1470

Leu Glu Asn Lys Arg Arg Arg Leu Leu Ala Ser Ala Lys Glu Leu
1475 1480 1485

Gln Lys Gly Asn Glu Leu Val Ile Pro Gln Arg Phe Thr Thr Leu
1490 1495 1500

Leu Tyr His Ser Tyr Arg Ile Glu Lys Asp Tyr Glu Pro Glu His
1505 1510 1515

Arg Glu Tyr Val Glu Lys His Lys Asp Glu Phe Lys Glu Leu Leu
1520 1525 1530

Glu Tyr Ile Ser Val Phe Ser Arg Lys Tyr Val Leu Ala Asp Asn
1535 1540 1545

Asn Leu Thr Lys Ile Glu Met Leu Phe Ser Lys Asn Lys Asp Ala
1550 1555 1560

Glu Val Ser Ser Leu Ala Lys Ser Phe Ile Ser Leu Leu Thr Phe
1565 1570 1575

Thr Ala Phe Gly Ala Pro Ala Ala Phe Asn Phe Phe Gly Glu Asn

1580 1585 1590

Ile Asp Arg Lys Arg Tyr Thr Ser Val Thr Glu Cys Leu Asn Ala
1595 1600 1605
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1610 1615 1620
Asp Leu Ser Lys Leu Gly Glu Asp Gly Lys Arg Pro Ala Ala Thr
1625 1630 1635
Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Ser Gly Gly
1640 1645 1650

Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu
1655 1660 1665
Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met
1670 1675 1680
Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser
1685 1690 1695
Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn
1700 1705 1710

Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala
1715 1720 1725
Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu
1730 1735 1740
Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp
1745 1750 1755
Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser
1760 1765 1770

Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys
1775 1780 1785
Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr
1790 1795 1800
Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys
1805 1810 1815

Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile

1820 1825 1830

Lys Met Leu Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr

1835 1840 1845

Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1850 1855 1860

<210> 22

<211> 208

<212> PRT

<213> Petromyzon marinus

<400> 22

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr

1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg

20 25 30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys

35 40 45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly

50 55 60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg

65 70 75 80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro

85 90 95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu

100 105 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr

115 120 125

Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn

130 135 140

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg

145 150 155 160

Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175
Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser
180 185 190
Ile Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val
195 200 205

<210> 23
<211> 145
<212> PRT
<213> Homo sapiens

<400> 23
Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45
Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95
Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110
Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Ala Pro
130 135 140
Val
145

<210> 24

<211> 198

<212> PRT

<213> Homo sapiens

<400> 24

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

<210> 25

<211> 198

<212> PRT

<213> Mus sp.

<400> 25

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

<210> 26

<211> 198

<212> PRT

<213> Canis lupus familiaris

<400> 26

Met Asp Ser Leu Leu Met Lys Gln Arg Lys Phe Leu Tyr His Phe Lys

1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val

20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly His

35 40 45

Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp

65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp

85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Ala Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg

115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr

130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Lys Thr Phe Lys

145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu

165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala

180 185 190

Phe Arg Thr Leu Gly Leu

195

<210> 27

<211> 199

<212> PRT

<213> Bos sp.

<400> 27

Met Asp Ser Leu Leu Lys Lys Gln Arg Gln Phe Leu Tyr Gln Phe Lys
1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30

Val Lys Arg Arg Asp Ser Pro Thr Ser Phe Ser Leu Asp Phe Gly His
35 40 45

Leu Arg Asn Lys Ala Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr

50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110

Leu Tyr Phe Cys Asp Lys Glu Arg Lys Ala Glu Pro Glu Gly Leu Arg

115 120 125

Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp
130 135 140

Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe
145 150 155 160

Lys Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln
165 170 175

Leu Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp

180 185 190

Ala Phe Arg Thr Leu Gly Leu
195

<210> 28

<211> 239

<212> PRT

<213> Rattus sp.

<400> 28

Met Ala Val Gly Ser Lys Pro Lys Ala Ala Leu Val Gly Pro His Trp
1 5 10 15

Glu Arg Glu Arg Ile Trp Cys Phe Leu Cys Ser Thr Gly Leu Gly Thr
20 25 30

Gln Gln Thr Gly Gln Thr Ser Arg Trp Leu Arg Pro Ala Ala Thr Gln
35 40 45

Asp Pro Val Ser Pro Pro Arg Ser Leu Leu Met Lys Gln Arg Lys Phe
50 55 60

Leu Tyr His Phe Lys Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr
65 70 75 80

Tyr Leu Cys Tyr Val Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser
85 90 95

Leu Asp Phe Gly Tyr Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu
100 105 110

Leu Phe Leu Arg Tyr Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys
115 120 125

Tyr Arg Val Thr Trp Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala
130 135 140

Arg His Val Ala Asp Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg
145 150 155 160

Ile Phe Thr Ala Arg Leu Thr Gly Trp Gly Ala Leu Pro Ala Gly Leu
165 170 175

Met Ser Pro Ala Arg Pro Ser Asp Tyr Phe Tyr Cys Trp Asn Thr Phe
180 185 190

Val Glu Asn His Glu Arg Thr Phe Lys Ala Trp Glu Gly Leu His Glu
195 200 205

Asn Ser Val Arg Leu Ser Arg Arg Leu Arg Arg Ile Leu Leu Pro Leu
210 215 220

Tyr Glu Val Asp Asp Leu Arg Asp Ala Phe Arg Thr Leu Gly Leu

225 230 235

<210> 29

<211> 198

<212> PRT

<213> Canis lupus familiaris

<400> 29

Met Asp Ser Leu Leu Met Lys Gln Arg Lys Phe Leu Tyr His Phe Lys
1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly His
35 40 45

Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Ala Ala Arg
100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr
130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Lys Thr Phe Lys
145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190

Phe Arg Thr Leu Gly Leu
195

<210> 30

<211> 199

<212> PRT

<213> Bos taurus

<400> 30

Met Asp Ser Leu Leu Lys Lys Gln Arg Gln Phe Leu Tyr Gln Phe Lys
1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30

Val Lys Arg Arg Asp Ser Pro Thr Ser Phe Ser Leu Asp Phe Gly His
35 40 45

Leu Arg Asn Lys Ala Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95

Phe Leu Arg Gly Tyr Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110

Leu Tyr Phe Cys Asp Lys Glu Arg Lys Ala Glu Pro Glu Gly Leu Arg
115 120 125

Arg Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp
130 135 140

Tyr Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe
145 150 155 160

Lys Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln
165 170 175

Leu Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp
180 185 190

Ala Phe Arg Thr Leu Gly Leu
195

<210> 31

<211> 198

<212> PRT

<213> Mus musculus

<400> 31

Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15

Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys Tyr Val
20 25 30

Val Lys Arg Arg Asp Ser Ala Thr Ser Phe Ser Leu Asp Phe Gly Tyr
35 40 45

Leu Arg Asn Lys Asn Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60

Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80

Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Asp
85 90 95

Phe Leu Arg Gly Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg
100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125

Leu His Arg Ala Gly Val Gln Ile Ala Ile Met Thr Phe Lys Asp Tyr
130 135 140

Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn His Glu Arg Thr Phe Lys
145 150 155 160

Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Ser Arg Gln Leu
165 170 175

Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190

Phe Arg Thr Leu Gly Leu
195

<210> 32

<211> 229

<212> PRT

<213> Rattus norvegicus

<

400> 32

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys

225

<210> 33

<211> 229

<212> PRT

<213> Mesocricetus auratus

<400> 33

Met Ser Ser Glu Thr Gly Pro Val Val Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Asp Ala Phe Phe Asp Gln Gly Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Arg Trp Gly Gly Arg His
35 40 45

Asn Ile Trp Arg His Thr Gly Gln Asn Thr Ser Arg His Val Glu Ile
50 55 60

Asn Phe Ile Glu Lys Phe Thr Ser Glu Arg Tyr Phe Tyr Pro Ser Thr
65 70 75 80

Arg Cys Ser Ile Val Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Lys Ala Ile Thr Glu Phe Leu Ser Gly His Pro Asn Val Thr Leu
100 105 110

Phe Ile Tyr Ala Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Arg Gly Val Thr Ile Arg Ile Met
130 135 140

Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160

Pro Ser Asn Glu Val Tyr Trp Pro Arg Tyr Pro Asn Leu Trp Met Arg
165 170 175

Leu Tyr Ala Leu Glu Leu Tyr Cys Ile His Leu Gly Leu Pro Pro Cys
180 185 190

Leu Lys Ile Lys Arg Arg His Gln Tyr Pro Leu Thr Phe Phe Arg Leu

195 200 205

Asn Leu Gln Ser Cys His Tyr Gln Arg Ile Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Phe Ile
225

<210> 34
<211> 236
<212> PRT
<213> Pongo pygmaeus
<400> 34

Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Ser Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg
35 40 45
Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val
50 55 60
Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Arg Phe His Ser Ser Ile
65 70 75 80
Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
85 90 95

Ser Gln Ala Ile Arg Glu Phe Leu Ser Gln His Pro Gly Val Thr Leu
100 105 110
Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160

Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met

165 170 175

Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys

180 185 190

Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Ala Phe Phe Arg Leu

195 200 205

His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu

210 215 220

Ala Thr Gly Leu Ile His Pro Ser Val Thr Trp Arg

225 230 235

<210> 35

<211> 236

<212> PRT

<213> Oryctolagus cuniculus

<400> 35

Met Ala Ser Glu Lys Gly Pro Ser Asn Lys Asp Tyr Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro Trp Glu Phe Glu Val Phe Phe Asp Pro Gln Glu Leu

20 25 30

Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Ala Ser Ser

35 40 45

Lys Thr Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val

50 55 60

Asn Phe Leu Glu Lys Leu Thr Ser Glu Gly Arg Leu Gly Pro Ser Thr

65 70 75 80

Cys Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys

85 90 95

Ser Met Ala Ile Arg Glu Phe Leu Ser Gln His Pro Gly Val Thr Leu

100 105 110

Ile Ile Phe Val Ala Arg Leu Phe Gln His Met Asp Arg Arg Asn Arg

115 120 125

Gln Gly Leu Lys Asp Leu Val Thr Ser Gly Val Thr Val Arg Val Met

130 135 140

Ser Val Ser Glu Tyr Cys Tyr Cys Trp Glu Asn Phe Val Asn Tyr Pro
145 150 155 160
Pro Gly Lys Ala Ala Gln Trp Pro Arg Tyr Pro Pro Arg Trp Met Leu
165 170 175

Met Tyr Ala Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Lys Ile Ser Arg Arg His Gln Lys Gln Leu Thr Phe Phe Ser Leu
195 200 205
Thr Pro Gln Tyr Cys His Tyr Lys Met Ile Pro Pro Tyr Ile Leu Leu
210 215 220
Ala Thr Gly Leu Leu Gln Pro Ser Val Pro Trp Arg
225 230 235
<210> 36
<211
> 235
<212> PRT
<213> Monodelphis domestica
<400> 36
Met Asn Ser Lys Thr Gly Pro Ser Val Gly Asp Ala Thr Leu Arg Arg
1 5 10 15
Arg Ile Lys Pro Trp Glu Phe Val Ala Phe Phe Asn Pro Gln Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys Trp Gly Asn Gln Asn
35 40 45
Ile Trp Arg His Ser Asn Gln Asn Thr Ser Gln His Ala Glu Ile Asn

50 55 60
Phe Met Glu Lys Phe Thr Ala Glu Arg His Phe Asn Ser Ser Val Arg
65 70 75 80
Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser
85 90 95
Lys Ala Ile Arg Lys Phe Leu Asp His Tyr Pro Asn Val Thr Leu Ala
100 105 110

Ile Phe Ile Ser Arg Leu Tyr Trp His Met Asp Gln Gln His Arg Gln

115 120 125

Gly Leu Lys Glu Leu Val His Ser Gly Val Thr Ile Gln Ile Met Ser

130 135 140

Tyr Ser Glu Tyr His Tyr Cys Trp Arg Asn Phe Val Asp Tyr Pro Gln

145 150 155 160

Gly Glu Glu Asp Tyr Trp Pro Lys Tyr Pro Tyr Leu Trp Ile Met Leu

165 170 175

Tyr Val Leu Glu Leu His Cys Ile Ile Leu Gly Leu Pro Pro Cys Leu

180 185 190

Lys Ile Ser Gly Ser His Ser Asn Gln Leu Ala Leu Phe Ser Leu Asp

195 200 205

Leu Gln Asp Cys His Tyr Gln Lys Ile Pro Tyr Asn Val Leu Val Ala

210 215 220

Thr Gly Leu Val Gln Pro Phe Val Thr Trp Arg

225 230 235

<210> 37

<211> 224

<212> PRT

<213> Pongo pygmaeus

<400> 37

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Thr Glu Ala Ala Ser Gln

1 5 10 15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu

20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro

35 40 45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly

50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly

65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala
85 90 95
Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110
Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125
Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn

130 135 140
Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Leu
145 150 155 160
Glu Ile Gln Asp Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175
Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val
180 185 190
Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile

195 200 205
Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220
<210> 38
<211> 224
<212> PRT
<213> Bos taurus
<400> 38
Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Ala Glu Pro Ala Ser Gln
1 5 10 15
Asn Gly Glu Glu Val Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45
Ala His Tyr Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60
Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly

65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Thr
85 90 95

Asn His Ala Glu Glu Ala Phe Phe Asn Ser Ile Met Pro Thr Phe Asp
100 105 110

Pro Ala Leu Arg Tyr Met Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125

Ala Ala Cys Ala Asp Arg Ile Val Lys Thr Leu Asn Lys Thr Lys Asn
130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160

Glu Ile Gln Ala Ala Leu Arg Lys Leu Lys Glu Ala Gly Cys Arg Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 39

<211> 440

<212> PRT

<213> Mus musculus

<400> 39

Met Gln Pro Gln Arg Leu Gly Pro Arg Ala Gly Met Gly Pro Phe Cys
1 5 10 15

Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro Ile Arg Asn Leu Ile
20 25 30

Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn Leu Gly Tyr Ala Lys
35 40 45

Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val Thr Arg Lys Asp Cys

50 55 60

Asp Ser Pro Val Ser Leu His His Gly Val Phe Lys Asn Lys Asp Asn

65 70 75 80

Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe His Asp Lys Val Leu

85 90 95

Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile Thr Trp Tyr Met Ser

100 105 110

Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile Val Arg Phe Leu Ala

115 120 125

Thr His His Asn Leu Ser Leu Asp Ile Phe Ser Ser Arg Leu Tyr Asn

130 135 140

Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys Arg Leu Val Gln Glu

145 150 155 160

Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu Phe Lys Lys Cys Trp

165 170 175

Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe Arg Pro Trp Lys Arg

180 185 190

Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys Leu Gln Glu Ile Leu

195 200 205

Arg Pro Cys Tyr Ile Ser Val Pro Ser Ser Ser Ser Ser Thr Leu Ser

210 215 220

Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr Arg Phe Trp Val Glu

225 230 235 240

Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu Phe Tyr Ser Gln Phe

245 250 255

Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr His Arg Met Lys Pro

260 265 270

Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly Gln Ala Pro Leu Lys

275 280 285

Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His Ala Glu Ile Leu Phe

290 295 300

Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln Val Thr Ile Thr Cys

305 310 315 320

Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala Trp Gln Leu Ala Ala

325 330 335

Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His Ile Tyr Thr Ser Arg

340 345 350

Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys Gly Leu Cys Ser Leu

355 360 365

Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp Leu Pro Gln Phe Thr

370 375 380

Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg Pro Phe Trp Pro Trp

385 390 395 400

Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln Arg Arg Leu Arg Arg

405 410 415

Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val Asn Asp Phe Gly Asn

420 425 430

Leu Gln Leu Gly Pro Pro Met Ser

435 440

<210> 40

<211> 429

<212> PRT

<213> Mus sp.

<400> 40

Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro

1 5 10 15

Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn

20 25 30

Leu Gly Tyr Ala Lys Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val

35 40 45

Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val Phe

50 55 60

Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe
65 70 75 80
His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile
85 90 95
Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile
100 105 110
Val Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe Ser

115 120 125
Ser Arg Leu Tyr Asn Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys
130 135 140
Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu
145 150 155 160
Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe
165 170 175
Arg Pro Trp Lys Arg Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys

180 185 190
Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser
195 200 205
Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr
210 215 220
Arg Phe Cys Val Glu Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu
225 230 235 240
Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr

245 250 255
His Arg Met Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly
260 265 270
Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His
275 280 285
Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln
290 295 300
Val Thr Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala

305 310 315 320

Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His

325 330 335

Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys

340 345 350

Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp

355 360 365

Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg

370 375 380

Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln

385 390 395 400

Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val

405 410 415

Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser

420 425

<210> 41

<211> 430

<212> PRT

<213> Rattus sp.

<400> 41

Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro

1 5 10 15

Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn

20 25 30

Arg Leu Arg Tyr Ala Ile Asp Arg Lys Asp Thr Phe Leu Cys Tyr Glu

35 40 45

Val Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val

50 55 60

Phe Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp

65 70 75 80

Phe His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys

85 90 95

Ile Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln
100 105 110
Val Leu Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe
115 120 125
Ser Ser Arg Leu Tyr Asn Ile Arg Asp Pro Glu Asn Gln Gln Asn Leu

130 135 140
Cys Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr
145 150 155 160
Glu Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg
165 170 175
Phe Arg Pro Trp Lys Lys Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser
180 185 190
Lys Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser

195 200 205
Ser Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu
210 215 220
Thr Arg Phe Cys Val Glu Arg Arg Arg Val His Leu Leu Ser Glu Glu
225 230 235 240
Glu Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr
245 250 255
Tyr His Gly Val Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn

260 265 270
Gly Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln
275 280 285
His Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser
290 295 300
Gln Val Ile Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys
305 310 315 320
Ala Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu

325 330 335
His Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln

340 345 350

Lys Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met

355 360 365

Asp Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys

370 375 380

Arg Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr

385 390 395 400

Gln Arg Arg Leu His Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu

405 410 415

Val Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser

420 425 430

<210> 42

<211> 199

<212> PRT

<213> Homo sapiens

<400> 42

Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His

1 5 10 15

Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Thr Tyr

20 25 30

Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met

35 40 45

Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys

50 55 60

Gly Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro

65 70 75 80

Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile

85 90 95

Ser Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala

100 105 110

Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg

115 120 125

Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg
130 135 140

Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His
145 150 155 160
Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp
165 170 175
Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala
180 185 190
Ile Leu Gln Asn Gln Gly Asn
195

<210> 43
<211> 373
<212> PRT
<213> Homo sapiens

<400> 43
Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1 5 10 15
Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
20 25 30
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Arg
35 40 45
Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Gln Pro Glu His
50 55 60

His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu Pro
65 70 75 80
Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro Cys
85 90 95
Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ala Glu His Pro Asn
100 105 110
Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg
115 120 125

Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg Val
130 135 140
Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe Val
145 150 155 160
Tyr Ser Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn
165 170 175
Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met
180 185 190

Glu Ala Met Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Arg
195 200 205
Lys Ala Tyr Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val
210 215 220
Val Lys His His Ser Pro Val Ser Trp Lys Arg Gly Val Phe Arg Asn
225 230 235 240
Gln Val Asp Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser
245 250 255

Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr
260 265 270
Trp Tyr Thr Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala
275 280 285
Glu Phe Leu Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala
290 295 300
Arg Leu Tyr Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser
305 310 315 320

Leu Ser Gln Glu Gly Ala Ser Val Glu Ile Met Gly Tyr Lys Asp Phe
325 330 335
Lys Tyr Cys Trp Glu Asn Phe Val Tyr Asn Asp Asp Glu Pro Phe Lys
340 345 350
Pro Trp Lys Gly Leu Lys Tyr Asn Phe Leu Phe Leu Asp Ser Lys Leu
355 360 365
Gln Glu Ile Leu Glu
370

<210> 44

<211> 370

<212> PRT

<213> Macaca mulatta

<400> 44

Met Val Glu Pro Met Asp Pro Arg Thr Phe Val Ser Asn Phe Asn Asn
1 5 10 15
Arg Pro Ile Leu Ser Gly Leu Asn Thr Val Trp Leu Cys Cys Glu Val
20 25 30
Lys Thr Lys Asp Pro Ser Gly Pro Pro Leu Asp Ala Lys Ile Phe Gln
35 40 45
Gly Lys Val Tyr Ser Lys Ala Lys Tyr His Pro Glu Met Arg Phe Leu
50 55 60

Arg Trp Phe His Lys Trp Arg Gln Leu His His Asp Gln Glu Tyr Lys
65 70 75 80
Val Thr Trp Tyr Val Ser Trp Ser Pro Cys Thr Arg Cys Ala Asn Ser
85 90 95
Val Ala Thr Phe Leu Ala Lys Asp Pro Lys Val Thr Leu Thr Ile Phe
100 105 110
Val Ala Arg Leu Tyr Tyr Phe Trp Lys Pro Asp Tyr Gln Gln Ala Leu
115 120 125

Arg Ile Leu Cys Gln Lys Arg Gly Gly Pro His Ala Thr Met Lys Ile
130 135 140
Met Asn Tyr Asn Glu Phe Gln Asp Cys Trp Asn Lys Phe Val Asp Gly
145 150 155 160
Arg Gly Lys Pro Phe Lys Pro Arg Asn Asn Leu Pro Lys His Tyr Thr
165 170 175
Leu Leu Gln Ala Thr Leu Gly Glu Leu Leu Arg His Leu Met Asp Pro
180 185 190

Gly Thr Phe Thr Ser Asn Phe Asn Asn Lys Pro Trp Val Ser Gly Gln
195 200 205

His Glu Thr Tyr Leu Cys Tyr Lys Val Glu Arg Leu His Asn Asp Thr
210 215 220
Trp Val Pro Leu Asn Gln His Arg Gly Phe Leu Arg Asn Gln Ala Pro
225 230 235 240
Asn Ile His Gly Phe Pro Lys Gly Arg His Ala Glu Leu Cys Phe Leu
245 250 255

Asp Leu Ile Pro Phe Trp Lys Leu Asp Gly Gln Gln Tyr Arg Val Thr
260 265 270
Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln Glu Met Ala
275 280 285
Lys Phe Ile Ser Asn Asn Glu His Val Ser Leu Cys Ile Phe Ala Ala
290 295 300
Arg Ile Tyr Asp Asp Gln Gly Arg Tyr Gln Glu Gly Leu Arg Ala Leu
305 310 315 320

His Arg Asp Gly Ala Lys Ile Ala Met Met Asn Tyr Ser Glu Phe Glu
325 330 335
Tyr Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe Gln Pro
340 345 350
Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg
355 360 365
Ala Ile
370
<210> 45
<211> 384
<212> PRT
<213> Pan troglodytes
<400> 45
Met Lys Pro His Phe Arg Asn Pro Val Glu Arg Met Tyr Gln Asp Thr

1 5 10 15
Phe Ser Asp Asn Phe Tyr Asn Arg Pro Ile Leu Ser His Arg Asn Thr
20 25 30
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro

35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Lys Leu Lys Tyr

50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu

65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro

85 90 95

Cys Thr Lys Cys Thr Arg Asp Val Ala Thr Phe Leu Ala Glu Asp Pro

100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp

115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly

130 135 140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys

145 150 155 160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn

165 170 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile

180 185 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Ser Asn Phe Asn Asn

195 200 205

Glu Leu Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val

210 215 220

Glu Arg Leu His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly

225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg

245 250 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp

260 265 270

Leu His Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys

275 280 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Asn Asn Lys His
290 295 300
Val Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg
305 310 315 320
Cys Gln Glu Gly Leu Arg Thr Leu Ala Lys Ala Gly Ala Lys Ile Ser

325 330 335
Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
340 345 350
His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser
355 360 365
Gln Ala Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
370 375 380

<210> 46
<211> 377
<212> PRT
<213> Chlorocebus sp.

<400> 46
Met Asn Pro Gln Ile Arg Asn Met Val Glu Gln Met Glu Pro Asp Ile
1 5 10 15
Phe Val Tyr Tyr Phe Asn Asn Arg Pro Ile Leu Ser Gly Arg Asn Thr
20 25 30
Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Asp Pro Ser Gly Pro Pro
35 40 45
Leu Asp Ala Asn Ile Phe Gln Gly Lys Leu Tyr Pro Glu Ala Lys Asp
50 55 60

His Pro Glu Met Lys Phe Leu His Trp Phe Arg Lys Trp Arg Gln Leu
65 70 75 80
His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Val Ser Trp Ser Pro
85 90 95
Cys Thr Arg Cys Ala Asn Ser Val Ala Thr Phe Leu Ala Glu Asp Pro
100 105 110
Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Lys

115 120 125

Pro Asp Tyr Gln Gln Ala Leu Arg Ile Leu Cys Gln Glu Arg Gly Gly
130 135 140
Pro His Ala Thr Met Lys Ile Met Asn Tyr Asn Glu Phe Gln His Cys
145 150 155 160
Trp Asn Glu Phe Val Asp Gly Gln Gly Lys Pro Phe Lys Pro Arg Lys
165 170 175
Asn Leu Pro Lys His Tyr Thr Leu Leu His Ala Thr Leu Gly Glu Leu
180 185 190

Leu Arg His Val Met Asp Pro Gly Thr Phe Thr Ser Asn Phe Asn Asn
195 200 205
Lys Pro Trp Val Ser Gly Gln Arg Glu Thr Tyr Leu Cys Tyr Lys Val
210 215 220
Glu Arg Ser His Asn Asp Thr Trp Val Leu Leu Asn Gln His Arg Gly
225 230 235 240
Phe Leu Arg Asn Gln Ala Pro Asp Arg His Gly Phe Pro Lys Gly Arg
245 250 255

His Ala Glu Leu Cys Phe Leu Asp Leu Ile Pro Phe Trp Lys Leu Asp
260 265 270
Asp Gln Gln Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe
275 280 285
Ser Cys Ala Gln Lys Met Ala Lys Phe Ile Ser Asn Asn Lys His Val
290 295 300
Ser Leu Cys Ile Phe Ala Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys
305 310 315 320

Gln Glu Gly Leu Arg Thr Leu His Arg Asp Gly Ala Lys Ile Ala Val
325 330 335
Met Asn Tyr Ser Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Asp Arg
340 345 350
Gln Gly Arg Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln
355 360 365

Ala Leu Ser Gly Arg Leu Arg Ala Ile
370 375

<210> 47

<211> 384

<212> PRT

<213

> Homo sapiens

<400> 47

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro
35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr
50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
85 90 95

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro
100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp
115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
130 135 140

Pro Arg Ala Thr Met Lys Ile Met Asn Tyr Asp Glu Phe Gln His Cys
145 150 155 160

Trp Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn
165 170 175

Asn Leu Pro Lys Tyr Tyr Ile Leu Leu His Ile Met Leu Gly Glu Ile
180 185 190

Leu Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn
195 200 205

Glu Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val
210 215 220

Glu Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly
225 230 235 240

Phe Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg
245 250 255

His Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp
260 265 270

Leu Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys
275 280 285

Phe Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His
290 295 300

Val Ser Leu Cys Ile Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg
305 310 315 320

Cys Gln Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser
325 330 335

Ile Met Thr Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp
340 345 350

His Gln Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser
355 360 365

Gln Asp Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn
370 375 380

<210> 48
<211> 373
<212> PRT
<213> Homo sapiens
<400> 48

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr

20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Arg

35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Gln Pro Glu His

50 55 60

His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu Pro

65 70 75 80

Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro Cys

85 90 95

Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ala Glu His Pro Asn

100 105 110

Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu Arg

115 120 125

Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg Val

130 135 140

Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe Val

145 150 155 160

Tyr Ser Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn

165 170 175

Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met

180 185 190

Glu Ala Met Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Arg

195 200 205

Lys Ala Tyr Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val

210 215 220

Val Lys His His Ser Pro Val Ser Trp Lys Arg Gly Val Phe Arg Asn

225 230 235 240

Gln Val Asp Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser

245 250 255

Trp Phe Cys Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr

260 265 270

Trp Tyr Thr Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala
275 280 285
Glu Phe Leu Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala
290 295 300
Arg Leu Tyr Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Arg Ser

305 310 315 320
Leu Ser Gln Glu Gly Ala Ser Val Glu Ile Met Gly Tyr Lys Asp Phe
325 330 335
Lys Tyr Cys Trp Glu Asn Phe Val Tyr Asn Asp Asp Glu Pro Phe Lys
340 345 350
Pro Trp Lys Gly Leu Lys Tyr Asn Phe Leu Phe Leu Asp Ser Lys Leu
355 360 365
Gln Glu Ile Leu Glu
370

<210
> 49
<211> 382
<212> PRT
<213> Homo sapiens
<400> 49
Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr
1 5 10 15
Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30
Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu
35 40 45
Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Val Tyr Phe Lys Pro Gln

50 55 60
Tyr His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu
65 70 75 80
Pro Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro
85 90 95
Cys Pro Asp Cys Val Ala Lys Leu Ala Glu Phe Leu Ser Glu His Pro

100 105 110

Asn Val Thr Leu Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Trp Glu

115 120 125

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg

130 135 140

Val Thr Ile Met Asp Tyr Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe

145 150 155 160

Val Tyr Asn Glu Gly Gln Gln Phe Met Pro Trp Tyr Lys Phe Asp Glu

165 170 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Leu Arg Tyr Leu

180 185 190

Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val

195 200 205

Leu Arg Arg Arg Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp

210 215 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn

225 230 235 240

Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu

245 250 255

Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile

260 265 270

Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly

275 280 285

Cys Ala Gly Glu Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg

290 295 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys

305 310 315 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met

325 330 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln

340 345 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala
355 360 365
Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn

370 375 380
<210> 50
<211> 395
<212> PRT
<213> Rattus sp.
<400> 50
Met Gln Pro Gln Gly Leu Gly Pro Asn Ala Gly Met Gly Pro Val Cys
1 5 10 15
Leu Gly Cys Ser His Arg Arg Pro Tyr Ser Pro Ile Arg Asn Pro Leu
20 25 30
Lys Lys Leu Tyr Gln Gln Thr Phe Tyr Phe His Phe Lys Asn Val Arg
35 40 45

Tyr Ala Trp Gly Arg Lys Asn Asn Phe Leu Cys Tyr Glu Val Asn Gly
50 55 60
Met Asp Cys Ala Leu Pro Val Pro Leu Arg Gln Gly Val Phe Arg Lys
65 70 75 80
Gln Gly His Ile His Ala Glu Leu Cys Phe Ile Tyr Trp Phe His Asp
85 90 95
Lys Val Leu Arg Val Leu Ser Pro Met Glu Glu Phe Lys Val Thr Trp
100 105 110

Tyr Met Ser Trp Ser Pro Cys Ser Lys Cys Ala Glu Gln Val Ala Arg
115 120 125
Phe Leu Ala Ala His Arg Asn Leu Ser Leu Ala Ile Phe Ser Ser Arg
130 135 140
Leu Tyr Tyr Tyr Leu Arg Asn Pro Asn Tyr Gln Gln Lys Leu Cys Arg
145 150 155 160
Leu Ile Gln Glu Gly Val His Val Ala Ala Met Asp Leu Pro Glu Phe
165 170 175

Lys Lys Cys Trp Asn Lys Phe Val Asp Asn Asp Gly Gln Pro Phe Arg
180 185 190
Pro Trp Met Arg Leu Arg Ile Asn Phe Ser Phe Tyr Asp Cys Lys Leu
195 200 205
Gln Glu Ile Phe Ser Arg Met Asn Leu Leu Arg Glu Asp Val Phe Tyr
210 215 220
Leu Gln Phe Asn Asn Ser His Arg Val Lys Pro Val Gln Asn Arg Tyr
225 230 235 240

Tyr Arg Arg Lys Ser Tyr Leu Cys Tyr Gln Leu Glu Arg Ala Asn Gly
245 250 255
Gln Glu Pro Leu Lys Gly Tyr Leu Leu Tyr Lys Lys Gly Glu Gln His
260 265 270
Val Glu Ile Leu Phe Leu Glu Lys Met Arg Ser Met Glu Leu Ser Gln
275 280 285
Val Arg Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
290 295 300

Arg Gln Leu Ala Ala Phe Lys Lys Asp His Pro Asp Leu Ile Leu Arg
305 310 315 320
Ile Tyr Thr Ser Arg Leu Tyr Phe Trp Arg Lys Lys Phe Gln Lys Gly
325 330 335
Leu Cys Thr Leu Trp Arg Ser Gly Ile His Val Asp Val Met Asp Leu
340 345 350
Pro Gln Phe Ala Asp Cys Trp Thr Asn Phe Val Asn Pro Gln Arg Pro
355 360 365

Phe Arg Pro Trp Asn Glu Leu Glu Lys Asn Ser Trp Arg Ile Gln Arg
370 375 380
Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu
385 390 395
<210> 51
<211> 227
<212> PRT
<213> Bos sp.

<400> 51

Met Asp Gly Trp Glu Val Ala Phe Arg Ser Gly Thr Val Leu Lys Ala
1 5 10 15
Gly Val Leu Gly Val Ser Met Thr Glu Gly Trp Ala Gly Ser Gly His
20 25 30

Pro Gly Gln Gly Ala Cys Val Trp Thr Pro Gly Thr Arg Asn Thr Met
35 40 45
Asn Leu Leu Arg Glu Val Leu Phe Lys Gln Gln Phe Gly Asn Gln Pro
50 55 60
Arg Val Pro Ala Pro Tyr Tyr Arg Arg Lys Thr Tyr Leu Cys Tyr Gln
65 70 75 80
Leu Lys Gln Arg Asn Asp Leu Thr Leu Asp Arg Gly Cys Phe Arg Asn
85 90 95

Lys Lys Gln Arg His Ala Glu Arg Phe Ile Asp Lys Ile Asn Ser Leu
100 105 110
Asp Leu Asn Pro Ser Gln Ser Tyr Lys Ile Ile Cys Tyr Ile Thr Trp
115 120 125
Ser Pro Cys Pro Asn Cys Ala Asn Glu Leu Val Asn Phe Ile Thr Arg
130 135 140
Asn Asn His Leu Lys Leu Glu Ile Phe Ala Ser Arg Leu Tyr Phe His
145 150 155 160

Trp Ile Lys Ser Phe Lys Met Gly Leu Gln Asp Leu Gln Asn Ala Gly
165 170 175
Ile Ser Val Ala Val Met Thr His Thr Glu Phe Glu Asp Cys Trp Glu
180 185 190
Gln Phe Val Asp Asn Gln Ser Arg Pro Phe Gln Pro Trp Asp Lys Leu
195 200 205
Glu Gln Tyr Ser Ala Ser Ile Arg Arg Arg Leu Gln Arg Ile Leu Thr
210 215 220

Ala Pro Ile
225

<210> 52

<211> 490

<212> PRT

<213> Pan troglodytes

<400> 52

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Trp Met Tyr Gln Arg Thr
1 5 10 15

Phe Tyr Tyr Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30

Thr Trp Leu Cys Tyr Glu Val Lys Ile Arg Arg Gly His Ser Asn Leu
35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Gln Met Tyr Ser Gln Pro Glu
50 55 60

His His Ala Glu Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Gln Leu
65 70 75 80

Ser Ala Tyr Lys Cys Phe Gln Ile Thr Trp Phe Val Ser Trp Thr Pro
85 90 95

Cys Pro Asp Cys Val Ala Lys Leu Ala Lys Phe Leu Ala Glu His Pro
100 105 110

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

Arg Asp Tyr Arg Arg Ala Leu Cys Arg Leu Ser Gln Ala Gly Ala Arg
130 135 140

Val Lys Ile Met Asp Asp Glu Glu Phe Ala Tyr Cys Trp Glu Asn Phe
145 150 155 160

Val Tyr Asn Glu Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp
165 170 175

Asn Tyr Ala Phe Leu His Arg Thr Leu Lys Glu Ile Ile Arg His Leu
180 185 190

Met Asp Pro Asp Thr Phe Thr Phe Asn Phe Asn Asn Asp Pro Leu Val
195 200 205

Leu Arg Arg His Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp

210 215 220

Asn Gly Thr Trp Val Leu Met Asp Gln His Met Gly Phe Leu Cys Asn

225 230 235 240

Glu Ala Lys Asn Leu Leu Cys Gly Phe Tyr Gly Arg His Ala Glu Leu

245 250 255

Arg Phe Leu Asp Leu Val Pro Ser Leu Gln Leu Asp Pro Ala Gln Ile

260 265 270

Tyr Arg Val Thr Trp Phe Ile Ser Trp Ser Pro Cys Phe Ser Trp Gly

275 280 285

Cys Ala Gly Gln Val Arg Ala Phe Leu Gln Glu Asn Thr His Val Arg

290 295 300

Leu Arg Ile Phe Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys

305 310 315 320

Glu Ala Leu Gln Met Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met

325 330 335

Thr Tyr Asp Glu Phe Glu Tyr Cys Trp Asp Thr Phe Val Tyr Arg Gln

340 345 350

Gly Cys Pro Phe Gln Pro Trp Asp Gly Leu Glu Glu His Ser Gln Ala

355 360 365

Leu Ser Gly Arg Leu Arg Ala Ile Leu Gln Val Arg Ala Ser Ser Leu

370 375 380

Cys Met Val Pro His Arg Pro Pro Pro Pro Pro Gln Ser Pro Gly Pro

385 390 395 400

Cys Leu Pro Leu Cys Ser Glu Pro Pro Leu Gly Ser Leu Leu Pro Thr

405 410 415

Gly Arg Pro Ala Pro Ser Leu Pro Phe Leu Leu Thr Ala Ser Phe Ser

420 425 430

Phe Pro Pro Pro Ala Ser Leu Pro Pro Leu Pro Ser Leu Ser Leu Ser

435 440 445

Pro Gly His Leu Pro Val Pro Ser Phe His Ser Leu Thr Ser Cys Ser

450 455 460

Ile Gln Pro Pro Cys Ser Ser Arg Ile Arg Glu Thr Glu Gly Trp Ala
465 470 475 480
Ser Val Ser Lys Glu Gly Arg Asp Leu Gly
485 490

<210> 53
<211> 190
<212> PRT
<213> Homo sapiens
<400> 53

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

180 185 190

<210> 54

<211> 190

<212> PRT

<213> Gorilla gorilla

<400> 54

Met Asn Pro Gln Ile Arg Asn Pro Met Lys Ala Met Tyr Pro Gly Thr
1 5 10 15
Phe Tyr Phe Gln Phe Lys Asn Leu Trp Glu Ala Asn Asp Arg Asn Glu
20 25 30
Thr Trp Leu Cys Phe Thr Val Glu Gly Ile Lys Arg Arg Ser Val Val
35 40 45
Ser Trp Lys Thr Gly Val Phe Arg Asn Gln Val Asp Ser Glu Thr His
50 55 60

Cys His Ala Glu Arg Cys Phe Leu Ser Trp Glu Cys Asp Asp Ile Leu
65 70 75 80
Ser Pro Asn Thr Asn Tyr Gln Val Thr Trp Tyr Thr Ser Trp Ser Pro
85 90 95
Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu Ala Arg His Ser
100 105 110
Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Tyr Tyr Phe Gln Asp
115 120 125

Thr Asp Tyr Gln Glu Gly Leu Arg Ser Leu Ser Gln Glu Gly Val Ala
130 135 140
Val Lys Ile Met Asp Tyr Lys Asp Phe Lys Tyr Cys Trp Glu Asn Phe
145 150 155 160
Val Tyr Asn Asp Asp Glu Pro Phe Lys Pro Trp Lys Gly Leu Lys Tyr
165 170 175
Asn Phe Arg Phe Leu Lys Arg Arg Leu Gln Glu Ile Leu Glu
180 185 190

<210> 55

<211> 199

<212> PRT

<213> Homo sapiens

<400> 55

Met Glu Ala Ser Pro Ala Ser Gly Pro Arg His Leu Met Asp Pro His
1 5 10 15
Ile Phe Thr Ser Asn Phe Asn Asn Gly Ile Gly Arg His Lys Thr Tyr
20 25 30
Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Ser Val Lys Met
35 40 45
Asp Gln His Arg Gly Phe Leu His Asn Gln Ala Lys Asn Leu Leu Cys

50 55 60
Gly Phe Tyr Gly Arg His Ala Glu Leu Arg Phe Leu Asp Leu Val Pro
65 70 75 80
Ser Leu Gln Leu Asp Pro Ala Gln Ile Tyr Arg Val Thr Trp Phe Ile
85 90 95
Ser Trp Ser Pro Cys Phe Ser Trp Gly Cys Ala Gly Glu Val Arg Ala
100 105 110
Phe Leu Gln Glu Asn Thr His Val Arg Leu Arg Ile Phe Ala Ala Arg

115 120 125
Ile Tyr Asp Tyr Asp Pro Leu Tyr Lys Glu Ala Leu Gln Met Leu Arg
130 135 140
Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Asp Glu Phe Lys His
145 150 155 160
Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe Gln Pro Trp
165 170 175
Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg Leu Arg Ala

180 185 190
Ile Leu Gln Asn Gln Gly Asn
195

<210> 56

<211> 202

<212> PRT

<213> Macaca mulatta

<400> 56

Met Asp Gly Ser Pro Ala Ser Arg Pro Arg His Leu Met Asp Pro Asn
1 5 10 15

Thr Phe Thr Phe Asn Phe Asn Asn Asp Leu Ser Val Arg Gly Arg His
20 25 30

Gln Thr Tyr Leu Cys Tyr Glu Val Glu Arg Leu Asp Asn Gly Thr Trp
35 40 45

Val Pro Met Asp Glu Arg Arg Gly Phe Leu Cys Asn Lys Ala Lys Asn
50 55 60

Val Pro Cys Gly Asp Tyr Gly Cys His Val Glu Leu Arg Phe Leu Cys
65 70 75 80

Glu Val Pro Ser Trp Gln Leu Asp Pro Ala Gln Thr Tyr Arg Val Thr
85 90 95

Trp Phe Ile Ser Trp Ser Pro Cys Phe Arg Arg Gly Cys Ala Gly Gln
100 105 110

Val Arg Val Phe Leu Gln Glu Asn Lys His Val Arg Leu Arg Ile Phe
115 120 125

Ala Ala Arg Ile Tyr Asp Tyr Asp Pro Leu Tyr Gln Glu Ala Leu Arg
130 135 140

Thr Leu Arg Asp Ala Gly Ala Gln Val Ser Ile Met Thr Tyr Glu Glu
145 150 155 160

Phe Lys His Cys Trp Asp Thr Phe Val Asp Arg Gln Gly Arg Pro Phe
165 170 175

Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Ala Leu Ser Gly Arg
180 185 190

Leu Arg Ala Ile Leu Gln Asn Gln Gly Asn
195 200

<210> 57

<211> 185

<212> PRT

<213> Bos sp.

<400> 57

Met Asp Glu Tyr Thr Phe Thr Glu Asn Phe Asn Asn Gln Gly Trp Pro
1 5 10 15
Ser Lys Thr Tyr Leu Cys Tyr Glu Met Glu Arg Leu Asp Gly Asp Ala

20 25 30
Thr Ile Pro Leu Asp Glu Tyr Lys Gly Phe Val Arg Asn Lys Gly Leu
35 40 45
Asp Gln Pro Glu Lys Pro Cys His Ala Glu Leu Tyr Phe Leu Gly Lys
50 55 60
Ile His Ser Trp Asn Leu Asp Arg Asn Gln His Tyr Arg Leu Thr Cys
65 70 75 80
Phe Ile Ser Trp Ser Pro Cys Tyr Asp Cys Ala Gln Lys Leu Thr Thr

85 90 95
Phe Leu Lys Glu Asn His His Ile Ser Leu His Ile Leu Ala Ser Arg
100 105 110
Ile Tyr Thr His Asn Arg Phe Gly Cys His Gln Ser Gly Leu Cys Glu
115 120 125
Leu Gln Ala Ala Gly Ala Arg Ile Thr Ile Met Thr Phe Glu Asp Phe
130 135 140
Lys His Cys Trp Glu Thr Phe Val Asp His Lys Gly Lys Pro Phe Gln

145 150 155 160
Pro Trp Glu Gly Leu Asn Val Lys Ser Gln Ala Leu Cys Thr Glu Leu
165 170 175
Gln Ala Ile Leu Lys Thr Gln Gln Asn
180 185

<210> 58

<211> 200

<212> PRT

<213> Homo sapiens

<400> 58

Met Ala Leu Leu Thr Ala Glu Thr Phe Arg Leu Gln Phe Asn Asn Lys

1 5 10 15

Arg Arg Leu Arg Arg Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr

20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly Tyr Phe Glu

35 40 45

Asn Lys Lys Lys Cys His Ala Glu Ile Cys Phe Ile Asn Glu Ile Lys

50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu

65 70 75 80

Thr Trp Ser Pro Cys Ser Ser Cys Ala Trp Glu Leu Val Asp Phe Ile

85 90 95

Lys Ala His Asp His Leu Asn Leu Gly Ile Phe Ala Ser Arg Leu Tyr

100 105 110

Tyr His Trp Cys Lys Pro Gln Gln Lys Gly Leu Arg Leu Leu Cys Gly

115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Phe Pro Lys Phe Ala Asp Cys

130 135 140

Trp Glu Asn Phe Val Asp His Glu Lys Pro Leu Ser Phe Asn Pro Tyr

145 150 155 160

Lys Met Leu Glu Glu Leu Asp Lys Asn Ser Arg Ala Ile Lys Arg Arg

165 170 175

Leu Glu Arg Ile Lys Ile Pro Gly Val Arg Ala Gln Gly Arg Tyr Met

180 185 190

Asp Ile Leu Cys Asp Ala Glu Val

195 200

<210> 59

<211> 210

<212> PRT

<213> Macaca mulatta

<400> 59

Met Ala Leu Leu Thr Ala Lys Thr Phe Ser Leu Gln Phe Asn Asn Lys

1 5 10 15

Arg Arg Val Asn Lys Pro Tyr Tyr Pro Arg Lys Ala Leu Leu Cys Tyr

20 25 30

Gln Leu Thr Pro Gln Asn Gly Ser Thr Pro Thr Arg Gly His Leu Lys

35 40 45

Asn Lys Lys Lys Asp His Ala Glu Ile Arg Phe Ile Asn Lys Ile Lys

50 55 60

Ser Met Gly Leu Asp Glu Thr Gln Cys Tyr Gln Val Thr Cys Tyr Leu

65 70 75 80

Thr Trp Ser Pro Cys Pro Ser Cys Ala Gly Glu Leu Val Asp Phe Ile

85 90 95

Lys Ala His Arg His Leu Asn Leu Arg Ile Phe Ala Ser Arg Leu Tyr

100 105 110

Tyr His Trp Arg Pro Asn Tyr Gln Glu Gly Leu Leu Leu Leu Cys Gly

115 120 125

Ser Gln Val Pro Val Glu Val Met Gly Leu Pro Glu Phe Thr Asp Cys

130 135 140

Trp Glu Asn Phe Val Asp His Lys Glu Pro Pro Ser Phe Asn Pro Ser

145 150 155 160

Glu Lys Leu Glu Glu Leu Asp Lys Asn Ser Gln Ala Ile Lys Arg Arg

165 170 175

Leu Glu Arg Ile Lys Ser Arg Ser Val Asp Val Leu Glu Asn Gly Leu

180 185 190

Arg Ser Leu Gln Leu Gly Pro Val Thr Pro Ser Ser Ser Ile Arg Asn

195 200 205

Ser Arg

210

<210> 60

<211> 386

<212> PRT

<213> Homo sapiens

<400> 60

Met Asn Pro Gln Ile Arg Asn Pro Met Glu Arg Met Tyr Arg Asp Thr
1 5 10 15
Phe Tyr Asp Asn Phe Glu Asn Glu Pro Ile Leu Tyr Gly Arg Ser Tyr
20 25 30
Thr Trp Leu Cys Tyr Glu Val Lys Ile Lys Arg Gly Arg Ser Asn Leu
35 40 45

Leu Trp Asp Thr Gly Val Phe Arg Gly Pro Val Leu Pro Lys Arg Gln
50 55 60
Ser Asn His Arg Gln Glu Val Tyr Phe Arg Phe Glu Asn His Ala Glu
65 70 75 80
Met Cys Phe Leu Ser Trp Phe Cys Gly Asn Arg Leu Pro Ala Asn Arg
85 90 95
Arg Phe Gln Ile Thr Trp Phe Val Ser Trp Asn Pro Cys Leu Pro Cys
100 105 110

Val Val Lys Val Thr Lys Phe Leu Ala Glu His Pro Asn Val Thr Leu
115 120 125
Thr Ile Ser Ala Ala Arg Leu Tyr Tyr Tyr Arg Asp Arg Asp Trp Arg
130 135 140
Trp Val Leu Leu Arg Leu His Lys Ala Gly Ala Arg Val Lys Ile Met
145 150 155 160
Asp Tyr Glu Asp Phe Ala Tyr Cys Trp Glu Asn Phe Val Cys Asn Glu
165 170 175

Gly Gln Pro Phe Met Pro Trp Tyr Lys Phe Asp Asp Asn Tyr Ala Ser
180 185 190
Leu His Arg Thr Leu Lys Glu Ile Leu Arg Asn Pro Met Glu Ala Met
195 200 205
Tyr Pro His Ile Phe Tyr Phe His Phe Lys Asn Leu Leu Lys Ala Cys
210 215 220
Gly Arg Asn Glu Ser Trp Leu Cys Phe Thr Met Glu Val Thr Lys His
225 230 235 240

His Ser Ala Val Phe Arg Lys Arg Gly Val Phe Arg Asn Gln Val Asp

245 250 255

Pro Glu Thr His Cys His Ala Glu Arg Cys Phe Leu Ser Trp Phe Cys

260 265 270

Asp Asp Ile Leu Ser Pro Asn Thr Asn Tyr Glu Val Thr Trp Tyr Thr

275 280 285

Ser Trp Ser Pro Cys Pro Glu Cys Ala Gly Glu Val Ala Glu Phe Leu

290 295 300

Ala Arg His Ser Asn Val Asn Leu Thr Ile Phe Thr Ala Arg Leu Cys

305 310 315 320

Tyr Phe Trp Asp Thr Asp Tyr Gln Glu Gly Leu Cys Ser Leu Ser Gln

325 330 335

Glu Gly Ala Ser Val Lys Ile Met Gly Tyr Lys Asp Phe Val Ser Cys

340 345 350

Trp Lys Asn Phe Val Tyr Ser Asp Asp Glu Pro Phe Lys Pro Trp Lys

355 360 365

Gly Leu Gln Thr Asn Phe Arg Leu Leu Lys Arg Arg Leu Arg Glu Ile

370 375 380

Leu Gln

385

<210> 61

<211> 236

<212> PRT

<213> Homo sapiens

<400> 61

Met Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro Trp Glu Phe Asp Val Phe Tyr Asp Pro Arg Glu Leu

20 25 30

Arg Lys Glu Ala Cys Leu Leu Tyr Glu Ile Lys Trp Gly Met Ser Arg

35 40 45

Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr Thr Asn His Val Glu Val

50 55 60

Asn Phe Ile Lys Lys Phe Thr Ser Glu Arg Asp Phe His Pro Ser Met
65 70 75 80
Ser Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Trp Glu Cys
85 90 95
Ser Gln Ala Ile Arg Glu Phe Leu Ser Arg His Pro Gly Val Thr Leu

100 105 110
Val Ile Tyr Val Ala Arg Leu Phe Trp His Met Asp Gln Gln Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Arg Ala Ser Glu Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro
145 150 155 160
Pro Gly Asp Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met

165 170 175
Leu Tyr Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys
180 185 190
Leu Lys Ile Ser Arg Arg Trp Gln Asn His Leu Thr Phe Phe Arg Leu
195 200 205
His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu Leu
210 215 220
Ala Thr Gly Leu Ile His Pro Ser Val Ala Trp Arg

225 230 235
<210> 62
<211> 229
<212> PRT
<213> Mus sp.
<400> 62
Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His

35 40 45

Ser Val Trp Arg His Thr Ser Gln Asn Thr Ser Asn His Val Glu Val

50 55 60

Asn Phe Leu Glu Lys Phe Thr Thr Glu Arg Tyr Phe Arg Pro Asn Thr

65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys

85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg His Pro Tyr Val Thr Leu

100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Thr Asp Gln Arg Asn Arg

115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Tyr Cys Tyr Cys Trp Arg Asn Phe Val Asn Tyr Pro

145 150 155 160

Pro Ser Asn Glu Ala Tyr Trp Pro Arg Tyr Pro His Leu Trp Val Lys

165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys

180 185 190

Leu Lys Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile

195 200 205

Thr Leu Gln Thr Cys His Tyr Gln Arg Ile Pro Pro His Leu Leu Trp

210 215 220

Ala Thr Gly Leu Lys

225

<210> 63

<211> 229

<212> PRT

<213> Rattus sp.

<400> 63

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg

1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205

Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys
225
<210> 64
<211> 224

<212> PRT

<213> Homo sapiens

<400> 64

Met Ala Gln Lys Glu Glu Ala Ala Val Ala Thr Glu Ala Ala Ser Gln
1 5 10 15

Asn Gly Glu Asp Leu Glu Asn Leu Asp Asp Pro Glu Lys Leu Lys Glu
20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45

Ala Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Gly Lys Gly
65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Ala
85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110

Pro Ala Leu Arg Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125

Ala Ala Cys Ala Asp Arg Ile Ile Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160

Glu Ile Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Val Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Gln Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 65

<211> 224

<212> PRT

<213> Mus sp.

<400> 65

Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln
1 5 10 15

Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
20 25 30

Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro
35 40 45

Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Val Gln Ser Lys Gly
65 70 75 80

Gly Gln Ala Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly
85 90 95

Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110

Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125

Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140

Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160

Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 66

<211> 224

<212> PRT

<213> Rattus sp.

<400> 66

Met Ala Gln Lys Glu Glu Ala Ala Glu Ala Ala Ala Pro Ala Ser Gln
1 5 10 15
Asn Gly Asp Asp Leu Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu

20 25 30
Leu Ile Asp Leu Pro Pro Phe Glu Ile Val Thr Gly Val Arg Leu Pro
35 40 45
Val Asn Phe Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60
Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly
65 70 75 80
Gly Gln Val Gln Ala Thr Gln Gly Tyr Leu Glu Asp Glu His Ala Gly

85 90 95
Ala His Ala Glu Glu Ala Phe Phe Asn Thr Ile Leu Pro Ala Phe Asp
100 105 110
Pro Ala Leu Lys Tyr Asn Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125
Ala Ala Cys Ala Asp Arg Ile Leu Lys Thr Leu Ser Lys Thr Lys Asn
130 135 140
Leu Arg Leu Leu Ile Leu Val Ser Arg Leu Phe Met Trp Glu Glu Pro

145 150 155 160
Glu Val Gln Ala Ala Leu Lys Lys Leu Lys Glu Ala Gly Cys Lys Leu
165 170 175
Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Leu Trp Gln Asn Phe Val
180 185 190
Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys

210 215 220

<210> 67

<211> 224

<212> PRT

<213> Bos sp.

<400> 67

Met Ala Gln Lys Glu Glu Ala Ala Ala Ala Ala Glu Pro Ala Ser Gln
1 5 10 15

Asn Gly Glu Glu Val Glu Asn Leu Glu Asp Pro Glu Lys Leu Lys Glu
20 25 30

Leu Ile Glu Leu Pro Pro Phe Glu Ile Val Thr Gly Glu Arg Leu Pro
35 40 45

Ala His Tyr Phe Lys Phe Gln Phe Arg Asn Val Glu Tyr Ser Ser Gly
50 55 60

Arg Asn Lys Thr Phe Leu Cys Tyr Val Val Glu Ala Gln Ser Lys Gly
65 70 75 80

Gly Gln Val Gln Ala Ser Arg Gly Tyr Leu Glu Asp Glu His Ala Thr
85 90 95

Asn His Ala Glu Glu Ala Phe Phe Asn Ser Ile Met Pro Thr Phe Asp
100 105 110

Pro Ala Leu Arg Tyr Met Val Thr Trp Tyr Val Ser Ser Ser Pro Cys
115 120 125

Ala Ala Cys Ala Asp Arg Ile Val Lys Thr Leu Asn Lys Thr Lys Asn
130 135 140

Leu Arg Leu Leu Ile Leu Val Gly Arg Leu Phe Met Trp Glu Glu Pro
145 150 155 160

Glu Ile Gln Ala Ala Leu Arg Lys Leu Lys Glu Ala Gly Cys Arg Leu
165 170 175

Arg Ile Met Lys Pro Gln Asp Phe Glu Tyr Ile Trp Gln Asn Phe Val
180 185 190

Glu Gln Glu Glu Gly Glu Ser Lys Ala Phe Glu Pro Trp Glu Asp Ile
195 200 205

Gln Glu Asn Phe Leu Tyr Tyr Glu Glu Lys Leu Ala Asp Ile Leu Lys
210 215 220

<210> 68

<211> 208

<212> PRT

<213> Petromyzon marinus

<400> 68

Met Thr Asp Ala Glu Tyr Val Arg Ile His Glu Lys Leu Asp Ile Tyr

1 5 10 15

Thr Phe Lys Lys Gln Phe Phe Asn Asn Lys Lys Ser Val Ser His Arg
20 25 30

Cys Tyr Val Leu Phe Glu Leu Lys Arg Arg Gly Glu Arg Arg Ala Cys
35 40 45

Phe Trp Gly Tyr Ala Val Asn Lys Pro Gln Ser Gly Thr Glu Arg Gly
50 55 60

Ile His Ala Glu Ile Phe Ser Ile Arg Lys Val Glu Glu Tyr Leu Arg

65 70 75 80

Asp Asn Pro Gly Gln Phe Thr Ile Asn Trp Tyr Ser Ser Trp Ser Pro
85 90 95

Cys Ala Asp Cys Ala Glu Lys Ile Leu Glu Trp Tyr Asn Gln Glu Leu
100 105 110

Arg Gly Asn Gly His Thr Leu Lys Ile Trp Ala Cys Lys Leu Tyr Tyr
115 120 125

Glu Lys Asn Ala Arg Asn Gln Ile Gly Leu Trp Asn Leu Arg Asp Asn

130 135 140

Gly Val Gly Leu Asn Val Met Val Ser Glu His Tyr Gln Cys Cys Arg
145 150 155 160

Lys Ile Phe Ile Gln Ser Ser His Asn Gln Leu Asn Glu Asn Arg Trp
165 170 175

Leu Glu Lys Thr Leu Lys Arg Ala Glu Lys Arg Arg Ser Glu Leu Ser

180 185 190

Phe Met Ile Gln Val Lys Ile Leu His Thr Thr Lys Ser Pro Ala Val

195 200 205

<210> 69

<211> 381

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 69

Met Lys Pro His Phe Arg Asn Thr Val Glu Arg Met Tyr Arg Asp Thr
1 5 10 15

Phe Ser Tyr Asn Phe Tyr Asn Arg Pro Ile Leu Ser Arg Arg Asn Thr
20 25 30

Val Trp Leu Cys Tyr Glu Val Lys Thr Lys Gly Pro Ser Arg Pro Pro

35 40 45

Leu Asp Ala Lys Ile Phe Arg Gly Gln Val Tyr Ser Glu Leu Lys Tyr
50 55 60

His Pro Glu Met Arg Phe Phe His Trp Phe Ser Lys Trp Arg Lys Leu
65 70 75 80

His Arg Asp Gln Glu Tyr Glu Val Thr Trp Tyr Ile Ser Trp Ser Pro
85 90 95

Cys Thr Lys Cys Thr Arg Asp Met Ala Thr Phe Leu Ala Glu Asp Pro

100 105 110

Lys Val Thr Leu Thr Ile Phe Val Ala Arg Leu Tyr Tyr Phe Trp Asp

115 120 125

Pro Asp Tyr Gln Glu Ala Leu Arg Ser Leu Cys Gln Lys Arg Asp Gly
130 135 140

Pro Arg Ala Thr Met Lys Phe Asn Tyr Asp Glu Phe Gln His Cys Trp
145 150 155 160

Ser Lys Phe Val Tyr Ser Gln Arg Glu Leu Phe Glu Pro Trp Asn Asn

165 170 175

Leu Pro Lys Tyr Tyr Ile Leu Leu His Phe Met Leu Gly Glu Ile Leu

180 185 190

Arg His Ser Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu

195 200 205

Pro Trp Val Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu

210 215 220

Arg Met His Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe

225 230 235 240

Leu Cys Asn Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His

245 250 255

Ala Glu Leu Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu

260 265 270

Asp Gln Asp Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe

275 280 285

Ser Cys Ala Gln Glu Met Ala Lys Phe Ile Ser Lys Lys His Val Ser

290 295 300

Leu Cys Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln

305 310 315 320

Glu Gly Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Thr

325 330 335

Tyr Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly

340 345 350

Cys Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu

355 360 365

Ser Gly Arg Leu Arg Ala Ile Leu Gln Asn Gln Glu Asn

370 375 380

<210> 70

<211> 182

<212> PRT

<213> Homo sapiens

<400> 70

Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Trp
1 5 10 15
Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His Asn
20 25 30
Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn Gln

35 40 45
Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu Cys
50 55 60
Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp Tyr
65 70 75 80
Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala Gln
85 90 95
Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys Ile

100 105 110
Phe Thr Ala Arg Ile Tyr Asp Asp Gln Gly Arg Cys Gln Glu Gly Leu
115 120 125
Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Thr Tyr Ser Glu
130 135 140
Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys Pro Phe
145 150 155 160
Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser Gly Arg

165 170 175
Leu Arg Ala Ile Leu Gln
180

<210> 71
<211> 184
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 71
Met Asp Pro Pro Thr Phe Thr Phe Asn Phe Asn Asn Glu Pro Trp Val

1 5 10 15

Arg Gly Arg His Glu Thr Tyr Leu Cys Tyr Glu Val Glu Arg Met His

20 25 30

Asn Asp Thr Trp Val Leu Leu Asn Gln Arg Arg Gly Phe Leu Cys Asn

35 40 45

Gln Ala Pro His Lys His Gly Phe Leu Glu Gly Arg His Ala Glu Leu

50 55 60

Cys Phe Leu Asp Val Ile Pro Phe Trp Lys Leu Asp Leu Asp Gln Asp

65 70 75 80

Tyr Arg Val Thr Cys Phe Thr Ser Trp Ser Pro Cys Phe Ser Cys Ala

85 90 95

Gln Glu Met Ala Lys Phe Ile Ser Lys Asn Lys His Val Ser Leu Cys

100 105 110

Ile Phe Thr Ala Arg Ile Tyr Arg Arg Gln Gly Arg Cys Gln Glu Gly

115 120 125

Leu Arg Thr Leu Ala Glu Ala Gly Ala Lys Ile Ser Phe Met Thr Tyr

130 135 140

Ser Glu Phe Lys His Cys Trp Asp Thr Phe Val Asp His Gln Gly Cys

145 150 155 160

Pro Phe Gln Pro Trp Asp Gly Leu Asp Glu His Ser Gln Asp Leu Ser

165 170 175

Gly Arg Leu Arg Ala Ile Leu Gln

180

<210> 72

<211> 367

<212> PRT

<213> Homo sapiens

<400> 72

Met Glu Pro Ile Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val

1 5 10 15

Lys Pro Tyr Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro

20 25 30

Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe
35 40 45
Cys Gln Ile Phe Gly Phe Pro Tyr Gly Thr Thr Phe Pro Gln Thr Lys
50 55 60
His Leu Thr Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln
65 70 75 80
Lys Gly His Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser
85 90 95

Met Leu Phe Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn
100 105 110
Asp Ser Ile Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asn
115 120 125
Glu Ala Asn His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Ile Thr
130 135 140
Tyr Pro Gly Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr
145 150 155 160

Glu Met Asp Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser
165 170 175
Leu Ala Ser Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly
180 185 190
Ile Trp His Ser Val Leu His Ser Phe Ile Ser Gly Val Ser Gly Ser
195 200 205
His Val Phe Gln Pro Ile Leu Thr Gly Arg Ala Leu Ala Asp Arg His
210 215 220

Asn Ala Tyr Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Tyr Phe Thr
225 230 235 240
Asp Val Leu Leu Gln Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Glu
245 250 255
Ala Leu Glu Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Phe Phe
260 265 270

Gln Met Pro Ser Gly Gln Leu Gln Pro Asn Leu Pro Pro Asp Leu Arg
275 280 285

Ala Pro Val Val Phe Val Leu Val Pro Leu Arg Asp Leu Pro Pro Met
290 295 300
His Met Gly Gln Asn Pro Asn Lys Pro Arg Asn Ile Val Arg His Leu
305 310 315 320
Asn Met Pro Gln Met Ser Phe Gln Glu Thr Lys Asp Leu Gly Arg Leu
325 330 335
Pro Thr Gly Arg Ser Val Glu Ile Val Glu Ile Thr Glu Gln Phe Ala
340 345 350

Ser Ser Lys Glu Ala Asp Glu Lys Lys Lys Lys Lys Gly Lys Lys
355 360 365
<210> 73
<211> 198
<212> PRT
<213> Mus musculus
<400> 73
Met Asp Ser Leu Leu Met Lys Gln Lys Lys Phe Leu Tyr His Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg His Glu Thr Tyr Leu Cys Tyr Val
20 25 30
Val Lys Arg Arg Asp Ser Ala Thr Ser Cys Ser Leu Asp Phe Gly His

35 40 45
Leu Arg Asn Lys Ser Gly Cys His Val Glu Leu Leu Phe Leu Arg Tyr
50 55 60
Ile Ser Asp Trp Asp Leu Asp Pro Gly Arg Cys Tyr Arg Val Thr Trp
65 70 75 80
Phe Thr Ser Trp Ser Pro Cys Tyr Asp Cys Ala Arg His Val Ala Glu
85 90 95
Phe Leu Arg Trp Asn Pro Asn Leu Ser Leu Arg Ile Phe Thr Ala Arg

100 105 110

Leu Tyr Phe Cys Glu Asp Arg Lys Ala Glu Pro Glu Gly Leu Arg Arg
115 120 125
Leu His Arg Ala Gly Val Gln Ile Gly Ile Met Thr Phe Lys Asp Tyr
130 135 140
Phe Tyr Cys Trp Asn Thr Phe Val Glu Asn Arg Glu Arg Thr Phe Lys
145 150 155 160
Ala Trp Glu Gly Leu His Glu Asn Ser Val Arg Leu Thr Arg Gln Leu

165 170 175
Arg Arg Ile Leu Leu Pro Leu Tyr Glu Val Asp Asp Leu Arg Asp Ala
180 185 190
Phe Arg Met Leu Gly Phe
195

<210> 74
<211> 388
<212> PRT
<213> Rattus norvegicus
<400> 74
Met Glu Pro Leu Tyr Glu Glu Tyr Leu Thr His Ser Gly Thr Ile Val
1 5 10 15
Lys Pro Tyr Tyr Trp Leu Ser Val Ser Leu Asn Cys Thr Asn Cys Pro

20 25 30
Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Pro Tyr Thr Glu Phe
35 40 45
His Gln Thr Phe Gly Phe Pro Trp Ser Thr Tyr Pro Gln Thr Lys His
50 55 60
Leu Thr Phe Tyr Glu Leu Arg Ser Ser Ser Gly Asn Leu Ile Gln Lys
65 70 75 80
Gly Leu Ala Ser Asn Cys Thr Gly Ser His Thr His Pro Glu Ser Met

85 90 95
Leu Phe Glu Arg Asp Gly Tyr Leu Asp Ser Leu Ile Phe His Asp Ser
100 105 110
Asn Ile Arg His Ile Ile Leu Tyr Ser Asn Asn Ser Pro Cys Asp Glu

115 120 125

Ala Asn His Cys Cys Ile Ser Lys Met Tyr Asn Phe Leu Met Asn Tyr

130 135 140

Pro Glu Val Thr Leu Ser Val Phe Phe Ser Gln Leu Tyr His Thr Glu

145 150 155 160

Asn Gln Phe Pro Thr Ser Ala Trp Asn Arg Glu Ala Leu Arg Gly Leu

165 170 175

Ala Ser Leu Trp Pro Gln Val Thr Leu Ser Ala Ile Ser Gly Gly Ile

180 185 190

Trp Gln Ser Ile Leu Glu Thr Phe Val Ser Gly Ile Ser Glu Gly Leu

195 200 205

Thr Ala Val Arg Pro Phe Thr Ala Gly Arg Thr Leu Thr Asp Arg Tyr

210 215 220

Asn Ala Tyr Glu Ile Asn Cys Ile Thr Glu Val Lys Pro Tyr Phe Thr

225 230 235 240

Asp Ala Leu His Ser Trp Gln Lys Glu Asn Gln Asp Gln Lys Val Trp

245 250 255

Ala Ala Ser Glu Asn Gln Pro Leu His Asn Thr Thr Pro Ala Gln Trp

260 265 270

Gln Pro Asp Met Ser Gln Asp Cys Arg Thr Pro Ala Val Phe Met Leu

275 280 285

Val Pro Tyr Arg Asp Leu Pro Pro Ile His Val Asn Pro Ser Pro Gln

290 295 300

Lys Pro Arg Thr Val Val Arg His Leu Asn Thr Leu Gln Leu Ser Ala

305 310 315 320

Ser Lys Val Lys Ala Leu Arg Lys Ser Pro Ser Gly Arg Pro Val Lys

325 330 335

Lys Glu Glu Ala Arg Lys Gly Ser Thr Arg Ser Gln Glu Ala Asn Glu

340 345 350

Thr Asn Lys Ser Lys Trp Lys Lys Gln Thr Leu Phe Ile Lys Ser Asn

355 360 365

Ile Cys His Leu Leu Glu Arg Glu Gln Lys Lys Ile Gly Ile Leu Ser
370 375 380

Ser Trp Ser Val
385

<210> 75

<211> 363

<212> PRT

<213> Macaca fascicularis

<400> 75

Met Glu Pro Thr Tyr Glu Glu Tyr Leu Ala Asn His Gly Thr Ile Val
1 5 10 15

Lys Pro Tyr Tyr Trp Leu Ser Phe Ser Leu Asp Cys Ser Asn Cys Pro
20 25 30

Tyr His Ile Arg Thr Gly Glu Glu Ala Arg Val Ser Leu Thr Glu Phe
35 40 45

Cys Gln Ile Phe Gly Phe Pro Tyr Gly Thr Thr Tyr Pro Gln Thr Lys
50 55 60

His Leu Thr Phe Tyr Glu Leu Lys Thr Ser Ser Gly Ser Leu Val Gln
65 70 75 80

Lys Gly His Ala Ser Ser Cys Thr Gly Asn Tyr Ile His Pro Glu Ser
85 90 95

Met Leu Phe Glu Met Asn Gly Tyr Leu Asp Ser Ala Ile Tyr Asn Asn
100 105 110

Asp Ser Ile Arg His Ile Ile Leu Tyr Cys Asn Asn Ser Pro Cys Asn
115 120 125

Glu Ala Asn His Cys Cys Ile Ser Lys Val Tyr Asn Phe Leu Ile Thr
130 135 140

Tyr Pro Gly Ile Thr Leu Ser Ile Tyr Phe Ser Gln Leu Tyr His Thr
145 150 155 160

Glu Met Asp Phe Pro Ala Ser Ala Trp Asn Arg Glu Ala Leu Arg Ser
165 170 175

Leu Ala Ser Leu Trp Pro Arg Val Val Leu Ser Pro Ile Ser Gly Gly

180 185 190

Ile Trp His Ser Val Leu His Ser Phe Val Ser Gly Val Ser Gly Ser

195 200 205

His Val Phe Gln Pro Ile Leu Thr Gly Arg Ala Leu Thr Asp Arg Tyr

210 215 220

Asn Ala Tyr Glu Ile Asn Ala Ile Thr Gly Val Lys Pro Phe Phe Thr

225 230 235 240

Asp Val Leu Leu His Thr Lys Arg Asn Pro Asn Thr Lys Ala Gln Met

245 250 255

Ala Leu Glu Ser Tyr Pro Leu Asn Asn Ala Phe Pro Gly Gln Ser Phe

260 265 270

Gln Met Thr Ser Gly Ile Pro Pro Asp Leu Arg Ala Pro Val Val Phe

275 280 285

Val Leu Leu Pro Leu Arg Asp Leu Pro Pro Met His Met Gly Gln Asp

290 295 300

Pro Asn Lys Pro Arg Asn Ile Ile Arg His Leu Asn Met Pro Gln Met

305 310 315 320

Ser Phe Gln Glu Thr Lys Asp Leu Glu Arg Leu Pro Thr Arg Arg Ser

325 330 335

Val Glu Thr Val Glu Ile Thr Glu Arg Phe Ala Ser Ser Lys Gln Ala

340 345 350

Glu Glu Lys Thr Lys Lys Lys Lys Gly Lys Lys

355 360

<210> 76

<211> 224

<212> PRT

<213> Petromyzon marinus

<400> 76

Met Ala Gly Tyr Glu Cys Val Arg Val Ser Glu Lys Leu Asp Phe Asp

1 5 10 15

Thr Phe Glu Phe Gln Phe Glu Asn Leu His Tyr Ala Thr Glu Arg His

20 25 30

Arg Thr Tyr Val Ile Phe Asp Val Lys Pro Gln Ser Ala Gly Gly Arg
35 40 45
Ser Arg Arg Leu Trp Gly Tyr Ile Ile Asn Asn Pro Asn Val Cys His
50 55 60
Ala Glu Leu Ile Leu Met Ser Met Ile Asp Arg His Leu Glu Ser Asn
65 70 75 80
Pro Gly Val Tyr Ala Met Thr Trp Tyr Met Ser Trp Ser Pro Cys Ala
85 90 95

Asn Cys Ser Ser Lys Leu Asn Pro Trp Leu Lys Asn Leu Leu Glu Glu
100 105 110
Gln Gly His Thr Leu Thr Met His Phe Ser Arg Ile Tyr Asp Arg Asp
115 120 125
Arg Glu Gly Asp His Arg Gly Leu Arg Gly Leu Lys His Val Ser Asn
130 135 140
Ser Phe Arg Met Gly Val Val Gly Arg Ala Glu Val Lys Glu Cys Leu
145 150 155 160

Ala Glu Tyr Val Glu Ala Ser Arg Arg Thr Leu Thr Trp Leu Asp Thr
165 170 175
Thr Glu Ser Met Ala Ala Lys Met Arg Arg Lys Leu Phe Cys Ile Leu
180 185 190
Val Arg Cys Ala Gly Met Arg Glu Ser Gly Ile Pro Leu His Leu Phe
195 200 205
Thr Leu Gln Thr Pro Leu Leu Ser Gly Arg Val Val Trp Trp Arg Val
210 215 220

<210> 77
<211> 331
<212> PRT
<213> Petromyzon marinus
<400> 77
Met Glu Leu Arg Glu Val Val Asp Cys Ala Leu Ala Ser Cys Val Arg

1 5 10 15

His Glu Pro Leu Ser Arg Val Ala Phe Leu Arg Cys Phe Ala Ala Pro

20 25 30

Ser Gln Lys Pro Arg Gly Thr Val Ile Leu Phe Tyr Val Glu Gly Ala

35 40 45

Gly Arg Gly Val Thr Gly Gly His Ala Val Asn Tyr Asn Lys Gln Gly

50 55 60

Thr Ser Ile His Ala Glu Val Leu Leu Leu Ser Ala Val Arg Ala Ala

65 70 75 80

Leu Leu Arg Arg Arg Arg Cys Glu Asp Gly Glu Glu Ala Thr Arg Gly

85 90 95

Cys Thr Leu His Cys Tyr Ser Thr Tyr Ser Pro Cys Arg Asp Cys Val

100 105 110

Glu Tyr Ile Gln Glu Phe Gly Ala Ser Thr Gly Val Arg Val Val Ile

115 120 125

His Cys Cys Arg Leu Tyr Glu Leu Asp Val Asn Arg Arg Arg Ser Glu

130 135 140

Ala Glu Gly Val Leu Arg Ser Leu Ser Arg Leu Gly Arg Asp Phe Arg

145 150 155 160

Leu Met Gly Pro Arg Asp Ala Ile Ala Leu Leu Leu Gly Gly Arg Leu

165 170 175

Ala Asn Thr Ala Asp Gly Glu Ser Gly Ala Ser Gly Asn Ala Trp Val

180 185 190

Thr Glu Thr Asn Val Val Glu Pro Leu Val Asp Met Thr Gly Phe Gly

195 200 205

Asp Glu Asp Leu His Ala Gln Val Gln Arg Asn Lys Gln Ile Arg Glu

210 215 220

Ala Tyr Ala Asn Tyr Ala Ser Ala Val Ser Leu Met Leu Gly Glu Leu

225 230 235 240

His Val Asp Pro Asp Lys Phe Pro Phe Leu Ala Glu Phe Leu Ala Gln

245 250 255

Thr Ser Val Glu Pro Ser Gly Thr Pro Arg Glu Thr Arg Gly Arg Pro
260 265 270
Arg Gly Ala Ser Ser Arg Gly Pro Glu Ile Gly Arg Gln Arg Pro Ala
275 280 285
Asp Phe Glu Arg Ala Leu Gly Ala Tyr Gly Leu Phe Leu His Pro Arg
290 295 300
Ile Val Ser Arg Glu Ala Asp Arg Glu Glu Ile Lys Arg Asp Leu Ile

305 310 315 320
Val Val Met Arg Lys His Asn Tyr Gln Gly Pro
325 330

<210> 78
<211> 209
<212> PRT
<213> Petromyzon marinus
<400> 78
Met Ala Gly Asp Glu Asn Val Arg Val Ser Glu Lys Leu Asp Phe Asp
1 5 10 15
Thr Phe Glu Phe Gln Phe Glu Asn Leu His Tyr Ala Thr Glu Arg His
20 25 30
Arg Thr Tyr Val Ile Phe Asp Val Lys Pro Gln Ser Ala Gly Gly Arg

35 40 45
Ser Arg Arg Leu Trp Gly Tyr Ile Ile Asn Asn Pro Asn Val Cys His
50 55 60
Ala Glu Leu Ile Leu Met Ser Met Ile Asp Arg His Leu Glu Ser Asn
65 70 75 80
Pro Gly Val Tyr Ala Met Thr Trp Tyr Met Ser Trp Ser Pro Cys Ala
85 90 95
Asn Cys Ser Ser Lys Leu Asn Pro Trp Leu Lys Asn Leu Leu Glu Glu

100 105 110
Gln Gly His Thr Leu Met Met His Phe Ser Arg Ile Tyr Asp Arg Asp
115 120 125
Arg Glu Gly Asp His Arg Gly Leu Arg Gly Leu Lys His Val Ser Asn

130 135 140

Ser Phe Arg Met Gly Val Val Gly Arg Ala Glu Val Lys Glu Cys Leu

145 150 155 160

Ala Glu Tyr Val Glu Ala Ser Arg Arg Thr Leu Thr Trp Leu Asp Thr

165 170 175

Thr Glu Ser Met Ala Ala Lys Met Arg Arg Lys Leu Phe Cys Ile Leu

180 185 190

Val Arg Cys Ala Gly Met Arg Glu Ser Gly Met Pro Leu His Leu Phe

195 200 205

Thr

<210> 79

<211> 158

<212> PRT

<213> Saccharomyces cerevisiae

<400> 79

Met Val Thr Gly Gly Met Ala Ser Lys Trp Asp Gln Lys Gly Met Asp

1 5 10 15

Ile Ala Tyr Glu Glu Ala Ala Leu Gly Tyr Lys Glu Gly Gly Val Pro

20 25 30

Ile Gly Gly Cys Leu Ile Asn Asn Lys Asp Gly Ser Val Leu Gly Arg

35 40 45

Gly His Asn Met Arg Phe Gln Lys Gly Ser Ala Thr Leu His Gly Glu

50 55 60

Ile Ser Thr Leu Glu Asn Cys Gly Arg Leu Glu Gly Lys Val Tyr Lys

65 70 75 80

Asp Thr Thr Leu Tyr Thr Thr Leu Ser Pro Cys Asp Met Cys Thr Gly

85 90 95

Ala Ile Ile Met Tyr Gly Ile Pro Arg Cys Val Val Gly Glu Asn Val

100 105 110

Asn Phe Lys Ser Lys Gly Glu Lys Tyr Leu Gln Thr Arg Gly His Glu

115 120 125

Val Val Val Val Asp Asp Glu Arg Cys Lys Lys Ile Met Lys Gln Phe
130 135 140

Ile Asp Glu Arg Pro Gln Asp Trp Phe Glu Asp Ile Gly Glu
145 150 155

<210> 80

<211> 218

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 80

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15

Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30

Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60

Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80

Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95

Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125

Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Gly Leu Pro Pro Cys Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu
180 185 190
Thr Phe Phe Thr Ile Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro
195 200 205
Pro His Ile Leu Trp Ala Thr Gly Leu Lys
210 215

<210> 81

<211> 218

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 81

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45

Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110

Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met

130 135 140

Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160

Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175

Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190

Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln His Tyr Gln Arg Leu Pro
195 200 205

Pro His Ile Leu Trp Ala Thr Gly Leu Lys
210 215

<210> 82

<211> 429

<212> PRT

<213> Mus sp.

<400> 82

Met Gly Pro Phe Cys Leu Gly Cys Ser His Arg Lys Cys Tyr Ser Pro
1 5 10 15

Ile Arg Asn Leu Ile Ser Gln Glu Thr Phe Lys Phe His Phe Lys Asn
20 25 30

Leu Gly Tyr Ala Lys Gly Arg Lys Asp Thr Phe Leu Cys Tyr Glu Val
35 40 45

Thr Arg Lys Asp Cys Asp Ser Pro Val Ser Leu His His Gly Val Phe
50 55 60

Lys Asn Lys Asp Asn Ile His Ala Glu Ile Cys Phe Leu Tyr Trp Phe
65 70 75 80

His Asp Lys Val Leu Lys Val Leu Ser Pro Arg Glu Glu Phe Lys Ile
85 90 95

Thr Trp Tyr Met Ser Trp Ser Pro Cys Phe Glu Cys Ala Glu Gln Ile
100 105 110

Val Arg Phe Leu Ala Thr His His Asn Leu Ser Leu Asp Ile Phe Ser
115 120 125

Ser Arg Leu Tyr Asn Val Gln Asp Pro Glu Thr Gln Gln Asn Leu Cys
130 135 140

Arg Leu Val Gln Glu Gly Ala Gln Val Ala Ala Met Asp Leu Tyr Glu
145 150 155 160
Phe Lys Lys Cys Trp Lys Lys Phe Val Asp Asn Gly Gly Arg Arg Phe
165 170 175
Arg Pro Trp Lys Arg Leu Leu Thr Asn Phe Arg Tyr Gln Asp Ser Lys
180 185 190
Leu Gln Glu Ile Leu Arg Pro Cys Tyr Ile Pro Val Pro Ser Ser Ser
195 200 205

Ser Ser Thr Leu Ser Asn Ile Cys Leu Thr Lys Gly Leu Pro Glu Thr
210 215 220
Arg Phe Cys Val Glu Gly Arg Arg Met Asp Pro Leu Ser Glu Glu Glu
225 230 235 240
Phe Tyr Ser Gln Phe Tyr Asn Gln Arg Val Lys His Leu Cys Tyr Tyr
245 250 255
His Arg Met Lys Pro Tyr Leu Cys Tyr Gln Leu Glu Gln Phe Asn Gly
260 265 270

Gln Ala Pro Leu Lys Gly Cys Leu Leu Ser Glu Lys Gly Lys Gln His
275 280 285
Ala Glu Ile Leu Phe Leu Asp Lys Ile Arg Ser Met Glu Leu Ser Gln
290 295 300
Val Thr Ile Thr Cys Tyr Leu Thr Trp Ser Pro Cys Pro Asn Cys Ala
305 310 315 320
Trp Gln Leu Ala Ala Phe Lys Arg Asp Arg Pro Asp Leu Ile Leu His
325 330 335

Ile Tyr Thr Ser Arg Leu Tyr Phe His Trp Lys Arg Pro Phe Gln Lys
340 345 350
Gly Leu Cys Ser Leu Trp Gln Ser Gly Ile Leu Val Asp Val Met Asp
355 360 365
Leu Pro Gln Phe Thr Asp Cys Trp Thr Asn Phe Val Asn Pro Lys Arg

370 375 380

Pro Phe Trp Pro Trp Lys Gly Leu Glu Ile Ile Ser Arg Arg Thr Gln
385 390 395 400

Arg Arg Leu Arg Arg Ile Lys Glu Ser Trp Gly Leu Gln Asp Leu Val
405 410 415

Asn Asp Phe Gly Asn Leu Gln Leu Gly Pro Pro Met Ser
420 425

<210> 83

<211> 167

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 83

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu
1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala
20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala
35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg
50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Ile Asp Ala Thr Leu
65 70 75 80

Tyr Val Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His
85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly
100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His Tyr Pro Gly Met Asn His
115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Tyr Phe Phe Arg Met Pro Arg Gln Val Phe Asn Ala Gln Lys
145 150 155 160
Lys Ala Gln Ser Ser Thr Asp
165

<210> 84
<211> 1378
<212> PRT
<213> Lactobacillus fermentum
<400> 84

Met Lys Glu Tyr His Ile Gly Leu Asp Ile Gly Thr Ser Ser Ile Gly
1 5 10 15
Trp Ala Val Thr Asp Ser Gln Phe Lys Leu Met Arg Ile Lys Gly Lys
20 25 30

Thr Ala Ile Gly Val Arg Leu Phe Glu Glu Gly Lys Thr Ala Ala Glu
35 40 45
Arg Arg Thr Phe Arg Thr Thr Arg Arg Arg Leu Lys Arg Arg Lys Trp
50 55 60
Arg Leu His Tyr Leu Asp Glu Ile Phe Ala Pro His Leu Gln Glu Val
65 70 75 80
Asp Glu Asn Phe Leu Arg Arg Leu Lys Gln Ser Asn Ile His Pro Glu
85 90 95

Asp Pro Ala Lys Asn Gln Ala Phe Ile Gly Lys Leu Leu Phe Pro Asp
100 105 110
Leu Leu Lys Lys Asn Glu Arg Gly Tyr Pro Thr Leu Ile Lys Met Arg
115 120 125
Asp Glu Leu Pro Val Glu Gln Arg Ala His Tyr Pro Val Thr Asn Ile
130 135 140
Tyr Lys Leu Arg Glu Ala Met Ile Asn Glu Asp Arg Gln Phe Asp Leu
145 150 155 160

Arg Glu Val Tyr Leu Ala Val His His Ile Val Lys Tyr Arg Gly His

165 170 175

Phe Leu Asn Asn Ala Ser Val Asp Lys Phe Lys Val Gly Arg Ile Asp

180 185 190

Phe Asp Lys Ser Phe Asn Val Leu Asn Glu Ala Tyr Glu Glu Leu Gln

195 200 205

Asn Gly Glu Gly Ser Phe Thr Ile Glu Pro Ser Lys Val Glu Lys Ile

210 215 220

Gly Gln Leu Leu Leu Asp Thr Lys Met Arg Lys Leu Asp Arg Gln Lys

225 230 235 240

Ala Val Ala Lys Leu Leu Glu Val Lys Val Ala Asp Lys Glu Glu Thr

245 250 255

Lys Arg Asn Lys Gln Ile Ala Thr Ala Met Ser Lys Leu Val Leu Gly

260 265 270

Tyr Lys Ala Asp Phe Ala Thr Val Ala Met Ala Asn Gly Asn Glu Trp

275 280 285

Lys Ile Asp Leu Ser Ser Glu Thr Ser Glu Asp Glu Ile Glu Lys Phe

290 295 300

Arg Glu Glu Leu Ser Asp Ala Gln Asn Asp Ile Leu Thr Glu Ile Thr

305 310 315 320

Ser Leu Phe Ser Gln Ile Met Leu Asn Glu Ile Val Pro Asn Gly Met

325 330 335

Ser Ile Ser Glu Ser Met Met Asp Arg Tyr Trp Thr His Glu Arg Gln

340 345 350

Leu Ala Glu Val Lys Glu Tyr Leu Ala Thr Gln Pro Ala Ser Ala Arg

355 360 365

Lys Glu Phe Asp Gln Val Tyr Asn Lys Tyr Ile Gly Gln Ala Pro Lys

370 375 380

Glu Lys Gly Phe Asp Leu Glu Lys Gly Leu Lys Lys Ile Leu Ser Lys

385 390 395 400

Lys Glu Asn Trp Lys Glu Ile Asp Glu Leu Leu Lys Ala Gly Asp Phe

405 410 415

Leu Pro Lys Gln Arg Thr Ser Ala Asn Gly Val Ile Pro His Gln Met
420 425 430

His Gln Gln Glu Leu Asp Arg Ile Ile Glu Lys Gln Ala Lys Tyr Tyr
435 440 445

Pro Trp Leu Ala Thr Glu Asn Pro Ala Thr Gly Glu Arg Asp Arg His
450 455 460

Gln Ala Lys Tyr Glu Leu Asp Gln Leu Val Ser Phe Arg Ile Pro Tyr
465 470 475 480

Tyr Val Gly Pro Leu Val Thr Pro Glu Val Gln Lys Ala Thr Ser Gly
485 490 495

Ala Lys Phe Ala Trp Ala Lys Arg Lys Glu Asp Gly Glu Ile Thr Pro
500 505 510

Trp Asn Leu Trp Asp Lys Ile Asp Arg Ala Glu Ser Ala Glu Ala Phe
515 520 525

Ile Lys Arg Met Thr Val Lys Asp Thr Tyr Leu Leu Asn Glu Asp Val
530 535 540

Leu Pro Ala Asn Ser Leu Leu Tyr Gln Lys Tyr Asn Val Leu Asn Glu
545 550 555 560

Leu Asn Asn Val Arg Val Asn Gly Arg Arg Leu Ser Val Gly Ile Lys
565 570 575

Gln Asp Ile Tyr Thr Glu Leu Phe Lys Lys Lys Lys Thr Val Lys Ala
580 585 590

Gly Asp Val Ala Ser Leu Val Met Ala Lys Thr Arg Gly Val Asn Lys
595 600 605

Pro Ser Val Glu Gly Leu Ser Asp Pro Lys Lys Phe Asn Ser Asn Leu
610 615 620

Ala Thr Tyr Leu Asp Leu Lys Ser Ile Val Gly Asp Lys Val Asp Asp
625 630 635 640

Asn Arg Tyr Gln Met Asp Leu Glu Asn Ile Ile Glu Trp Arg Ser Val
645 650 655

Phe Glu Asp Gly Glu Ile Phe Ala Asp Lys Leu Thr Glu Val Glu Trp

660 665 670

Leu Thr Asp Glu Gln Arg Ser Ala Leu Val Lys Lys Arg Tyr Lys Gly
675 680 685
Trp Gly Arg Leu Ser Lys Lys Leu Leu Thr Gly Ile Val Asp Glu Asn
690 695 700
Gly Gln Arg Ile Ile Asp Leu Met Trp Asn Thr Asp Gln Asn Phe Met
705 710 715 720
Gln Ile Val Asn Gln Pro Val Phe Lys Glu Gln Ile Asp Gln Leu Asn
725 730 735

Gln Lys Ala Ile Thr Asn Asp Gly Met Thr Leu Arg Glu Arg Val Glu
740 745 750
Ser Val Leu Asp Asp Ala Tyr Thr Ser Pro Gln Asn Lys Lys Ala Ile
755 760 765
Trp Gln Val Val Arg Val Val Glu Asp Ile Val Lys Ala Val Gly Asn
770 775 780
Ala Pro Lys Ser Ile Ser Ile Glu Phe Ala Arg Asn Glu Gly Asn Lys
785 790 795 800

Gly Glu Ile Thr Arg Ser Arg Arg Thr Gln Leu Gln Lys Leu Phe Glu
805 810 815
Asp Gln Ala His Glu Leu Val Lys Asp Thr Ser Leu Thr Glu Glu Leu
820 825 830
Glu Lys Ala Pro Asp Leu Ser Asp Arg Tyr Tyr Phe Tyr Phe Thr Gln
835 840 845
Gly Gly Lys Asp Met Tyr Thr Gly Asp Pro Ile Asn Phe Asp Glu Ile
850 855 860

Ser Thr Lys Tyr Asp Ile Asp His Ile Leu Pro Gln Ser Phe Val Lys
865 870 875 880
Asp Asp Ser Leu Asp Asn Arg Val Leu Val Ser Arg Ala Glu Asn Asn
885 890 895
Lys Lys Ser Asp Arg Val Pro Ala Lys Leu Tyr Ala Ala Lys Met Lys
900 905 910

Pro Tyr Trp Asn Gln Leu Leu Lys Gln Gly Leu Ile Thr Gln Arg Lys
915 920 925

Phe Glu Asn Leu Thr Met Asp Val Asp Gln Thr Ile Lys Tyr Arg Ser
930 935 940
Leu Gly Phe Val Lys Arg Gln Leu Val Glu Thr Arg Gln Val Ile Lys
945 950 955 960
Leu Thr Ala Asn Ile Leu Gly Ser Met Tyr Gln Glu Ala Gly Thr Asp
965 970 975
Ile Ile Glu Thr Arg Ala Gly Leu Thr Lys Gln Leu Arg Glu Glu Phe
980 985 990

Asp Leu Pro Lys Val Arg Glu Val Asn Asp Tyr His His Ala Val Asp
995 1000 1005
Ala Tyr Leu Thr Thr Phe Ala Gly Gln Tyr Leu Asn Arg Arg Tyr
1010 1015 1020
Pro Lys Leu Arg Ser Phe Phe Val Tyr Gly Glu Tyr Met Lys Phe
1025 1030 1035
Lys His Gly Ser Asp Leu Lys Leu Arg Asn Phe Asn Phe Phe His
1040 1045 1050

Glu Leu Met Glu Gly Asp Lys Ser Gln Gly Lys Val Val Asp Gln
1055 1060 1065
Gln Thr Gly Glu Leu Ile Thr Thr Arg Asp Glu Val Ala Asp Tyr
1070 1075 1080
Phe Asp Trp Val Ile Asn Leu Lys Val Met Leu Ile Ser Asn Glu
1085 1090 1095
Thr Tyr Glu Glu Thr Gly Lys Tyr Phe Asp Ala Ser His Glu Ser
1100 1105 1110

Ser Ser Leu Tyr Leu Lys Asn Gln Asn Lys Lys Ser Lys Leu Val
1115 1120 1125
Val Pro Leu Lys Asn Lys Leu Gln Pro Glu Tyr Tyr Gly Ala Tyr
1130 1135 1140
Thr Gly Ile Thr Gln Gly Tyr Met Val Ile Leu Lys Leu Leu Asp

1145 1150 1155

Lys Lys Gly Gly Phe Gly Val Tyr Arg Ile Pro Arg Tyr Ala Ala

1160 1165 1170

Asp Ile Leu Asn Lys Cys His Asp Glu Val Ala Tyr Arg Asn Lys

1175 1180 1185

Ile Ala Glu Ile Ile Ser Ser Asp Pro Arg Ala Pro Lys Ser Phe

1190 1195 1200

Glu Val Val Val Pro Arg Val Leu Lys Gly Thr Phe Leu Val Asp

1205 1210 1215

Gly Glu Glu Lys Phe Ile Leu Ser Ser Tyr Arg Tyr Lys Val Asn

1220 1225 1230

Ala Thr Gln Leu Ile Leu Pro Val Ser Asp Ile Lys Leu Ile Gln

1235 1240 1245

Asp Asn Phe Lys Ala Leu Lys Lys Leu Asn Val Glu Met Gln Thr

1250 1255 1260

Lys Lys Leu Ile Glu Ile Tyr Asp Asn Ile Leu Arg Gln Val Asp

1265 1270 1275

Lys Tyr Tyr Lys Leu Tyr Asp Ile Asn Lys Phe Arg Ala Lys Leu

1280 1285 1290

His Asp Gly Arg Ser Lys Phe Val Glu Leu Asp Asp Phe Gly Gln

1295 1300 1305

Asp Ala Ser Lys Glu Lys Val Ile Ile Lys Ile Leu Arg Gly Leu

1310 1315 1320

His Phe Gly Ser Asp Leu Gln Asn Leu Lys Glu Ile Gly Phe Gly

1325 1330 1335

Thr Thr Pro Leu Gly Gln Phe Gln Val Ser Glu Ala Gly Ile Arg

1340 1345 1350

Leu Ser Asn Thr Ala Phe Ile Ile Phe Lys Ser Pro Thr Gly Leu

1355 1360 1365

Phe Asn Arg Lys Leu Tyr Leu Lys Asn Leu

1370 1375

<210> 85

<211> 1432

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 85

Met Pro Lys Lys Lys Arg Lys Val Gly Lys Glu Tyr His Ile Gly Leu
1 5 10 15
Asp Ile Gly Thr Ser Ser Ile Gly Trp Ala Val Thr Asp Ser Gln Phe

20 25 30
Lys Leu Met Arg Ile Lys Gly Lys Thr Ala Ile Gly Val Arg Leu Phe
35 40 45
Glu Glu Gly Lys Thr Ala Ala Glu Arg Arg Thr Phe Arg Thr Thr Arg
50 55 60
Arg Arg Leu Lys Arg Arg Lys Trp Arg Leu His Tyr Leu Asp Glu Ile
65 70 75 80
Phe Ala Pro His Leu Gln Glu Val Asp Glu Asn Phe Leu Arg Arg Leu

85 90 95
Lys Gln Ser Asn Ile His Pro Glu Asp Pro Ala Lys Asn Gln Ala Phe
100 105 110
Ile Gly Lys Leu Leu Phe Pro Asp Leu Leu Lys Lys Asn Glu Arg Gly
115 120 125
Tyr Pro Thr Leu Ile Lys Met Arg Asp Glu Leu Pro Val Glu Gln Arg
130 135 140
Ala His Tyr Pro Val Thr Asn Ile Tyr Lys Leu Arg Glu Ala Met Ile

145 150 155 160
Asn Glu Asp Arg Gln Phe Asp Leu Arg Glu Val Tyr Leu Ala Val His
165 170 175
His Ile Val Lys Tyr Arg Gly His Phe Leu Asn Asn Ala Ser Val Asp
180 185 190
Lys Phe Lys Val Gly Arg Ile Asp Phe Asp Lys Ser Phe Asn Val Leu

195 200 205

Asn Glu Ala Tyr Glu Glu Leu Gln Asn Gly Glu Gly Ser Phe Thr Ile

210 215 220

Glu Pro Ser Lys Val Glu Lys Ile Gly Gln Leu Leu Leu Asp Thr Lys

225 230 235 240

Met Arg Lys Leu Asp Arg Gln Lys Ala Val Ala Lys Leu Leu Glu Val

245 250 255

Lys Val Ala Asp Lys Glu Glu Thr Lys Arg Asn Lys Gln Ile Ala Thr

260 265 270

Ala Met Ser Lys Leu Val Leu Gly Tyr Lys Ala Asp Phe Ala Thr Val

275 280 285

Ala Met Ala Asn Gly Asn Glu Trp Lys Ile Asp Leu Ser Ser Glu Thr

290 295 300

Ser Glu Asp Glu Ile Glu Lys Phe Arg Glu Glu Leu Ser Asp Ala Gln

305 310 315 320

Asn Asp Ile Leu Thr Glu Ile Thr Ser Leu Phe Ser Gln Ile Met Leu

325 330 335

Asn Glu Ile Val Pro Asn Gly Met Ser Ile Ser Glu Ser Met Met Asp

340 345 350

Arg Tyr Trp Thr His Glu Arg Gln Leu Ala Glu Val Lys Glu Tyr Leu

355 360 365

Ala Thr Gln Pro Ala Ser Ala Arg Lys Glu Phe Asp Gln Val Tyr Asn

370 375 380

Lys Tyr Ile Gly Gln Ala Pro Lys Glu Lys Gly Phe Asp Leu Glu Lys

385 390 395 400

Gly Leu Lys Lys Ile Leu Ser Lys Lys Glu Asn Trp Lys Glu Ile Asp

405 410 415

Glu Leu Leu Lys Ala Gly Asp Phe Leu Pro Lys Gln Arg Thr Ser Ala

420 425 430

Asn Gly Val Ile Pro His Gln Met His Gln Gln Glu Leu Asp Arg Ile

435 440 445

Ile Glu Lys Gln Ala Lys Tyr Tyr Pro Trp Leu Ala Thr Glu Asn Pro

450 455 460

Ala Thr Gly Glu Arg Asp Arg His Gln Ala Lys Tyr Glu Leu Asp Gln

465 470 475 480

Leu Val Ser Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Val Thr Pro

485 490 495

Glu Val Gln Lys Ala Thr Ser Gly Ala Lys Phe Ala Trp Ala Lys Arg

500 505 510

Lys Glu Asp Gly Glu Ile Thr Pro Trp Asn Leu Trp Asp Lys Ile Asp

515 520 525

Arg Ala Glu Ser Ala Glu Ala Phe Ile Lys Arg Met Thr Val Lys Asp

530 535 540

Thr Tyr Leu Leu Asn Glu Asp Val Leu Pro Ala Asn Ser Leu Leu Tyr

545 550 555 560

Gln Lys Tyr Asn Val Leu Asn Glu Leu Asn Asn Val Arg Val Asn Gly

565 570 575

Arg Arg Leu Ser Val Gly Ile Lys Gln Asp Ile Tyr Thr Glu Leu Phe

580 585 590

Lys Lys Lys Lys Thr Val Lys Ala Gly Asp Val Ala Ser Leu Val Met

595 600 605

Ala Lys Thr Arg Gly Val Asn Lys Pro Ser Val Glu Gly Leu Ser Asp

610 615 620

Pro Lys Lys Phe Asn Ser Asn Leu Ala Thr Tyr Leu Asp Leu Lys Ser

625 630 635 640

Ile Val Gly Asp Lys Val Asp Asp Asn Arg Tyr Gln Met Asp Leu Glu

645 650 655

Asn Ile Ile Glu Trp Arg Ser Val Phe Glu Asp Gly Glu Ile Phe Ala

660 665 670

Asp Lys Leu Thr Glu Val Glu Trp Leu Thr Asp Glu Gln Arg Ser Ala

675 680 685

Leu Val Lys Lys Arg Tyr Lys Gly Trp Gly Arg Leu Ser Lys Lys Leu

690 695 700

Leu Thr Gly Ile Val Asp Glu Asn Gly Gln Arg Ile Ile Asp Leu Met

705 710 715 720

Trp Asn Thr Asp Gln Asn Phe Met Gln Ile Val Asn Gln Pro Val Phe

725 730 735

Lys Glu Gln Ile Asp Gln Leu Asn Gln Lys Ala Ile Thr Asn Asp Gly

740 745 750

Met Thr Leu Arg Glu Arg Val Glu Ser Val Leu Asp Asp Ala Tyr Thr

755 760 765

Ser Pro Gln Asn Lys Lys Ala Ile Trp Gln Val Val Arg Val Val Glu

770 775 780

Asp Ile Val Lys Ala Val Gly Asn Ala Pro Lys Ser Ile Ser Ile Glu

785 790 795 800

Phe Ala Arg Asn Glu Gly Asn Lys Gly Glu Ile Thr Arg Ser Arg Arg

805 810 815

Thr Gln Leu Gln Lys Leu Phe Glu Asp Gln Ala His Glu Leu Val Lys

820 825 830

Asp Thr Ser Leu Thr Glu Glu Leu Glu Lys Ala Pro Asp Leu Ser Asp

835 840 845

Arg Tyr Tyr Phe Tyr Phe Thr Gln Gly Gly Lys Asp Met Tyr Thr Gly

850 855 860

Asp Pro Ile Asn Phe Asp Glu Ile Ser Thr Lys Tyr Asp Ile Asp His

865 870 875 880

Ile Leu Pro Gln Ser Phe Val Lys Asp Asp Ser Leu Asp Asn Arg Val

885 890 895

Leu Val Ser Arg Ala Glu Asn Asn Lys Lys Ser Asp Arg Val Pro Ala

900 905 910

Lys Leu Tyr Ala Ala Lys Met Lys Pro Tyr Trp Asn Gln Leu Leu Lys

915 920 925

Gln Gly Leu Ile Thr Gln Arg Lys Phe Glu Asn Leu Thr Met Asp Val

930 935 940

Asp Gln Thr Ile Lys Tyr Arg Ser Leu Gly Phe Val Lys Arg Gln Leu
945 950 955 960
Val Glu Thr Arg Gln Val Ile Lys Leu Thr Ala Asn Ile Leu Gly Ser
965 970 975
Met Tyr Gln Glu Ala Gly Thr Asp Ile Ile Glu Thr Arg Ala Gly Leu

980 985 990
Thr Lys Gln Leu Arg Glu Glu Phe Asp Leu Pro Lys Val Arg Glu Val
995 1000 1005
Asn Asp Tyr His His Ala Val Asp Ala Tyr Leu Thr Thr Phe Ala
1010 1015 1020
Gly Gln Tyr Leu Asn Arg Arg Tyr Pro Lys Leu Arg Ser Phe Phe
1025 1030 1035
Val Tyr Gly Glu Tyr Met Lys Phe Lys His Gly Ser Asp Leu Lys

1040 1045 1050
Leu Arg Asn Phe Asn Phe Phe His Glu Leu Met Glu Gly Asp Lys
1055 1060 1065
Ser Gln Gly Lys Val Val Asp Gln Gln Thr Gly Glu Leu Ile Thr
1070 1075 1080
Thr Arg Asp Glu Val Ala Asp Tyr Phe Asp Trp Val Ile Asn Leu
1085 1090 1095
Lys Val Met Leu Ile Ser Asn Glu Thr Tyr Glu Glu Thr Gly Lys

1100 1105 1110
Tyr Phe Asp Ala Ser His Glu Ser Ser Ser Leu Tyr Leu Lys Asn
1115 1120 1125
Gln Asn Lys Lys Ser Lys Leu Val Val Pro Leu Lys Asn Lys Leu
1130 1135 1140
Gln Pro Glu Tyr Tyr Gly Ala Tyr Thr Gly Ile Thr Gln Gly Tyr
1145 1150 1155
Met Val Ile Leu Lys Leu Leu Asp Lys Lys Gly Gly Phe Gly Val

1160 1165 1170
Tyr Arg Ile Pro Arg Tyr Ala Ala Asp Ile Leu Asn Lys Cys His

1175 1180 1185

Asp Glu Val Ala Tyr Arg Asn Lys Ile Ala Glu Ile Ile Ser Ser

1190 1195 1200

Asp Pro Arg Ala Pro Lys Ser Phe Glu Val Val Val Pro Arg Val

1205 1210 1215

Leu Lys Gly Thr Phe Leu Val Asp Gly Glu Glu Lys Phe Ile Leu

1220 1225 1230

Ser Ser Tyr Arg Tyr Lys Val Asn Ala Thr Gln Leu Ile Leu Pro

1235 1240 1245

Val Ser Asp Ile Lys Leu Ile Gln Asp Asn Phe Lys Ala Leu Lys

1250 1255 1260

Lys Leu Asn Val Glu Met Gln Thr Lys Lys Leu Ile Glu Ile Tyr

1265 1270 1275

Asp Asn Ile Leu Arg Gln Val Asp Lys Tyr Tyr Lys Leu Tyr Asp

1280 1285 1290

Ile Asn Lys Phe Arg Ala Lys Leu His Asp Gly Arg Ser Lys Phe

1295 1300 1305

Val Glu Leu Asp Asp Phe Gly Gln Asp Ala Ser Lys Glu Lys Val

1310 1315 1320

Ile Ile Lys Ile Leu Arg Gly Leu His Phe Gly Ser Asp Leu Gln

1325 1330 1335

Asn Leu Lys Glu Ile Gly Phe Gly Thr Thr Pro Leu Gly Gln Phe

1340 1345 1350

Gln Val Ser Glu Ala Gly Ile Arg Leu Ser Asn Thr Ala Phe Ile

1355 1360 1365

Ile Phe Lys Ser Pro Thr Gly Leu Phe Asn Arg Lys Leu Tyr Leu

1370 1375 1380

Lys Asn Leu Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln

1385 1390 1395

Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr

1400 1405 1410

Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val
1415 1420 1425

Pro Asp Tyr Ala
1430

<210> 86

<211> 1359

<212> PRT

<213> Peptoniphilus sp.

<400> 86

Met Glu Lys Lys Thr Asn Tyr Thr Ile Gly Leu Asp Ile Gly Thr Asp
1 5 10 15

Ser Val Gly Trp Ala Val Val Lys Asp Asp Leu Glu Leu Val Lys Lys
20 25 30

Arg Met Lys Val Leu Gly Asn Thr Glu Thr Asn Tyr Ile Lys Lys Asn
35 40 45

Leu Trp Gly Ser Leu Leu Phe Glu Ser Gly Gln Thr Ala Lys Asp Arg
50 55 60

Arg Leu Lys Arg Val Ala Arg Arg Arg Tyr Glu Arg Arg Arg Asn Arg
65 70 75 80

Leu Thr Glu Leu Gln Lys Ile Phe Ala Pro Ala Ile Asp Glu Val Asp
85 90 95

Glu Asn Phe Phe Phe Arg Leu Asn Glu Ser Phe Leu Val Pro Glu Asp
100 105 110

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

Thr Tyr Tyr Lys Thr Tyr Pro Thr Ile Tyr His Leu Arg Gln His Leu
130 135 140

Ala Asp Ser Glu Glu Lys Ala Asp Val Arg Leu Ile Tyr Leu Ala Leu
145 150 155 160

Ala His Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Lys Leu
165 170 175

Asp Thr Glu His Ile Ala Ile Asn Glu Asn Leu Glu Gln Phe Phe Glu

180 185 190

Ser Tyr Asn Ala Leu Phe Ser Glu Glu Pro Ile Glu Leu Arg Lys Glu

195 200 205

Glu Leu Ile Ala Ile Glu Asn Ile Leu Arg Glu Lys Asn Ser Arg Thr

210 215 220

Val Lys Glu Lys Arg Ile Thr Ser Phe Leu Lys Asp Ile Gly Arg Ala

225 230 235 240

Asn Lys Gln Ser Pro Met Met Ala Phe Ile Thr Leu Ile Val Gly Lys

245 250 255

Lys Ala Lys Phe Lys Ala Ala Phe Asn Leu Glu Glu Glu Ile Ser Leu

260 265 270

Asn Leu Thr Asp Asp Ser Tyr Asp Glu Asn Leu Glu Ile Leu Leu Asn

275 280 285

Thr Ile Gly Ser Asp Phe Ala Asp Leu Phe Asp His Ala Gln Arg Val

290 295 300

Tyr Asn Ala Val Glu Leu Ala Gly Ile Leu Ser Gly Asp Val Lys Asn

305 310 315 320

Thr His Ala Lys Leu Ser Ala Gln Met Val Ala Met Tyr Glu Arg His

325 330 335

Lys Glu Gln Leu Lys Glu Tyr Lys Ser Phe Ile Lys Ala Asn Leu Pro

340 345 350

Asp Gln Tyr Asp Met Thr Phe Val Ala Pro Lys Asp Ala Gln Lys Lys

355 360 365

Asp Leu Lys Gly Tyr Ala Gly Tyr Ile Asp Gly Asn Met Ser Gln Asp

370 375 380

Ser Phe Tyr Lys Phe Val Lys Asp Gln Leu Lys Glu Val Pro Gly Ser

385 390 395 400

Glu Lys Phe Leu Asp Ser Ile Glu Lys Glu Asp Phe Leu Arg Lys Gln

405 410 415

Arg Ser Phe Tyr Asn Gly Val Ile Pro Asn Gln Val His Leu Ala Glu

420 425 430

Met Glu Ala Ile Leu Asp Arg Gln Glu Asn Tyr Tyr Pro Trp Leu Lys
435 440 445
Glu Asn Arg Glu Lys Ile Ile Ser Leu Leu Thr Phe Arg Ile Pro Tyr
450 455 460
Tyr Val Gly Pro Leu Ala Asp Gly Gln Ser Glu Phe Ala Trp Leu Glu
465 470 475 480

Arg Lys Ser Asp Glu Lys Ile Lys Pro Trp Asn Phe Ser Asp Val Val
485 490 495
Asp Leu Asp Arg Ser Ala Glu Lys Phe Ile Glu Gln Leu Ile Gly Arg
500 505 510
Asp Thr Tyr Leu Pro Asp Glu Tyr Val Leu Pro Lys Lys Ser Leu Ile
515 520 525
Tyr Gln Lys Tyr Met Val Phe Asn Glu Leu Thr Lys Ile Ala Tyr Leu
530 535 540

Asp Glu Arg Gln Lys Arg Met Asn Leu Ser Ser Val Glu Lys Lys Glu
545 550 555 560
Ile Phe Glu Thr Leu Phe Lys Lys Arg Ser Lys Val Thr Glu Lys Gln
565 570 575
Leu Val Lys Phe Phe Glu Asn Tyr Leu Gln Ile Asp Asn Pro Thr Ile
580 585 590
Phe Gly Ile Glu Asp Ala Phe Asn Ala Asp Tyr Ser Thr Tyr Val Glu
595 600 605

Leu Ala Lys Val Pro Gly Met Lys Ser Met Met Asp Asp Pro Asp Asn
610 615 620
Glu Asp Leu Met Glu Glu Ile Val Lys Ile Leu Thr Val Phe Glu Asp
625 630 635 640
Arg Lys Met Arg Arg Lys Gln Leu Glu Lys Tyr Lys Glu Arg Leu Ser
645 650 655
Pro Glu Gln Ile Lys Glu Leu Ala Lys Lys His Tyr Thr Gly Trp Gly
660 665 670

Arg Leu Ser Lys Lys Leu Leu Val Gly Ile Arg Asp Lys Glu Thr Gln

675 680 685

Lys Thr Ile Leu Asp Tyr Leu Val Glu Asp Asp Asn His Ser Gly Gly

690 695 700

Arg Gln His Leu Asn Arg Asn Leu Met Gln Leu Ile Asn Asp Asp Arg

705 710 715 720

Leu Ser Phe Lys Lys Thr Ile Ala Glu Leu Gln Met Ile Asp Pro Ser

725 730 735

Ala Asp Leu Tyr Ala Gln Val Gln Glu Ile Ala Gly Ser Pro Ala Ile

740 745 750

Lys Lys Gly Ile Leu Leu Gly Leu Lys Ile Val Asp Glu Ile Ile Arg

755 760 765

Val Met Gly Glu Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu

770 775 780

Asn Gln Thr Thr Ala Arg Gly Lys Ala Leu Ser Lys Arg Arg Glu Ala

785 790 795 800

Lys Ile Lys Glu Gly Leu Ala Ala Leu Gly Ser Ser Leu Leu Lys Glu

805 810 815

Asn Leu Pro Gly Asn Ala Asp Leu Ser Gln Arg Lys Ile Tyr Leu Tyr

820 825 830

Tyr Thr Gln Asn Gly Lys Asp Ile Tyr Leu Asp Glu Pro Leu Asp Phe

835 840 845

Asp Arg Leu Ser Gln Tyr Asp Glu Asp His Ile Ile Pro Gln Ser Phe

850 855 860

Thr Val Asp Asn Ser Leu Asp Asn Leu Val Leu Thr Asn Ser Ser Gln

865 870 875 880

Asn Arg Gly Asn Lys Lys Asp Asp Val Pro Ser Leu Glu Val Val Asn

885 890 895

Arg Gln Leu Ala Tyr Trp Arg Ser Leu Lys Asp Ala Gly Leu Met Thr

900 905 910

Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Met Arg Gly Gly Leu Thr

915 920 925

Asp Lys Asp Arg Glu Arg Phe Ile Gln Arg Gln Leu Val Glu Thr Arg
930 935 940
Gln Ile Thr Lys Asn Val Ala Lys Leu Leu Asp Met Arg Leu Asn Asp
945 950 955 960
Lys Lys Asp Glu Ala Gly Asn Lys Ile Arg Glu Thr Asn Ile Val Leu
965 970 975
Leu Lys Ser Ala Met Ala Ser Glu Phe Arg Lys Met Phe Arg Leu Tyr
980 985 990

Lys Val Arg Glu Leu Asn Asp Tyr His His Ala His Asp Ala Tyr Leu
995 1000 1005
Asn Ala Ala Ile Ala Ile Asn Leu Leu Ala Leu Tyr Pro Tyr Met
1010 1015 1020
Ala Asp Asp Phe Val Tyr Gly Glu Phe Arg Tyr Lys Lys Lys Pro
1025 1030 1035
Gln Ala Glu Lys Ala Thr Tyr Glu Lys Leu Arg Gln Trp Asn Leu
1040 1045 1050

Ile Lys Arg Phe Gly Glu Lys Gln Leu Phe Thr Pro Asp His Glu
1055 1060 1065
Asp Cys Trp Asn Lys Glu Arg Asp Ile Lys Thr Ile Lys Lys Val
1070 1075 1080
Met Gly Tyr Arg Gln Val Asn Val Val Lys Lys Ala Glu Glu Arg
1085 1090 1095
Thr Gly Met Leu Phe Lys Glu Thr Ile Asn Gly Lys Thr Asn Lys
1100 1105 1110

Gly Ser Arg Ile Pro Ile Lys Lys Asp Leu Asp Pro Ser Lys Tyr
1115 1120 1125
Gly Gly Tyr Ile Glu Glu Lys Met Ala Tyr Tyr Ala Val Ile Ser
1130 1135 1140
Tyr Glu Asp Lys Lys Lys Lys Pro Gly Lys Thr Ile Val Gly Ile
1145 1150 1155
Ser Ile Met Asp Lys Lys Glu Phe Glu Tyr Asp Ser Ile Ser Tyr

1160 1165 1170

Leu Gly Lys Leu Gly Phe Ser Asn Pro Val Val Gln Ile Ile Leu
1175 1180 1185
Lys Asn Tyr Ser Leu Ile Ala Tyr Pro Asp Gly Arg Arg Arg Tyr
1190 1195 1200
Ile Thr Gly Ala Thr Lys Thr Thr Lys Gly Lys Val Glu Leu Gln
1205 1210 1215
Lys Ala Asn Gln Ile Ala Met Glu Gln Asp Leu Val Asn Phe Ile
1220 1225 1230

Tyr His Leu Lys Asn Tyr Asp Glu Ile Ser His Pro Glu Ser Tyr
1235 1240 1245
Ala Phe Val Gln Ser His Thr Asp Tyr Phe Asp Arg Leu Phe Asp
1250 1255 1260
Ser Ile Glu His Tyr Thr Arg Arg Phe Leu Asp Ala Glu Thr Asn
1265 1270 1275
Ile Asn Arg Leu Arg Arg Ile Tyr Glu Glu Glu Lys Lys Lys Asp
1280 1285 1290

Pro Val Asp Ile Glu Ala Leu Val Ala Ser Phe Ile Glu Leu Leu
1295 1300 1305
Lys Leu Thr Ser Ala Gly Ala Pro Ala Asp Phe Ile Phe Met Gly
1310 1315 1320
Glu Ala Ile Ser Arg Arg Arg Tyr Asn Ser Met Thr Gly Leu Phe
1325 1330 1335
Asp Gly Gln Val Ile Tyr Gln Ser Leu Thr Gly Leu Tyr Glu Thr
1340 1345 1350

Arg Met Arg Phe Glu Asp
1355

<210> 87
<211> 1413
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 87

Met Pro Lys Lys Lys Arg Lys Val Gly Glu Lys Lys Thr Asn Tyr Thr
1 5 10 15

Ile Gly Leu Asp Ile Gly Thr Asp Ser Val Gly Trp Ala Val Val Lys
20 25 30

Asp Asp Leu Glu Leu Val Lys Lys Arg Met Lys Val Leu Gly Asn Thr
35 40 45

Glu Thr Asn Tyr Ile Lys Lys Asn Leu Trp Gly Ser Leu Leu Phe Glu
50 55 60

Ser Gly Gln Thr Ala Lys Asp Arg Arg Leu Lys Arg Val Ala Arg Arg
65 70 75 80

Arg Tyr Glu Arg Arg Arg Asn Arg Leu Thr Glu Leu Gln Lys Ile Phe
85 90 95

Ala Pro Ala Ile Asp Glu Val Asp Glu Asn Phe Phe Phe Arg Leu Asn
100 105 110

Glu Ser Phe Leu Val Pro Glu Asp Lys Ala Phe Ser Lys Asn Pro Ile
115 120 125

Phe Gly Thr Leu Gly Glu Asp Lys Thr Tyr Tyr Lys Thr Tyr Pro Thr
130 135 140

Ile Tyr His Leu Arg Gln His Leu Ala Asp Ser Glu Glu Lys Ala Asp
145 150 155 160

Val Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly
165 170 175

His Phe Leu Ile Glu Gly Lys Leu Asp Thr Glu His Ile Ala Ile Asn
180 185 190

Glu Asn Leu Glu Gln Phe Phe Glu Ser Tyr Asn Ala Leu Phe Ser Glu
195 200 205

Glu Pro Ile Glu Leu Arg Lys Glu Glu Leu Ile Ala Ile Glu Asn Ile
210 215 220

Leu Arg Glu Lys Asn Ser Arg Thr Val Lys Glu Lys Arg Ile Thr Ser

225 230 235 240

Phe Leu Lys Asp Ile Gly Arg Ala Asn Lys Gln Ser Pro Met Met Ala

245 250 255

Phe Ile Thr Leu Ile Val Gly Lys Lys Ala Lys Phe Lys Ala Ala Phe

260 265 270

Asn Leu Glu Glu Glu Ile Ser Leu Asn Leu Thr Asp Asp Ser Tyr Asp

275 280 285

Glu Asn Leu Glu Ile Leu Leu Asn Thr Ile Gly Ser Asp Phe Ala Asp

290 295 300

Leu Phe Asp His Ala Gln Arg Val Tyr Asn Ala Val Glu Leu Ala Gly

305 310 315 320

Ile Leu Ser Gly Asp Val Lys Asn Thr His Ala Lys Leu Ser Ala Gln

325 330 335

Met Val Ala Met Tyr Glu Arg His Lys Glu Gln Leu Lys Glu Tyr Lys

340 345 350

Ser Phe Ile Lys Ala Asn Leu Pro Asp Gln Tyr Asp Met Thr Phe Val

355 360 365

Ala Pro Lys Asp Ala Gln Lys Lys Asp Leu Lys Gly Tyr Ala Gly Tyr

370 375 380

Ile Asp Gly Asn Met Ser Gln Asp Ser Phe Tyr Lys Phe Val Lys Asp

385 390 395 400

Gln Leu Lys Glu Val Pro Gly Ser Glu Lys Phe Leu Asp Ser Ile Glu

405 410 415

Lys Glu Asp Phe Leu Arg Lys Gln Arg Ser Phe Tyr Asn Gly Val Ile

420 425 430

Pro Asn Gln Val His Leu Ala Glu Met Glu Ala Ile Leu Asp Arg Gln

435 440 445

Glu Asn Tyr Tyr Pro Trp Leu Lys Glu Asn Arg Glu Lys Ile Ile Ser

450 455 460

Leu Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Asp Gly

465 470 475 480

Gln Ser Glu Phe Ala Trp Leu Glu Arg Lys Ser Asp Glu Lys Ile Lys

485 490 495

Pro Trp Asn Phe Ser Asp Val Val Asp Leu Asp Arg Ser Ala Glu Lys

500 505 510

Phe Ile Glu Gln Leu Ile Gly Arg Asp Thr Tyr Leu Pro Asp Glu Tyr

515 520 525

Val Leu Pro Lys Lys Ser Leu Ile Tyr Gln Lys Tyr Met Val Phe Asn

530 535 540

Glu Leu Thr Lys Ile Ala Tyr Leu Asp Glu Arg Gln Lys Arg Met Asn

545 550 555 560

Leu Ser Ser Val Glu Lys Lys Glu Ile Phe Glu Thr Leu Phe Lys Lys

565 570 575

Arg Ser Lys Val Thr Glu Lys Gln Leu Val Lys Phe Phe Glu Asn Tyr

580 585 590

Leu Gln Ile Asp Asn Pro Thr Ile Phe Gly Ile Glu Asp Ala Phe Asn

595 600 605

Ala Asp Tyr Ser Thr Tyr Val Glu Leu Ala Lys Val Pro Gly Met Lys

610 615 620

Ser Met Met Asp Asp Pro Asp Asn Glu Asp Leu Met Glu Glu Ile Val

625 630 635 640

Lys Ile Leu Thr Val Phe Glu Asp Arg Lys Met Arg Arg Lys Gln Leu

645 650 655

Glu Lys Tyr Lys Glu Arg Leu Ser Pro Glu Gln Ile Lys Glu Leu Ala

660 665 670

Lys Lys His Tyr Thr Gly Trp Gly Arg Leu Ser Lys Lys Leu Leu Val

675 680 685

Gly Ile Arg Asp Lys Glu Thr Gln Lys Thr Ile Leu Asp Tyr Leu Val

690 695 700

Glu Asp Asp Asn His Ser Gly Gly Arg Gln His Leu Asn Arg Asn Leu

705 710 715 720

Met Gln Leu Ile Asn Asp Asp Arg Leu Ser Phe Lys Lys Thr Ile Ala
725 730 735
Glu Leu Gln Met Ile Asp Pro Ser Ala Asp Leu Tyr Ala Gln Val Gln

740 745 750
Glu Ile Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Leu Gly Leu
755 760 765
Lys Ile Val Asp Glu Ile Ile Arg Val Met Gly Glu Lys Pro Glu Asn
770 775 780
Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Ala Arg Gly Lys
785 790 795 800
Ala Leu Ser Lys Arg Arg Glu Ala Lys Ile Lys Glu Gly Leu Ala Ala

805 810 815
Leu Gly Ser Ser Leu Leu Lys Glu Asn Leu Pro Gly Asn Ala Asp Leu
820 825 830
Ser Gln Arg Lys Ile Tyr Leu Tyr Tyr Thr Gln Asn Gly Lys Asp Ile
835 840 845
Tyr Leu Asp Glu Pro Leu Asp Phe Asp Arg Leu Ser Gln Tyr Asp Glu
850 855 860
Asp His Ile Ile Pro Gln Ser Phe Thr Val Asp Asn Ser Leu Asp Asn

865 870 875 880
Leu Val Leu Thr Asn Ser Ser Gln Asn Arg Gly Asn Lys Lys Asp Asp
885 890 895
Val Pro Ser Leu Glu Val Val Asn Arg Gln Leu Ala Tyr Trp Arg Ser
900 905 910
Leu Lys Asp Ala Gly Leu Met Thr Gln Arg Lys Phe Asp Asn Leu Thr
915 920 925
Lys Ala Met Arg Gly Gly Leu Thr Asp Lys Asp Arg Glu Arg Phe Ile

930 935 940
Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys Asn Val Ala Lys
945 950 955 960
Leu Leu Asp Met Arg Leu Asn Asp Lys Lys Asp Glu Ala Gly Asn Lys

965 970 975

Ile Arg Glu Thr Asn Ile Val Leu Leu Lys Ser Ala Met Ala Ser Glu

980 985 990

Phe Arg Lys Met Phe Arg Leu Tyr Lys Val Arg Glu Leu Asn Asp Tyr

995 1000 1005

His His Ala His Asp Ala Tyr Leu Asn Ala Ala Ile Ala Ile Asn

1010 1015 1020

Leu Leu Ala Leu Tyr Pro Tyr Met Ala Asp Asp Phe Val Tyr Gly

1025 1030 1035

Glu Phe Arg Tyr Lys Lys Lys Pro Gln Ala Glu Lys Ala Thr Tyr

1040 1045 1050

Glu Lys Leu Arg Gln Trp Asn Leu Ile Lys Arg Phe Gly Glu Lys

1055 1060 1065

Gln Leu Phe Thr Pro Asp His Glu Asp Cys Trp Asn Lys Glu Arg

1070 1075 1080

Asp Ile Lys Thr Ile Lys Lys Val Met Gly Tyr Arg Gln Val Asn

1085 1090 1095

Val Val Lys Lys Ala Glu Glu Arg Thr Gly Met Leu Phe Lys Glu

1100 1105 1110

Thr Ile Asn Gly Lys Thr Asn Lys Gly Ser Arg Ile Pro Ile Lys

1115 1120 1125

Lys Asp Leu Asp Pro Ser Lys Tyr Gly Gly Tyr Ile Glu Glu Lys

1130 1135 1140

Met Ala Tyr Tyr Ala Val Ile Ser Tyr Glu Asp Lys Lys Lys Lys

1145 1150 1155

Pro Gly Lys Thr Ile Val Gly Ile Ser Ile Met Asp Lys Lys Glu

1160 1165 1170

Phe Glu Tyr Asp Ser Ile Ser Tyr Leu Gly Lys Leu Gly Phe Ser

1175 1180 1185

Asn Pro Val Val Gln Ile Ile Leu Lys Asn Tyr Ser Leu Ile Ala

1190 1195 1200

Tyr Pro Asp Gly Arg Arg Arg Tyr Ile Thr Gly Ala Thr Lys Thr

1205 1210 1215

Thr Lys Gly Lys Val Glu Leu Gln Lys Ala Asn Gln Ile Ala Met

1220 1225 1230

Glu Gln Asp Leu Val Asn Phe Ile Tyr His Leu Lys Asn Tyr Asp

1235 1240 1245

Glu Ile Ser His Pro Glu Ser Tyr Ala Phe Val Gln Ser His Thr

1250 1255 1260

Asp Tyr Phe Asp Arg Leu Phe Asp Ser Ile Glu His Tyr Thr Arg

1265 1270 1275

Arg Phe Leu Asp Ala Glu Thr Asn Ile Asn Arg Leu Arg Arg Ile

1280 1285 1290

Tyr Glu Glu Glu Lys Lys Lys Asp Pro Val Asp Ile Glu Ala Leu

1295 1300 1305

Val Ala Ser Phe Ile Glu Leu Leu Lys Leu Thr Ser Ala Gly Ala

1310 1315 1320

Pro Ala Asp Phe Ile Phe Met Gly Glu Ala Ile Ser Arg Arg Arg

1325 1330 1335

Tyr Asn Ser Met Thr Gly Leu Phe Asp Gly Gln Val Ile Tyr Gln

1340 1345 1350

Ser Leu Thr Gly Leu Tyr Glu Thr Arg Met Arg Phe Glu Asp Gly

1355 1360 1365

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys

1370 1375 1380

Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr

1385 1390 1395

Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1400 1405 1410

<210> 88

<211> 1628

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 88

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Ser Glu Val Glu Phe
1 5 10 15

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala
20 25 30

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn
35 40 45

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro
50 55 60

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met
65 70 75 80

Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro
85 90 95

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
100 105 110

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
115 120 125

Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
130 135 140

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg
145 150 155 160

Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr
165 170 175

Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr
180 185 190

Pro Glu Ser Ser Gly Pro Lys Lys Lys Arg Lys Val Gly Lys Glu Tyr
195 200 205

His Ile Gly Leu Ala Ile Gly Thr Ser Ser Ile Gly Trp Ala Val Thr

210 215 220

Asp Ser Gln Phe Lys Leu Met Arg Ile Lys Gly Lys Thr Ala Ile Gly

225 230 235 240

Val Arg Leu Phe Glu Glu Gly Lys Thr Ala Ala Glu Arg Arg Thr Phe

245 250 255

Arg Thr Thr Arg Arg Arg Leu Lys Arg Arg Lys Trp Arg Leu His Tyr

260 265 270

Leu Asp Glu Ile Phe Ala Pro His Leu Gln Glu Val Asp Glu Asn Phe

275 280 285

Leu Arg Arg Leu Lys Gln Ser Asn Ile His Pro Glu Asp Pro Ala Lys

290 295 300

Asn Gln Ala Phe Ile Gly Lys Leu Leu Phe Pro Asp Leu Leu Lys Lys

305 310 315 320

Asn Glu Arg Gly Tyr Pro Thr Leu Ile Lys Met Arg Asp Glu Leu Pro

325 330 335

Val Glu Gln Arg Ala His Tyr Pro Val Thr Asn Ile Tyr Lys Leu Arg

340 345 350

Glu Ala Met Ile Asn Glu Asp Arg Gln Phe Asp Leu Arg Glu Val Tyr

355 360 365

Leu Ala Val His His Ile Val Lys Tyr Arg Gly His Phe Leu Asn Asn

370 375 380

Ala Ser Val Asp Lys Phe Lys Val Gly Arg Ile Asp Phe Asp Lys Ser

385 390 395 400

Phe Asn Val Leu Asn Glu Ala Tyr Glu Glu Leu Gln Asn Gly Glu Gly

405 410 415

Ser Phe Thr Ile Glu Pro Ser Lys Val Glu Lys Ile Gly Gln Leu Leu

420 425 430

Leu Asp Thr Lys Met Arg Lys Leu Asp Arg Gln Lys Ala Val Ala Lys

435 440 445

Leu Leu Glu Val Lys Val Ala Asp Lys Glu Glu Thr Lys Arg Asn Lys

450 455 460

Gln Ile Ala Thr Ala Met Ser Lys Leu Val Leu Gly Tyr Lys Ala Asp
465 470 475 480
Phe Ala Thr Val Ala Met Ala Asn Gly Asn Glu Trp Lys Ile Asp Leu
485 490 495
Ser Ser Glu Thr Ser Glu Asp Glu Ile Glu Lys Phe Arg Glu Glu Leu

500 505 510
Ser Asp Ala Gln Asn Asp Ile Leu Thr Glu Ile Thr Ser Leu Phe Ser
515 520 525
Gln Ile Met Leu Asn Glu Ile Val Pro Asn Gly Met Ser Ile Ser Glu
530 535 540
Ser Met Met Asp Arg Tyr Trp Thr His Glu Arg Gln Leu Ala Glu Val
545 550 555 560
Lys Glu Tyr Leu Ala Thr Gln Pro Ala Ser Ala Arg Lys Glu Phe Asp

565 570 575
Gln Val Tyr Asn Lys Tyr Ile Gly Gln Ala Pro Lys Glu Lys Gly Phe
580 585 590
Asp Leu Glu Lys Gly Leu Lys Lys Ile Leu Ser Lys Lys Glu Asn Trp
595 600 605
Lys Glu Ile Asp Glu Leu Leu Lys Ala Gly Asp Phe Leu Pro Lys Gln
610 615 620
Arg Thr Ser Ala Asn Gly Val Ile Pro His Gln Met His Gln Gln Glu

625 630 635 640
Leu Asp Arg Ile Ile Glu Lys Gln Ala Lys Tyr Tyr Pro Trp Leu Ala
645 650 655
Thr Glu Asn Pro Ala Thr Gly Glu Arg Asp Arg His Gln Ala Lys Tyr
660 665 670
Glu Leu Asp Gln Leu Val Ser Phe Arg Ile Pro Tyr Tyr Val Gly Pro
675 680 685
Leu Val Thr Pro Glu Val Gln Lys Ala Thr Ser Gly Ala Lys Phe Ala

690 695 700
Trp Ala Lys Arg Lys Glu Asp Gly Glu Ile Thr Pro Trp Asn Leu Trp

705 710 715 720

Asp Lys Ile Asp Arg Ala Glu Ser Ala Glu Ala Phe Ile Lys Arg Met

725 730 735

Thr Val Lys Asp Thr Tyr Leu Leu Asn Glu Asp Val Leu Pro Ala Asn

740 745 750

Ser Leu Leu Tyr Gln Lys Tyr Asn Val Leu Asn Glu Leu Asn Asn Val

755 760 765

Arg Val Asn Gly Arg Arg Leu Ser Val Gly Ile Lys Gln Asp Ile Tyr

770 775 780

Thr Glu Leu Phe Lys Lys Lys Lys Thr Val Lys Ala Gly Asp Val Ala

785 790 795 800

Ser Leu Val Met Ala Lys Thr Arg Gly Val Asn Lys Pro Ser Val Glu

805 810 815

Gly Leu Ser Asp Pro Lys Lys Phe Asn Ser Asn Leu Ala Thr Tyr Leu

820 825 830

Asp Leu Lys Ser Ile Val Gly Asp Lys Val Asp Asp Asn Arg Tyr Gln

835 840 845

Met Asp Leu Glu Asn Ile Ile Glu Trp Arg Ser Val Phe Glu Asp Gly

850 855 860

Glu Ile Phe Ala Asp Lys Leu Thr Glu Val Glu Trp Leu Thr Asp Glu

865 870 875 880

Gln Arg Ser Ala Leu Val Lys Lys Arg Tyr Lys Gly Trp Gly Arg Leu

885 890 895

Ser Lys Lys Leu Leu Thr Gly Ile Val Asp Glu Asn Gly Gln Arg Ile

900 905 910

Ile Asp Leu Met Trp Asn Thr Asp Gln Asn Phe Met Gln Ile Val Asn

915 920 925

Gln Pro Val Phe Lys Glu Gln Ile Asp Gln Leu Asn Gln Lys Ala Ile

930 935 940

Thr Asn Asp Gly Met Thr Leu Arg Glu Arg Val Glu Ser Val Leu Asp

945 950 955 960

Asp Ala Tyr Thr Ser Pro Gln Asn Lys Lys Ala Ile Trp Gln Val Val
965 970 975
Arg Val Val Glu Asp Ile Val Lys Ala Val Gly Asn Ala Pro Lys Ser
980 985 990
Ile Ser Ile Glu Phe Ala Arg Asn Glu Gly Asn Lys Gly Glu Ile Thr
995 1000 1005
Arg Ser Arg Arg Thr Gln Leu Gln Lys Leu Phe Glu Asp Gln Ala

1010 1015 1020
His Glu Leu Val Lys Asp Thr Ser Leu Thr Glu Glu Leu Glu Lys
1025 1030 1035
Ala Pro Asp Leu Ser Asp Arg Tyr Tyr Phe Tyr Phe Thr Gln Gly
1040 1045 1050
Gly Lys Asp Met Tyr Thr Gly Asp Pro Ile Asn Phe Asp Glu Ile
1055 1060 1065
Ser Thr Lys Tyr Asp Ile Asp His Ile Leu Pro Gln Ser Phe Val

1070 1075 1080
Lys Asp Asp Ser Leu Asp Asn Arg Val Leu Val Ser Arg Ala Glu
1085 1090 1095
Asn Asn Lys Lys Ser Asp Arg Val Pro Ala Lys Leu Tyr Ala Ala
1100 1105 1110
Lys Met Lys Pro Tyr Trp Asn Gln Leu Leu Lys Gln Gly Leu Ile
1115 1120 1125
Thr Gln Arg Lys Phe Glu Asn Leu Thr Met Asp Val Asp Gln Thr

1130 1135 1140
Ile Lys Tyr Arg Ser Leu Gly Phe Val Lys Arg Gln Leu Val Glu
1145 1150 1155
Thr Arg Gln Val Ile Lys Leu Thr Ala Asn Ile Leu Gly Ser Met
1160 1165 1170
Tyr Gln Glu Ala Gly Thr Asp Ile Ile Glu Thr Arg Ala Gly Leu
1175 1180 1185
Thr Lys Gln Leu Arg Glu Glu Phe Asp Leu Pro Lys Val Arg Glu

1190 1195 1200

Val Asn Asp Tyr His His Ala Val Asp Ala Tyr Leu Thr Thr Phe

1205 1210 1215

Ala Gly Gln Tyr Leu Asn Arg Arg Tyr Pro Lys Leu Arg Ser Phe

1220 1225 1230

Phe Val Tyr Gly Glu Tyr Met Lys Phe Lys His Gly Ser Asp Leu

1235 1240 1245

Lys Leu Arg Asn Phe Asn Phe Phe His Glu Leu Met Glu Gly Asp

1250 1255 1260

Lys Ser Gln Gly Lys Val Val Asp Gln Gln Thr Gly Glu Leu Ile

1265 1270 1275

Thr Thr Arg Asp Glu Val Ala Asp Tyr Phe Asp Trp Val Ile Asn

1280 1285 1290

Leu Lys Val Met Leu Ile Ser Asn Glu Thr Tyr Glu Glu Thr Gly

1295 1300 1305

Lys Tyr Phe Asp Ala Ser His Glu Ser Ser Ser Leu Tyr Leu Lys

1310 1315 1320

Asn Gln Asn Lys Lys Ser Lys Leu Val Val Pro Leu Lys Asn Lys

1325 1330 1335

Leu Gln Pro Glu Tyr Tyr Gly Ala Tyr Thr Gly Ile Thr Gln Gly

1340 1345 1350

Tyr Met Val Ile Leu Lys Leu Leu Asp Lys Lys Gly Gly Phe Gly

1355 1360 1365

Val Tyr Arg Ile Pro Arg Tyr Ala Ala Asp Ile Leu Asn Lys Cys

1370 1375 1380

His Asp Glu Val Ala Tyr Arg Asn Lys Ile Ala Glu Ile Ile Ser

1385 1390 1395

Ser Asp Pro Arg Ala Pro Lys Ser Phe Glu Val Val Val Pro Arg

1400 1405 1410

Val Leu Lys Gly Thr Phe Leu Val Asp Gly Glu Glu Lys Phe Ile

1415 1420 1425

Leu Ser Ser Tyr Arg Tyr Lys Val Asn Ala Thr Gln Leu Ile Leu

1430 1435 1440

Pro Val Ser Asp Ile Lys Leu Ile Gln Asp Asn Phe Lys Ala Leu

1445 1450 1455

Lys Lys Leu Asn Val Glu Met Gln Thr Lys Lys Leu Ile Glu Ile

1460 1465 1470

Tyr Asp Asn Ile Leu Arg Gln Val Asp Lys Tyr Tyr Lys Leu Tyr

1475 1480 1485

Asp Ile Asn Lys Phe Arg Ala Lys Leu His Asp Gly Arg Ser Lys

1490 1495 1500

Phe Val Glu Leu Asp Asp Phe Gly Gln Asp Ala Ser Lys Glu Lys

1505 1510 1515

Val Ile Ile Lys Ile Leu Arg Gly Leu His Phe Gly Ser Asp Leu

1520 1525 1530

Gln Asn Leu Lys Glu Ile Gly Phe Gly Thr Thr Pro Leu Gly Gln

1535 1540 1545

Phe Gln Val Ser Glu Ala Gly Ile Arg Leu Ser Asn Thr Ala Phe

1550 1555 1560

Ile Ile Phe Lys Ser Pro Thr Gly Leu Phe Asn Arg Lys Leu Tyr

1565 1570 1575

Leu Lys Asn Leu Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly

1580 1585 1590

Gln Ala Lys Lys Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp

1595 1600 1605

Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp

1610 1615 1620

Val Pro Asp Tyr Ala

1625

<210> 89

<211> 1624

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 89

Met Pro Lys Lys Lys Arg Lys Val Gly Lys Glu Tyr His Ile Gly Leu
1 5 10 15

Ala Ile Gly Thr Ser Ser Ile Gly Trp Ala Val Thr Asp Ser Gln Phe
20 25 30

Lys Leu Met Arg Ile Lys Gly Lys Thr Ala Ile Gly Val Arg Leu Phe
35 40 45

Glu Glu Gly Lys Thr Ala Ala Glu Arg Arg Thr Phe Arg Thr Thr Arg
50 55 60

Arg Arg Leu Lys Arg Arg Lys Trp Arg Leu His Tyr Leu Asp Glu Ile
65 70 75 80

Phe Ala Pro His Leu Gln Glu Val Asp Glu Asn Phe Leu Arg Arg Leu
85 90 95

Lys Gln Ser Asn Ile His Pro Glu Asp Pro Ala Lys Asn Gln Ala Phe
100 105 110

Ile Gly Lys Leu Leu Phe Pro Asp Leu Leu Lys Lys Asn Glu Arg Gly
115 120 125

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

Ala His Tyr Pro Val Thr Asn Ile Tyr Lys Leu Arg Glu Ala Met Ile
145 150 155 160

Asn Glu Asp Arg Gln Phe Asp Leu Arg Glu Val Tyr Leu Ala Val His
165 170 175

His Ile Val Lys Tyr Arg Gly His Phe Leu Asn Asn Ala Ser Val Asp
180 185 190

Lys Phe Lys Val Gly Arg Ile Asp Phe Asp Lys Ser Phe Asn Val Leu
195 200 205

Asn Glu Ala Tyr Glu Glu Leu Gln Asn Gly Glu Gly Ser Phe Thr Ile

210 215 220

Glu Pro Ser Lys Val Glu Lys Ile Gly Gln Leu Leu Leu Asp Thr Lys
225 230 235 240
Met Arg Lys Leu Asp Arg Gln Lys Ala Val Ala Lys Leu Leu Glu Val
245 250 255
Lys Val Ala Asp Lys Glu Glu Thr Lys Arg Asn Lys Gln Ile Ala Thr
260 265 270
Ala Met Ser Lys Leu Val Leu Gly Tyr Lys Ala Asp Phe Ala Thr Val
275 280 285

Ala Met Ala Asn Gly Asn Glu Trp Lys Ile Asp Leu Ser Ser Glu Thr
290 295 300
Ser Glu Asp Glu Ile Glu Lys Phe Arg Glu Glu Leu Ser Asp Ala Gln
305 310 315 320
Asn Asp Ile Leu Thr Glu Ile Thr Ser Leu Phe Ser Gln Ile Met Leu
325 330 335
Asn Glu Ile Val Pro Asn Gly Met Ser Ile Ser Glu Ser Met Met Asp
340 345 350

Arg Tyr Trp Thr His Glu Arg Gln Leu Ala Glu Val Lys Glu Tyr Leu
355 360 365
Ala Thr Gln Pro Ala Ser Ala Arg Lys Glu Phe Asp Gln Val Tyr Asn
370 375 380
Lys Tyr Ile Gly Gln Ala Pro Lys Glu Lys Gly Phe Asp Leu Glu Lys
385 390 395 400
Gly Leu Lys Lys Ile Leu Ser Lys Lys Glu Asn Trp Lys Glu Ile Asp
405 410 415

Glu Leu Leu Lys Ala Gly Asp Phe Leu Pro Lys Gln Arg Thr Ser Ala
420 425 430
Asn Gly Val Ile Pro His Gln Met His Gln Gln Glu Leu Asp Arg Ile
435 440 445
Ile Glu Lys Gln Ala Lys Tyr Tyr Pro Trp Leu Ala Thr Glu Asn Pro
450 455 460

Ala Thr Gly Glu Arg Asp Arg His Gln Ala Lys Tyr Glu Leu Asp Gln
465 470 475 480

Leu Val Ser Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Val Thr Pro
485 490 495
Glu Val Gln Lys Ala Thr Ser Gly Ala Lys Phe Ala Trp Ala Lys Arg
500 505 510
Lys Glu Asp Gly Glu Ile Thr Pro Trp Asn Leu Trp Asp Lys Ile Asp
515 520 525
Arg Ala Glu Ser Ala Glu Ala Phe Ile Lys Arg Met Thr Val Lys Asp
530 535 540

Thr Tyr Leu Leu Asn Glu Asp Val Leu Pro Ala Asn Ser Leu Leu Tyr
545 550 555 560
Gln Lys Tyr Asn Val Leu Asn Glu Leu Asn Asn Val Arg Val Asn Gly
565 570 575
Arg Arg Leu Ser Val Gly Ile Lys Gln Asp Ile Tyr Thr Glu Leu Phe
580 585 590
Lys Lys Lys Lys Thr Val Lys Ala Gly Asp Val Ala Ser Leu Val Met
595 600 605

Ala Lys Thr Arg Gly Val Asn Lys Pro Ser Val Glu Gly Leu Ser Asp
610 615 620
Pro Lys Lys Phe Asn Ser Asn Leu Ala Thr Tyr Leu Asp Leu Lys Ser
625 630 635 640
Ile Val Gly Asp Lys Val Asp Asp Asn Arg Tyr Gln Met Asp Leu Glu
645 650 655
Asn Ile Ile Glu Trp Arg Ser Val Phe Glu Asp Gly Glu Ile Phe Ala
660 665 670

Asp Lys Leu Thr Glu Val Glu Trp Leu Thr Asp Glu Gln Arg Ser Ala
675 680 685
Leu Val Lys Lys Arg Tyr Lys Gly Trp Gly Arg Leu Ser Lys Lys Leu
690 695 700
Leu Thr Gly Ile Val Asp Glu Asn Gly Gln Arg Ile Ile Asp Leu Met

705 710 715 720

Trp Asn Thr Asp Gln Asn Phe Met Gln Ile Val Asn Gln Pro Val Phe

725 730 735

Lys Glu Gln Ile Asp Gln Leu Asn Gln Lys Ala Ile Thr Asn Asp Gly

740 745 750

Met Thr Leu Arg Glu Arg Val Glu Ser Val Leu Asp Asp Ala Tyr Thr

755 760 765

Ser Pro Gln Asn Lys Lys Ala Ile Trp Gln Val Val Arg Val Val Glu

770 775 780

Asp Ile Val Lys Ala Val Gly Asn Ala Pro Lys Ser Ile Ser Ile Glu

785 790 795 800

Phe Ala Arg Asn Glu Gly Asn Lys Gly Glu Ile Thr Arg Ser Arg Arg

805 810 815

Thr Gln Leu Gln Lys Leu Phe Glu Asp Gln Ala His Glu Leu Val Lys

820 825 830

Asp Thr Ser Leu Thr Glu Glu Leu Glu Lys Ala Pro Asp Leu Ser Asp

835 840 845

Arg Tyr Tyr Phe Tyr Phe Thr Gln Gly Gly Lys Asp Met Tyr Thr Gly

850 855 860

Asp Pro Ile Asn Phe Asp Glu Ile Ser Thr Lys Tyr Asp Ile Asp His

865 870 875 880

Ile Leu Pro Gln Ser Phe Val Lys Asp Asp Ser Leu Asp Asn Arg Val

885 890 895

Leu Val Ser Arg Ala Glu Asn Asn Lys Lys Ser Asp Arg Val Pro Ala

900 905 910

Lys Leu Tyr Ala Ala Lys Met Lys Pro Tyr Trp Asn Gln Leu Leu Lys

915 920 925

Gln Gly Leu Ile Thr Gln Arg Lys Phe Glu Asn Leu Thr Met Asp Val

930 935 940

Asp Gln Thr Ile Lys Tyr Arg Ser Leu Gly Phe Val Lys Arg Gln Leu

945 950 955 960

Val Glu Thr Arg Gln Val Ile Lys Leu Thr Ala Asn Ile Leu Gly Ser
965 970 975
Met Tyr Gln Glu Ala Gly Thr Asp Ile Ile Glu Thr Arg Ala Gly Leu
980 985 990

Thr Lys Gln Leu Arg Glu Glu Phe Asp Leu Pro Lys Val Arg Glu Val
995 1000 1005
Asn Asp Tyr His His Ala Val Asp Ala Tyr Leu Thr Thr Phe Ala
1010 1015 1020
Gly Gln Tyr Leu Asn Arg Arg Tyr Pro Lys Leu Arg Ser Phe Phe
1025 1030 1035
Val Tyr Gly Glu Tyr Met Lys Phe Lys His Gly Ser Asp Leu Lys
1040 1045 1050

Leu Arg Asn Phe Asn Phe Phe His Glu Leu Met Glu Gly Asp Lys
1055 1060 1065
Ser Gln Gly Lys Val Val Asp Gln Gln Thr Gly Glu Leu Ile Thr
1070 1075 1080
Thr Arg Asp Glu Val Ala Asp Tyr Phe Asp Trp Val Ile Asn Leu
1085 1090 1095
Lys Val Met Leu Ile Ser Asn Glu Thr Tyr Glu Glu Thr Gly Lys
1100 1105 1110

Tyr Phe Asp Ala Ser His Glu Ser Ser Ser Leu Tyr Leu Lys Asn
1115 1120 1125
Gln Asn Lys Lys Ser Lys Leu Val Val Pro Leu Lys Asn Lys Leu
1130 1135 1140
Gln Pro Glu Tyr Tyr Gly Ala Tyr Thr Gly Ile Thr Gln Gly Tyr
1145 1150 1155
Met Val Ile Leu Lys Leu Leu Asp Lys Lys Gly Gly Phe Gly Val
1160 1165 1170

Tyr Arg Ile Pro Arg Tyr Ala Ala Asp Ile Leu Asn Lys Cys His
1175 1180 1185
Asp Glu Val Ala Tyr Arg Asn Lys Ile Ala Glu Ile Ile Ser Ser

1190 1195 1200

Asp Pro Arg Ala Pro Lys Ser Phe Glu Val Val Val Pro Arg Val

1205 1210 1215

Leu Lys Gly Thr Phe Leu Val Asp Gly Glu Glu Lys Phe Ile Leu

1220 1225 1230

Ser Ser Tyr Arg Tyr Lys Val Asn Ala Thr Gln Leu Ile Leu Pro

1235 1240 1245

Val Ser Asp Ile Lys Leu Ile Gln Asp Asn Phe Lys Ala Leu Lys

1250 1255 1260

Lys Leu Asn Val Glu Met Gln Thr Lys Lys Leu Ile Glu Ile Tyr

1265 1270 1275

Asp Asn Ile Leu Arg Gln Val Asp Lys Tyr Tyr Lys Leu Tyr Asp

1280 1285 1290

Ile Asn Lys Phe Arg Ala Lys Leu His Asp Gly Arg Ser Lys Phe

1295 1300 1305

Val Glu Leu Asp Asp Phe Gly Gln Asp Ala Ser Lys Glu Lys Val

1310 1315 1320

Ile Ile Lys Ile Leu Arg Gly Leu His Phe Gly Ser Asp Leu Gln

1325 1330 1335

Asn Leu Lys Glu Ile Gly Phe Gly Thr Thr Pro Leu Gly Gln Phe

1340 1345 1350

Gln Val Ser Glu Ala Gly Ile Arg Leu Ser Asn Thr Ala Phe Ile

1355 1360 1365

Ile Phe Lys Ser Pro Thr Gly Leu Phe Asn Arg Lys Leu Tyr Leu

1370 1375 1380

Lys Asn Leu Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala

1385 1390 1395

Lys Lys Lys Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser

1400 1405 1410

Ala Thr Pro Glu Ser Ser Gly Ser Glu Val Glu Phe Ser His Glu

1415 1420 1425

Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp
1430 1435 1440
Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn Arg
1445 1450 1455
Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro
1460 1465 1470

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val
1475 1480 1485
Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe
1490 1495 1500
Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile
1505 1510 1515
Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala
1520 1525 1530

Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His Arg
1535 1540 1545
Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu
1550 1555 1560
Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln
1565 1570 1575
Lys Lys Ala Gln Ser Ser Thr Asp Pro Ala Ala Lys Arg Val Lys
1580 1585 1590

Leu Asp Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro
1595 1600 1605
Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr
1610 1615 1620
Ala

<210> 90
<211> 1888
<212> PRT
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 90

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Thr Ser Glu Lys Gly Pro
1 5 10 15

Ser Thr Gly Asp Pro Thr Leu Arg Arg Arg Ile Glu Ser Trp Glu Phe
20 25 30

Asp Val Phe Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu
35 40 45

Tyr Glu Ile Lys Trp Gly Met Ser Arg Lys Ile Trp Arg Ser Ser Gly
50 55 60

Lys Asn Thr Thr Asn His Val Glu Val Asn Phe Ile Lys Lys Phe Thr
65 70 75 80

Ser Glu Arg Arg Phe His Ser Ser Ile Ser Cys Ser Ile Thr Trp Phe
85 90 95

Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser Gln Ala Ile Arg Glu Phe
100 105 110

Leu Ser Gln His Pro Gly Val Thr Leu Val Ile Tyr Val Ala Arg Leu
115 120 125

Phe Trp His Met Asp Gln Arg Asn Arg Gln Gly Leu Arg Asp Leu Val
130 135 140

Asn Ser Gly Val Thr Ile Gln Ile Met Arg Ala Ser Glu Tyr Tyr His
145 150 155 160

Cys Trp Arg Asn Phe Val Asn Tyr Pro Pro Gly Asp Glu Ala His Trp
165 170 175

Pro Gln Tyr Pro Pro Leu Trp Met Met Leu Tyr Ala Leu Glu Leu His
180 185 190

Cys Ile Ile Leu Ser Leu Pro Pro Cys Leu Lys Ile Ser Arg Arg Trp
195 200 205

Gln Asn His Leu Ala Phe Phe Arg Leu His Leu Gln Asn Cys His Tyr
210 215 220

Gln Thr Ile Pro Pro His Ile Leu Leu Ala Thr Gly Leu Ile His Pro
225 230 235 240
Ser Val Thr Trp Arg Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser
245 250 255
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
260 265 270

Gly Gly Ser Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val Gly Lys
275 280 285
Glu Tyr His Ile Gly Leu Ala Ile Gly Thr Ser Ser Ile Gly Trp Ala
290 295 300
Val Thr Asp Ser Gln Phe Lys Leu Met Arg Ile Lys Gly Lys Thr Ala
305 310 315 320
Ile Gly Val Arg Leu Phe Glu Glu Gly Lys Thr Ala Ala Glu Arg Arg
325 330 335

Thr Phe Arg Thr Thr Arg Arg Arg Leu Lys Arg Arg Lys Trp Arg Leu
340 345 350
His Tyr Leu Asp Glu Ile Phe Ala Pro His Leu Gln Glu Val Asp Glu
355 360 365
Asn Phe Leu Arg Arg Leu Lys Gln Ser Asn Ile His Pro Glu Asp Pro
370 375 380
Ala Lys Asn Gln Ala Phe Ile Gly Lys Leu Leu Phe Pro Asp Leu Leu
385 390 395 400

Lys Lys Asn Glu Arg Gly Tyr Pro Thr Leu Ile Lys Met Arg Asp Glu
405 410 415
Leu Pro Val Glu Gln Arg Ala His Tyr Pro Val Thr Asn Ile Tyr Lys
420 425 430
Leu Arg Glu Ala Met Ile Asn Glu Asp Arg Gln Phe Asp Leu Arg Glu
435 440 445
Val Tyr Leu Ala Val His His Ile Val Lys Tyr Arg Gly His Phe Leu
450 455 460

Asn Asn Ala Ser Val Asp Lys Phe Lys Val Gly Arg Ile Asp Phe Asp

465 470 475 480

Lys Ser Phe Asn Val Leu Asn Glu Ala Tyr Glu Glu Leu Gln Asn Gly

485 490 495

Glu Gly Ser Phe Thr Ile Glu Pro Ser Lys Val Glu Lys Ile Gly Gln

500 505 510

Leu Leu Leu Asp Thr Lys Met Arg Lys Leu Asp Arg Gln Lys Ala Val

515 520 525

Ala Lys Leu Leu Glu Val Lys Val Ala Asp Lys Glu Glu Thr Lys Arg

530 535 540

Asn Lys Gln Ile Ala Thr Ala Met Ser Lys Leu Val Leu Gly Tyr Lys

545 550 555 560

Ala Asp Phe Ala Thr Val Ala Met Ala Asn Gly Asn Glu Trp Lys Ile

565 570 575

Asp Leu Ser Ser Glu Thr Ser Glu Asp Glu Ile Glu Lys Phe Arg Glu

580 585 590

Glu Leu Ser Asp Ala Gln Asn Asp Ile Leu Thr Glu Ile Thr Ser Leu

595 600 605

Phe Ser Gln Ile Met Leu Asn Glu Ile Val Pro Asn Gly Met Ser Ile

610 615 620

Ser Glu Ser Met Met Asp Arg Tyr Trp Thr His Glu Arg Gln Leu Ala

625 630 635 640

Glu Val Lys Glu Tyr Leu Ala Thr Gln Pro Ala Ser Ala Arg Lys Glu

645 650 655

Phe Asp Gln Val Tyr Asn Lys Tyr Ile Gly Gln Ala Pro Lys Glu Lys

660 665 670

Gly Phe Asp Leu Glu Lys Gly Leu Lys Lys Ile Leu Ser Lys Lys Glu

675 680 685

Asn Trp Lys Glu Ile Asp Glu Leu Leu Lys Ala Gly Asp Phe Leu Pro

690 695 700

Lys Gln Arg Thr Ser Ala Asn Gly Val Ile Pro His Gln Met His Gln

705 710 715 720

Gln Glu Leu Asp Arg Ile Ile Glu Lys Gln Ala Lys Tyr Tyr Pro Trp
725 730 735
Leu Ala Thr Glu Asn Pro Ala Thr Gly Glu Arg Asp Arg His Gln Ala
740 745 750
Lys Tyr Glu Leu Asp Gln Leu Val Ser Phe Arg Ile Pro Tyr Tyr Val
755 760 765
Gly Pro Leu Val Thr Pro Glu Val Gln Lys Ala Thr Ser Gly Ala Lys
770 775 780

Phe Ala Trp Ala Lys Arg Lys Glu Asp Gly Glu Ile Thr Pro Trp Asn
785 790 795 800
Leu Trp Asp Lys Ile Asp Arg Ala Glu Ser Ala Glu Ala Phe Ile Lys
805 810 815
Arg Met Thr Val Lys Asp Thr Tyr Leu Leu Asn Glu Asp Val Leu Pro
820 825 830
Ala Asn Ser Leu Leu Tyr Gln Lys Tyr Asn Val Leu Asn Glu Leu Asn
835 840 845

Asn Val Arg Val Asn Gly Arg Arg Leu Ser Val Gly Ile Lys Gln Asp
850 855 860
Ile Tyr Thr Glu Leu Phe Lys Lys Lys Lys Thr Val Lys Ala Gly Asp
865 870 875 880
Val Ala Ser Leu Val Met Ala Lys Thr Arg Gly Val Asn Lys Pro Ser
885 890 895
Val Glu Gly Leu Ser Asp Pro Lys Lys Phe Asn Ser Asn Leu Ala Thr
900 905 910

Tyr Leu Asp Leu Lys Ser Ile Val Gly Asp Lys Val Asp Asp Asn Arg
915 920 925
Tyr Gln Met Asp Leu Glu Asn Ile Ile Glu Trp Arg Ser Val Phe Glu
930 935 940
Asp Gly Glu Ile Phe Ala Asp Lys Leu Thr Glu Val Glu Trp Leu Thr
945 950 955 960
Asp Glu Gln Arg Ser Ala Leu Val Lys Lys Arg Tyr Lys Gly Trp Gly

965 970 975

Arg Leu Ser Lys Lys Leu Leu Thr Gly Ile Val Asp Glu Asn Gly Gln
980 985 990
Arg Ile Ile Asp Leu Met Trp Asn Thr Asp Gln Asn Phe Met Gln Ile
995 1000 1005
Val Asn Gln Pro Val Phe Lys Glu Gln Ile Asp Gln Leu Asn Gln
1010 1015 1020
Lys Ala Ile Thr Asn Asp Gly Met Thr Leu Arg Glu Arg Val Glu
1025 1030 1035

Ser Val Leu Asp Asp Ala Tyr Thr Ser Pro Gln Asn Lys Lys Ala
1040 1045 1050
Ile Trp Gln Val Val Arg Val Val Glu Asp Ile Val Lys Ala Val
1055 1060 1065
Gly Asn Ala Pro Lys Ser Ile Ser Ile Glu Phe Ala Arg Asn Glu
1070 1075 1080
Gly Asn Lys Gly Glu Ile Thr Arg Ser Arg Arg Thr Gln Leu Gln
1085 1090 1095

Lys Leu Phe Glu Asp Gln Ala His Glu Leu Val Lys Asp Thr Ser
1100 1105 1110
Leu Thr Glu Glu Leu Glu Lys Ala Pro Asp Leu Ser Asp Arg Tyr
1115 1120 1125
Tyr Phe Tyr Phe Thr Gln Gly Gly Lys Asp Met Tyr Thr Gly Asp
1130 1135 1140
Pro Ile Asn Phe Asp Glu Ile Ser Thr Lys Tyr Asp Ile Asp His
1145 1150 1155

Ile Leu Pro Gln Ser Phe Val Lys Asp Asp Ser Leu Asp Asn Arg
1160 1165 1170
Val Leu Val Ser Arg Ala Glu Asn Asn Lys Lys Ser Asp Arg Val
1175 1180 1185
Pro Ala Lys Leu Tyr Ala Ala Lys Met Lys Pro Tyr Trp Asn Gln
1190 1195 1200

Leu Leu Lys Gln Gly Leu Ile Thr Gln Arg Lys Phe Glu Asn Leu
1205 1210 1215

Thr Met Asp Val Asp Gln Thr Ile Lys Tyr Arg Ser Leu Gly Phe
1220 1225 1230
Val Lys Arg Gln Leu Val Glu Thr Arg Gln Val Ile Lys Leu Thr
1235 1240 1245
Ala Asn Ile Leu Gly Ser Met Tyr Gln Glu Ala Gly Thr Asp Ile
1250 1255 1260
Ile Glu Thr Arg Ala Gly Leu Thr Lys Gln Leu Arg Glu Glu Phe
1265 1270 1275

Asp Leu Pro Lys Val Arg Glu Val Asn Asp Tyr His His Ala Val
1280 1285 1290
Asp Ala Tyr Leu Thr Thr Phe Ala Gly Gln Tyr Leu Asn Arg Arg
1295 1300 1305
Tyr Pro Lys Leu Arg Ser Phe Phe Val Tyr Gly Glu Tyr Met Lys
1310 1315 1320
Phe Lys His Gly Ser Asp Leu Lys Leu Arg Asn Phe Asn Phe Phe
1325 1330 1335

His Glu Leu Met Glu Gly Asp Lys Ser Gln Gly Lys Val Val Asp
1340 1345 1350
Gln Gln Thr Gly Glu Leu Ile Thr Thr Arg Asp Glu Val Ala Asp
1355 1360 1365
Tyr Phe Asp Trp Val Ile Asn Leu Lys Val Met Leu Ile Ser Asn
1370 1375 1380
Glu Thr Tyr Glu Glu Thr Gly Lys Tyr Phe Asp Ala Ser His Glu
1385 1390 1395

Ser Ser Ser Leu Tyr Leu Lys Asn Gln Asn Lys Lys Ser Lys Leu
1400 1405 1410
Val Val Pro Leu Lys Asn Lys Leu Gln Pro Glu Tyr Tyr Gly Ala
1415 1420 1425
Tyr Thr Gly Ile Thr Gln Gly Tyr Met Val Ile Leu Lys Leu Leu

1430 1435 1440

Asp Lys Lys Gly Gly Phe Gly Val Tyr Arg Ile Pro Arg Tyr Ala

1445 1450 1455

Ala Asp Ile Leu Asn Lys Cys His Asp Glu Val Ala Tyr Arg Asn

1460 1465 1470

Lys Ile Ala Glu Ile Ile Ser Ser Asp Pro Arg Ala Pro Lys Ser

1475 1480 1485

Phe Glu Val Val Val Pro Arg Val Leu Lys Gly Thr Phe Leu Val

1490 1495 1500

Asp Gly Glu Glu Lys Phe Ile Leu Ser Ser Tyr Arg Tyr Lys Val

1505 1510 1515

Asn Ala Thr Gln Leu Ile Leu Pro Val Ser Asp Ile Lys Leu Ile

1520 1525 1530

Gln Asp Asn Phe Lys Ala Leu Lys Lys Leu Asn Val Glu Met Gln

1535 1540 1545

Thr Lys Lys Leu Ile Glu Ile Tyr Asp Asn Ile Leu Arg Gln Val

1550 1555 1560

Asp Lys Tyr Tyr Lys Leu Tyr Asp Ile Asn Lys Phe Arg Ala Lys

1565 1570 1575

Leu His Asp Gly Arg Ser Lys Phe Val Glu Leu Asp Asp Phe Gly

1580 1585 1590

Gln Asp Ala Ser Lys Glu Lys Val Ile Ile Lys Ile Leu Arg Gly

1595 1600 1605

Leu His Phe Gly Ser Asp Leu Gln Asn Leu Lys Glu Ile Gly Phe

1610 1615 1620

Gly Thr Thr Pro Leu Gly Gln Phe Gln Val Ser Glu Ala Gly Ile

1625 1630 1635

Arg Leu Ser Asn Thr Ala Phe Ile Ile Phe Lys Ser Pro Thr Gly

1640 1645 1650

Leu Phe Asn Arg Lys Leu Tyr Leu Lys Asn Leu Gly Lys Arg Pro

1655 1660 1665

Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser
1670 1675 1680
Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp
1685 1690 1695

Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser
1700 1705 1710
Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys
1715 1720 1725
Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr
1730 1735 1740
Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys
1745 1750 1755

Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile
1760 1765 1770
Lys Met Leu Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn
1775 1780 1785
Leu Ser Asp Ile Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile
1790 1795 1800
Gln Glu Ser Ile Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile
1805 1810 1815

Gly Asn Lys Pro Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp
1820 1825 1830
Glu Ser Thr Asp Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro
1835 1840 1845
Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu
1850 1855 1860
Asn Lys Ile Lys Met Leu Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1865 1870 1875

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr
1880 1885
<210> 91

<211> 1609

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 91

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Ser Glu Val Glu Phe
1 5 10 15

Ser His Glu Tyr Trp Met Arg His Ala Leu Thr Leu Ala Lys Arg Ala
20 25 30

Arg Asp Glu Arg Glu Val Pro Val Gly Ala Val Leu Val Leu Asn Asn
35 40 45

Arg Val Ile Gly Glu Gly Trp Asn Arg Ala Ile Gly Leu His Asp Pro
50 55 60

Thr Ala His Ala Glu Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met
65 70 75 80

Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro
85 90 95

Cys Val Met Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val
100 105 110

Val Phe Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met
115 120 125

Asp Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu
130 135 140

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe Arg
145 150 155 160

Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser Ser Thr
165 170 175

Asp Gly Ser Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr
180 185 190

Pro Glu Ser Ser Gly Pro Lys Lys Lys Arg Lys Val Gly Glu Lys Lys
195 200 205

Thr Asn Tyr Thr Ile Gly Leu Ala Ile Gly Thr Asp Ser Val Gly Trp

210 215 220

Ala Val Val Lys Asp Asp Leu Glu Leu Val Lys Lys Arg Met Lys Val

225 230 235 240

Leu Gly Asn Thr Glu Thr Asn Tyr Ile Lys Lys Asn Leu Trp Gly Ser

245 250 255

Leu Leu Phe Glu Ser Gly Gln Thr Ala Lys Asp Arg Arg Leu Lys Arg

260 265 270

Val Ala Arg Arg Arg Tyr Glu Arg Arg Arg Asn Arg Leu Thr Glu Leu

275 280 285

Gln Lys Ile Phe Ala Pro Ala Ile Asp Glu Val Asp Glu Asn Phe Phe

290 295 300

Phe Arg Leu Asn Glu Ser Phe Leu Val Pro Glu Asp Lys Ala Phe Ser

305 310 315 320

Lys Asn Pro Ile Phe Gly Thr Leu Gly Glu Asp Lys Thr Tyr Tyr Lys

325 330 335

Thr Tyr Pro Thr Ile Tyr His Leu Arg Gln His Leu Ala Asp Ser Glu

340 345 350

Glu Lys Ala Asp Val Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile

355 360 365

Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Lys Leu Asp Thr Glu His

370 375 380

Ile Ala Ile Asn Glu Asn Leu Glu Gln Phe Phe Glu Ser Tyr Asn Ala

385 390 395 400

Leu Phe Ser Glu Glu Pro Ile Glu Leu Arg Lys Glu Glu Leu Ile Ala

405 410 415

Ile Glu Asn Ile Leu Arg Glu Lys Asn Ser Arg Thr Val Lys Glu Lys

420 425 430

Arg Ile Thr Ser Phe Leu Lys Asp Ile Gly Arg Ala Asn Lys Gln Ser

435 440 445

Pro Met Met Ala Phe Ile Thr Leu Ile Val Gly Lys Lys Ala Lys Phe

450 455 460

Lys Ala Ala Phe Asn Leu Glu Glu Glu Ile Ser Leu Asn Leu Thr Asp

465 470 475 480

Asp Ser Tyr Asp Glu Asn Leu Glu Ile Leu Leu Asn Thr Ile Gly Ser

485 490 495

Asp Phe Ala Asp Leu Phe Asp His Ala Gln Arg Val Tyr Asn Ala Val

500 505 510

Glu Leu Ala Gly Ile Leu Ser Gly Asp Val Lys Asn Thr His Ala Lys

515 520 525

Leu Ser Ala Gln Met Val Ala Met Tyr Glu Arg His Lys Glu Gln Leu

530 535 540

Lys Glu Tyr Lys Ser Phe Ile Lys Ala Asn Leu Pro Asp Gln Tyr Asp

545 550 555 560

Met Thr Phe Val Ala Pro Lys Asp Ala Gln Lys Lys Asp Leu Lys Gly

565 570 575

Tyr Ala Gly Tyr Ile Asp Gly Asn Met Ser Gln Asp Ser Phe Tyr Lys

580 585 590

Phe Val Lys Asp Gln Leu Lys Glu Val Pro Gly Ser Glu Lys Phe Leu

595 600 605

Asp Ser Ile Glu Lys Glu Asp Phe Leu Arg Lys Gln Arg Ser Phe Tyr

610 615 620

Asn Gly Val Ile Pro Asn Gln Val His Leu Ala Glu Met Glu Ala Ile

625 630 635 640

Leu Asp Arg Gln Glu Asn Tyr Tyr Pro Trp Leu Lys Glu Asn Arg Glu

645 650 655

Lys Ile Ile Ser Leu Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro

660 665 670

Leu Ala Asp Gly Gln Ser Glu Phe Ala Trp Leu Glu Arg Lys Ser Asp

675 680 685

Glu Lys Ile Lys Pro Trp Asn Phe Ser Asp Val Val Asp Leu Asp Arg

690 695 700

Ser Ala Glu Lys Phe Ile Glu Gln Leu Ile Gly Arg Asp Thr Tyr Leu

705 710 715 720

Pro Asp Glu Tyr Val Leu Pro Lys Lys Ser Leu Ile Tyr Gln Lys Tyr

725 730 735

Met Val Phe Asn Glu Leu Thr Lys Ile Ala Tyr Leu Asp Glu Arg Gln

740 745 750

Lys Arg Met Asn Leu Ser Ser Val Glu Lys Lys Glu Ile Phe Glu Thr

755 760 765

Leu Phe Lys Lys Arg Ser Lys Val Thr Glu Lys Gln Leu Val Lys Phe

770 775 780

Phe Glu Asn Tyr Leu Gln Ile Asp Asn Pro Thr Ile Phe Gly Ile Glu

785 790 795 800

Asp Ala Phe Asn Ala Asp Tyr Ser Thr Tyr Val Glu Leu Ala Lys Val

805 810 815

Pro Gly Met Lys Ser Met Met Asp Asp Pro Asp Asn Glu Asp Leu Met

820 825 830

Glu Glu Ile Val Lys Ile Leu Thr Val Phe Glu Asp Arg Lys Met Arg

835 840 845

Arg Lys Gln Leu Glu Lys Tyr Lys Glu Arg Leu Ser Pro Glu Gln Ile

850 855 860

Lys Glu Leu Ala Lys Lys His Tyr Thr Gly Trp Gly Arg Leu Ser Lys

865 870 875 880

Lys Leu Leu Val Gly Ile Arg Asp Lys Glu Thr Gln Lys Thr Ile Leu

885 890 895

Asp Tyr Leu Val Glu Asp Asp Asn His Ser Gly Gly Arg Gln His Leu

900 905 910

Asn Arg Asn Leu Met Gln Leu Ile Asn Asp Asp Arg Leu Ser Phe Lys

915 920 925

Lys Thr Ile Ala Glu Leu Gln Met Ile Asp Pro Ser Ala Asp Leu Tyr

930 935 940

Ala Gln Val Gln Glu Ile Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

945 950 955 960

Leu Leu Gly Leu Lys Ile Val Asp Glu Ile Ile Arg Val Met Gly Glu

965 970 975

Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr

980 985 990

Ala Arg Gly Lys Ala Leu Ser Lys Arg Arg Glu Ala Lys Ile Lys Glu

995 1000 1005

Gly Leu Ala Ala Leu Gly Ser Ser Leu Leu Lys Glu Asn Leu Pro

1010 1015 1020

Gly Asn Ala Asp Leu Ser Gln Arg Lys Ile Tyr Leu Tyr Tyr Thr

1025 1030 1035

Gln Asn Gly Lys Asp Ile Tyr Leu Asp Glu Pro Leu Asp Phe Asp

1040 1045 1050

Arg Leu Ser Gln Tyr Asp Glu Asp His Ile Ile Pro Gln Ser Phe

1055 1060 1065

Thr Val Asp Asn Ser Leu Asp Asn Leu Val Leu Thr Asn Ser Ser

1070 1075 1080

Gln Asn Arg Gly Asn Lys Lys Asp Asp Val Pro Ser Leu Glu Val

1085 1090 1095

Val Asn Arg Gln Leu Ala Tyr Trp Arg Ser Leu Lys Asp Ala Gly

1100 1105 1110

Leu Met Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Met Arg

1115 1120 1125

Gly Gly Leu Thr Asp Lys Asp Arg Glu Arg Phe Ile Gln Arg Gln

1130 1135 1140

Leu Val Glu Thr Arg Gln Ile Thr Lys Asn Val Ala Lys Leu Leu

1145 1150 1155

Asp Met Arg Leu Asn Asp Lys Lys Asp Glu Ala Gly Asn Lys Ile

1160 1165 1170

Arg Glu Thr Asn Ile Val Leu Leu Lys Ser Ala Met Ala Ser Glu

1175 1180 1185

Phe Arg Lys Met Phe Arg Leu Tyr Lys Val Arg Glu Leu Asn Asp
1190 1195 1200
Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Ala Ile Ala Ile
1205 1210 1215
Asn Leu Leu Ala Leu Tyr Pro Tyr Met Ala Asp Asp Phe Val Tyr
1220 1225 1230
Gly Glu Phe Arg Tyr Lys Lys Lys Pro Gln Ala Glu Lys Ala Thr

1235 1240 1245
Tyr Glu Lys Leu Arg Gln Trp Asn Leu Ile Lys Arg Phe Gly Glu
1250 1255 1260
Lys Gln Leu Phe Thr Pro Asp His Glu Asp Cys Trp Asn Lys Glu
1265 1270 1275
Arg Asp Ile Lys Thr Ile Lys Lys Val Met Gly Tyr Arg Gln Val
1280 1285 1290
Asn Val Val Lys Lys Ala Glu Glu Arg Thr Gly Met Leu Phe Lys

1295 1300 1305
Glu Thr Ile Asn Gly Lys Thr Asn Lys Gly Ser Arg Ile Pro Ile
1310 1315 1320
Lys Lys Asp Leu Asp Pro Ser Lys Tyr Gly Gly Tyr Ile Glu Glu
1325 1330 1335
Lys Met Ala Tyr Tyr Ala Val Ile Ser Tyr Glu Asp Lys Lys Lys
1340 1345 1350
Lys Pro Gly Lys Thr Ile Val Gly Ile Ser Ile Met Asp Lys Lys

1355 1360 1365
Glu Phe Glu Tyr Asp Ser Ile Ser Tyr Leu Gly Lys Leu Gly Phe
1370 1375 1380
Ser Asn Pro Val Val Gln Ile Ile Leu Lys Asn Tyr Ser Leu Ile
1385 1390 1395
Ala Tyr Pro Asp Gly Arg Arg Arg Tyr Ile Thr Gly Ala Thr Lys
1400 1405 1410
Thr Thr Lys Gly Lys Val Glu Leu Gln Lys Ala Asn Gln Ile Ala

1415 1420 1425

Met Glu Gln Asp Leu Val Asn Phe Ile Tyr His Leu Lys Asn Tyr

1430 1435 1440

Asp Glu Ile Ser His Pro Glu Ser Tyr Ala Phe Val Gln Ser His

1445 1450 1455

Thr Asp Tyr Phe Asp Arg Leu Phe Asp Ser Ile Glu His Tyr Thr

1460 1465 1470

Arg Arg Phe Leu Asp Ala Glu Thr Asn Ile Asn Arg Leu Arg Arg

1475 1480 1485

Ile Tyr Glu Glu Glu Lys Lys Lys Asp Pro Val Asp Ile Glu Ala

1490 1495 1500

Leu Val Ala Ser Phe Ile Glu Leu Leu Lys Leu Thr Ser Ala Gly

1505 1510 1515

Ala Pro Ala Asp Phe Ile Phe Met Gly Glu Ala Ile Ser Arg Arg

1520 1525 1530

Arg Tyr Asn Ser Met Thr Gly Leu Phe Asp Gly Gln Val Ile Tyr

1535 1540 1545

Gln Ser Leu Thr Gly Leu Tyr Glu Thr Arg Met Arg Phe Glu Asp

1550 1555 1560

Gly Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys

1565 1570 1575

Lys Lys Gly Ser Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro

1580 1585 1590

Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr

1595 1600 1605

Ala

<210> 92

<211> 1604

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 92

Met Pro Lys Lys Lys Arg Lys Val Glu Lys Lys Thr Asn Tyr Thr Ile
1 5 10 15

Gly Leu Ala Ile Gly Thr Asp Ser Val Gly Trp Ala Val Val Lys Asp
20 25 30

Asp Leu Glu Leu Val Lys Lys Arg Met Lys Val Leu Gly Asn Thr Glu
35 40 45

Thr Asn Tyr Ile Lys Lys Asn Leu Trp Gly Ser Leu Leu Phe Glu Ser
50 55 60

Gly Gln Thr Ala Lys Asp Arg Arg Leu Lys Arg Val Ala Arg Arg Arg
65 70 75 80

Tyr Glu Arg Arg Arg Asn Arg Leu Thr Glu Leu Gln Lys Ile Phe Ala
85 90 95

Pro Ala Ile Asp Glu Val Asp Glu Asn Phe Phe Phe Arg Leu Asn Glu
100 105 110

Ser Phe Leu Val Pro Glu Asp Lys Ala Phe Ser Lys Asn Pro Ile Phe
115 120 125

Gly Thr Leu Gly Glu Asp Lys Thr Tyr Tyr Lys Thr Tyr Pro Thr Ile
130 135 140

Tyr His Leu Arg Gln His Leu Ala Asp Ser Glu Glu Lys Ala Asp Val
145 150 155 160

Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Tyr Arg Gly His
165 170 175

Phe Leu Ile Glu Gly Lys Leu Asp Thr Glu His Ile Ala Ile Asn Glu
180 185 190

Asn Leu Glu Gln Phe Phe Glu Ser Tyr Asn Ala Leu Phe Ser Glu Glu
195 200 205

Pro Ile Glu Leu Arg Lys Glu Glu Leu Ile Ala Ile Glu Asn Ile Leu
210 215 220

Arg Glu Lys Asn Ser Arg Thr Val Lys Glu Lys Arg Ile Thr Ser Phe

225 230 235 240

Leu Lys Asp Ile Gly Arg Ala Asn Lys Gln Ser Pro Met Met Ala Phe

245 250 255

Ile Thr Leu Ile Val Gly Lys Lys Ala Lys Phe Lys Ala Ala Phe Asn

260 265 270

Leu Glu Glu Glu Ile Ser Leu Asn Leu Thr Asp Asp Ser Tyr Asp Glu

275 280 285

Asn Leu Glu Ile Leu Leu Asn Thr Ile Gly Ser Asp Phe Ala Asp Leu

290 295 300

Phe Asp His Ala Gln Arg Val Tyr Asn Ala Val Glu Leu Ala Gly Ile

305 310 315 320

Leu Ser Gly Asp Val Lys Asn Thr His Ala Lys Leu Ser Ala Gln Met

325 330 335

Val Ala Met Tyr Glu Arg His Lys Glu Gln Leu Lys Glu Tyr Lys Ser

340 345 350

Phe Ile Lys Ala Asn Leu Pro Asp Gln Tyr Asp Met Thr Phe Val Ala

355 360 365

Pro Lys Asp Ala Gln Lys Lys Asp Leu Lys Gly Tyr Ala Gly Tyr Ile

370 375 380

Asp Gly Asn Met Ser Gln Asp Ser Phe Tyr Lys Phe Val Lys Asp Gln

385 390 395 400

Leu Lys Glu Val Pro Gly Ser Glu Lys Phe Leu Asp Ser Ile Glu Lys

405 410 415

Glu Asp Phe Leu Arg Lys Gln Arg Ser Phe Tyr Asn Gly Val Ile Pro

420 425 430

Asn Gln Val His Leu Ala Glu Met Glu Ala Ile Leu Asp Arg Gln Glu

435 440 445

Asn Tyr Tyr Pro Trp Leu Lys Glu Asn Arg Glu Lys Ile Ile Ser Leu

450 455 460

Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Asp Gly Gln

465 470 475 480

Ser Glu Phe Ala Trp Leu Glu Arg Lys Ser Asp Glu Lys Ile Lys Pro

485 490 495

Trp Asn Phe Ser Asp Val Val Asp Leu Asp Arg Ser Ala Glu Lys Phe

500 505 510

Ile Glu Gln Leu Ile Gly Arg Asp Thr Tyr Leu Pro Asp Glu Tyr Val

515 520 525

Leu Pro Lys Lys Ser Leu Ile Tyr Gln Lys Tyr Met Val Phe Asn Glu

530 535 540

Leu Thr Lys Ile Ala Tyr Leu Asp Glu Arg Gln Lys Arg Met Asn Leu

545 550 555 560

Ser Ser Val Glu Lys Lys Glu Ile Phe Glu Thr Leu Phe Lys Lys Arg

565 570 575

Ser Lys Val Thr Glu Lys Gln Leu Val Lys Phe Phe Glu Asn Tyr Leu

580 585 590

Gln Ile Asp Asn Pro Thr Ile Phe Gly Ile Glu Asp Ala Phe Asn Ala

595 600 605

Asp Tyr Ser Thr Tyr Val Glu Leu Ala Lys Val Pro Gly Met Lys Ser

610 615 620

Met Met Asp Asp Pro Asp Asn Glu Asp Leu Met Glu Glu Ile Val Lys

625 630 635 640

Ile Leu Thr Val Phe Glu Asp Arg Lys Met Arg Arg Lys Gln Leu Glu

645 650 655

Lys Tyr Lys Glu Arg Leu Ser Pro Glu Gln Ile Lys Glu Leu Ala Lys

660 665 670

Lys His Tyr Thr Gly Trp Gly Arg Leu Ser Lys Lys Leu Leu Val Gly

675 680 685

Ile Arg Asp Lys Glu Thr Gln Lys Thr Ile Leu Asp Tyr Leu Val Glu

690 695 700

Asp Asp Asn His Ser Gly Gly Arg Gln His Leu Asn Arg Asn Leu Met

705 710 715 720

Gln Leu Ile Asn Asp Asp Arg Leu Ser Phe Lys Lys Thr Ile Ala Glu

725 730 735

Leu Gln Met Ile Asp Pro Ser Ala Asp Leu Tyr Ala Gln Val Gln Glu

740 745 750

Ile Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Leu Gly Leu Lys

755 760 765

Ile Val Asp Glu Ile Ile Arg Val Met Gly Glu Lys Pro Glu Asn Ile

770 775 780

Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Ala Arg Gly Lys Ala

785 790 795 800

Leu Ser Lys Arg Arg Glu Ala Lys Ile Lys Glu Gly Leu Ala Ala Leu

805 810 815

Gly Ser Ser Leu Leu Lys Glu Asn Leu Pro Gly Asn Ala Asp Leu Ser

820 825 830

Gln Arg Lys Ile Tyr Leu Tyr Tyr Thr Gln Asn Gly Lys Asp Ile Tyr

835 840 845

Leu Asp Glu Pro Leu Asp Phe Asp Arg Leu Ser Gln Tyr Asp Glu Asp

850 855 860

His Ile Ile Pro Gln Ser Phe Thr Val Asp Asn Ser Leu Asp Asn Leu

865 870 875 880

Val Leu Thr Asn Ser Ser Gln Asn Arg Gly Asn Lys Lys Asp Asp Val

885 890 895

Pro Ser Leu Glu Val Val Asn Arg Gln Leu Ala Tyr Trp Arg Ser Leu

900 905 910

Lys Asp Ala Gly Leu Met Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys

915 920 925

Ala Met Arg Gly Gly Leu Thr Asp Lys Asp Arg Glu Arg Phe Ile Gln

930 935 940

Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys Asn Val Ala Lys Leu

945 950 955 960

Leu Asp Met Arg Leu Asn Asp Lys Lys Asp Glu Ala Gly Asn Lys Ile

965 970 975

Arg Glu Thr Asn Ile Val Leu Leu Lys Ser Ala Met Ala Ser Glu Phe

980 985 990

Arg Lys Met Phe Arg Leu Tyr Lys Val Arg Glu Leu Asn Asp Tyr His

995 1000 1005

His Ala His Asp Ala Tyr Leu Asn Ala Ala Ile Ala Ile Asn Leu

1010 1015 1020

Leu Ala Leu Tyr Pro Tyr Met Ala Asp Asp Phe Val Tyr Gly Glu

1025 1030 1035

Phe Arg Tyr Lys Lys Lys Pro Gln Ala Glu Lys Ala Thr Tyr Glu

1040 1045 1050

Lys Leu Arg Gln Trp Asn Leu Ile Lys Arg Phe Gly Glu Lys Gln

1055 1060 1065

Leu Phe Thr Pro Asp His Glu Asp Cys Trp Asn Lys Glu Arg Asp

1070 1075 1080

Ile Lys Thr Ile Lys Lys Val Met Gly Tyr Arg Gln Val Asn Val

1085 1090 1095

Val Lys Lys Ala Glu Glu Arg Thr Gly Met Leu Phe Lys Glu Thr

1100 1105 1110

Ile Asn Gly Lys Thr Asn Lys Gly Ser Arg Ile Pro Ile Lys Lys

1115 1120 1125

Asp Leu Asp Pro Ser Lys Tyr Gly Gly Tyr Ile Glu Glu Lys Met

1130 1135 1140

Ala Tyr Tyr Ala Val Ile Ser Tyr Glu Asp Lys Lys Lys Lys Pro

1145 1150 1155

Gly Lys Thr Ile Val Gly Ile Ser Ile Met Asp Lys Lys Glu Phe

1160 1165 1170

Glu Tyr Asp Ser Ile Ser Tyr Leu Gly Lys Leu Gly Phe Ser Asn

1175 1180 1185

Pro Val Val Gln Ile Ile Leu Lys Asn Tyr Ser Leu Ile Ala Tyr

1190 1195 1200

Pro Asp Gly Arg Arg Arg Tyr Ile Thr Gly Ala Thr Lys Thr Thr

1205 1210 1215

Lys Gly Lys Val Glu Leu Gln Lys Ala Asn Gln Ile Ala Met Glu

1220 1225 1230

Gln Asp Leu Val Asn Phe Ile Tyr His Leu Lys Asn Tyr Asp Glu

1235 1240 1245

Ile Ser His Pro Glu Ser Tyr Ala Phe Val Gln Ser His Thr Asp

1250 1255 1260

Tyr Phe Asp Arg Leu Phe Asp Ser Ile Glu His Tyr Thr Arg Arg

1265 1270 1275

Phe Leu Asp Ala Glu Thr Asn Ile Asn Arg Leu Arg Arg Ile Tyr

1280 1285 1290

Glu Glu Glu Lys Lys Lys Asp Pro Val Asp Ile Glu Ala Leu Val

1295 1300 1305

Ala Ser Phe Ile Glu Leu Leu Lys Leu Thr Ser Ala Gly Ala Pro

1310 1315 1320

Ala Asp Phe Ile Phe Met Gly Glu Ala Ile Ser Arg Arg Arg Tyr

1325 1330 1335

Asn Ser Met Thr Gly Leu Phe Asp Gly Gln Val Ile Tyr Gln Ser

1340 1345 1350

Leu Thr Gly Leu Tyr Glu Thr Arg Met Arg Phe Glu Asp Lys Arg

1355 1360 1365

Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Ser

1370 1375 1380

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser

1385 1390 1395

Ser Gly Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His

1400 1405 1410

Ala Leu Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro

1415 1420 1425

Val Gly Ala Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly

1430 1435 1440

Trp Asn Arg Ala Ile Gly Leu His Asp Pro Thr Ala His Ala Glu

1445 1450 1455

Ile Met Ala Leu Arg Gln Gly Gly Leu Val Met Gln Asn Tyr Arg

1460 1465 1470

Leu Tyr Asp Ala Thr Leu Tyr Val Thr Phe Glu Pro Cys Val Met

1475 1480 1485

Cys Ala Gly Ala Met Ile His Ser Arg Ile Gly Arg Val Val Phe

1490 1495 1500

Gly Val Arg Asn Ala Lys Thr Gly Ala Ala Gly Ser Leu Met Asp

1505 1510 1515

Val Leu His His Pro Gly Met Asn His Arg Val Glu Ile Thr Glu

1520 1525 1530

Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu Leu Cys Arg Phe Phe

1535 1540 1545

Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys Lys Ala Gln Ser

1550 1555 1560

Ser Thr Asp Pro Ala Ala Lys Arg Val Lys Leu Asp Gly Ser Tyr

1565 1570 1575

Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp

1580 1585 1590

Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1595 1600

<210> 93

<211> 1869

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 93

Met Pro Ala Ala Lys Arg Val Lys Leu Asp Thr Ser Glu Lys Gly Pro

1 5 10 15

Ser Thr Gly Asp Pro Thr Leu Arg Arg Arg Ile Glu Ser Trp Glu Phe

```
            20                  25                  30
Asp Val Phe Tyr Asp Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu
        35                  40                  45

Tyr Glu Ile Lys Trp Gly Met Ser Arg Lys Ile Trp Arg Ser Ser Gly
    50                  55                  60
Lys Asn Thr Thr Asn His Val Glu Val Asn Phe Ile Lys Lys Phe Thr
65                  70                  75                  80
Ser Glu Arg Arg Phe His Ser Ser Ile Ser Cys Ser Ile Thr Trp Phe
                85                  90                  95
Leu Ser Trp Ser Pro Cys Trp Glu Cys Ser Gln Ala Ile Arg Glu Phe
            100                 105                 110

Leu Ser Gln His Pro Gly Val Thr Leu Val Ile Tyr Val Ala Arg Leu
        115                 120                 125
Phe Trp His Met Asp Gln Arg Asn Arg Gln Gly Leu Arg Asp Leu Val
    130                 135                 140
Asn Ser Gly Val Thr Ile Gln Ile Met Arg Ala Ser Glu Tyr Tyr His
145                 150                 155                 160
Cys Trp Arg Asn Phe Val Asn Tyr Pro Pro Gly Asp Glu Ala His Trp
                165                 170                 175

Pro Gln Tyr Pro Pro Leu Trp Met Met Leu Tyr Ala Leu Glu Leu His
            180                 185                 190
Cys Ile Ile Leu Ser Leu Pro Pro Cys Leu Lys Ile Ser Arg Arg Trp
        195                 200                 205
Gln Asn His Leu Ala Phe Phe Arg Leu His Leu Gln Asn Cys His Tyr
    210                 215                 220
Gln Thr Ile Pro Pro His Ile Leu Leu Ala Thr Gly Leu Ile His Pro
225                 230                 235                 240

Ser Val Thr Trp Arg Leu Lys Ser Gly Gly Ser Ser Gly Gly Ser Ser
                245                 250                 255
Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser
            260                 265                 270
```

Gly Gly Ser Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val Gly Glu
275 280 285
Lys Lys Thr Asn Tyr Thr Ile Gly Leu Ala Ile Gly Thr Asp Ser Val
290 295 300

Gly Trp Ala Val Val Lys Asp Asp Leu Glu Leu Val Lys Lys Arg Met
305 310 315 320
Lys Val Leu Gly Asn Thr Glu Thr Asn Tyr Ile Lys Lys Asn Leu Trp
325 330 335
Gly Ser Leu Leu Phe Glu Ser Gly Gln Thr Ala Lys Asp Arg Arg Leu
340 345 350
Lys Arg Val Ala Arg Arg Arg Tyr Glu Arg Arg Arg Asn Arg Leu Thr
355 360 365

Glu Leu Gln Lys Ile Phe Ala Pro Ala Ile Asp Glu Val Asp Glu Asn
370 375 380
Phe Phe Phe Arg Leu Asn Glu Ser Phe Leu Val Pro Glu Asp Lys Ala
385 390 395 400
Phe Ser Lys Asn Pro Ile Phe Gly Thr Leu Gly Glu Asp Lys Thr Tyr
405 410 415
Tyr Lys Thr Tyr Pro Thr Ile Tyr His Leu Arg Gln His Leu Ala Asp
420 425 430

Ser Glu Glu Lys Ala Asp Val Arg Leu Ile Tyr Leu Ala Leu Ala His
435 440 445
Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Lys Leu Asp Thr
450 455 460
Glu His Ile Ala Ile Asn Glu Asn Leu Glu Gln Phe Phe Glu Ser Tyr
465 470 475 480
Asn Ala Leu Phe Ser Glu Glu Pro Ile Glu Leu Arg Lys Glu Glu Leu
485 490 495

Ile Ala Ile Glu Asn Ile Leu Arg Glu Lys Asn Ser Arg Thr Val Lys
500 505 510
Glu Lys Arg Ile Thr Ser Phe Leu Lys Asp Ile Gly Arg Ala Asn Lys

515 520 525

Gln Ser Pro Met Met Ala Phe Ile Thr Leu Ile Val Gly Lys Lys Ala

530 535 540

Lys Phe Lys Ala Ala Phe Asn Leu Glu Glu Glu Ile Ser Leu Asn Leu

545 550 555 560

Thr Asp Asp Ser Tyr Asp Glu Asn Leu Glu Ile Leu Leu Asn Thr Ile

565 570 575

Gly Ser Asp Phe Ala Asp Leu Phe Asp His Ala Gln Arg Val Tyr Asn

580 585 590

Ala Val Glu Leu Ala Gly Ile Leu Ser Gly Asp Val Lys Asn Thr His

595 600 605

Ala Lys Leu Ser Ala Gln Met Val Ala Met Tyr Glu Arg His Lys Glu

610 615 620

Gln Leu Lys Glu Tyr Lys Ser Phe Ile Lys Ala Asn Leu Pro Asp Gln

625 630 635 640

Tyr Asp Met Thr Phe Val Ala Pro Lys Asp Ala Gln Lys Lys Asp Leu

645 650 655

Lys Gly Tyr Ala Gly Tyr Ile Asp Gly Asn Met Ser Gln Asp Ser Phe

660 665 670

Tyr Lys Phe Val Lys Asp Gln Leu Lys Glu Val Pro Gly Ser Glu Lys

675 680 685

Phe Leu Asp Ser Ile Glu Lys Glu Asp Phe Leu Arg Lys Gln Arg Ser

690 695 700

Phe Tyr Asn Gly Val Ile Pro Asn Gln Val His Leu Ala Glu Met Glu

705 710 715 720

Ala Ile Leu Asp Arg Gln Glu Asn Tyr Tyr Pro Trp Leu Lys Glu Asn

725 730 735

Arg Glu Lys Ile Ile Ser Leu Leu Thr Phe Arg Ile Pro Tyr Tyr Val

740 745 750

Gly Pro Leu Ala Asp Gly Gln Ser Glu Phe Ala Trp Leu Glu Arg Lys

755 760 765

Ser Asp Glu Lys Ile Lys Pro Trp Asn Phe Ser Asp Val Val Asp Leu
770 775 780
Asp Arg Ser Ala Glu Lys Phe Ile Glu Gln Leu Ile Gly Arg Asp Thr
785 790 795 800
Tyr Leu Pro Asp Glu Tyr Val Leu Pro Lys Lys Ser Leu Ile Tyr Gln
805 810 815

Lys Tyr Met Val Phe Asn Glu Leu Thr Lys Ile Ala Tyr Leu Asp Glu
820 825 830
Arg Gln Lys Arg Met Asn Leu Ser Ser Val Glu Lys Lys Glu Ile Phe
835 840 845
Glu Thr Leu Phe Lys Lys Arg Ser Lys Val Thr Glu Lys Gln Leu Val
850 855 860
Lys Phe Phe Glu Asn Tyr Leu Gln Ile Asp Asn Pro Thr Ile Phe Gly
865 870 875 880

Ile Glu Asp Ala Phe Asn Ala Asp Tyr Ser Thr Tyr Val Glu Leu Ala
885 890 895
Lys Val Pro Gly Met Lys Ser Met Met Asp Asp Pro Asp Asn Glu Asp
900 905 910
Leu Met Glu Glu Ile Val Lys Ile Leu Thr Val Phe Glu Asp Arg Lys
915 920 925
Met Arg Arg Lys Gln Leu Glu Lys Tyr Lys Glu Arg Leu Ser Pro Glu
930 935 940

Gln Ile Lys Glu Leu Ala Lys Lys His Tyr Thr Gly Trp Gly Arg Leu
945 950 955 960
Ser Lys Lys Leu Leu Val Gly Ile Arg Asp Lys Glu Thr Gln Lys Thr
965 970 975
Ile Leu Asp Tyr Leu Val Glu Asp Asp Asn His Ser Gly Gly Arg Gln
980 985 990
His Leu Asn Arg Asn Leu Met Gln Leu Ile Asn Asp Asp Arg Leu Ser
995 1000 1005

Phe Lys Lys Thr Ile Ala Glu Leu Gln Met Ile Asp Pro Ser Ala

1010 1015 1020

Asp Leu Tyr Ala Gln Val Gln Glu Ile Ala Gly Ser Pro Ala Ile

1025 1030 1035

Lys Lys Gly Ile Leu Leu Gly Leu Lys Ile Val Asp Glu Ile Ile

1040 1045 1050

Arg Val Met Gly Glu Lys Pro Glu Asn Ile Val Ile Glu Met Ala

1055 1060 1065

Arg Glu Asn Gln Thr Thr Ala Arg Gly Lys Ala Leu Ser Lys Arg

1070 1075 1080

Arg Glu Ala Lys Ile Lys Glu Gly Leu Ala Ala Leu Gly Ser Ser

1085 1090 1095

Leu Leu Lys Glu Asn Leu Pro Gly Asn Ala Asp Leu Ser Gln Arg

1100 1105 1110

Lys Ile Tyr Leu Tyr Tyr Thr Gln Asn Gly Lys Asp Ile Tyr Leu

1115 1120 1125

Asp Glu Pro Leu Asp Phe Asp Arg Leu Ser Gln Tyr Asp Glu Asp

1130 1135 1140

His Ile Ile Pro Gln Ser Phe Thr Val Asp Asn Ser Leu Asp Asn

1145 1150 1155

Leu Val Leu Thr Asn Ser Ser Gln Asn Arg Gly Asn Lys Lys Asp

1160 1165 1170

Asp Val Pro Ser Leu Glu Val Val Asn Arg Gln Leu Ala Tyr Trp

1175 1180 1185

Arg Ser Leu Lys Asp Ala Gly Leu Met Thr Gln Arg Lys Phe Asp

1190 1195 1200

Asn Leu Thr Lys Ala Met Arg Gly Gly Leu Thr Asp Lys Asp Arg

1205 1210 1215

Glu Arg Phe Ile Gln Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

1220 1225 1230

Lys Asn Val Ala Lys Leu Leu Asp Met Arg Leu Asn Asp Lys Lys

1235 1240 1245

Asp Glu Ala Gly Asn Lys Ile Arg Glu Thr Asn Ile Val Leu Leu
1250 1255 1260
Lys Ser Ala Met Ala Ser Glu Phe Arg Lys Met Phe Arg Leu Tyr
1265 1270 1275
Lys Val Arg Glu Leu Asn Asp Tyr His His Ala His Asp Ala Tyr
1280 1285 1290
Leu Asn Ala Ala Ile Ala Ile Asn Leu Leu Ala Leu Tyr Pro Tyr
1295 1300 1305

Met Ala Asp Asp Phe Val Tyr Gly Glu Phe Arg Tyr Lys Lys Lys
1310 1315 1320
Pro Gln Ala Glu Lys Ala Thr Tyr Glu Lys Leu Arg Gln Trp Asn
1325 1330 1335
Leu Ile Lys Arg Phe Gly Glu Lys Gln Leu Phe Thr Pro Asp His
1340 1345 1350
Glu Asp Cys Trp Asn Lys Glu Arg Asp Ile Lys Thr Ile Lys Lys
1355 1360 1365

Val Met Gly Tyr Arg Gln Val Asn Val Val Lys Lys Ala Glu Glu
1370 1375 1380
Arg Thr Gly Met Leu Phe Lys Glu Thr Ile Asn Gly Lys Thr Asn
1385 1390 1395
Lys Gly Ser Arg Ile Pro Ile Lys Lys Asp Leu Asp Pro Ser Lys
1400 1405 1410
Tyr Gly Gly Tyr Ile Glu Glu Lys Met Ala Tyr Tyr Ala Val Ile
1415 1420 1425

Ser Tyr Glu Asp Lys Lys Lys Lys Pro Gly Lys Thr Ile Val Gly
1430 1435 1440
Ile Ser Ile Met Asp Lys Lys Glu Phe Glu Tyr Asp Ser Ile Ser
1445 1450 1455
Tyr Leu Gly Lys Leu Gly Phe Ser Asn Pro Val Val Gln Ile Ile
1460 1465 1470
Leu Lys Asn Tyr Ser Leu Ile Ala Tyr Pro Asp Gly Arg Arg Arg

1475 1480 1485

Tyr Ile Thr Gly Ala Thr Lys Thr Thr Lys Gly Lys Val Glu Leu
1490 1495 1500
Gln Lys Ala Asn Gln Ile Ala Met Glu Gln Asp Leu Val Asn Phe
1505 1510 1515
Ile Tyr His Leu Lys Asn Tyr Asp Glu Ile Ser His Pro Glu Ser
1520 1525 1530
Tyr Ala Phe Val Gln Ser His Thr Asp Tyr Phe Asp Arg Leu Phe
1535 1540 1545

Asp Ser Ile Glu His Tyr Thr Arg Arg Phe Leu Asp Ala Glu Thr
1550 1555 1560
Asn Ile Asn Arg Leu Arg Arg Ile Tyr Glu Glu Glu Lys Lys Lys
1565 1570 1575
Asp Pro Val Asp Ile Glu Ala Leu Val Ala Ser Phe Ile Glu Leu
1580 1585 1590
Leu Lys Leu Thr Ser Ala Gly Ala Pro Ala Asp Phe Ile Phe Met
1595 1600 1605

Gly Glu Ala Ile Ser Arg Arg Arg Tyr Asn Ser Met Thr Gly Leu
1610 1615 1620
Phe Asp Gly Gln Val Ile Tyr Gln Ser Leu Thr Gly Leu Tyr Glu
1625 1630 1635
Thr Arg Met Arg Phe Glu Asp Gly Lys Arg Pro Ala Ala Thr Lys
1640 1645 1650
Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly Ser Ser Gly Gly Ser
1655 1660 1665

Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys
1670 1675 1680
Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu
1685 1690 1695
Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp
1700 1705 1710

Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val
1715 1720 1725

Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu
1730 1735 1740
Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser
1745 1750 1755
Gly Gly Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile
1760 1765 1770
Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile
1775 1780 1785

Leu Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro
1790 1795 1800
Glu Ser Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp
1805 1810 1815
Glu Asn Val Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro
1820 1825 1830
Trp Ala Leu Val Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys
1835 1840 1845

Met Leu Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp
1850 1855 1860
Val Pro Asp Tyr Ala Tyr
1865
<210> 94
<211> 1873
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polypeptide
<400> 94
Met Pro Ala Ala Lys Arg Val Lys Leu Asp Thr Asn Leu Ser Asp Ile
1 5 10 15

Ile Glu Lys Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu

20 25 30

Met Leu Pro Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser

35 40 45

Asp Ile Leu Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val

50 55 60

Met Leu Leu Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val

65 70 75 80

Ile Gln Asp Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly

85 90 95

Ser Gly Gly Ser Gly Gly Ser Thr Asn Leu Ser Asp Ile Ile Glu Lys

100 105 110

Glu Thr Gly Lys Gln Leu Val Ile Gln Glu Ser Ile Leu Met Leu Pro

115 120 125

Glu Glu Val Glu Glu Val Ile Gly Asn Lys Pro Glu Ser Asp Ile Leu

130 135 140

Val His Thr Ala Tyr Asp Glu Ser Thr Asp Glu Asn Val Met Leu Leu

145 150 155 160

Thr Ser Asp Ala Pro Glu Tyr Lys Pro Trp Ala Leu Val Ile Gln Asp

165 170 175

Ser Asn Gly Glu Asn Lys Ile Lys Met Leu Ser Gly Gly Ser Gly Gly

180 185 190

Ser Gly Gly Ser Pro Lys Lys Lys Arg Lys Val Glu Lys Lys Thr Asn

195 200 205

Tyr Thr Ile Gly Leu Ala Ile Gly Thr Asp Ser Val Gly Trp Ala Val

210 215 220

Val Lys Asp Asp Leu Glu Leu Val Lys Lys Arg Met Lys Val Leu Gly

225 230 235 240

Asn Thr Glu Thr Asn Tyr Ile Lys Lys Asn Leu Trp Gly Ser Leu Leu

245 250 255

Phe Glu Ser Gly Gln Thr Ala Lys Asp Arg Arg Leu Lys Arg Val Ala

260 265 270

Arg Arg Arg Tyr Glu Arg Arg Arg Asn Arg Leu Thr Glu Leu Gln Lys

275 280 285

Ile Phe Ala Pro Ala Ile Asp Glu Val Asp Glu Asn Phe Phe Phe Arg

290 295 300

Leu Asn Glu Ser Phe Leu Val Pro Glu Asp Lys Ala Phe Ser Lys Asn

305 310 315 320

Pro Ile Phe Gly Thr Leu Gly Glu Asp Lys Thr Tyr Tyr Lys Thr Tyr

325 330 335

Pro Thr Ile Tyr His Leu Arg Gln His Leu Ala Asp Ser Glu Glu Lys

340 345 350

Ala Asp Val Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile Lys Tyr

355 360 365

Arg Gly His Phe Leu Ile Glu Gly Lys Leu Asp Thr Glu His Ile Ala

370 375 380

Ile Asn Glu Asn Leu Glu Gln Phe Phe Glu Ser Tyr Asn Ala Leu Phe

385 390 395 400

Ser Glu Glu Pro Ile Glu Leu Arg Lys Glu Glu Leu Ile Ala Ile Glu

405 410 415

Asn Ile Leu Arg Glu Lys Asn Ser Arg Thr Val Lys Glu Lys Arg Ile

420 425 430

Thr Ser Phe Leu Lys Asp Ile Gly Arg Ala Asn Lys Gln Ser Pro Met

435 440 445

Met Ala Phe Ile Thr Leu Ile Val Gly Lys Lys Ala Lys Phe Lys Ala

450 455 460

Ala Phe Asn Leu Glu Glu Glu Ile Ser Leu Asn Leu Thr Asp Asp Ser

465 470 475 480

Tyr Asp Glu Asn Leu Glu Ile Leu Leu Asn Thr Ile Gly Ser Asp Phe

485 490 495

Ala Asp Leu Phe Asp His Ala Gln Arg Val Tyr Asn Ala Val Glu Leu

500 505 510

Ala Gly Ile Leu Ser Gly Asp Val Lys Asn Thr His Ala Lys Leu Ser
515 520 525
Ala Gln Met Val Ala Met Tyr Glu Arg His Lys Glu Gln Leu Lys Glu

530 535 540
Tyr Lys Ser Phe Ile Lys Ala Asn Leu Pro Asp Gln Tyr Asp Met Thr
545 550 555 560
Phe Val Ala Pro Lys Asp Ala Gln Lys Lys Asp Leu Lys Gly Tyr Ala
565 570 575
Gly Tyr Ile Asp Gly Asn Met Ser Gln Asp Ser Phe Tyr Lys Phe Val
580 585 590
Lys Asp Gln Leu Lys Glu Val Pro Gly Ser Glu Lys Phe Leu Asp Ser

595 600 605
Ile Glu Lys Glu Asp Phe Leu Arg Lys Gln Arg Ser Phe Tyr Asn Gly
610 615 620
Val Ile Pro Asn Gln Val His Leu Ala Glu Met Glu Ala Ile Leu Asp
625 630 635 640
Arg Gln Glu Asn Tyr Tyr Pro Trp Leu Lys Glu Asn Arg Glu Lys Ile
645 650 655
Ile Ser Leu Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala

660 665 670
Asp Gly Gln Ser Glu Phe Ala Trp Leu Glu Arg Lys Ser Asp Glu Lys
675 680 685
Ile Lys Pro Trp Asn Phe Ser Asp Val Val Asp Leu Asp Arg Ser Ala
690 695 700
Glu Lys Phe Ile Glu Gln Leu Ile Gly Arg Asp Thr Tyr Leu Pro Asp
705 710 715 720
Glu Tyr Val Leu Pro Lys Lys Ser Leu Ile Tyr Gln Lys Tyr Met Val

725 730 735
Phe Asn Glu Leu Thr Lys Ile Ala Tyr Leu Asp Glu Arg Gln Lys Arg
740 745 750
Met Asn Leu Ser Ser Val Glu Lys Lys Glu Ile Phe Glu Thr Leu Phe

```
        755                 760                 765
Lys Lys Arg Ser Lys Val Thr Glu Lys Gln Leu Val Lys Phe Phe Glu
    770                 775                 780
Asn Tyr Leu Gln Ile Asp Asn Pro Thr Ile Phe Gly Ile Glu Asp Ala

785                 790                 795                 800
Phe Asn Ala Asp Tyr Ser Thr Tyr Val Glu Leu Ala Lys Val Pro Gly
                805                 810                 815
Met Lys Ser Met Met Asp Asp Pro Asp Asn Glu Asp Leu Met Glu Glu
            820                 825                 830
Ile Val Lys Ile Leu Thr Val Phe Glu Asp Arg Lys Met Arg Arg Lys
        835                 840                 845
Gln Leu Glu Lys Tyr Lys Glu Arg Leu Ser Pro Glu Gln Ile Lys Glu

    850                 855                 860
Leu Ala Lys Lys His Tyr Thr Gly Trp Gly Arg Leu Ser Lys Lys Leu
865                 870                 875                 880
Leu Val Gly Ile Arg Asp Lys Glu Thr Gln Lys Thr Ile Leu Asp Tyr
                885                 890                 895
Leu Val Glu Asp Asp Asn His Ser Gly Gly Arg Gln His Leu Asn Arg
            900                 905                 910
Asn Leu Met Gln Leu Ile Asn Asp Asp Arg Leu Ser Phe Lys Lys Thr

        915                 920                 925
Ile Ala Glu Leu Gln Met Ile Asp Pro Ser Ala Asp Leu Tyr Ala Gln
    930                 935                 940
Val Gln Glu Ile Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Leu
945                 950                 955                 960
Gly Leu Lys Ile Val Asp Glu Ile Ile Arg Val Met Gly Glu Lys Pro
                965                 970                 975
Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Ala Arg

            980                 985                 990
Gly Lys Ala Leu Ser Lys Arg Arg  Glu Ala Lys Ile Lys  Glu Gly Leu
        995                 1000                1005
```

Ala Ala Leu Gly Ser Ser Leu Leu Lys Glu Asn Leu Pro Gly Asn
1010 1015 1020
Ala Asp Leu Ser Gln Arg Lys Ile Tyr Leu Tyr Tyr Thr Gln Asn
1025 1030 1035
Gly Lys Asp Ile Tyr Leu Asp Glu Pro Leu Asp Phe Asp Arg Leu

1040 1045 1050
Ser Gln Tyr Asp Glu Asp His Ile Ile Pro Gln Ser Phe Thr Val
1055 1060 1065
Asp Asn Ser Leu Asp Asn Leu Val Leu Thr Asn Ser Ser Gln Asn
1070 1075 1080
Arg Gly Asn Lys Lys Asp Asp Val Pro Ser Leu Glu Val Val Asn
1085 1090 1095
Arg Gln Leu Ala Tyr Trp Arg Ser Leu Lys Asp Ala Gly Leu Met

1100 1105 1110
Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Met Arg Gly Gly
1115 1120 1125
Leu Thr Asp Lys Asp Arg Glu Arg Phe Ile Gln Arg Gln Leu Val
1130 1135 1140
Glu Thr Arg Gln Ile Thr Lys Asn Val Ala Lys Leu Leu Asp Met
1145 1150 1155
Arg Leu Asn Asp Lys Lys Asp Glu Ala Gly Asn Lys Ile Arg Glu

1160 1165 1170
Thr Asn Ile Val Leu Leu Lys Ser Ala Met Ala Ser Glu Phe Arg
1175 1180 1185
Lys Met Phe Arg Leu Tyr Lys Val Arg Glu Leu Asn Asp Tyr His
1190 1195 1200
His Ala His Asp Ala Tyr Leu Asn Ala Ala Ile Ala Ile Asn Leu
1205 1210 1215
Leu Ala Leu Tyr Pro Tyr Met Ala Asp Asp Phe Val Tyr Gly Glu

1220 1225 1230
Phe Arg Tyr Lys Lys Lys Pro Gln Ala Glu Lys Ala Thr Tyr Glu

1235 1240 1245

Lys Leu Arg Gln Trp Asn Leu Ile Lys Arg Phe Gly Glu Lys Gln

1250 1255 1260

Leu Phe Thr Pro Asp His Glu Asp Cys Trp Asn Lys Glu Arg Asp

1265 1270 1275

Ile Lys Thr Ile Lys Lys Val Met Gly Tyr Arg Gln Val Asn Val

1280 1285 1290

Val Lys Lys Ala Glu Glu Arg Thr Gly Met Leu Phe Lys Glu Thr

1295 1300 1305

Ile Asn Gly Lys Thr Asn Lys Gly Ser Arg Ile Pro Ile Lys Lys

1310 1315 1320

Asp Leu Asp Pro Ser Lys Tyr Gly Gly Tyr Ile Glu Glu Lys Met

1325 1330 1335

Ala Tyr Tyr Ala Val Ile Ser Tyr Glu Asp Lys Lys Lys Lys Pro

1340 1345 1350

Gly Lys Thr Ile Val Gly Ile Ser Ile Met Asp Lys Lys Glu Phe

1355 1360 1365

Glu Tyr Asp Ser Ile Ser Tyr Leu Gly Lys Leu Gly Phe Ser Asn

1370 1375 1380

Pro Val Val Gln Ile Ile Leu Lys Asn Tyr Ser Leu Ile Ala Tyr

1385 1390 1395

Pro Asp Gly Arg Arg Arg Tyr Ile Thr Gly Ala Thr Lys Thr Thr

1400 1405 1410

Lys Gly Lys Val Glu Leu Gln Lys Ala Asn Gln Ile Ala Met Glu

1415 1420 1425

Gln Asp Leu Val Asn Phe Ile Tyr His Leu Lys Asn Tyr Asp Glu

1430 1435 1440

Ile Ser His Pro Glu Ser Tyr Ala Phe Val Gln Ser His Thr Asp

1445 1450 1455

Tyr Phe Asp Arg Leu Phe Asp Ser Ile Glu His Tyr Thr Arg Arg

1460 1465 1470

Phe Leu Asp Ala Glu Thr Asn Ile Asn Arg Leu Arg Arg Ile Tyr
1475 1480 1485
Glu Glu Glu Lys Lys Lys Asp Pro Val Asp Ile Glu Ala Leu Val
1490 1495 1500
Ala Ser Phe Ile Glu Leu Leu Lys Leu Thr Ser Ala Gly Ala Pro
1505 1510 1515
Ala Asp Phe Ile Phe Met Gly Glu Ala Ile Ser Arg Arg Arg Tyr

1520 1525 1530
Asn Ser Met Thr Gly Leu Phe Asp Gly Gln Val Ile Tyr Gln Ser
1535 1540 1545
Leu Thr Gly Leu Tyr Glu Thr Arg Met Arg Phe Glu Asp Lys Arg
1550 1555 1560
Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Gly
1565 1570 1575
Ser Ser Gly Gly Ser Ser Gly Gly Ser Ser Gly Ser Glu Thr Pro

1580 1585 1590
Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser Gly Gly Ser Ser
1595 1600 1605
Gly Gly Ser Thr Ser Glu Lys Gly Pro Ser Thr Gly Asp Pro Thr
1610 1615 1620
Leu Arg Arg Arg Ile Glu Ser Trp Glu Phe Asp Val Phe Tyr Asp
1625 1630 1635
Pro Arg Glu Leu Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Lys

1640 1645 1650
Trp Gly Met Ser Arg Lys Ile Trp Arg Ser Ser Gly Lys Asn Thr
1655 1660 1665
Thr Asn His Val Glu Val Asn Phe Ile Lys Lys Phe Thr Ser Glu
1670 1675 1680
Arg Arg Phe His Ser Ser Ile Ser Cys Ser Ile Thr Trp Phe Leu
1685 1690 1695
Ser Trp Ser Pro Cys Trp Glu Cys Ser Gln Ala Ile Arg Glu Phe

1700 1705 1710

Leu Ser Gln His Pro Gly Val Thr Leu Val Ile Tyr Val Ala Arg

1715 1720 1725

Leu Phe Trp His Met Asp Gln Arg Asn Arg Gln Gly Leu Arg Asp

1730 1735 1740

Leu Val Asn Ser Gly Val Thr Ile Gln Ile Met Arg Ala Ser Glu

1745 1750 1755

Tyr Tyr His Cys Trp Arg Asn Phe Val Asn Tyr Pro Pro Gly Asp

1760 1765 1770

Glu Ala His Trp Pro Gln Tyr Pro Pro Leu Trp Met Met Leu Tyr

1775 1780 1785

Ala Leu Glu Leu His Cys Ile Ile Leu Ser Leu Pro Pro Cys Leu

1790 1795 1800

Lys Ile Ser Arg Arg Trp Gln Asn His Leu Ala Phe Phe Arg Leu

1805 1810 1815

His Leu Gln Asn Cys His Tyr Gln Thr Ile Pro Pro His Ile Leu

1820 1825 1830

Leu Ala Thr Gly Leu Ile His Pro Ser Val Thr Trp Arg Tyr Pro

1835 1840 1845

Tyr Asp Val Pro Asp Tyr Ala Tyr Pro Tyr Asp Val Pro Asp Tyr

1850 1855 1860

Ala Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1865 1870

<210> 95

<211> 125

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 95

gucuuggaug agugugaaaa cacucauagu caagaucaaa cgagugguuu uccacgaguu 60

auuacuuuug aggucuuaua uggcccauac auaaaaagga gucggaauuu ccggcuccuu 120

uucuu 125

<210> 96

<211> 88

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 96

guuuuagagc cauguagaaa uacauugcaa guuaaaauaa ggcuuugucc guaaucaacu 60

ugaaaaagug gcgcuguuuc ggcgcuuu 88

<210> 97

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 97

gaaacaatga taacaagacc 20

<210> 98

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 98

gtggcccctg tgcccagccc 20

<210> 99

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 99

gtcccaaata tgtagctgtt 20

<210> 100

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 100

gctcccatca catcaaccgg 20

<210> 101

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 101

gatgtcacct ccaatgacta 20

<210> 102

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 102

gttgaagatg aagcccagag 20

<210> 103

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 103

gccaacacca accagaactt 20

<210> 104

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 104

tgctgcacac agcaggcctt 20

<210> 105

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 105

gaacacaaag catagactgc 20

<210> 106

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 106

gaaacaatga taacaagacc 20

<210> 107

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 107

gtggcccctg tgcccagccc 20

<210> 108

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 108

gtcccaaata tgtagctgtt 20

<210> 109

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 109

agagggacac acagatctat 20

<210> 110

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 110

gctcccatca catcaaccgg 20

<210> 111

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 111

gatgtcacct ccaatgacta 20

<210> 112

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 112

gggcaaccac aaacccacga 20

<210> 113

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 113

gttgaagatg aagcccagag 20

<210> 114

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 114

gccaacacca accagaactt 20

<210> 115

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 115

tgctgcacac agcaggcctt 20

<210> 116

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 116

gtgccagaaa caggggtgac 20

<210> 117

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 117

cctgcctcag ctgctcactt 20

<210> 118

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 118

aaacggtccc cagagggttc 20

<210> 119

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 119

gccaccggtt gatgtgatgg 20

<210> 120

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 120

aagtggtccc aggcctcagc 20

<210> 121

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 121

agagaaaatg aaactttcaa 20

<210> 122

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 122

ccaaacccaa ctccatctac 20

<210> 123

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 123

ggtccttgaa ttgcagtatc 20

<210> 124

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 124

gcatagactg cggggcgggc 20

<210> 125

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 125

ggaaactgga acacaaagca 20

<210> 126

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 126

gacagcatgt ggtaattttc 20

<210> 127

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 127

gcccccggaa actctgtcca 20

<210> 128

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 128

tcgaccccca ccaaggttca 20

<210> 129

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 129

tgttaacagc tgacccaata 20

<210> 130

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 130

gttactcgcc tgtcaagtgg 20

<210> 131

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 131

gggctcccat cacatcaacc 20

<210> 132

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 132

gctttgggga ggcctggagt 20

<210> 133

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 133

tagctgccaa tgactatagc 20

<210> 134

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 134

ttaaaatagg atctacatca 20

<210> 135

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 135

gaatcctgcc atacactttg 20

<210> 136

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 136

ctgcggggcg ggccagcctg 20

<210> 137

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 137

acattgtcag agggacacac 20

<210> 138

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 138

agcaactcca gtcccaaata 20

<210> 139

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 139

gtggtggccg agcgcccct 20

<210> 140

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 140

cattcaccca gcttccctgt 20

<210> 141

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 141

gaacacaaag catagactgc 20

<210> 142

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 142

gaaacaatga taacaagacc 20

<210> 143

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 143

gtggcccctg tgcccagccc 20

<210> 144

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 144

gtcccaaata tgtagctgtt 20

<210> 145

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 145

agagggacac acagatctat 20

<210> 146

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 146

gctcccatca catcaaccgg 20

<210> 147

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 147

gatgtcacct ccaatgacta 20

<210> 148

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 148

gggcaaccac aaacccacga 20

<210> 149

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 149

gttgaagatg aagcccagag 20

<210> 150

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 150

gccaacacca accagaactt 20

<210> 151

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 151

tgctgcacac agcaggcctt 20

<210> 152

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 152

gtgccagaaa caggggtgac 20

<210> 153

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 153

tcacggagac tgaacactcc 20

<210> 154

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 154

gtaacagaca tggaccatca 20

<210> 155

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 155

gggagggagg ggcacagatg 20

<210> 156

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 156

tgtggttcca gaaccggagg 20

<210> 157

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 157

aatgagagaa aatgaaactt 20

<210> 158

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 158

ggccatcaag gatgcccacg 20

<210> 159

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 159

aaattgtcca gccccatctg 20

<210> 160

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 160

cctgtaaagg aaactggaac 20

<210> 161

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 161

tacatgaagc aactccagtc 20

<210> 162

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 162

aaactccccc cacccccttt 20

<210> 163

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 163

gagttgggtt tggtgctcaa 20

<210> 164

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 164

gcgggccagc ctgaatagct 20

<210> 165

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 165

gaaacaatga taacaagacc 20

<210> 166

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 166

gtggcccctg tgcccagccc 20

<210> 167

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 167

gtcccaaata tgtagctgtt 20

<210> 168

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 168

gctcccatca catcaaccgg 20

<210> 169

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 169

gatgtcacct ccaatgacta 20

<210> 170

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 170

gttgaagatg aagcccagag 20

<210> 171

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 171

gccaacacca accagaactt 20

<210> 172

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 172

tgctgcacac agcaggcctt 20

<210> 173

<211> 1368

<212> PRT

<213> Streptococcus pyogenes

<400> 173

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr

545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp

565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly

580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp

595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr

610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr

645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp

660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe

690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu

705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly

725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly

740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln

755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro

785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 174
<211> 60
<212> RNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide
<400> 174
guuuuagagc ugugcuguuu aaacaacaca gcaaguuaaa auaaggcuuu guccguacuc 60

<210> 175
<211> 1377
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 175
Met Pro Lys Lys Lys Arg Lys Val Gly Met Gly Lys Pro Tyr Ser Ile
1 5 10 15
Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Val Thr Asp
20 25 30
Asp Tyr Lys Val Pro Ala Lys Lys Met Lys Val Leu Gly Asn Thr Asp

35 40 45

Lys Gln Ser Ile Lys Lys Asn Leu Leu Gly Ala Leu Leu Phe Asp Ser
50 55 60
Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg
65 70 75 80
Tyr Thr Arg Arg Lys Asn Arg Leu Arg Tyr Leu Gln Glu Ile Phe Thr
85 90 95
Gly Glu Met Asn Lys Val Asp Glu Asn Phe Phe Gln Arg Leu Asp Asp
100 105 110

Ser Phe Leu Val Asp Glu Asp Lys Arg Gly Glu His His Pro Ile Phe
115 120 125
Gly Asn Ile Ala Ala Glu Val Lys Tyr His Asp Asp Phe Pro Thr Ile
130 135 140
Tyr His Leu Arg Arg His Leu Ala Asp Thr Ser Lys Lys Ala Asp Leu
145 150 155 160
Arg Leu Val Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His
165 170 175

Phe Leu Tyr Glu Gly Asp Leu Lys Ala Glu Asn Thr Asp Val Gln Ala
180 185 190
Leu Phe Lys Asp Phe Val Glu Glu Tyr Asp Lys Thr Ile Glu Glu Ser
195 200 205
His Leu Ser Glu Ile Thr Val Asp Ala Leu Ser Ile Leu Thr Glu Lys
210 215 220
Val Ser Lys Ser Ser Arg Leu Glu Asn Leu Ile Ala His Tyr Pro Thr
225 230 235 240

Glu Lys Lys Asn Thr Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Asp
245 250 255
Leu His Pro Asn Phe Lys Thr Asn Phe Gln Leu Ser Glu Asp Ala Lys
260 265 270
Leu Gln Phe Ser Lys Asp Thr Tyr Glu Glu Asp Leu Glu Gly Phe Leu
275 280 285

Gly Glu Val Gly Asp Glu Tyr Ala Asp Leu Phe Ala Ser Ala Lys Asn
290 295 300

Leu Tyr Asp Ala Ile Leu Leu Ser Gly Ile Leu Thr Val Asp Asp Asn
305 310 315 320
Ser Thr Lys Ala Pro Leu Ser Ala Ser Met Val Lys Arg Tyr Glu Glu
325 330 335
His Gln Lys Asp Leu Lys Lys Leu Lys Asp Phe Ile Lys Val Asn Ala
340 345 350
Pro Asp Gln Tyr Asn Ala Ile Phe Lys Asp Lys Asn Lys Lys Gly Tyr
355 360 365

Ala Ser Tyr Ile Glu Ser Gly Val Lys Gln Asp Glu Phe Tyr Lys Tyr
370 375 380
Leu Lys Gly Ile Leu Leu Lys Ile Asn Gly Ser Gly Asp Phe Leu Asp
385 390 395 400
Lys Ile Asp Arg Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn
405 410 415
Gly Ser Ile Pro His Gln Ile His Leu Gln Glu Met His Ala Ile Leu
420 425 430

Arg Arg Gln Gly Glu His Tyr Pro Phe Leu Lys Glu Asn Gln Asp Lys
435 440 445
Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu
450 455 460
Ala Arg Lys Gly Ser Arg Phe Ala Trp Ala Glu Tyr Lys Ala Asp Glu
465 470 475 480
Lys Ile Thr Pro Trp Asn Phe Asp Asp Ile Leu Asp Lys Glu Lys Ser
485 490 495

Ala Glu Lys Phe Ile Thr Arg Met Thr Leu Asn Asp Leu Tyr Leu Pro
500 505 510
Glu Glu Lys Val Leu Pro Lys His Ser Pro Leu Tyr Glu Ala Phe Thr
515 520 525
Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Asn Glu Gln Gly Glu

530 535 540

Ala Lys Phe Phe Asp Thr Asn Met Lys Gln Glu Ile Phe Asp His Val

545 550 555 560

Phe Lys Glu Asn Arg Lys Val Thr Lys Asp Lys Leu Leu Asn Tyr Leu

565 570 575

Asn Lys Glu Phe Glu Glu Phe Arg Ile Val Asn Leu Thr Gly Leu Asp

580 585 590

Lys Glu Asn Lys Ala Phe Asn Ser Ser Leu Gly Thr Tyr His Asp Leu

595 600 605

Arg Lys Ile Leu Asp Lys Ser Phe Leu Asp Asp Lys Ala Asn Glu Lys

610 615 620

Thr Ile Glu Asp Ile Ile Gln Thr Leu Thr Leu Phe Glu Asp Arg Glu

625 630 635 640

Met Ile Arg Gln Arg Leu Gln Lys Tyr Ser Asp Ile Phe Thr Lys Ala

645 650 655

Gln Leu Lys Lys Leu Glu Arg Arg His Tyr Thr Gly Trp Gly Arg Leu

660 665 670

Ser Tyr Lys Leu Ile Asn Gly Ile Arg Asn Lys Glu Asn Lys Lys Thr

675 680 685

Ile Leu Asp Tyr Leu Ile Asp Asp Gly Tyr Ala Asn Arg Asn Phe Met

690 695 700

Gln Leu Ile Asn Asp Asp Ala Leu Ser Phe Lys Glu Glu Ile Ala Arg

705 710 715 720

Ala Gln Ile Ile Asp Asp Val Asp Asp Ile Ala Asn Val Val His Asp

725 730 735

Leu Pro Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Val Lys

740 745 750

Ile Val Asp Glu Leu Val Lys Val Met Gly His Asn Pro Ala Asn Ile

755 760 765

Ile Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Asp Lys Gly Arg Arg

770 775 780

Asn Ser Gln Gln Arg Leu Lys Leu Leu Gln Asp Ser Leu Lys Asn Leu
785 790 795 800
Asp Asn Pro Val Asn Ile Lys Asn Val Glu Asn Gln Gln Leu Gln Asn
805 810 815

Asp Arg Leu Phe Leu Tyr Tyr Ile Gln Asn Gly Lys Asp Met Tyr Thr
820 825 830
Gly Glu Thr Leu Asp Ile Asn Asn Leu Ser Gln Tyr Asp Ile Asp His
835 840 845
Ile Ile Pro Gln Ala Phe Ile Lys Asp Asn Ser Leu Asp Asn Arg Val
850 855 860
Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asp Val Pro Ser
865 870 875 880

Ile Glu Val Val His Glu Met Lys Ser Phe Trp Ser Lys Leu Leu Ser
885 890 895
Val Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu
900 905 910
Arg Gly Gly Leu Thr Glu Glu Asp Lys Ala Gly Phe Ile Lys Arg Gln
915 920 925
Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp
930 935 940

Glu Arg Phe Asn Thr Glu Phe Asp Gly Asn Lys Arg Arg Ile Arg Asn
945 950 955 960
Val Lys Ile Ile Thr Leu Lys Ser Asn Leu Val Ser Asn Phe Arg Lys
965 970 975
Glu Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Tyr His His Ala
980 985 990
His Asp Ala Tyr Leu Asn Ala Val Val Gly Asn Ala Leu Leu Leu Lys
995 1000 1005

Tyr Pro Gln Leu Glu Pro Glu Phe Val Tyr Gly Glu Tyr Pro Lys
1010 1015 1020
Tyr Asn Ser Tyr Arg Ser Arg Lys Ser Ala Thr Glu Lys Phe Leu

1025 1030 1035

Phe Tyr Ser Asn Ile Leu Arg Phe Phe Lys Lys Glu Asp Ile Gln

1040 1045 1050

Thr Asn Glu Asp Gly Glu Ile Ala Trp Asn Lys Glu Lys His Ile

1055 1060 1065

Lys Ile Leu Arg Lys Val Leu Ser Tyr Pro Gln Val Asn Ile Val

1070 1075 1080

Lys Lys Thr Glu Glu Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1085 1090 1095

Leu Pro Lys Gly Glu Ser Asp Lys Leu Ile Pro Arg Lys Thr Lys

1100 1105 1110

Asn Ser Tyr Trp Asp Pro Lys Lys Tyr Gly Gly Phe Met Gln Pro

1115 1120 1125

Val Val Ala Tyr Ser Ile Leu Val Phe Ala Asp Val Glu Lys Gly

1130 1135 1140

Lys Ser Lys Lys Leu Arg Lys Val Gln Asp Met Val Gly Ile Thr

1145 1150 1155

Ile Met Glu Lys Lys Arg Phe Glu Lys Asn Pro Val Asp Phe Leu

1160 1165 1170

Glu Gln Arg Gly Tyr Arg Asn Val Arg Leu Glu Lys Ile Ile Lys

1175 1180 1185

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Lys Arg Arg Arg

1190 1195 1200

Leu Leu Ala Ser Ala Lys Phe Leu Gln Lys Gly Asn Glu Leu Val

1205 1210 1215

Ile Pro Gln Arg Phe Thr Thr Leu Leu Tyr His Ser Tyr Arg Ile

1220 1225 1230

Glu Lys Asp Tyr Glu Pro Glu His Arg Glu Tyr Val Glu Lys His

1235 1240 1245

Lys Asp Glu Phe Lys Glu Leu Leu Glu Tyr Ile Ser Val Phe Ser

1250 1255 1260

Arg Lys Tyr Val Leu Ala Asp Asn Asn Leu Thr Lys Ile Glu Met
1265 1270 1275
Leu Phe Ser Lys Asn Lys Asp Ala Glu Val Ser Ser Leu Ala Lys
1280 1285 1290
Ser Phe Ile Ser Leu Leu Thr Phe Thr Ala Phe Gly Ala Pro Arg
1295 1300 1305

Ala Phe Asn Phe Phe Gly Glu Asn Ile Ala Arg Lys Glu Tyr Arg
1310 1315 1320
Ser Val Thr Glu Cys Leu Asn Ala Thr Leu Ile His Gln Ser Ile
1325 1330 1335
Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Lys Leu Gly Glu
1340 1345 1350
Asp Gly Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe
1355 1360 1365

Glu Ser Pro Lys Lys Lys Arg Lys Val
1370 1375

<210> 176
<211> 1377
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 176
Met Pro Lys Lys Lys Arg Lys Val Gly Met Gly Lys Pro Tyr Ser Ile
1 5 10 15
Gly Leu Asp Ile Gly Thr Asn Ser Val Gly Trp Ala Val Val Thr Asp
20 25 30
Asp Tyr Lys Val Pro Ala Lys Lys Met Lys Val Leu Gly Asn Thr Asp

35 40 45
Lys Gln Ser Ile Lys Lys Asn Leu Leu Gly Ala Leu Leu Phe Asp Ser
50 55 60
Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg

65 70 75 80

Tyr Thr Arg Arg Lys Asn Arg Leu Arg Tyr Leu Gln Glu Ile Phe Thr

85 90 95

Gly Glu Met Asn Lys Val Asp Glu Asn Phe Phe Gln Arg Leu Asp Asp

100 105 110

Ser Phe Leu Val Asp Glu Asp Lys Arg Gly Glu His His Pro Ile Phe

115 120 125

Gly Asn Ile Ala Ala Glu Val Lys Tyr His Asp Asp Phe Pro Thr Ile

130 135 140

Tyr His Leu Arg Arg His Leu Ala Asp Thr Ser Lys Lys Ala Asp Leu

145 150 155 160

Arg Leu Val Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His

165 170 175

Phe Leu Tyr Glu Gly Asp Leu Lys Ala Glu Asn Thr Asp Val Gln Ala

180 185 190

Leu Phe Lys Asp Phe Val Glu Glu Tyr Asp Lys Thr Ile Glu Glu Ser

195 200 205

His Leu Ser Glu Ile Thr Val Asp Ala Leu Ser Ile Leu Thr Glu Lys

210 215 220

Val Ser Lys Ser Ser Arg Leu Glu Asn Leu Ile Ala His Tyr Pro Thr

225 230 235 240

Glu Lys Lys Asn Thr Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Asp

245 250 255

Leu His Pro Asn Phe Lys Thr Asn Phe Gln Leu Ser Glu Asp Ala Lys

260 265 270

Leu Gln Phe Ser Lys Asp Thr Tyr Glu Glu Asp Leu Glu Gly Phe Leu

275 280 285

Gly Glu Val Gly Asp Glu Tyr Ala Asp Leu Phe Ala Ser Ala Lys Asn

290 295 300

Leu Tyr Asp Ala Ile Leu Leu Ser Gly Ile Leu Thr Val Asp Asp Asn

305 310 315 320

Ser Thr Lys Ala Pro Leu Ser Ala Ser Met Val Lys Arg Tyr Glu Glu
325 330 335
His Gln Lys Asp Leu Lys Lys Leu Lys Asp Phe Ile Lys Val Asn Ala
340 345 350
Pro Asp Gln Tyr Asn Ala Ile Phe Lys Asp Lys Asn Lys Lys Gly Tyr

355 360 365
Ala Ser Tyr Ile Glu Ser Gly Val Lys Gln Asp Glu Phe Tyr Lys Tyr
370 375 380
Leu Lys Gly Ile Leu Leu Lys Ile Asn Gly Ser Gly Asp Phe Leu Asp
385 390 395 400
Lys Ile Asp Arg Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn
405 410 415
Gly Ile Ile Pro His Gln Ile His Leu Gln Glu Met His Ala Ile Leu

420 425 430
Arg Arg Gln Gly Glu His Tyr Pro Phe Leu Lys Glu Asn Gln Asp Lys
435 440 445
Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu
450 455 460
Ala Arg Lys Gly Ser Arg Phe Ala Trp Ala Glu Tyr Lys Ala Asp Glu
465 470 475 480
Lys Ile Thr Pro Trp Asn Phe Asp Asp Ile Leu Asp Lys Glu Lys Ser

485 490 495
Ala Glu Lys Phe Ile Thr Arg Met Thr Leu Asn Asp Leu Tyr Leu Pro
500 505 510
Glu Glu Lys Val Leu Pro Lys His Ser Pro Leu Tyr Glu Ala Phe Thr
515 520 525
Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val Asn Glu Gln Gly Glu
530 535 540
Ala Lys Phe Phe Asp Thr Asn Met Lys Gln Glu Ile Phe Asp His Val

545 550 555 560
Phe Lys Glu Asn Arg Lys Val Thr Lys Asp Lys Leu Leu Asn Tyr Leu

565 570 575

Asn Lys Glu Phe Glu Glu Phe Arg Ile Val Asn Leu Thr Gly Leu Asp

580 585 590

Lys Glu Asn Lys Ala Phe Asn Ser Ser Leu Gly Thr Tyr His Asp Leu

595 600 605

Arg Lys Ile Leu Asp Lys Ser Phe Leu Asp Asp Lys Ala Asn Glu Lys

610 615 620

Thr Ile Glu Asp Ile Ile Gln Thr Leu Thr Leu Phe Glu Asp Arg Glu

625 630 635 640

Met Ile Arg Gln Arg Leu Gln Lys Tyr Ser Asp Ile Phe Thr Lys Ala

645 650 655

Gln Leu Lys Lys Leu Glu Arg Leu His Tyr Thr Gly Trp Gly Arg Leu

660 665 670

Ser Tyr Lys Leu Ile Asn Gly Ile Arg Asn Lys Glu Asn Lys Lys Thr

675 680 685

Ile Leu Asp Tyr Leu Ile Asp Asp Gly Tyr Ala Asn Arg Asn Phe Met

690 695 700

Gln Leu Ile Asn Asp Asp Ala Leu Ser Phe Lys Glu Glu Ile Ala Arg

705 710 715 720

Ala Gln Ile Ile Asp Asp Val Asp Asp Ile Ala Asn Val Val His Asp

725 730 735

Leu Pro Gly Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Val Lys

740 745 750

Ile Val Asp Glu Leu Val Lys Val Met Gly His Asn Pro Ala Asn Ile

755 760 765

Ile Ile Glu Met Ala Arg Glu Asn Gln Thr Thr Asp Lys Gly Arg Arg

770 775 780

Asn Ser Gln Gln Arg Leu Lys Leu Leu Gln Asp Ser Leu Lys Asn Leu

785 790 795 800

Asp Asn Pro Val Asn Ile Lys Asn Val Glu Asn Gln Gln Leu Gln Asn

805 810 815

Asp Arg Leu Phe Leu Tyr Tyr Ile Gln Asn Gly Lys Asp Met Tyr Thr
820 825 830
Gly Glu Thr Leu Asp Ile Asn Asn Leu Ser Gln Tyr Asp Ile Asp His
835 840 845
Ile Ile Pro Gln Ala Phe Ile Lys Asp Asn Ser Leu Asp Asn Arg Val
850 855 860
Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asp Val Pro Ser

865 870 875 880
Ile Glu Val Val His Glu Met Lys Ser Phe Trp Ser Lys Leu Leu Ser
885 890 895
Val Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu
900 905 910
Arg Gly Gly Leu Thr Glu Glu Asp Lys Ala Gly Phe Ile Lys Arg Gln
915 920 925
Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp

930 935 940
Glu Arg Phe Asn Thr Glu Phe Asp Gly Asn Lys Arg Arg Ile Arg Asn
945 950 955 960
Val Lys Ile Ile Thr Leu Lys Ser Asn Leu Val Ser Asn Phe Arg Lys
965 970 975
Glu Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Tyr His His Ala
980 985 990
His Asp Ala Tyr Leu Asn Ala Val Val Gly Asn Ala Leu Leu Leu Lys

995 1000 1005
Tyr Pro Gln Leu Glu Pro Glu Phe Val Tyr Gly Glu Tyr Pro Lys
1010 1015 1020
Tyr Asn Ser Tyr Arg Ser Arg Lys Ser Ala Thr Glu Lys Phe Leu
1025 1030 1035
Phe Tyr Ser Asn Ile Leu Arg Phe Phe Lys Lys Glu Asp Ile Gln
1040 1045 1050
Thr Asn Glu Asp Gly Glu Ile Ala Trp Asn Lys Glu Lys His Ile

1055 1060 1065

Lys Ile Leu Arg Lys Val Leu Ser Tyr Pro Gln Val Asn Ile Val

1070 1075 1080

Lys Lys Thr Glu Glu Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile

1085 1090 1095

Leu Pro Lys Gly Glu Ser Asp Lys Leu Ile Pro Arg Lys Thr Lys

1100 1105 1110

Asn Ser Tyr Trp Asp Pro Lys Lys Tyr Gly Gly Phe Met Gln Pro

1115 1120 1125

Val Val Ala Tyr Ser Ile Leu Val Phe Ala Asp Val Glu Lys Gly

1130 1135 1140

Lys Ser Lys Lys Leu Arg Lys Val Gln Asp Met Val Gly Ile Thr

1145 1150 1155

Ile Met Glu Lys Lys Arg Phe Glu Lys Asn Pro Val Asp Phe Leu

1160 1165 1170

Glu Gln Arg Gly Tyr Arg Asn Val Arg Leu Glu Lys Ile Ile Lys

1175 1180 1185

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Lys Arg Arg Arg

1190 1195 1200

Leu Leu Ala Ser Ala Lys Phe Leu Gln Lys Gly Asn Glu Leu Val

1205 1210 1215

Ile Pro Gln Arg Phe Thr Thr Leu Leu Tyr His Ser Tyr Arg Ile

1220 1225 1230

Glu Lys Asp Tyr Glu Pro Glu His Arg Glu Tyr Val Glu Lys His

1235 1240 1245

Lys Asp Glu Phe Lys Glu Leu Leu Glu Tyr Ile Ser Val Phe Ser

1250 1255 1260

Arg Lys Tyr Val Leu Ala Asp Asn Asn Leu Thr Lys Ile Glu Met

1265 1270 1275

Leu Phe Ser Lys Asn Lys Asp Ala Glu Val Ser Ser Leu Ala Lys

1280 1285 1290

Ser Phe Ile Ser Leu Leu Thr Phe Thr Ala Phe Gly Ala Pro Arg

1295 1300 1305

Ala Phe Asn Phe Phe Gly Glu Asn Ile Ala Arg Lys Glu Tyr Arg

1310 1315 1320

Ser Val Thr Glu Cys Leu Asn Ala Thr Leu Ile His Gln Ser Ile

1325 1330 1335

Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Lys Leu Gly Glu

1340 1345 1350

Asp Gly Glu Gly Ala Asp Lys Arg Thr Ala Asp Gly Ser Glu Phe

1355 1360 1365

Glu Ser Pro Lys Lys Lys Arg Lys Val

1370 1375

<210> 177

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 177

atcgacaaga aagggactga 20

<210> 178

<211> 1566

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polypeptide

<400> 178

Met Ser Glu Val Glu Phe Ser His Glu Tyr Trp Met Arg His Ala Leu

1 5 10 15

Thr Leu Ala Lys Arg Ala Arg Asp Glu Arg Glu Val Pro Val Gly Ala

20 25 30

Val Leu Val Leu Asn Asn Arg Val Ile Gly Glu Gly Trp Asn Arg Ala

35 40 45

Ile Gly Leu His Asp Pro Thr Ala His Ala Glu Ile Met Ala Leu Arg

50 55 60

Gln Gly Gly Leu Val Met Gln Asn Tyr Arg Leu Tyr Asp Ala Thr Leu

65 70 75 80

Tyr Ser Thr Phe Glu Pro Cys Val Met Cys Ala Gly Ala Met Ile His

85 90 95

Ser Arg Ile Gly Arg Val Val Phe Gly Val Arg Asn Ala Lys Thr Gly

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

180 185 190

Gly Gly Ser Ser Gly Gly Ser Gly Lys Pro Tyr Ser Ile Gly Leu Ala

195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Val Thr Asp Asp Tyr Lys

210 215 220

Val Pro Ala Lys Lys Met Lys Val Leu Gly Asn Thr Asp Lys Gln Ser

225 230 235 240

Ile Lys Lys Asn Leu Leu Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

245 250 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

260 265 270

Arg Lys Asn Arg Leu Arg Tyr Leu Gln Glu Ile Phe Thr Gly Glu Met

275 280 285

Asn Lys Val Asp Glu Asn Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu
290 295 300
Val Asp Glu Asp Lys Arg Gly Glu His His Pro Ile Phe Gly Asn Ile
305 310 315 320

Ala Ala Glu Val Lys Tyr His Asp Asp Phe Pro Thr Ile Tyr His Leu
325 330 335
Arg Arg His Leu Ala Asp Thr Ser Lys Lys Ala Asp Leu Arg Leu Val
340 345 350
Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Tyr
355 360 365
Glu Gly Asp Leu Lys Ala Glu Asn Thr Asp Val Gln Ala Leu Phe Lys
370 375 380

Asp Phe Val Glu Glu Tyr Asp Lys Thr Ile Glu Glu Ser His Leu Ser
385 390 395 400
Glu Ile Thr Val Asp Ala Leu Ser Ile Leu Thr Glu Lys Val Ser Lys
405 410 415
Ser Ser Arg Leu Glu Asn Leu Ile Ala His Tyr Pro Thr Glu Lys Lys
420 425 430
Asn Thr Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Asp Leu His Pro
435 440 445

Asn Phe Lys Thr Asn Phe Gln Leu Ser Glu Asp Ala Lys Leu Gln Phe
450 455 460
Ser Lys Asp Thr Tyr Glu Glu Asp Leu Glu Gly Phe Leu Gly Glu Val
465 470 475 480
Gly Asp Glu Tyr Ala Asp Leu Phe Ala Ser Ala Lys Asn Leu Tyr Asp
485 490 495
Ala Ile Leu Leu Ser Gly Ile Leu Thr Val Asp Asp Asn Ser Thr Lys
500 505 510

Ala Pro Leu Ser Ala Ser Met Val Lys Arg Tyr Glu Glu His Gln Lys
515 520 525
Asp Leu Lys Lys Leu Lys Asp Phe Ile Lys Val Asn Ala Pro Asp Gln

530 535 540

Tyr Asn Ala Ile Phe Lys Asp Lys Asn Lys Lys Gly Tyr Ala Ser Tyr

545 550 555 560

Ile Glu Ser Gly Val Lys Gln Asp Glu Phe Tyr Lys Tyr Leu Lys Gly

565 570 575

Ile Leu Leu Lys Ile Asn Gly Ser Gly Asp Phe Leu Asp Lys Ile Asp

580 585 590

Arg Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile

595 600 605

Pro His Gln Ile His Leu Gln Glu Met His Ala Ile Leu Arg Arg Gln

610 615 620

Gly Glu His Tyr Pro Phe Leu Lys Glu Asn Gln Asp Lys Ile Glu Lys

625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Lys

645 650 655

Gly Ser Arg Phe Ala Trp Ala Glu Tyr Lys Ala Asp Glu Lys Ile Thr

660 665 670

Pro Trp Asn Phe Asp Asp Ile Leu Asp Lys Glu Lys Ser Ala Glu Lys

675 680 685

Phe Ile Thr Arg Met Thr Leu Asn Asp Leu Tyr Leu Pro Glu Glu Lys

690 695 700

Val Leu Pro Lys His Ser Pro Leu Tyr Glu Ala Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Asn Glu Gln Gly Glu Ala Lys Phe

725 730 735

Phe Asp Thr Asn Met Lys Gln Glu Ile Phe Asp His Val Phe Lys Glu

740 745 750

Asn Arg Lys Val Thr Lys Asp Lys Leu Leu Asn Tyr Leu Asn Lys Glu

755 760 765

Phe Glu Glu Phe Arg Ile Val Asn Leu Thr Gly Leu Asp Lys Glu Asn

770 775 780

Lys Ala Phe Asn Ser Ser Leu Gly Thr Tyr His Asp Leu Arg Lys Ile
785 790 795 800
Leu Asp Lys Ser Phe Leu Asp Asp Lys Ala Asn Glu Lys Thr Ile Glu
805 810 815
Asp Ile Ile Gln Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Arg
820 825 830

Gln Arg Leu Gln Lys Tyr Ser Asp Ile Phe Thr Lys Ala Gln Leu Lys
835 840 845
Lys Leu Glu Arg Arg His Tyr Thr Gly Trp Gly Arg Leu Ser Tyr Lys
850 855 860
Leu Ile Asn Gly Ile Arg Asn Lys Glu Asn Lys Lys Thr Ile Leu Asp
865 870 875 880
Tyr Leu Ile Asp Asp Gly Tyr Ala Asn Arg Asn Phe Met Gln Leu Ile
885 890 895

Asn Asp Asp Ala Leu Ser Phe Lys Glu Glu Ile Ala Arg Ala Gln Ile
900 905 910
Ile Asp Asp Val Asp Asp Ile Ala Asn Val Val His Asp Leu Pro Gly
915 920 925
Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Val Lys Ile Val Asp
930 935 940
Glu Leu Val Lys Val Met Gly His Asn Pro Ala Asn Ile Ile Ile Glu
945 950 955 960

Met Ala Arg Glu Asn Gln Thr Thr Asp Lys Gly Arg Arg Asn Ser Gln
965 970 975
Gln Arg Leu Lys Leu Leu Gln Asp Ser Leu Lys Asn Leu Asp Asn Pro
980 985 990
Val Asn Ile Lys Asn Val Glu Asn Gln Gln Leu Gln Asn Asp Arg Leu
995 1000 1005
Phe Leu Tyr Tyr Ile Gln Asn Gly Lys Asp Met Tyr Thr Gly Glu
1010 1015 1020

Thr Leu Asp Ile Asn Asn Leu Ser Gln Tyr Asp Ile Asp His Ile

1025 1030 1035

Ile Pro Gln Ala Phe Ile Lys Asp Asn Ser Leu Asp Asn Arg Val

1040 1045 1050

Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asp Val Pro

1055 1060 1065

Ser Ile Glu Val Val His Glu Met Lys Ser Phe Trp Ser Lys Leu

1070 1075 1080

Leu Ser Val Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr

1085 1090 1095

Lys Ala Glu Arg Gly Gly Leu Thr Glu Glu Asp Lys Ala Gly Phe

1100 1105 1110

Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val

1115 1120 1125

Ala Gln Ile Leu Asp Glu Arg Phe Asn Thr Glu Phe Asp Gly Asn

1130 1135 1140

Lys Arg Arg Ile Arg Asn Val Lys Ile Ile Thr Leu Lys Ser Asn

1145 1150 1155

Leu Val Ser Asn Phe Arg Lys Glu Phe Glu Leu Tyr Lys Val Arg

1160 1165 1170

Glu Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn Ala

1175 1180 1185

Val Val Gly Asn Ala Leu Leu Leu Lys Tyr Pro Gln Leu Glu Pro

1190 1195 1200

Glu Phe Val Tyr Gly Glu Tyr Pro Lys Tyr Asn Ser Tyr Arg Ser

1205 1210 1215

Arg Lys Ser Ala Thr Glu Lys Phe Leu Phe Tyr Ser Asn Ile Leu

1220 1225 1230

Arg Phe Phe Lys Lys Glu Asp Ile Gln Thr Asn Glu Asp Gly Glu

1235 1240 1245

Ile Ala Trp Asn Lys Glu Lys His Ile Lys Ile Leu Arg Lys Val

1250 1255 1260

Leu Ser Tyr Pro Gln Val Asn Ile Val Lys Lys Thr Glu Glu Gln
1265 1270 1275
Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Gly Glu Ser
1280 1285 1290
Asp Lys Leu Ile Pro Arg Lys Thr Lys Asn Ser Tyr Trp Asp Pro
1295 1300 1305
Lys Lys Tyr Gly Gly Phe Met Gln Pro Val Val Ala Tyr Ser Ile
1310 1315 1320

Leu Val Phe Ala Asp Val Glu Lys Gly Lys Ser Lys Lys Leu Arg
1325 1330 1335
Lys Val Gln Asp Met Val Gly Ile Thr Ile Met Glu Lys Lys Arg
1340 1345 1350
Phe Glu Lys Asn Pro Val Asp Phe Leu Glu Gln Arg Gly Tyr Arg
1355 1360 1365
Asn Val Arg Leu Glu Lys Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1370 1375 1380

Phe Glu Leu Glu Asn Lys Arg Arg Arg Leu Leu Ala Ser Ala Lys
1385 1390 1395
Phe Leu Gln Lys Gly Asn Glu Leu Val Ile Pro Gln Arg Phe Thr
1400 1405 1410
Thr Leu Leu Tyr His Ser Tyr Arg Ile Glu Lys Asp Tyr Glu Pro
1415 1420 1425
Glu His Arg Glu Tyr Val Glu Lys His Lys Asp Glu Phe Lys Glu
1430 1435 1440

Leu Leu Glu Tyr Ile Ser Val Phe Ser Arg Lys Tyr Val Leu Ala
1445 1450 1455
Asp Asn Asn Leu Thr Lys Ile Glu Met Leu Phe Ser Lys Asn Lys
1460 1465 1470
Asp Ala Glu Val Ser Ser Leu Ala Lys Ser Phe Ile Ser Leu Leu
1475 1480 1485
Thr Phe Thr Ala Phe Gly Ala Pro Arg Ala Phe Asn Phe Phe Gly

1490 1495 1500

Glu Asn Ile Ala Arg Lys Glu Tyr Arg Ser Val Thr Glu Cys Leu
1505 1510 1515
Asn Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr
1520 1525 1530
Arg Ile Asp Leu Ser Lys Leu Gly Glu Asp Gly Glu Gly Ala Asp
1535 1540 1545
Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys
1550 1555 1560

Arg Lys Val
1565

<210> 179
<211> 1566
<212> PRT
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic polypeptide
<400> 179

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

100 105 110

Ala Ala Gly Ser Leu Met Asp Val Leu His His Pro Gly Met Asn His

115 120 125

Arg Val Glu Ile Thr Glu Gly Ile Leu Ala Asp Glu Cys Ala Ala Leu

130 135 140

Leu Cys Arg Phe Phe Arg Met Pro Arg Arg Val Phe Asn Ala Gln Lys

145 150 155 160

Lys Ala Gln Ser Ser Thr Asp Ser Gly Gly Ser Ser Gly Gly Ser Ser

165 170 175

Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Ser

180 185 190

Gly Gly Ser Ser Gly Gly Ser Gly Lys Pro Tyr Ser Ile Gly Leu Ala

195 200 205

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Val Thr Asp Asp Tyr Lys

210 215 220

Val Pro Ala Lys Lys Met Lys Val Leu Gly Asn Thr Asp Lys Gln Ser

225 230 235 240

Ile Lys Lys Asn Leu Leu Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

245 250 255

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

260 265 270

Arg Lys Asn Arg Leu Arg Tyr Leu Gln Glu Ile Phe Thr Gly Glu Met

275 280 285

Asn Lys Val Asp Glu Asn Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu

290 295 300

Val Asp Glu Asp Lys Arg Gly Glu His His Pro Ile Phe Gly Asn Ile

305 310 315 320

Ala Ala Glu Val Lys Tyr His Asp Asp Phe Pro Thr Ile Tyr His Leu

325 330 335

Arg Arg His Leu Ala Asp Thr Ser Lys Lys Ala Asp Leu Arg Leu Val

340 345 350

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Tyr

355 360 365

Glu Gly Asp Leu Lys Ala Glu Asn Thr Asp Val Gln Ala Leu Phe Lys

370 375 380

Asp Phe Val Glu Glu Tyr Asp Lys Thr Ile Glu Glu Ser His Leu Ser

385 390 395 400

Glu Ile Thr Val Asp Ala Leu Ser Ile Leu Thr Glu Lys Val Ser Lys

405 410 415

Ser Ser Arg Leu Glu Asn Leu Ile Ala His Tyr Pro Thr Glu Lys Lys

420 425 430

Asn Thr Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Asp Leu His Pro

435 440 445

Asn Phe Lys Thr Asn Phe Gln Leu Ser Glu Asp Ala Lys Leu Gln Phe

450 455 460

Ser Lys Asp Thr Tyr Glu Glu Asp Leu Glu Gly Phe Leu Gly Glu Val

465 470 475 480

Gly Asp Glu Tyr Ala Asp Leu Phe Ala Ser Ala Lys Asn Leu Tyr Asp

485 490 495

Ala Ile Leu Leu Ser Gly Ile Leu Thr Val Asp Asp Asn Ser Thr Lys

500 505 510

Ala Pro Leu Ser Ala Ser Met Val Lys Arg Tyr Glu Glu His Gln Lys

515 520 525

Asp Leu Lys Lys Leu Lys Asp Phe Ile Lys Val Asn Ala Pro Asp Gln

530 535 540

Tyr Asn Ala Ile Phe Lys Asp Lys Asn Lys Lys Gly Tyr Ala Ser Tyr

545 550 555 560

Ile Glu Ser Gly Val Lys Gln Asp Glu Phe Tyr Lys Tyr Leu Lys Gly

565 570 575

Ile Leu Leu Lys Ile Asn Gly Ser Gly Asp Phe Leu Asp Lys Ile Asp

580 585 590

Arg Glu Asp Phe Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ile Ile

595 600 605

Pro His Gln Ile His Leu Gln Glu Met His Ala Ile Leu Arg Arg Gln

610 615 620

Gly Glu His Tyr Pro Phe Leu Lys Glu Asn Gln Asp Lys Ile Glu Lys

625 630 635 640

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Lys

645 650 655

Gly Ser Arg Phe Ala Trp Ala Glu Tyr Lys Ala Asp Glu Lys Ile Thr

660 665 670

Pro Trp Asn Phe Asp Asp Ile Leu Asp Lys Glu Lys Ser Ala Glu Lys

675 680 685

Phe Ile Thr Arg Met Thr Leu Asn Asp Leu Tyr Leu Pro Glu Glu Lys

690 695 700

Val Leu Pro Lys His Ser Pro Leu Tyr Glu Ala Phe Thr Val Tyr Asn

705 710 715 720

Glu Leu Thr Lys Val Lys Tyr Val Asn Glu Gln Gly Glu Ala Lys Phe

725 730 735

Phe Asp Thr Asn Met Lys Gln Glu Ile Phe Asp His Val Phe Lys Glu

740 745 750

Asn Arg Lys Val Thr Lys Asp Lys Leu Leu Asn Tyr Leu Asn Lys Glu

755 760 765

Phe Glu Glu Phe Arg Ile Val Asn Leu Thr Gly Leu Asp Lys Glu Asn

770 775 780

Lys Ala Phe Asn Ser Ser Leu Gly Thr Tyr His Asp Leu Arg Lys Ile

785 790 795 800

Leu Asp Lys Ser Phe Leu Asp Asp Lys Ala Asn Glu Lys Thr Ile Glu

805 810 815

Asp Ile Ile Gln Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Arg

820 825 830

Gln Arg Leu Gln Lys Tyr Ser Asp Ile Phe Thr Lys Ala Gln Leu Lys

835 840 845

Lys Leu Glu Arg Leu His Tyr Thr Gly Trp Gly Arg Leu Ser Tyr Lys
850 855 860
Leu Ile Asn Gly Ile Arg Asn Lys Glu Asn Lys Lys Thr Ile Leu Asp

865 870 875 880
Tyr Leu Ile Asp Asp Gly Tyr Ala Asn Arg Asn Phe Met Gln Leu Ile
885 890 895
Asn Asp Asp Ala Leu Ser Phe Lys Glu Glu Ile Ala Arg Ala Gln Ile
900 905 910
Ile Asp Asp Val Asp Asp Ile Ala Asn Val Val His Asp Leu Pro Gly
915 920 925
Ser Pro Ala Ile Lys Lys Gly Ile Leu Gln Ser Val Lys Ile Val Asp

930 935 940
Glu Leu Val Lys Val Met Gly His Asn Pro Ala Asn Ile Ile Ile Glu
945 950 955 960
Met Ala Arg Glu Asn Gln Thr Thr Asp Lys Gly Arg Arg Asn Ser Gln
965 970 975
Gln Arg Leu Lys Leu Leu Gln Asp Ser Leu Lys Asn Leu Asp Asn Pro
980 985 990
Val Asn Ile Lys Asn Val Glu Asn Gln Gln Leu Gln Asn Asp Arg Leu

995 1000 1005
Phe Leu Tyr Tyr Ile Gln Asn Gly Lys Asp Met Tyr Thr Gly Glu
1010 1015 1020
Thr Leu Asp Ile Asn Asn Leu Ser Gln Tyr Asp Ile Asp His Ile
1025 1030 1035
Ile Pro Gln Ala Phe Ile Lys Asp Asn Ser Leu Asp Asn Arg Val
1040 1045 1050
Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asp Val Pro

1055 1060 1065
Ser Ile Glu Val Val His Glu Met Lys Ser Phe Trp Ser Lys Leu
1070 1075 1080
Leu Ser Val Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr

1085 1090 1095

Lys Ala Glu Arg Gly Gly Leu Thr Glu Glu Asp Lys Ala Gly Phe

1100 1105 1110

Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val

1115 1120 1125

Ala Gln Ile Leu Asp Glu Arg Phe Asn Thr Glu Phe Asp Gly Asn

1130 1135 1140

Lys Arg Arg Ile Arg Asn Val Lys Ile Ile Thr Leu Lys Ser Asn

1145 1150 1155

Leu Val Ser Asn Phe Arg Lys Glu Phe Glu Leu Tyr Lys Val Arg

1160 1165 1170

Glu Ile Asn Asp Tyr His His Ala His Asp Ala Tyr Leu Asn Ala

1175 1180 1185

Val Val Gly Asn Ala Leu Leu Leu Lys Tyr Pro Gln Leu Glu Pro

1190 1195 1200

Glu Phe Val Tyr Gly Glu Tyr Pro Lys Tyr Asn Ser Tyr Arg Ser

1205 1210 1215

Arg Lys Ser Ala Thr Glu Lys Phe Leu Phe Tyr Ser Asn Ile Leu

1220 1225 1230

Arg Phe Phe Lys Lys Glu Asp Ile Gln Thr Asn Glu Asp Gly Glu

1235 1240 1245

Ile Ala Trp Asn Lys Glu Lys His Ile Lys Ile Leu Arg Lys Val

1250 1255 1260

Leu Ser Tyr Pro Gln Val Asn Ile Val Lys Lys Thr Glu Glu Gln

1265 1270 1275

Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Gly Glu Ser

1280 1285 1290

Asp Lys Leu Ile Pro Arg Lys Thr Lys Asn Ser Tyr Trp Asp Pro

1295 1300 1305

Lys Lys Tyr Gly Gly Phe Met Gln Pro Val Val Ala Tyr Ser Ile

1310 1315 1320

Leu Val Phe Ala Asp Val Glu Lys Gly Lys Ser Lys Lys Leu Arg

1325 1330 1335

Lys Val Gln Asp Met Val Gly Ile Thr Ile Met Glu Lys Lys Arg

1340 1345 1350

Phe Glu Lys Asn Pro Val Asp Phe Leu Glu Gln Arg Gly Tyr Arg

1355 1360 1365

Asn Val Arg Leu Glu Lys Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1370 1375 1380

Phe Glu Leu Glu Asn Lys Arg Arg Arg Leu Leu Ala Ser Ala Lys

1385 1390 1395

Phe Leu Gln Lys Gly Asn Glu Leu Val Ile Pro Gln Arg Phe Thr

1400 1405 1410

Thr Leu Leu Tyr His Ser Tyr Arg Ile Glu Lys Asp Tyr Glu Pro

1415 1420 1425

Glu His Arg Glu Tyr Val Glu Lys His Lys Asp Glu Phe Lys Glu

1430 1435 1440

Leu Leu Glu Tyr Ile Ser Val Phe Ser Arg Lys Tyr Val Leu Ala

1445 1450 1455

Asp Asn Asn Leu Thr Lys Ile Glu Met Leu Phe Ser Lys Asn Lys

1460 1465 1470

Asp Ala Glu Val Ser Ser Leu Ala Lys Ser Phe Ile Ser Leu Leu

1475 1480 1485

Thr Phe Thr Ala Phe Gly Ala Pro Arg Ala Phe Asn Phe Phe Gly

1490 1495 1500

Glu Asn Ile Ala Arg Lys Glu Tyr Arg Ser Val Thr Glu Cys Leu

1505 1510 1515

Asn Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr

1520 1525 1530

Arg Ile Asp Leu Ser Lys Leu Gly Glu Asp Gly Glu Gly Ala Asp

1535 1540 1545

Lys Arg Thr Ala Asp Gly Ser Glu Phe Glu Ser Pro Lys Lys Lys

1550 1555 1560

Arg Lys Val

1565

<210> 180

<211> 98

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 180

guuuuagagc ugugcuguuu aaacaacaca gcaaguuaaa auaaggcuuu guccguacuc 60

aagcuugcaa aagcgugcac cgauucggug cuuuuuuu 98

<210> 181

<211> 107

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<220><223> Description of Combined DNA/RNA Molecule: Synthetic

oligonucleotide

<400> 181

guuugagagu agugugaaaa cauuacgagu ucaaauacaa auuaauuuac aaugccuucg 60

ggcugcccga cguagggcac cuacucucaa uucuucggaa uugagtt 107

<210> 182

<211> 128

<212> RNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 182

gucuuggaug agugugaaaa cacucauagu caagaucaaa cgagugguuu uccacgaguu 60

auuacuuuug aggucuuaua uggcccauac auaaaaagga gucggaauuu ccggcuccuu 120

uucuuuuu 128