Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3231712 A1 2023/03/16

(21) 3 231 712

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2022/09/07
(87) **Date publication PCT/PCT Publication Date:** 2023/03/16
(85) **Entrée phase nationale/National Entry:** 2024/03/07
(86) **N° demande PCT/PCT Application No.:** US 2022/076058
(87) **N° publication PCT/PCT Publication No.:** 2023/039435
(30) **Priorités/Priorities:** 2021/09/08 (US63/241,897); 2022/01/27 (US63/303,927); 2022/06/24 (US63/367,025)

(51) **Cl.Int./Int.Cl.** ***C12N 15/86*** (2006.01), ***A61K 48/00*** (2006.01)

(71) **Demandeur/Applicant:**
FLAGSHIP PIONEERING INNOVATIONS VI, LLC, US

(72) **Inventeurs/Inventors:**
ALTSHULER, ROBERT CHARLES, US;
BOTHMER, ANNE HELEN, US;
CHEE, DANIEL RAYMOND, US;
COTTA-RAMUSINO, CECILIA GIOVANNA SILVIA, US;
KIM, KYUSIK, US;
KOTLAR, RANDI MICHELLE, US;
MCALLISTER, GREGORY DAVID, US;
...

(74) **Agent:** SMART & BIGGAR LP

(54) **Titre : COMPOSITIONS ET PROCEDES DE MODULATION DE PAH**
(54) **Title: PAH-MODULATING COMPOSITIONS AND METHODS**

(57) **Abrégé/Abstract:**
The disclosure provides, e.g., compositions, systems, and methods for targeting, editing, modifying, or manipulating a host cell's genome at one or more locations in a DNA sequence in a cell, tissue, or subject. Gene modifying systems for treating phenylketonuria (PKU) are described.

(21) 3 231 712

(13) A1

(72) **Inventeurs(suite)/Inventors(continued):** RAY, ANANYA, US; ROQUET, NATHANIEL, US; SANCHEZ, CARLOS, US; STEINBERG, BARRETT ETHAN, US; SALOMON, WILLIAM EDWARD, US; CITORIK, ROBERT JAMES, US; QUERBES, WILLIAM, US; APPONI, LUCIANO HENRIQUE, US; WANG, ZHAN, US

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 March 2023 (16.03.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/039435 A3

(51) International Patent Classification:
C12N 15/113 (2010.01)

(21) International Application Number:
PCT/US2022/076058

(22) International Filing Date:
07 September 2022 (07.09.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

63/241,897	08 September 2021 (08.09.2021)	US
63/303,927	27 January 2022 (27.01.2022)	US
63/367,025	24 June 2022 (24.06.2022)	US

(71) Applicant: **FLAGSHIP PIONEERING INNOVATIONS VI, LLC** [US/US]; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US).

(72) Inventors: **ALTSHULER, Robert Charles**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **BOTHMER, Anne Helen**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **CHEE, Daniel Raymond**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **COTTA-RAMUSINO, Cecilia Giovanna Silvia**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **KIM, Kyusik**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **KOTLAR, Randi Michelle**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **MCALLISTER, Gregory David**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **RAY, Ananya**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **ROQUET, Nathaniel**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **SANCHEZ, Carlos**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **STEINBERG, Barrett Ethan**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **SALOMON, William Edward**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **CITORIK, Robert James**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **QUERBES, William**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **APPONI, Luciano Henrique**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US). **WANG, Zhan**; 55 Cambridge Parkway, Suite 800E, Cambridge, Massachusetts 02142 (US).

(74) Agent: **REPASS, Maria A. DeCicco** et al.; K&L Gates LLP, One Lincoln Street, Boston, Massachusetts 02111 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:
27 July 2023 (27.07.2023)

WO 2023/039435 A3

(54) Title: PAH-MODULATING COMPOSITIONS AND METHODS

(57) Abstract: The disclosure provides, e.g., compositions, systems, and methods for targeting, editing, modifying, or manipulating a host cell's genome at one or more locations in a DNA sequence in a cell, tissue, or subject. Gene modifying systems for treating phenylketonuria (PKU) are described.

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 9
CONTENANT LES PAGES 1 À 217

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 9
CONTAINING PAGES 1 TO 217

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

PAH-MODULATING COMPOSITIONS AND METHODS

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in XML format compliant with WIPO Standard ST.26 and is hereby incorporated by reference in its entirety. Said XLM copy, created on September 8, 2022, is named V2065-7025WO_SL.xml and is 15,726,993 kb in size.

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 63/241,897, filed September 8, 2021, U.S. Provisional Application No. 63/303,927, filed January 27, 2022, and U.S. Provisional Application No. 63/367,025, filed June 24, 2022. The contents of the aforementioned applications are hereby incorporated by reference in their entirety.

BACKGROUND

Integration of a nucleic acid of interest into a genome occurs at low frequency and with little site specificity, in the absence of a specialized protein to promote the insertion event. Some existing approaches, like CRISPR/Cas9, are more suited for small edits that rely on host repair pathways, and are less effective at integrating longer sequences. Other existing approaches, like Cre/loxP, require a first step of inserting a loxP site into the genome and then a second step of inserting a sequence of interest into the loxP site. There is a need in the art for improved compositions (e.g., proteins and nucleic acids) and methods for inserting, altering, or deleting sequences of interest in a genome.

PKU is an inherited disorder involving an autosomal recessive inborn error of metabolism caused by a deficiency in the hepatic enzyme PAH. PAH catalyzes the hydroxylation of phenylalanine to tyrosine, the rate-limiting step in phenylalanine metabolism. The reaction is dependent on tetrahydrobiopterin (BH4), as a cofactor, molecular oxygen, and iron. Loss-of-function mutations in one, or both, copies of the PAH gene lead to a non-functional, or less efficient enzyme. This ultimately results in phenotypically severe forms of PKU where phenylalanine in the blood can accumulate to toxic concentrations, with impaired levels of plasma tyrosine.

Additionally, the deficiency prevents normal synthesis of downstream products, including dopamine, norepinephrine, and melanin.

The *PAH* genomic sequence and its flanking regions span about 171 kb, containing 13 exons. Study of pathogenic allelic variants have identified more than 500 different disease-causing mutations in the *PAH* gene (Mitchell, *et al. Genet Med.* 2011; 13:697-707). Of these mutations, approximately 62% have been characterized as missense, 13% deletions, 11% splice, 6% silent, 5% nonsense, 2% insertion, and < 1% deletion or duplication of exons. The identification of several PAH mutations have been described for their effects on enzymatic activity using enzyme kinetics and crystallographic studies. Mutations affecting the catalytic binding mode, including Y138F, S23A, and Y377F, were observed with reduced propensity for tetramer formation (Flydal, *et al. PNAS.* 2019; 116(23):11229-34). Other residues that interact with BH_4 in the precatalytic conformation (amino acids 245–255, 286, 322, and 325) also interact with BH_4 in the catalytic conformation, and, in addition, these sites are actually associated with severe destabilization of PAH.

Naturally occurring N-terminal PAH mutations have been determined to be distributed in a nonrandom pattern, clustering within residues 46-48 (GAL motif) and 65-69 (IESRP motif), both motifs highly conserved in pyruvate dehydrogenase (PDH) (Gjetting, *et al. Am./ J. Hum. Genet.* 2001; 68:1353-60). Structure-function studies demonstrated that mutations in these regions drastically reduced phenylalanine binding. Most missense mutations identified in PKU to date result in phenotypic outcomes associated with misfolding of the PAH enzyme, increased protein turnover, and loss of enzymatic function. Residues in exons 7-9 and in interdomain regions within the subunit appear to play an important structural role and constitute hotspots for destabilization. Additionally, using recombinant forms of hPAH, mutations in BH_4 responsive domains, including R408W and Y414C showed residual activity, but had perturbed allostery suggesting altered protein conformation (Gersting, *et al. Hum. Genet.* 2008; 83:5-17). Mutation analyses and structure-function analyses have identified a robust genotype-phenotype mapping for PAH's role in PKU; however, outside of lifetime symptom management strategies, there has not been a successful cure.

Dietary therapy of phenylalanine (Phe) remains to be the mainstay treatment for PKU since its introduction in 1953. In the 1970s, tetrahydrobiopterin (BH_4) and neurotransmitter precursor (L-dopa/carbidopa and 5-hydroxytryptophan) combination therapy showed promise in modulating PKU. Since its institution as a therapy, synthetics such as sapropterin have been formulated for as

small molecule isomers of BH_4. Although, this form of therapy is generally only useful in patients with mild subsets of PAH-deficient PKU. It is thought that the therapy responsiveness is associated with mutations in the PAH gene resulting in some residual enzyme activity. At high blood concentrations, Phe in the blood will compete with other large neutral amino acids (LNAAs) for transport across the blood-brain barrier. LNAA supplementation has been shown to reduce cerebral Phe concentrations despite the observed increase in plasma Phe levels. Likewise, dietary supplementation with glycomacropeptides (GMP) has been observed to significantly reduce ureagenesis, improved protein retention, and Phe utilization. Although, these strategies do little to address the increased blood levels of Phe or the genotypic drivers.

Modern non-dietary approaches include the development of PAH-based fusion proteins and enzyme substitution therapies. Enzyme substitution therapies can include administration of phenylalanine ammonia-lyase (PAL) to a patient. PAL is an enzyme which catalyzes the conversion of Phe to transcinnamic acid and insignificant amounts of ammonia. Early studies using PAL administered in enteric-coated gelatin capsules to PKU patients, showed reductions in Phe levels; however, repeated dosing *in vivo* resulted in mounting of immune responses. Although, these approaches are not practical from a clinical perspective as several intravenous injections would be required due to the limited half-life of circulating enzymes. Gene therapy has shown some promise, for example using viral vectors, in rescuing PAH functionality. However, the efficacy of this strategy is hampered by the very low gene transfer rate and transient transgene expression. Accordingly, there is a need for new and more effective treatments for targeting *PAH* in PKU.

SUMMARY OF THE INVENTION

This disclosure relates to novel compositions, systems, and methods for altering a genome at one or more locations in a host cell, tissue, or subject, *in vivo* or *in vitro*. The disclosure provides gene modifying systems that are capable of modulating (e.g., inserting, altering, or deleting sequences of interest) phenylalanine hydroxylase (PAH) activity and methods of treating phenylketonuria (PKU) by administering one or more such systems to alter a genomic sequence, such as to correct mutations, within the *PAH* gene on the human chromosome 12q23.2 involved as a genetic driver in PKU.

In one aspect, the disclosure relates to a system for modifying DNA to correct a human PAH gene mutation causing PKU comprising (a) a nucleic acid encoding a gene modifying

polypeptide capable of target primed reverse transcription, the polypeptide comprising (i) a reverse transcriptase domain and (ii) a Cas9 nickase that binds DNA and has endonuclease activity, and (b) a template RNA comprising (i) a gRNA spacer that is complementary to a first portion of the human PAH gene, (ii) a gRNA scaffold that binds the polypeptide, (iii) a heterologous object sequence comprising a mutation region to correct the mutation, and (iv) a primer binding site (PBS) sequence comprising at least 3, 4, 5, 6, 7, or 8 bases of 100% homology to a target DNA strand at the 3′ end of the template RNA. In some embodiments, the PAH gene may comprise a R408W mutation. In some embodiments, the PAH gene may comprise a R261Q mutation. In some embodiments, the PAH gene may comprise a R243Q mutation. In some embodiments, the PAH gene may comprise a IVS10-11G>A mutation. The template RNA sequence may comprise a sequence described herein, e.g., in Table 1A, 1B, 1C, 1D, 3A, 3B, 3C, 3D, 4A, 4B, 4C, 4D, 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

The gRNA spacer may comprise at least 15 bases of 100% homology to the target DNA at the 5′ end of the template RNA. The template RNA may further comprise a PBS sequence comprising at least 5 bases of at least 80% homology to the target DNA strand. The template RNA may comprise one or more chemical modifications.

The domains of the gene modifying polypeptide may be joined by a peptide linker. The polypeptide may comprise one or more peptide linkers. The gene modifying polypeptide may further comprise a nuclear localization signal. The polypeptide may comprise more than one nuclear localization signal, e.g., multiple adjacent nuclear localization signals or one or more nuclear localization signals in different regions of the polypeptide, e.g., one or more nuclear localization signals in the N-terminus of the polypeptide and one or more nuclear localization signals in the C-terminus of the polypeptide. The nucleic acid encoding the gene modifying polypeptide may encode one or more intein domains.

Introduction of the system into a target cell may result in insertion of at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, or 1000 base pairs of exogenous DNA. Introduction of the system into a target cell may result in deletion, wherein the deletion is less than 2, 3, 4, 5, 10, 50, or 100 base pairs of genomic DNA upstream or downstream of the insertion. Introduction of the system into a target cell may result in substitution, e.g., substitution of 1, 2, or 3 nucleotides, e.g., consecutive nucleotides.

The heterologous object sequence may be at least 5, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, or 700 base pairs.

In one aspect, the disclosure relates to a pharmaceutical composition comprising the system described above and a pharmaceutically acceptable excipient or carrier, wherein the pharmaceutically acceptable excipient or carrier is selected from the group consisting of a plasmid vector, a viral vector, a vesicle, and a lipid nanoparticle. In one aspect, the disclosure relates to a pharmaceutical composition comprising the system described above and multiple pharmaceutically acceptable excipients or carriers, wherein the pharmaceutically acceptable excipients or carriers are selected from the group consisting of a plasmid vector, a viral vector, a vesicle, and a lipid nanoparticle, e.g., where the system described above is delivered by two distinct excipients or carriers, e.g., two lipid nanoparticles, two viral vectors, or one lipid nanoparticle and one viral vector. The viral vector may be an adeno-associated virus (AAV).

In one aspect, the disclosure relates to a host cell (e.g., a mammalian cell, e.g., a human cell) comprising the system described above.

In one aspect, the disclosure relates to a method of correcting a mutation in the human PAH gene in a cell, tissue or subject, the method comprising administering the system described above to the cell, tissue or subject, wherein optionally the correction of the mutant PAH gene comprises an amino acid substitution of W408R, Q261R, and/or Q243R (reversing the pathogenic substitution which is R408W, R261Q, or R243Q). In another aspect, the correction of the mutant PAH gene comprises a nucleic acid substitution of IVS10-11A>G (reversing the pathogenic substitution which is IVS10-11G>A). The system may be introduced *in vivo*, *in vitro*, *ex vivo*, or *in situ*. The nucleic acid of (a) may be integrated into the genome of the host cell. In some embodiments, the nucleic acid of (a) is not integrated into the genome of the host cell. In some embodiments, the heterologous object sequence is inserted at only one target site in the host cell genome. The heterologous object sequence may be inserted at two or more target sites in the host cell genome, e.g., at the same corresponding site in two homologous chromosomes or at two different sites on the same or different chromosomes. The heterologous object sequence may encode a mammalian polypeptide, or a fragment or a variant thereof. The components of the system may be delivered on 1, 2, 3, 4, or more distinct nucleic acid molecules. The system may be introduced into a host cell by electroporation or by using at least one vehicle selected from a plasmid vector, a viral vector, a vesicle, and a lipid nanoparticle.

Features of the compositions or methods can include one or more of the following enumerated embodiments.

Enumerated Embodiments

1. A template RNA comprising, e.g., from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a sequence comprising the core nucleotides of a gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer (e.g., comprises one or more flanking nucleotides that are adjacent to the core nucleotides), or wherein the gRNA spacer has a sequence of a spacer chosen from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A;

(ii) a gRNA scaffold that binds a gene modifying polypeptide (e.g., binds the Cas domain of the gene modifying polypeptide),

(iii) a heterologous object sequence comprising a mutation region to introduce a mutation into (e.g., to correct a mutation in) a second portion of the human PAH gene (wherein optionally the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, a mutation region, and a pre-edit homology region), and

(iv) a primer binding site (PBS) sequence comprising at least 3, 4, 5, 6, 7, or 8 bases with 100% identity to a third portion of the human PAH gene.

2. The template RNA of embodiment 1, wherein the heterologous object sequence comprises the core nucleotides of an RT template sequence from Table 3A, Table 3B, Table 3C, or Table 3D, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or wherein the heterologous object sequence comprises a sequence of an RT template sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

3. The template RNA of embodiment 1, wherein the heterologous object sequence comprises the core nucleotides of the RT template sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the gRNA spacer sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence (e.g., comprises one or more flanking nucleotides that are adjacent to the core nucleotides), or wherein the heterologous object sequence comprises a sequence of an RT template sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

4. The template RNA according to any one of embodiments 1-3 wherein the PBS sequence has a sequence comprising the core nucleotides of the PBS sequence from the same row of Table 3A, Table 3B, Table 3C, or Table 3D as the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence (e.g., comprises one or more flanking nucleotides that are adjacent to the core nucleotides).

5. The template RNA according to any one of embodiments 1-3, wherein the PBS sequence has a sequence comprising the core nucleotides of a PBS sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, the gRNA spacer sequence, or both, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence, or wherein the PBS sequence comprises a PBS sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, the gRNA spacer sequence, or both.

6. The template RNA according to any of embodiments 1-5, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

7. The template RNA according to any of embodiments 1-5, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12 that corresponds to the RT template

sequence, the gRNA spacer sequence, or both, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

8. A template RNA comprising, e.g., from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene,

(ii) a gRNA scaffold that binds a gene modifying polypeptide (e.g., binds the Cas domain of the gene modifying polypeptide),

(iii) a heterologous object sequence comprising a mutation region to introduce a mutation into a second portion of the human PAH gene, wherein the heterologous object sequence comprises the core nucleotides of an RT template sequence of Table 3A, Table 3B, Table 3C, or Table 3D, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or wherein the heterologous object sequence comprises an RT template sequence of Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A; and

(iv) a PBS sequence comprising at least 3, 4, 5, 6, 7, or 8 bases of 100% identity to a third portion of the human PAH gene.

9. The template RNA of embodiment 8, wherein the gRNA spacer comprises the core nucleotides of a gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer sequence, or wherein the gRNA spacer comprises a gRNA spacer sequence of Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

10. The template RNA of any one of embodiments 1-9, wherein the gRNA spacer comprises ACCTCAATCCTTTGGGTGTA (SEQ ID NO: 16355), or a sequence having 1, 2, or 3 substitutions thereto.

11. The template RNA of any one of embodiments 1-9, wherein the gRNA spacer comprises CCTCAATCCTTTGGGTGTAT (SEQ ID NO: 16332) , or a sequence having 1, 2, or 3 substitutions thereto.

12. The template RNA of any one of embodiments 1-9, wherein the gRNA spacer comprises TGGGTCGTAGCGAACTGAGA (SEQ ID NO: 16102), or a sequence having 1, 2, or 3 substitutions thereto.

13. The template RNA of any one of embodiments 1-9, wherein the gRNA spacer comprises GGGTCGTAGCGAACTGAGAA (SEQ ID NO: 16084), or a sequence having 1, 2, or 3 substitutions thereto.

14. The template RNA of any one of embodiments 1-9, wherein the gRNA spacer comprises TAGCGAACTGAGAAGGGCCA (SEQ ID NO: 16011), or a sequence having 1, 2, or 3 substitutions thereto.

15. The template RNA of any one of embodiments 1-9, wherein the gRNA spacer comprises ACTTTGCTGCCACAATACCT (SEQ ID NO: 16032), or a sequence having 1, 2, or 3 substitutions thereto.

16. The template RNA of embodiment 8, wherein the heterologous object sequence comprises the core nucleotides of the gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer sequence, or wherein the heterologous object sequence comprises the nucleotides of the gRNA spacer sequence of Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto.

17. The template RNA according to any one of embodiments 8-16, wherein the PBS sequence has a sequence comprising the core nucleotides of the PBS sequence from the same

row of Table 3A, Table 3B, Table 3C, or Table 3D as the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence.

18. The template RNA according to any one of embodiments 8-16, wherein the PBS sequence has a sequence comprising the core nucleotides of a PBS sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, the gRNA spacer sequence, or both, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence, or wherein the PBS sequence has a sequence comprising the a PBS sequence of Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, the gRNA spacer sequence, or both.

19. The template RNA according to any of embodiments 8-18, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

20. The template RNA according to any of embodiments 8-18, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12 that corresponds to the RT template sequence, the gRNA spacer sequence, or both, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

21. A gene modifying system for modifying DNA, comprising:

(a) a first RNA comprising, from 5' to 3, (i) a guide RNA sequence that is complementary to a first portion of the human PAH gene, wherein the guide RNA sequence has a sequence comprising the core nucleotides of a spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the guide RNA sequence, or wherein the guide RNA sequence has a sequence of a spacer chosen from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A; and (ii) a sequence (e.g., a

scaffold region) that binds a gene modifying polypeptide (e.g., binds the Cas domain of the gene modifying polypeptide), and

(b) a second RNA comprising (iii) a heterologous object sequence comprising a nucleotide substitution to introduce a mutation into a second portion of the human PAH gene (wherein optionally the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, a mutation region, and a pre-edit homology region), (iv) a primer region comprising at least 5, 6, 7, or 8 bases of 100% identity to a third portion of the human PAH gene, and (v) an RRS (RNA binding protein recognition sequence) that binds a gene modifying protein.

22. The gene modifying system of embodiment 21, wherein the heterologous object sequence comprises the core nucleotides of an RT template sequence from Table 3A, Table 3B, Table 3C, or Table 3D, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or wherein the heterologous object sequence comprises a sequence of an RT template sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A, or a sequence having 1, 2, or 3 substitutions thereto.

23. The gene modifying system of embodiment 21, wherein the heterologous object sequence comprises the core nucleotides of the RT template sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the gRNA spacer sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or wherein the heterologous object sequence comprises a sequence of an RT template sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the gRNA spacer sequence, or a sequence having 1, 2, or 3 substitutions thereto.

24. The gene modifying system of any one of embodiments 21-23, wherein the PBS sequence has a sequence comprising the core nucleotides of the PBS sequence from the same row of Table 3A, Table 3B, Table 3C, or Table 3D as the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive

nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence, or wherein the PBS sequence has a sequence of a PBS sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A, or a sequence having 1, 2, or 3 substitutions thereto.

25. The gene modifying system of one of embodiments 21-23, wherein the PBS sequence has a sequence comprising the core nucleotides of a PBS sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, the gRNA spacer sequence, or both, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence, or wherein the PBS sequence has a sequence of a PBS sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto.

26. The gene modifying system of any one of embodiments 21-25, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

27. The gene modifying system of any one of embodiments 21-25, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12 that corresponds to the RT template sequence, the gRNA spacer sequence, or both, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

28. A gene modifying system for modifying DNA, comprising:

(a) a first RNA comprising, from 5' to 3, (i) a guide RNA sequence that is complementary to a first portion of the human PAH gene, and (ii) a sequence (e.g., a scaffold region) that binds a gene modifying polypeptide (e.g., binds the Cas domain of the gene modifying polypeptide), and

(b) a second RNA comprising (iii) a heterologous object sequence comprising a nucleotide substitution to introduce a mutation into a second portion of the human PAH gene, wherein the heterologous object sequence comprises the core nucleotides of an RT template sequence of Table 3A, Table 3B, Table 3C, or Table 3D, or a sequence having 1, 2, or 3

substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or wherein the heterologous object sequence comprises an RT sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A, or a sequence having 1, 2, or 3 substitutions thereto, and (iv) a primer region comprising at least 5, 6, 7, or 8 bases of 100% homology to a third portion of the human PAH gene, and (v) an RRS (RNA binding protein recognition sequence) that binds a gene modifying protein.

29. The gene modifying system of embodiment 28, wherein the gRNA spacer comprises the core nucleotides of a gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer sequence, or wherein the gRNA spacer comprises a gRNA spacer sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A, or a sequence having 1, 2, or 3 substitutions thereto.

30. The gene modifying system of embodiment 28, wherein the heterologous object sequence comprises the core nucleotides of the gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer sequence, or wherein the heterologous object sequence comprises a gRNA spacer sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto.

31. The gene modifying system of any one of embodiments 28-30, wherein the PBS sequence has a sequence comprising the core nucleotides of the PBS sequence from the same row of Table 3A, Table 3B, Table 3C, or Table 3D as the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence, or wherein the PBS sequence has a sequence comprising a PBS sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A, or a sequence having 1, 2, or 3 substitutions thereto.

32. The gene modifying system of any one of embodiments 28-30, wherein the PBS sequence has a sequence comprising the core nucleotides of a PBS sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the RT template sequence, the gRNA spacer sequence, or both, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence, or wherein the PBS sequence has a sequence comprising a PBS sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, the gRNA spacer sequence, or both, or a sequence having 1, 2, or 3 substitutions thereto.

33. The gene modifying system of any one of embodiments 28-32, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

34. The gene modifying system of any one of embodiments 28-32, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12 that corresponds to the RT template sequence, the gRNA spacer sequence, or both, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

35. A gRNA comprising (i) a gRNA spacer sequence that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a sequence comprising the core nucleotides of a gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D, Table 2A, Table 2B, Table 2C, or Table 2D, or Table 4A, Table 4B, Table 4C, or Table 4D, or a sequence having 1, 2, or 3 substitutions thereto and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer sequence, or wherein the gRNA spacer has a sequence comprising a gRNA spacer sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A or a sequence having 1, 2, or 3 substitutions thereto; and (ii) a gRNA scaffold.

36. The gRNA of embodiment 35, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

37. The gRNA of embodiment 35, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12 that corresponds to the gRNA spacer sequence, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

38. A template RNA comprising: (iii) a heterologous object sequence comprising a mutation region to introduce a mutation into a second portion of the human PAH gene, wherein the heterologous object sequence comprises the core nucleotides of an RT template sequence of Table 3A, Table 3B, Table 3C, or Table 3D, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or wherein the heterologous object sequence comprises an RT template sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A or a sequence having 1, 2, or 3 substitutions thereto, and (iv) a PBS sequence comprising at least 5, 6, 7, or 8 bases of 100% homology to a third portion of the human PAH gene.

39. The template RNA according to embodiment 38, wherein the PBS sequence has a sequence comprising the core nucleotides of the PBS sequence from the same row of Table 3A, Table 3B, Table 3C, or Table 3D as the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence.

40. The template RNA according to embodiment 38, wherein the PBS sequence has a sequence comprising the core nucleotides of a PBS sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence, or wherein the PBS sequence comprises a PBS sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto.

41. The template RNA according to any one of embodiments 1-20 or 38-40, the gene modifying system of any one of embodiments 21-35, or the gRNA of any one of embodiments 35-37, wherein the mutation introduced by the system is a W408R, Q261R, Q243R, and/or IVS10-11A>G mutation (e.g., to correct a pathogenic R408W, R261Q, R243Q, and/or IVS10-11G>A mutation) of the PAH gene.

42. The template RNA according to any one of embodiments 1-20 or 38-41 or the gene modifying system of any one of embodiments 35-37 or 41, wherein the pre-edit sequence comprises between about 1 nucleotide to about 35 nucleotides (e.g., comprises about 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, or 30-35 nucleotides) in length.

43. The template RNA according to any one of embodiments 1-20 or 38-42 or the gene modifying system of any one of embodiments 35-37, 41, or 42, wherein the mutation region comprises a single nucleotide.

44. The template RNA according to any one of embodiments 1-20 or 38-42 or the gene modifying system of any one of embodiments35-37, 41, or 42, wherein the mutation region is at least two nucleotides in length.

45. The template RNA according to any one of embodiments 1-20, 38-42, or 44 or the gene modifying system of any one of embodiments 35-37, 41, 42, or 44, wherein the mutation region is up to 32 (e.g., up to 5, 10, 15, 20, 25, 30, or 32) nucleotides in length and comprises one, two, or three sequence differences relative to a second portion of the human PAH gene.

46. The template RNA according to any one of embodiments 1-20, 38-42, 44, or 45 or the gene modifying system of any one of embodiments 35-37, 41, 42, 44, or 45, wherein the mutation region comprises two sequences differences relative to a second portion of the human PAH gene.

47. The template RNA according to any one of embodiments 1-20, 38-42, or 44-46 or the gene modifying system of any one of embodiments 35-37, 41, 42, or 44-46, wherein the

mutation region comprises a first region (e.g., a first nucleotide) designed to correct a pathogenic mutation in the PAH gene and a second region (e.g., a second nucleotide) designed to inactivate a PAM sequence (e.g., a "PAM-kill" mutation as described herein).

48. The template RNA according to any one of embodiments 1-20 or 38-46 or the gene modifying system of any one of embodiments 35-37 or 41-46, wherein the mutation region comprises less than 80%, 70%, 60%, 50%, 40%, or 30% identity to corresponding portion of the human PAH gene.

49. The template RNA of any one of the preceding embodiments, wherein the template RNA comprises one or more silent mutations (e.g., silent substitutions), e.g., as exemplified in Tables 7A-7C, 8A-8D, E6, or E6A.

50. The template RNA of any of the preceding embodiments, wherein the mutation region comprises a first region designed to correct a pathogenic mutation in the PAH gene and a second region designed to introduce a silent substitution.

51. The template RNA of any one of the preceding embodiments, which comprises one or more chemically modified nucleotides.

52. A gene modifying system comprising:

a template RNA of any of embodiments 1-20, 38-46, or a system of any of embodiments 35-37 or 41-46, and

a gene modifying polypeptide, or a nucleic acid (e.g., RNA) encoding the gene modifying polypeptide.

53. The gene modifying system of embodiment 52, wherein the gene modifying polypeptide comprises:

a reverse transcriptase (RT) domain (e.g., an RT domain from a retrovirus, or a polypeptide domain having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acids sequence identity thereto); and

a Cas domain that binds to the target DNA molecule and is heterologous to the RT domain (e.g., a Cas9 domain); and

optionally, a linker disposed between the RT domain and the Cas domain.

54. The gene modifying system of embodiment 53, wherein the RT domain comprises:

(a) an RT domain of Table 6; or

(b) an RT domain from a murine leukemia virus (MMLV), a porcine endogenous retrovirus (PERV); Avian reticuloendotheliosis virus (AVIRE), a feline leukemia virus (FLV), simian foamy virus (SFV) (e.g., SFV3L), bovine leukemia virus (BLV), Mason-Pfizer monkey virus (MPMV), human foamy virus (HFV), or bovine foamy/syncytial virus (BFV/BSV).

55. The gene modifying system of embodiment 53 or 54, wherein the Cas domain comprises a Cas domain of Table 7 or Table 8.

56. The gene modifying system of any one of embodiments 53-55, wherein the Cas domain:

(a) is a Cas9 domain;

(b) is a SpCas9 domain, a BlatCas9 domain, a Nme2Cas9 domain, a PnpCas9 domain, a SauCas9 domain, a SauCas9-KKH domain, a SauriCas9 domain, a SauriCas9-KKH domain, a ScaCas9-Sc++ domain, a SpyCas9 domain, a SpyCas9-NG domain, a SpyCas9-SpRY domain, or a St1Cas9 domain; and/or

(c) is a Cas9 domain comprising an N670A mutation, an N611A mutation, an N605A mutation, an N580A mutation, an N588A mutation, an N872A mutation, an N863 mutation, an N622A mutation, or an H840A mutation.

57. The gene modifying system of embodiment 56, wherein the Cas9 domain binds a PAM sequence listed in Table 7 or Table 12.

58. The gene modifying system of embodiment 57, wherein a second portion of the human PAH gene overlaps with a PAM recognized by the Cas domain, e.g., wherein the second portion of the human PAH gene is within the PAM or wherein the PAM is within the second portion of the human PAH gene).

59. The gene modifying system any one of embodiments 52-58, wherein the gRNA spacer is a gRNA spacer according to Table 1A, Table 1B, Table 1C, or Table 1D, and the Cas domain comprises a Cas domain listed in the same row of Table 1A, Table 1B, Table 1C, or Table 1D, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

60. The gene modifying system of any one of embodiments 52-58, wherein the template RNA comprises a sequence of a template RNA sequence of Table 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

61. The gene modifying system of any one of embodiments 52-60, wherein:

(a) the template RNA comprises a sequence of a template RNA sequence of Tables 3A-3D, 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A;

(b) the Cas domain comprises a Cas domain of Table 7 or Table 8;

(c) the linker comprises a linker sequence of Table 10 (e.g., of any of SEQ ID NOs: 5217, 5106, 5190, and 5218); and

(d) the gene modifying polypeptide comprises one or two NLS sequences from Table 11 (e.g., of any of SEQ ID NOs: 5245, 5290, 5323, 5330, 5349, 5350, 5351, and 4001).

62. The gene modifying system of any of embodiments 52-61, which produces a first nick in a first strand of the human PAH gene.

63. The gene modifying system of embodiment 62, which further comprises a second strand-targeting gRNA that directs a second nick to the second strand of the human PAH gene.

64. The gene modifying system of embodiment 63, wherein the second strand-targeting gRNA comprises:

(i) a sequence comprising the core nucleotides of a left gRNA spacer sequence or a right gRNA spacer sequence from Table 2A, Table 2B, Table 2C, or Table 2D, and optionally

comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the left gRNA spacer sequence or right gRNA spacer sequence; or

(ii) a second -strand-targeting gRNA comprising a spacer sequence of Table 6A, or a spacer sequence having 1, 2, or 3 substitutions thereto.

65. The gene modifying system of embodiment 63, wherein the second strand-targeting gRNA comprises a sequence comprising the core nucleotides of a left gRNA spacer sequence or a right gRNA spacer sequence from Table 2A, Table 2B, Table 2C, or Table 2D that corresponds to the gRNA spacer sequence of (i), and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the left gRNA spacer sequence or right gRNA spacer sequence.

66. The gene modifying system of embodiment 63, wherein the second strand-targeting gRNA comprises:

(i) a sequence comprising the core nucleotides of a second nick gRNA sequence from Table 4A, Table 4B, Table 4C, Table 4D, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the second nick gRNA sequence; or

(ii) a second -strand-targeting gRNA comprising a spacer sequence from Table 6A or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

67. The gene modifying system of embodiment 63, wherein the second strand-targeting gRNA comprises a sequence comprising the core nucleotides of the second nick gRNA sequence from Table 4A, Table 4B, Table 4C, or Table 4D that corresponds to the gRNA spacer sequence of (i), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the second nick gRNA sequence.

68. The gene modifying system of any one of embodiments 52-67, wherein the second strand-targeting gRNA has a "PAM-in orientation" with the template RNA of the gene modifying system, e.g., as exemplified in Tables 2A-2D, 4A-4D, or 6A.

69. The gene modifying system of any one of embodiments 52-68, the second strand-targeting gRNA targets a sequence overlapping the target mutation of the template RNA.

70. The gene modifying system of embodiment 69, wherein second strand-targeting gRNA comprises:

(i) a sequence (e.g., a spacer sequence) complementary to the PAH mutation;

(ii) a sequence (e.g., a spacer sequence) complementary to the wild-type sequence at the target locus;

(iii) a sequence (e.g., a spacer sequence) complementary to a SNP proximal to the target locus, e.g., a SNP contained in the genomic DNA of a subject (e.g., a patient);

(iv) a sequence (e.g., spacer sequence) complementary to or comprising one or more silent substitutions proximal to the target locus.

71. The template RNA, gene modifying system, or gRNA, of any one of the preceding embodiments, wherein the gRNA spacer comprises about 1, 2, 3, or more flanking nucleotides of the gRNA spacer.

72. The template RNA or gene modifying system of any one of the preceding embodiments, wherein the heterologous object sequence comprises about 2, 3, 4, 5, 10, 20, 30, 40, or more flanking nucleotides of the RT template sequence.

73. The template RNA or gene modifying system, of any one of the preceding embodiments, wherein the heterologous object sequence comprises between about 8-30, 9-25, 10-20, 11-16, or 12-15 (e.g., about 11-16) nucleotides.

74. The template RNA or gene modifying system, of any one of the preceding embodiments, wherein the mutation region comprises 1, 2, or 3 nucleotide positions of sequence differences relative to the corresponding portion of the human PAH gene.

75. The template RNA or gene modifying system of any one of the preceding embodiments, wherein the mutation region comprises at least 2 nucleotide positions of sequence difference relative to the corresponding portion of the human PAH gene.

76. The template RNA or gene modifying system, of any one of the preceding embodiments, wherein the post-edit homology region and/or pre-edit homology region comprises 100% identity to the PAH gene.

77. The template RNA or gene modifying system of any one of the preceding embodiments, wherein the PBS sequence additionally comprises about 1, 2, 3, 4, 5, 6, 7, or more flanking nucleotides.

78. The template RNA or gene modifying system of any one of the preceding embodiments, wherein the PBS sequence comprises about 5-20, 8-16, 8-14, 8-13, 9-13, 9-12, or 10-12 (e.g., about 9-12) nucleotides.

79. The template RNA or gene modifying system of any one of the preceding embodiments, wherein the PBS sequence binds within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nick site in the PAH gene.

80. The gene modifying system of any one of the preceding embodiments, wherein the domains of the gene modifying polypeptide are joined by a peptide linker.

81. The gene modifying system of embodiment 80, wherein the linker comprises a sequence of a linker of Table 10 (e.g., of any of SEQ ID NOs: 5217, 5106, 5190, and 5218).

82. The gene modifying system of any one of the preceding embodiments, wherein the gene modifying polypeptide further comprise one or more nuclear localization sequences (NLS).

83. The gene modifying system of embodiment 82, wherein the gene modifying polypeptide comprises a first NLS and a second NLS.

84. The gene modifying system of embodiment 82 or 83, wherein the NLS comprises a sequence of a NLS of Table 11 (e.g., of any of SEQ ID NOs: 5245, 5290, 5323, 5330, 5349, 5350, 5351, and 4001).

85. A template RNA comprising a sequence of a template RNA of Table 4A-4D, 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

86. A template RNA comprising a sequence of a template RNA of Tables 4A-4D, 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

87. A gene modifying system comprising:

(i) a template RNA comprising a sequence of a template RNA of Table 4A, Table 4B, Table 4C, or Table 4D, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(ii) a second-nick gRNA sequence from the same row of Table 4A, Table 4B, Table 4C, or Table 4D as (i), a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

88. A gene modifying system comprising:

(i) a template RNA comprising a sequence of a template RNA of Table 4A, Table 4B, Table 4C, or Table 4D; and

(ii) a second-nick gRNA sequence from the same row of Table 4A, Table 4B, Table 4C, or Table 4D as (i).

89. A DNA encoding the template RNA of any one of embodiments 1-20, 38-48, 71-79, 85, or 86, or the gRNA of any one of embodiments 34-37.

90. A pharmaceutical composition, comprising the system of any one of embodiments 49-84, 87, or 88, or one or more nucleic acids encoding the same, and a pharmaceutically acceptable excipient or carrier.

91. The pharmaceutical composition of embodiment 90, wherein the pharmaceutically acceptable excipient or carrier is selected from the group consisting of a plasmid vector, a viral vector, a vesicle, and a lipid nanoparticle.

92. The pharmaceutical composition of embodiment 91, wherein the viral vector is an adeno-associated virus.

93. A host cell (e.g., a mammalian cell, e.g., a human cell) comprising the template RNA or gene modifying system of any one of the preceding embodiments.

94. A method of making the template RNA of any one of embodiments 1-20, 38-48, 71-79, 85, or 86, the method comprising synthesizing the template RNA by *in vitro* transcription (e.g., solid state synthesis) or by introducing a DNA encoding the template RNA into a host cell under conditions that allow for production of the template RNA.

95. A method for modifying a target site in the human PAH gene in a cell, the method comprising contacting the cell with the gene modifying system of any one of embodiments 49-84, 87, or 88, or DNA encoding the same, thereby modifying the target site in the human PAH gene in a cell.

96. A method for modifying a target site in the human PAH gene in a cell, the method comprising contacting the cell with: (i) the template RNA of any one of embodiments 49-84, 87, or 88, or DNA encoding the same; and (ii) a gene modifying polypeptide or a nucleic acid

encoding a gene modifying polypeptide, thereby modifying the target site in the human PAH gene in a cell.

97. A method for treating a subject having a disease or condition associated with a mutation in the human PAH gene, the method comprising administering to the subject the gene modifying system of any one of embodiments 49-84, 87, or 88, or DNA encoding the same, thereby treating the subject having a disease or condition associated with a mutation in the human PAH gene.

98. A method for treating a subject having a disease or condition associated with a mutation in the human PAH gene, the method comprising administering to the subject the template RNA of any one of embodiments 49-84, 87, or 88, or DNA encoding the same; and (ii) a gene modifying polypeptide or a nucleic acid encoding a gene modifying polypeptide, thereby treating the subject having a disease or condition associated with a mutation in the human PAH gene.

99. The method of embodiment 97 or 98, wherein the disease or condition is phenylketonuria (PKU).

100. The method of any one of embodiments 97-99, wherein the subject has a R408W, R261Q, R243Q, and/or IVS10-11G>A mutation.

101. A method for treating a subject having PKU the method comprising administering to the subject the gene modifying system of any one of embodiments 49-84, 87, or 88, or DNA encoding the same, thereby treating the subject having PKU.

102. A method for treating a subject having PKU the method comprising administering to the subject (i) the template RNA of any one of embodiments 49-84, 87, or 88, or DNA encoding the same, and (ii) a gene modifying polypeptide or a nucleic acid encoding a gene modifying polypeptide, thereby treating the subject having PKU.

103. The gene modifying system or method of any one of the preceding embodiments, wherein introduction of the system into a target cell results in a correction of a pathogenic mutation in the PAH gene.

104. The gene modifying system or method of any one of the preceding embodiments, wherein the pathogenic mutation is a W408R, Q261R, Q243R, and/or IVS10-11A>G mutation, and wherein the correction comprises an amino acid substitution of R408W, R261Q, or R243Q, or a nucleic acid substitution of IVS10-11G>A.

105. The gene modifying system or method of any of the preceding embodiments, wherein correction of the mutation occurs in at least 30% (e.g., 30%, 40%, 50%, 60%, 70%, or more) of target nucleic acids.

106. The gene modifying system or method of any of the preceding embodiments, wherein correction of the mutation occurs in at least 30% (e.g., 30%, 40%, 50%, 60%, 70%, or more) of target cells.

107. The gene modifying system or method of any of the preceding embodiments, wherein the gene modifying system comprises a second strand-targeting gRNA, and wherein correction of the mutation in a population of target cells is increased relative to a population of target cells treated with a gene modifying system comprising a template RNA without a second strand-targeting gRNA.

108. The gene modifying system or method of any of the preceding embodiments, wherein the template RNA comprises one or more silent substitutions (e.g., as exemplified in Tables 7A, X4, and X4A), and wherein correction of the mutation in a population of target cells is increased relative to a population of target cells treated with a gene modifying system comprising a template RNA that does not comprise one or more silent substitutions.

109. The method of any of the preceding embodiments, wherein the cell is a mammalian cell, such as a human cell.

110. The method of any one of the preceding embodiments, wherein the subject is a human.

111. The method of any of the preceding embodiments, wherein the contacting occurs *ex vivo*, e.g., wherein the cell's or subject's DNA is modified *ex vivo*.

112. The method of any of the preceding embodiments, wherein the contacting occurs *in vivo*, e.g., wherein the cell's or subject's DNA is modified *in vivo*.

113. The method of any of the preceding embodiments, wherein contacting the cell or the subject with the system comprises contacting the cell or a cell within the subject with a nucleic acid (e.g., DNA or RNA) encoding the gene modifying polypeptide under conditions that allow for production of the gene modifying polypeptide.

Additional Enumerated Embodiments

A1. A template RNA comprising, from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a nucleotide sequence comprising ACCTCAATCCTTTGGGTGTA (SEQ ID NO: 16355), or a nucleotide sequence having 1, 2, or 3 substitution thereto;

(ii) a gRNA scaffold that binds a Cas domain of a gene modifying polypeptide,

(iii) a heterologous object sequence comprising a mutation region to correct a mutation in a second portion of the human PAH gene, and

(iv) a primer binding site (PBS) sequence comprising at least 5 bases with 100% identity to a third portion of the human PAH gene.

A2. The template RNA of embodiment A1, wherein the gRNA spacer has a nucleotide sequence comprising ACCTCAATCCTTTGGGTGTA (SEQ ID NO: 16355).

A3. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 30 nucleotides from the 3' end of a sequence

according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttctcagttcgctacgacccatac (SEQ ID NO: 24984), or a sequence having 1, 2, 3, or 4 substitutions thereto.

A4. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttctcagttcgctacgacccatac (SEQ ID NO: 24984), or comprises at least 40, 50, 60, or 70 nucleotides from the 3' end of said sequence, or a sequence having 1, 2, 3, or 4 substitutions thereto.

A5. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 30 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttctcagttcgctacgacccatac (SEQ ID NO: 24984).

A6. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttctcagttcgctacgacccatac (SEQ ID NO: 24984), or comprises at least 40, 50, 60, or 70 nucleotides from the 3' end of said sequence.

A7. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of at least 5, 6, 7, or 8 nucleotides from the 5' end of a sequence according to acccaaagg (SEQ ID NO: 37628), or a sequence having 1 substitution thereto.

A8. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of acccaaagg (SEQ ID NO: 37628)

A9. A template RNA comprising, from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a nucleotide sequence comprising CCTCAATCCTTTGGGTGTAT (SEQ ID NO: 16332), or a nucleotide sequence having 1, 2, or 3 substitution thereto;

(ii) a gRNA scaffold that binds a Cas domain of a gene modifying polypeptide,

(iii) a heterologous object sequence comprising a mutation region to correct a mutation in a second portion of the human PAH gene, and

(iv) a primer binding site (PBS) sequence comprising at least 5 bases with 100% identity to a third portion of the human PAH gene.

A10. The template RNA of embodiment A9, wherein the gRNA spacer has a nucleotide sequence comprising CCTCAATCCTTTGGGTGTAT (SEQ ID NO: 16332).

A11. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 50 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttctcagttcgctacgacccata (SEQ ID NO: 24975), or a sequence having 1, 2, 3, or 4 substitutions thereto.

A12. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttctcagttcgctacgacccata (SEQ ID NO: 24975), or comprises at least 60 or 70 nucleotides from the 3' end of said sequence, or a sequence having 1, 2, 3, or 4 substitutions thereto.

A13. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 50 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttctcagttcgctacgacccata (SEQ ID NO: 24975).

A14. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttctcagttcgctacgacccata (SEQ ID NO: 24975), or comprises at least 60 or 70 nucleotides from the 3' end of said sequence.

A15. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of at least 5, 6, 7, or 8 nucleotides from the 5' end of a sequence according to cacccaaag (SEQ ID NO: 37629), or a sequence having 1 substitution thereto.

A16. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of cacccaaag (SEQ ID NO: 37629).

A17. A template RNA comprising, from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a nucleotide sequence comprising TGGGTCGTAGCGAACTGAGA (SEQ ID NO: 16102), or a nucleotide sequence having 1, 2, or 3 substitution thereto;

(ii) a gRNA scaffold that binds a Cas domain of a gene modifying polypeptide,

(iii) a heterologous object sequence comprising a mutation region to correct a mutation in a second portion of the human PAH gene, and

(iv) a primer binding site (PBS) sequence comprising at least 5 bases with 100% identity to a third portion of the human PAH gene.

A18. The template RNA of embodiment A17, wherein the gRNA spacer has a nucleotide sequence comprising TGGGTCGTAGCGAACTGAGA (SEQ ID NO: 16102).

A19. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 10 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttct (SEQ ID NO: 24863), or a sequence having 1, 2, 3, or 4 substitutions thereto.

A20. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttct (SEQ ID NO: 24863), or comprises at least 20, 30, 40, of 50 nucleotides from the 3' end of said sequence, or a sequence having 1, 2, 3, or 4 substitutions thereto.

A21. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 10 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttct (SEQ ID NO: 24863).

A22. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttct (SEQ ID NO: 24863), or comprises at least 20, 30, 40, of 50 nucleotides from the 3' end of said sequence.

A23. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of at least 5, 6, 7, or 8 nucleotides from the 5' end of a sequence according to cagttcgct (SEQ ID NO: 37630), or a sequence having 1 substitution thereto.

A24. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of cagttcgct (SEQ ID NO: 37630).

A25. A template RNA comprising, from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a nucleotide sequence comprising GGGTCGTAGCGAACTGAGAA (SEQ ID NO: 16084), or a nucleotide sequence having 1, 2, or 3 substitution thereto;

(ii) a gRNA scaffold that binds a Cas domain of a gene modifying polypeptide,

(iii) a heterologous object sequence comprising a mutation region to correct a mutation in a second portion of the human PAH gene, and

(iv) a primer binding site (PBS) sequence comprising at least 5 bases with 100% identity to a third portion of the human PAH gene.

A26. The template RNA of embodiment A25, wherein the gRNA spacer has a nucleotide sequence comprising GGGTCGTAGCGAACTGAGAA (SEQ ID NO: 16084).

A27. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 9 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttc (SEQ ID NO: 24856), or a sequence having 1, 2, 3, or 4 substitutions thereto.

A28. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttc (SEQ ID NO: 24856), or comprises at least 10, 20, 30, 40, or 50 nucleotides from the 3' end of said sequence, or a sequence having 1, 2, 3, or 4 substitutions thereto.

A29. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 9 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttc (SEQ ID NO: 24856).

A30. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCggcccttc (SEQ ID NO: 24856), or comprises at least 10, 20, 30, 40, or 50 nucleotides from the 3' end of said sequence.

A31. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of at least 5, 6, 7, or 8 nucleotides from the 5' end of a sequence according to tcagttcgc (SEQ ID NO: 37631), or a sequence having 1 substitution thereto.

A32. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of tcagttcgc (SEQ ID NO: 37631)

A33. A template RNA comprising, from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a nucleotide sequence comprising

TAGCGAACTGAGAAGGGCCA (SEQ ID NO: 16011), or a nucleotide sequence having 1, 2, or 3 substitution thereto;

(ii) a gRNA scaffold that binds a Cas domain of a gene modifying polypeptide,

(iii) a heterologous object sequence comprising a mutation region to correct a mutation in a second portion of the human PAH gene, and

(iv) a primer binding site (PBS) sequence comprising at least 5 bases with 100% identity to a third portion of the human PAH gene.

A34. The template RNA of embodiment A33, wherein the gRNA spacer has a nucleotide sequence comprising TAGCGAACTGAGAAGGGCCA (SEQ ID NO: 16011).

A35. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 3 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCgg (SEQ I DNO: 24817), or a sequence having 1, 2, 3, or 4 substitutions thereto.

A36. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCgg (SEQ I DNO: 24817), or comprises at least 5, 10, 20, 30, 40, or 50 nucleotides from the 3' end of said sequence, or a sequence having 1, 2, 3, or 4 substitutions thereto.

A37. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 3 nucleotides from the 3' end of a sequence according to tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCgg (SEQ I DNO: 24817).

A38. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of tcactcaagcctgtggttttggtcttaggaactttgctgccacaatacctCgg (SEQ ID NO: 24817), or comprises at least 5, 10, 20, 30, 40, or 50 nucleotides from the 3' end of said sequence.

A39. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of at least 5, 6, 7, or 8 nucleotides from the 5' end of a sequence according to cccttctca (SEQ ID NO: 37632), or a sequence having 1 substitution thereto.

A40. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of cccttctca (SEQ ID NO: 37632).

A41. A template RNA comprising, from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a nucleotide sequence comprising ACTTTGCTGCCACAATACCT (SEQ ID NO: 16032), or a nucleotide sequence having 1, 2, or 3 substitution thereto;

(ii) a gRNA scaffold that binds a Cas domain of a gene modifying polypeptide,

(iii) a heterologous object sequence comprising a mutation region to correct a mutation in a second portion of the human PAH gene, and

(iv) a primer binding site (PBS) sequence comprising at least 5 bases with 100% identity to a third portion of the human PAH gene.

A42. The template RNA of embodiment A41, wherein the gRNA spacer has a nucleotide sequence comprising ACTTTGCTGCCACAATACCT (SEQ ID NO: 16032).

A43. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 4 nucleotides from the 3' end of a sequence according to caagacctcaatcctttgggtgtatgggtcgtagcgaactgagaagggccGagg (SEQ ID NO: 24825), or a sequence having 1, 2, 3, or 4 substitutions thereto.

A44. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of caagacctcaatcctttgggtgtatgggtcgtagcgaactgagaagggccGagg (SEQ ID NO: 24825), or comprises at least 5, 10, 20, 30, 40, or 50 nucleotides from the 3' end of said sequence, or a sequence having 1, 2, 3, or 4 substitutions thereto.

A45. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of at least 4 nucleotides from the 3' end of a sequence according to caagacctcaatcctttgggtgtatgggtcgtagcgaactgagaagggccGagg (SEQ ID NO: 24825).

A46. The template RNA of any of the preceding embodiments, wherein the heterologous object sequence comprises a sequence of caagacctcaatcctttgggtgtatgggtcgtagcgaactgagaagggccGagg (SEQ ID NO: 24825), or comprises at least 5, 10, 20, 30, 40, or 50 nucleotides from the 3' end of said sequence.

A47. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of at least 5, 6, 7, 8, 9, 10, or 15 nucleotides from the 5' end of a sequence according to tattgtggcagcaaagt (SEQ ID NO: 37633), or a sequence having 1 substitution thereto.

A48. The template RNA of any of the preceding embodiments, wherein the PBS comprises a sequence of tattgtggcagcaaagt (SEQ ID NO: 37633).

A49. The template RNA of any of the preceding embodiments, wherein the gRNA scaffold has a sequence according to GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTGC (SEQ ID NO: 37627), or a sequence having at least 90% identity thereto.

A50. The template RNA of any of the preceding embodiments, wherein the gRNA scaffold has a sequence according to GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTGC (SEQ ID NO: 37627).

A51. The template RNA of any of the preceding embodiments, wherein the mutation region comprises a single nucleotide.

A52. The template RNA of any of embodiments A1-51, wherein the mutation region is at least two nucleotides in length.

A53. The template RNA of any of the preceding embodiments, wherein the mutation region is up to 20 nucleotides in length and comprises one, two, or three sequence differences relative to the second portion of the human PAH gene.

A54. The template RNA of any of embodiments A1-53, wherein the mutation region comprises a first region designed to correct a pathogenic mutation in the PAH gene and a second region designed to inactivate a PAM sequence.

A55. The template RNA of any of embodiments A1-54, wherein the mutation region comprises a first region designed to correct a pathogenic mutation in the PAH gene and a second region designed to introduce a silent substitution.

A56. The template RNA of any of the preceding embodiments, which is configured to edit a pathogenic R408W mutation in the human PAH gene.

A57. The template RNA of embodiment A56, which is configured to convert an R408W mutation to arginine.

A58. The template RNA of any of the preceding embodiments, which comprises one or more chemically modified nucleotides.

A59. A gene modifying system comprising:
a template RNA of any of the preceding embodiments, and
a gene modifying polypeptide, or a nucleic acid encoding the gene modifying polypeptide.

A60. The gene modifying system of embodiment A59, wherein the gene modifying polypeptide comprises an RT domain having a sequence according to SEQ ID NO: 8,003, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

A61. The gene modifying system of embodiment A59, wherein the gene modifying polypeptide comprises an RT domain having a sequence according to SEQ ID NO: 8,020, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

A62. The gene modifying system of embodiment 5A9, wherein the gene modifying polypeptide comprises an RT domain having a sequence according to SEQ ID NO: 8,074, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

A63. The gene modifying system of embodiment A59, wherein the gene modifying polypeptide comprises an RT domain having a sequence according to SEQ ID NO: 8,113, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

A64. The gene modifying system of embodiment A59, wherein the gene modifying polypeptide comprises DNA binding domain having a sequence of a Cas9 nickase comprising an N863A mutation, e.g., a sequence according to SEQ ID NO: 11,096, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

A65. The gene modifying system of embodiment A59, which produces a first nick in a first strand of the human PAH gene.

A66. The gene modifying system of embodiment A65, which further comprises a second strand-targeting gRNA that directs a second nick to the second strand of the human PAH gene.

A67. The gene modifying system of embodiment A66, wherein the first nick and the second nick are 80-120 nucleotides apart.

A68. The gene modifying system of embodiment A66, wherein the template RNA and the second strand-targeting gRNA are configured to produce an outward nick orientation.

A69. The gene modifying system of embodiment A66, wherein the second strand-targeting gRNA comprises a spacer sequence that is complementary to a human PAH gene having a disease mutation or a wild-type sequence.

A70. A method for modifying a target site in the human PAH gene in a cell, the method comprising contacting the cell with the gene modifying system of embodiment 59, thereby modifying the target site in the human PAH gene in a cell.

A71. The method of embodiment A70, wherein correction of the mutation occurs in at least 30% of target nucleic acids.

A72. A method for treating a subject having a disease or condition associated with a mutation in the human PAH gene, wherein the disease or condition is phenylketonuria (PKU) or hyperphenylalaninemia (e.g., mild or severe hyperphenylalaninemia), the method comprising administering to the subject the gene modifying system of embodiment 59, thereby treating the subject having a disease or condition associated with a mutation in the human PAH gene.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIG. 1 depicts a gene modifying system as described herein. The left hand diagram shows the gene modifying polypeptide, which comprises a Cas nickase domain (e.g., spCas9 N863A) and a reverse transcriptase domain (RT domain) which are linked by a linker. The right hand diagram shows the template RNA which comprises, from 5' to 3', a gRNA spacer, a gRNA scaffold, a heterologous object sequence, and a primer binding site sequence (PBS sequence). The heterologous object sequence can comprise a mutation region that comprises one or more sequence differences relative to the target site. The heterologous object sequence can also comprise a pre-edit homology region and a post-edit homology region, which flank the mutation region. Without wishing to be bound by theory, it is thought that the gRNA spacer of the template RNA binds to the second strand of a target site in the genome, and the gRNA

scaffold of the template RNA binds to the gene modifying polypeptide, e.g., localizing the gene modifying polypeptide to the target site in the genome. It is thought that the Cas domain of the gene modifying polypeptide nicks the target site (e.g., the first strand of the target site), e.g., allowing the PBS sequence to bind to a sequence adjacent to the site to be altered on the first strand of the target site. It is thought that the RT domain of the gene modifying polypeptide uses the first strand of the target site that is bound to the complementary sequence comprising the PBS sequence of the template RNA as a primer and the heterologous object sequence of the template RNA as a template to, e.g., polymerize a sequence complementary to the heterologous object sequence. Without wishing to be bound by theory, it is thought that reverse transcription can then proceed through the pre-edit homology region, then through the mutation region, and then through the post-edit homology region, thereby producing a DNA strand comprising a mutation specified by the heterologous object sequence.

FIG. 2 is a graph showing the percent rewriting achieved using the RNAV209-013 or RNAV214-040 gene modifying polypeptides with the indicated template RNAs.

FIG. 3 is a graph showing the amount of *Fah* mRNA relative to wild type when template RNAs are used with the RNAV209-013 or RNAV214-040 gene modifying polypeptides.

FIG.4 is a graph showing the percentage of Cas9-positive hepatocytes 6 hours following dosing with LNPs containing various gene modifying polypeptides and template RNAs.

FIG. 5 is a graph showing the rewrite levels in liver samples 6 days following dosing with LNPs containing various gene modifying polypeptides and template RNAs.

FIG. 6 is a graph showing wild type *Fah* mRNA restoration compared to littermate heterozygous mice in liver samples following dosing with LNPs containing various gene modifying polypeptides and template RNAs.

FIG. 7 is a graph showing Fah protein distribution in liver samples following dosing with LNPs containing various gene modifying polypeptides and template RNAs.

FIG. 8 is a series of western blots showing Cas9-RT Expression 6 hours after infusion of Cas9-RT mRNA + TTR guide LNP. Each lane represents an individual animal where 20 μg of tissue homogenate was added per lane. Positive control was from an *in vitro* cell experiment where Cas9-RT was expressed (described previously). GAPDH was used as a loading control for each sample. n=4 per group, vehicle or treated.

FIG. 9 is a graph showing gene editing of TTR locus after treatment with Cas9-RT mRNA + TTR guide LNP. Level of indels detected at the TTR locus measured by TIDE analysis of Sanger sequencing of the TTR locus where the protospacer targets.

FIG. 10 is a graph showing that TTR Serum levels decrease after treatment with Cas9-RT mRNA + TTR guide LNP. Measurement of circulating TTR levels 5 days after mice were treated with LNPs encapsulating Cas9-RT + TTR guide RNA.

FIG. 11 is a graph showing Cas9-RT Expression after infusion of Cas9-RT mRNA + TTR guide LNP. Relative expression quantified by ProteinSimple Jess capillary electrophoresis Western blot. Numbers in the symbols are animal number in group. Vehicle n=2, Cas9-RT + TTR guide n=3.

FIG. 12 is a graph showing gene editing of TTR locus after infusion of Cas9-RT mRNA + TTR guide LNP. Level of indels detected at the TTR locus were measured by amplicon sequencing of the TTR locus where the protospacer targets. Each animal had 8 different biopsies taken across the liver where amplicon sequencing measured the percentage of reads showing an indel.

FIG. 13 is a graph showing percent rewriting in primary mouse hepatocytes nucleofected with various gene modifying systems.

FIG. 14 is a graph showing percent editing in primary mouse hepatocytes nucleofected with various gene modifying systems containing second-nick gRNAs.

FIG. 15 is a heat map showing rewriting efficiency of various gene modifying systems with or without second-nick gRNAs.

FIG. 16 is a graph showing the percent of mouse hepatocytes expressing Cas9 six hours post-dosing with various gene modifying systems.

FIG. 17 is a pair of western blots showing expression of Cas9 in mouse liver samples six hours post-dosing with various gene modifying systems.

FIG. 18 is a graph showing the level of phenylalanine (Phe) present in plasma samples 7 days post-dosing with various gene modifying systems.

FIGs. 19A-19B are graphs showing percent rewriting (**FIG. 19A**) and percent indel (**FIG. 19B**) in mouse liver 7 days post-dosing with various gene modifying systems.

FIGs. 20A-20C are graphs showing percent rewriting in liver samples (**FIG. 20A**),levels of Phe in plasma (**FIG. 20B**), and percent indels in mouse liver (**FIG. 20C**) 7 days post-dosing with various gene modifying systems.

FIGs. 21A-21B are a pair of graphs showing percent rewriting and percent indel in liver samples (**FIG. 21A**) and levels of Phe in plasma (**FIG. 21B**) 7 days post-dosing with various gene modifying systems with or without second-nick gRNAs.

FIG. 22 is a graph showing the level of phenylalanine (Phe) in the plasma versus percent rewriting in samples obtained from mice treated with various gene modifying systems.

FIG. 23 is a graph showing percent rewriting in HEK293T cells containing the ***M fascicularis*** PAH gene for four different mutation types using template RNAs containing four different spacer sequences.

FIGs. 24A-24C are graphs showing percent rewriting (FIG. 24A) and percent indels (FIG. 24B) in mouse liver cells, or concentration of Phe in plasma (FIG. 24C) days post-dosing with LNPs comprising various gene modifying systems.

FIGs. 25A-25C are heat maps showing percent rewriting for each combination of template RNA and second strand-targeting RNA in primary human hepatocytes (**FIG. 25A**) and primary mouse hepatocytes (**FIG. 25C**) following transfection with (**FIGs. 25A** and **25B**) or LNP delivery of (**FIG. 25C**) various gene modifying systems.

FIGs. 26A-26B are graphs showing percent rewriting (FIG. 26A) and percent indels (FIG. 26B) in 7- and 28-day liver samples following LNP delivery of gene modifying systems to mice.

FIG. 27 is a graph showing the concentration of Phe in 7- and 28-day plasma samples following LNP delivery of gene modifying systems to mice.

FIG. 28 is a graph showing the concentration of Phe in 7- and 28-day brain samples following LNP delivery of gene modifying systems to mice.

FIG. 29 is a graph showing the concentration of Phe in the brain versus concentration of Phe in the plasma from samples used to generate **FIGs. 27** and **28**.

FIGs. 30A-30H are heat maps showing percent rewriting for each combination of template RNA and second strand-targeting RNA following mRNA delivery of gene modifying systems to primary cyno hepatocytes.

FIGs. 31A-31B are a graph stratified by silent substitution (**FIG. 31A**) showing percent total rewriting following mRNA delivery of various gene modifying systems utilizing the hPKU3

template RNAs comprising various silent substitutions into human iPSC-derived hepatoblasts and a chart (**FIG. 31B**) showing the particular silent substitutions utilized in **FIG. 31A**.

FIGs. 32A-32B are a graph stratified by silent substitution (**FIG. 32A**) showing percent total rewriting following mRNA delivery of various gene modifying systems utilizing the hPKU4 template RNAs comprising various silent substitutions into human iPSC-derived hepatoblasts and a chart (**FIG. 32B**) showing the particular silent substitutions utilized in **FIG. 32A**.

FIGs. 33A-33B are a graph (33A) and a chart (33B) showing are a graph stratified by silent substitution (**FIG. 33A**) showing percent total rewriting following mRNA delivery of various gene modifying systems utilizing the hPKU5 template RNAs comprising various silent substitutions into human iPSC-derived hepatoblasts and a chart (**FIG. 33B**) showing the particular silent substitutions utilized in **FIG. 33A**.

FIGs. 34A-34B are a graph stratified by silent substitution (**FIG. 34A**) showing percent total rewriting following mRNA delivery of various gene modifying systems utilizing the hPKU6 template RNAs comprising various silent substitutions into human iPSC-derived hepatoblasts and a chart (**FIG. 34B**) showing the particular silent substitutions utilized in **FIG. 34A**.

FIG. 35 is a graph showing serum levels of Phe in mice following treatment with LNPs comprising various gene modifying systems.

FIGs. 36A-36B are graphs showing percent rewriting (**FIG. 36A**) and percent indels (**FIG. 36B**) in mouse liver following treatment with LNPs comprising various gene modifying systems.

DETAILED DESCRIPTION

Definitions

The term "expression cassette," as used herein, refers to a nucleic acid construct comprising nucleic acid elements sufficient for the expression of the nucleic acid molecule of the instant invention.

A "gRNA spacer", as used herein, refers to a portion of a nucleic acid that has complementarity to a target nucleic acid and can, together with a gRNA scaffold, target a Cas protein to the target nucleic acid.

A "gRNA scaffold", as used herein, refers to a portion of a nucleic acid that can bind a Cas protein and can, together with a gRNA spacer, target the Cas protein to the target nucleic

acid. In some embodiments, the gRNA scaffold comprises a crRNA sequence, tetraloop, and tracrRNA sequence.

A "gene modifying polypeptide", as used herein, refers to a polypeptide comprising a retroviral reverse transcriptase, or a polypeptide comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to a retroviral reverse transcriptase, which is capable of integrating a nucleic acid sequence (e.g., a sequence provided on a template nucleic acid) into a target DNA molecule (e.g., in a mammalian host cell, such as a genomic DNA molecule in the host cell). In some embodiments, the gene modifying polypeptide is capable of integrating the sequence substantially without relying on host machinery. In some embodiments, the gene modifying polypeptide integrates a sequence into a random position in a genome, and in some embodiments, the gene modifying polypeptide integrates a sequence into a specific target site. In some embodiments, a gene modifying polypeptide includes one or more domains that, collectively, facilitate 1) binding the template nucleic acid, 2) binding the target DNA molecule, and 3) facilitate integration of the at least a portion of the template nucleic acid into the target DNA. Gene modifying polypeptides include both naturally occurring polypeptides as well as engineered variants of the foregoing, e.g., having one or more amino acid substitutions to the naturally occurring sequence. Gene modifying polypeptides also include heterologous constructs, e.g., where one or more of the domains recited above are heterologous to each other, whether through a heterologous fusion (or other conjugate) of otherwise wild-type domains, as well as fusions of modified domains, e.g., by way of replacement or fusion of a heterologous sub-domain or other substituted domain. Exemplary gene modifying polypeptides, and systems comprising them and methods of using them, that can be used in the methods provided herein are described, e.g., in PCT/US2021/020948, which is incorporated herein by reference with respect to gene modifying polypeptides that comprise a retroviral reverse transcriptase domain. In some embodiments, a gene modifying polypeptide integrates a sequence into a gene. In some embodiments, a gene modifying polypeptide integrates a sequence into a sequence outside of a gene. A "gene modifying system," as used herein, refers to a system comprising a gene modifying polypeptide and a template nucleic acid.

The term "domain" as used herein refers to a structure of a biomolecule that contributes to a specified function of the biomolecule. A domain may comprise a contiguous region (e.g., a

contiguous sequence) or distinct, non-contiguous regions (e.g., non-contiguous sequences) of a biomolecule. Examples of protein domains include, but are not limited to, an endonuclease domain, a DNA binding domain, a reverse transcription domain; an example of a domain of a nucleic acid is a regulatory domain, such as a transcription factor binding domain. In some embodiments, a domain (e.g., a Cas domain) can comprise two or more smaller domains (e.g., a DNA binding domain and an endonuclease domain).

As used herein, the term "exogenous", when used with reference to a biomolecule (such as a nucleic acid sequence or polypeptide) means that the biomolecule was introduced into a host genome, cell or organism by the hand of man. For example, a nucleic acid that is as added into an existing genome, cell, tissue or subject using recombinant DNA techniques or other methods is exogenous to the existing nucleic acid sequence, cell, tissue or subject.

As used herein, "first strand" and "second strand", as used to describe the individual DNA strands of target DNA, distinguish the two DNA strands based upon which strand the reverse transcriptase domain initiates polymerization, e.g., based upon where target primed synthesis initiates. The first strand refers to the strand of the target DNA upon which the reverse transcriptase domain initiates polymerization, e.g., where target primed synthesis initiates. The second strand refers to the other strand of the target DNA. First and second strand designations do not describe the target site DNA strands in other respects; for example, in some embodiments the first and second strands are nicked by a polypeptide described herein, but the designations 'first' and 'second' strand have no bearing on the order in which such nicks occur.

The term "heterologous," as used herein to describe a first element in reference to a second element means that the first element and second element do not exist in nature disposed as described. For example, a heterologous polypeptide, nucleic acid molecule, construct or sequence refers to (a) a polypeptide, nucleic acid molecule or portion of a polypeptide or nucleic acid molecule sequence that is not native to a cell in which it is expressed, (b) a polypeptide or nucleic acid molecule or portion of a polypeptide or nucleic acid molecule that has been altered or mutated relative to its native state, or (c) a polypeptide or nucleic acid molecule with an altered expression as compared to the native expression levels under similar conditions. For example, a heterologous regulatory sequence (e.g., promoter, enhancer) may be used to regulate expression of a gene or a nucleic acid molecule in a way that is different than the gene or a nucleic acid molecule is normally expressed in nature. In another example, a heterologous

domain of a polypeptide or nucleic acid sequence (e.g., a DNA binding domain of a polypeptide or nucleic acid encoding a DNA binding domain of a polypeptide) may be disposed relative to other domains or may be a different sequence or from a different source, relative to other domains or portions of a polypeptide or its encoding nucleic acid. In certain embodiments, a heterologous nucleic acid molecule may exist in a native host cell genome, but may have an altered expression level or have a different sequence or both. In other embodiments, heterologous nucleic acid molecules may not be endogenous to a host cell or host genome but instead may have been introduced into a host cell by transformation (e.g., transfection, electroporation), wherein the added molecule may integrate into the host genome or can exist as extra-chromosomal genetic material either transiently (e.g., mRNA) or semi-stably for more than one generation (e.g., episomal viral vector, plasmid or other self-replicating vector).

As used herein, "insertion" of a sequence into a target site refers to the net addition of DNA sequence at the target site, e.g., where there are new nucleotides in the heterologous object sequence with no cognate positions in the unedited target site. In some embodiments, a nucleotide alignment of the PBS sequence and heterologous object sequence to the target nucleic acid sequence would result in an alignment gap in the target nucleic acid sequence.

As used herein, a "deletion" generated by a heterologous object sequence in a target site refers to the net deletion of DNA sequence at the target site, e.g., where there are nucleotides in the unedited target site with no cognate positions in the heterologous object sequence. In some embodiments, a nucleotide alignment of the PBS sequence and heterologous object sequence to the target nucleic acid sequence would result in an alignment gap in the molecule comprising the PBS sequence and heterologous object sequence.

The term "inverted terminal repeats" or "ITRs" as used herein refers to AAV viral cis-elements named so because of their symmetry. These elements promote efficient multiplication of an AAV genome. It is hypothesized that the minimal elements for ITR function are a Rep-binding site (RBS; 5′-GCGCGCTCGCTCGCTC-3′ for AAV2; SEQ ID NO: 4601) and a terminal resolution site (TRS; 5′-AGTTGG-3′ for AAV2; SEQ ID NO: 4602) plus a variable palindromic sequence allowing for hairpin formation. According to the present invention, an ITR comprises at least these three elements (RBS, TRS, and sequences allowing the formation of an hairpin). In addition, in the present invention, the term "ITR" refers to ITRs of known natural AAV serotypes (e.g. ITR of a serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 AAV), to chimeric ITRs

formed by the fusion of ITR elements derived from different serotypes, and to functional variants thereof. "Functional variant" refers to a sequence presenting a sequence identity of at least 80%, 85%, 90%, preferably of at least 95% with a known ITR and allowing multiplication of the sequence that includes said ITR in the presence of Rep proteins.

The term "mutation region," as used herein, refers to a region in a template RNA having one or more sequence difference relative to the corresponding sequence in a target nucleic acid. The sequence difference may comprise, for example, a substitution, insertion, frameshift, or deletion.

The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence are inserted, deleted, or changed compared to a reference (e.g., native) nucleic acid sequence. A single alteration may be made at a locus (a point mutation), or multiple nucleotides may be inserted, deleted, or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art.

"Nucleic acid molecule" refers to both RNA and DNA molecules including, without limitation, complementary DNA ("cDNA"), genomic DNA ("gDNA"), and messenger RNA ("mRNA"), and also includes synthetic nucleic acid molecules, such as those that are chemically synthesized or recombinantly produced, such as RNA templates, as described herein. The nucleic acid molecule can be double-stranded or single-stranded, circular, or linear. If single-stranded, the nucleic acid molecule can be the sense strand or the antisense strand. Unless otherwise indicated, and as an example for all sequences described herein under the general format "SEQ ID NO:," or "nucleic acid comprising SEQ ID NO:1" refers to a nucleic acid, at least a portion which has either (i) the sequence of SEQ ID NO:1, or (ii) a sequence complimentary to SEQ ID NO:1. The choice between the two is dictated by the context in which SEQ ID NO:1 is used. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complementary to the desired target. Nucleic acid sequences of the present disclosure may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more naturally occurring nucleotides with an analog, inter-nucleotide modifications such as uncharged linkages (for example, methyl phosphonates, phosphotriesters,

phosphoramidates, carbamates, etc.), charged linkages (for example, phosphorothioates, phosphorodithioates, etc.), pendant moieties, (for example, polypeptides), intercalators (for example, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (for example, alpha anomeric nucleic acids, etc.). Also included are chemically modified bases (see, for example, Table 13), backbones (see, for example, Table 14), and modified caps (see, for example, Table 15). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of a molecule, e.g., peptide nucleic acids (PNAs). Other modifications can include, for example, analogs in which the ribose ring contains a bridging moiety or other structure such as modifications found in "locked" nucleic acids (LNAs). In various embodiments, the nucleic acids are in operative association with additional genetic elements, such as tissue-specific expression-control sequence(s) (e.g., tissue-specific promoters and tissue-specific microRNA recognition sequences), as well as additional elements, such as inverted repeats (e.g., inverted terminal repeats, such as elements from or derived from viruses, e.g., AAV ITRs) and tandem repeats, inverted repeats/direct repeats, homology regions (segments with various degrees of homology to a target DNA), untranslated regions (UTRs) (5′, 3′, or both 5′ and 3′ UTRs), and various combinations of the foregoing. The nucleic acid elements of the systems provided by the invention can be provided in a variety of topologies, including single-stranded, double-stranded, circular, linear, linear with open ends, linear with closed ends, and particular versions of these, such as doggybone DNA (dbDNA), closed-ended DNA (ceDNA).

As used herein, a "gene expression unit" is a nucleic acid sequence comprising at least one regulatory nucleic acid sequence operably linked to at least one effector sequence. A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if the promoter or enhancer affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be contiguous or non-contiguous. Where necessary to join two protein-coding regions, operably linked sequences may be in the same reading frame.

The terms "host genome" or "host cell", as used herein, refer to a cell and/or its genome into which protein and/or genetic material has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell and/or genome, but to the progeny of such a cell and/or the genome of the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A host genome or host cell may be an isolated cell or cell line grown in culture, or genomic material isolated from such a cell or cell line, or may be a host cell or host genome which composing living tissue or an organism. In some instances, a host cell may be an animal cell or a plant cell, e.g., as described herein. In certain instances, a host cell may be a mammalian cell, a human cell, avian cell, reptilian cell, bovine cell, horse cell, pig cell, goat cell, sheep cell, chicken cell, or turkey cell. In certain instances, a host cell may be a corn cell, soy cell, wheat cell, or rice cell.

As used herein, "operative association" describes a functional relationship between two nucleic acid sequences, such as a 1) promoter and 2) a heterologous object sequence, and means, in such example, the promoter and heterologous object sequence (e.g., a gene of interest) are oriented such that, under suitable conditions, the promoter drives expression of the heterologous object sequence. For instance, a template nucleic acid carrying a promoter and a heterologous object sequence may be single-stranded, e.g., either the (+) or (-) orientation. An "operative association" between the promoter and the heterologous object sequence in this template means that, regardless of whether the template nucleic acid will be transcribed in a particular state, when it is in the suitable state (e.g., is in the (+) orientation, in the presence of required catalytic factors, and NTPs, etc.), it is accurately transcribed. Operative association applies analogously to other pairs of nucleic acids, including other tissue-specific expression control sequences (such as enhancers, repressors and microRNA recognition sequences), IR/DR, ITRs, UTRs, or homology regions and heterologous object sequences or sequences encoding a retroviral RT domain.

The term "primer binding site sequence" or "PBS sequence," as used herein, refers to a portion of a template RNA capable of binding to a region comprised in a target nucleic acid sequence. In some instances, a PBS sequence is a nucleic acid sequence comprising at least 3, 4, 5, 6, 7, or 8 bases with 100% identity to the region comprised in the target nucleic acid sequence.

In some embodiments the primer region comprises at least 5, 6, 7, 8 bases with 100% identity to the region comprised in the target nucleic acid sequence. Without wishing to be bound by theory, in some embodiments when a template RNA comprises a PBS sequence and a heterologous object sequence, the PBS sequence binds to a region comprised in a target nucleic acid sequence, allowing a reverse transcriptase domain to use that region as a primer for reverse transcription, and to use the heterologous object sequence as a template for reverse transcription.

As used herein, a "stem-loop sequence" refers to a nucleic acid sequence (e.g., RNA sequence) with sufficient self-complementarity to form a stem-loop, e.g., having a stem comprising at least two (e.g., 3, 4, 5, 6, 7, 8, 9, or 10) base pairs, and a loop with at least three (e.g., four) base pairs. The stem may comprise mismatches or bulges.

As used herein, a "tissue-specific expression-control sequence" means nucleic acid elements that increase or decrease the level of a transcript comprising the heterologous object sequence in a target tissue in a tissue-specific manner, e.g., preferentially in on-target tissue(s), relative to off-target tissue(s). In some embodiments, a tissue-specific expression-control sequence preferentially drives or represses transcription, activity, or the half-life of a transcript comprising the heterologous object sequence in the target tissue in a tissue-specific manner, e.g., preferentially in an on-target tissue(s), relative to an off-target tissue(s). Exemplary tissue-specific expression-control sequences include tissue-specific promoters, repressors, enhancers, or combinations thereof, as well as tissue-specific microRNA recognition sequences. Tissue specificity refers to on-target (tissue(s) where expression or activity of the template nucleic acid is desired or tolerable) and off-target (tissue(s) where expression or activity of the template nucleic acid is not desired or is not tolerable). For example, a tissue-specific promoter drives expression preferentially in on-target tissues, relative to off-target tissues. In contrast, a microRNA that binds the tissue-specific microRNA recognition sequences is preferentially expressed in off-target tissues, relative to on-target tissues, thereby reducing expression of a template nucleic acid in off-target tissues. Accordingly, a promoter and a microRNA recognition sequence that are specific for the same tissue, such as the target tissue, have contrasting functions (promote and repress, respectively, with concordant expression levels, i.e., high levels of the microRNA in off-target tissues and low levels in on-target tissues, while promoters drive high expression in on-target tissues and low expression in off-target tissues) with regard to the transcription, activity, or half-life of an associated sequence in that tissue.

Table of Contents

6) Kits, Articles of Manufacture, and Pharmaceutical Compositions
7) Chemistry, Manufacturing, and Controls (CMC)

Introduction

This disclosure relates to methods for treating phenylketonuria (PKU) and compositions for targeting, editing, modifying or manipulating a DNA sequence (e.g., inserting a heterologous object sequence into a target site of a mammalian genome) at one or more locations in a DNA sequence in a cell, tissue or subject, e.g., *in vivo* or *in vitro*. The heterologous object DNA sequence may include, e.g., a substitution.

More specifically, the disclosure provides methods for treating PKU using reverse transcriptase-based systems for altering a genomic DNA sequence of interest, e.g., by inserting, deleting, or substituting one or more nucleotides into/from the sequence of interest.

The disclosure provides, in part, methods for treating PKU using a gene modifying system comprising a gene modifying polypeptide component and a template nucleic acid (e.g., template RNA) component. In some embodiments, a gene modifying system can be used to introduce an alteration into a target site in a genome. In some embodiments, the gene modifying polypeptide component comprises a writing domain (e.g., a reverse transcriptase domain), a DNA-binding domain, and an endonuclease domain (e.g., nickase domain). In some embodiments, the template nucleic acid (e.g., template RNA) comprises a sequence (e.g., a gRNA spacer) that binds a target site in the genome (e.g., that binds to a second strand of the target site), a sequence (e.g., a gRNA scaffold) that binds the gene modifying polypeptide component, a heterologous object sequence, and a PBS sequence. Without wishing to be bound by theory, it is thought that the template nucleic acid (e.g., template RNA) binds to the second strand of a target site in the genome, and binds to the gene modifying polypeptide component (e.g., localizing the polypeptide component to the target site in the genome). It is thought that the endonuclease (e.g., nickase) of the gene modifying polypeptide component cuts the target site (e.g., the first strand of the target site), e.g., allowing the PBS sequence to bind to a sequence adjacent to the site to be altered on the first strand of the target site. It is thought that the writing domain (e.g., reverse transcriptase domain) of the polypeptide component uses the first strand of the target site that is bound to the complementary sequence comprising the PBS sequence of the template nucleic acid as a primer and the heterologous object sequence of the template nucleic

acid as a template to, e.g., polymerize a sequence complementary to the heterologous object sequence. Without wishing to be bound by theory, it is thought that selection of an appropriate heterologous object sequence can result in substitution, deletion, and/or insertion of one or more nucleotides at the target site.

Gene modifying systems

In some embodiments, a gene modifying system described herein comprises: (A) a gene modifying polypeptide or a nucleic acid encoding the gene modifying polypeptide, wherein the gene modifying polypeptide comprises (i) a reverse transcriptase domain, and either (x) an endonuclease domain that contains DNA binding functionality or (y) an endonuclease domain and separate DNA binding domain; and (B) a template RNA. A gene modifying polypeptide, in some embodiments, acts as a substantially autonomous protein machine capable of integrating a template nucleic acid sequence into a target DNA molecule (e.g., in a mammalian host cell, such as a genomic DNA molecule in the host cell), substantially without relying on host machinery. For example, the gene modifying protein may comprise a DNA-binding domain, a reverse transcriptase domain, and an endonuclease domain. In some embodiments, the DNA-binding function may involve an RNA component that directs the protein to a DNA sequence, e.g., a gRNA spacer. In other embodiments, the gene modifying polypeptide may comprise a reverse transcriptase domain and an endonuclease domain. The RNA template element of a gene modifying system is typically heterologous to the gene modifying polypeptide element and provides an object sequence to be inserted (reverse transcribed) into the host genome. In some embodiments, the gene modifying polypeptide is capable of target primed reverse transcription. In some embodiments, the gene modifying polypeptide is capable of second-strand synthesis.

In some embodiments the gene modifying system is combined with a second polypeptide. In some embodiments, the second polypeptide may comprise an endonuclease domain. In some embodiments, the second polypeptide may comprise a polymerase domain, e.g., a reverse transcriptase domain. In some embodiments, the second polypeptide may comprise a DNA-dependent DNA polymerase domain. In some embodiments, the second polypeptide aids in completion of the genome edit, e.g., by contributing to second-strand synthesis or DNA repair resolution.

A functional gene modifying polypeptide can be made up of unrelated DNA binding, reverse transcription, and endonuclease domains. This modular structure allows combining of functional domains, e.g., dCas9 (DNA binding), MMLV reverse transcriptase (reverse transcription), FokI (endonuclease). In some embodiments, multiple functional domains may arise from a single protein, e.g., Cas9 or Cas9 nickase (DNA binding, endonuclease).

In some embodiments, a gene modifying polypeptide includes one or more domains that, collectively, facilitate 1) binding the template nucleic acid, 2) binding the target DNA molecule, and 3) facilitate integration of the at least a portion of the template nucleic acid into the target DNA. In some embodiments, the gene modifying polypeptide is an engineered polypeptide that comprises one or more amino acid substitutions to a corresponding naturally occurring sequence. In some embodiments, the gene modifying polypeptide comprises two or more domains that are heterologous relative to each other, e.g., through a heterologous fusion (or other conjugate) of otherwise wild-type domains, or well as fusions of modified domains, e.g., by way of replacement or fusion of a heterologous sub-domain or other substituted domain. For instance, in some embodiments, one or more of: the RT domain is heterologous to the DBD; the DBD is heterologous to the endonuclease domain; or the RT domain is heterologous to the endonuclease domain.

In some embodiments, a template RNA molecule for use in the system comprises, from 5′ to 3′ (1) a gRNA spacer; (2) a gRNA scaffold; (3) heterologous object sequence (4) a primer binding site (PBS) sequence. In some embodiments:

(1) Is a gRNA spacer of ~18-22 nt, e.g., is 20 nt

(2) Is a gRNA scaffold comprising one or more hairpin loops, e.g., 1, 2, of 3 loops for associating the template with a Cas domain, e.g., a nickase Cas9 domain. In some embodiments, the gRNA scaffold comprises the sequence, from 5′ to 3′, GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT GAAAAAGTGGGACCGAGTCGGTCC (SEQ ID NO: 5008).

(3) In some embodiments, the heterologous object sequence is, e.g., 7-74, e.g., 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, or 70-80 nt or, 80-90 nt in length. In some embodiments, the first (most 5′) base of the sequence is not C.

(4) In some embodiments, the PBS sequence that binds the target priming sequence after nicking occurs is e.g., 3-20 nt, e.g., 7-15 nt, e.g., 12-14 nt. In some embodiments, the PBS sequence has 40-60% GC content.

In some embodiments, a second gRNA associated with the system may help drive complete integration. In some embodiments, the second gRNA may target a location that is 0-200 nt away from the first-strand nick, e.g., 0-50, 50-100, 100-200 nt away from the first-strand nick. In some embodiments, the second gRNA can only bind its target sequence after the edit is made, e.g., the gRNA binds a sequence present in the heterologous object sequence, but not in the initial target sequence.

In some embodiments, a gene modifying system described herein is used to make an edit in HEK293, K562, U2OS, or HeLa cells. In some embodiment, a gene modifying system is used to make an edit in primary cells, e.g., primary cortical neurons from E18.5 mice.

In some embodiments, a gene modifying polypeptide as described herein comprises a reverse transcriptase or RT domain (e.g., as described herein) that comprises a MoMLV RT sequence or variant thereof. In embodiments, the MoMLV RT sequence comprises one or more mutations selected from D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, and K103L. In embodiments, the MoMLV RT sequence comprises a combination of mutations, such as D200N, L603W, and T330P, optionally further including T306K and/or W313F.

In some embodiments, an endonuclease domain (e.g., as described herein) nCas9, e.g., comprising an N863A mutation (e.g., in spCas9) or a H840A mutation.

In some embodiments, the heterologous object sequence (e.g., of a system as described herein) is about 1-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, or more, nucleotides in length.

In some embodiments, the RT and endonuclease domains are joined by a flexible linker, e.g., comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 5006).

In some embodiments, the endonuclease domain is N-terminal relative to the RT domain. In some embodiments, the endonuclease domain is C-terminal relative to the RT domain.

In some embodiments, the system incorporates a heterologous object sequence into a target site by TPRT, e.g., as described herein.

In some embodiments, a gene modifying polypeptide comprises a DNA binding domain. In some embodiments, a gene modifying polypeptide comprises an RNA binding domain. In some embodiments, the RNA binding domain comprises an RNA binding domain of B-box protein, MS2 coat protein, dCas, or an element of a sequence of a table herein. In some embodiments, the RNA binding domain is capable of binding to a template RNA with greater affinity than a reference RNA binding domain.

In some embodiments, a gene modifying system is capable of producing an insertion into the target site of at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a gene modifying system is capable of producing an insertion into the target site of at least 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, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a gene modifying system is capable of producing an insertion into the target site of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases). In some embodiments, a gene modifying system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a gene modifying system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a gene modifying system is capable of producing a deletion of at least 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, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a gene modifying system is capable of producing a deletion of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases). In some embodiments, a gene modifying system is capable of producing a substitution into the target site of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 or more

nucleotides. In some embodiments, a gene modifying system is capable of producing a substitution in the target site of 1-2, 2-3, 3-4, 4-5, 5-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, or 90-100 nucleotides.

In some embodiments, the substitution is a transition mutation. In some embodiments, the substitution is a transversion mutation. In some embodiments, the substitution converts an adenine to a thymine, an adenine to a guanine, an adenine to a cytosine, a guanine to a thymine, a guanine to a cytosine, a guanine to an adenine, a thymine to a cytosine, a thymine to an adenine, a thymine to a guanine, a cytosine to an adenine, a cytosine to a guanine, or a cytosine to a thymine.

In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a gene by altering, adding, or deleting sequences in a promoter or enhancer, e.g. sequences that bind transcription factors. In some embodiments, an insertion, deletion, substitution, or combination thereof alters translation of a gene (e.g. alters an amino acid sequence), inserts or deletes a start or stop codon, alters or fixes the translation frame of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof alters splicing of a gene, e.g. by inserting, deleting, or altering a splice acceptor or donor site. In some embodiments, an insertion, deletion, substitution, or combination thereof alters transcript or protein half-life. In some embodiments, an insertion, deletion, substitution, or combination thereof alters protein localization in the cell (e.g. from the cytoplasm to a mitochondria, from the cytoplasm into the extracellular space (e.g. adds a secretion tag)). In some embodiments, an insertion, deletion, substitution, or combination thereof alters (e.g. improves) protein folding (e.g. to prevent accumulation of misfolded proteins). In some embodiments, an insertion, deletion, substitution, or combination thereof, alters, increases, decreases the activity of a gene, e.g. a protein encoded by the gene.

Exemplary gene modifying polypeptides, and systems comprising them and methods of using them are described, e.g., in PCT/US2021/020948, which is incorporated herein by reference with respect to retroviral RT domains, including the amino acid and nucleic acid sequences therein.

Exemplary gene modifying polypeptides and retroviral RT domain sequences are also described, e.g., in International Application No. PCT/US21/20948 filed March 4, 2021, e.g., at Table 30, Table 31, and Table 44 therein; the entire application is incorporated by reference herein with respect to retroviral RTs, e.g., in said sequences and tables. Accordingly, a gene modifying polypeptide described herein may comprise an amino acid sequence according to any of the Tables mentioned in this paragraph, or a domain thereof (e.g., a retroviral RT domain), or a functional fragment or variant of any of the foregoing, or an amino acid sequence having at least 70%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In some embodiments, a polypeptide for use in any of the systems described herein can be a molecular reconstruction or ancestral reconstruction based upon the aligned polypeptide sequence of multiple homologous proteins. In some embodiments, a reverse transcriptase domain for use in any of the systems described herein can be a molecular reconstruction or an ancestral reconstruction, or can be modified at particular residues, based upon alignments of reverse transcriptase domains from the same or different sources. A skilled artisan can, based on the Accession numbers provided herein, align polypeptides or nucleic acid sequences, e.g., by using routine sequence analysis tools as Basic Local Alignment Search Tool (BLAST) or CD-Search for conserved domain analysis. Molecular reconstructions can be created based upon sequence consensus, e.g. using approaches described in Ivics et al., Cell 1997, 501 – 510 ; Wagstaff et al., Molecular Biology and Evolution 2013, 88-99.

Polypeptide components of gene modifying systems

In some embodiments, the gene modifying polypeptide possesses the functions of DNA target site binding, template nucleic acid (e.g., RNA) binding, DNA target site cleavage, and template nucleic acid (e.g., RNA) writing, e.g., reverse transcription. In some embodiments, each functions is contained within a distinct domain. In some embodiments, a function may be attributed to two or more domains (e.g., two or more domains, together, exhibit the functionality). In some embodiments, two or more domains may have the same or similar function (e.g., two or more domains each independently have DNA-binding functionality, e.g., for two different DNA sequences). In other embodiments, one or more domains may be capable of enabling one or more functions, e.g., a Cas9 domain enabling both DNA binding and target site cleavage. In some embodiments, the domains are all located within a single polypeptide. In

some embodiments, a first domain is in one polypeptide and a second domain is in a second polypeptide. For example, in some embodiments, the sequences may be split between a first polypeptide and a second polypeptide, e.g., wherein the first polypeptide comprises a reverse transcriptase (RT) domain and wherein the second polypeptide comprises a DNA-binding domain and an endonuclease domain, e.g., a nickase domain. As a further example, in some embodiments, the first polypeptide and the second polypeptide each comprise a DNA binding domain (e.g., a first DNA binding domain and a second DNA binding domain). In some embodiments, the first and second polypeptide may be brought together post-translationally via a split-intein to form a single gene modifying polypeptide.

In some aspects, a gene modifying polypeptide described herein comprises (e.g., a system described herein comprises a gene modifying polypeptide that comprises): 1) a Cas domain (e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); 2) a reverse transcriptase (RT) domain of Table D, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto, wherein the RT domain is C-terminal of the Cas domain; and a linker disposed between the RT domain and the Cas domain, wherein the linker has a sequence from the same row of Table D as the RT domain, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto.

In some embodiments, the RT domain has a sequence with 100% identity to the RT domain of Table D and the linker has a sequence with 100% identity to the linker sequence from the same row of Table D as the RT domain. In some embodiments, the Cas domain comprises a sequence of Table 8, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto. In some embodiments, the gene modifying polypeptide comprises an amino acid sequence according to any of SEQ ID NOs: 1-3332 in the sequence listing, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto.

In some embodiments, the gene modifying polypeptide comprises a GG amino acid sequence between the Cas domain and the linker, an AG amino acid sequence between the RT domain and the second NLS, and/or a GG amino acid sequence between the linker and the RT domain. In some embodiments, the gene modifying polypeptide comprises a sequence of SEQ ID NO: 4000 which comprises the first NLS and the Cas domain, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto. In some embodiments, the gene modifying polypeptide comprises a sequence of SEQ ID NO: 4001 which comprises the

second NLS, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

Exemplary N-terminal NLS-Cas9 domain

MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGG (SEQ ID NO: 4000)

Exemplary C-terminal sequence comprising an NLS

AGKRTADGSEFEKRTADGSEFESPKKKAKVE (SEQ ID NO: 4001)

Writing domain (RT Domain)

In certain aspects of the present invention, the writing domain of the gene modifying system possesses reverse transcriptase activity and is also referred to as a reverse transcriptase domain (a RT domain). In some embodiments, the RT domain comprises an RT catalytic portion and RNA-binding region (e.g., a region that binds the template RNA).

In some embodiments, a nucleic acid encoding the reverse transcriptase is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some

embodiments the reverse transcriptase domain is a heterologous reverse transcriptase from a retrovirus. In some embodiments, the RT domain comprising a gene modifying polypeptide has been mutated from its original amino acid sequence, e.g., has at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions. In some embodiments, the RT domain is derived from the RT of a retrovirus, e.g., HIV-1 RT, Moloney Murine Leukemia Virus (MMLV) RT, avian myeloblastosis virus (AMV) RT, or Rous Sarcoma Virus (RSV) RT.

In some embodiments, the retroviral reverse transcriptase (RT) domain exhibits enhanced stringency of target-primed reverse transcription (TPRT) initiation, e.g., relative to an endogenous RT domain. In some embodiments, the RT domain initiates TPRT when the 3 nt in the target site immediately upstream of the first strand nick, e.g., the genomic DNA priming the RNA template, have at least 66% or 100% complementarity to the 3 nt of homology in the RNA template. In some embodiments, the RT domain initiates TPRT when there are less than 5 nt mismatched (e.g., less than 1, 2, 3, 4, or 5 nt mismatched) between the template RNA homology and the target DNA priming reverse transcription. In some embodiments, the RT domain is modified such that the stringency for mismatches in priming the TPRT reaction is increased, e.g., wherein the RT domain does not tolerate any mismatches or tolerates fewer mismatches in the priming region relative to a wild-type (e.g., unmodified) RT domain. In some embodiments, the RT domain comprises a HIV-1 RT domain. In embodiments, the HIV-1 RT domain initiates lower levels of synthesis even with three nucleotide mismatches relative to an alternative RT domain (e.g., as described by Jamburuthugoda and Eickbush J Mol Biol 407(5):661-672 (2011); incorporated herein by reference in its entirety). In some embodiments, the RT domain forms a dimer (e.g., a heterodimer or homodimer). In some embodiments, the RT domain is monomeric. In some embodiments, an RT domain, naturally functions as a monomer or as a dimer (e.g., heterodimer or homodimer). In some embodiments, an RT domain naturally functions as a monomer, e.g., is derived from a virus wherein it functions as a monomer. In embodiments, the RT domain is selected from an RT domain from murine leukemia virus (MLV; sometimes referred to as MoMLV) (e.g., P03355), porcine endogenous retrovirus (PERV) (e.g., UniProt Q4VFZ2), mouse mammary tumor virus (MMTV) (e.g., UniProt P03365), Avian reticuloendotheliosis virus (AVIRE) (e.g., UniProtKB accession: P03360); Feline leukemia virus (FLV or FeLV) (e.g., e.g., UniProtKB accession: P10273); Mason-Pfizer monkey virus (MPMV) (e.g., UniProt P07572), bovine leukemia virus (BLV) (e.g., UniProt P03361), human T-cell

leukemia virus-1 (HTLV-1) (e.g., UniProt P03362), human foamy virus (HFV) (e.g., UniProt P14350), simian foamy virus (SFV) (e.g., SFV3L) (e.g., UniProt P23074 or P27401), or bovine foamy/syncytial virus (BFV/BSV) (e.g., UniProt O41894), or a functional fragment or variant thereof (e.g., an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto). In some embodiments, an RT domain is dimeric in its natural functioning. In some embodiments, the RT domain is derived from a virus wherein it functions as a dimer. In embodiments, the RT domain is selected from an RT domain from avian sarcoma/leukemia virus (ASLV) (e.g., UniProt A0A142BKH1), Rous sarcoma virus (RSV) (e.g., UniProt P03354), avian myeloblastosis virus (AMV) (e.g., UniProt Q83133), human immunodeficiency virus type I (HIV-1) (e.g., UniProt P03369), human immunodeficiency virus type II (HIV-2) (e.g., UniProt P15833), simian immunodeficiency virus (SIV) (e.g., UniProt P05896), bovine immunodeficiency virus (BIV) (e.g., UniProt P19560), equine infectious anemia virus (EIAV) (e.g., UniProt P03371), or feline immunodeficiency virus (FIV) (e.g., UniProt P16088) (Herschhorn and Hizi *Cell Mol Life Sci* 67(16):2717-2747 (2010)), or a functional fragment or variant thereof (e.g., an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto). Naturally heterodimeric RT domains may, in some embodiments, also be functional as homodimers. In some embodiments, dimeric RT domains are expressed as fusion proteins, e.g., as homodimeric fusion proteins or heterodimeric fusion proteins. In some embodiments, the RT function of the system is fulfilled by multiple RT domains (e.g., as described herein). In further embodiments, the multiple RT domains are fused or separate, e.g., may be on the same polypeptide or on different polypeptides.

In some embodiments, a gene modifying system described herein comprises an integrase domain, e.g., wherein the integrase domain may be part of the RT domain. In some embodiments, an RT domain (e.g., as described herein) comprises an integrase domain. In some embodiments, an RT domain (e.g., as described herein) lacks an integrase domain, or comprises an integrase domain that has been inactivated by mutation or deleted. In some embodiment, a gene modifying system described herein comprises an RNase H domain, e.g., wherein the RNase H domain may be part of the RT domain. In some embodiments, the RNase H domain is not part of the RT domain and is covalently linked via a flexible linker. In some embodiments, an RT domain (e.g., as described herein) comprises an RNase H domain, e.g., an endogenous RNAse H domain or a heterologous RNase H domain. In some embodiments, an RT domain (e.g., as

described herein) lacks an RNase H domain. In some embodiments, an RT domain (e.g., as described herein) comprises an RNase H domain that has been added, deleted, mutated, or swapped for a heterologous RNase H domain. In some embodiments, the polypeptide comprises an inactivated endogenous RNase H domain. In some embodiments, an endogenous RNase H domain from one of the other domains of the polypeptide is genetically removed such that it is not included in the polypeptide, e.g., the endogenous RNase H domain is partially or completely truncated from the comprising domain. In some embodiments, mutation of an RNase H domain yields a polypeptide exhibiting lower RNase activity, e.g., as determined by the methods described in Kotewicz et al. *Nucleic Acids Res* 16(1):265-277 (1988) (incorporated herein by reference in its entirety), e.g., lower by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% compared to an otherwise similar domain without the mutation. In some embodiments, RNase H activity is abolished.

In some embodiments, an RT domain is mutated to increase fidelity compared to an otherwise similar domain without the mutation. For instance, in some embodiments, a YADD or YMDD motif in an RT domain (e.g., in a reverse transcriptase) is replaced with YVDD. In embodiments, replacement of the YADD or YMDD or YVDD results in higher fidelity in retroviral reverse transcriptase activity (e.g., as described in Jamburuthugoda and Eickbush J Mol Biol 2011; incorporated herein by reference in its entirety).

In some embodiments, a gene modifying polypeptide described herein comprises an RT domain having an amino acid sequence according to Table 6, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto. In some embodiments, a nucleic acid described herein encodes an RT domain having an amino acid sequence according to Table 6, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto.

Table 6: Exemplary reverse transcriptase domains from retroviruses

RT Name	SEQ ID NO:	RT amino acid sequence
AVIRE_P03360	8,001	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPV RKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFD EALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQV REFLGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSK RLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLD TLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIY RERGLLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
AVIRE_P03360_3mut	8,002	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPV RKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFN EALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQV REFLGTIGYCRLWIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSK RLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLD

RT Name	SEQ ID NO:	RT amino acid sequence
		TLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIY RERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
AVIRE_P03360_3mutA	8,003	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPV RKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFN EALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQV REFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKR LDPVAAGWPPCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDT LDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIY RERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
BAEVM_P10272	8,004	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVK KPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQGFKNSPTLFD EALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPRNPRE VREFLGTAGFCRLWIPGFAELAAPLYALTKESTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKGVLTQKLGPWKRPVAYLSKK LDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQPSPHDCR QVLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIALTKALELSKGKKANIYTDSRYAFATAHTH GSIYERRGLLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BAEVM_P10272_3mut	8,005	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVK KPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQGFKNSPTLFN EALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPRNPRE VREFLGTAGFCRLWIPGFAELAAPLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKGVLTQKLGPWKRPVAYLSKK LDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQPSPHDCR QVLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIALTKALELSKGKKANIYTDSRYAFATAHTH GSIYERRGWLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BAEVM_P10272_3mutA	8,006	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVK KPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQGFKNSPTLFN EALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPRNPRE VREFLGKAGFCRLFIPGFAELAAPLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKGVLTQKLGPWKRPVAYLSKKL DPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQPSPHDCRQ VLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIALTKALELSKGKKANIYTDSRYAFATAHTHG SIYERRGWLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BLVAU_P25059	8,007	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRVTNALTKPIPALSPGPPDLTAIPT HLPHIICLDLKDAFFQIPVEDRFRSYFAFTLPTPGGLQPHRRFAWRVLPQGFINSPALFERALQEPLRQVSAAFSQSLLVSYMDDILYVSPTEEQRLQCY QTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQLQTVLGDLQWVSRGTPTTRRPLQLLYSSLKGIDDPRAIIHLSP EQQQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALSSYAKTILKYYHNLPK TSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQFSPEPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAPESAQKGELA GLLAGLAAAPPEPLNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIFVGHVRSHSSASHPIASLNNYVDQL
BLVAU_P25059_2mut	8,008	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRVTNALTKPIPALSPGPPDLTAIPT HLPHIICLDLKDAFFQIPVEDRFRSYFAFTLPTPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYVSPTEEQRLQCY QTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQLQTVLGDLQWVSRGTPTTRRPLQLLYSSLKPIDDPRAIIHLSP EQQQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALSSYAKTILKYYHNLPK TSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQFSPEPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAPESAQKGELA GLLAGLAAAPPEPLNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIFVGHVRSHSSASHPIASLNNYVDQL
BLVJ_P03361	8,009	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDLTAIPT HPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFERALQEPLRQVSAAFSQSLLVSYMDDILYASPTEEQRSQCY QALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDPRAIIQLSPE QLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAKPILKYYHNLPKTS LDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPESAQKGELAGL LAGLAAAPPEPVNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDQL
BLVJ_P03361_2mut	8,010	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDLTAIPT HPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFNRALQEPLRQVSAAFSQSLLVSYMDDILYASPTEEQRSQCY QALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDPRAIIQLSPE QLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAKPILKYYHNLPKTS LDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPESAQKGELAGL LAGLAAAPPEPVNIWVDSKYLYSLLRTWVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDQL
BLVJ_P03361_2mutB	8,011	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDLTAPP THPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYASPTEEQRSQC YQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDPRAIIQLSP EQLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSSYAKPILKYYHNLPKT SLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPESAQKGELAG LLAGLAAAPPEPVNIWVDSKYLYSLLRTWVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLNNYVDQL
FFV_O93209	8,012	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLD LEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPV PKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFTGDVVDL LQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGLLNFARNFIPD FTELIAPLYALIPKSTKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKG LLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKE

RT Name	SEQ ID NO:	RT amino acid sequence
		GHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSV ADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O 93209_ 2mut	8,013	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLD LEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPV PKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFNGDVVDL LQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGLLNFARNFIPD FTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKG LLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKE GHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSV ADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O 93209_ 2mutA	8,014	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLD LEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPV PKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFNGDVVDL LQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGKLNFARNFIPD FTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKG LLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKE GHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSV ADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O 93209- Pro	8,015	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGV LIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGF LNSPGLFTGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQ SILGLLNFARNFIPDFTELIAPLYALIPKSTKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELK FTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHI FYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNR KKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O 93209- Pro_2m ut	8,016	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGV LIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGF LNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQ SILGLLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELK FTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHI FYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNR KKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O 93209- Pro_2m utA	8,017	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGV LIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGF LNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQ SILGKLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELK FTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHI FYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNR KKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FLV_P 10273	8,018	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLP VKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQGFKNSPTL FDEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSR QVREFLGTAGYCRLWIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQKLGPWKRPVAYLSK KLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDC LQILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIALTQALKMAEGKKLTVYTDSRYAFATTHVH GEIYRRRGLLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FLV_P 10273_ 3mut	8,019	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLP VKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQGFKNSPTL FNEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSR QVREFLGTAGYCRLWIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQKLGPWKRPVAYLSK KLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDC LQILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIALTQALKMAEGKKLTVYTDSRYAFATTHVH GEIYRRRGWLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FLV_P 10273_ 3mutA	8,020	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLP VKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQGFKNSPTL FNEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSR QVREFLGKAGYCRLFIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKGVLVQKLGPWKRPVAYLSKK LDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCL QILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIALTQALKMAEGKKLTVYTDSRYAFATTHVHG EIYRRRGWLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FOAM V_P14 350	8,021	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTIL VPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPV YPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADV VDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFAR NFIPNFAELVQPLYNLIASAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKL LTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAI

RT Name	SEQ ID NO:	RT amino acid sequence
		KSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISK WKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350_2mut	8,022	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTIL VPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPV YPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADV VDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFAR NFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKL LTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAI KSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISK WKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350_2mutA	8,023	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTIL VPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPV YPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADV VDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGKLNFAR NFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKL LTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAI KSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISK WKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350-Pro	8,024	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQ GFLNSPALFTADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLK QLQSILGLLNFARNFIPNFAELVQPLYNLIASAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVF SKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPS QYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGF VNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350-Pro_2mut	8,025	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQ GFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDL KQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYV FSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHP SQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNG FVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350-Pro_2mutA	8,026	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQ GFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDL KQLQSILGKLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYV FSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHP SQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNG FVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
GALV_P21414	8,027	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGFKNSP TLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVP TTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLTQTLGPWRRPVAY LSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVH RCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKNINIYTDSRYAFATAHIH GAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
GALV_P21414_3mut	8,028	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGFKNSP TLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVP TTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLTQTLGPWRRPVAY LSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVH RCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKNINIYTDSRYAFATAHIH GAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
GALV_P21414_3mutA	8,029	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQGFKNSP TLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVP TTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVARGVLTQTLGPWRRPVAYL SKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVH RCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKNINIYTDSRYAFATAHIH GAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
HTL1A_P03362	8,030	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTI DLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEATMASLI SHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQSLVQL RQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQTFNQFIQTS DHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLL HGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL

RT Name	SEQ ID NO:	RT amino acid sequence
HTL1A_P03362_2mut	8,031	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTI DLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEATMASLI SHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQL RQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQTFNQFIQTS DHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLL HGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1A_P03362_2mutB	8,032	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSPPTTLAHLQTI DLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEATMASLI SHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQL RQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQTFNQFIQTS DHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLL HGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1C_P14078	8,033	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTI DLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWRVLPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLSEATMASLI SHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPKVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQSLVQL RQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQTFNQFIQTS DHPSVPILLHHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQRAELLGLL HGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1C_P14078_2mut	8,034	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTI DLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWRVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLSEATMASLI SHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPKVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQL RQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISTQTFNQFIQTS DHPSVPILLHHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQRAELLGLL HGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQL
HTL1L_P0C211	8,035	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSLPTTLAHLQTIDLK DAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFEMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEATMASLISH GLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQ QALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISIQTFNQFIQTSD HPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLH GLSSARSWHCLNIFLDSKYLYHYLRTLALGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPIL
HTL1L_P0C211_2mut	8,036	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSLPTTLAHLQTIDLK DAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEATMASLISH GLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQ QALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISIQTFNQFIQTSD HPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLH GLSSARSWHCLNIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPIL
HTL1L_P0C211_2mutB	8,037	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSPPTTLAHLQTIDLK DAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEATMASLISH GLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQ QALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTIHHNISIQTFNQFIQTSD HPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLH GLSSARSWHCLNIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNALTDALLITPIL
HTL32_Q0R5R2	8,038	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSLPQGLPHLRTIDLT DAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFEQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTNALTKEGL PLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKALT LNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQALTYYLHTSDQSSV AILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWDRQVIHQQVLSLPSTCSAQAGELFGLLAGLQK SQPWVALNIFLDSKFLIGHLRRMALGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTL32_Q0R5R2_2mut	8,039	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSLPQGLPHLRTIDLT DAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFEQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTNALTKEGL PLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKALT LNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQALTYYLHTSDQSSV AILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWDRQVIHQQVLSLPSTCSAQAGELFGLLAGLQK SQPWVALNIFLDSKFLIGHLRRMALGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTL32_Q0R5R2_2mutB	8,040	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSPPQGLPHLRTIDL TDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFQQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTNALTKEG LPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKAL TLNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNISNQALTYYLHTSDQSS VAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWDRQVIHQQVLSLPSTCSAQAGELFGLLAGLQ KSQPWVALNIFLDSKFLIGHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTL3P_Q4U0X6	8,041	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSLPQDLPHLRTIDLT DAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFEQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVTNALTKEGL PMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALKTIQKALA LNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQALTYYLHTSDQSSVAIL LQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSKP WPALNIFLDSKFLIGHLRRMALGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL

RT Name	SEQ ID NO:	RT amino acid sequence
HTL3P_Q4U0X6_2mut	8,042	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSLPQDLPHLRTIDLT DAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVTNALTKEG LPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALKTIQKAL ALNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQALTYYLHTSDQSSVAI LLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSK PWPALNIFLDSKFLIGHLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTL3P_Q4U0X6_2mutB	8,043	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSPPQDLPHLRTIDLT DAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVTNALTKEG LPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALKTIQKAL ALNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQALTYYLHTSDQSSVAI LLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSK PWPALNIFLDSKFLIGHLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL
HTLV2_P03363_2mut	8,044	HLPPPPQVDQFPLNLPERLQALNDLVSKALEAGHIEPYSGPGNNPVFPVKKPNGKWRFIHDLRATNAITTTLTSPSPGPPDLTSLPTALPHLQTIDLTDA FFQIPLPKQYQPYFAFTIPQPCNYGPGTRYAWTVLPQGFKNSPTLFQQQLAAVLNPMRKMFPTSTIVQYMDDILLASPTNEELQQLSQLTLQALTTHGL PISQEKTQQTPGQIRFLGQVISPNHITYESTPTIPIKSQWTLTELQVILGEIQWVSKGTPILRKHLQSLYSALHPYRDPRACITLTPQQLHALHAIQQALQH NCRGRLNPALPLLGLISLSTSGTTSVIFQPKQNWPLAWLHTPHPPTSLCPWGHLLACTILTLDKYTLQHYGQLCQSFHHNMSKQALCDFLRNSPHPSV GILIHHMGRFHNLGSQPSGPWKTLLHLPTLLQEPRLLRPIFTLSPVVLDTAPCLFSDGSPQKAAYVLWDQTILQQDITPLPSHETHSAQKGELLALICGLR AAKPWPSLNIFLDSKYLIKYLHSLAIGAFLGTSAHQTLQAALPPLLQGKTIYLHHVRSHTNLPDPISTFNEYTDSLILAPLVPL
JSRV_P31623	8,045	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQLRLGHIEPSTSAWNSPIFVIKKKSGKWRLLQDLRKVNETMMHMGALQPGLPT PSAIPDKSYIIVIDLKDCFYTIPLAPQDCKRFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFPQLYLVHYMDDILLAHTDEHLL YQAFSILKQHLSLNGLVIADEKIQTHFPYNYLGFSLYPRVVYNTQLVKLQTDHLKTLNDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASPRTLSLE GRTALQSIEEAIRQQQITYCDYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYYELVAKIIAKGRHEAIQYFGMEPPFICVPYALEQQDWL FQFSDNWSIAFANYPGQITHHYPSDKLLQFASSHAFIFPKIVRRQPIPEATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQALLTVPTSFNL FTDSSYVVVGALQMIETVPIIGTTSPEVLNLFTLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVLTKQVFFQS
JSRV_P31623_2mutB	8,046	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQLRLGHIEPSTSAWNSPIFVIKKKSGKWRLLQDLRKVNETMMHMGALQPGLPT PSPIPDKSYIIVIDLKDCFYTIPLAPQDCKRFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFPQLYLVHYMDDILLAHTDEHLL YQAFSILKQHLSLNGLVIADEKIQTHFPYNYLGFSLYPRVVYNTQLVKLQTDHLKTLNDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASPRTLSLE GRTALQSIEEAIRQQQITYCDYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYYELVAKIIAKGRHEAIQYFGMEPPFICVPYALEQQDWL FQFSDNWSIAFANYPGQITHHYPSDKLLQFASSHAFIFPKIVRRQPIPEATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQALLTVPTSFNL FTDSSYVVVGALQMIETVPIIGTTSPEVLNLFTLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVLTKQVFFQS
KORV_Q9TTC1	8,047	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLT KLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGI RPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEW RDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYL GYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFAL YVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLN ERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALT QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTET TKN
KORV_Q9TTC1_3mut	8,048	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLT KLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGI RPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEW RDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYL GYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFAL YVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLN ERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALT QALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTE TTKN
KORV_Q9TTC1_3mutA	8,049	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLT KLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGI RPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEW RDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYL GYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALY VDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLNE RVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQ ALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTET TKN
KORV_Q9TTC1-Pro	8,050	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPM SKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQ PLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLC REEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQEAFGRIKEALLSAPALALPD LTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTH YQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQ KAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQ STRILTETTKN
KORV_Q9TTC	8,051	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPM SKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQ

RT Name	SEQ ID NO:	RT amino acid sequence
1-Pro_3mut		PLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLC REEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPD LTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTH YQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQ KAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAA QSTRILTETTKN
KORV_Q9TTC1-Pro_3mutA	8,052	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPM SKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQ PLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLC REEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPDL TKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHY QSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQK AELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQ STRILTETTKN
MLVAV_P03356	8,053	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSP TLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEG APHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAF ATAHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVAV_P03356_3mut	8,054	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSP TLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPV AYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEE GAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYA FATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVAV_P03356_3mutA	8,055	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLL PVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSP TLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEG APHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAF ATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7	8,056	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSPT LFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAP HDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFAT AHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7	8,057	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSPT LFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAP HDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFAT AHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7_3mut	8,058	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGA PHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7_3mut	8,059	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGA PHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7_3	8,060	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSPTL FNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKTP RQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYL

RT Name	SEQ ID NO:	RT amino acid sequence
mutA_WS		SKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAP HDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
MLVBM_Q7SVK7_3mutA_WS	8,061	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPV KKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQGFKNSPTL FNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKTP RQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYL SKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAP HDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
MLVCB_P08361	8,062	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFDEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQ HDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVCB_P08361_3mut	8,063	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGL QHDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAF ATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVCB_P08361_3mutA	8,064	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKT PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQ HDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810	8,065	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQGFKNSPT LFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGLCRLWIPGFAEMAAPLYPLTKTGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQ HDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAAGKKLNVYTDSRYAFAT AHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810_3mut	8,066	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQ HDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAAGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810_3mutA	8,067	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGKAGLCRLFIPGFAEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQ HDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAAGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVFF_P26809_3mut	8,068	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQ HDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNRAEARGNRMADQAAREVATRETPETSTLL
MLVFF_P26809_3mutA	8,069	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQ HDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNRAEARGNRMADQAAREVATRETPETSTLL

RT Name	SEQ ID NO:	RT amino acid sequence
MLVM S_P03 355	8,070	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGL QHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_refer ence	8,137	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQ HNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSP
MLVM S_P03 355	8,071	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGL QHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P03 355_3 mut	8,072	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGL QHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P03 355_3 mut	8,073	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGL QHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFA TAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P03 355_3 mutA_ WS	8,074	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQ HNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P03 355_PL V919	8,076	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQ HNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEF E
MLVM S_P03 355_PL V919	8,077	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQ HNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEF E

RT Name	SEQ ID NO:	RT amino acid sequence
MLVRD_P11227	8,078	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQGLREVNKRVEDIHPTVPNPYNLLSGLPTSHRWYTVLDLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLPQGFKNSPT LFDEALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPRFAEMAAPLYPLTKTGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAP HDCLEILAETHGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFATA HIHGEIYKRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVRD_P11227_3mut	8,079	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQGLREVNKRVEDIHPTVPNPYNLLSGLPTSHRWYTVLDLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLPQGFKNSPT LFNEALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPTPKT PRQLREFLGTAGFCRLWIPRFAEMAAPLYPLTKPGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAY LSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAP HDCLEILAETHGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKRLNVYTDSRYAFATA HIHGEIYKRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MMTVB_P03365	8,080	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAV NATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDS YIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLF EILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSK DPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQA EIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365	8,081	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAV NATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDS YIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLF EILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSK DPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQA EIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365_2mut	8,082	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAV NATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDS YIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLF EILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSK DPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQA EIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365_2mut_WS	8,083	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDI KVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNA TMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIV HYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEIL NPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDP DYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIV AVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTVB_P03365_2mut_WS	8,084	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDI KVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNA TMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIV HYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEIL NPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDP DYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIV AVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTVB_P03365_2mutB	8,085	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAV NATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDS YIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLF EILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSK DPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQA EIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365_2mutB	8,086	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMK DIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAV NATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDS YIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLF EILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSK DPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQA EIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365_2	8,087	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDI KVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNA TMHDMGALQPGLPSPPAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIV HYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEIL

RT Name	SEQ ID NO:	RT amino acid sequence
mutB_WS		NPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDP DYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIV AVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P03365_2mutB_WS	8,088	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDI KVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNA TMHDMGALQPGLPSPPAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIV HYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEIL NPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDP DYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIV AVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P03365_WS	8,089	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDI KVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNA TMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIV HYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEIL NGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDP DYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIV AVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P03365_WS	8,090	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDI KVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNA TMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIV HYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEIL NGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDP DYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIV AVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P03365-Pro	8,091	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVR DKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTT GELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHR SKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQ NTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro	8,092	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVR DKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTT GELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHR SKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQ NTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro_2mut	8,093	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVR DKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTT GELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHR SKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQ NTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro_2mut	8,094	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVR DKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTT GELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHR SKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQ NTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro_2mutB	8,095	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVR DKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTT GELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHR SKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQ NTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro_2mutB	8,096	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVR DKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTT GELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHR SKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQ NTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MPMV_P07572	8,097	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLP SPVAIPQGYLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDILIAGKDGQ QVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKGDSDPNSHR SLSKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILGRDHSKKYFGIEPSTIIQPYSKSQIDW LMQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFVFPQIISKTPLNNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPL NIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAQGNQRADLATKIVASNINT

RT Name	SEQ ID NO:	RT amino acid sequence
MPMV_P07572_2mutB	8,098	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLP SPVAPPQGYLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDILIAGKDGQ QVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKPDSDPNSHRS LSKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILGRDHSKKYFGIEPSTIIQPYSKSQIDWL MQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFVFPQIISKTPLNNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPL NIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAQGNQRADLATKIVASNINT
PERV_Q4VFZ2	8,099	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLL PVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNS PTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPT TAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVA YLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTH DCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAH VHGAIYKQRGLLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_Q4VFZ2	8,100	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLL PVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNS PTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPT TAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVA YLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTH DCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAH VHGAIYKQRGLLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_Q4VFZ2_3mut	8,101	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLL PVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNS PTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPT TAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVA YLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTH DCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAH VHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_Q4VFZ2_3mut	8,102	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLL PVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNS PTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPT TAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVA YLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTH DCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAH VHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_Q4VFZ2_3mutA_WS	8,103	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVR KPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIF NEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAK QVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQ LLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAI YKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLP
PERV_Q4VFZ2_3mutA_WS	8,104	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVR KPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIF NEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAK QVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSK KLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQ LLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAI YKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLP
SFV1_P23074	8,105	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQL TVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNT PVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTAD VVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFAR NFIPNYSELVKPLYTIVANANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLL TTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIK HPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKW KSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074_2mut	8,106	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQL TVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNT PVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNAD VVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFAR NFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLT TMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKH PDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWK SIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074_2mutA	8,107	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQL TVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNT PVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNAD

RT Name	SEQ ID NO:	RT amino acid sequence
		VVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGKLNFAR NFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLT TMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKH PDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWK SIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074-Pro 8,108		VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQ GFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQ LQSILGLLNFARNFIPNYSELVKPLYTIVANANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKA EAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFA MVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNK KKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074-Pro_2mut 8,109		VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQ GFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLK QLQSILGLLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSK AEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEF AMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNN KKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074-Pro_2mutA 8,110		VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQ GFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLK QLQSILGKLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSK AEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEF AMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNN KKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV3L_P27401 8,111		MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTT LVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTP VYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPALFTADV VDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGLLNFAR NFIPNFSELVKPLYNIIATANGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLL TTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHP NVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWK SIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401_2mut 8,112		MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTT LVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTP VYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPALFNADV VDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGLLNFAR NFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLL TTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHP NVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWK SIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401_2mutA 8,113		MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTT LVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTP VYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPALFNADV VDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGKLNFA RNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKL LTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKH PNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKW KSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401-Pro 8,114		IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQ GVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQ GFLNSPALFTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDL KQLQSILGLLNFARNFIPNFSELVKPLYNIIATANGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVY TKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEF SMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFN NKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401-Pro_2mut 8,115		IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQ GVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQ GFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDL KQLQSILGLLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVY TKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEF SMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFN NKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401-	8,116	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQ GVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQ GFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDL

RT Name	SEQ ID NO:	RT amino acid sequence
Pro_2mutA		KQLQSILGKLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVY TKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEF SMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFN NKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFVCP_Q87040	8,117	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTI LVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTP VYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSPALFTAD AVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQSILGLLNF ARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLE KLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSA IKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISK WKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040_2mut	8,118	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTI LVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTP VYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSPALFNAD AVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQSILGLLNF ARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLE KLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSA IKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISK WKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040_2mutA	8,119	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTI LVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTP VYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSPALFNAD AVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQSILGKLNF ARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLE KLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSA IKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISK WKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro	8,120	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQ GFLNSPALFTADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDL KQLQSILGLLNFARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYV FSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPS QYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGF VNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro_2mut	8,121	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQ GFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDL KQLQSILGLLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYV FSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPS QYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGF VNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro_2mutA	8,122	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQ GFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDL KQLQSILGKLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYV FSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPS QYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGF VNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SMRVH_P03364	8,123	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPV AIPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLACDSAEAAK ACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKGDPNPLSVRALTPE AKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAMLIIKGRYTGRQLFGRDPHSIIIPY TQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTV LAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAEGNALADAATQIFPIISD
SMRVH_P03364_2mut	8,124	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPV AIPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLACDSAEAAK ACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKPDPNPLSVRALTPE AKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAMLIIKGRYTGRQLFGRDPHSIIIPY TQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTV LAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAEGNALADAATQIFPIISD
SMRVH_P03364_2mutB	8,125	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPV APPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDILLACDSAEAAK ACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKPDPNPLSVRALTPE AKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAMLIIKGRYTGRQLFGRDPHSIIIPY TQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTV LAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAEGNALADAATQIFPIISD

RT Name	SEQ ID NO:	RT amino acid sequence
SRV2_P51517	8,126	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLP SPVAIPQGYFKIVIDLKDCFFTIPLQPVDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYMDDILIAGKLGE QVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRDKLQTLNDFQKLLGDINWLRPYLHLTTGDLKPLFDILKGDSNPNSPRS LSEAALASLQKVETAIAEQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILGRDNSKKYFGLEPSTIIQPYSKSQIH WLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAVVFPRIISKTPLDNALLVFTDGSSTGIAAYTFEKTTVRFKTSHTSAQLVELQALIAVLSAFPHR ALNVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGPLSQGNHITDLATKVVATTLTT
SRV2_P51517_2mutB	8,127	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLP SPVAPPQGYFKIVIDLKDCFFTIPLQPVDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYMDDILIAGKLGE QVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRDKLQTLNDFQKLLGDINWLRPYLHLTTGDLKPLFDILKGDSNPNSPRS LSEAALASLQKVETAIAEQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILGRDNSKKYFGLEPSTIIQPYSKSQIH WLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAVVFPRIISKTPLDNALLVFTDGSSTGIAAYTFEKTTVRFKTSHTSAQLVELQALIAVLSAFPHR ALNVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGPLSQGNHITDLATKVVATTLTT
WDSV_O92815	8,128	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHD LRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFSQALYQSLHKIKFKISSEICIYMD DVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQIRAFLGLVGYCRHWIPEFSIHSKFL EKQLKKDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFDAIESGLPPCLKACASIHRSLTQA DSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSD GSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQ IMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WDSV_O92815_2mut	8,129	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHD LRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQSLHKIKFKISSEICIYMD DVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQIRAFLGLVGYCRHWIPEFSIHSKFL EKQLKPDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFDAIESGLPPCLKACASIHRSLTQA DSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSD GSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQ IMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WDSV_O92815_2mutA	8,130	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHD LRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQSLHKIKFKISSEICIYMD DVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQIRAFLGKVGYCRHFIPEFSIHSKFL EKQLKPDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFDAIESGLPPCLKACASIHRSLTQA DSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSD GSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQ IMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WMSV_P03359	8,131	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSP TLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVMKIPPP TTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAY LSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVH RCSEILAEETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKDINIYTDSRYAFATAHI HGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
WMSV_P03359_3mut	8,132	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSP TLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVMKIPPP TTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAY LSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVH RCSEILAEETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKDINIYTDSRYAFATAHI HGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
WMSV_P03359_3mutA	8,133	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLL PVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSP TLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVMKIPPP TTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAY LSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVH RCSEILAEETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKDINIYTDSRYAFATAHI HGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
XMRV6_A1Z651	8,134	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEA PHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHVHGEIYRRRGLLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREAAMKAVLETSTLL
XMRV6_A1Z651_3mut	8,135	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPV AYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKE

RT Name	SEQ ID NO:	RT amino acid sequence
		APHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAF ATAHVHGEIYRRRGWLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREAAMKAVLETSTLL
XMRV6_A1Z651_3mutA	8,136	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPT LFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEA PHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHVHGEIYRRRGWLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREAAMKAVLETSTLL

In some embodiments, reverse transcriptase domains are modified, for example by site-specific mutation. In some embodiments, reverse transcriptase domains are engineered to have improved properties, e.g. SuperScript IV (SSIV) reverse transcriptase derived from the MMLV RT. In some embodiments, the reverse transcriptase domain may be engineered to have lower error rates, e.g., as described in WO2001068895, incorporated herein by reference. In some embodiments, the reverse transcriptase domain may be engineered to be more thermostable. In some embodiments, the reverse transcriptase domain may be engineered to be more processive. In some embodiments, the reverse transcriptase domain may be engineered to have tolerance to inhibitors. In some embodiments, the reverse transcriptase domain may be engineered to be faster. In some embodiments, the reverse transcriptase domain may be engineered to better tolerate modified nucleotides in the RNA template. In some embodiments, the reverse transcriptase domain may be engineered to insert modified DNA nucleotides. In some embodiments, the reverse transcriptase domain is engineered to bind a template RNA. In some embodiments, one or more mutations are chosen from D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K, H8Y, T306K, or D653N in the RT domain of murine leukemia virus reverse transcriptase or a corresponding mutation at a corresponding position of another RT domain.

In some embodiments, a gene modifying polypeptide comprises the RT domain from a retroviral reverse transcriptase, e.g., a wild-type M-MLV RT, e.g., comprising the following sequence:

M-MLV (WT):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK

RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY
AFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 5002)

In some embodiments, a gene modifying polypeptide comprises the RT domain from a retroviral reverse transcriptase, e.g., an M-MLV RT, e.g., comprising the following sequence:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG
TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY
AFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLL (SEQ ID NO: 5003)

In some embodiments, a gene modifying polypeptide comprises the RT domain from a retroviral reverse transcriptase comprising the sequence of amino acids 659-1329 of NP_057933. In embodiments, the gene modifying polypeptide further comprises one additional amino acid at the N-terminus of the sequence of amino acids 659-1329 of NP_057933, e.g., as shown below:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP**VKKPGTNDYRPVQDLREVN**
KRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP
EMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAAT
SELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKE
TVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAY
QEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPV
AAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTH
YQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDAD
HTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK
KLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPG
HQKGHSAEARGNRMADQAARKAA (SEQ ID NO: 5004)

Core RT (bold), annotated per above
RNAseH (underlined), annotated per above

In embodiments, the gene modifying polypeptide further comprises one additional amino acid at the C-terminus of the sequence of amino acids 659-1329 of NP_057933. In embodiments, the gene modifying polypeptide comprises an RNaseH1 domain (e.g., amino acids 1178-1318 of NP_057933).

In some embodiments, a retroviral reverse transcriptase domain, e.g., M-MLV RT, may comprise one or more mutations from a wild-type sequence that may improve features of the RT, e.g., thermostability, processivity, and/or template binding. In some embodiments, an M-MLV RT domain comprises, relative to the M-MLV (WT) sequence above, one or more mutations, e.g., selected from D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, K103L, e.g., a combination of mutations, such as D200N, L603W, and T330P, optionally further including T306K and W313F. In some embodiments, an M-MLV RT used herein comprises the mutations D200N, L603W, T330P, T306K and W313F. In embodiments, the mutant M-MLV RT comprises the following amino acid sequence:

M-MLV (PE2):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI
KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGIS
GQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQG

TRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKT
PRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAP
ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRM
VAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDR
VQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL
LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRY
AFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR
GNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 5005)

In some embodiments, a writing domain (e.g., RT domain) comprises an RNA-binding domain, e.g., that specifically binds to an RNA sequence. In some embodiments, a template RNA comprises an RNA sequence that is specifically bound by the RNA-binding domain of the writing domain.

In some embodiments, the reverse transcription domain only recognizes and reverse transcribes a specific template, e.g., a template RNA of the system. In some embodiments, the template comprises a sequence or structure that enables recognition and reverse transcription by a reverse transcription domain. In some embodiments, the template comprises a sequence or structure that enables association with an RNA-binding domain of a polypeptide component of a genome engineering system described herein. In some embodiments, the genome engineering system reverse preferably transcribes a template comprising an association sequence over a template lacking an association sequence.

The writing domain may also comprise DNA-dependent DNA polymerase activity, e.g., comprise enzymatic activity capable of writing DNA into the genome from a template DNA sequence. In some embodiments, DNA-dependent DNA polymerization is employed to complete second-strand synthesis of a target site edit. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a DNA polymerase domain in the polypeptide. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a reverse transcriptase domain that is also capable of DNA-dependent DNA polymerization, e.g., second-strand synthesis. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a second polypeptide of the system. In some embodiments, the DNA-

dependent DNA polymerase activity is provided by an endogenous host cell polymerase that is optionally recruited to the target site by a component of the genome engineering system.

In some embodiments, the reverse transcriptase domain has a lower probability of premature termination rate (P_{off}) *in vitro* relative to a reference reverse transcriptase domain. In some embodiments, the reference reverse transcriptase domain is a viral reverse transcriptase domain, e.g., the RT domain from M-MLV.

In some embodiments, the reverse transcriptase domain has a lower probability of premature termination rate (P_{off}) *in vitro* of less than about 5 x 10^{-3}/nt, 5 x 10^{-4}/nt, or 5 x 10^{-6}/nt, e.g., as measured on a 1094 nt RNA. In embodiments, the *in vitro* premature termination rate is determined as described in Bibillo and Eickbush (2002) J Biol Chem 277(38):34836-34845 (incorporated by reference herein its entirety).

In some embodiments, the reverse transcriptase domain is able to complete at least about 30% or 50% of integrations in cells. The percent of complete integrations can be measured by dividing the number of substantially full-length integration events (e.g., genomic sites that comprise at least 98% of the expected integrated sequence) by the number of total (including substantially full-length and partial) integration events in a population of cells. In embodiments, the integrations in cells is determined (e.g., across the integration site) using long-read amplicon sequencing, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety).

In embodiments, quantifying integrations in cells comprises counting the fraction of integrations that contain at least about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the DNA sequence corresponding to the template RNA (e.g., a template RNA having a length of at least 0.05, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 3, 4, or 5 kb, e.g., a length between 0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-0.9, 1.0-1.2, 1.2-1.4, 1.4-1.6, 1.6-1.8, 1.8-2.0, 2-3, 3-4, or 4-5 kb).

In some embodiments, the reverse transcriptase domain is capable of polymerizing dNTPs *in vitro*. In embodiments, the reverse transcriptase domain is capable of polymerizing dNTPs *in vitro* at a rate between 0.1 – 50 nt/sec (e.g., between 0.1-1, 1-10, or 10-50 nt/sec). In embodiments, polymerization of dNTPs by the reverse transcriptase domain is measured by a single-molecule assay, e.g., as described in Schwartz and Quake (2009) *PNAS* 106(48):20294-20299 (incorporated by reference in its entirety).

In some embodiments, the reverse transcriptase domain has an *in vitro* error rate (e.g., misincorporation of nucleotides) of between $1 \times 10^{-3} - 1 \times 10^{-4}$ or $1 \times 10^{-4} - 1 \times 10^{-5}$ substitutions/nt , e.g., as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (incorporated herein by reference in its entirety). In some embodiments, the reverse transcriptase domain has an error rate (e.g., misincorporation of nucleotides) in cells (e.g., HEK293T cells) of between $1 \times 10^{-3} - 1 \times 10^{-4}$ or $1 \times 10^{-4} - 1 \times 10^{-5}$ substitutions/nt, e.g., by long-read amplicon sequencing, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety).

In some embodiments, the reverse transcriptase domain is capable of performing reverse transcription of a target RNA *in vitro*. In some embodiments, the reverse transcriptase requires a primer of at least 3 nucleotides to initiate reverse transcription of a template. In some embodiments, reverse transcription of the target RNA is determined by detection of cDNA from the target RNA (e.g., when provided with a ssDNA primer, e.g., which anneals to the target with at least 3, 4, 5, 6, 7, 8, 9, or 10 nt at the 3′ end), e.g., as described in Bibillo and Eickbush (2002) *J Biol Chem* 277(38):34836-34845 (incorporated herein by reference in its entirety).

In some embodiments, the reverse transcriptase domain performs reverse transcription at least 5 or 10 times more efficiently (e.g., by cDNA production), e.g., when converting its RNA template to cDNA, for example, as compared to an RNA template lacking the protein binding motif (e.g., a 3′ UTR). In embodiments, efficiency of reverse transcription is measured as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (incorporated by reference herein in its entirety).

In some embodiments, the reverse transcriptase domain specifically binds a specific RNA template with higher frequency (e.g., about 5 or 10-fold higher frequency) than any endogenous cellular RNA, e.g., when expressed in cells (e.g., HEK293T cells). In embodiments, frequency of specific binding between the reverse transcriptase domain and the template RNA are measured by CLIP-seq, e.g., as described in Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (incorporated herein by reference in its entirety).

In some embodiments, an RT domain (e.g., as listed in Table 6) comprises one or more mutations as listed in Table 2 below. In some embodiment, an RT domain as listed in Table 6 comprises one, two, three, four, five, or six of the mutations listed in the corresponding row of Table 2 below.

Table 2. Exemplary RT domain mutations (relative to corresponding wild-type sequences as listed in the corresponding row of Table 6)

RT Domain Name	Mutation(s)					
AVIRE_P03360						
AVIRE_P03360_3mut	D200N	G330P	L605W			
AVIRE_P03360_3mutA	D200N	G330P	L605W	T306K	W313F	
BAEVM_P10272						
BAEVM_P10272_3mut	D198N	E328P	L602W			
BAEVM_P10272_3mutA	D198N	E328P	L602W	T304K	W311F	
BLVAU_P25059						
BLVAU_P25059_2mut	E159Q	G286P				
BLVJ_P03361						
BLVJ_P03361_2mut	E159Q	L524W				
BLVJ_P03361_2mutB	E159Q	L524W	I97P			
FFV_O93209	D21N					
FFV_O93209_2mut	D21N	T293N	T419P			
FFV_O93209_2mutA	D21N	T293N	T419P	L393K		
FFV_O93209-Pro						
FFV_O93209-Pro_2mut	T207N	T333P				
FFV_O93209-Pro_2mutA	T207N	T333P	L307K			
FLV_P10273						
FLV_P10273_3mut	D199N	L602W				
FLV_P10273_3mutA	D199N	L602W	T305K	W312F		
FOAMV_P14350	D24N					
FOAMV_P14350_2mut	D24N	T296N	S420P			
FOAMV_P14350_2mutA	D24N	T296N	S420P	L396K		
FOAMV_P14350-Pro						
FOAMV_P14350-Pro_2mut	T207N	S331P				
FOAMV_P14350-Pro_2mutA	T207N	S331P	L307K			
GALV_P21414						
GALV_P21414_3mut	D198N	E328P	L600W			
GALV_P21414_3mutA	D198N	E328P	L600W	T304K	W311F	
HTL1A_P03362						
HTL1A_P03362_2mut	E152Q	R279P				
HTL1A_P03362_2mutB	E152Q	R279P	L90P			
HTL1C_P14078						
HTL1C_P14078_2mut	E152Q	R279P				
HTL1L_P0C211						

HTL1L_P0C211_2mut	E149Q	L527W				
HTL1L_P0C211_2mutB	E149Q	L527W	L87P			
HTL32_Q0R5R2						
HTL32_Q0R5R2_2mut	E149Q	L526W				
HTL32_Q0R5R2_2mutB	E149Q	L526W	L87P			
HTL3P_Q4U0X6						
HTL3P_Q4U0X6_2mut	E149Q	L526W				
HTL3P_Q4U0X6_2mutB	E149Q	L526W	L87P			
HTLV2_P03363_2mut	E147Q	G274P				
JSRV_P31623						
JSRV_P31623_2mutB	A100P					
KORV_Q9TTC1	D32N					
KORV_Q9TTC1_3mut	D32N	D322N	E452P	L724W		
KORV_Q9TTC1_3mutA	D32N	D322N	E452P	L724W	T428K	W435F
KORV_Q9TTC1-Pro						
KORV_Q9TTC1-Pro_3mut	D231N	E361P	L633W			
KORV_Q9TTC1-Pro_3mutA	D231N	E361P	L633W	T337K	W344F	
MLVAV_P03356						
MLVAV_P03356_3mut	D200N	T330P	L603W			
MLVAV_P03356_3mutA	D200N	T330P	L603W	T306K	W313F	
MLVBM_Q7SVK7						
MLVBM_Q7SVK7						
MLVBM_Q7SVK7_3mut	D200N	T330P	L603W			
MLVBM_Q7SVK7_3mut	D200N	T330P	L603W			
MLVBM_Q7SVK7_3mutA_WS	D199N	T329P	L602W	T305K	W312F	
MLVBM_Q7SVK7_3mutA_WS	D199N	T329P	L602W	T305K	W312F	
MLVCB_P08361						
MLVCB_P08361_3mut	D200N	T330P	L603W			
MLVCB_P08361_3mutA	D200N	T330P	L603W	T306K	W313F	
MLVF5_P26810						
MLVF5_P26810_3mut	D200N	T330P	L603W			
MLVF5_P26810_3mutA	D200N	T330P	L603W	T306K	W313F	
MLVFF_P26809_3mut	D200N	T330P	L603W			
MLVFF_P26809_3mutA	D200N	T330P	L603W	T306K	W313F	
MLVMS_P03355						
MLVMS_P03355						
MLVMS_P03355_3mut	D200N	T330P	L603W			
MLVMS_P03355_3mut	D200N	T330P	L603W			
MLVMS_P03355_3mutA_WS	D200N	T330P	L603W	T306K	W313F	

MLVMS_P03355_3mutA_WS	D200N	T330P	L603W	T306K	W313F	
MLVMS_P03355_PLV919	D200N	T330P	L603W	T306K	W313F	H8Y
MLVMS_P03355_PLV919	D200N	T330P	L603W	T306K	W313F	H8Y
MLVRD_P11227						
MLVRD_P11227_3mut	D200N	T330P	L603W			
MMTVB_P03365	D26N					
MMTVB_P03365	D26N					
MMTVB_P03365_2mut	D26N	G401P				
MMTVB_P03365_2mut_WS	G400P					
MMTVB_P03365_2mut_WS	G400P					
MMTVB_P03365_2mutB	D26N	G401P	V215P			
MMTVB_P03365_2mutB	D26N	G401P	V215P			
MMTVB_P03365_2mutB_WS	G400P	V212P				
MMTVB_P03365_2mutB_WS	G400P	V212P				
MMTVB_P03365_WS						
MMTVB_P03365_WS						
MMTVB_P03365-Pro						
MMTVB_P03365-Pro						
MMTVB_P03365-Pro_2mut	G309P					
MMTVB_P03365-Pro_2mut	G309P					
MMTVB_P03365-Pro_2mutB	G309P	V123P				
MMTVB_P03365-Pro_2mutB	G309P	V123P				
MPMV_P07572						
MPMV_P07572_2mutB	G289P	I103P				
PERV_Q4VFZ2						
PERV_Q4VFZ2						
PERV_Q4VFZ2_3mut	D199N	E329P	L602W			
PERV_Q4VFZ2_3mut	D199N	E329P	L602W			
PERV_Q4VFZ2_3mutA_WS	D196N	E326P	L599W	T302K	W309F	
PERV_Q4VFZ2_3mutA_WS	D196N	E326P	L599W	T302K	W309F	
SFV1_P23074	D24N					
SFV1_P23074_2mut	D24N	T296N	N420P			
SFV1_P23074_2mutA	D24N	T296N	N420P	L396K		
SFV1_P23074-Pro						
SFV1_P23074-Pro_2mut	T207N	N331P				
SFV1_P23074-Pro_2mutA	T207N	N331P	L307K			
SFV3L_P27401	D24N					
SFV3L_P27401_2mut	D24N	T296N	N422P			
SFV3L_P27401_2mutA	D24N	T296N	N422P	L396K		

SFV3L_P27401-Pro						
SFV3L_P27401-Pro_2mut	T307N	N333P				
SFV3L_P27401-Pro_2mutA	T307N	N333P	L307K			
SFVCP_Q87040	D24N					
SFVCP_Q87040_2mut	D24N	T296N	K422P			
SFVCP_Q87040_2mutA	D24N	T296N	K422P	L396K		
SFVCP_Q87040-Pro						
SFVCP_Q87040-Pro_2mut	T207N	K333P				
SFVCP_Q87040-Pro_2mutA	T207N	K333P	L307K			
SMRVH_P03364						
SMRVH_P03364_2mut	G288P					
SMRVH_P03364_2mutB	G288P	I102P				
SRV2_P51517						
SRV2_P51517_2mutB	I103P					
WDSV_O92815						
WDSV_O92815_2mut	S183N	K312P				
WDSV_O92815_2mutA	S183N	K312P	L288K	W295F		
WMSV_P03359						
WMSV_P03359_3mut	D198N	E328P	L600W			
WMSV_P03359_3mutA	D198N	E328P	L600W	T304K	W311F	
XMRV6_A1Z651						
XMRV6_A1Z651_3mut	D200N	T330P	L603W			
XMRV6_A1Z651_3mutA	D200N	T330P	L603W	T306K	W313F	

Template nucleic acid binding domain

The gene modifying polypeptide typically contains regions capable of associating with the template nucleic acid (e.g., template RNA). In some embodiments, the template nucleic acid binding domain is an RNA binding domain. In some embodiments, the RNA binding domain is a modular domain that can associate with RNA molecules containing specific signatures, e.g., structural motifs. In other embodiments, the template nucleic acid binding domain (e.g., RNA binding domain) is contained within the reverse transcription domain, e.g., the reverse transcriptase-derived component has a known signature for RNA preference.

In other embodiments, the template nucleic acid binding domain (e.g., RNA binding domain) is contained within the target DNA binding domain. For example, in some embodiments, the DNA binding domain is a CRISPR-associated protein that recognizes the

structure of a template nucleic acid (e.g., template RNA) comprising a gRNA. In some embodiments, a gene modifying polypeptide comprises a DNA-binding domain comprising a CRISPR-associated protein that associates with a gRNA scaffold that allows the DNA-binding domain to bind a target genomic DNA sequence. In some embodiments, the gRNA scaffold and gRNA spacer is comprised within the template nucleic acid (e.g., template RNA), thus the DNA-binding domain is also the template nucleic acid binding domain. In some embodiments, the polypeptide possesses RNA binding function in multiple domains, e.g., can bind a gRNA structure in a CRISPR-associated DNA binding domain and an additional sequence or structure in a reverse transcriptase domain.

In some embodiments, the RNA binding domain is capable of binding to a template RNA with greater affinity than a reference RNA binding domain. In some embodiments, the reference RNA binding domain is an RNA binding domain from Cas9 of S. pyogenes. In some embodiments, the RNA binding domain is capable of binding to a template RNA with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM). In some embodiments, the affinity of a RNA binding domain for its template RNA is measured *in vitro*, e.g., by thermophoresis, e.g., as described in Asmari et al. Methods 146:107-119 (2018) (incorporated by reference herein in its entirety). In some embodiments, the affinity of a RNA binding domain for its template RNA is measured in cells (e.g., by FRET or CLIP-Seq).

In some embodiments, the RNA binding domain is associated with the template RNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled RNA. In some embodiments, the frequency of association between the RNA binding domain and the template RNA or scrambled RNA is measured by CLIP-seq, e.g., as described in Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (incorporated by reference herein in its entirety). In some embodiments, the RNA binding domain is associated with the template RNA in cells (e.g., in HEK293T cells) at a frequency at least about 5-fold or 10-fold higher than with a scrambled RNA. In some embodiments, the frequency of association between the RNA binding domain and the template RNA or scrambled RNA is measured by CLIP-seq, e.g., as described in Lin and Miles (2019), *supra.*

Endonuclease domains and DNA binding domains

In some embodiments, a gene modifying polypeptide possesses the function of DNA target site cleavage via an endonuclease domain. In some embodiments, a gene modifying polypeptide comprises a DNA binding domain, e.g., for binding to a target nucleic acid. In some embodiments, a domain (e.g., a Cas domain) of the gene modifying polypeptide comprises two or more smaller domains, e.g., a DNA binding domain and an endonuclease domain. It is understood that when a DNA binding domain (e.g., a Cas domain) is said to bind to a target nucleic acid sequence, in some embodiments, the binding is mediated by a gRNA.

In some embodiments, a domain has two functions. For example, in some embodiments, the endonuclease domain is also a DNA-binding domain. In some embodiments, the endonuclease domain is also a template nucleic acid (e.g., template RNA) binding domain. For example, in some embodiments, a polypeptide comprises a CRISPR-associated endonuclease domain that binds a template RNA comprising a gRNA, binds a target DNA sequence (e.g., with complementarity to a portion of the gRNA), and cuts the target DNA sequence. In some embodiments, an endonuclease domain or endonuclease/DNA-binding domain from a heterologous source can be used or can be modified (e.g., by insertion, deletion, or substitution of one or more residues) in a gene modifying system described herein.

In some embodiments, a nucleic acid encoding the endonuclease domain or endonuclease/DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments, the endonuclease element is a heterologous endonuclease element, such as a Cas endonuclease (e.g., Cas9), a type-II restriction endonuclease (e.g., Fok1), a meganuclease (e.g., I-SceI), or other endonuclease domain.

In certain aspects, the DNA-binding domain of a gene modifying polypeptide described herein is selected, designed, or constructed for binding to a desired host DNA target sequence. In certain embodiments, the DNA-binding domain of the polypeptide is a heterologous DNA-binding element. In some embodiments the heterologous DNA binding element is a zinc-finger element or a TAL effector element, e.g., a zinc-finger or TAL polypeptide or functional fragment thereof. In some embodiments the heterologous DNA binding element is a sequence-guided DNA binding element, such as Cas9, Cpf1, or other CRISPR-related protein that has been altered to have no endonuclease activity. In some embodiments the heterologous DNA binding element retains endonuclease activity. In some embodiments, the heterologous DNA binding element retains partial endonuclease activity to cleave ssDNA, e.g., possesses nickase activity. In

specific embodiments, the heterologous DNA-binding domain can be any one or more of Cas9, TAL domain, ZF domain, Myb domain, combinations thereof, or multiples thereof.

In some embodiments, DNA-binding domains are modified, for example by site-specific mutation, increasing or decreasing DNA-binding elements (for example, number and/or specificity of zinc fingers), etc., to alter DNA-binding specificity and affinity. In some embodiments a nucleic acid sequence encoding the DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In embodiments, the DNA binding domain comprises one or more modifications relative to a wild-type DNA binding domain, e.g., a modification via directed evolution, e.g., phage-assisted continuous evolution (PACE).

In some embodiments, the DNA binding domain comprises a meganuclease domain (e.g., as described herein, e.g., in the endonuclease domain section), or a functional fragment thereof. In some embodiments, the meganuclease domain possesses endonuclease activity, e.g., double-strand cleavage and/or nickase activity. In other embodiments, the meganuclease domain has reduced activity, e.g., lacks endonuclease activity, e.g., the meganuclease is catalytically inactive. In some embodiments, a catalytically inactive meganuclease is used as a DNA binding domain, e.g., as described in Fonfara et al. Nucleic Acids Res 40(2):847-860 (2012), incorporated herein by reference in its entirety.

In some embodiments, a gene modifying polypeptide comprises a modification to a DNA-binding domain, e.g., relative to the wild-type polypeptide. In some embodiments, the DNA-binding domain comprises an addition, deletion, replacement, or modification to the amino acid sequence of the original DNA-binding domain. In some embodiments, the DNA-binding domain is modified to include a heterologous functional domain that binds specifically to a target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the functional domain replaces at least a portion (e.g., the entirety of) the prior DNA-binding domain of the polypeptide. In some embodiments, the functional domain comprises a zinc finger (e.g., a zinc finger that specifically binds to the target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the functional domain comprises a Cas domain (e.g., a Cas domain that specifically binds to the target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the Cas domain comprises a Cas9 or a mutant or variant thereof (e.g., as described herein). In embodiments, the Cas domain is associated with a guide RNA (gRNA), e.g., as

described herein. In embodiments, the Cas domain is directed to a target nucleic acid (e.g., DNA) sequence of interest by the gRNA. In embodiments, the Cas domain is encoded in the same nucleic acid (e.g., RNA) molecule as the gRNA. In embodiments, the Cas domain is encoded in a different nucleic acid (e.g., RNA) molecule from the gRNA.

In some embodiments, the DNA binding domain is capable of binding to a target sequence (e.g., a dsDNA target sequence) with greater affinity than a reference DNA binding domain. In some embodiments, the reference DNA binding domain is a DNA binding domain from Cas9 of S. pyogenes. In some embodiments, the DNA binding domain is capable of binding to a target sequence (e.g., a dsDNA target sequence) with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM).

In some embodiments, the affinity of a DNA binding domain for its target sequence (e.g., dsDNA target sequence) is measured in vitro, e.g., by thermophoresis, e.g., as described in Asmari et al. Methods 146:107-119 (2018) (incorporated by reference herein in its entirety).

In embodiments, the DNA binding domain is capable of binding to its target sequence (e.g., dsDNA target sequence), e.g, with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM) in the presence of a molar excess of scrambled sequence competitor dsDNA, e.g., of about 100-fold molar excess.

In some embodiments, the DNA binding domain is found associated with its target sequence (e.g., dsDNA target sequence) more frequently than any other sequence in the genome of a target cell, e.g., human target cell, e.g., as measured by ChIP-seq (e.g., in HEK293T cells), e.g., as described in He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (incorporated herein by reference in its entirety). In some embodiments, the DNA binding domain is found associated with its target sequence (e.g., dsDNA target sequence) at least about 5-fold or 10-fold, more frequently than any other sequence in the genome of a target cell, e.g., as measured by ChIP-seq (e.g., in HEK293T cells), e.g., as described in He and Pu (2010), *supra*.

In some embodiments, the endonuclease domain has nickase activity and cleaves one strand of a target DNA. In some embodiments, nickase activity reduces the formation of double-stranded breaks at the target site. In some embodiments, the endonuclease domain creates a staggered nick structure in the first and second strands of a target DNA. In some embodiments, a staggered nick structure generates free 3' overhangs at the target site. In some embodiments,

free 3' overhangs at the target site improve editing efficiency, e.g., by enhancing access and annealing of a 3' homology region of a template nucleic acid. In some embodiments, a staggered nick structure reduces the formation of double-stranded breaks at the target site.

In some embodiments, the endonuclease domain cleaves both strands of a target DNA, e.g., results in blunt-end cleavage of a target with no ssDNA overhangs on either side of the cut-site. The amino acid sequence of an endonuclease domain of a gene modifying system described herein may be at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identical to the amino acid sequence of an endonuclease domain described herein, e.g., an endonuclease domain from Table 8.

In certain embodiments, the heterologous endonuclease is Fok1 or a functional fragment thereof. In certain embodiments, the heterologous endonuclease is a Holliday junction resolvase or homolog thereof, such as the Holliday junction resolving enzyme from Sulfolobus solfataricus—Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). In certain embodiments, the heterologous endonuclease is the endonuclease of the large fragment of a spliceosomal protein, such as Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). In certain embodiments, the heterologous endonuclease is derived from a CRISPR-associated protein, e.g., Cas9. In certain embodiments, the heterologous endonuclease is engineered to have only ssDNA cleavage activity, e.g., only nickase activity, e.g., be a Cas9 nickase, e.g., SpCas9 with D10A, H840A, or N863A mutations. Table 8 provides exemplary Cas proteins and mutations associated with nickase activity. In still other embodiments, homologous endonuclease domains are modified, for example by site-specific mutation, to alter DNA endonuclease activity. In still other embodiments, endonuclease domains are modified to reduce DNA-sequence specificity, e.g., by truncation to remove domains that confer DNA-sequence specificity or mutation to inactivate regions conferring DNA-sequence specificity.

In some embodiments, the endonuclease domain has nickase activity and does not form double-stranded breaks. In some embodiments, the endonuclease domain forms single-stranded breaks at a higher frequency than double-stranded breaks, e.g., at least 90%, 95%, 96%, 97%, 98%, or 99% of the breaks are single-stranded breaks, or less than 10%, 5%, 4%, 3%, 2%, or 1% of the breaks are double-stranded breaks. In some embodiments, the endonuclease forms

substantially no double-stranded breaks. In some embodiments, the endonuclease does not form detectable levels of double-stranded breaks.

In some embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the first strand; e.g., in some embodiments, the endonuclease domain cuts the genomic DNA of the target site near to the site of alteration on the strand that will be extended by the writing domain. In some embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the first strand and does not nick the target site DNA of the second strand. For example, when a polypeptide comprises a CRISPR-associated endonuclease domain having nickase activity, in some embodiments, said CRISPR-associated endonuclease domain nicks the target site DNA strand containing the PAM site (e.g., and does not nick the target site DNA strand that does not contain the PAM site). As a further example, when a polypeptide comprises a CRISPR-associated endonuclease domain having nickase activity, in some embodiments, said CRISPR-associated endonuclease domain nicks the target site DNA strand not containing the PAM site (e.g., and does not nick the target site DNA strand that contains the PAM site).

In some other embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the first strand and the second strand. Without wishing to be bound by theory, after a writing domain (e.g., RT domain) of a polypeptide described herein polymerizes (e.g., reverse transcribes) from the heterologous object sequence of a template nucleic acid (e.g., template RNA), the cellular DNA repair machinery must repair the nick on the first DNA strand. The target site DNA now contains two different sequences for the first DNA strand: one corresponding to the original genomic DNA (e.g., having a free 5′ end) and a second corresponding to that polymerized from the heterologous object sequence (e.g., having a free 3′ end). It is thought that the two different sequences equilibrate with one another, first one hybridizing the second strand, then the other, and which sequence the cellular DNA repair apparatus incorporates into its repaired target site may be a stochastic process. Without wishing to be bound by theory, it is thought that introducing an additional nick to the second-strand may bias the cellular DNA repair machinery to adopt the heterologous object sequence-based sequence more frequently than the original genomic sequence (Anzalone et al. Nature 576:149-157 (2019)). In some embodiments, the additional nick is positioned at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140,

145, or 150 nucleotides 5′ or 3′ of the target site modification (e.g., the insertion, deletion, or substitution) or to the nick on the first strand.

Alternatively or additionally, without wishing to be bound by theory, it is thought that an additional nick to the second strand may promote second-strand synthesis. In some embodiments, where the gene modifying system has inserted or substituted a portion of the first strand, synthesis of a new sequence corresponding to the insertion/substitution in the second strand is necessary.

In some embodiments, the polypeptide comprises a single domain having endonuclease activity (e.g., a single endonuclease domain) and said domain nicks both the first strand and the second strand. For example, in such an embodiment the endonuclease domain may be a CRISPR-associated endonuclease domain, and the template nucleic acid (e.g., template RNA) comprises a gRNA spacer that directs nicking of the first strand and an additional gRNA spacer that directs nicking of the second strand. In some embodiments, the polypeptide comprises a plurality of domains having endonuclease activity, and a first endonuclease domain nicks the first strand and a second endonuclease domain nicks the second strand (optionally, the first endonuclease domain does not (e.g., cannot) nick the second strand and the second endonuclease domain does not (e.g., cannot) nick the first strand).

In some embodiments, the endonuclease domain is capable of nicking a first strand and a second strand. In some embodiments, the first and second strand nicks occur at the same position in the target site but on opposite strands. In some embodiments, the second strand nick occurs in a staggered location, e.g., upstream or downstream, from the first nick. In some embodiments, the endonuclease domain generates a target site deletion if the second strand nick is upstream of the first strand nick. In some embodiments, the endonuclease domain generates a target site duplication if the second strand nick is downstream of the first strand nick. In some embodiments, the endonuclease domain generates no duplication and/or deletion if the first and second strand nicks occur in the same position of the target site. In some embodiments, the endonuclease domain has altered activity depending on protein conformation or RNA-binding status, e.g., which promotes the nicking of the first or second strand (e.g., as described in Christensen et al. PNAS 2006; incorporated by reference herein in its entirety).

In some embodiments, the endonuclease domain comprises a meganuclease, or a functional fragment thereof. In some embodiments, the endonuclease domain comprises a

homing endonuclease, or a functional fragment thereof. In some embodiments, the endonuclease domain comprises a meganuclease from the LAGLIDADG, GIY-YIG, HNH, His-Cys Box, or PD-(D/E) XK families, or a functional fragment or variant thereof, e.g., which possess conserved amino acid motifs, e.g., as indicated in the family names. In some embodiments, the endonuclease domain comprises a meganuclease, or fragment thereof, chosen from, e.g., I-SmaMI (Uniprot F7WD42), I-SceI (Uniprot P03882), I-AniI (Uniprot P03880), I-DmoI (Uniprot P21505), I-CreI (Uniprot P05725), I-TevI (Uniprot P13299), I-OnuI (Uniprot Q4VWW5), or I-BmoI (Uniprot Q9ANR6). In some embodiments, the meganuclease is naturally monomeric, e.g., I-SceI, I-TevI, or dimeric, e.g., I-CreI, in its functional form. For example, the LAGLIDADG meganucleases with a single copy of the LAGLIDADG motif generally form homodimers, whereas members with two copies of the LAGLIDADG motif are generally found as monomers. In some embodiments, a meganuclease that normally forms as a dimer is expressed as a fusion, e.g., the two subunits are expressed as a single ORF and, optionally, connected by a linker, e.g., an I-CreI dimer fusion (Rodriguez-Fornes et al. Gene Therapy 2020; incorporated by reference herein in its entirety). In some embodiments, a meganuclease, or a functional fragment thereof, is altered to favor nickase activity for one strand of a double-stranded DNA molecule, e.g., I-SceI (K122I and/or K223I) (Niu et al. J Mol Biol 2008), I-AniI (K227M) (McConnell Smith et al. PNAS 2009), I-DmoI (Q42A and/or K120M) (Molina et al. J Biol Chem 2015). In some embodiments, a meganuclease or functional fragment thereof possessing this preference for single-strand cleavage is used as an endonuclease domain, e.g., with nickase activity. In some embodiments, an endonuclease domain comprises a meganuclease, or a functional fragment thereof, which naturally targets or is engineered to target a safe harbor site, e.g., an I-CreI targeting SH6 site (Rodriguez-Fornes et al., *supra*). In some embodiments, an endonuclease domain comprises a meganuclease, or a functional fragment thereof, with a sequence tolerant catalytic domain, e.g., I-TevI recognizing the minimal motif CNNNG (Kleinstiver et al. PNAS 2012). In some embodiments, a target sequence tolerant catalytic domain is fused to a DNA binding domain, e.g., to direct activity, e.g., by fusing I-TevI to: (i) zinc fingers to create Tev-ZFEs (Kleinstiver et al. PNAS 2012), (ii) other meganucleases to create MegaTevs (Wolfs et al. Nucleic Acids Res 2014), and/or (iii) Cas9 to create TevCas9 (Wolfs et al. PNAS 2016).

In some embodiments, the endonuclease domain comprises a restriction enzyme, e.g., a Type IIS or Type IIP restriction enzyme. In some embodiments, the endonuclease domain comprises a Type IIS restriction enzyme, e.g., FokI, or a fragment or variant thereof. In some embodiments, the endonuclease domain comprises a Type IIP restriction enzyme, e.g., PvuII, or a fragment or variant thereof. In some embodiments, a dimeric restriction enzyme is expressed as a fusion such that it functions as a single chain, e.g., a FokI dimer fusion (Minczuk et al. Nucleic Acids Res 36(12):3926-3938 (2008)).

The use of additional endonuclease domains is described, for example, in Guha and Edgell Int J Mol Sci 18(22):2565 (2017), which is incorporated herein by reference in its entirety.

In some embodiments, a gene modifying polypeptide comprises a modification to an endonuclease domain, e.g., relative to a wild-type Cas protein. In some embodiments, the endonuclease domain comprises an addition, deletion, replacement, or modification to the amino acid sequence of the wild-type Cas protein. In some embodiments, the endonuclease domain is modified to include a heterologous functional domain that binds specifically to and/or induces endonuclease cleavage of a target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the endonuclease domain comprises a zinc finger. In embodiments, the endonuclease domain comprising the Cas domain is associated with a guide RNA (gRNA), e.g., as described herein. In some embodiments, the endonuclease domain is modified to include a functional domain that does not target a specific target nucleic acid (e.g., DNA) sequence. In embodiments, the endonuclease domain comprises a Fok1 domain.

In some embodiments, the endonuclease domain is associated with the target dsDNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled dsDNA. In some embodiments, the endonuclease domain is associated with the target dsDNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled dsDNA, e.g., in a cell (e.g., a HEK293T cell). In some embodiments, the frequency of association between the endonuclease domain and the target DNA or scrambled DNA is measured by ChIP-seq, e.g., as described in He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (incorporated by reference herein in its entirety).

In some embodiments, the endonuclease domain can catalyze the formation of a nick at a target sequence, e.g., to an increase of at least about 5-fold or 10-fold relative to a non-target

sequence (e.g., relative to any other genomic sequence in the genome of the target cell). In some embodiments, the level of nick formation is determined using NickSeq, e.g., as described in Elacqua et al. (2019) *bioRxiv* doi.org/10.1101/867937 (incorporated herein by reference in its entirety).

In some embodiments, the endonuclease domain is capable of nicking DNA *in vitro*. In embodiments, the nick results in an exposed base. In embodiments, the exposed base can be detected using a nuclease sensitivity assay, e.g., as described in Chaudhry and Weinfeld (1995) *Nucleic Acids Res* 23(19):3805-3809 (incorporated by reference herein in its entirety). In embodiments, the level of exposed bases (e.g., detected by the nuclease sensitivity assay) is increased by at least 10%, 50%, or more relative to a reference endonuclease domain. In some embodiments, the reference endonuclease domain is an endonuclease domain from Cas9 of S. pyogenes.

In some embodiments, the endonuclease domain is capable of nicking DNA in a cell. In embodiments, the endonuclease domain is capable of nicking DNA in a HEK293T cell. In embodiments, an unrepaired nick that undergoes replication in the absence of Rad51 results in increased NHEJ rates at the site of the nick, which can be detected, e.g., by using a Rad51 inhibition assay, e.g., as described in Bothmer et al. (2017) *Nat Commun* 8:13905 (incorporated by reference herein in its entirety). In embodiments, NHEJ rates are increased above 0-5%. In embodiments, NHEJ rates are increased to 20-70% (e.g., between 30%-60% or 40-50%), e.g., upon Rad51 inhibition.

In some embodiments, the endonuclease domain releases the target after cleavage. In some embodiments, release of the target is indicated indirectly by assessing for multiple turnovers by the enzyme, e.g., as described in Yourik at al. RNA 25(1):35-44 (2019) (incorporated herein by reference in its entirety) and shown in FIG. 2. In some embodiments, the k_{exp} of an endonuclease domain is $1 \times 10^{-3} - 1 \times 10^{-5}$ min-1 as measured by such methods.

In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about 1×10^8 s^{-1} M^{-1} *in vitro*. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1×10^5, 1×10^6, 1×10^7, or 1×10^8, s^{-1} M^{-1} *in vitro*. In embodiments, catalytic efficiency is determined as described in Chen et al. (2018) *Science* 360(6387):436-439 (incorporated herein by reference in its entirety). In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about 1×10^8 s^{-1} M^{-1} in

cells. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1 x 10^5, 1 x 10^6, 1 x 10^7, or 1 x 10^8 s^{-1} M^{-1} in cells.

Gene modifying polypeptides comprising Cas domains

In some embodiments, a gene modifying polypeptide described herein comprises a Cas domain. In some embodiments, the Cas domain can direct the gene modifying polypeptide to a target site specified by a gRNA spacer, thereby modifying a target nucleic acid sequence in "cis". In some embodiments, a gene modifying polypeptide is fused to a Cas domain. In some embodiments, a gene modifying polypeptide comprises a CRISPR/Cas domain (also referred to herein as a CRISPR-associated protein). In some embodiments, a CRISPR/Cas domain comprises a protein involved in the clustered regulatory interspaced short palindromic repeat (CRISPR) system, e.g., a Cas protein, and optionally binds a guide RNA, e.g., single guide RNA (sgRNA).

CRISPR systems are adaptive defense systems originally discovered in bacteria and archaea. CRISPR systems use RNA-guided nucleases termed CRISPR-associated or "Cas" endonucleases (e. g., Cas9 or Cpf1) to cleave foreign DNA. For example, in a typical CRISPR-Cas system, an endonuclease is directed to a target nucleotide sequence (e. g., a site in the genome that is to be sequence-edited) by sequence-specific, non-coding "guide RNAs" that target single- or double-stranded DNA sequences. Three classes (I-III) of CRISPR systems have been identified. The class II CRISPR systems use a single Cas endonuclease (rather than multiple Cas proteins). One class II CRISPR system includes a type II Cas endonuclease such as Cas9, a CRISPR RNA ("crRNA"), and a trans-activating crRNA ("tracrRNA"). The crRNA contains a "spacer" sequence, a typically about 20-nucleotide RNA sequence that corresponds to a target DNA sequence ("protospacer"). In the wild-type system, and in some engineered systems, crRNA also contains a region that binds to the tracrRNA to form a partially double-stranded structure that is cleaved by RNase III, resulting in a crRNA/tracrRNA hybrid molecule. A crRNA/tracrRNA hybrid then directs the Cas endonuclease to recognize and cleave a target DNA sequence. A target DNA sequence is generally adjacent to a "protospacer adjacent motif" ("PAM") that is specific for a given Cas endonuclease and required for cleavage activity at a target site matching the spacer of the crRNA. CRISPR endonucleases identified from various prokaryotic species have unique PAM sequence requirements, e.g., as listed for exemplary Cas enzymes in Table 7; examples of PAM sequences include 5′-NGG (Streptococcus pyogenes;

SEQ ID NO: 11,019), 5′-NNAGAA (Streptococcus thermophilus CRISPR1; SEQ ID NO: 11,020), 5′-NGGNG (Streptococcus thermophilus CRISPR3; SEQ ID NO: 11,021), and 5′-NNNGATT (Neisseria meningiditis; SEQ ID NO: 11,022). Some endonucleases, e.g., Cas9 endonucleases, are associated with G-rich PAM sites, e. g., 5′-NGG (SEQ ID NO: 11,023), and perform blunt-end cleaving of the target DNA at a location 3 nucleotides upstream from (5′ from) the PAM site. Another class II CRISPR system includes the type V endonuclease Cpf1, which is smaller than Cas9; examples include AsCpf1 (from Acidaminococcus sp.) and LbCpf1 (from Lachnospiraceae sp.). Cpf1-associated CRISPR arrays are processed into mature crRNAs without the requirement of a tracrRNA; in other words, a Cpf1 system, in some embodiments, comprises only Cpf1 nuclease and a crRNA to cleave a target DNA sequence. Cpf1 endonucleases, are typically associated with T-rich PAM sites, e. g., 5′-TTN. Cpf1 can also recognize a 5′-CTA PAM motif. Cpf1 typically cleaves a target DNA by introducing an offset or staggered double-strand break with a 4- or 5-nucleotide 5′ overhang, for example, cleaving a target DNA with a 5-nucleotide offset or staggered cut located 18 nucleotides downstream from (3′ from) from a PAM site on the coding strand and 23 nucleotides downstream from the PAM site on the complimentary strand; the 5-nucleotide overhang that results from such offset cleavage allows more precise genome editing by DNA insertion by homologous recombination than by insertion at blunt-end cleaved DNA. See, e.g., Zetsche et al. (2015) Cell, 163:759 – 771.

A variety of CRISPR associated (Cas) genes or proteins can be used in the technologies provided by the present disclosure and the choice of Cas protein will depend upon the particular conditions of the method. Specific examples of Cas proteins include class II systems including Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cpf1, C2C1, or C2C3. In some embodiments, a Cas protein, e.g., a Cas9 protein, may be from any of a variety of prokaryotic species. In some embodiments a particular Cas protein, e.g., a particular Cas9 protein, is selected to recognize a particular protospacer-adjacent motif (PAM) sequence. In some embodiments, a DNA-binding domain or endonuclease domain includes a sequence targeting polypeptide, such as a Cas protein, e.g., Cas9. In certain embodiments a Cas protein, e.g., a Cas9 protein, may be obtained from a bacteria or archaea or synthesized using known methods. In certain embodiments, a Cas protein may be from a gram-positive bacteria or a gram-negative bacteria. In certain embodiments, a Cas protein may be from a Streptococcus (e.g., a S. pyogenes, or a S. thermophilus), a Francisella (e.g., an F. novicida), a Staphylococcus (e.g., an S. aureus), an

Acidaminococcus (e.g., an Acidaminococcus sp. BV3L6), a Neisseria (e.g., an N. meningitidis), a Cryptococcus, a Corynebacterium, a Haemophilus, a Eubacterium, a Pasteurella, a Prevotella, a Veillonella, or a Marinobacter.

In some embodiments, a gene modifying polypeptide may comprise the amino acid sequence of SEQ ID NO: 4000 below, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto. In embodiments, the amino acid sequence of SEQ ID NO: 4000 below, or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto, is positioned at the N-terminal end of the gene modifying polypeptide. In embodiments, the amino acid sequence of SEQ ID NO: 4000 below, or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto, is positioned within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 amino acids of the N-terminal end of the gene modifying polypeptide.

Exemplary N-terminal NLS-Cas9 domain

MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI
KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG
VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLF
DDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGT

ALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSF
EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV
NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA
YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGDGG (SEQ ID NO: 4000)

In some embodiments, a gene modifying polypeptide may comprise the amino acid sequence of SEQ ID NO: 4001 below, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto. In embodiments, the amino acid sequence of SEQ ID NO: 4001 below, or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto, is positioned at the C-terminal end of the gene modifying polypeptide. In embodiments, the amino acid sequence of SEQ ID NO: 4001 below, or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto, is positioned within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 amino acids of the C-terminal end of the gene modifying polypeptide.

Exemplary C-terminal sequence comprising an NLS

AGKRTADGSEFEKRTADGSEFESPKKKAKVE (SEQ ID NO: 4001)

Exemplary benchmarking sequence

MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI
KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLE
NLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL
PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR
TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAW
MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISG

VEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLF
DDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR
DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKK
MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS
RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGT
ALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG
EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSF
EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV
NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA
YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG
LYETRIDLSQLGGDGGSGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGTLNIEDEYRL
HETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQE
ARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTV
PNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL
PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLG
NLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLG
KAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTK
PFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTK
DAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVAL
NPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAG
AAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHG
EIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQA
ARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEAGKRTADGSEFEKRTADGSEFESPK
KKAKVE (SEQ ID NO: 4002)

In some embodiments, a gene modifying polypeptide may comprise a Cas domain as listed in Table 7 or 8, or a functional fragment thereof, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto.

Table 7. CRISPR/Cas Proteins, Species, and Mutations

Name	*Enzyme*	*Species*	*# of AAs*	*PAM*	*SEQ ID NO:*	*Mutations to alter PAM recognition*	*Mutations to make catalytically dead*
FnCas9	Cas9	*Francisella novicida*	1629	5'-NGG-3'	11,024	Wt	D11A/H969A/N995A
FnCas9 RHA	Cas9	*Francisella novicida*	1629	5'-YG-3'	11,025	E1369R/E1449H/R1556A	D11A/H969A/N995A
SaCas9	Cas9	*Staphylococcus aureus*	1053	5'-NNGRRT-3'	11,026	Wt	D10A/H557A
SaCas9 KKH	Cas9	*Staphylococcus aureus*	1053	5'-NNNRRT-3'	11,027	E782K/N968K/R1015H	D10A/H557A
SpCas9	Cas9	*Streptococcus pyogenes*	1368	5'-NGG-3'	11,028	Wt	D10A/D839A/H840A/N863A
SpCas9 VQR	Cas9	*Streptococcus pyogenes*	1368	5'-NGA-3'	11,029	D1135V/R1335Q/T1337R	D10A/D839A/H840A/N863A
AsCpf1 RR	Cpf1	*Acidaminococcus sp. BV3L6*	1307	5'-TYCV-3'	11,030	S542R/K607R	E993A
AsCpf1 RVR	Cpf1	*Acidaminococcus sp. BV3L6*	1307	5'-TATV-3'	11,031	S542R/K548V/N552R	E993A
FnCpf1	Cpf1	*Francisella novicida*	1300	5'-NTTN-3'	11,032	Wt	D917A/E1006A/D1255A
NmCas9	Cas9	*Neisseria meningitidis*	1082	5'-NNNGATT-3'	11,033	Wt	D16A/D587A/H588A/N611A

Table 8 Amino Acid Sequences of CRISPR/Cas Proteins, Species, and Mutations

Variant	**Parental Host(s)**	**Protein Sequence**	**SEQ ID NO:**	**Nickase (HNH)**	**Nickase (HNH)**	**Nickase (RuvC)**
Nme2Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPK TGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKS LPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELG ALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKD LQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCT FEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRK SKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEG LKDKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKF VQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRN PVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENR KDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNE KGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSR EWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVA DHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACS TVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEV MIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNR KMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIEL YEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNK KNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKG YRIDDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGS KEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVR	9,001	N611A	H588A	D16A
PpnCas9	Pasteurella pneumotropica	MQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAKTGE SLALSRRLARSSRRLIKRRAERLKKAKRLLKAEKILHSIDEKLPINVWQLRVKGL KEKLERQEWAAVLLHLSKHRGYLSQRKNEGKSDNKELGALLSGIASNHQML QSSEYRTPAEIAVKKFQVEEGHIRNQRGSYTHTFSRLDLLAEMELLFQRQAEL GNSYTSTTLLENLTALLMWQKPALAGDAILKMLGKCTFEPSEYKAAKNSYSA ERFVWLTKLNNLRILENGTERALNDNERFALLEQPYEKSKLTYAQVRAMLAL SDNAIFKGVRYLGEDKKTVESKTTLIEMKFYHQIRKTLGSAELKKEWNELKGN	9,002	N605A	H582A	D13A

		SDLLDEIGTAFSLYKTDDDICRYLEGKLPERVLNALLENLNFDKFIQLSLKALHQ ILPLMLQGQRYDEAVSAIYGDHYGKKSTETTRLLPTIPADEIRNPVVLRTLTQA RKVINAVVRLYGSPARIHIETAREVGKSYQDRKKLEKQQEDNRKQRESAVKK FKEMFPHFVGEPKGKDILKMRLYELQQAKCLYSGKSLELHRLLEKGYVEVDH ALPFSRTWDDSFNNKVLVLANENQNKGNLTPYEWLDGKNNSERWQHFVV RVQTSGFSYAKKQRILNHKLDEKGFIERNLNDTRYVARFLCNFIADNMLLVG KGKRNVFASNGQITALLRHRWGLQKVREQNDRHHALDAVVVACSTVAMQ QKITRFVRYNEGNVFSGERIDRETGEIIPLHFPSPWAFFKENVEIRIFSENPKLE LENRLPDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSAKRLNEGLS VLKVPLTQLKLSDLERMVNRDREIALYESLKARLEQFGNDPAKAFAEPFYKKG GALVKAVRLEQTQKSGVLVRDGNGVADNASMVRVDVFTKGGKYFLVPIYT WQVAKGILPNRAATQGKDENDWDIMDEMATFQFSLCQNDLIKLVTKKKTI FGYFNGLNRATSNINIKEHDLDKSKGKLGIYLEVGVKLAISLEKYQVDELGKNI RPCRPTKRQHVR				
SauCas9	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVN NLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP RIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	9,003	N580A	H557A	D10A
SauCas9-KKH	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	9,004	N580A	H557A	D10A
SauriCas9	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNR RSKRGARRLKRRRIHRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPL TKEEFAIALLHIAKRRGLHNISVSMGDEEQDNELSTKQQLQKNAQQLQDKY VCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQY	9,005	N588A	H565A	D15A

		IDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGV QDYDIRGYRITKSGKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQ DEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPENQ MEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSIKVINAVINRFGL PEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIEKI KLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQ SENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEER DINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAVKVKTINGGFTNH LRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTHKALRRTDKILEQPGLE VNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRQLINDTL YSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLM TILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVS NKYPETQNKLVKLSLKSFRFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYE AEKQKKKIKESDLFVGSFYYNDLIMYEDELFRVIGVNSDINNLVELNMVDITY KDFCEVNNVTGEKRIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL				
SauriCas9-KKH	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNR RSKRGARRLKRRRIHRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPL TKEEFAIALLHIAKRRGLHNISVSMGDEEQDNELSTKQQLQKNAQQLQDKY VCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQY IDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGV QDYDIRGYRITKSGKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQ DEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPENQ MEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSIKVINAVINRFGL PEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIEKI KLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQ SENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEER DINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAVKVKTINGGFTNH LRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTHKALRRTDKILEQPGLE VNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRKLINDTL YSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLM TILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVS NKYPETQNKLVKLSLKSFRFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYE AEKQKKKIKESDLFVGSFYKNDLIMYEDELFRVIGVNSDINNLVELNMVDITY KDFCEVNNVTGEKHIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL	9,006	N588A	H565A	D15A
ScaCas9-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALL FDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESF LVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALA HIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSA RLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKD TYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMV KRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRS GKLATEEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLK ELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEA ITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNEL TKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE ERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKS DGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELE SQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKN DKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKL ANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTG GFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKL KSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRR	9,007	N872A	H849A	D10A

		MLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIF EKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD				
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	9,008	N863A	H840A	D10A
SpyCas9-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES IRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF KYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	9,009	N863A	H840A	D10A
SpyCas9-SpRY	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAERTRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH	9,010	N863A	H840A	D10A

		MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES IRPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTRLGAPRAF KYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD				
St1Cas9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFK APYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADE TYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPN KQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDIT PKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQ EKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKH YVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGN QHIIKNEGDKPKLDF	9,011	N622A	H599A	D9A
BlatCas9	Brevibacillus laterosporus	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKNGEALAVPRRE ARSSRRRLRRKKHRIERLKHMFVRNGLAVDIQHLEQTLRSQNEIDVWQLRV DGLDRMLTQKEWLRVLIHLAQRRGFQSNRKTDGSSEDGQVLVNVTENDRL MEEKDYRTVAEMMVKDEKFSDHKRNKGNYHGVVSRSSLLVEIHTLFETQ RQHHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTFLPKEKRAPKAS WHFQYFMLLQTINHIRITNVQGTRSLNKEEIEQVVNMALTKSKVSYHDTRKI LDLSEEYQFVGLDYGKEDEKKKVESKETIIKLDDYHKLNKIFNEVELAKGETWE ADDYDTVAYALTFFKDDEDIRDYLQNKYKDSKNRLVKNLANKEYTNELIGKV STLSFRKVGHLSLKALRKIIPFLEQGMTYDKACQAAGFDFQGISKKKRSVVLP VIDQISNPVVNRALTQTRKVINALIKKYGSPETIHIETARELSKTFDERKNITKD YKENRDKNEHAKKHLSELGIINPTGLDIVKYKLWCEQQGRCMYSNQPISFER LKESGYTEVDHIIPYSRSMNDSYNNRVLVMTRENREKGNQTPFEYMGNDT QRWYEFEQRVTTNPQIKKEKRQNLLLKGFTNRRELEMLERNLNDTRYITKYL SHFISTNLEFSPSDKKKKVVNTSGRITSHLRSRWGLEKNRGQNDLHHAMDAI	9,012	N607A	H584A	D8A

		VIAVTSDSFIQQVTNYYKRKERRELNGDDKFPLPWKFFREEVIARLSPNPKEQ IEALPNHFYSEDELADLQPIFVSRMPKRSITGEAHQAQFRRVVGKTKEGKNIT AKKTALVDISYDKNGDFNMYGRETDPATYEAIKERYLEFGGNVKKAFSTDLH KPKKDGTKGPLIKSVRIMENKTLVHPVNKGKGVVYNSSIVRTDVFQRKEKYY LLPVYVTDVTKGKLPNKVIVAKKGYHDWIEVDDSFTFLFSLYPNDLIFIRQNPK KKISLKKRIESHSISDSKEVQEIHAYYKGVDSSTAAIEFIIHDGSYYAKGVGVQN LDCFEKYQVDILGNYFKVKGEKRLELETSDSNHKGKDVNSIKSTSR				
cCas9-v16	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNSDKNNLIEVNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	9,013	N580A	H557A	D10A
cCas9-v17	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNNSTRNIVELNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	9,014	N580A	H557A	D10A
cCas9-v21	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ	9,015	N580A	H557A	D10A

		YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNSDDRNIIELNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG				
cCas9-v42	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNNNRLNKIELNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	9,016	N580A	H557A	D10A
CdiCas9	Corynebacterium diphtheriae	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVT RLASSGIARRTRRLYRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVR AELAASYIADEKERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFK AIREEIKRASGQPVPETATVGQMVTLCELGTLKLRGEGGVLSARLQQSDYAR EIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKAS DAFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIA EILGIDRGQLIGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTA SALEQHAMVKALSNAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSLHLPV GRAAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPTPRIGEPVGNP AVDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVTEKRAREMDGDMRR RAARNAKLFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQCAYCGSPITF SNSEMDHIVPRAGQGSTNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEG VSVKEAVERTRHWVTDTGMRSTDFKKFTKAVVERFQRATMDEEIDARSME SVAWMANELRSRVAQHFASHGTTVRVYRGSLTAEARRASGISGKLKFFDGV GKSRLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPQWREFT GKDAEHRAAWRVWCQKMEKLSALLTEDLRDDRVVVMSNVRLRLGNGSA HKETIGKLSKVKLSSQLSVSDIDKASSEALWCALTREPGFDPKEGLPANPERHI RVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSFHHARVYKITSGKKPAF AMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKLRDALATGNAEYLG WLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFFSPSKLRLRPLQM SKEGIKKESAPELSKIIDRPGWLPAVNKLFSDGNVTVVRRDSLGRVRLESTAH LPVTWKVQ	9,017	N597A	H573A	D8A
CjeCas9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSA RKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRA LNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQS VGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFG FSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVAL TRIINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFK GEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLN	9,018	N582A	H559A	D8A

		QNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDK KDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVG KNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAY SGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFE AFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYI ARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTW GFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELD YKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSY GGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDF ALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFV YYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEK YIVSALGEVTKAEFRQREDFKK				
GeoCas9	Geobacillus stearothermophilus	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLA RSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDR KLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRAILSSYRTV GEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEF ENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHIN KLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDR GESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKD DADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLKALRS ILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPNIPPIANPVVMRALTQA RKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQL MEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPY SRSLDDSYTNKVLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFS KKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQK VYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFY QRREQNKELAKKTEPHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQ KLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKL DASGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGP VIRTVKIIDTKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIM KGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEE INVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNI YKVRGEKRVGLASSAHSKPGKTIRPLQSTRD	9,019	N605A	H582A	D8A
iSpyMacCas9	Streptococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGG LFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISV MNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEI HKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEIISFSKKC KLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQ	9,020	N863A	H840A	D10A

		KQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGEDSGGSGGSKRTADGSEFES				
NmeCas9	Neisseria meningitidis	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETARDVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR	9,021	N611A	H588A	D16A
ScaCas9	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVDGGASQEEFYKFIKPILEKMDGAEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNEKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	9,022	N872A	H849A	D10A
ScaCas9-HiFi-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLKRTARRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVDGADKKLRKRSGKLATEEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEA	9,023	N872A	H849A	D10A

		ITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNEL TKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE ERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKS DGFSNANFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELE SQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKN DKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKL ANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTG GFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKL KSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRR MLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIF EKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD				
SpyCas9-3var-NRRH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGVLHKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGVPAA FKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	9,024	N863A	H840A	D10A
SpyCas9-3var-NRTH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF	9,025	N863A	H840A	D10A

		LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS ASVLHKGNELALPSKYVNFLYLASHYEKLKGSSEDNKQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGASAAF KYFDTTIGRKLYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD				
SpyCas9-3var-NRCH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRIDLSQLGGD	9,026	N863A	H840A	D10A
SpyCas9-HF1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA	9,027	N863A	H840A	D10A

		GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD				
SpyCas9-QQR1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADAQLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTFKQKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	9,028	N863A	H840A	D10A
SpyCas9-SpG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	9,029	N863A	H840A	D10A
SpyCas9-VQR	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH	9,030	N863A	H840A	D10A

		MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD				
SpyCas9-VRER	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	9,031	N863A	H840A	D10A
SpyCas9-xCas	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEK VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE	9,032	N863A	H840A	D10A

		GMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD				
SpyCas9-xCas-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEK VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES IRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF KYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	9,033	N863A	H840A	D10A
St1Cas9-CNRZ1066	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDET YVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNK	9,034	N622A	H599A	D9A

		QMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDI TPENSKNKVVLQSLKPWRTDVYFNKATGKYEILGLKYADLQFEKGTGTYKIS QEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPKQK HYVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTDVLG NQHIIKNEGDKPKLDF				
St1Cas9-LMG1831	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDET YVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNK QMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDI TPENSKNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYADLQFEKKTGTYKISQ EKYNGIMKEEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPNVK YYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISIYKVRTDVLGNQH IIKNEGDKPKLDF	9,035	N622A	H599A	D9A
St1Cas9-MTH17CL396	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFK APYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADE TYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPN KQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDIT PKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISK EQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHYV ELKPYNRQKFEGSEYLIKSLGTVAKGGQCIKGLGKSNISIYKVRTDVLGNQHII KNEGDKPKLDF	9,036	N622A	H599A	D9A
St1Cas9-TH1477	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY	9,037	N622A	H599A	D9A

		NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFK APYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADE TYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPN KQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDIT PKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISK EQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHYV ELKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSNISIYKVRTDVLGNQHIIK NEGDKPKLDF				
sRGN3.1	Staphylococcus spp.	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIAL LHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLE NEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREYF EGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALN DLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRI TKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQ LEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYL NMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDIIIE LARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQ QEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSK KSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFE VQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLRKV WKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETKQLDI QVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRKK DNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYA NEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQFKSST KKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKKKI KDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNIK GEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	9,038	N585A	H562A	D10A
sRGN3.3	Staphylococcus spp.	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGS RRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIAL LHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLE NEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREYF EGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALN DLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRI TKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQ LEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYL NMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDIIIE LARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQ QEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSK KSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFE VQKEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKV WRFDKYRNHGYKHHAEDALIIANADFLFKENKKLQNTNKILEKPTIENNTKK VTVEKEEDYNNVFETPKLVEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTRM KDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIMKQ YSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYEN STKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKK KIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNI KGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	9,039	N585A	H562A	D10A

In some embodiments, a Cas protein requires a protospacer adjacent motif (PAM) to be present in or adjacent to a target DNA sequence for the Cas protein to bind and/or function. In

some embodiments, the PAM is or comprises, from 5′ to 3′, NGG (SEQ ID NO: 11,024), YG (SEQ ID NO: 11,025), NNGRRT (SEQ ID NO: 11,026), NNNRRT (SEQ ID NO: 11,027), NGA (SEQ ID NO: 11,029), TYCV (SEQ ID NO: 11,030), TATV (SEQ ID NO: 11,031), NTTN (SEQ ID NO: 11,032), or NNNGATT (SEQ ID NO: 11,033), where N stands for any nucleotide, Y stands for C or T, R stands for A or G, and V stands for A or C or G.In some embodiments, a Cas protein is a protein listed in Table 7 or 8. In some embodiments, a Cas protein comprises one or more mutations altering its PAM. In some embodiments, a Cas protein comprises E1369R, E1449H, and R1556A mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises E782K, N968K, and R1015H mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises D1135V, R1335Q, and T1337R mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises S542R and K607R mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises S542R, K548V, and N552R mutations or analogous substitutions to the amino acids corresponding to said positions. Exemplary advances in the engineering of Cas enzymes to recognize altered PAM sequences are reviewed in Collias et al Nature Communications 12:555 (2021), incorporated herein by reference in its entirety.

In some embodiments, the Cas protein is catalytically active and cuts one or both strands of the target DNA site. In some embodiments, cutting the target DNA site is followed by formation of an alteration, e.g., an insertion or deletion, e.g., by the cellular repair machinery.

In some embodiments, the Cas protein is modified to deactivate or partially deactivate the nuclease, e.g., nuclease-deficient Cas9. Whereas wild-type Cas9 generates double-strand breaks (DSBs) at specific DNA sequences targeted by a gRNA, a number of CRISPR endonucleases having modified functionalities are available, for example: a “nickase” version of Cas9 that has been partially deactivated generates only a single-strand break; a catalytically inactive Cas9 (“dCas9”) does not cut target DNA. In some embodiments, dCas9 binding to a DNA sequence may interfere with transcription at that site by steric hindrance. In some embodiments, dCas9 binding to an anchor sequence may interfere with (e.g., decrease or prevent) genomic complex (e.g., ASMC) formation and/or maintenance. In some embodiments, a DNA-binding domain comprises a catalytically inactive Cas9, e.g., dCas9. Many catalytically inactive Cas9 proteins

are known in the art. In some embodiments, dCas9 comprises mutations in each endonuclease domain of the Cas protein, e.g., D10A and H840A or N863A mutations. In some embodiments, a catalytically inactive or partially inactive CRISPR/Cas domain comprises a Cas protein comprising one or more mutations, e.g., one or more of the mutations listed in Table 7. In some embodiments, a Cas protein described on a given row of Table 7 comprises one, two, three, or all of the mutations listed in the same row of Table 7. In some embodiments, a Cas protein, e.g., not described in Table 7, comprises one, two, three, or all of the mutations listed in a row of Table 7 or a corresponding mutation at a corresponding site in that Cas protein.

In some embodiments, a catalytically inactive, e.g., dCas9, or partially deactivated Cas9 protein comprises a D11 mutation (e.g., D11A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H969 mutation (e.g., H969A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a N995 mutation (e.g., N995A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises mutations at one, two, or three of positions D11, H969, and N995 (e.g., D11A, H969A, and N995A mutations) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D10 mutation (e.g., a D10A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H557 mutation (e.g., a H557A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D10 mutation (e.g., a D10A mutation) and a H557 mutation (e.g., a H557A mutation) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D839 mutation (e.g., a D839A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a

H840 mutation (e.g., a H840A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a N863 mutation (e.g., a N863A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D10 mutation (e.g., D10A), a D839 mutation (e.g., D839A), a H840 mutation (e.g., H840A), and a N863 mutation (e.g., N863A) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a E993 mutation (e.g., a E993A mutation) or an analogous substitution to the amino acid corresponding to said position.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D917 mutation (e.g., a D917A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a a E1006 mutation (e.g., a E1006A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D1255 mutation (e.g., a D1255A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D917 mutation (e.g., D917A), a E1006 mutation (e.g., E1006A), and a D1255 mutation (e.g., D1255A) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D16 mutation (e.g., a D16A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D587 mutation (e.g., a D587A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a partially deactivated Cas domain has nickase activity. In some embodiments, a partially deactivated Cas9 domain is a Cas9 nickase domain. In some embodiments, the catalytically inactive Cas domain or dead Cas domain produces no detectable double strand break formation. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H588

mutation (e.g., a H588A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a N611 mutation (e.g., a N611A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D16 mutation (e.g., D16A), a D587 mutation (e.g., D587A), a H588 mutation (e.g., H588A), and a N611 mutation (e.g., N611A) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a DNA-binding domain or endonuclease domain may comprise a Cas molecule comprising or linked (e.g., covalently) to a gRNA (e.g., a template nucleic acid, e.g., template RNA, comprising a gRNA).

In some embodiments, an endonuclease domain or DNA binding domain comprises a *Streptococcus pyogenes* Cas9 (SpCas9) or a functional fragment or variant thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises a modified SpCas9. In embodiments, the modified SpCas9 comprises a modification that alters protospacer-adjacent motif (PAM) specificity. In embodiments, the PAM has specificity for the nucleic acid sequence 5′-NGT-3′. In embodiments, the modified SpCas9 comprises one or more amino acid substitutions, e.g., at one or more of positions L1111, D1135, G1218, E1219, A1322, of R1335, e.g., selected from L1111R, D1135V, G1218R, E1219F, A1322R, R1335V. In embodiments, the modified SpCas9 comprises the amino acid substitution T1337R and one or more additional amino acid substitutions, e.g., selected from L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereto. In embodiments, the modified SpCas9 comprises: (i) one or more amino acid substitutions selected from D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337; and (ii) one or more amino acid substitutions selected from L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereto.

In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas domain, e.g., a Cas9 domain. In embodiments, the endonuclease domain or DNA binding domain comprises a nuclease-active Cas domain, a Cas nickase (nCas) domain, or a nuclease-

inactive Cas (dCas) domain. In embodiments, the endonuclease domain or DNA binding domain comprises a nuclease-active Cas9 domain, a Cas9 nickase (nCas9) domain, or a nuclease-inactive Cas9 (dCas9) domain. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 domain of Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises an S. pyogenes or an S. thermophilus Cas9, or a functional fragment thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 sequence, e.g., as described in Chylinski, Rhun, and Charpentier (2013) RNA Biology 10:5, 726-737; incorporated herein by reference. In some embodiments, the endonuclease domain or DNA binding domain comprises the HNH nuclease subdomain and/or the RuvC1 subdomain of a Cas, e.g., Cas9, e.g., as described herein, or a variant thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas polypeptide (e.g., enzyme), or a functional fragment thereof. In embodiments, the Cas polypeptide (e.g., enzyme) is selected from Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (e.g., Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARF, DinG, Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12b/C2c1, Cas12c/C2c3, SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, hyper accurate Cas9 variant (HypaCas9), homologues thereof, modified or engineered versions thereof, and/or functional fragments thereof. In embodiments, the Cas9 comprises one or more substitutions, e.g., selected from H840A, D10A, P475A, W476A, N477A, D1125A, W1126A, and D1127A. In embodiments, the Cas9 comprises one or more mutations at positions selected from: D10, G12,

G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987, e.g., one or more substitutions selected from D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas (e.g., Cas9) sequence from Corynebacterium ulcerans, Corynebacterium diphtheria, Spiroplasma syrphidicola, Prevotella intermedia, Spiroplasma taiwanense, Streptococcus iniae, Belliella baltica, Psychroflexus torquis, Streptococcus thermophilus, Listeria innocua, Campylobacter jejuni, Neisseria meningitidis, Streptococcus pyogenes, or Staphylococcus aureus, or a fragment or variant thereof.

In some embodiments, the endonuclease domain or DNA binding domain comprises a Cpf1 domain, e.g., comprising one or more substitutions, e.g., at position D917, E1006A, D1255 or any combination thereof, e.g., selected from D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, and D917A/E1006A/D1255A.

In some embodiments, the endonuclease domain or DNA binding domain comprises spCas9, spCas9-VRQR(SEQ ID NO: 5019), spCas9- VRER(SEQ ID NO: 5020), xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER(SEQ ID NO: 5021), spCas9-LRKIQK(SEQ ID NO: 5022), or spCas9- LRVSQL(SEQ ID NO: 5023).

In some embodiments, a gene modifying polypeptide has an endonuclease domain comprising a Cas9 nickase, e.g., Cas9 H840A. In embodiments, the Cas9 H840A has the following amino acid sequence:

Cas9 nickase (H840A):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGN
IVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG
YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAI
LRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV
DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS
GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII

KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRE
RMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDV
DAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK
VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWD
KGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED
NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ
ID NO: 11,001)

In some embodiments, a gene modifying polypeptide comprises a dCas9 sequence comprising a D10A and/or H840A mutation, e.g., the following sequence:

SMDKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI
FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG
YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE
LHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE
EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS
DYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI
TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK

KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG
SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAE
NIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
(SEQ ID NO: 5007)

TAL Effectors and Zinc Finger Nucleases

In some embodiments, an endonuclease domain or DNA-binding domain comprises a TAL effector molecule. A TAL effector molecule, e.g., a TAL effector molecule that specifically binds a DNA sequence, typically comprises a plurality of TAL effector domains or fragments thereof, and optionally one or more additional portions of naturally occurring TAL effectors (e.g., N- and/or C-terminal of the plurality of TAL effector domains). Many TAL effectors are known to those of skill in the art and are commercially available, e.g., from Thermo Fisher Scientific.

Naturally occurring TALEs are natural effector proteins secreted by numerous species of bacterial pathogens including the plant pathogen *Xanthomonas* which modulates gene expression in host plants and facilitates bacterial colonization and survival. The specific binding of TAL effectors is based on a central repeat domain of tandemly arranged nearly identical repeats of typically 33 or 34 amino acids (the repeat-variable di-residues, RVD domain).

Members of the TAL effectors family differ mainly in the number and order of their repeats. The number of repeats typically ranges from 1.5 to 33.5 repeats and the C-terminal repeat is usually shorter in length (e.g., about 20 amino acids) and is generally referred to as a "half-repeat." Each repeat of the TAL effector generally features a one-repeat-to-one-base-pair correlation with different repeat types exhibiting different base-pair specificity (one repeat recognizes one base-pair on the target gene sequence). Generally, the smaller the number of repeats, the weaker the protein-DNA interactions. A number of 6.5 repeats has been shown to be sufficient to activate transcription of a reporter gene (Scholze et al., 2010).

Repeat to repeat variations occur predominantly at amino acid positions 12 and 13, which have therefore been termed "hypervariable" and which are responsible for the specificity of the interaction with the target DNA promoter sequence, as shown in Table 9 listing exemplary repeat variable diresidues (RVD) and their correspondence to nucleic acid base targets.

Table 9 – RVDs and Nucleic Acid Base Specificity

Target	Possible RVD Amino Acid Combinations												
A	NI	NN	CI	HI	KI								
G	NN	GN	SN	VN	LN	DN	QN	EN	HN	RH	NK	AN	FN
C	HD	RD	KD	ND	AD								
T	NG	HG	VG	IG	EG	MG	YG	AA	EP	VA	QG	KG	RG

Accordingly, it is possible to modify the repeats of a TAL effector to target specific DNA sequences. Further studies have shown that the RVD NK can target G. Target sites of TAL effectors also tend to include a T flanking the 5′ base targeted by the first repeat, but the exact mechanism of this recognition is not known. More than 113 TAL effector sequences are known to date. Non-limiting examples of TAL effectors from *Xanthomonas* include, Hax2, Hax3, Hax4, AvrXa7, AvrXa10 and AvrBs3.

Accordingly, the TAL effector domain of a TAL effector molecule described herein may be derived from a TAL effector from any bacterial species (e.g., *Xanthomonas* species such as the African strain of *Xanthomonas oryzae* pv. *Oryzae* (Yu et al. 2011), *Xanthomonas campestris* pv. *raphani* strain 756C and *Xanthomonas oryzae* pv. *oryzicola*strain BLS256 (Bogdanove et al. 2011). In some embodiments, the TAL effector domain comprises an RVD domain as well as flanking sequence(s) (sequences on the N-terminal and/or C-terminal side of the RVD domain) also from the naturally occurring TAL effector. It may comprise more or fewer repeats than the RVD of the naturally occurring TAL effector. The TAL effector molecule can be designed to target a given DNA sequence based on the above code and others known in the art. The number of TAL effector domains (e.g., repeats (monomers or modules)) and their specific sequence can beselected based on the desired DNA target sequence. For example, TAL effector domains, e.g., repeats, may be removed or added in order to suit a specific target sequence. In an embodiment, the TAL effector molecule of the present invention comprises between 6.5 and 33.5 TAL effector domains, e.g., repeats. In an embodiment, TAL effector molecule of the present invention comprises between 8 and 33.5 TAL effector domains, e.g., repeats, e.g., between 10

and 25 TAL effector domains, e.g., repeats, e.g., between 10 and 14 TAL effector domains, e.g., repeats.

In some embodiments, the TAL effector molecule comprises TAL effector domains that correspond to a perfect match to the DNA target sequence. In some embodiments, a mismatch between a repeat and a target base-pair on the DNA target sequence is permitted as along as it allows for the function of the polypeptide comprising the TAL effector molecule. In general, TALE binding is inversely correlated with the number of mismatches. In some embodiments, the TAL effector molecule of a polypeptide of the present invention comprises no more than 7 mismatches, 6 mismatches, 5 mismatches, 4 mismatches, 3 mismatches, 2 mismatches, or 1 mismatch, and optionally no mismatch, with the target DNA sequence. Without wishing to be bound by theory, in general the smaller the number of TAL effector domains in the TAL effector molecule, the smaller the number of mismatches will be tolerated and still allow for the function of the polypeptide comprising the TAL effector molecule. The binding affinity is thought to depend on the sum of matching repeat-DNA combinations. For example, TAL effector molecules having 25 TAL effector domains or more may be able to tolerate up to 7 mismatches.

In addition to the TAL effector domains, the TAL effector molecule of the present invention may comprise additional sequences derived from a naturally occurring TAL effector. The length of the C-terminal and/or N-terminal sequence(s) included on each side of the TAL effector domain portion of the TAL effector molecule can vary and be selected by one skilled in the art, for example based on the studies of Zhang et al. (2011). Zhang et al., have characterized a number of C-terminal and N-terminal truncation mutants in Hax3 derived TAL-effector based proteins and have identified key elements, which contribute to optimal binding to the target sequence and thus activation of transcription. Generally, it was found that transcriptional activity is inversely correlated with the length of N-terminus. Regarding the C-terminus, an important element for DNA binding residues within the first 68 amino acids of the Hax 3 sequence was identified. Accordingly, in some embodiments, the first 68 amino acids on the C-terminal side of the TAL effector domains of the naturally occurring TAL effector is included in the TAL effector molecule. Accordingly, in an embodiment, a TAL effector molecule comprises 1) one or more TAL effector domains derived from a naturally occurring TAL effector; 2) at least 70, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260, 270, 280 or more amino acids from the naturally occurring TAL effector on the N-terminal side of the

TAL effector domains; and/or 3) at least 68, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260 or more amino acids from the naturally occurring TAL effector on the C-terminal side of the TAL effector domains.

In some embodiments, an endonuclease domain or DNA-binding domain is or comprises a Zn finger molecule. A Zn finger molecule comprises a Zn finger protein, e.g., a naturally occurring Zn finger protein or engineered Zn finger protein, or fragment thereof. Many Zn finger proteins are known to those of skill in the art and are commercially available, e.g., from Sigma-Aldrich.

In some embodiments, a Zn finger molecule comprises a non-naturally occurring Zn finger protein that is engineered to bind to a target DNA sequence of choice. See, for example, Beerli, et al. (2002) Nature Biotechnol. 20:135-141; Pabo, et al. (2001) Ann. Rev. Biochem. 70:313-340; Isalan, et al. (2001) Nature Biotechnol. 19:656-660; Segal, et al. (2001) Curr. Opin. Biotechnol. 12:632-637; Choo, et al. (2000) Curr. Opin. Struct. Biol. 10:411-416; U.S. Pat. Nos. 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136; 7,067,317; 7,262,054; 7,070,934; 7,361,635; 7,253,273; and U.S. Patent Publication Nos. 2005/0064474; 2007/0218528; 2005/0267061, all incorporated herein by reference in their entireties.

An engineered Zn finger protein may have a novel binding specificity, compared to a naturally-occurring Zn finger protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual Zn finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, for example, U.S. Pat. Nos. 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in U.S. Pat. Nos. 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as International Patent Publication Nos. WO 98/37186; WO 98/53057; WO 00/27878; and WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger proteins has been described, for example, in International Patent Publication No. WO 02/077227.

In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Pat. Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned International Patent Publication No. WO 02/077227.

Zn finger proteins and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Pat. Nos. 6,140,0815; 789,538; 6,453,242; 6,534,261; 5,925,523; 6,007,988; 6,013,453; and 6,200,759; International Patent Publication Nos. WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536; and WO 03/016496.

In addition, as disclosed in these and other references, Zn finger proteins and/or multi-fingered Zn finger proteins may be linked together, e.g., as a fusion protein, using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Pat. Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The Zn finger molecules described herein may include any combination of suitable linkers between the individual zinc finger proteins and/or multi-fingered Zn finger proteins of the Zn finger molecule.

In certain embodiments, the DNA-binding domain or endonuclease domain comprises a Zn finger molecule comprising an engineered zinc finger protein that binds (in a sequence-specific manner) to a target DNA sequence. In some embodiments, the Zn finger molecule comprises one Zn finger protein or fragment thereof. In other embodiments, the Zn finger molecule comprises a plurality of Zn finger proteins (or fragments thereof), e.g., 2, 3, 4, 5, 6 or more Zn finger proteins (and optionally no more than 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 Zn finger proteins). In some embodiments, the Zn finger molecule comprises at least three Zn finger proteins. In some embodiments, the Zn finger molecule comprises four, five or six fingers. In some embodiments, the Zn finger molecule comprises 8, 9, 10, 11 or 12 fingers. In some embodiments, a Zn finger molecule comprising three Zn finger proteins recognizes a target DNA

sequence comprising 9 or 10 nucleotides. In some embodiments, a Zn finger molecule comprising four Zn finger proteins recognizes a target DNA sequence comprising 12 to 14 nucleotides. In some embodiments, a Zn finger molecule comprising six Zn finger proteins recognizes a target DNA sequence comprising 18 to 21 nucleotides.

In some embodiments, a Zn finger molecule comprises a two-handed Zn finger protein. Two handed zinc finger proteins are those proteins in which two clusters of zinc finger proteins are separated by intervening amino acids so that the two zinc finger domains bind to two discontinuous target DNA sequences. An example of a two handed type of zinc finger binding protein is SIP1, where a cluster of four zinc finger proteins is located at the amino terminus of the protein and a cluster of three Zn finger proteins is located at the carboxyl terminus (see Remade, et al. (1999) EMBO Journal 18(18):5073-5084). Each cluster of zinc fingers in these proteins is able to bind to a unique target sequence and the spacing between the two target sequences can comprise many nucleotides.

Linkers

In some embodiments, a gene modifying polypeptide may comprise a linker, e.g., a peptide linker, e.g., a linker as described in Table 10. In some embodiments, a gene modifying polypeptide comprises, in an N-terminal to C-terminal direction, a Cas domain (e.g., a Cas domain of Table 8), a linker of Table 10 (or a sequence having at least 70%, 80%, 85%, 90%, 95%, or 99% identity thereto), and an RT domain (e.g., an RT domain of Table 6). In some embodiments, a gene modifying polypeptide comprises a flexible linker between the endonuclease and the RT domain, e.g., a linker comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 11,002). In some embodiments, an RT domain of a gene modifying polypeptide may be located C-terminal to the endonuclease domain. In some embodiments, an RT domain of a gene modifying polypeptide may be located N-terminal to the endonuclease domain.

Table 10 Exemplary linker sequences

Amino Acid Sequence	SEQ ID NO
GGS	5101
GGSGGS	5102
GGSGGSGGS	5103

Amino Acid Sequence	SEQ ID NO
GGSGGSGGSGGS	5104
GGSGGSGGSGGSGGS	5105
GGSGGSGGSGGSGGSGGS	5106
GGGGS	5107
GGGGSGGGGS	5108
GGGGSGGGGSGGGGS	5109
GGGGSGGGGSGGGGSGGGGS	5110
GGGGSGGGGSGGGGSGGGGSGGGGS	5111
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	5112
GGG	5113
GGGG	5114
GGGGG	5115
GGGGGG	5116
GGGGGGG	5117
GGGGGGGG	5118
GSS	5119
GSSGSS	5120
GSSGSSGSS	5121
GSSGSSGSSGSS	5122
GSSGSSGSSGSSGSS	5123
GSSGSSGSSGSSGSSGSS	5124
EAAAK	5125
EAAAKEAAAK	5126
EAAAKEAAAKEAAAK	5127
EAAAKEAAAKEAAAKEAAAK	5128
EAAAKEAAAKEAAAKEAAAKEAAAK	5129
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	5130
PAP	5131
PAPAP	5132
PAPAPAP	5133
PAPAPAPAP	5134
PAPAPAPAPAP	5135
PAPAPAPAPAPAP	5136
GGSGGG	5137
GGGGGS	5138
GGSGSS	5139
GSSGGS	5140
GGSEAAAK	5141

Amino Acid Sequence	SEQ ID NO
EAAAKGGS	5142
GGSPAP	5143
PAPGGS	5144
GGGGSS	5145
GSSGGG	5146
GGGEAAAK	5147
EAAAKGGG	5148
GGGPAP	5149
PAPGGG	5150
GSSEAAAK	5151
EAAAKGSS	5152
GSSPAP	5153
PAPGSS	5154
EAAAKPAP	5155
PAPEAAAK	5156
GGSGGGGSS	5157
GGSGSSGGG	5158
GGGGGSGSS	5159
GGGGSSGGS	5160
GSSGGSGGG	5161
GSSGGGGGS	5162
GGSGGGEAAAK	5163
GGSEAAAKGGG	5164
GGGGGSEAAAK	5165
GGGEAAAKGGS	5166
EAAAKGGSGGG	5167
EAAAKGGGGGS	5168
GGSGGGPAP	5169
GGSPAPGGG	5170
GGGGGSPAP	5171
GGGPAPGGS	5172
PAPGGSGGG	5173
PAPGGGGGS	5174
GGSGSSEAAAK	5175
GGSEAAAKGSS	5176
GSSGGSEAAAK	5177
GSSEAAAKGGS	5178
EAAAKGGSGSS	5179

Amino Acid Sequence	SEQ ID NO
EAAAKGSSGGS	5180
GGSGSSPAP	5181
GGSPAPGSS	5182
GSSGGSPAP	5183
GSSPAPGGS	5184
PAPGGSGSS	5185
PAPGSSGGS	5186
GGSEAAAKPAP	5187
GGSPAPEAAAK	5188
EAAAKGGSPAP	5189
EAAAKPAPGGS	5190
PAPGGSEAAAK	5191
PAPEAAAKGGS	5192
GGGGSSEAAAK	5193
GGGEAAAKGSS	5194
GSSGGGEAAAK	5195
GSSEAAAKGGG	5196
EAAAKGGGGSS	5197
EAAAKGSSGGG	5198
GGGGSSPAP	5199
GGGPAPGSS	5200
GSSGGGPAP	5201
GSSPAPGGG	5202
PAPGGGGSS	5203
PAPGSSGGG	5204
GGGEAAAKPAP	5205
GGGPAPEAAAK	5206
EAAAKGGGPAP	5207
EAAAKPAPGGG	5208
PAPGGGEAAAK	5209
PAPEAAAKGGG	5210
GSSEAAAKPAP	5211
GSSPAPEAAAK	5212
EAAAKGSSPAP	5213
EAAAKPAPGSS	5214
PAPGSSEAAAK	5215
PAPEAAAKGSS	5216
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	5217

Amino Acid Sequence	SEQ ID NO
GGGGSEAAAKGGGGS	5218
EAAAKGGGGSEAAAK	5219
SGSETPGTSESATPES	5220
GSAGSAAGSGEF	5221
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	5222

In some embodiments, a linker of a gene modifying polypeptide comprises a motif chosen from: $(SGGS)_n$ (SEQ ID NO: 5025), $(GGGS)_n$ (SEQ ID NO: 5026), $(GGGGS)_n$ (SEQ ID NO: 5027), $(G)_n$, $(EAAAK)_n$ (SEQ ID NO: 5028), $(GGS)_n$, or $(XP)_n$.

Gene modifying polypeptide selection by pooled screening

Candidate gene modifying polypeptides may be screened to evaluate a candidate's gene editing ability. For example, an RNA gene modifying system designed for the targeted editing of a coding sequence in the human genome may be used. In certain embodiments, such a gene modifying system may be used in conjunction with a pooled screening approach.

For example, a library of gene modifying polypeptide candidates and a template guide RNA (tgRNA) may be introduced into mammalian cells to test the candidates' gene editing abilities by a pooled screening approach. In specific embodiments, a library of gene modifying polypeptide candidates is introduced into mammalian cells followed by introduction of the tgRNA into the cells.

Representative, non-limiting examples of mammalian cells that may be used in screening include HEK293T cells, U2OS cells, HeLa cells, HepG2 cells, Huh7 cells, K562 cells, or iPS cells.

A gene modifying polypeptide candidate may comprise 1) a Cas-nuclease, for example a wild-type Cas nuclease, e.g., a wild-type Cas9 nuclease, a mutant Cas nuclease, e.g., a Cas nickase, for example, a Cas9 nickase such as a Cas9 N863A nickase, or a Cas nuclease selected from **Table 7 or Table 8**, 2) a peptide linker, e.g., a sequence from **Table D or Table 10,** that may exhibit varying degrees of length, flexibility, hydrophobicity, and/or secondary structure; and 3) a reverse transcriptase (RT), e.g. an RT domain from **Table D or Table 6**. A gene modifying polypeptide candidate library comprises: a plurality of different gene modifying polypeptide candidates that differ from each other with respect to one, two or all three of the Cas nuclease, peptide linker or

RT domain components, or a plurality of nucleic acid expression vectors that encode such gene modifying polypeptide candidates.

For screening of gene modifying polypeptide candidates, a two-component system may be used that comprises a gene modifying polypeptide component and a tgRNA component. A gene modifying component may comprise, for example, an expression vector, e.g., an expression plasmid or lentiviral vector, that encodes a gene modifying polypeptide candidate, for example, comprises a human codon-optimized nucleic acid that encodes a gene modifying polypeptide candidate, e.g., a Cas-linker-RT fusion as described above. In a particular embodiment, a lentiviral cassette is utilized that comprises: (i) a promoter for expression in mammalian cells, e.g., a CMV promoter; (ii) a gene modifying library candidate, e.g. a Cas-linker-RT fusion comprising a Cas nuclease of **Table 7 or Table 8**, a peptide linker of **Table 10**, and an RT of **Table 6**, for example a Cas-linker-RT fusion as in **Table D**; (iii) a self-cleaving polypeptide, e.g., a T2A peptide; (iv) a marker enabling selection in mammalian cells, e.g., a puromycin resistance gene; and (v) a termination signal, e.g., a poly A tail.

The tgRNA component may comprise a tgRNA or expression vector, e.g., an expression plasmid, that produces the tgRNA, for example, utilizes a U6 promoter to drive expression of the tgRNA, wherein the tgRNA is a non-coding RNA sequence that is recognized by Cas and localizes it to the genomic locus of interest, and that also templates reverse transcription of the desired edit into the genome by the RT domain.

To prepare a pool of cells expressing gene modifying polypeptide library candidates, mammalian cells, e.g., HEK293T or U2OS cells, may be transduced with pooled gene modifying polypeptide candidate expression vector preparations, e.g., lentiviral preparations, of the gene modifying candidate polypeptide library. In a particular embodiment, lentiviral plasmids are utilized, and HEK293 Lenti-X cells are seeded in 15 cm plates (~$12x10^6$ cells) prior to lentiviral plasmid transfection. In such an embodiment, lentiviral plasmid transfection may be performed using the Lentiviral Packaging Mix (Biosettia) and transfection of the plasmid DNA for the gene modifying candidate library is performed the following day using Lipofectamine 2000 and Opti-MEM media according to the manufacturer's protocol. In such an embodiment, extracellular DNA may be removed by a full media change the next day and virus-containing media may be harvested 48 hours after. Lentiviral media may be concentrated using Lenti-X Concentrator (TaKaRa

Biosciences) and 5 mL lentiviral aliquots may be made and stored at -80°C. Lentiviral titering is performed by enumerating colony forming units post-selection, e.g., post Puromycin selection.

For monitoring gene editing of a target DNA, mammalian cells, e.g., HEK293T or U2OS cells, carrying a target DNA may be utilized. In other embodiments for monitoring gene editing of a target DNA, mammalian cells, e.g., HEK293T or U2OS cells, carrying a target DNA genomic landing pad may be utilized. In particular embodiments, the target DNA genomic landing pad may comprise a gene to be edited for treatment of a disease or disorder of interest. In other particular embodiments, the target DNA is a gene sequence that expresses a protein that exhibits detectable characteristics that may be monitored to determine whether gene editing has occurred. For example, in certain embodiments, a blue fluorescence protein (BFP)- or green fluorescence protein (GFP)-expressing genomic landing pad is utilized. In certain embodiments, mammalian cells, e.g., HEK293T or U2OS cells, comprising a target DNA, e.g., a target DNA genomic landing pad, are seeded in culture plates at 500x-3000x cells per gene modifying library candidate and transduced at a 0.2-0.3 multiplicity of infection (MOI) to minimize multiple infections per cell. Puromycin (2.5 ug/mL) may be added 48 hours post infection to allow for selection of infected cells. In such an embodiment, cells may be kept under puromycin selection for at least 7 days and then scaled up for tgRNA introduction, e.g., tgRNA electroporation.

To ascertain whether gene editing occurs, mammalian cells containing a target DNA to be edited may be infected with gene modifying polypeptide library candidates then transfected with tgRNA designed for use in editing of the target DNA. Subsequently, the cells may be analyzed to determine whether editing of the target locus has occurred according to the designed outcome, or whether no editing or imperfect editing has occurred, e.g., by using cell sorting and sequence analysis.

In a particular embodiment, to ascertain whether genome editing occurs, BFP- or GFP-expressing mammalian cells, e.g., HEK293T or U2OS cells, may be infected with gene modifying library candidates and then transfected or electroporated with tgRNA plasmid or RNA, e.g., by electroporation of 250,000 cells/well with 200 ng of a tgRNA plasmid designed to convert BFP-to-GFP or GFP-to-BFP, at a cell count ensuring >250x-1000x coverage per library candidate. In such an embodiment, the genome-editing capacity of the various constructs in this assay may be assessed by sorting the cells by Fluorescence-Activated Cell Sorting (FACS) for expression of the color-converted fluorescent protein (FP) at 4-10 days post-electroporation. Cells are sorted and

harvested as distinct populations of unedited cells (exhibiting original florescence protein signal), edited cells (exhibiting converted fluorescence protein signal), and imperfect edit (exhibiting no florescence protein signal) cells. A sample of unsorted cells may also be harvested as the input population to determine candidate enrichment during analysis.

To determine which gene modifying library candidates exhibit genome-editing capacity in an assay, genomic DNA (gDNA) is harvested from the sorted cell populations, and analyzed by sequencing the gene modifying library candidates in each population. Briefly, gene modifying candidates may be amplified from the genome using primers specific to the gene modifying polypeptide expression vector, e.g., the lentiviral cassette, amplified in a second round of PCR to dilute genomic DNA, and then sequenced, for example, sequenced by a next-generation sequencing platform. After quality control of sequencing reads, reads of at least about 1500 nucleotides and generally no more than about 3200 nucleotides are mapped to the gene modifying polypeptide library sequences and those containing a minimum of about an 80% match to a library sequence are considered to be successfully aligned to a given candidate for purposes of this pooled screen. In order to identify candidates capable of performing gene editing in the assay, e.g., the BFP-to-GFP or GFP-to-BFP edit, the read count of each library candidate in the edited population is compared to its read count in the initial, unsorted population.

For purposes of pooled screening, gene modifying candidates with genome-editing capacity are identified based on enrichment in the edited (converted FP) population relative to unsorted (input) cells. In some embodiments, an enrichment of at least 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or at least 100-fold over the input indicates potentially useful gene editing activity, e.g., at least 2-fold enrichment. In some embodiments, the enrichment is converted to a log-value by taking the log base 2 of the enrichment ratio. In some embodiments, a log2 enrichment score of at least 0, 1, 2, 3, 4, 5, 5.5, 6.0, 6.2, 6.3, 6.4, 6.5, or at least 6.6 indicates potentially useful gene editing activity, e.g., a log2 enrichment score of at least 1.0. In particular embodiments, enrichment values observed for gene modifying candidates may be compared to enrichment values observed under similar conditions utilizing a reference, e.g., Element ID No: 17380.

In some embodiments, multiple tgRNAs may be used to screen the gene modifying candidate library. In particular embodiments, a plurality of tgRNAs may be utilized to optimize template/Cas-linker-RT fusion pairs, e.g., for gene editing of particular target genes, for example,

gene targets for the treatment of disease. In specific embodiments, a pooled approach to screening gene modifying candidates may be performed using a multiplicity of different tgRNAs in an arrayed format.

In some embodiments, multiple types of edits, e.g., insertions, substitutions, and/or deletions of different lengths, may be used to screen the gene modifying candidate library.

In some embodiments, multiple target sequences, e.g., different fluorescent proteins, may be used to screen the gene modifying candidate library. In some embodiments, multiple target sequences, e.g., different fluorescent proteins, may be used to screen the gene modifying candidate library. In some embodiments, multiple cell types, e.g., HEK293T or U2OS, may be used to screen the gene modifying candidate library. The person of ordinary skill in the art will appreciate that a given candidate may exhibit altered editing capacity or even the gain or loss of any observable or useful activity across different conditions, including tgRNA sequence (e.g., nucleotide modifications, PBS length, RT template length), target sequence, target location, type of edit, location of mutation relative to the first-strand nick of the gene modifying polypeptide, or cell type. Thus, in some embodiments, gene modifying library candidates are screened across multiple parameters, e.g., with at least two distinct tgRNAs in at least two cell types, and gene editing activity is identified by enrichment in any single condition. In other embodiments, a candidate with more robust activity across different tgRNA and cell types is identified by enrichment in at least two conditions, e.g., in all conditions screened. For clarity, candidates found to exhibit little to no enrichment under any given condition are not assumed to be inactive across all conditions and may be screened with different parameters or reconfigured at the polypeptide level, e.g., by swapping, shuffling, or evolving domains (e.g., RT domain), linkers, or other signals (e.g., NLS).

Sequences of exemplary Cas9-linker-RT fusions

In some embodiments, a gene modifying polypeptide comprises a linker sequence and an RT sequence. In some embodiments, a gene modifying polypeptide comprises a linker sequence as listed in Table D, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises the amino acid sequence of an RT domain as listed in Table D, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises a linker sequence as listed in

Table D, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and the amino acid sequence of an RT domain as listed in Table D, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises: (i) a linker sequence as listed in a row of Table D, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and (ii) the amino acid sequence of an RT domain as listed in the same row of Table D, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

Exemplary Gene Modifying Polypeptides

In some embodiments, a gene modifying polypeptide (e.g., a gene modifying polypeptide that is part of a system described herein) comprises an amino acid sequence of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises an amino acid sequence of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 80% identity thereto. In some embodiments, a gene modifying polypeptide comprises an amino acid sequence of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 90% identity thereto. In some embodiments, a gene modifying polypeptide comprises an amino acid sequence of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 95% identity thereto. In some embodiments, a gene modifying polypeptide comprises an amino acid sequence of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises an amino acid sequence of any one of SEQ ID NOs: 1-7743. In some embodiments, a gene modifying polypeptide comprises an amino acid sequence of any one of SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises an amino acid sequence of any one of SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide comprises an amino acid sequence as listed in Table A1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide comprises an amino acid sequence as listed in Table T1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises a linker comprising a linker sequence as listed in Table T1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises an RT domain comprising an RT domain sequence as listed in Table T1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises: (i) a linker comprising a linker sequence as listed in a row of Table T1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto; and (ii) an RT domain comprising an RT domain sequence as listed in the same row of Table T1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

Table T1. Selection of exemplary gene modifying polypeptides

SEQ ID NO: for Full Polypeptide Sequence	Linker Sequence	SEQ ID NO: of linker	RT name
1372	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,401	AVIRE_P03360_3mutA
1197	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,402	FLV_P10273_3mutA
2784	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,403	MLVMS_P03355_3mutA_WS
647	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,404	SFV3L_P27401_2mutA

In some embodiments, a gene modifying polypeptide comprises an amino acid sequence as listed in Table T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises a linker comprising a linker sequence as listed in Table T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises an RT domain comprising an RT domain sequence as listed in Table T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises: (i) a linker

comprising a linker sequence as listed in a row of Table T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto; and (ii) an RT domain comprising an RT domain sequence as listed in the same row of Table T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

Table T2. Selection of exemplary gene modifying polypeptides

SEQ ID NO: for Full Polypeptide Sequence	Linker Sequence	SEQ ID NO: of linker	RT name
2311	GGGGSGGGGSGGGGSGGGGS	15,405	MLVCB_P08361_3mutA
1373	GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	15,406	AVIRE_P03360_3mutA
2644	GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	15,407	MLVMS_P03355_PLV919
2304	GSSGSSGSSGSSGSSGSS	15,408	MLVCB_P08361_3mutA
2325	EAAAKEAAAKEAAAKEAAAK	15,409	MLVCB_P08361_3mutA
2322	EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	15,410	MLVCB_P08361_3mutA
2187	PAPAPAPAPAP	15,411	MLVBM_Q7SVK7_3mut
2309	PAPAPAPAPAPAP	15,412	MLVCB_P08361_3mutA
2534	PAPAPAPAPAPAP	15,413	MLVFF_P26809_3mutA
2797	PAPAPAPAPAPAP	15,414	MLVMS_P03355_3mutA_WS
3084	PAPAPAPAPAPAP	15,415	MLVMS_P03355_3mutA_WS
2868	PAPAPAPAPAPAP	15,416	MLVMS_P03355_PLV919
126	EAAAKGGG	15,417	PERV_Q4VFZ2_3mut
306	EAAAKGGG	15,418	PERV_Q4VFZ2_3mut
1410	PAPGGG	15,419	AVIRE_P03360_3mutA
804	GGGGSSGGS	15,420	WMSV_P03359_3mut
1937	GGGGGSEAAAK	15,421	BAEVM_P10272_3mutA
2721	GGGEAAAKGGS	15,422	MLVMS_P03355_3mut
3018	GGGEAAAKGGS	15,423	MLVMS_P03355_3mut
1018	GGGEAAAKGGS	15,424	XMRV6_A1Z651_3mutA
2317	GGSGGGPAP	15,425	MLVCB_P08361_3mutA
2649	PAPGGSGGG	15,426	MLVMS_P03355_PLV919
2878	PAPGGSGGG	15,427	MLVMS_P03355_PLV919
912	GGSEAAAKPAP	15,428	WMSV_P03359_3mutA
2338	GGSPAPEAAAK	15,429	MLVCB_P08361_3mutA
2527	GGSPAPEAAAK	15,430	MLVFF_P26809_3mutA
141	EAAAKGGSPAP	15,431	PERV_Q4VFZ2_3mut
341	EAAAKGGSPAP	15,432	PERV_Q4VFZ2_3mut

2315	EAAAKPAPGGS	15,433	MLVCB_P08361_3mutA
3080	EAAAKPAPGGS	15,434	MLVMS_P03355_3mutA_WS
2688	GGGGSSEAAAK	15,435	MLVMS_P03355_PLV919
2885	GGGGSSEAAAK	15,436	MLVMS_P03355_PLV919
2810	GSSGGGEAAAK	15,437	MLVMS_P03355_3mutA_WS
3057	GSSGGGEAAAK	15,438	MLVMS_P03355_3mutA_WS
1861	GSSEAAAKGGG	15,439	MLVAV_P03356_3mutA
3056	GSSGGGPAP	15,440	MLVMS_P03355_3mutA_WS
1038	GSSPAPGGG	15,441	XMRV6_A1Z651_3mutA
2308	PAPGGGGSS	15,442	MLVCB_P08361_3mutA
1672	GGGEAAAKPAP	15,443	KORV_Q9TTC1-Pro_3mutA
2526	GGGEAAAKPAP	15,444	MLVFF_P26809_3mutA
1938	GGGPAPEAAAK	15,445	BAEVM_P10272_3mutA
2641	GSSEAAAKPAP	15,446	MLVMS_P03355_PLV919
2891	GSSEAAAKPAP	15,447	MLVMS_P03355_PLV919
1225	GSSPAPEAAAK	15,448	FLV_P10273_3mutA
2839	GSSPAPEAAAK	15,449	MLVMS_P03355_3mutA_WS
3127	GSSPAPEAAAK	15,450	MLVMS_P03355_3mutA_WS
2798	PAPGSSEAAAK	15,451	MLVMS_P03355_3mutA_WS
3091	PAPGSSEAAAK	15,452	MLVMS_P03355_3mutA_WS
1372	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,453	AVIRE_P03360_3mutA
1197	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,454	FLV_P10273_3mutA
2611	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,455	MLVMS_P03355_PLV919
2784	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,456	MLVMS_P03355_3mutA_WS
480	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,457	SFV1_P23074_2mutA
647	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,458	SFV3L_P27401_2mutA
1006	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	15,459	XMRV6_A1Z651_3mutA
2518	SGSETPGTSESATPES	15,460	MLVFF_P26809_3mutA

Subsequences of Exemplary Gene Modifying Polypeptides

In some embodiments, the gene modifying polypeptide comprises, in N-terminal to C-terminal order, one or more (e.g., 1, 2, 3, 4, 5, or all 6) of an N-terminal methionine residue, a first nuclear localization signal (NLS), a DNA binding domain, a linker, an RT domain, and/or a second NLS. In some embodiments, a gene modifying polypeptide comprises, in N-terminal to C-terminal order, a NLS (e.g., a first NLS), a DNA binding domain, a linker, and an RT domain, wherein the linker and RT domain are the linker and RT domain of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said linker and RT domain. In some embodiments, a gene modifying polypeptide comprises, in N-terminal to C-terminal order, a DNA binding domain, a linker, an RT domain, and an NLS (e.g., a second NLS) wherein the linker and RT domain are the linker and RT domain of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said linker and RT domain. In some embodiments, a gene modifying polypeptide comprises, in N-terminal to C-terminal order, a first NLS, a DNA binding domain, a linker, an RT domain, and a second NLS, wherein the linker and RT domain are the linker and RT domain of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said linker and RT domain. In some embodimetns, the gene modifying polypeptide further comprises an N-terminal methionine residue.

In some embodiments, the gene modifying polypeptide comprises, in N-terminal to C-terminal order, one or more (e.g., 1, 2, 3, 4, 5, or all 6) of an N-terminal methionine residue, a first nuclear localization signal (NLS) (e.g., of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743 and/or as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto), a DNA binding domain (e.g., a Cas domain, e.g., a SpyCas9 domain, e.g., as listed in Table 8, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto; or a DNA binding domain of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743 and/or as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto), a linker (e.g., of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743 and/or as listed in any of Tables A1, T1, or T2, or an amino acid

sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto), an RT domain (e.g., of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743 and/or as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto), and a second NLS (e.g., of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743 and/or as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto). In some embodiments, the gene modifying polypeptide further comprises (e.g., C-terminal to the second NLS) a T2A sequence and/or a puromycin sequence (e.g., of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743 and/or as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto). In some embodiments, a nucleic acid encoding a gene modifying polypeptide (e.g., as described herein) encodes a T2A sequence, e.g., wherein the T2A sequence is situated between a region encoding the gene modifying polypeptide and a second region, wherein the second region optionally encodes a selectable marker, e.g., puromycin.

In certain embodiments, the first NLS comprises a first NLS sequence of a gene modifying polypeptide having an amino acid sequence of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the first NLS comprises a first NLS sequence of a gene modifying polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the first NLS sequence comprises a C-myc NLS. In certain embodiments, the first NLS comprises the amino acid sequence PAAKRVKLD (SEQ ID NO: 11,095) , or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the gene modifying polypeptide further comprises a spacer sequence between the first NLS and the DNA binding domain. In certain embodiments, the spacer sequence between the first NLS and the DNA binding domain comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids. In certain embodiments, the spacer sequence between the first NLS and the DNA binding domain comprises the amino acid sequence GG.

In certain embodiments, the DNA binding domain comprises a DNA binding domain of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain

embodiments, the DNA binding domain comprises a DNA binding domain of a gene modifying polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the DNA binding domain comprises a Cas domain (e.g., as listed in Table 8). In certain embodiments, the DNA binding domain comprises the amino acid sequence of a SpyCas9 polypeptide (e.g., as listed in Table 8, e.g., a Cas9 N863A polypeptide), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the DNA binding domain comprises the amino acid sequence:

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFD
LAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASM
IKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG
TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP
YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL
LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS
VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH
LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL
SDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF
DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSK
LVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGK
SKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA
GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI
LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA
TLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 11,096),

or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the gene modifying polypeptide further comprises a spacer sequence between the DNA binding domain and the linker. In certain embodiments, the spacer sequence between the DNA binding domain and the linker comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids. In certain embodiments, the spacer sequence between the DNA binding domain and the linker comprises the amino acid sequence GG.

In certain embodiments, the linker comprises a linker sequence of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the linker comprises a linker sequence of a gene modifying polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the linker comprises an amino acid sequence as listed in Table D or 10, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the gene modifying polypeptide further comprises a spacer sequence between the linker and the RT domain. In certain embodiments, the spacer sequence between the linker and the RT domain comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids. In certain embodiments, the spacer sequence between the linker and the RT domain comprises the amino acid sequence GG.

In certain embodiments, the RT domain comprises a RT domain sequence of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the RT domain comprises a RT domain sequence of a gene modifying polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the RT domain comprises an amino acid sequence as listed in Table D or 6, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain has a length of about 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 amino acids.

In certain embodiments, the gene modifying polypeptide further comprises a spacer sequence between the RT domain and the second NLS. In certain embodiments, the spacer

sequence between the RT domain and the second NLS comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids. In certain embodiments, the spacer sequence between the RT domain and the second NLS comprises the amino acid sequence AG.

In certain embodiments, the second NLS comprises a second NLS sequence of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743. In certain embodiments, the second NLS comprises a second NLS sequence of a gene modifying polypeptide as listed in any of Tables A1, T1, or T2. In certain embodiments, the second NLS sequence comprises a plurality of partial NLS sequences. In embodiments, the NLS sequence, e.g., the second NLS sequence, comprises a first partial NLS sequence, e.g., comprising the amino acid sequence KRTADGSEFE (SEQ ID NO: 11,097), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In embodiments, the NLS sequence, e.g., the second NLS sequence, comprises a second partial NLS sequence. In embodiments, the NLS sequence, e.g., the second NLS sequence, comprises an SV40A5 NLS, e.g., a bipartite SV40A5 NLS, e.g., comprising the amino acid sequence KRTADGSEFESPKKKAKVE (SEQ ID NO: 11,098), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the NLS sequence, e.g., the second NLS sequence, comprises the amino acid sequence KRTADGSEFEKRTADGSEFESPKKKAKVE (SEQ ID NO: 11,099), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the gene modifying polypeptide further comprises a spacer sequence between the second NLS and the T2A sequence and/or puromycin sequence. In certain embodiments, the spacer sequence between the second NLS and the T2A sequence and/or puromycin sequence comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids. In certain embodiments, the spacer sequence between the second NLS and the T2A sequence and/or puromycin sequence comprises the amino acid sequence GSG.

Linkers and RT domains

In some embodiments, the gene modifying polypeptide comprises a linker (e.g., as described herein) and an RT domain (e.g., as described herein). In certain embodiments, the gene modifying polypeptide comprises, in N-terminal to C-terminal order, a linker (e.g., as described herein) and an RT domain (e.g., as described herein).

In certain embodiments, the linker comprises a linker sequence as listed in Table 10, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the linker comprises a linker sequence of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the linker comprises a linker sequence of any one of SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the linker comprises a linker sequence of any one of SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the linker comprises a linker sequence of an exemplary gene modifying polypeptide listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the RT domain comprises an RT domain sequence as listed in Table 6, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the RT domain comprises an RT domain sequence of an exemplary gene modifying polypeptide listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide comprises a portion of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion.

In some embodiments, a gene modifying polypeptide comprises a linker of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said linker. In some embodiments, a gene modifying polypeptide comprises a linker of a gene modifying polypeptide of any one of SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said linker. In some embodiments, a gene modifying polypeptide comprises a linker of a gene modifying polypeptide of any one of SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said linker. In some embodiments, a gene modifying polypeptide comprises a linker of a gene

modifying polypeptide as listed in any of Tables A1, T1, or T2, or a linker comprising an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide comprises an RT domain of a gene modifying polypeptide of any one of SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said RT domain. In some embodiments, a gene modifying polypeptide comprises an RT domain of a gene modifying polypeptide of any one of SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity said RT domain. In some embodiments, a gene modifying polypeptide comprises an RT domain of a gene modifying polypeptide of any one of SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity said RT domain. In some embodiments, a gene modifying polypeptide comprises an RT domain of a gene modifying polypeptide as listed in any of Tables A1, T1, or T2, or an RT domain comprising an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise the amino acid sequences of a linker and RT domain (or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto) of a gene modifying polypeptide having the amino acid sequence of any one of SEQ ID NOs: 1-7743. In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise amino acid sequences of a linker and RT domain having at least 80% identity to the linker and RT domains of any one of SEQ ID NOs: 1-7743. In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise amino acid sequences of a linker and RT domain having at least 90% identity to the linker and RT domains of any one of SEQ ID NOs: 1-7743. In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise amino acid sequences of a linker and RT domain having at least 95% identity to the linker and RT domains of any one of SEQ ID NOs: 1-7743. In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise amino acid sequences of a linker and RT domain having at least 99% identity to the linker and RT domains of any one of SEQ ID NOs: 1-7743. In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise the amino acid sequences of a linker and RT domain (or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto) of a gene

modifying polypeptide having the amino acid sequence of any one of SEQ ID NOs: 6001-7743. In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise the amino acid sequences of a linker and RT domain (or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto) of a gene modifying polypeptide having the amino acid sequence of any one of SEQ ID NOs: 4501-4541. In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise the amino acid sequences of a linker and RT domain (or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto) from a single row of any of Tables A1, T1, or T2 (e.g., from a single exemplary gene modifying polypeptide as listed in any of Tables A1, T1, or T2).

In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise the amino acid sequences of a linker and RT domain (or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto) from two different amino acid sequences selected from SEQ ID NOs: 1-7743. In certain embodiments, the linker and the RT domain of a gene modifying polypeptide comprise the amino acid sequences of a linker and RT domain (or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto) from different rows of any of Tables A1, T1, or T2.

In certain embodiments, the gene modifying polypeptide further comprises a first NLS (e.g., a 5' NLS), e.g., as described herein. In certain embodiments, the gene modifying polypeptide further comprises a second NLS (e.g., a 3' NLS), e.g., as described herein. In certain embodiments, the gene modifying polypeptide further comprises an N-terminal methionine residue.

RT Families and Mutants

In certain embodiments, a gene modifying polypeptide comprises comprises the amino acid sequence of an RT domain sequence from a family selected from: AVIRE, BAEVM, FFV, FLV, FOAMV, GALV, KORV, MLVAV, MLVBM, MLVCB, MLVFF, MLVMS, PERV, SFV1, SFV3L, WMSV, XMRV6, BLVAU, BLVJ, HTL1A, HTL1C, HTL1L, HTL32, HTL3P, HTLV2, JSRV, MLVF5, MLVRD, MMTVB, MPMV, SFVCP, SMRVH, SRV1, SRV2, and WDSV. In certain embodiments, a gene modifying polypeptide comprises comprises the amino acid sequence of an RT domain sequence from a family selected from: AVIRE, BAEVM, FFV,

FLV, FOAMV, GALV, KORV, MLVAV, MLVBM, MLVCB, MLVFF, MLVMS, PERV, SFV1, SFV3L, WMSV, and XMRV6.

In certain embodiments, a gene modifying polypeptide comprises comprises the amino acid sequence of an RT domain sequence from an MLVMS RT domain. In embodiments, the amino acid sequence of an RT domain sequence comprises one or more point mutations as listed in column 1 of Table M1, or a point mutation corresponding thereto. In embodiments, the amino acid sequence of an RT domain sequence comprises one or more point mutations as listed in column 3 of Table M1 (Gen1 MLVMS), or a point mutation corresponding thereto. In embodiments, the amino acid sequence of an RT domain sequence comprises one or more point mutations at an amino acid position of the RT domain as listed in columns 1 and 2 of Table M2, or an amino acid position corresponding thereto.

In certain embodiments, a gene modifying polypeptide comprises comprises the amino acid sequence of an RT domain sequence from an AVIRE RT domain. In embodiments, the amino acid sequence of an RT domain sequence comprises one or more point mutations as listed in column 2 of Table M1, or a point mutation corresponding thereto. In embodiments, the amino acid sequence of an RT domain sequence comprises one or more point mutations as listed in column 4 of Table M1 (Gen2 AVIRE), or a point mutation corresponding thereto. In embodiments, the amino acid sequence of an RT domain sequence comprises one or more point mutations at an amino acid position of the RT domain as listed in columns 3 and 4 of Table M2, or an amino acid position corresponding thereto. In certain embodiments, the RT domain comprises an IENSSP (e.g., at the C-terminus).

Table M1. Exemplary point mutations in MLVMS and AVIRE RT domains

RT-linker filing (MLVMS)	Corresponding AVIRE	Gen1 MLVMS (PLV4921)	Gen2 AVIRE (PLV10990)
		H8Y	
P51L	Q51L		
S67R	T67R		
E67K	E67K		
E69K	E69K		
T197A	T197A		
D200N	D200N	D200N	D200N
H204R	N204R		

E302K	E302K		
		T306K	T306K
F309N	Y309N		
W313F	W313F	W313F	W313F
T330P	G330P	T330P	G330P
L435G	T436G		
N454K	N455K		
D524G	D526G		
E562Q	E564Q		
D583N	D585N		
H594Q	H596Q		
L603W	L605W	L603W	L605W
D653N	D655N		
L671P	L673P		
		IENSSP at C-term	

Table M2. Positions that can be mutated in exemplary MLVMS and AVIRE RT domains

WT residue & position			
MLVMS aa	**MLVMS position # ***	**AVIRE aa**	**AVIRE position # ***
H	8	Y	8
P	51	Q	51
S	67	T	67
E	69	E	69
T	197	T	197
D	200	D	200
H	204	N	204
E	302	E	302
T	306	T	306
F	309	Y	309
W	313	W	313
T	330	G	330
L	435	T	436
N	454	N	455
D	524	D	526
E	562	E	564
D	583	D	585

H	594	H	596
L	603	L	605
D	653	D	655
L	671	S	673

In certain embodiments, a gene modifying polypeptide comprises a gamma retrovirus derived RT domain. In certain embodiments, the gamma retrovirus-derived RT domain of a gene modifying polypeptide comprises the amino acid sequence of an RT domain sequence from a family selected from: AVIRE, BAEVM, FFV, FLV, FOAMV, GALV, KORV, MLVAV, MLVBM, MLVCB, MLVFF, MLVMS, PERV, SFV1, SFV3L, WMSV, and XMRV6. In some embodiments, the gamma retrovirus-derived RT domain of a gene modifying polypeptide is not derived from PERV. In some embodiments, said RT includes one, two, three, four, five, six or more mutations shown in Table 2 and corresponding to mutations D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K, or D653N in the RT domain of murine leukemia virus reverse transcriptase. In some embodiments, the gene modifying polypeptide further comprises a linker having at least 99% identity to a linker domains of any one of SEQ ID NOs: 1-7743. In some embodiments, the gene modifying polypeptide further comprises a linker having at least 99% or 100% identity to SEQ ID NO: 5217 or SEQ ID NO:11,041.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of an AVIRE RT (e.g., an AVIRE_P03360 sequence, e.g., SEQ ID NO: 8001), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of an AVIRE RT further comprising one, two, three, four, or five mutations selected from the group consisting of D200N, G330P, L605W, T306K, and W313F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an AVIRE RT further comprising one, two, or three mutations selected from the group consisting of D200N, G330P, and L605W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a BAEVM RT (e.g., an BAEVM_P10272 sequence, e.g., SEQ ID NO: 8004), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a BAEVM RT further

comprising one, two, three, four, or five mutations selected from the group consisting of D198N, E328P, L602W, T304K, and W311F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a BAEVM RT further comprising one, two, or three mutations selected from the group consisting of D198N, E328P, and L602W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of an FFV RT (e.g., an FFV_O93209 sequence, e.g., SEQ ID NO: 8012), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of an FFV RT further comprising one, two, three, or four mutations selected from the group consisting of D21N, T293N, T419P, and L393K, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an FFV RT further comprising one, two, or three mutations selected from the group consisting of D21N, T293N, and T419P, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an FFV RT further comprising the mutation D21N. In some embodiments, the RT domain comprises the amino acid sequence of an FFV RT further comprising one, two, or three mutations selected from the group consisting of T207N, T333P, and L307K, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an FFV RT further comprising one or two mutations selected from the group consisting of T207N and T333P, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of an FLV RT (e.g., an FLV_P10273 sequence, e.g., SEQ ID NO: 8019), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of an FLV RT further comprising one, two, three, or four mutations selected from the group consisting of D199N, L602W, T305K, and W312F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an FLV RT further comprising one or two mutations selected from the group consisting of D199N and L602W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a FOAMV RT (e.g., an FOAMV_P14350 sequence, e.g., SEQ ID NO: 8021), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of an FOAMV RT further comprising one, two, three, or four mutations selected from the group consisting of D24N, T296N, S420P, and L396K, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an FOAMV RT further comprising one, two, or three mutations selected from the group consisting of D24N, T296N, and S420P, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an FOAMV RT further comprising the mutation D24N, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an FOAMV RT further comprising one, two, or three mutations selected from the group consisting of T207N, S331P, and L307K, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of an FOAMV RT further comprising one or two mutations selected from the group consisting of T207N and S331P, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a GALV RT (e.g., an GALV_P21414 sequence, e.g., SEQ ID NO: 8027), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a GALV RT further comprising one, two, three, four, or five mutations selected from the group consisting of D198N, E328P, L600W, T304K, and W311F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a GALV RT further comprising one, two, or three mutations selected from the group consisting of D198N, E328P, and L600W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a KORV RT (e.g., an KORV_Q9TTC1 sequence, e.g., SEQ ID NO: 8047), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a GALV RT further comprising one, two, three, four, five, or six mutations selected from the group consisting of

D32N, D322N, E452P, L274W, T428K, and W435F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a GALV RT further comprising one, two, three, or four mutations selected from the group consisting of D32N, D322N, E452P, and L274W, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a GALV RT further comprising the mutation D32N. In some embodiments, the RT domain comprises the amino acid sequence of a KORV RT further comprising one, two, three, four, or five mutations selected from the group consisting of D231N, E361P, L633W, T337K, and W344F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a KORV RT further comprising one, two, or three mutations selected from the group consisting of D231N, E361P, and L633W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a MLVAV RT (e.g., an MLVAV_P03356 sequence, e.g., SEQ ID NO: 8053), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a MLVAV RT further comprising one, two, three, four, or five mutations selected from the group consisting of D200N, T330P, L603W, T306K, and W313F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a MLVAV RT further comprising one, two, or three mutations selected from the group consisting of D200N, T330P, and L603W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a MLVBM RT (e.g., an MLVBM_Q7SVK7 sequence, e.g., SEQ ID NO: 8056), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a MLVBM RT further comprising one, two, three, four, or five mutations selected from the group consisting of D199N, T329P, L602W, T305K, and W312F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a MLVBM RT further comprising one, two, and three mutations selected from the group consisting of D200N, T330P, and L603W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a MLVCB RT (e.g., an MLVCB_P08361 sequence, e.g., SEQ ID NO: 8062), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a MLVCB RT further comprising one, two, three, four, or five mutations selected from the group consisting of D200N, T330P, L603W, T306K, and W313F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a MLVCB RT further comprising one, two, and three mutations selected from the group consisting of D200N, T330P, and L603W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a MLVFF RT, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a MLVFF RT further comprising one, two, three, four, or five mutations selected from the group consisting of D200N, T330P, L603W, T306K, and W313F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a MLVFF RT further comprising one, two, and three mutations selected from the group consisting of D200N, T330P, and L603W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a MLVMS RT (e.g., an MLVMS_reference sequence, e.g., SEQ ID NO: 8137; or an MLVMS_P03355 sequence, e.g., SEQ ID NO: 8070), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a MLVMS RT further comprising one, two, three, four, five, or six mutations selected from the group consisting of D200N, T330P, L603W, T306K, W313F, and H8Y, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a MLVMS RT further comprising one, two, three, four, or five mutations selected from the group consisting of D200N, T330P, L603W, T306K, and W313F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a MLVMS RT further comprising one, two, or three mutations selected from the group consisting of D200N, T330P, and L603W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a PERV RT (e.g., an PERV_Q4VFZ2 sequence, e.g., SEQ ID NO: 8099), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a PERV RT further comprising one, two, three, four, or five mutations selected from the group consisting of D196N, E326P, L599W, T302K, and W309F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a PERV RT further comprising one, two, or three mutations selected from the group consisting of D196N, E326P, and L599W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a SFV1 RT (e.g., an SFV1_P23074 sequence, e.g., SEQ ID NO: 8105), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a SFV1 RT further comprising one, two, three, or four mutations selected from the group consisting of D24N, T296N, N420P, and L396K, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a SFV1 RT further comprising one, two, or three mutations selected from the group consisting of D24N, T296N, and N420P, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a SFV1 RT further comprising the D24N, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a SFV3L RT (e.g., an SFV3L_P27401 sequence, e.g., SEQ ID NO: 8111), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a SFV3L RT further comprising one, two, three, or four mutations selected from the group consisting of D24N, T296N, N422P, and L396K, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a SFV3L RT further comprising one, two, or three mutations selected from the group consisting of D24N, T296N, and N422P, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a SFV3L RT further comprising the mutation D24N, or a corresponding position in a homologous RT domain. In some embodiments, the RT

domain comprises the amino acid sequence of a SFV3L RT further comprising one, two, or three mutations selected from the group consisting of T307N, N333P, and L307K, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a SFV3L RT further comprising one or two mutations selected from the group consisting of T307N and N333P, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a WMSV RT (e.g., an WMSV_P03359 sequence, e.g., SEQ ID NO: 8131), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a WMSV RT further comprising one, two, three, four, or five mutations selected from the group consisting of D198N, E328P, L600W, T304K, and W311F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a WMSV RT further comprising one, two, or three mutations selected from the group consisting of D198N, E328P, and L600W, or a corresponding position in a homologous RT domain.

In embodiments, the RT domain comprises the amino acid sequence of an RT domain of a XMRV6 RT (e.g., an XMRV6_A1Z651 sequence, e.g., SEQ ID NO: 8134), or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the RT domain comprises the amino acid sequence of a XMRV6 RT further comprising one, two, three, four, or five mutations selected from the group consisting of D200N, T330P, L603W, T306K, and W313F, or a corresponding position in a homologous RT domain. In some embodiments, the RT domain comprises the amino acid sequence of a XMRV6 RT further comprising one, two, or three mutations selected from the group consisting of D200N, T330P, and L603W, or a corresponding position in a homologous RT domain.

In certain embodiments, the RT domain of a gene modifying polypeptide comprises the amino acid sequence of an RT domain of an AVIRE RT, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In embodiments, the RT domain comprises the amino acid sequence of an RT domain comprised in a sequence listed in column 1 of Table A5, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the gene modifying polypeptide further comprises a linker having at least 99% or 100% identity to SEQ ID NO: 5217 or SEQ ID NO:11,041.

In certain embodiments, the RT domain of a gene modifying polypeptide comprises the amino acid sequence of an RT domain of an MLVMS RT, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In embodiments, the RT domain comprises the amino acid sequence of an RT domain comprised in a sequence listed in any of columns 2-6 of Table A5, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the gene modifying polypeptide further comprises a linker having at least 99% or 100% identity to SEQ ID NO: 5217 or SEQ ID NO:11,041.

Table A5. Exemplary gene modifying polypeptides comprising an AVIRE RT domain or an MLVMS RT domain.

AVIRE SEQ ID NOs:	MLVMS SEQ ID NOs:				
1	2704	3007	3038	2638	2930
2	2706	3007	3038	2639	2930
3	2708	3008	3039	2639	2931
4	2709	3008	3039	2640	2931
5	2709	3009	3040	2640	2932
6	2710	3010	3040	2641	2932
7	2957	3010	3041	2641	2933
9	2957	3011	3041	2642	2933
10	2958	3012	3042	2642	2934
12	2959	3012	3042	2643	2934
13	2960	3013	3043	2643	2935
14	2962	3013	3043	2644	2935
6076	6042	3014	3044	2644	2936
6143	6068	3014	3044	2645	2936
6200	6097	3015	3045	2645	2937
6254	6136	3015	3045	2646	2937
6274	6156	3016	3046	2646	2938
6315	6215	3016	3046	2647	2938
6328	6216	3017	3047	2647	2939
6337	6301	3018	3047	2648	2939
6403	6352	3018	3048	2648	2940
6420	6365	3019	3048	2649	2940
6440	6411	3019	3049	2649	2941
6513	6436	3020	3049	2650	2941
6552	6458	3020	3050	2650	2942
6613	6459	3021	3051	2651	2942

6671	6524	3021	3051	2651	2943
6822	6562	3022	3052	2652	2943
6840	6563	3023	3052	2652	2944
6884	6699	3023	3053	2653	2945
6907	6865	3024	3053	2653	2945
6970	7022	3024	3054	2654	2946
7025	7037	3025	3054	2655	2946
7052	7088	3025	3055	2655	2947
7078	7116	3026	3055	2656	2947
7243	7175	3026	3056	2656	2948
7253	7200	3027	3056	2657	2948
7318	7206	3027	3057	2657	2949
7379	7277	3028	3057	2658	2949
7486	7294	3028	3058	2658	2950
7524	7330	3029	3058	2659	2950
7668	7411	3030	3059	2659	2951
7680	7455	3030	3059	2660	2951
7720	7477	3031	3060	2660	2952
1137	7511	3031	3060	2661	2952
1138	7538	3032	3061	2661	2953
1139	7559	3032	3061	2662	2953
1140	7560	3033	3062	2662	2954
1141	7593	3033	3062	2663	2954
1142	7594	3034	3063	2663	2955
1143	7607	3034	3063	2664	2955
1144	7623	6025	3064	2664	6485
1145	7638	6041	3064	2665	6486
1146	7717	6043	3065	2665	6504
1147	7731	6098	3065	2666	6505
1148	7732	6099	3066	2666	6595
1149	2711	6180	3066	2667	6596
1150	2711	6182	3067	2667	6751
1151	2712	6237	3067	2668	6752
1152	2712	6238	3068	2668	6777
1153	2713	6311	3068	2669	6778
1154	2713	6312	3069	2669	7172
1155	2714	6578	3069	2670	7174
1156	2714	6579	3070	2670	7313
1157	2715	6663	3070	2671	7314
1158	2715	6664	3071	2671	
1159	2716	6708	3071	2672	

1160	2716	6709	3072	2672	
1161	2717	6809	3072	2673	
1162	2717	6831	3073	2673	
1163	2718	6832	3073	2674	
1164	2718	6864	3074	2674	
1165	2719	6866	3074	2675	
1166	2719	7089	3075	2675	
1167	2720	7157	3075	2676	
6015	2720	7159	3076	2676	
6029	2721	7173	3076	2677	
6045	2721	7176	3077	2677	
6077	2722	7293	3077	2678	
6129	2722	7295	3078	2678	
6144	2723	7343	3078	2679	
6164	2723	7393	3079	2680	
6201	2724	7394	3079	2680	
6227	2724	7425	3080	2681	
6244	2725	7426	3080	2681	
6250	2725	7444	3081	2682	
6264	2726	7445	3081	2682	
6289	2726	7476	3082	2683	
6304	2727	7478	3082	2683	
6316	2727	7496	3083	2684	
6384	2728	7497	3083	2684	
6421	2728	7537	3084	2685	
6441	2729	7539	3084	2685	
6492	2729	2780	3085	2686	
6514	2730	2780	3085	2686	
6530	2730	2781	3086	2687	
6569	2731	2781	3086	2687	
6584	2731	2782	3087	2688	
6621	2732	2782	3087	2688	
6651	2732	2783	3088	2689	
6659	2733	2783	3088	2689	
6683	2734	2784	3089	2690	
6703	2734	2784	3089	2690	
6727	2735	2785	3090	2691	
6732	2735	2785	3090	2692	
6745	2736	2786	3091	2692	
6755	2736	2786	3091	2693	
6784	2737	2787	3092	2693	

6817	2737	2787	3092	2694	
6823	2738	2788	3093	2694	
6841	2739	2788	3093	2695	
6871	2740	2789	3094	2695	
6885	2740	2789	3095	2696	
6898	2741	2790	3095	2696	
6908	2741	2790	3096	2697	
6933	2742	2791	3096	2697	
6971	2742	2791	3097	2698	
7009	2743	2792	3097	2698	
7018	2743	2792	3098	2699	
7045	2744	2793	3098	2699	
7053	2744	2793	3099	2700	
7068	2745	2794	3099	2700	
7079	2745	2794	3100	2701	
7096	2746	2795	3100	2701	
7104	2746	2795	3101	2702	
7122	2747	2796	3101	2702	
7151	2747	2796	3102	2703	
7163	2748	2797	3102	2703	
7181	2748	2797	3103	2862	
7244	2749	2798	3103	2862	
7273	2750	2798	3104	2863	
7319	2750	2799	3104	2863	
7336	2751	2799	3105	2864	
7380	2751	2800	3105	2864	
7402	2752	2800	3106	2865	
7462	2752	2801	3106	2865	
7487	2753	2801	3107	2866	
7525	2753	2802	3107	2866	
7569	2754	2802	3108	2867	
7626	2754	2803	3108	2867	
7689	2755	2803	3109	2868	
7707	2755	2804	3109	2868	
7721	2756	2804	3110	2869	
1371	2756	2805	3110	2869	
1372	2757	2805	3111	2870	
1373	2758	2806	3111	2870	
1374	2758	2806	3112	2871	
1375	2759	2807	3112	2871	
1376	2759	2807	3113	2872	

1377	2760	2808	3113	2872	
1378	2760	2808	3114	2873	
1379	2761	2809	3114	2873	
1380	2761	2809	3115	2874	
1381	2762	2810	3115	2874	
1382	2762	2810	3116	2875	
1383	2763	2811	3116	2875	
1384	2763	2811	3117	2876	
1385	2764	2812	3117	2876	
1386	2764	2812	3118	2877	
1387	2765	2813	3118	2877	
1388	2765	2813	3119	2878	
1389	2766	2814	3119	2878	
1390	2766	2814	3120	2879	
1391	2767	2815	3120	2879	
1392	2767	2815	3121	2880	
1393	2768	2816	3121	2880	
1394	2768	2816	3122	2881	
1395	2769	2817	3122	2881	
1396	2769	2817	3123	2882	
1397	2770	2818	3123	2882	
1398	2770	2818	3124	2883	
1399	2771	2819	3124	2883	
1400	2771	2819	3125	2884	
1401	2772	2820	3125	2884	
1402	2773	2820	3126	2885	
1403	2773	2821	3126	2885	
1404	2774	2821	3127	2886	
1405	2774	2822	3127	2886	
1406	2775	2822	3128	2887	
1407	2775	2823	3128	2887	
1408	2776	2823	3129	2888	
1409	2776	2824	3129	2888	
1410	2777	2824	3130	2889	
1411	2777	2825	3130	2889	
1412	2778	2825	3131	2890	
1413	2779	2826	3131	2890	
1414	2779	2826	3132	2891	
1415	2965	2827	3133	2891	
1416	2965	2827	3133	2892	
1417	2966	2828	3134	2893	

1418	2966	2828	3134	2893	
1419	2967	2829	3135	2894	
1420	2968	2829	3135	2894	
1421	2968	2830	3136	2895	
1422	2969	2830	3136	2895	
1423	2969	2831	6181	2896	
1424	2970	2831	6183	2896	
1425	2970	2832	6284	2897	
1426	2971	2832	6285	2897	
1427	2971	2833	6760	2898	
1428	2972	2833	6761	2898	
1429	2972	2834	7036	2899	
1430	2973	2834	7038	2899	
1431	2974	2835	7158	2900	
1432	2974	2835	7160	2900	
1433	2975	2836	2610	2901	
1434	2976	2836	2610	2901	
1435	2976	2837	2611	2902	
1436	2977	2837	2611	2902	
1437	2977	2838	2612	2903	
1439	2978	2838	2612	2903	
1440	2978	2839	2613	2904	
1441	2979	2839	2613	2904	
1442	2979	2840	2614	2905	
1443	2980	2840	2614	2905	
1444	2980	2841	2615	2906	
1445	2981	2841	2615	2906	
1446	2981	2842	2616	2907	
1447	2982	2842	2616	2907	
6001	2982	2843	2617	2908	
6030	2983	2843	2617	2908	
6078	2983	2844	2618	2909	
6108	2984	2844	2618	2909	
6130	2985	2845	2619	2910	
6165	2985	2845	2619	2910	
6265	2986	2846	2620	2911	
6275	2987	2846	2620	2911	
6305	2987	2847	2621	2912	
6329	2988	2847	2621	2912	
6370	2988	2848	2622	2913	
6385	2989	2848	2622	2913	

6404	2989	2849	2623	2914	
6531	2990	2849	2623	2914	
6585	2990	2850	2624	2915	
6622	2991	2850	2624	2915	
6652	2991	2851	2625	2916	
6733	2992	2851	2625	2916	
6756	2992	2852	2626	2917	
6765	2993	2852	2626	2917	
6798	2993	2853	2627	2918	
6824	2994	2853	2627	2919	
6972	2994	2854	2628	2919	
7046	2995	2854	2628	2920	
7054	2995	2855	2629	2920	
7069	2996	2855	2629	2921	
7080	2996	2856	2630	2921	
7105	2997	2856	2630	2922	
7123	2998	2857	2631	2922	
7143	2998	2857	2631	2923	
7152	2999	2858	2632	2923	
7204	2999	2858	2632	2924	
7320	3001	2859	2633	2924	
7351	3001	2859	2633	2925	
7381	3002	2860	2634	2925	
7403	3002	2860	2634	2926	
7438	3003	2861	2635	2926	
7488	3003	2861	2635	2927	
7500	3004	3035	2636	2927	
7526	3004	3036	2636	2928	
7588	3005	3036	2637	2928	
7612	3005	3037	2637	2929	
7627	3006	3037	2638	2929	

Systems

In an aspect, the disclosure relates to a system comprising nucleic acid molecule encoding a gene modifying polypeptide (e.g., as described herein) and a template nucleic acid (e.g., a template RNA, e.g., as described herein). In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises one or more silent mutations in the coding region (e.g., in the sequence encoding the RT domain) relative to a nucleic acid molecule as described herein. In certain embodiments, the system further comprises a gRNA

(e.g., a gRNA that binds to a polypeptide that induces a nick, e.g., in the opposite strand of the target DNA bound by the gene modifying polypeptide).

In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide encodes a polypeptide having an amino acid sequence selected from SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide encodes a polypeptide having an amino acid sequence selected from SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide encodes a polypeptide having an amino acid sequence selected from SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide encodes a polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding a portion of an amino acid sequence selected from SEQ ID NOs: 1-7743, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding a portion of an amino acid sequence selected from SEQ ID NOs: 6001-7743, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding a portion of an amino acid sequence selected from SEQ ID NOs: 4501-4541, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding a portion of a polypeptide listed in any of Tables A1, T1, or T2, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion.

In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding the linker of an amino acid sequence selected from SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding the linker of a polypeptide having an amino acid sequence selected from SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding the linker of a polypeptide having an amino acid sequence selected from SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding the linker of a polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding the RT domain of an amino acid sequence selected from SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding the RT domain of a polypeptide having an amino acid sequence selected from SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding the RT domain of a polypeptide having an amino acid sequence selected from SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the nucleic acid molecule encoding the gene modifying polypeptide comprises a sequence encoding the RT domain of a polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In an aspect, the disclosure relates to a system comprising a gene modifying polypeptide (e.g., as described herein) and a template nucleic acid (e.g., a template RNA, e.g., as described herein).

In certain embodiments, the gene modifying polypeptide comprises a polypeptide having an amino acid sequence selected from SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises a polypeptide having an amino acid sequence selected from SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises a polypeptide having an amino acid sequence selected from SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises a polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the gene modifying polypeptide comprises a portion of an amino acid sequence selected from SEQ ID NOs: 1-7743, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion. In certain embodiments, the gene modifying polypeptide comprises a portion of an amino acid sequence selected from SEQ ID NOs: 6001-7743, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion. In certain embodiments, the gene modifying polypeptide comprises a portion of an amino acid sequence selected from SEQ ID NOs: 4501-4541, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion. In certain embodiments, the gene modifying polypeptide comprises a portion of a polypeptide listed in any of Tables A1, T1, or T2, wherein the portion comprises a linker and RT domain, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity to said portion.

In certain embodiments, the gene modifying polypeptide comprises the linker of an amino acid sequence selected from SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises a sequence encoding the linker of a polypeptide having an amino acid sequence selected from SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises a sequence encoding the linker of a

polypeptide having an amino acid sequence selected from SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises the linker of a polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In certain embodiments, the gene modifying polypeptide comprises the RT domain of an amino acid sequence selected from SEQ ID NOs: 1-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises a sequence encoding the RT domain of a polypeptide having an amino acid sequence selected from SEQ ID NOs: 6001-7743, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises a sequence encoding the RT domain of a polypeptide having an amino acid sequence selected from SEQ ID NOs: 4501-4541, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto. In certain embodiments, the gene modifying polypeptide comprises the RT domain of a polypeptide as listed in any of Tables A1, T1, or T2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, or 99% identity thereto.

Table A1. Exemplary amino acid sequences for gene modifying polypeptides comprising an RT domain and a linker sequence

SEQ ID NO:	Amino Acid Sequence
34	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGSSGGLDDEYRLYS
	PLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQ
	DLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQ
	HPQVTLLQYVDDLLLAGATKQDCLEGTKALLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLF
	IPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAI
	AAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
	LTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININTYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKN
	KEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
35	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK

	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGSPAPGGLDDEYR LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE IKNKEEILSLLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
35	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
36	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE

	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAA
	AKALEAEAAAKEAAAKEAAAKAGGLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIR
	PHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP
	GTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRD
	GQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKG
	VARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPA
	ALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEG
	KSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADG
	SEFEKRTADGSEFESPKKKAKVE
36	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAA
	AKALEAEAAAKEAAAKEAAAKAGGLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIR
	PHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP
	GTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRD
	GQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKG
	VARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPA
	ALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEG
	KSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADG
	SEFEKRTADGSEFESPKKKAKVE
37	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA

	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVVDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKEAAAKGGLD
	DEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGT
	NDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRD
	LANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGK
	AGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGW
	PVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVR
	KDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTS
	AGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
38	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVVDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGSGSSPAPGGLDDEYRLY
	SPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPV
	QDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRI
	QHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRL
	FIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKA
	IAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
	PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIK
	NKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
39	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK

	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKEAAA
	KEAAAKGGLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNT
	PLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSP
	TIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTT
	AKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLS
	KKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCH
	QLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGA
	IYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAK
	VE
40	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSSGGGEAAAKGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
40	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA

	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSSGGGEAAAKGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
41	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSSGGSSGSSGGGLDDEY
	RLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDY
	RPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLAN
	FRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGF
	CRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVC
	LKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDL
	TDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGR
	EIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
41	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ

	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGSGSGSGSGGLDDEY
	RLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDY
	RPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRI
	FRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRL
	CRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVC
	LKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDL
	TDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGR
	EIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
43	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGSGSGSGSGGLDDEY
	RLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDY
	RPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRI
	FRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRL
	CRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVC
	LKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDL
	TDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGR
	EIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
47	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL

	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSPAPGGLDDEYRLYSPL
	VKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGPTNDYRPVQ
	REVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQ
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSPAPGGLDDEYRLYSPL
	VKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGPTNDYRPVQ
	REVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQ
	HPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLF
	IPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAI
	AAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
	LTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKN
	KEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
48	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSPAPGGLDDEYRLYSPL
	PLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGPTNDYRPVQ
	DLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQ
	HPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLF
	IPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAI
	AAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
	LTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKN
	KEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
49	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL

	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGSGGGSGGGSGGLD
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSILRDGQRWLTEARKETVVQIPAPTTAKQVREFLGTA
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
51	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGSGGGSGGGSGGLD
	DEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGT
	NDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRD
	LANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSILRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGK
	AGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGW
	PVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVR
	KDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGAIYKQRGWLTS
	AGREIKNKEEILSLLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
62	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV

	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGGEAAAKGGSGGLDDEYR
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
65	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE

83	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGSSGGSGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGV
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFESPKKKAKVE
90	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGSETPGTSESATPESGGL
	DDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPG
	TNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHR
	DLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLG
	KAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASG
	WPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGV

	RKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLT
	SAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
97	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGGGSEAAAKGGTL
	QLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRK
	PGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEAL
	HRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREF
	LGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVA
	SGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEET
	GVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGL
	LTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
112	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGGPAPGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP

	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
113	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGSGSGSGSGSGSSG
	GTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLP
	VRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFN
	EALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQV
	REFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLD
	PVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLI
	EETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQ
	RGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
113	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGSGSGSGSGSSG
	GTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLP
	VRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFN
	EALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQV

	REFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLD
	PVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLI
	EETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQ
	RGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
117	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKEAAA
	KEAAAKEAAAKGGTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILV
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKEAAA
	KEAAAKEAAAKGGTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILV

	PVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRL PQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTV VQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPW RRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETD EPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAF ATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFE SPKKKAKVE
121	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
121	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK

	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGGSSEAAAKGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
122	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	LQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVR
	KPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEA
	LHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVRE
	FLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPV
	ASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEE
	TGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRG
	WLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
123	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE

	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGPAPGSSGGTLQLDDEYRLY
	SPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPV
	QDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRI
	QHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRL
	WIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKA
	IAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
	PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIK
	NKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
124	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGPAPEAAAKGGGGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
126	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM

	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGGGGTLQLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFC
	RLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
127	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGPAPGGGEAAAKGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
133	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL

	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSEAAAKGGGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
138	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGPAPGGSEAAAKGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
139	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA

	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGSSEAAAKGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFESPKKKAKVE
140	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGSSGGSSSGSETPGTSES
	ATPESSGGSSGSSGGSSGGTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQG
	ILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTW
	TRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARK
	KTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTL
	GPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPE
	ETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSR
	YAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGS
	EFESPKKKAKVE
141	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA

	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGSPAPGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKETVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
142	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGGSGGSGGSGGSGSG
	GTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLP
	VRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFN
	EALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQV
	REFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLD
	PVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLI
	EETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQ
	RGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
142	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA

	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGGSGGSGGSGGSG
	GTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLP
	VRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFN
	EALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQV
	REFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLD
	PVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLI
	EETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQ
	RGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
144	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGGSGGSGGSGGSG
	RLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDY
	RPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLAN
	FRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGF
	CRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVC
	LKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDL
	TDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGR
	EIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
147	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ

	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGGSGGTLQLDDEYRLYS
	PLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQ
	DLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQ
	HPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLW
	IPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAI
	AAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
	LTGEVLTWFTDGSSYVVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKN
	KEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
151	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGGSGGTLQLDDEYRLYS
	PLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQ
	DLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQ
	HPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLW
	IPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAI
	AAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
	LTGEVLTWFTDGSSYVVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKN
	KEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
156	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL

	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAAA
	AKALEAEAAAKEAAAKEAAAKEAAAKAGGTLQLDDEYRLVSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQE
	RKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFA
	PPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRL
	AEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTA
	DGSEFEKRTADGSEFESPKKKAKVE
156	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAAA AKALEAEAAAKEAAAKEAAAKEAAAKAGGTLQLDDEYRLVSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQE GIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYS LRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDE RKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFA PPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRL AEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTA DGSEFEKRTADGSEFESPKKKAKVE

157	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGSPAPEAAAKGGTLQLDD TAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASG WPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK RKDLTDIPLTGEVLTWFTDGSSYVVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGAIYKQRGWLT SAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
164	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGSPAPEAAAKGGTLQLDD EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK

	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
168	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKPAPGGSGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGF
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
173	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKPAPGGSGGTLQLDDEY
	RLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDY
	RPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLAN
	FRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGF
	CRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVC

	LKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDL
	TDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGR
	EIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFESPKKKAKVE
190	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAAAKEAA
	AKALEAEAAAKEAAAKEAAAKAGGLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIR
	PHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	GTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRD
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAAAKEAA
	GQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKG
	AKALEAEAAAKEAAAKEAAAKAGGLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIR
	VARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPA
	PHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP
	ALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEG
	KSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADG
	SEFEKRTADGSEFESPKKKAKVE
190	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAAAKEAA
	AKALEAEAAAKEAAAKEAAAKAGGLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIR
	PHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDP

	GTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRD
	GQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKG
	VARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPA
	ALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEG
	KSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADG
	SEFEKRTADGSEFESPKKKAKVE
191	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGGSGGSGGSG
	GLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRK
	PGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEAL
	HRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREF
	LGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVA
	SGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEET
	GVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGW
	LTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
192	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGPAPEAAAKGGGGGLDDEYR

	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFESPKKKAKVE
193	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKEAAAA KEAAAKEAAAKGGLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQ SPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQG FKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQI PAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRP VAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPV THDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATA HVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFESP KKKAKVE
195	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE

	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGSPAPGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
195	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGSPAPGGLDDEYR LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
196	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM

	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGSETPGTSESATPESGGL
	DDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPG
	TNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHR
	DLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLG
	KAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	WPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGV
	RKDLTDIPLTGEVLTWFTDGSSYVVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLT
	SAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
199	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGSGGEAAAKGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
199	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL

	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSGGGEAAAKGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKETVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
202	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSGGGSGSSGGSGGLD
	DEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGT
	NDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRD
	LANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGK
	AGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGW
	PVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVR
	KDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTS
	AGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
207	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA

	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGGSSPAPGGLDDEYRLY
	SPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPV
	QDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRI
	QHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRL
	FIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKA
	IAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
	PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIK
	NKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFESPKKKAKVE
208	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSGSSGSSSGGLDDEYRLY
	SPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPV
	QDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRI
	QHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRL
	FIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKA
	IAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
	PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIK
	NKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFESPKKKAKVE
212	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK

	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGSSGGLDDEYRLYSPL
	VKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDL
	REVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHP
	QVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIP
	GFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAA
	VAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLT
	GEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKE
	EILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
213	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSGSSGGLDDEYRLY
	SPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPV
	QDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRI
	QHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRL
	FIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKA
	IAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
	PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIK
	NKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
216	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA

	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGPAPGGSEAAAKGGLDDEYR KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGPAPGGSEAAAKGGLDDEYR LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
217	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGPAPGGSEAAAKGGLDDEYR LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
219	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA

	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGGGGLDDEYRLYS
	PLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQ
	DLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQ
	HPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLF
	IPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAI
	AAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIP
	LTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKN
	KEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
223	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSPAPGGLDDEYRLYSPL
	VKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDL
	REVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHP
	QVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIP
	GFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAA
	VAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLT
	GEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKE
	EILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
224	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ

	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKPAPGSSGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQGGLVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
225	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQGGLVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
228	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL

	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSGSGSSGGSGGLDDEY
	RLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDY
	RPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLAN
	FRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGF
	CRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVC
	LKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDL
	TDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGR
	EIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
228	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSGSGSSGGSGGLDDEY
	RLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDY
	RPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLAN
	FRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGF
	CRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVC
	LKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDL
	TDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGR
	EIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
229	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL

	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKEAAAKEAAA
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL
	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAK
232	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	PLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSP
	TIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTT
	AKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLS
	KKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCH
	QLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININIYTDSRYAFATAHVHGA
	IYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAK
	VE

235	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSGSSGGLDDEYRLYSPL VKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDL REVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHP QVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIP GFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAA VAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLT GEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKE EILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
239	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSEAAAKPAPGGLDDEYR LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT

	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
252	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGPAPGGSGGLDDEYRLYSPL
	VKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDL
	REVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHP
	QVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIP
	GFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAA
	VAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLT
	GEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKE
	EILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFEKRTADGSEFESPKKKAKVE
258	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSSEAAAKGGSGGLDDEYR
	LYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYR
	PVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANF
	RIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFC
	RLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCL

	KAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLT
	DIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGRE
	IKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFESPKKKAKVE
268	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGGSGSSPAPGGLDDEYRLY
	SPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPV
	QDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRI
	QHPQVTLLQYVDDLLLAGATKQDCLEGTKALLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRL
	FIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKA
	IAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDI
	PLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIK
	NKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLPAGKRTADGSEFESPKKKAKVE
278	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAKKGGGPAPGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA

	GFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
279	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKGGTL
	QLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRK
	PGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEAL
	HRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREF
	LGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVA
	SGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEET
	GVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGL
	LTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
280	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKEAAAKEAAA
	KEAAAKEAAAKGGTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILV
	PVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRL

	PQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTV VQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPW RRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETD EPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININYTDSRYAF ATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFE SPKKKAKVE
298	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKEAAAKEAAAKGGTL HRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREF LGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVA SGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEET GVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSININYTDSRYAFATAHVHGAIYKQRGW LTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
299	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKGGSGGGGGTLQLDD

	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
300	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAAAKEAA
	AKALEAEAAAKEAAAKEAAAKEAAAKAGGTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQE
	GIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEW
	RDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYS
	LRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDE
	RKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFA
	PPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRL
	AEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTA
	DGSEFEKRTADGSEFESPKKKAKVE
300	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE

	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAAAKEAAAKEAA
	AKALEAEAAAKEAAAKEAAAKAGGTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQE
	GIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEW
	RDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYS
	LRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDE
	RKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFA
	PPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRL
	AEGKSININYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTA
	DGSEFEKRTADGSEFESPKKKAKVE
302	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGGSPAPEAAAKGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
303	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL

	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSSGGSSGSETPGTSES
	ATPESSGGSSGGSSGGTLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQG
	ILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTW
	TRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARK
	KTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTL
	GPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPE
	ETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSR
	YAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGS
	EFESPKKKAKVE
304	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK
	ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNA
	KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
	NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
	PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKE
	VKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
	VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGEAAAKPAPGSSGGTLQLDD
	EYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTN
	DYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDL
	ANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTA
	GFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWP
	VCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRK
	DLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSA
	GREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLAGKRTADGSEFEKRTADGSEFESPKKKAKVE
305	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE
	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
	QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
	FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEEL
	LVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ
	SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA
	SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
	NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 9
CONTENANT LES PAGES 1 À 217

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 9
CONTAINING PAGES 1 TO 217

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHIER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

CLAIMS

1. A template RNA comprising, from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene, wherein the gRNA spacer has a sequence comprising the core nucleotides of a gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer, or wherein the gRNA spacer has a sequence of a spacer chosen from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A;

(ii) a gRNA scaffold that binds a gene modifying polypeptide (e.g., binds the Cas domain of the gene modifying polypeptide),

(iii) a heterologous object sequence comprising a mutation region to introduce a mutation into (e.g., to correct a mutation in) a second portion of the human PAH gene (wherein optionally the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, a mutation region, and a pre-edit homology region), and

(iv) a primer binding site (PBS) sequence comprising at least 5, 6, 7, or 8 bases with 100% identity to a third portion of the human PAH gene.

2. The template RNA of claim 1, wherein the heterologous object sequence comprises the core nucleotides of an RT template sequence from Table 3A, Table 3B, Table 3C, or Table 3D, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or wherein the heterologous object sequence comprises a sequence of an RT template sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

3. The template RNA of claim 1, wherein the heterologous object sequence comprises the core nucleotides of the RT template sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the gRNA spacer sequence, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or

wherein the heterologous object sequence comprises a sequence of an RT template sequence from Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

4. A template RNA comprising, from 5' to 3':

(i) a gRNA spacer that is complementary to a first portion of the human PAH gene,

(ii) a gRNA scaffold that binds a gene modifying polypeptide (e.g., binds the Cas domain of the gene modifying polypeptide),

(iii) a heterologous object sequence comprising a mutation region to introduce a mutation into (e.g., to correct a mutation in) a second portion of the human PAH gene, wherein the heterologous object sequence comprises the core nucleotides of an RT template sequence of Table 3A, Table 3B, Table 3C, or Table 3D, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the RT template sequence, or wherein the heterologous object sequence comprises an RT template sequence of Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A; and

(iv) a PBS sequence comprising at least 5, 6, 7, or 8 bases of 100% identity to a third portion of the human PAH gene.

5. The template RNA of claim 4, wherein the gRNA spacer comprises the core nucleotides of a gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer sequence, or wherein the gRNA spacer comprises a gRNA spacer sequence of Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

6. The template RNA of any one of claims 1-5, wherein the gRNA spacer comprises ACCTCAATCCTTTGGGTGTA (SEQ ID NO: 16355).

7. The template RNA of any one of claims 1-5, wherein the gRNA spacer comprises CCTCAATCCTTTGGGTGTAT (SEQ ID NO: 16332).

8. The template RNA of any one of claims 1-5, wherein the gRNA spacer comprises TGGGTCGTAGCGAACTGAGA (SEQ ID NO: 16102).

9. The template RNA of any one of claims 1-5, wherein the gRNA spacer comprises GGGTCGTAGCGAACTGAGAA (SEQ ID NO: 16084).

10. The template RNA of any one of claims 1-5, wherein the gRNA spacer comprises TAGCGAACTGAGAAGGGCCA (SEQ ID NO: 16011).

11. The template RNA of any one of claims 1-5, wherein the gRNA spacer comprises ACTTTGCTGCCACAATACCT (SEQ ID NO: 16032).

12. The template RNA of claim 5, wherein the heterologous object sequence comprises the core nucleotides of the gRNA spacer sequence of Table 1A, Table 1B, Table 1C, or Table 1D that corresponds to the RT template sequence, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the gRNA spacer sequence, or wherein the heterologous object sequence comprises the nucleotides of the gRNA spacer sequence of Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, or a sequence having 1, 2, or 3 substitutions thereto.

13. The template RNA according to any one of claims 1-12, wherein the PBS sequence has a sequence comprising the core nucleotides of the PBS sequence from the same row of Table 3A, Table 3B, Table 3C, or Table 3D as the RT template sequence, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence.

14. The template RNA according any one of claims 1-12, wherein the PBS sequence has a sequence comprising the core nucleotides of a PBS sequence of Table 3A, Table 3B, Table 3C, or Table 3D that corresponds to the RT template sequence, the gRNA spacer sequence, or both, and optionally comprises one or more consecutive nucleotides starting with the 5' end of the flanking nucleotides of the PBS sequence, or wherein the PBS sequence has a sequence

comprising the a PBS sequence of Tables 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A that corresponds to the RT template sequence, the gRNA spacer sequence, or both.

15. The template RNA according to any one of claims 1-14, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

16. The template RNA according to any one of claims 1-14, wherein the gRNA scaffold comprises a sequence of a gRNA scaffold of Table 12 that corresponds to the RT template sequence, the gRNA spacer sequence, or both, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

17. The template RNA according to any one of claims 1-16, wherein the mutation is a W408R, Q261R, Q243R, and/or IVS10-11A>G mutation (e.g., to correct a pathogenic R408W, R261Q, R243Q, and/or IVS10-11G>A mutation) of the PAH gene.

18. The template RNA of any one of claims 1-17, wherein the pre-edit sequence comprises between about 1 nucleotide to about 35 nucleotides (e.g., comprises about 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, or 30-35 nucleotides) in length.

19. The template RNA of any one of claims 1-18, wherein the mutation region comprises a single nucleotide.

20. The template RNA of any one of claims 1-18, wherein the mutation region is at least two nucleotides in length.

21. The template RNA of any one of claims 1-18 or 20, wherein the mutation region is up to 32 (e.g., up to 5, 10, 15, 20, 25, 30, or 32) nucleotides in length and comprises one, two, or three sequence differences relative to a second portion of the human PAH gene.

22. The template RNA of any one of claims 1-18, 20, or 21, wherein the mutation region comprises two sequences differences relative to a second portion of the human PAH gene.

23. The template RNA of any one of claims 1-18 or 20-22, wherein the mutation region comprises a first region (e.g., a first nucleotide) designed to correct a pathogenic mutation in the PAH gene and a second region (e.g., a second nucleotide) designed to inactivate a PAM sequence (e.g., a "PAM-kill" mutation as described herein).

24. The template RNA of any one of claims 1-23, wherein the mutation region comprises less than 80%, 70%, 60%, 50%, 40%, or 30% identity to the corresponding portion of the human PAH gene.

25. The template RNA of any one of claims 1-23, wherein the template RNA comprises one or more silent mutations (e.g., silent substitutions), e.g., as exemplified in Tables 7A-7C, 8A-8D, E6, or E6A.

26. The template RNA of any of the preceding claims, wherein the mutation region comprises a first region designed to correct a pathogenic mutation in the PAH gene and a second region designed to introduce a silent substitution.

27. The template RNA of any one of claims 1-26, which comprises one or more chemically modified nucleotides.

28. A gene modifying system comprising:
a template RNA of any of claims 1-27, and
a gene modifying polypeptide, or a nucleic acid (e.g., RNA) encoding the gene modifying polypeptide.

29. The gene modifying system of claim 28, wherein the gene modifying polypeptide comprises:

a reverse transcriptase (RT) domain (e.g., an RT domain from a retrovirus, or a polypeptide domain having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acids sequence identity thereto); and

a Cas domain that binds to the target DNA molecule and is heterologous to the RT domain (e.g., a Cas9 domain); and

optionally, a linker disposed between the RT domain and the Cas domain.

30. The gene modifying system of claim 29, wherein the RT domain comprises:

(a) an RT domain of Table 6; or

(b) an RT domain from a murine leukemia virus (MMLV), a porcine endogenous retrovirus (PERV); Avian reticuloendotheliosis virus (AVIRE), a feline leukemia virus (FLV), simian foamy virus (SFV) (e.g., SFV3L), bovine leukemia virus (BLV), Mason-Pfizer monkey virus (MPMV), human foamy virus (HFV), or bovine foamy/syncytial virus (BFV/BSV).

31. The gene modifying system of claim 29 or 30, wherein the Cas domain comprises a Cas domain of Table 7 or Table 8.

32. The gene modifying system of claim 29 or 30, wherein the Cas domain:

(a) is a Cas9 domain;

(b) is a SpCas9 domain, a BlatCas9 domain, a Nme2Cas9 domain, a PnpCas9 domain, a SauCas9 domain, a SauCas9-KKH domain, a SauriCas9 domain, a SauriCas9-KKH domain, a ScaCas9-Sc++ domain, a SpyCas9 domain, a SpyCas9-NG domain, a SpyCas9-SpRY domain, or a St1Cas9 domain; and/or

(c) is a Cas9 domain comprising an N670A mutation, an N611A mutation, an N605A mutation, an N580A mutation, an N588A mutation, an N872A mutation, an N863 mutation, an N622A mutation, or an H840A mutation.

33. The gene modifying system of claim 32, wherein the Cas9 domain binds a PAM sequence listed in Table 7 or Table 12.

34. The gene modifying system of claim 33, wherein a second portion of the human PAH gene overlaps with a PAM recognized by the Cas domain, e.g., wherein the second portion of the human PAH gene is within the PAM or wherein the PAM is within the second portion of the human PAH gene).

35. The gene modifying system of any one of claims 28-34, wherein the gRNA spacer is a gRNA spacer according to Table 1A, Table 1B, Table 1C, or Table 1D, and the Cas domain comprises a Cas domain listed in the same row of Table 1A, Table 1B, Table 1C, or Table 1D, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

36. The gene modifying system of any one of claims 28-34, wherein the template RNA comprises a sequence of a template RNA sequence of Table 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

37. The gene modifying system of any one of claims 28-36 wherein:

(a) the template RNA comprises a sequence of a template RNA sequence of Table 3A-3D, 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A;

(b) the Cas domain comprises a Cas domain of Table 7 or Table 8;

(c) the linker comprises a linker sequence of Table 10 (e.g., of any of SEQ ID NOs: 5217, 5106, 5190, and 5218); and

(d) the gene modifying polypeptide comprises one or two NLS sequences from Table 11 (e.g., of any of SEQ ID NOs: 5245, 5290, 5323, 5330, 5349, 5350, 5351, and 4001).

38. The gene modifying system of any of claims 28-37, which produces a first nick in a first strand of the human PAH gene.

39. The gene modifying system of claim 38, which further comprises a second strand-targeting gRNA that directs a second nick to the second strand of the human PAH gene.

40. The gene modifying system of claim 39, wherein the second strand-targeting gRNA comprises:

(i) a sequence comprising the core nucleotides of a left gRNA spacer sequence or a right gRNA spacer sequence from Table 2A, Table 2B, Table 2C, or Table 2D, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the left gRNA spacer sequence or right gRNA spacer sequence; or

(ii) a second -strand-targeting gRNA comprising a spacer sequence of Table 6A, or a spacer sequence having 1, 2, or 3 substitutions thereto.

41. The gene modifying system of claim 39, wherein the second strand-targeting gRNA comprises a sequence comprising the core nucleotides of a left gRNA spacer sequence or a right gRNA spacer sequence from Table 2A, Table 2B, Table 2C, or Table 2D that corresponds to the gRNA spacer sequence of (i), and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the left gRNA spacer sequence or right gRNA spacer sequence.

42. The gene modifying system of claim 39, wherein the second strand-targeting gRNA comprises a sequence comprising the core nucleotides of a second nick gRNA sequence from Table 4A, Table 4B, Table 4C, or Table 4D, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the second nick gRNA sequence; or

(ii) a second -strand-targeting gRNA comprising a spacer sequence from Table 6A or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

43. The gene modifying system of claim 39, wherein the second strand-targeting gRNA comprises a sequence comprising the core nucleotides of the second nick gRNA sequence from Table 4A, Table 4B, Table 4C, or Table 4D that corresponds to the gRNA spacer sequence of (i), or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity

thereto, and optionally comprises one or more consecutive nucleotides starting with the 3' end of the flanking nucleotides of the second nick gRNA sequence.

44. The gene modifying system of any one of claims 39-43, wherein the second strand-targeting gRNA has a "PAM-in orientation" with the template RNA of the gene modifying system, e.g., as exemplified in Tables 2A-2D, 4A-4D, or 6A.

45. The gene modifying system of any one of claims 39-44, the second strand-targeting gRNA targets a sequence overlapping the target mutation of the template RNA.

46. The gene modifying system of claim 45, wherein second strand-targeting gRNA comprises:

(i) a sequence (e.g., a spacer sequence) complementary to the PAH mutation;

(ii) a sequence (e.g., a spacer sequence) complementary to the wild-type sequence at the target locus;

(iii) a sequence (e.g., a spacer sequence) complementary to a SNP proximal to the target locus, e.g., a SNP contained in the genomic DNA of a subject (e.g., a patient);

(iv) a sequence (e.g., spacer sequence) complementary to or comprising one or more silent substitutions proximal to the target locus.

47. The template RNA or gene modifying system of any one of the preceding claims, wherein the gRNA spacer comprises about 1, 2, 3, or more flanking nucleotides of the gRNA spacer.

48. The template RNA or gene modifying system of any one of the preceding claims, wherein the heterologous object sequence comprises about 2, 3, 4, 5, 10, 20, 30, 40, or more flanking nucleotides of the RT template sequence.

49. The template RNA or gene modifying system of any one of the preceding claims, wherein the heterologous object sequence comprises between about 8-30, 9-25, 10-20, 11-16, or 12-15 (e.g., about 11-16) nucleotides.

50. The template RNA or gene modifying system of any one of the preceding claims, wherein the mutation region comprises 1, 2, or 3 nucleotide positions of sequence difference relative to the corresponding portion of the human PAH gene.

51. The template RNA or gene modifying system of any one of the preceding claims wherein the mutation region comprises at least 2 nucleotide positions of sequence difference relative to the corresponding portion of the human PAH gene.

52. The template RNA or gene modifying system, of any one of the preceding claims, wherein the post-edit homology region and/or pre-edit homology region comprises 100% identity to the PAH gene.

53. The template RNA or gene modifying system of any one of the preceding claims, wherein the PBS sequence additionally comprises about 1, 2, 3, 4, 5, 6, 7, or more flanking nucleotides.

54. The template RNA or gene modifying system of any one of the preceding claims, wherein the PBS sequence comprises about 5-20, 8-16, 8-14, 8-13, 9-13, 9-12, or 10-12 (e.g., about 9-12) nucleotides.

55. The template RNA or gene modifying system of any one of the preceding claims, wherein the PBS sequence binds within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nick site in the PAH gene.

56. The gene modifying system of any one of the preceding claims, wherein the domains of the gene modifying polypeptide are joined by a peptide linker.

57. The gene modifying system of claim 56, wherein the linker comprises a sequence of a linker of Table 10 (e.g., of any of SEQ ID NOs: 5217, 5106, 5190, and 5218).

58. The gene modifying system of any one of the preceding claims, wherein the gene modifying polypeptide further comprise one or more nuclear localization sequences (NLS).

59. The gene modifying system of claim 58, wherein the gene modifying polypeptide comprises a first NLS and a second NLS.

60. The gene modifying system of claim 58 or 59, wherein the NLS comprises a sequence of a NLS of Table 11 (e.g., of any of SEQ ID NOs: 5245, 5290, 5323, 5330, 5349, 5350, 5351, and 4001).

61. A template RNA comprising a sequence of a template RNA of Table 4A-4D 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

62. A template RNA comprising a sequence of a template RNA of Table 4A-4D, 5A-5F, 8A-8D, E3, E3A, BB, E5, E5A, E6, or E6A.

63. A gene modifying system comprising:

(iii) a template RNA comprising a sequence of a template RNA of Table 4A, Table 4B, Table 4C, or Table 4D, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(iv) a second-nick gRNA sequence from the same row of Table 4A, Table 4B, Table 4C, or Table 4D as (i), a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

64. A gene modifying system comprising:

(iii) a template RNA comprising a sequence of a template RNA of Table 4A, Table 4B, Table 4C, or Table 4D; and

(iv) a second-nick gRNA sequence from the same row of Table 4A, Table 4B, Table 4C, or Table 4D as (i).

65. A DNA encoding the template RNA of any one of claims 1-27, 48-55, 61, or 62, or the gene modifying system of any one of claims 28-60, 63, or 64.

66. A pharmaceutical composition, comprising the system of any one of claims 28-60, 63, or 64, or one or more nucleic acids encoding the same, and a pharmaceutically acceptable excipient or carrier.

67. The pharmaceutical composition of claim 66, wherein the pharmaceutically acceptable excipient or carrier is selected from the group consisting of a plasmid vector, a viral vector, a vesicle, and a lipid nanoparticle.

68. The pharmaceutical composition of claim 67, wherein the viral vector is an adeno-associated virus.

69. A host cell (e.g., a mammalian cell, e.g., a human cell) comprising the template RNA or gene modifying system of any one of the preceding claims.

70. A method of making the template RNA of any one of claims 1-27, 48-55, 61, or 62, the method comprising synthesizing the template RNA by *in vitro* transcription (e.g., solid state synthesis) or by introducing a DNA encoding the template RNA into a host cell under conditions that allow for production of the template RNA.

71. A method for modifying a target site in the human PAH gene in a cell, the method comprising contacting the cell with the gene modifying system of any one of claims 28-60, 63, or 64, or DNA encoding the same, thereby modifying the target site in the human PAH gene in a cell.

72. A method for treating a subject having a disease or condition associated with a mutation in the human PAH gene, the method comprising administering to the subject the gene modifying system of any one of claims 28-60, 63, or 64, or DNA encoding the same, thereby treating the subject having a disease or condition associated with a mutation in the human PAH gene.

73. The method of claim 71 or 72, wherein the disease or condition is phenylketonuria (PKU) or hyperphenylalaninemia (e.g., mild or severe hyperphenylalaninemia).

74. The method of any one of claims 72-73, wherein the subject has a R408W, R261Q, R243Q, and/or IVS10-11G>A mutation.

75. A method for treating a subject having PKU the method comprising administering to the subject the gene modifying system of any one of claims 28-60, 63, or 64, or DNA encoding the same, thereby treating the subject having PKU.

76. The gene modifying system or method of any one of the preceding claims, wherein introduction of the system into a target cell results in a correction of a pathogenic mutation in the PAH gene.

77. The gene modifying system or method of any one of the preceding claims, wherein the pathogenic mutation is a R408W, R261Q, R243Q, and/or IVS10-11G>A mutation, and wherein the correction comprises an amino acid substitution of W408R, Q261R, and/or Q243R, or a nucleotide substitution of IVS10-11A>G.

78. The gene modifying system or method of any one of the preceding claims, wherein introduction of the system into a target cell results in a mutation that causes the restoration of the function of the PAH gene.

79. The gene modifying system or method of any of the preceding claims, wherein correction of the mutation occurs in at least 30% (e.g., 30%, 40%, 50%, 60%, 70%, or more) of target nucleic acids.

80. The gene modifying system or method of any of the preceding claims, wherein correction of the mutation occurs in at least 30% (e.g., 30%, 40%, 50%, 60%, 70%, or more) of target cells.

81. The gene modifying system or method of any of the preceding claims, wherein the gene modifying system comprises a second strand-targeting gRNA, and wherein correction of the mutation in a population of target cells is increased relative to a population of target cells treated with a gene modifying system comprising a template RNA without a second strand-targeting gRNA.

82. The gene modifying system or method of any of the preceding claims, wherein the template RNA comprises one or more silent substitutions (e.g., as exemplified in Tables 7A, X4, and X4A), and wherein correction of the mutation in a population of target cells is increased relative to a population of target cells treated with a gene modifying system comprising a template RNA that does not comprise one or more silent substitutions.

83. The method of any of the preceding claims, wherein the cell is a mammalian cell, such as a human cell.

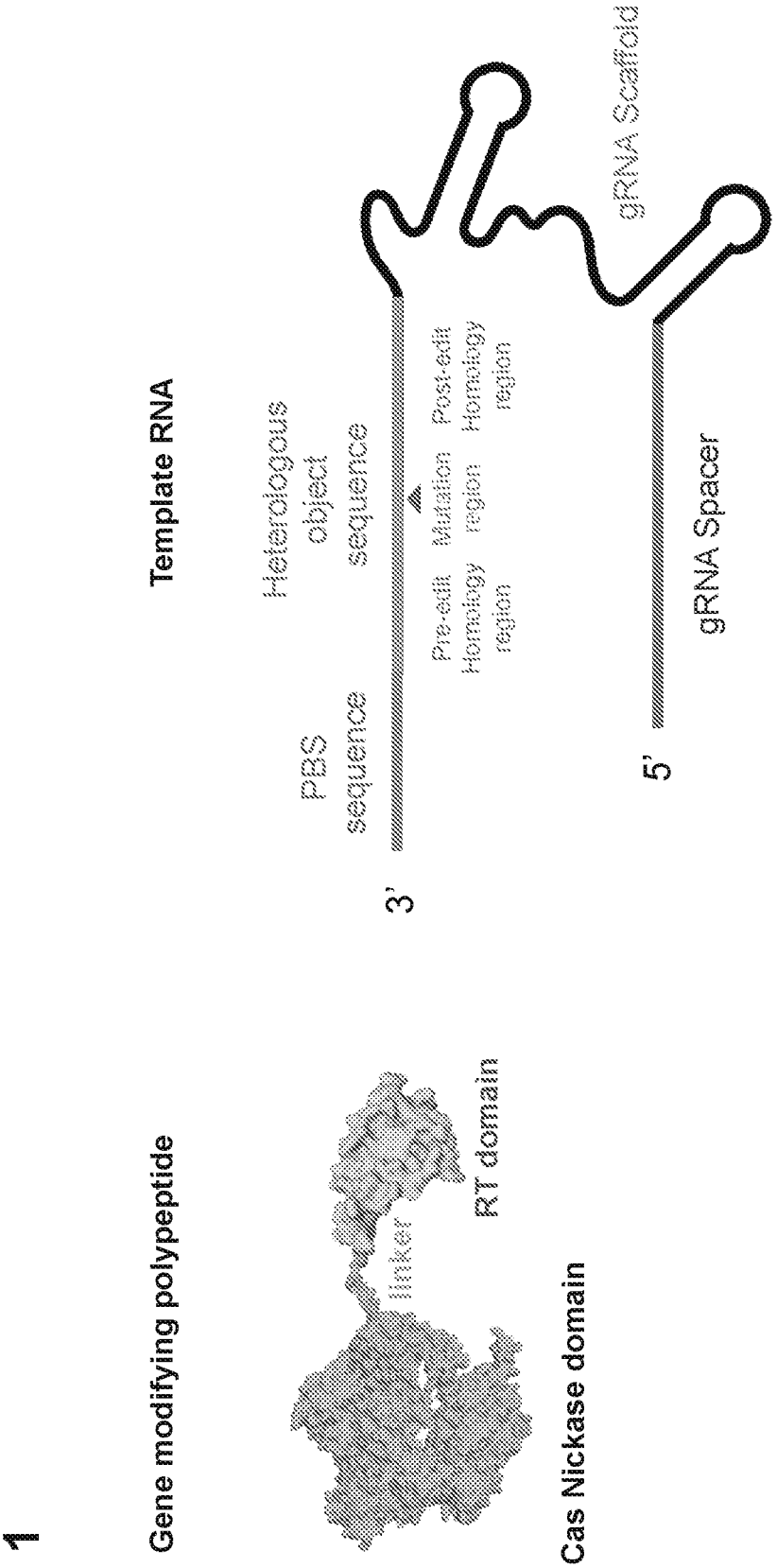
84. The method of any one of the preceding claims, wherein the subject is a human.

85. The method of any of the preceding claims, wherein the contacting occurs *ex vivo*, e.g., wherein the cell's or subject's DNA is modified *ex vivo*.

86. The method of any of the preceding claims, wherein the contacting occurs *in vivo*, e.g., wherein the cell's or subject's DNA is modified *in vivo*.

87. The method of any of the preceding claims, wherein contacting the cell or the subject with the system comprises contacting the cell or a cell within the subject with a nucleic acid (e.g., DNA or RNA) encoding the gene modifying polypeptide under conditions that allow for production of the gene modifying polypeptide.

FIG. 1


Gene modifying polypeptide
Cas Nickase domain
Linker
RT domain
Template RNA
3'
PBS sequence
Heterologous object sequence
Pre-edit Homology region
Mutation region
Post-edit Homology region
5'
gRNA Spacer
gRNA Scaffold

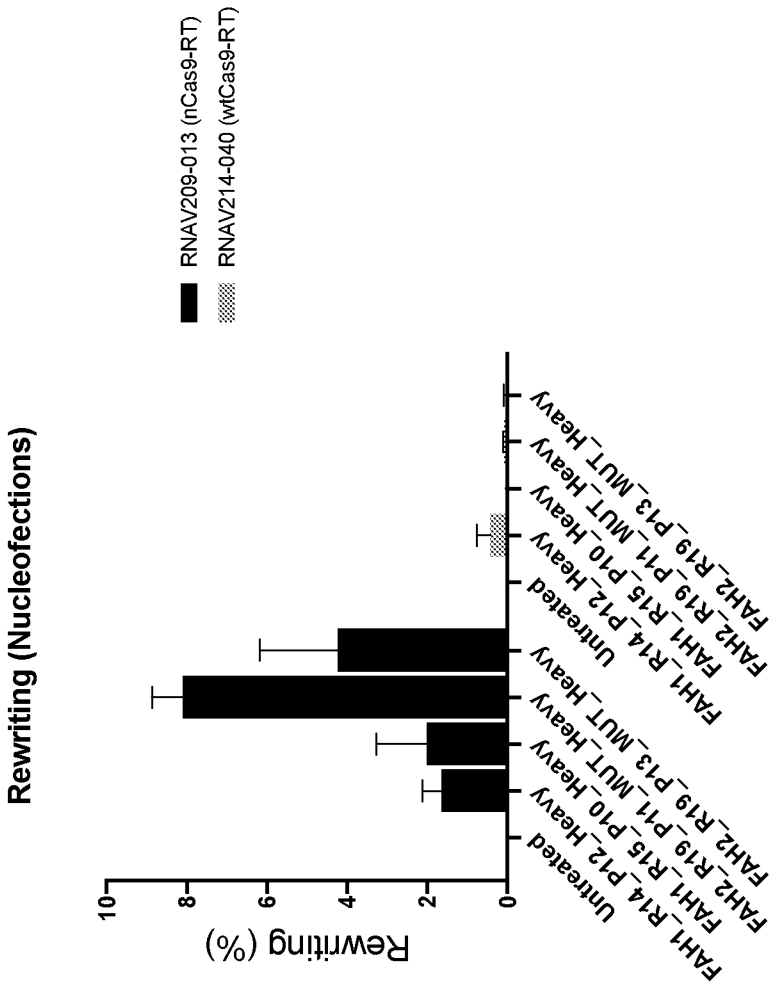

Rewriting (Nucleofections)
RNAV209-013 (nCas9-RT)
RNAV214-040 (wtCas9-RT)
Rewriting (%)
0
2
4
6
8
10
Untreated
FAH1_R14_P12_Heavy
FAH1_R15_P10_Heavy
FAH2_R19_P11_MUT_Heavy
FAH2_R19_P13_MUT_Heavy
Untreated
FAH1_R14_P12_Heavy
FAH1_R15_P10_Heavy
FAH2_R19_P11_MUT_Heavy
FAH2_R19_P13_MUT_Heavy

FIG. 2

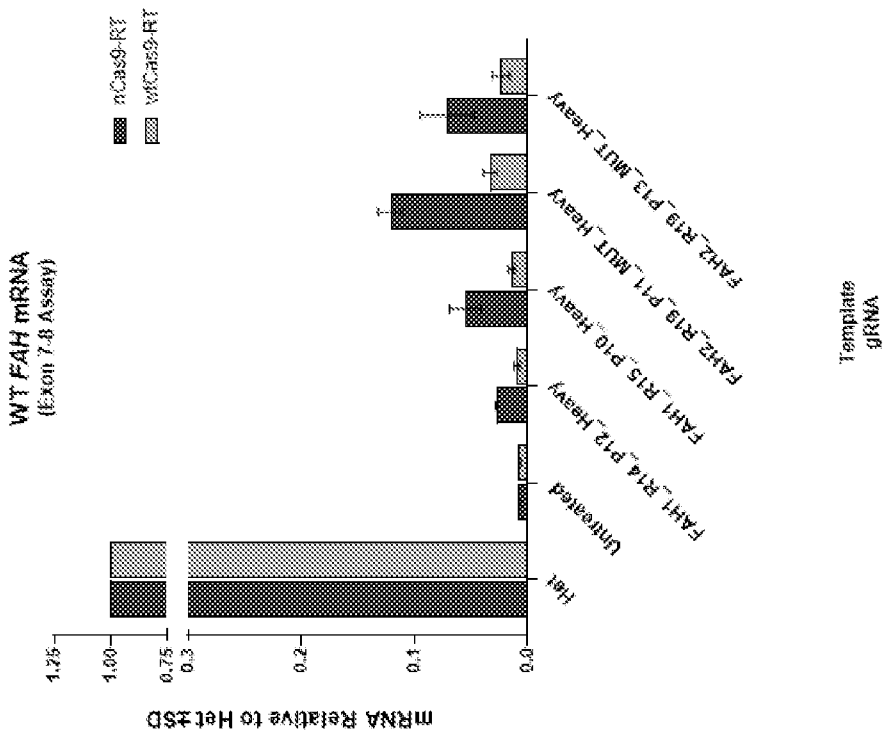
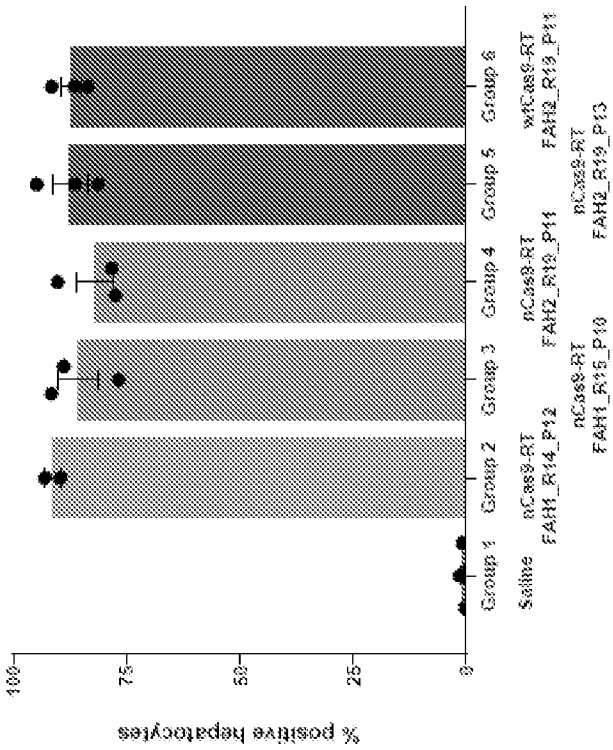

WT FAH mRNA
(Exon 7-8 Assay)
nCas9-RT
vitCas9-RT
mRNA Relative to Het±SD
1.25
1.00
0.75
0.3
0.2
0.1
0.0
Het
Untreated
FAH1_R14_P12_Heavy
FAH1_R15_P10_Heavy
FAH2_R19_P11_MULT_Heavy
FAH2_R19_P13_MULT_Heavy
Template
gRNA

FIG. 3

FIG. 4


Cas9 IHC
(6h post dosing)
% positive hepatocytes
100
75
50
25
0
Group 1
Saline
Group 2
nCas9-RT
FAH_R14_P12
Group 3
nCas9-RT
FAH1_R15_P10
Group 4
nCas9-RT
FAH2_R19_P11
Group 5
nCas9-RT
FAH1_R15_P13
Group 6
wtCas9-RT
FAH2_R19_P11
nCas9-RT
FAH2_R19_P13

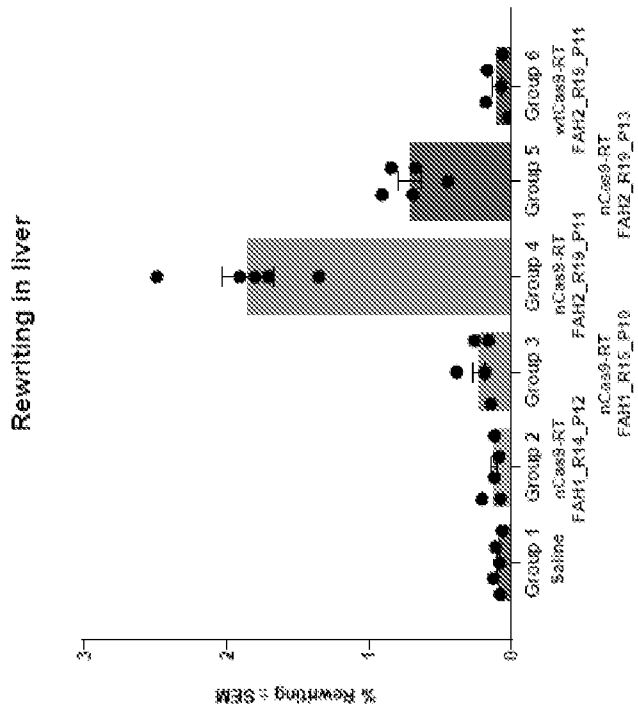

Rewriting in liver
% Rewriting ± SEM
0
1
2
3
Group 1
Saline
Group 2
nCas9-RT
FAH1_R14_P12
Group 3
nCas9-RT
FAH1_R19_P19
Group 4
nCas9-RT
FAH2_R19_P11
Group 5
nCas9-RT
FAH2_R19_P13
Group 6
wtCas9-RT
FAH2_R19_P11

FIG. 5

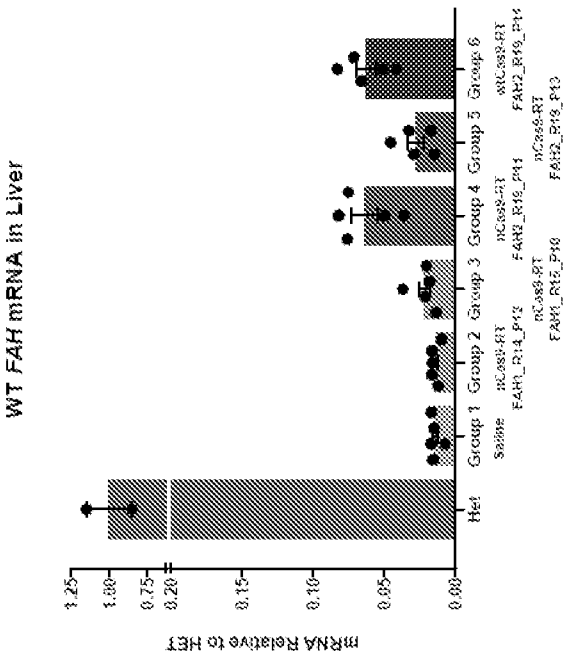

WT FAH mRNA in Liver
mRNA Relative to HET
3.25
1.50
0.75
0.20
0.15
0.10
0.05
0.00
Het
Group 1
Saline
Group 2
mCas9-RT
FAH_R14_P43
Group 3
mCas9-RT
FAH_R14_P43
Group 4
mCas9-RT
FAH_R15_P43
Group 5
mCas9-RT
FAH_R15_P43
Group 6
mCas9-RT
FAH_R15_P43

FIG. 6

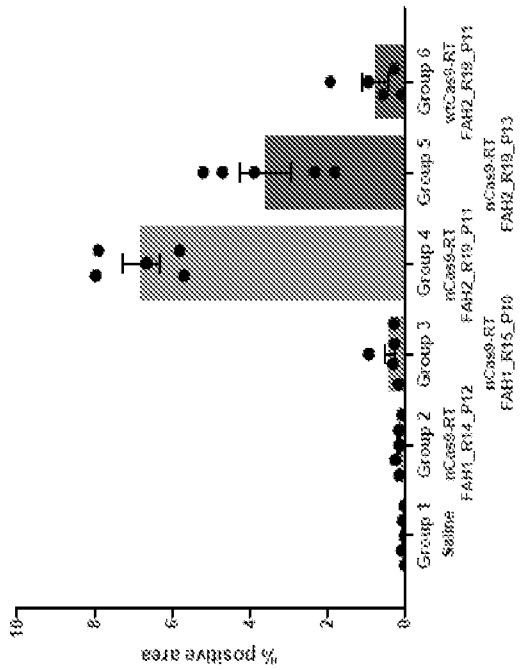
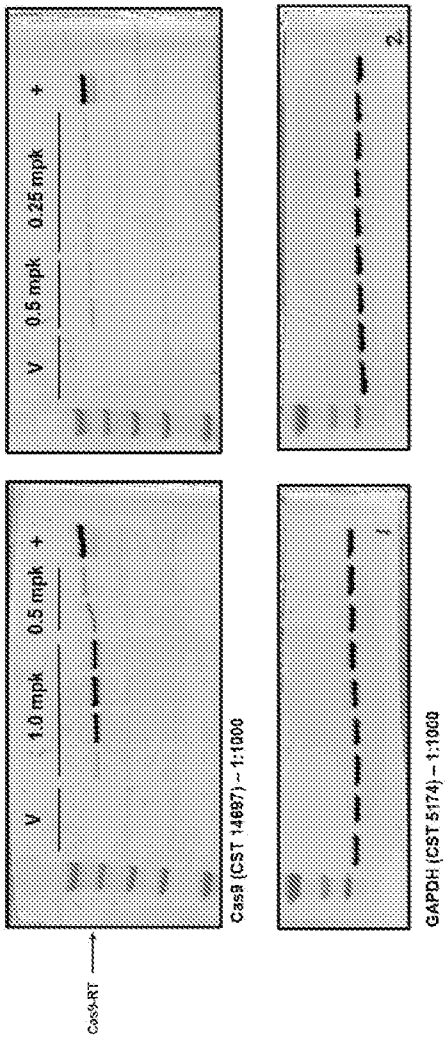

FAH IHC
(6 days post dosing)
% positive area
10
8
6
4
2
0
Group 1
Saline
Group 2
wtCas9-RT
FAH1_R14_P12
Group 3
wtCas9-RT
FAH1_R15_P16
Group 4
wtCas9-RT
FAH2_R18_P11
Group 5
wtCas9-RT
FAH2_R18_P11
Group 6
wtCas9-RT
FAH2_R18_P13

FIG. 7

FIG. 8


V
1.0 mpk
0.5 mpk
+
V
0.5 mpk
0.25 mpk
+
Cas9 (CST 14697) – 1:1000
GAPDH (CST 5174) – 1:1000
Cas9-RT

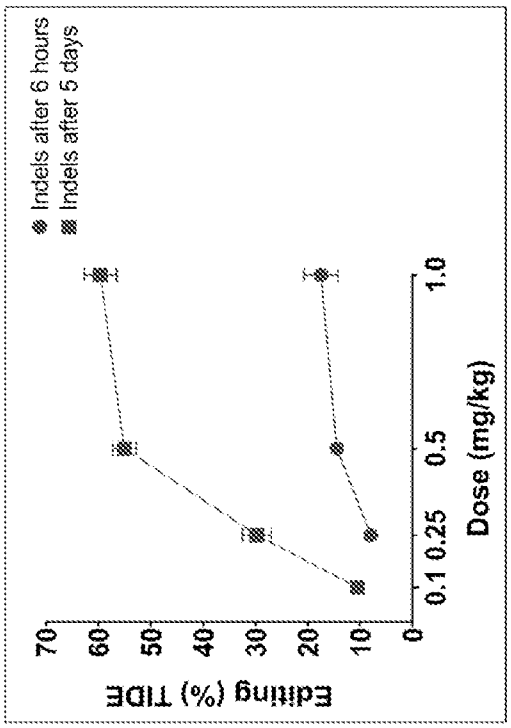

Indels after 6 hours
Indels after 5 days
Editing (%) TIDE
70
60
50
40
30
20
10
0
0.1 0.25 0.5 1.0
Dose (mg/kg)

FIG. 9

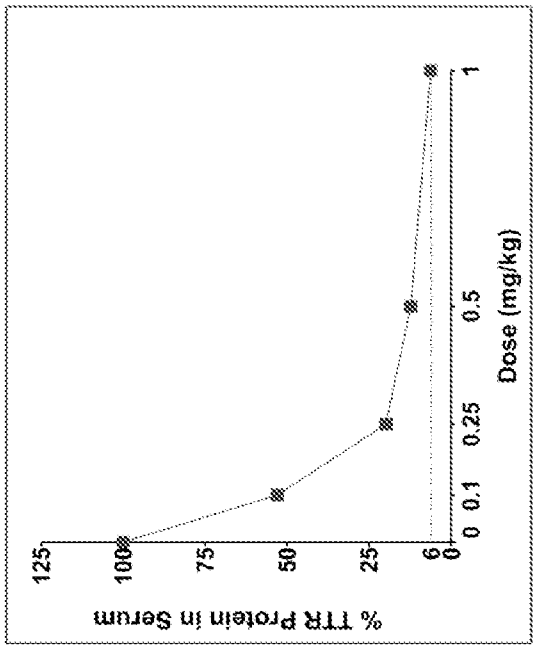


FIG. 10

FIG. 11

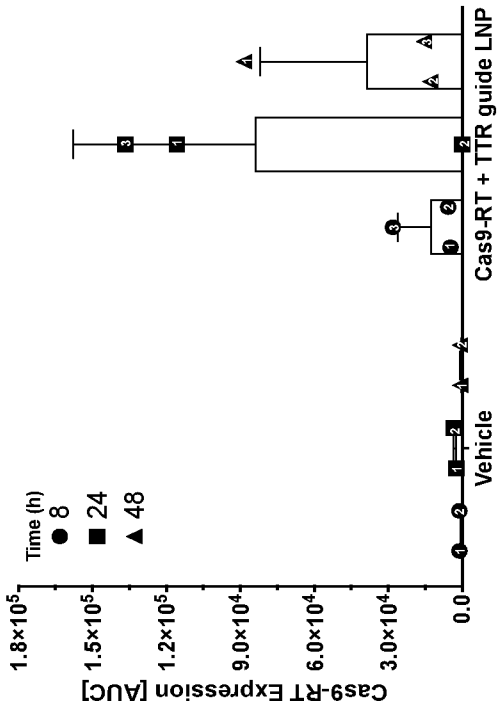
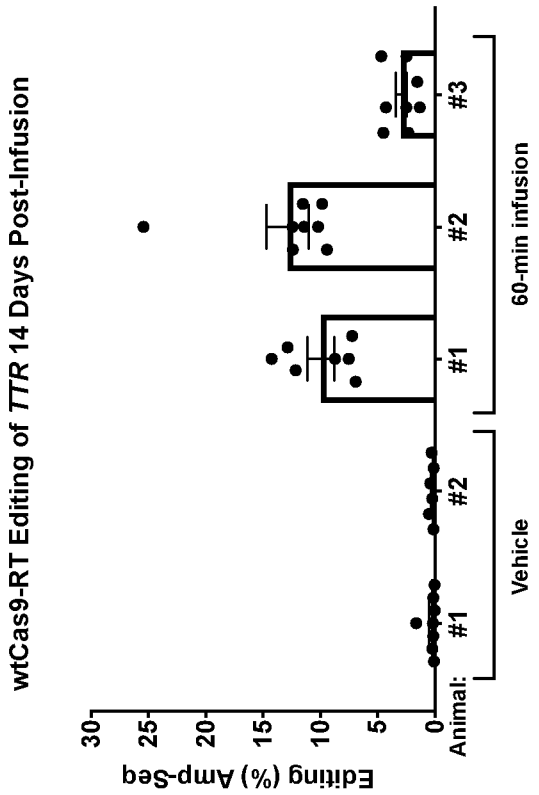

Cas9-RT Expression [AUC]
1.8×10^5
1.5×10^5
1.2×10^5
9.0×10^4
6.0×10^4
3.0×10^4
0.0
Time (h)
8
24
48
Vehicle
Cas9-RT + TTR guide LNP

FIG. 12


wtCas9-RT Editing of TTR 14 Days Post-Infusion
Editing (%) Amp-Seq
30
25
20
15
10
5
0
Animal: #1 #2 #1 #2 #3
Vehicle
60-min infusion

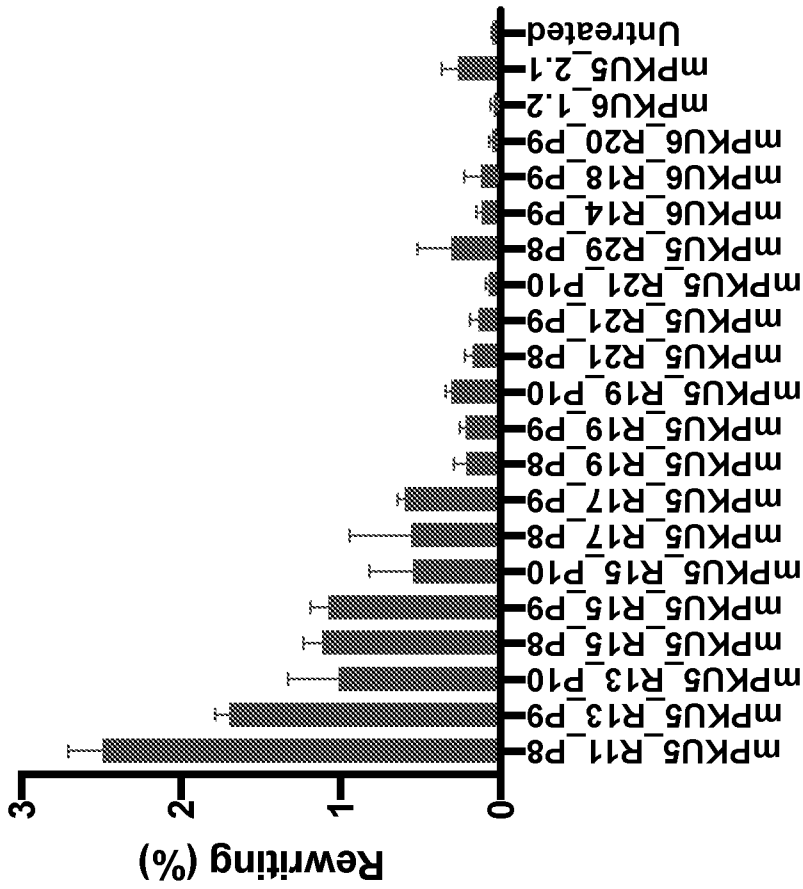
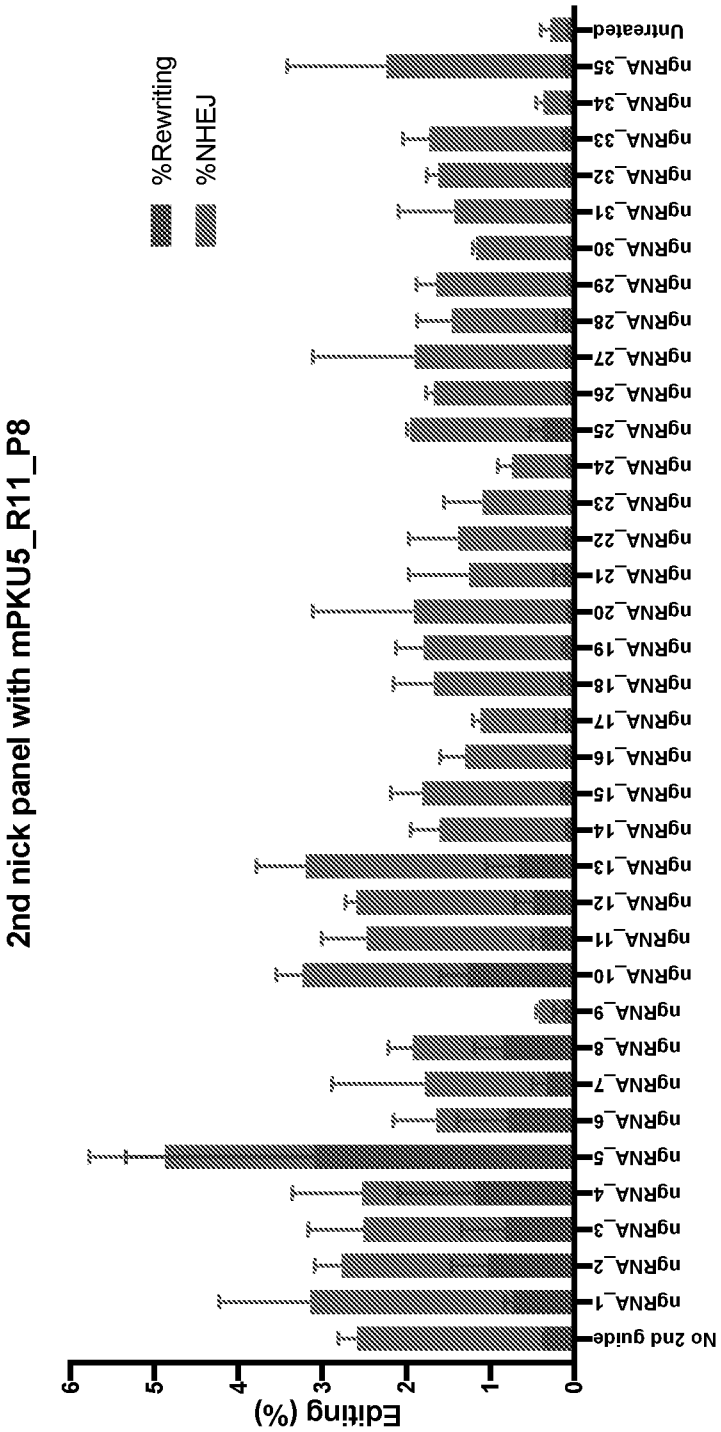

Rewriting (%)
0
1
2
3
mPKU5_R11_P8
mPKU5_R13_P9
mPKU5_R13_P10
mPKU5_R15_P8
mPKU5_R15_P9
mPKU5_R15_P10
mPKU5_R17_P8
mPKU5_R17_P9
mPKU5_R19_P8
mPKU5_R19_P9
mPKU5_R19_P10
mPKU5_R21_P8
mPKU5_R21_P9
mPKU5_R21_P10
mPKU5_R29_P8
mPKU6_R14_P9
mPKU6_R18_P9
mPKU6_R20_P9
mPKU6_1.2
mPKU5_2.1
Untreated

FIG. 13

FIG. 14


2nd nick panel with mPKU5_R11_P8
%Rewriting
%NHEJ
Editing (%)
0
1
2
3
4
5
6
No 2nd guide
ngRNA_1
ngRNA_2
ngRNA_3
ngRNA_4
ngRNA_5
ngRNA_6
ngRNA_7
ngRNA_8
ngRNA_9
ngRNA_10
ngRNA_11
ngRNA_12
ngRNA_13
ngRNA_14
ngRNA_15
ngRNA_16
ngRNA_17
ngRNA_18
ngRNA_19
ngRNA_20
ngRNA_21
ngRNA_22
ngRNA_23
ngRNA_24
ngRNA_25
ngRNA_26
ngRNA_27
ngRNA_28
ngRNA_29
ngRNA_30
ngRNA_31
ngRNA_32
ngRNA_33
ngRNA_34
ngRNA_35
Untreated

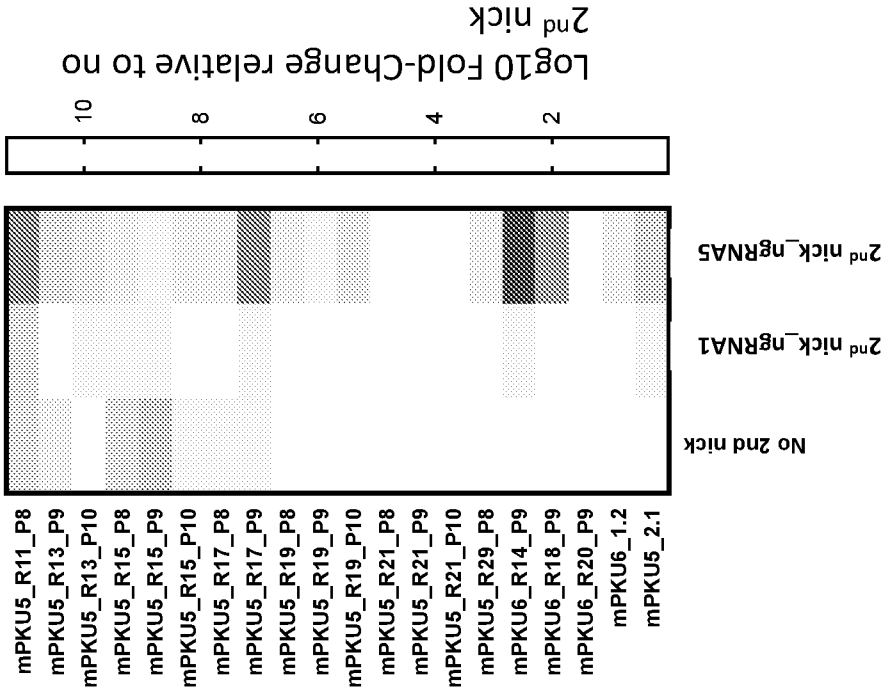

Log10 Fold-Change relative to no 2nd nick
10
8
6
4
2
No 2nd nick
2nd nick_ngRNA1
2nd nick_ngRNA5
mPKU5_R11_P8
mPKU5_R13_P9
mPKU5_R13_P10
mPKU5_R15_P8
mPKU5_R15_P9
mPKU5_R15_P10
mPKU5_R17_P8
mPKU5_R17_P9
mPKU5_R19_P8
mPKU5_R19_P9
mPKU5_R19_P10
mPKU5_R21_P8
mPKU5_R21_P9
mPKU5_R21_P10
mPKU5_R29_P8
mPKU6_R14_P9
mPKU6_R18_P9
mPKU6_R20_P9
mPKU6_1.2
mPKU5_2.1

FIG. 15

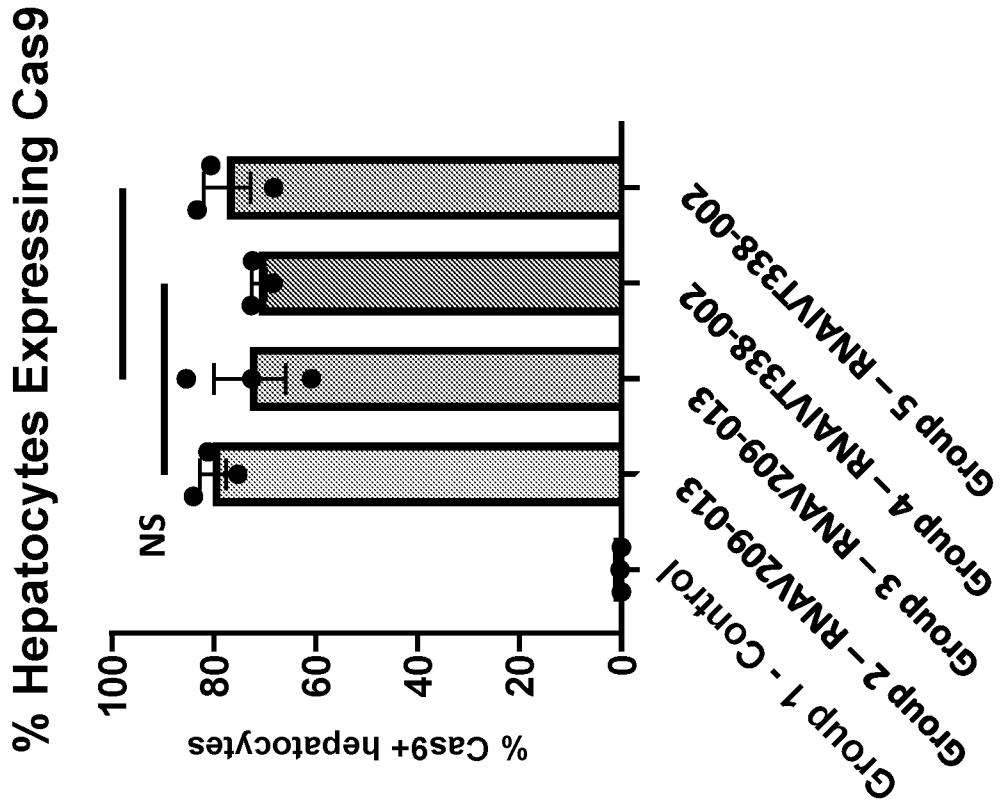
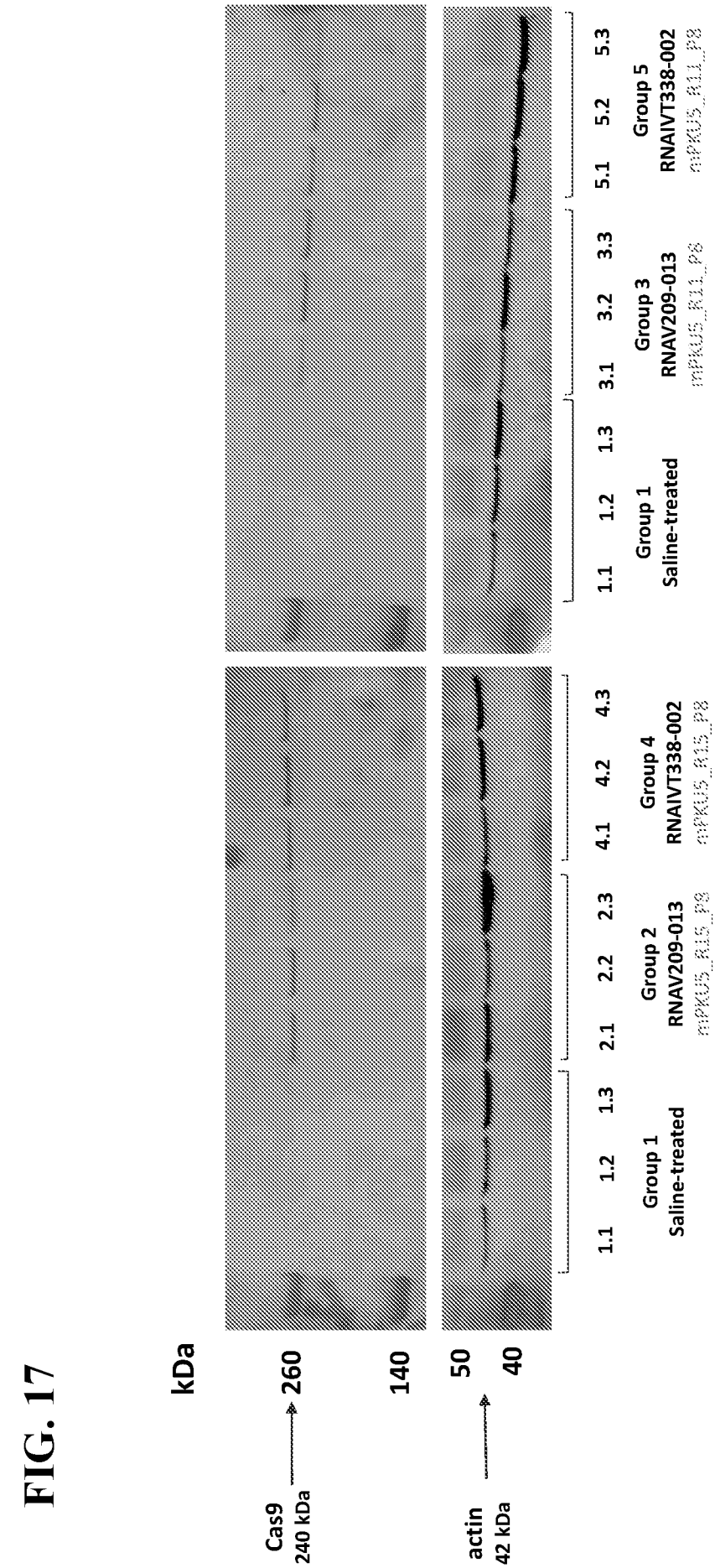

% Hepatocytes Expressing Cas9
% Cas9+ hepatocytes
100
80
60
40
20
0
NS
Group 1 - Control
Group 2 – RNAV209-013
Group 3 – RNAV209-013
Group 4 – RNAIVT338-002
Group 5 – RNAIVT338-002

FIG. 16

FIG. 17


kDa
260
140
50
40
Cas9
240 kDa
actin
42 kDa
1.1 1.2 1.3 2.1 2.2 2.3 4.1 4.2 4.3
Group 1
Saline-treated
Group 2
RNAV209-013
mPKUS_R13_P8
Group 4
RNAIVT338-002
mPKUS_R13_P8
1.1 1.2 1.3 3.1 3.2 3.3 5.1 5.2 5.3
Group 1
Saline-treated
Group 3
RNAV209-013
mPKUS_R11_P8
Group 5
RNAIVT338-002
mPKUS_R11_P8

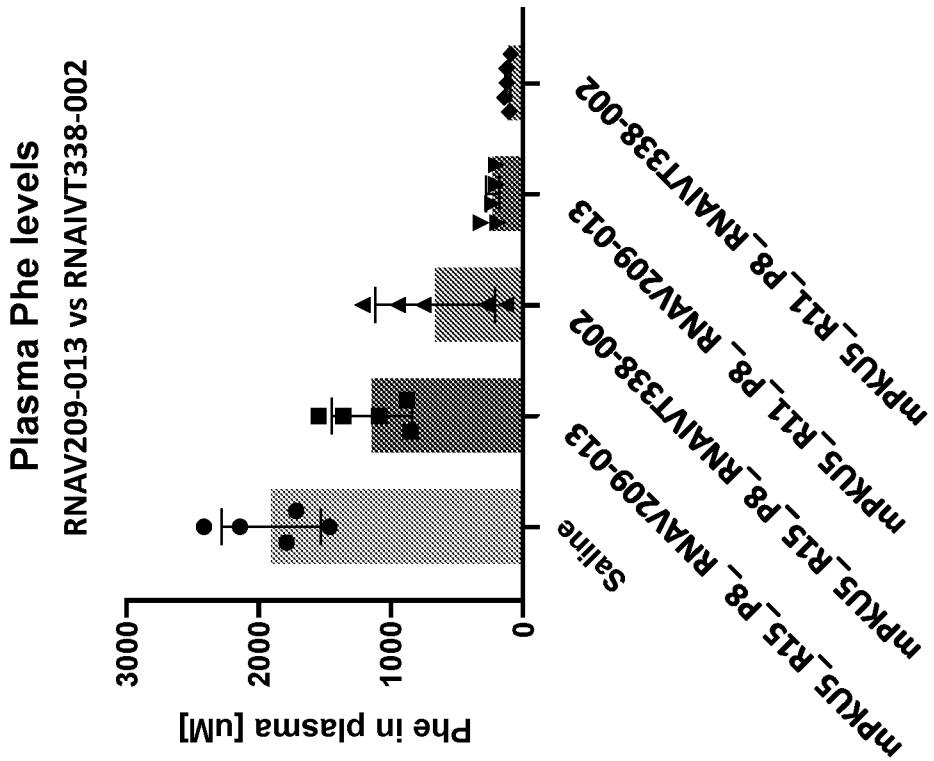

Plasma Phe levels
RNAV209-013 vs RNAIVT338-002
Phe in plasma [uM]
3000
2000
1000
0
Saline
mPKU5_R15_P8_RNAV209-013
mPKU5_R15_P8_RNAIVT338-002
mPKU5_R11_P8_RNAV209-013
mPKU5_R11_P8_RNAIVT338-002

FIG. 18

FIG. 19A

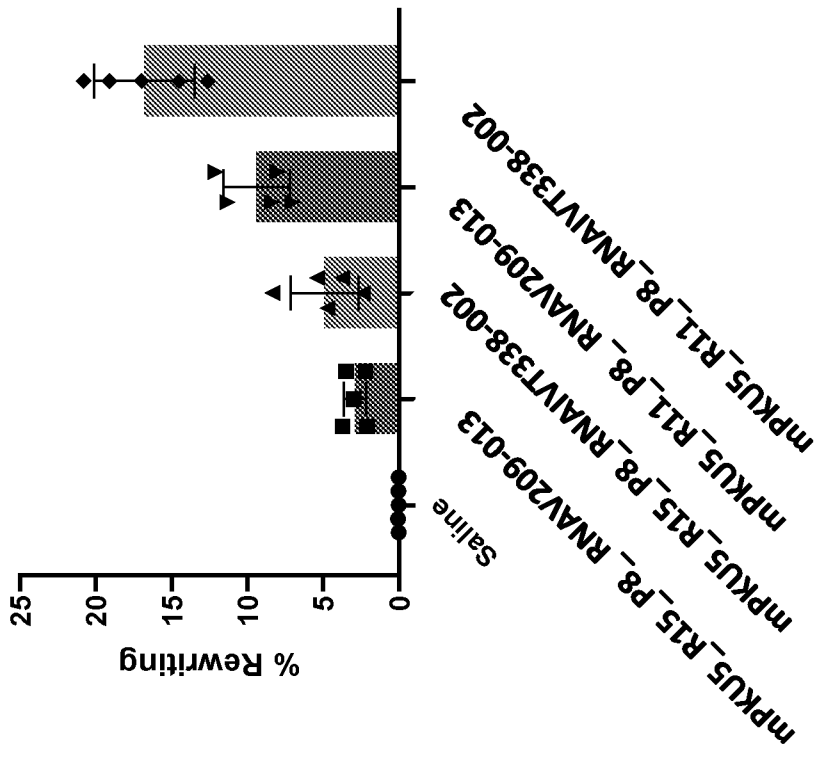


FIG. 19B

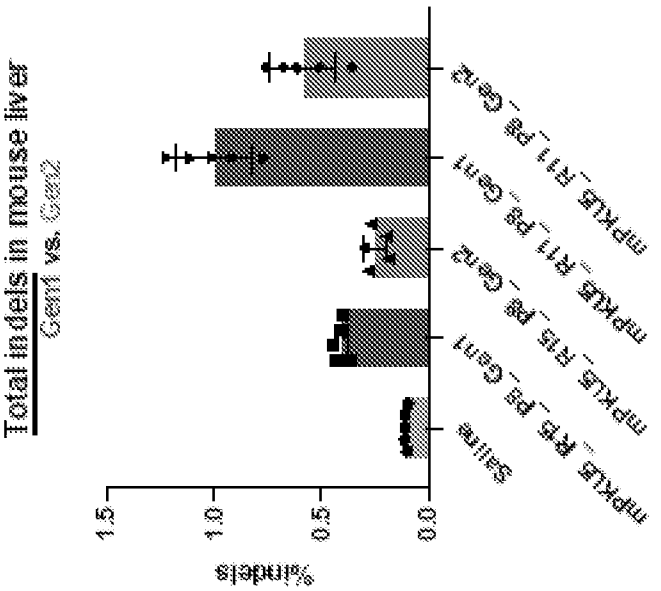


FIG. 20A

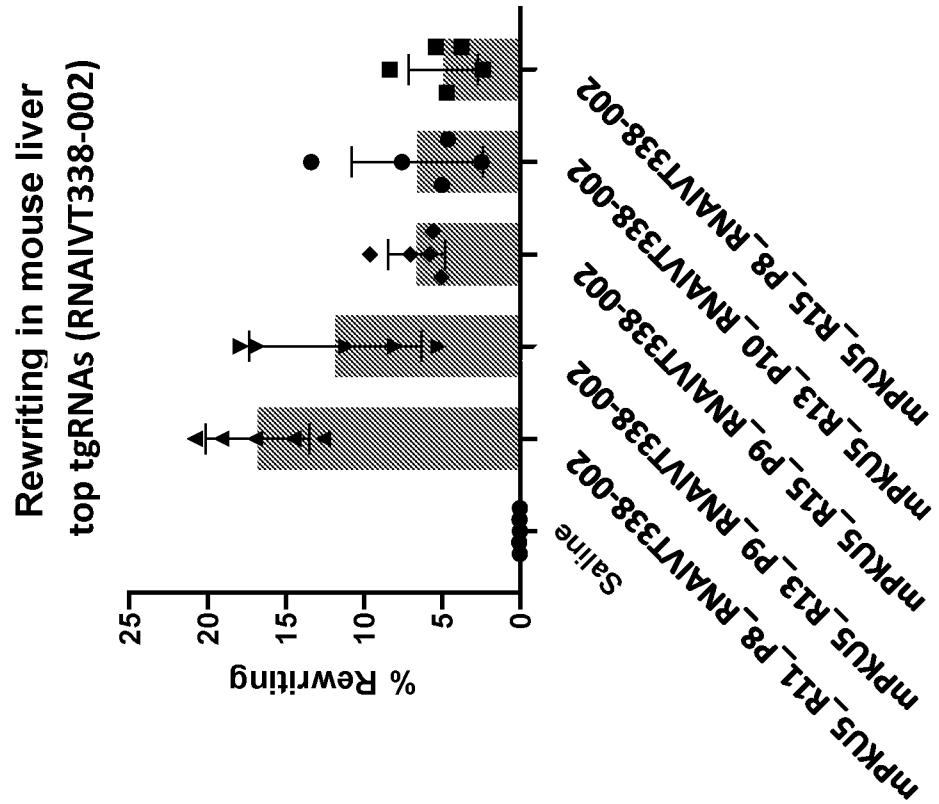


FIG. 20B

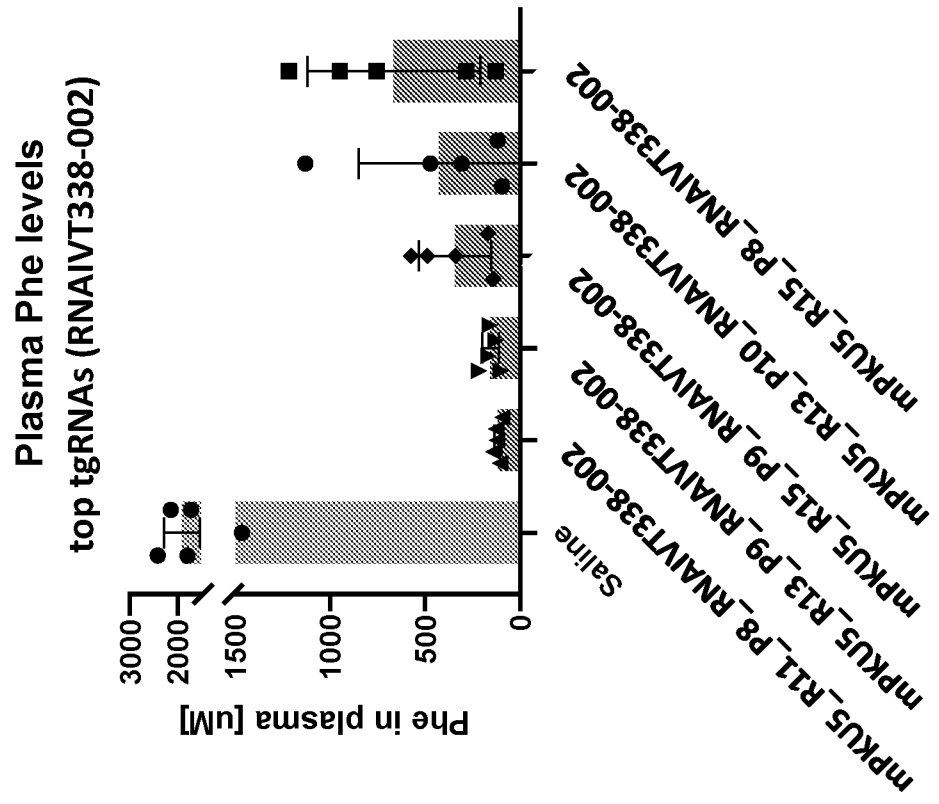


FIG. 20C

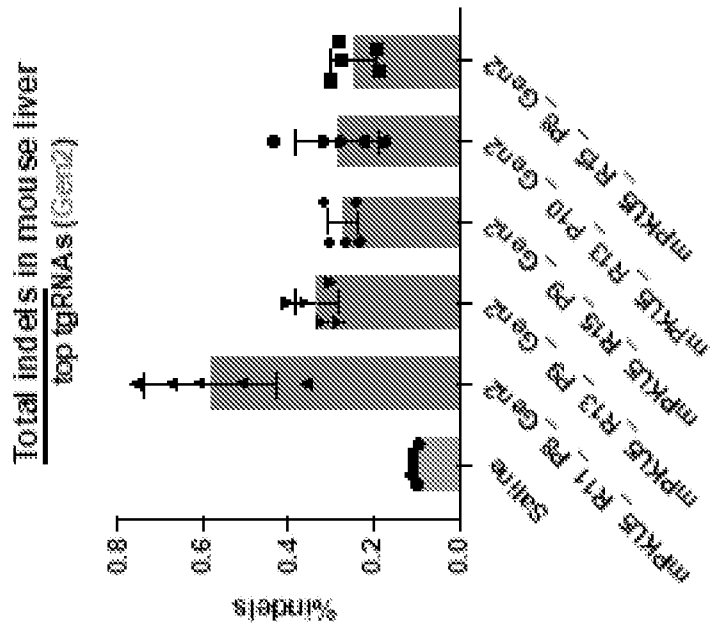

Total indels in mouse liver
top tgRNAs (Gene2)
%Indels
0.0
0.2
0.4
0.6
0.8
Saline
mPKLB_R11_P9_Gene2
mPKLB_R13_P9_Gene2
mPKLB_R15_P9_Gene2
mPKLB_R13_P10_Gene2
mPKLB_R15_P9_Gene2

FIG. 21A

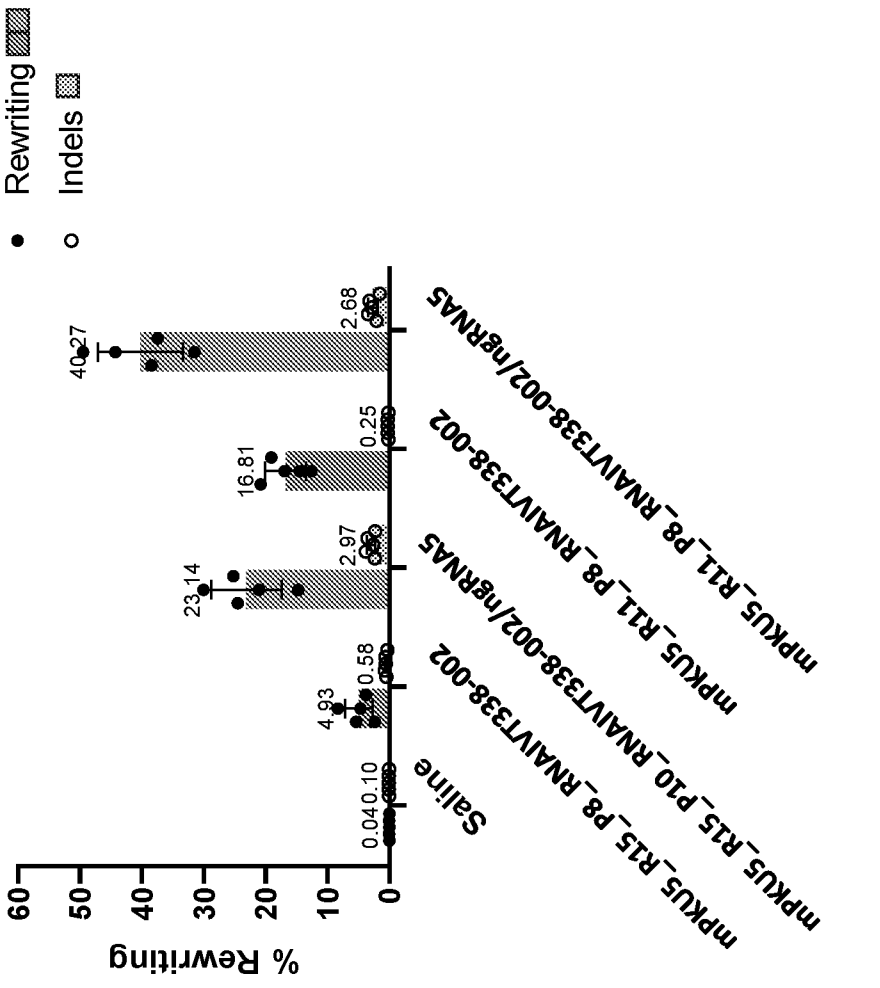


FIG. 21B

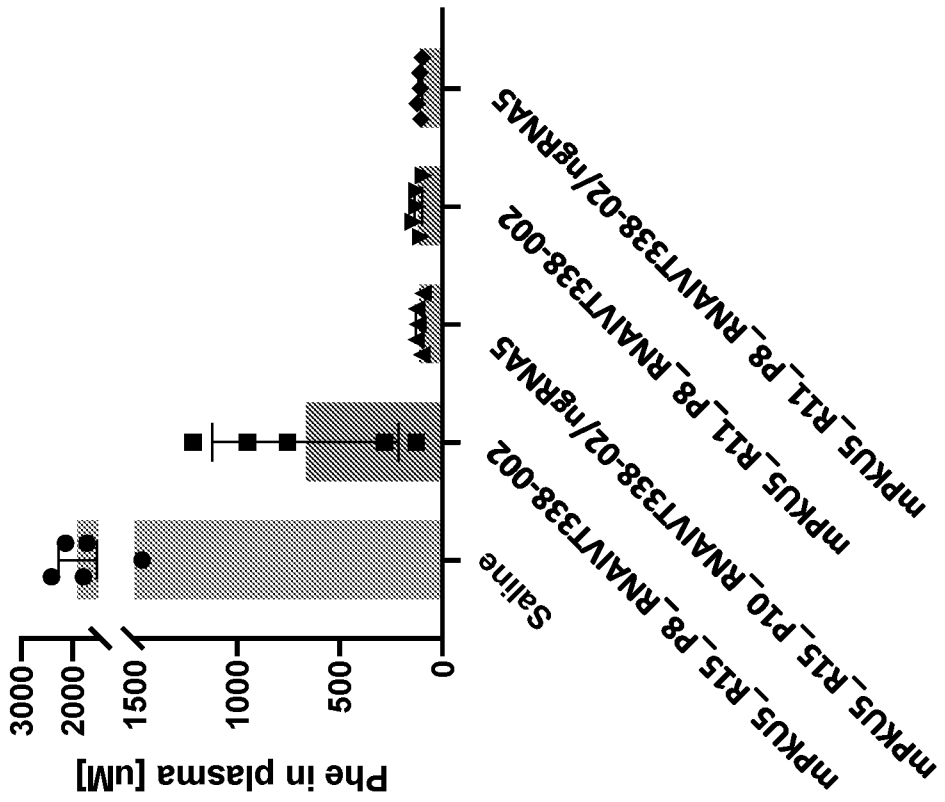
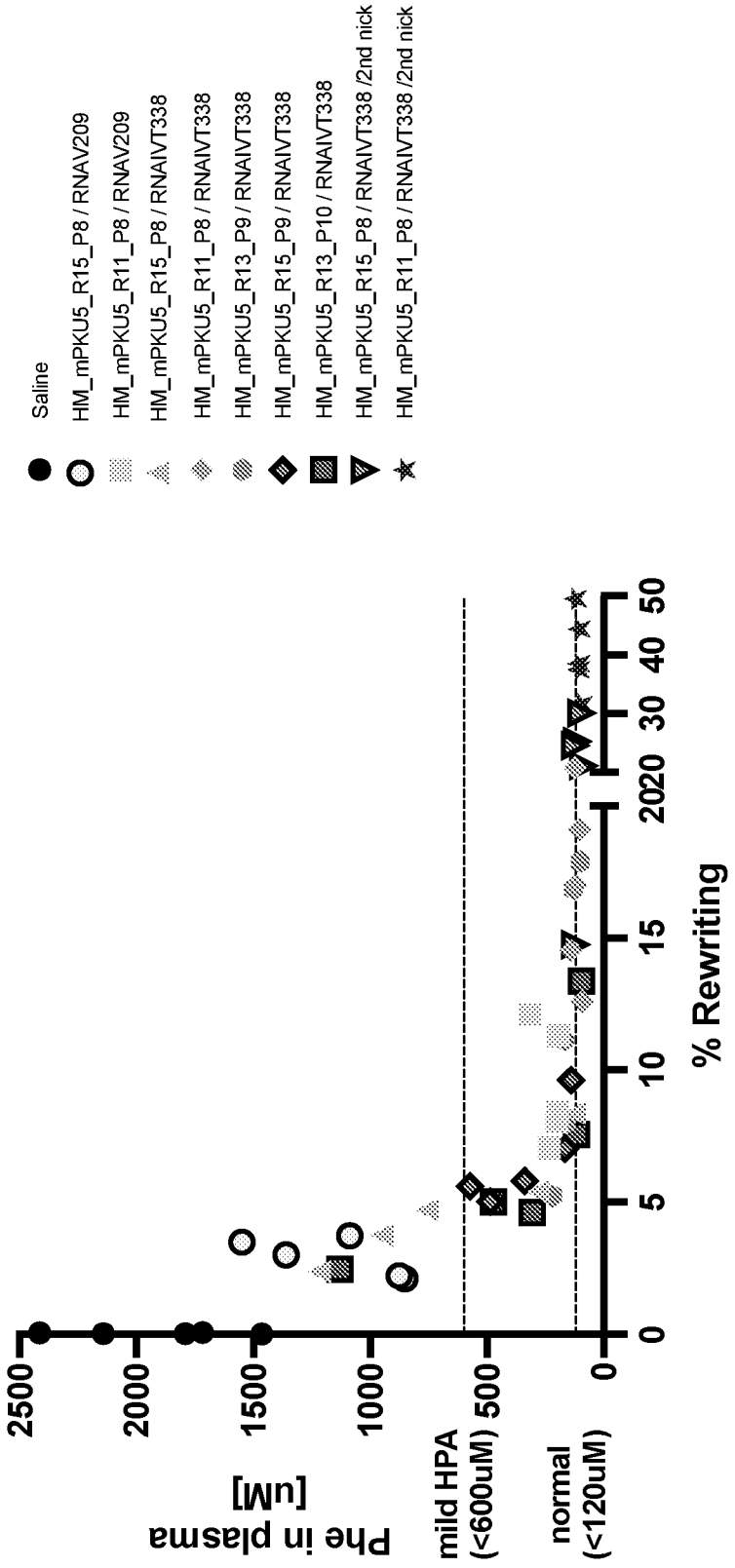


FIG. 22


AmpSeq vs Phe correlation
Saline
HM_mPKU5_R15_P8 / RNAV209
HM_mPKU5_R11_P8 / RNAV209
HM_mPKU5_R15_P8 / RNAIVT338
HM_mPKU5_R11_P8 / RNAIVT338
HM_mPKU5_R13_P9 / RNAIVT338
HM_mPKU5_R15_P9 / RNAIVT338
HM_mPKU5_R13_P10 / RNAIVT338
HM_mPKU5_R15_P8 / RNAIVT338 /2nd nick
HM_mPKU5_R11_P8 / RNAIVT338 /2nd nick
Phe in plasma [uM]
2500
2000
1500
1000
500
0
mild HPA (<600uM)
normal (<120uM)
0
5
10
15
2020
30
40
50
% Rewriting

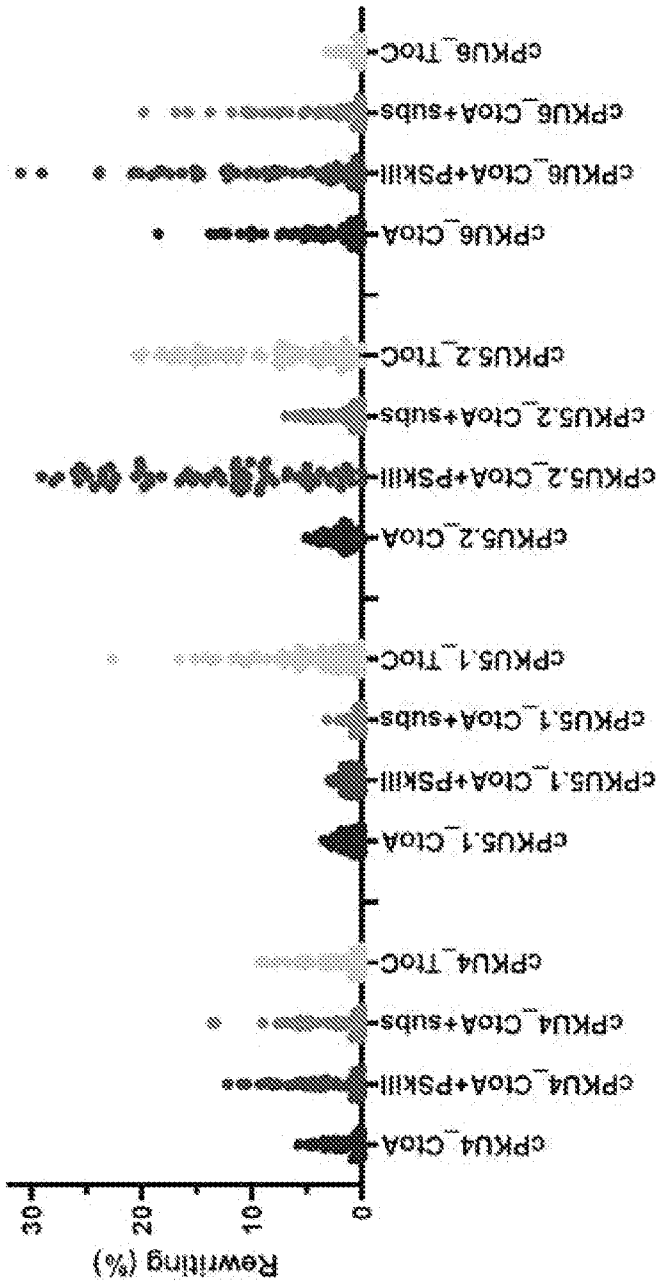

Rewriting (%)
0
10
20
30
cPKU4_CtoA
cPKU4_CtoA+PSkIII
cPKU4_CtoA+subs
cPKU4_TtoC
cPKU5.1_CtoA
cPKU5.1_CtoA+PSkIII
cPKU5.1_CtoA+subs
cPKU5.1_TtoC
cPKU5.2_CtoA
cPKU5.2_CtoA+PSkIII
cPKU5.2_CtoA+subs
cPKU5.2_TtoC
cPKU6_CtoA
cPKU6_CtoA+PSkIII
cPKU6_CtoA+subs
cPKU6_TtoC

FIG. 23

FIG. 24A

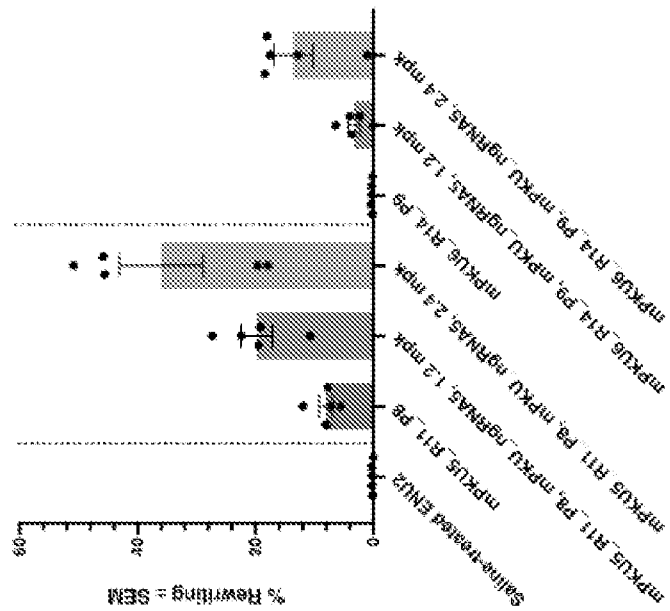


FIG. 24B

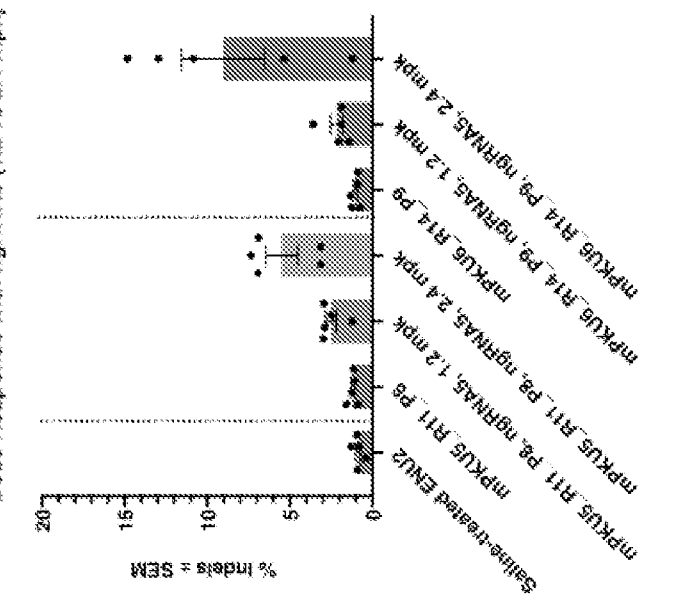


FIG. 24C

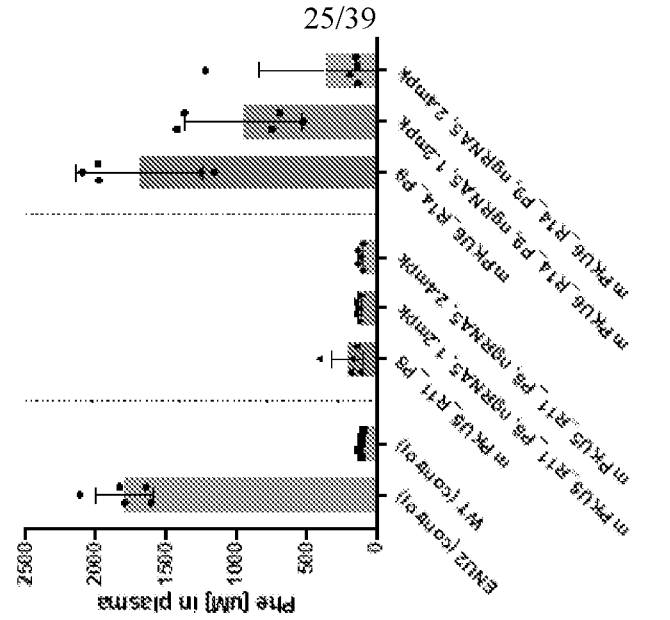


FIG. 25A

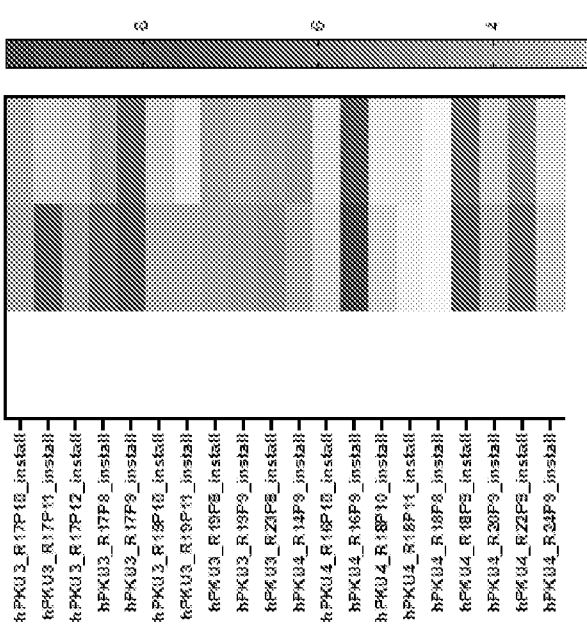


FIG. 25B

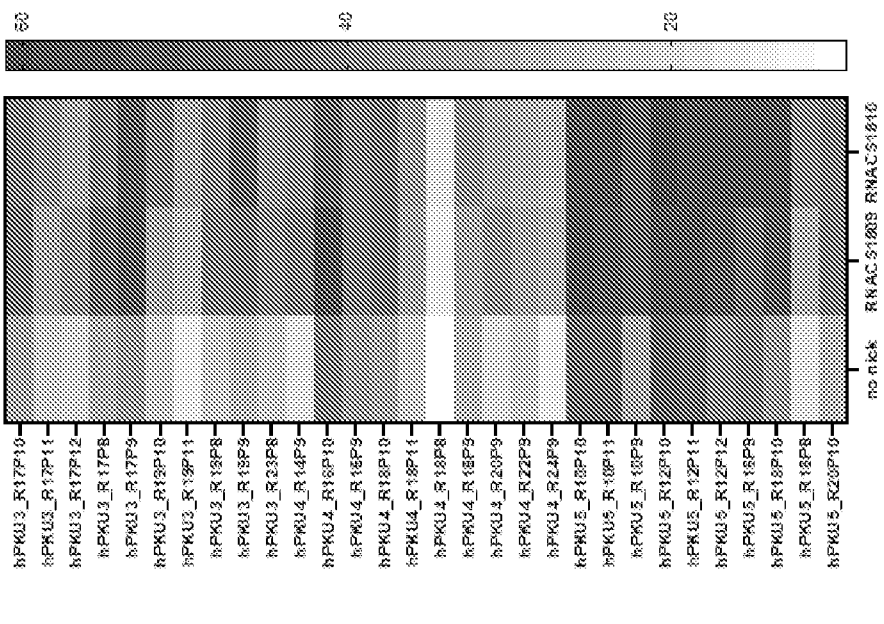


FIG. 25C

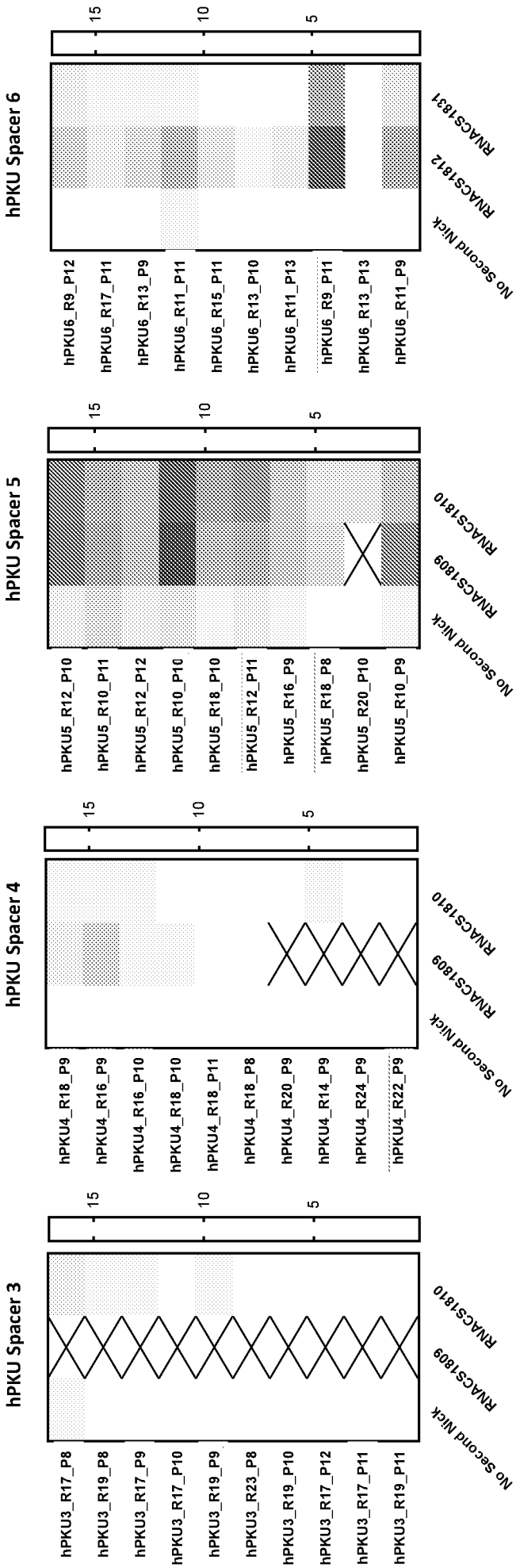

hPKU Spacer 3
hPKU3_R17_P8
hPKU3_R19_P8
hPKU3_R17_P9
hPKU3_R17_P10
hPKU3_R19_P9
hPKU3_R23_P8
hPKU3_R19_P10
hPKU3_R17_P12
hPKU3_R17_P11
hPKU3_R19_P11
hPKU Spacer 4
hPKU4_R18_P9
hPKU4_R16_P9
hPKU4_R16_P10
hPKU4_R18_P10
hPKU4_R18_P11
hPKU4_R18_P8
hPKU4_R20_P9
hPKU4_R14_P9
hPKU4_R24_P9
hPKU4_R22_P9
hPKU Spacer 5
hPKU5_R12_P10
hPKU5_R10_P11
hPKU5_R12_P12
hPKU5_R10_P10
hPKU5_R18_P10
hPKU5_R12_P11
hPKU5_R16_P9
hPKU5_R18_P8
hPKU5_R20_P10
hPKU5_R10_P9
hPKU Spacer 6
hPKU6_R9_P12
hPKU6_R17_P11
hPKU6_R13_P9
hPKU6_R11_P11
hPKU6_R15_P11
hPKU6_R13_P10
hPKU6_R11_P13
hPKU6_R9_P11
hPKU6_R13_P13
hPKU6_R11_P9
15
10
5
No Second Nick
RNACS1809
RNACS1810
RNACS1812
RNACS1831

FIG. 26B

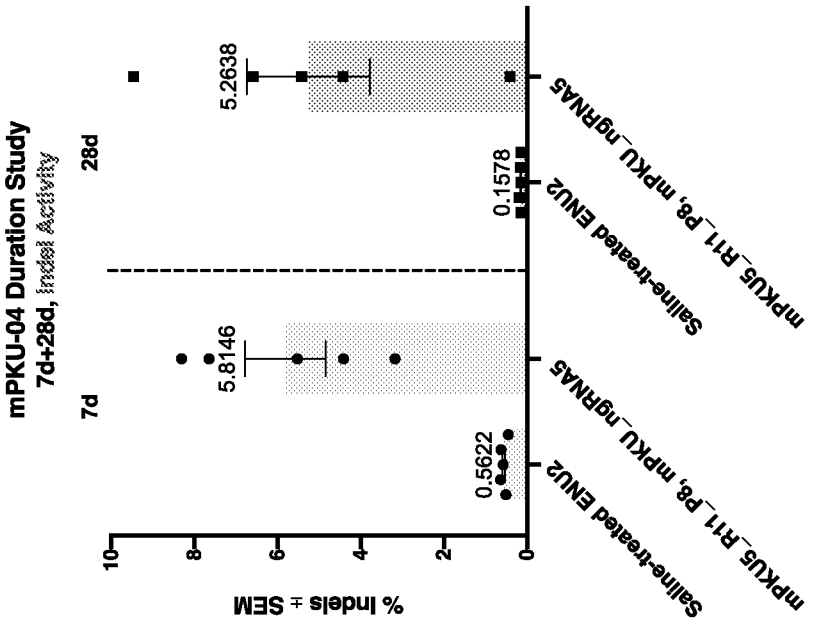


FIG. 26A

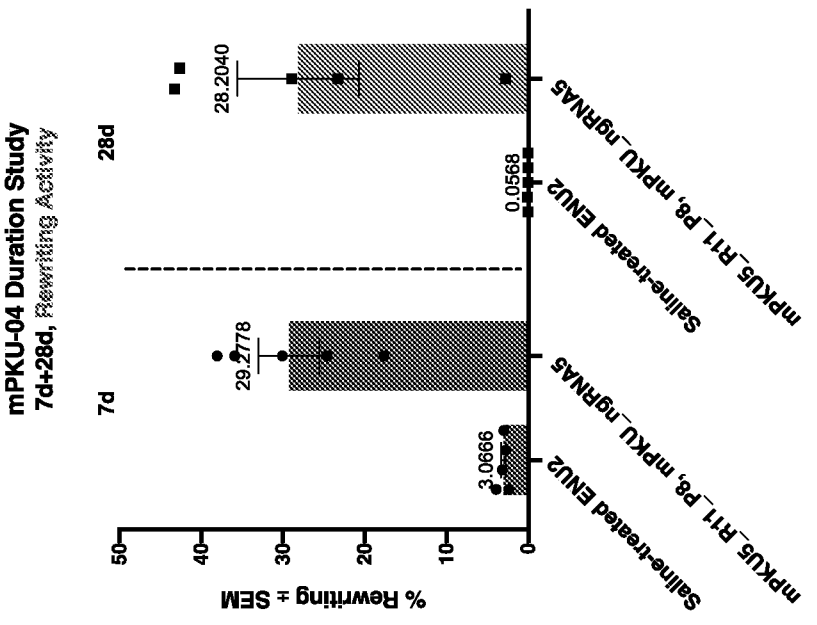


FIG. 27

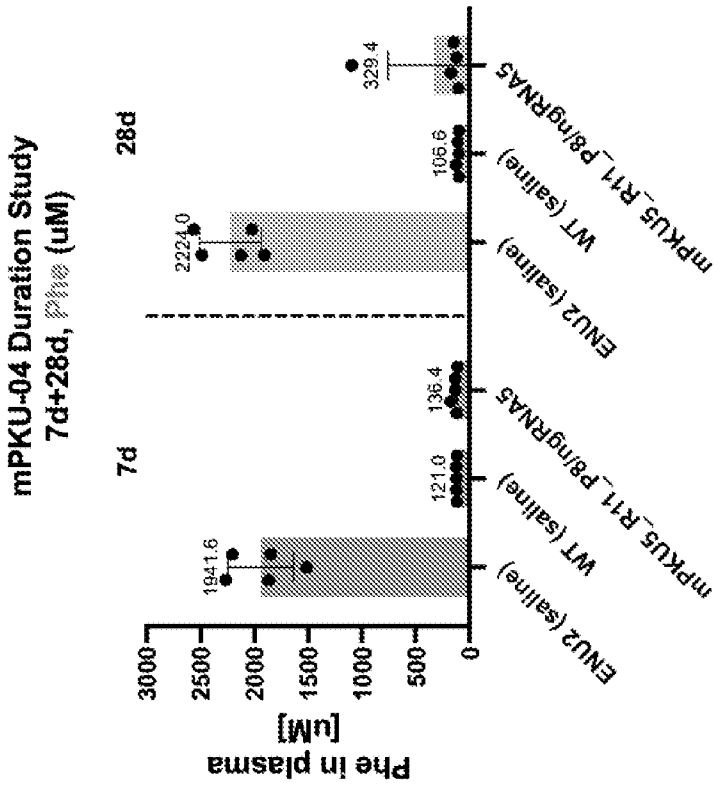

mPKU-04 Duration Study
7d+28d, Phe (uM)
7d
28d
Phe in plasma
[uM]
3000
2500
2000
1500
1000
500
0
1941.6
121.0
136.4
2224.0
106.6
329.4
ENU2 (saline)
WT (saline)
mPKU5_R11_P8/ngRNA5
ENU2 (saline)
WT (saline)
mPKU5_R11_P8/ngRNA5

FIG. 28

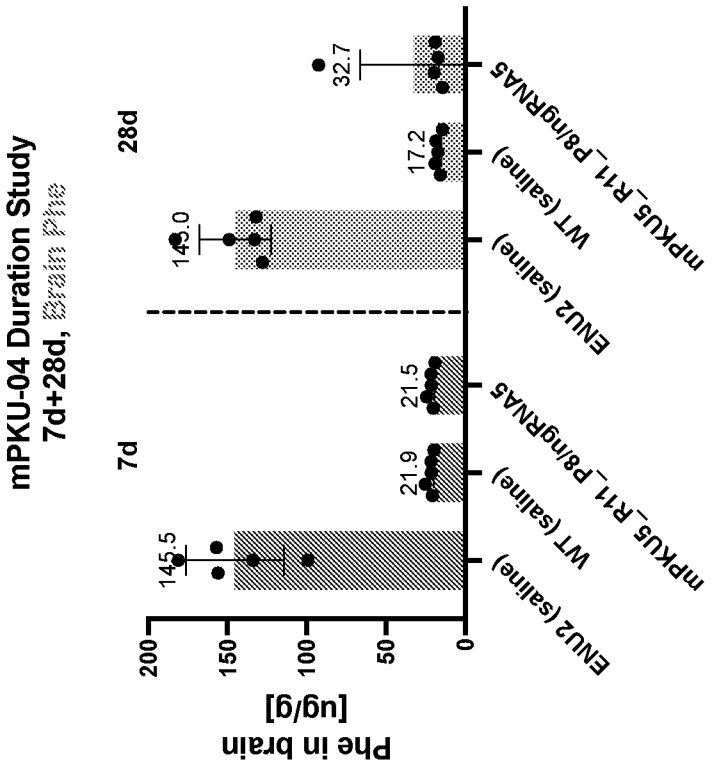

mPKU-04 Duration Study
7d+28d, Brain Phe
Phe in brain [ug/g]
200
150
100
50
0
7d
28d
145.5
21.9
21.5
145.0
17.2
32.7
ENU2 (saline)
WT (saline)
mPKU5_R11_P8/ngRNA5
ENU2 (saline)
WT (saline)
mPKU5_R11_P8/ngRNA5

FIG. 29

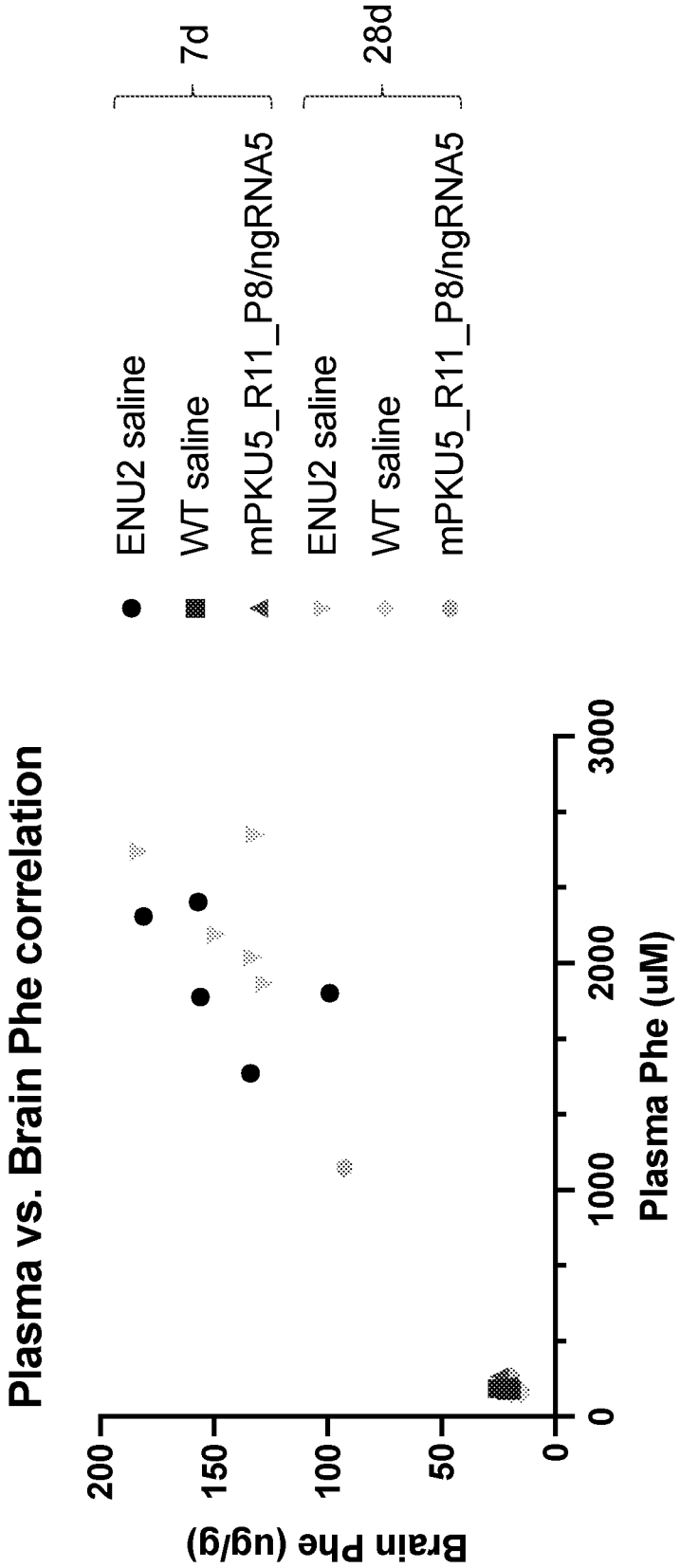

Plasma vs. Brain Phe correlation
Brain Phe (ug/g)
Plasma Phe (uM)
0
50
100
150
200
0
1000
2000
3000
7d
28d
ENU2 saline
WT saline
mPKU5_R11_P8/ngRNA5
ENU2 saline
WT saline
mPKU5_R11_P8/ngRNA5

FIG. 30A

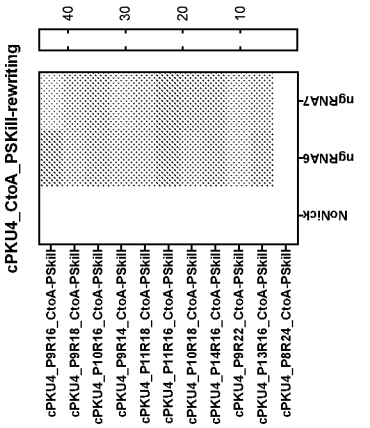


FIG. 30B

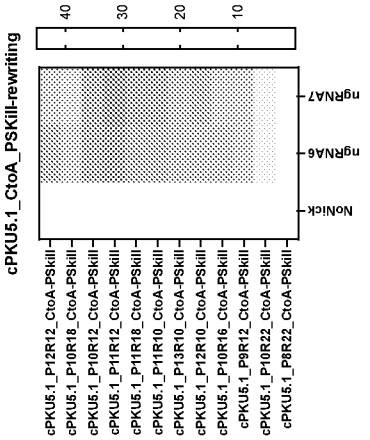


FIG. 30C

FIG. 30D

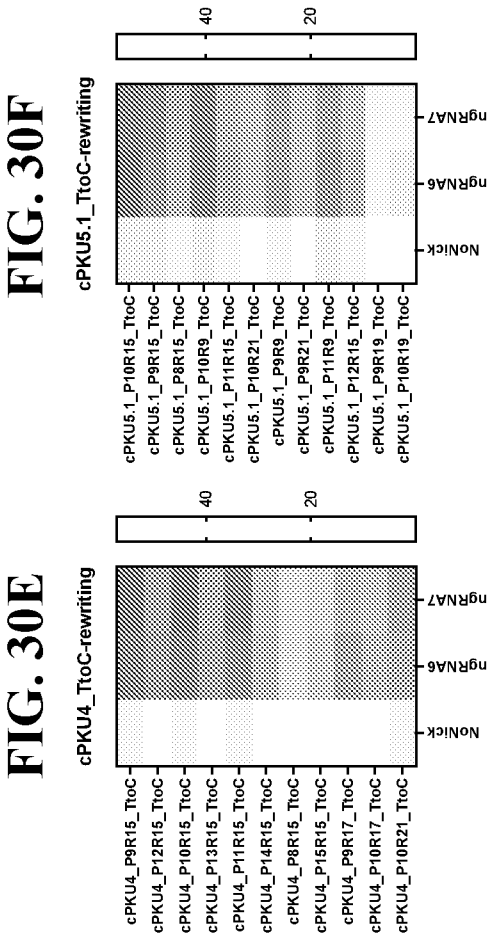


FIG. 30E

FIG. 30F

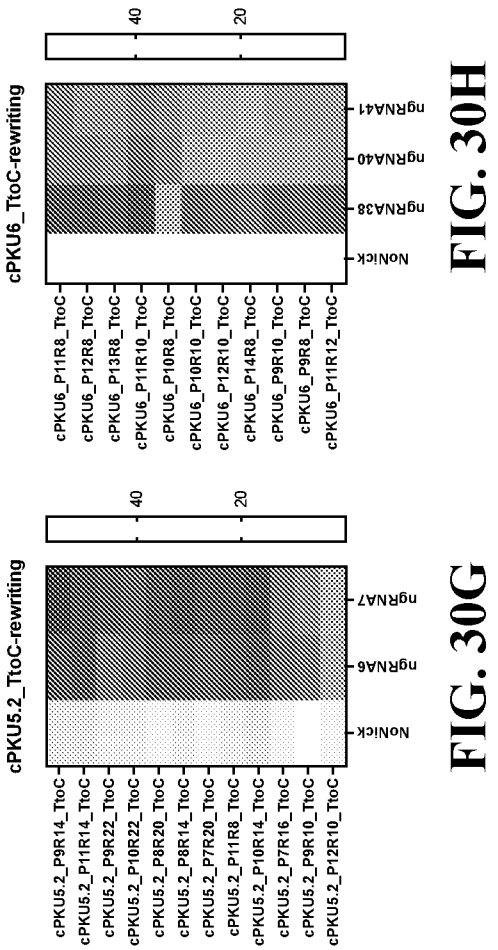


FIG. 30G

FIG. 30H

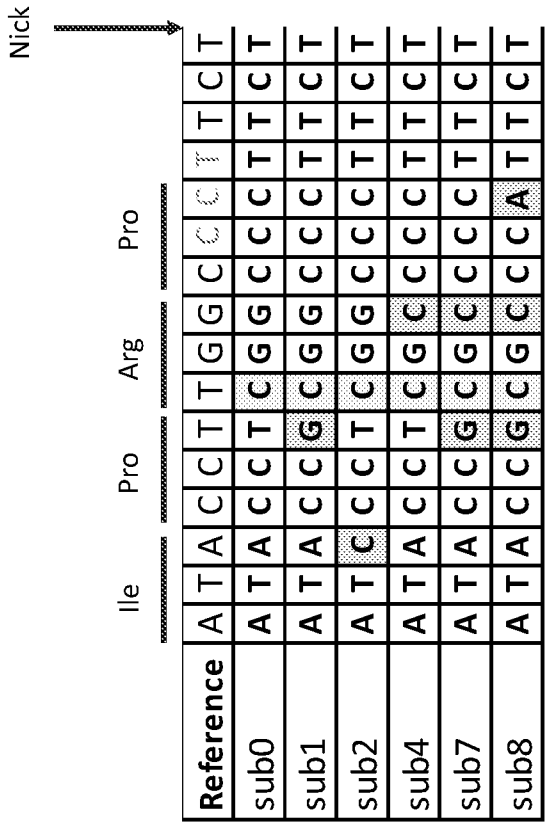


FIG. 31B

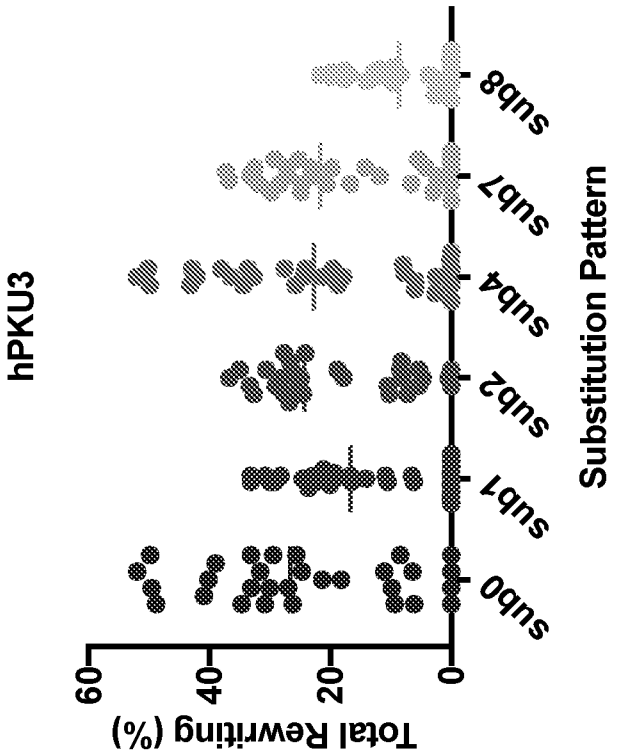


FIG. 31A

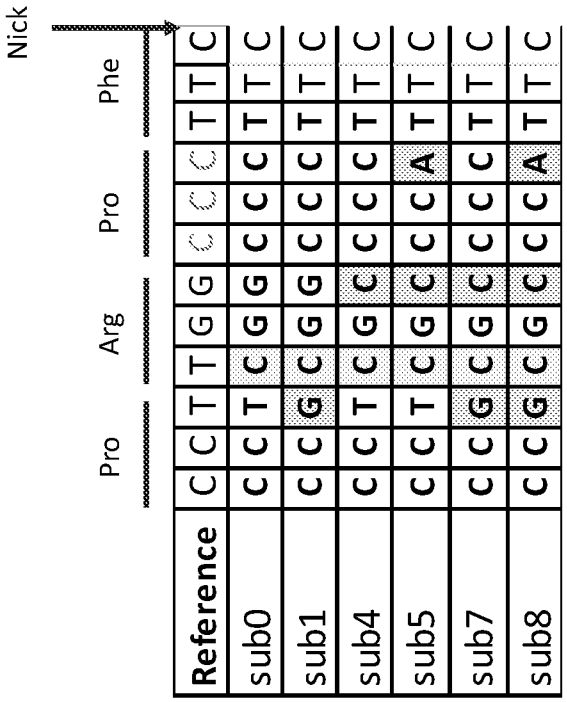


FIG. 32B

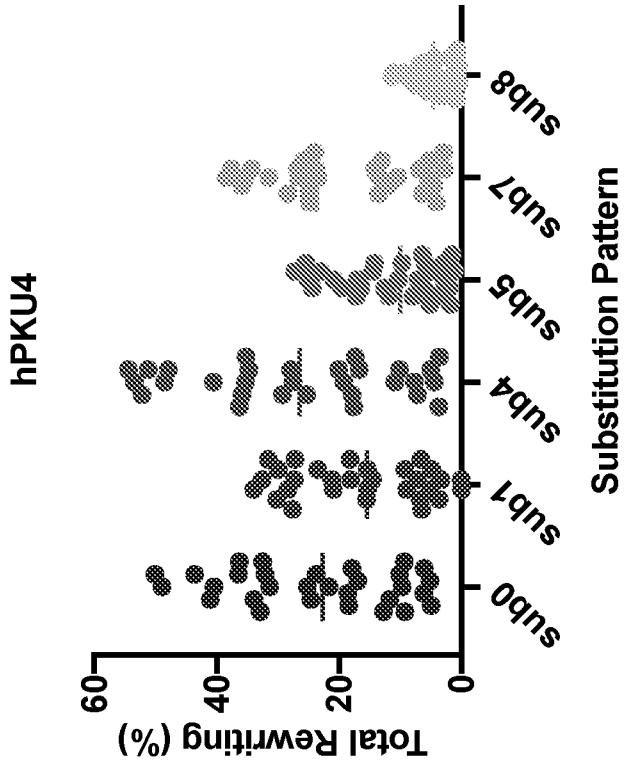


FIG. 32A

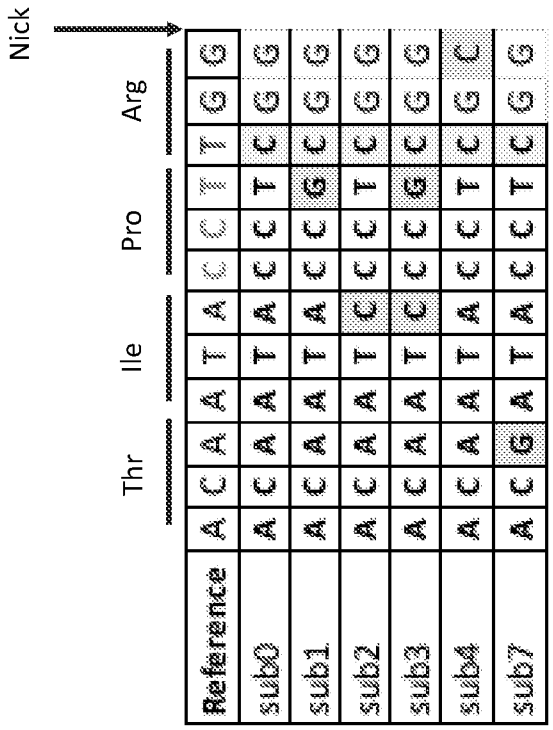


FIG. 33B

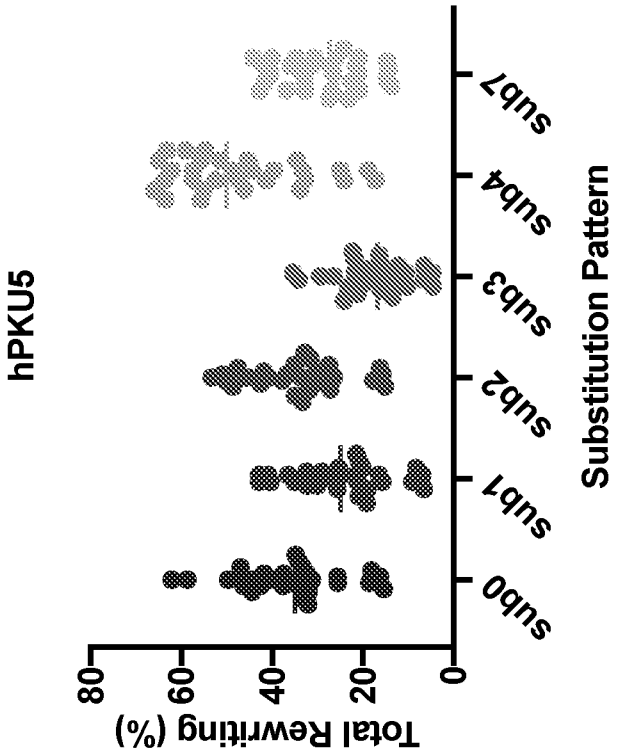


FIG. 33A

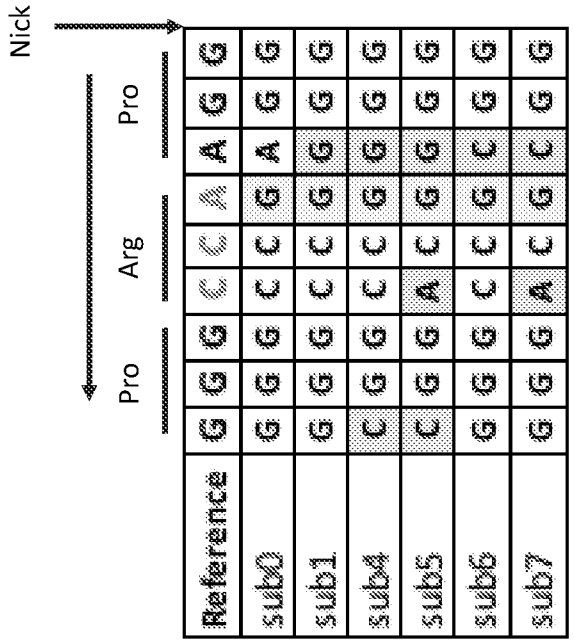


FIG. 34B

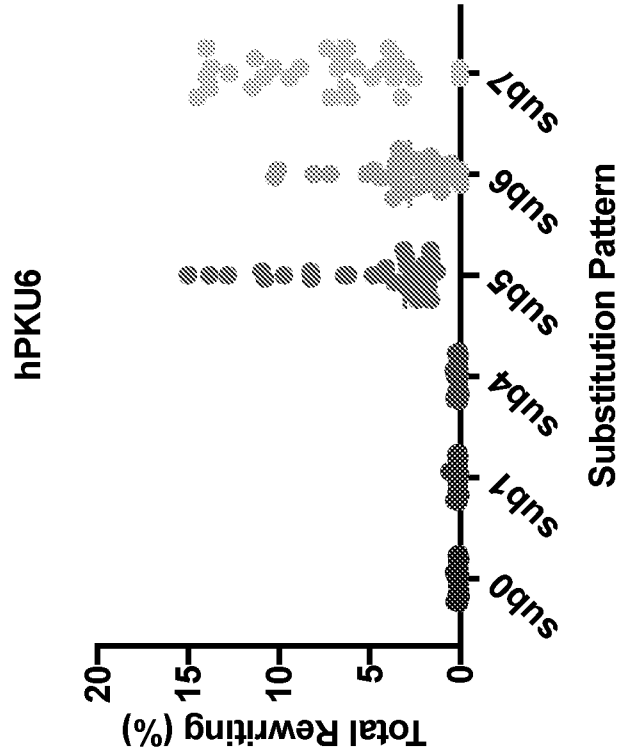


FIG. 34A

FIG. 35

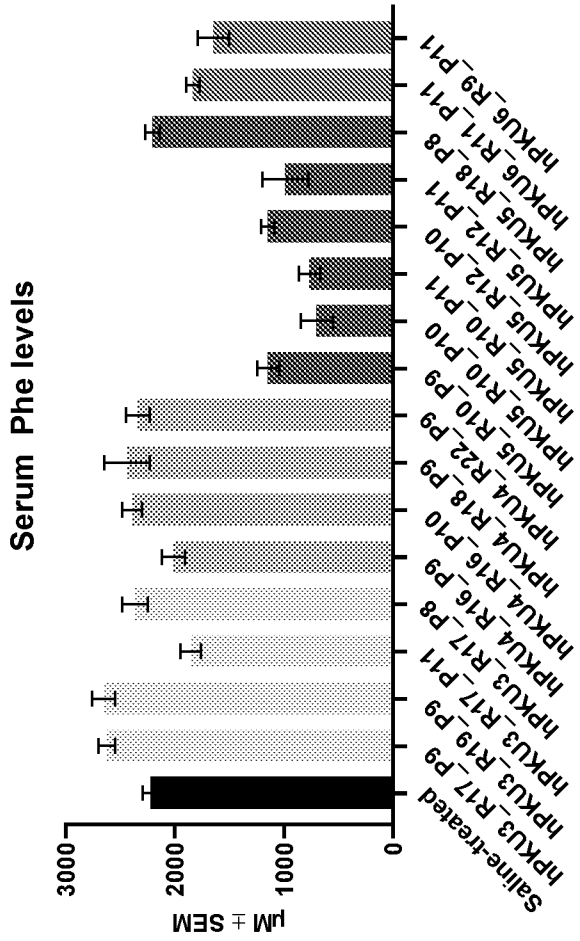

Serum Phe levels
μM ± SEM
3000
2000
1000
0
Saline-treated
hPKU3_R17_P9
hPKU3_R19_P9
hPKU3_R17_P11
hPKU4_R17_P8
hPKU4_R16_P9
hPKU4_R16_P10
hPKU4_R18_P9
hPKU4_R22_P9
hPKU5_R10_P9
hPKU5_R10_P10
hPKU5_R10_P11
hPKU5_R12_P10
hPKU5_R12_P11
hPKU5_R18_P8
hPKU6_R11_P11
hPKU6_R9_P11

FIG. 36A

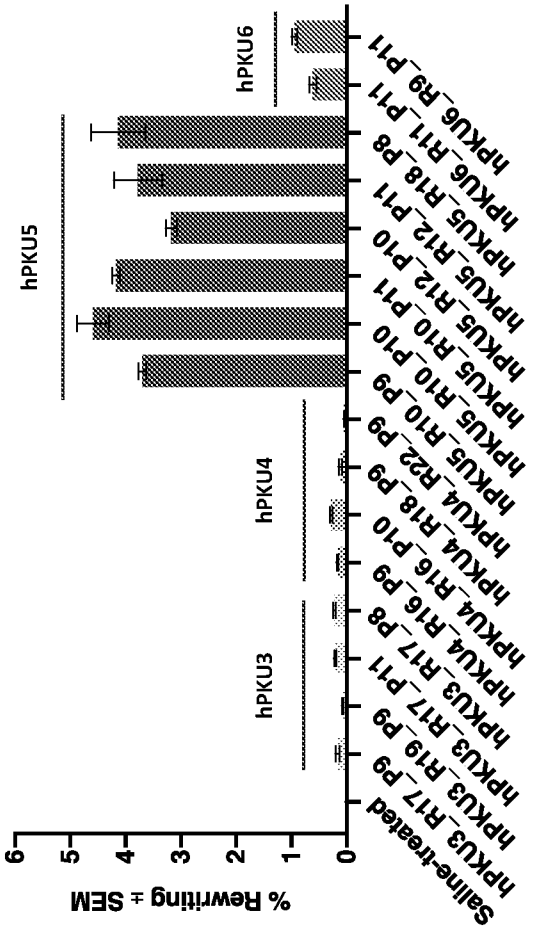


FIG. 36B