

MPOX VIRUS NUCLEIC ACID VACCINE

TECHNICAL FIELD

The present invention relates to the fields of biomedicine and virology, and in particular to a nucleic acid vaccine (preferably an mRNA vaccine). for preventing and/or treating mpox virus infections.

BACKGROUND ART

Monkeypox is a zoonotic disease caused by mpox virus (MPXV). Mpox virus is a class of enveloped DNA virus, with a diameter of about 200-250 nm, and is one of the members of the family Poxviridae and the genus Orthopoxvirus (see, for example, Alakunle E et al., Mpox virus in Nigeria: Infection Biology, Epidemiology, and Evolution. Viruses. 2020 Nov 5;12(11):1257. and Cho CT, Wenner HA. Mpox virus. Bacteriol Rev. 1973 Mar;37(1):1-18.). Although on May 12, 2023, the World Health Organization (World Health Organization, WHO) declared that the monkeypox outbreak no longer constituted a public health emergency of international concern, the number of monkeypox cases worldwide is still maintaining growth.

Transmission routes of mpox virus include direct contact transmission and droplet transmission, etc., and its incubation period is 7-14 days, and may be up to 21 days at most (see, for example, Moore MJ et al., Mpox (Monkeypox). 2023 May 3. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan.). The main symptoms of monkeypox include high fever, headache, skin rash, muscle pain and local lymph node enlargement, etc., most of which are transient, wherein the disease course lasts for 2-4 weeks and then may recover spontaneously, but patients with skin rash may have erythema, pigmentation or even scars remaining after scabs fall off, wherein the scars may last for as long as several years. For people with low immunity (such as AIDS patients), severe monkeypox infection may also cause death.

Mpox virus, highly pathogenic variola virus and vaccinia virus all belong to the genus Orthopoxvirus, and antigen proteins among the three viruses have relatively high homology. Considering that mpox virus needs to be operated in a biosafety level 3 laboratory, and that there is good cross-reactivity between vaccinia virus and mpox virus (see, for example, Zeng J et al., Mpox multi-antigen mRNA vaccine candidates by a simplified manufacturing strategy afford efficient protection against lethal orthopoxvirus challenge. Emerg Microbes Infect. 2023

Dec;12(1):2204151.), therefore, vaccinia virus is mostly used in currently published scientific articles to evaluate the immunogenicity of monkeypox vaccine molecules. In addition, there is also evidence showing that inoculation with one orthopoxvirus vaccine can provide cross-protective immunity against other orthopoxviruses (see, for example, Lum FM et al., Monkeypox: disease epidemiology, host immunity and clinical interventions. Nat Rev Immunol. 2022 Oct;22(10):597-613.).

At present, only two smallpox vaccines ACAM2000 and JYNNEOS developed based on vaccinia virus antigens have been approved by the United States Food and Drug Administration (Food and Drug Administration, FDA) for immunization for preventing monkeypox; however, because these two vaccines perform poorly in terms of safety and are associated with serious adverse events such as myocarditis and pericarditis, they are only approved for pre-exposure immunization of high-risk populations. Since the smallpox vaccination program was stopped in 1980, herd immunity of the general public to orthopoxviruses has declined, and studies have shown that monkeypox has a higher incidence rate in populations not inoculated with a smallpox vaccine (see, for example, Shah HH et al., Human monkeypox - After 40 years, an unintended consequence of smallpox eradication. Front Public Health. 2023 Jan 5;10:1082586. and Pattnaik H et al., Making Sense of Monkeypox: A Comparison of Other Poxviruses to the Monkeypox. Cureus. 2023 Apr 24;15(4):e38083.). Therefore, a safe and effective vaccine developed against mpox virus is still urgently needed.

SUMMARY OF THE INVENTION

In a first aspect, the present invention provides a polypeptide or polypeptide combination, comprising one or more peptide fragments, wherein the peptide fragments are each independently selected from peptide fragments of mpox virus A29L, A35R, B2R, B6R, C15L, M1R, and H3L proteins; optionally, the polypeptide or polypeptide combination further comprises one or more linkers, an IgE signal peptide and/or an HA polypeptide tag.

In some embodiments, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110, or an amino acid sequence having at least 80% identity to the amino

acid sequence as set forth in one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110.

In some embodiments, the polypeptide or polypeptide combination comprises two or more amino acid sequences selected from SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110, or an amino acid sequence having at least 80% identity to the amino acid sequence of one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110.

In a second aspect, the present invention provides a nucleic acid or nucleic acid combination, comprising a polynucleotide encoding the polypeptide or polypeptide combination of the present invention.

In some embodiments, the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA, and the polynucleotide comprises the nucleotide sequence as set forth in one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22, or a nucleotide sequence having at least 80% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22. In some embodiments, the polynucleotide is DNA, and the polynucleotide comprises the nucleotide sequence as set forth in one of SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44, or a nucleotide sequence having at least 80% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44.

In some embodiments, the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA, and the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22, or two or more nucleotide sequences having at least 80% identity to nucleotide sequences selected from SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22. In some embodiments, the polynucleotide is DNA, and the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44, or two or more nucleotide sequences having at least 80% identity to nucleotide sequences selected from SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44.

In some embodiments, the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA, and the polynucleotide comprises the nucleotide sequence as set forth in one of SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66, or a nucleotide sequence having at least 80% identity to the nucleotide sequence as set forth in one of
5 SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66. In some embodiments, the polynucleotide is DNA, and the polynucleotide comprises the nucleotide sequence as set forth in one of SEQ ID NOs: 67, 68, 69, 70, 71, 72, 73, 79, 80, 81, 82, 83, 84, 85, 86, 87 and 88, or a nucleotide sequence having at least 80% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 67, 68, 69, 70, 71, 72, 73, 79, 80, 81, 82, 83, 84, 85, 86, 87 and
10 88.

In some embodiments, the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA, and the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66, or two or more nucleotide sequences having at least 80% identity to nucleotide sequences selected
15 from SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66. In some embodiments, the polynucleotide is DNA, and the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 67, 68, 69, 70, 71, 72, 73, 79, 80, 81, 82, 83, 84, 85, 86, 87 and 88, or two or more nucleotide sequences having at least 80% identity to nucleotide sequences selected from SEQ ID NOs: 67, 68, 69, 70, 71, 72, 73, 79, 80, 81, 82, 83, 84, 85, 86, 87
20 and 88.

In a third aspect, the present invention provides a composition, comprising the nucleic acid or nucleic acid combination of the present invention.

In some embodiments, the composition comprises a first lipid, a second lipid, a phospholipid and a steroid, each as defined herein. In some embodiments, the first lipid is a cationic lipid. In
25 some embodiments, the composition is formulated as a lipid nanoparticle that encapsulates the nucleic acid or nucleic acid combination of the present invention in lipids.

In a fourth aspect, the present invention provides a pharmaceutical composition or kit, comprising the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination, or composition of the present invention.

In a fifth aspect, the present invention provides an expression vector, comprising the nucleic acid or nucleic acid combination of the present invention.

In a sixth aspect, the present invention provides a host cell, comprising the nucleic acid or nucleic acid combination or expression vector of the present invention.

5 In a seventh aspect, the present invention provides use of the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination, composition, pharmaceutical composition or kit of the present invention in the preparation of a vaccine for preventing and/or treating mpox virus infection in a subject.

DESCRIPTION OF THE DRAWINGS

10 FIG. 1 shows the results of detecting vaccinia virus cross-neutralizing antibody titers induced by candidate monkeypox vaccines in mice.

FIG. 2A and FIG. 2B show the results of detecting immune responses of mouse spleen T cells to a mpox virus mixed peptide pool.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

General Definitions and Terms

15 All patents, patent applications, scientific publications, manufacturers' instructions and guides, etc. cited herein, whether above or below, are incorporated herein by reference in their entirety. Nothing herein should be construed as an admission that the present disclosure is not entitled to antedate such disclosure.

20 Unless otherwise specified, scientific and technical terms used herein have meanings commonly understood by those skilled in the art. In addition, terms related to protein and nucleic acid chemistry, molecular biology, cell and tissue culture, and microbiology used herein are all terms widely used in the corresponding fields (see, for example, *Molecular Cloning: A Laboratory Manual*, 2nd Edition, J. Sambrook et al. eds., Cold Spring Harbor Laboratory Press, Cold Spring
25 Harbor 1989). Meanwhile, in order to better understand the present invention, definitions and explanations of relevant terms are provided below.

As used herein, the expressions "comprise", "include", "contain" and "have" are open-ended and mean the inclusion of recited elements, steps or components, but not the exclusion of other unrecited elements, steps or components. The expression "consist of" does not include any element,

step or component that is not specified. The expression "consist essentially of" means that the scope is limited to the specified elements, steps or components, plus optional elements, steps or components that do not significantly affect the basic and novel properties of the claimed subject matter. It should be understood that the expressions "consist essentially of" and "consist of" are encompassed within the meaning of the expression "comprise".

As used herein, unless the context indicates otherwise, the singular forms of the expressions "a", "an" or "the" include plural references. The terms "one or more" or "at least one" encompass 1, 2, 3, 4, 5, 6, 7, 8, 9 or more.

As used herein, the connective term "and/or" between a plurality of recited elements should be understood to include individual and combined options. In other words, "and/or" includes "and" as well as "or". For example, A and/or B includes A, B and A+B. A, B and/or C includes A, B, C and any combination thereof, for example A+B, A+C, B+C and A+B+C. More elements defined by "and/or" are understood in a similar manner and include any one thereof and any combination thereof.

The recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each distinct value falling within the range. Unless otherwise specified herein, each individual value is incorporated into the specification as if it were individually recited herein. Unless expressly indicated to the contrary, numerical values or ranges shown herein are modified by "about", which means $\pm 20\%$, $\pm 10\%$, $\pm 5\%$ or $\pm 3\%$ of the recited or claimed numerical value or range.

As used herein, a "nucleic acid combination" may mean a combination or collection comprising more than one nucleic acid. Such nucleic acids may be present separately, for example, in different compositions, different parts of pharmaceutical compositions or kits, or different vectors, or may be present simultaneously in the same composition, the same part of a pharmaceutical composition or kit, or the same vector. The nucleic acids in the nucleic acid combination may be the same or different from each other, and may each encode the same polypeptide (part) or may encode different polypeptides (parts).

As used herein, a "polypeptide combination" may mean a combination or collection comprising more than one polypeptide. Such polypeptides may be present separately, for example,

in different parts of different pharmaceutical compositions or kits, or may be present simultaneously in the same part of the same pharmaceutical composition or kit. The polypeptides in the polypeptide combination may be the same or different from each other, and may each comprise the same amino acid sequence (part) or may comprise different amino acid sequences (parts).

As used herein, the term "wild-type" means that the sequence is naturally occurring and has not been artificially modified, including naturally occurring mutants.

As used herein, the term "% identity" with respect to a sequence refers to the percentage of identical nucleotides or amino acids in the optimal alignment between sequences to be compared.

Differences between two sequences may be distributed over a local region (segment) or over the entire length of the sequences to be compared. Generally, identity between two sequences is determined after optimal alignment of a segment or "comparison window". Optimal alignment may be performed manually or with the aid of algorithms known in the art, including but not limited to Smith and Waterman, 1981, *Adv. App. Math.* 2, 482 and the local homology algorithm described in Neddleman and Wunsch, 1970, *J. Mol. Biol.* 48, 443, the similarity search method described in Pearson and Lipman, 1988, *Proc. Natl Acad. Sci. USA* 88, 2444, or using computer programs, for example GAP, BESTFIT, FASTA, BLAST P, BLAST N and TFASTA in Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis. For example, the percentage identity of two sequences may be determined using the publicly available BLASTN or BLASTP algorithm on the website of the National Center for Biotechnology Information (NCBI).

% identity is obtained by determining the number of identical positions corresponding to the sequences to be compared, dividing this number by the number of positions compared (for example, the number of positions in the reference sequence), and multiplying this result by 100. In some embodiments, a degree of identity is given over a region of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95% or about 100%. In some embodiments, a degree of identity is given over the entire length of the reference sequence. Alignments for determining sequence identity may be performed using tools known in the art, preferably using

optimal sequence alignment, for example, using Align, using standard settings, preferably EMBOSS:needle, Matrix: Blosum62, Gap Open 10.0, Gap Extend 0.5.

As used herein, "nucleotide" includes deoxyribonucleotides and ribonucleotides and derivatives thereof. As used herein, "ribonucleotide" is a constituent substance of ribonucleic acid (RNA), consisting of one molecule of base, one molecule of pentose and one molecule of phosphate, and refers to a nucleotide having a hydroxyl group at the 2' position of a β -D-ribofuranosyl group. A "deoxyribonucleotide" is a constituent substance of deoxyribonucleic acid (DNA), also consisting of one molecule of base, one molecule of pentose and one molecule of phosphate, and refers to a nucleotide in which the hydroxyl group at the 2' position of the β -D-ribofuranosyl group is substituted by hydrogen, and is the main chemical component of chromosomes. A "nucleotide" is generally referred to by a single letter representing the base therein: "A (a)" refers to deoxyadenylate or adenylate containing adenine, "C (c)" refers to deoxycytidylate or cytidylate containing cytosine, "G (g)" refers to deoxyguanylate or guanylate containing guanine, "U (u)" refers to uridylate containing uracil, and "T (t)" refers to deoxythymidylate containing thymine.

As used herein, the terms "polynucleotide" and "nucleic acid" may be used interchangeably to refer to a polymer of deoxyribonucleotides (deoxyribonucleic acid, DNA) or a polymer of ribonucleotides (ribonucleic acid, RNA). "Polynucleotide sequence", "nucleic acid sequence" and "nucleotide sequence" may be used interchangeably to indicate the ordering of nucleotides in a polynucleotide. Those skilled in the art will understand that a DNA coding strand (sense strand) and the RNA encoded thereby may be regarded as having the same nucleotide sequence, wherein deoxythymidylate in the DNA coding strand sequence corresponds to uridylate in the RNA sequence encoded thereby.

As used herein, the term "expression" includes transcription and/or translation of a nucleotide sequence. Thus, expression may involve production of a transcript and/or a polypeptide. The term "transcription" relates to the process of transcribing genetic code in a DNA sequence into RNA (transcript). The term "in vitro transcription" refers to in vitro synthesis of RNA, particularly mRNA, in a cell-free system (for example, in an appropriate cell extract) (see, for example, Pardi N., Muramatsu H., Weissman D., Karikó K. (2013). In: Rabinovich P. (eds) Synthetic Messenger

RNA and Cell Metabolism Modulation. Methods in Molecular Biology (Methods and Protocols), vol 969. Humana Press, Totowa, NJ.). A vector that can be used to produce a transcript is also referred to as a "transcription vector", which comprises regulatory sequences required for transcription. The term "transcription" encompasses "in vitro transcription".

5 As used herein, "encoding" refers to an inherent property of a particular nucleotide sequence in a nucleic acid, such as a gene, cDNA or mRNA, which can serve as a template for synthesizing polymers and macromolecules in other biological processes, provided that there is a defined nucleotide sequence or a defined amino acid sequence. Therefore, that a gene encodes a protein means that the mRNA of the gene produces the protein in cells or other biological systems through
10 transcription and translation.

As used herein, the term "host cell" refers to a cell used to receive, maintain, replicate and express a nucleic acid or vector. In some embodiments, the host cell may be a cell in which the polypeptide or polypeptide combination of the present invention is expressed.

The term "amino acid" is intended to encompass all molecules, whether natural or synthetic,
15 which include both an amino functional group and an acid functional group and are capable of being included in polymers of naturally occurring amino acids. Exemplary amino acids include naturally occurring amino acids; analogs, derivatives and congeners; amino acid analogs having variant side chains; and all stereoisomers of any of the foregoing. As used herein, the term "amino acid" includes D- or L- optical isomers and peptidomimetics.

20 The terms "polypeptide chain", "polypeptide", "peptide" and "protein" (if single-chain), used interchangeably herein, refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The term also encompasses an amino acid polymer that has been modified, where the modification may be, for example, disulfide bond formation, glycosylation, lipidation, acetylation,
25 phosphorylation or any other manipulation, such as conjugation with a labeling component. A polypeptide may be isolated from a natural source, may be produced by recombinant techniques from a eukaryotic or prokaryotic host, or may be the product of a synthetic method.

The terms "signal peptide" and "signal sequence" are used interchangeably herein, are generally located at the amino terminus of a protein, and are short peptide chains that guide newly

synthesized proteins to be transferred to the secretory pathway, with a length of about 5-30 amino acids.

As used herein, a "concatemer" refers to a polypeptide formed by linking 2, 3, 4, 5, 6, 7, 8, 9, 10 or more polypeptides, wherein a linker may or may not be included between each polypeptide and another polypeptide.

As used herein, an "antigen" refers to a molecule that, after entering the body, can elicit an acquired immune response in the body, wherein this immune response may involve production of antibodies, specific immunogenically active cells, or both. Those skilled in the art should understand that any macromolecule, including almost all proteins or peptide fragments, can serve as an antigen. Furthermore, an antigen may be derived from recombinant or genomic DNA or RNA. Those skilled in the art should understand that any DNA or RNA herein, the nucleotide sequence thereof or a partial nucleotide sequence thereof, can encode a protein capable of eliciting acquired immunity in the body. Furthermore, those skilled in the art should understand that an antigen need not individually encode only a full-length nucleotide sequence of one gene. Obviously, the present invention includes, but is not limited to, use of partial nucleotide sequences of more than one gene, and these nucleotide sequences form different mixtures to induce occurrence of a response. Furthermore, those skilled in the art should understand that an antigen need not be completely encoded by one gene. Obviously, an antigen may be synthetically produced or may be derived from a biological sample. Biological samples include, but are not limited to, tissue samples, tumor samples, cells or biological fluids.

As used herein, an "antibody" refers to a protein having a protective effect that is produced by the body due to stimulation by an antigen. It is an immunoglobulin produced by B lymphocytes. An antibody monomer is a Y-shaped molecule composed of four polypeptide chains. These include two identical heavy chains and two identical light chains, which are linked together by disulfide bonds. Each heavy chain is about 50 kDa, each light chain is about 25 kDa, and disulfide bonds are present between the light and heavy chains. Its uniqueness lies in its high affinity and specificity for a binding partner.

As used herein, a "vaccine" refers to a composition comprising an active ingredient (for example, the polypeptide or polypeptide combination or the nucleic acid or nucleic acid

combination of the present invention), which can elicit an immune response in an inoculated subject through inoculation. In specific embodiments, the immune response induced thereby can provide immune protection and is sufficient to prevent and/or alleviate at least one symptom associated with infection by a pathogen or disease. According to the present invention, the nucleic acid or nucleic acid combination or composition described herein can be used as a vaccine for providing prophylactic and/or therapeutic immunity against mpox virus in a subject in need thereof. As used herein, a "mpox virus RNA vaccine (for example, an mRNA vaccine, a circRNA vaccine or an saRNA vaccine)" provides a unique therapeutic alternative to peptide-based vaccines or DNA vaccines. The mpox virus RNA vaccine described herein (for example, an mRNA vaccine, a circRNA vaccine or an saRNA vaccine) comprises at least one ribonucleic acid (RNA) polynucleotide having an open reading frame encoding at least one mpox virus antigenic polypeptide or immunogenic fragment thereof (for example, an immunogenic fragment capable of inducing an immune response against mpox virus).

As used herein, the term "neutralizing antibody" refers to an antibody or fragment thereof capable of neutralizing, i.e., preventing, inhibiting, reducing or interfering with the ability of a pathogen to initiate and/or maintain infection in a host (for example, a host organism or host cell). According to the present invention, neutralizing antibody against mpox virus antigen proteins (or antigen polypeptide) can be produced in a subject inoculated with the vaccine of the present invention, for example, in immune serum of the subject. The level of neutralizing antibody titers in immune serum can be measured using methods known in the art.

Terms such as "treating" or "treatment" or "to treat" or "alleviating" or "to alleviate" refer to therapeutic measures that cure, slow, reduce symptoms of an existing diagnosed pathological condition or disorder, and/or arrest or slow progression of an existing diagnosed pathological condition or disorder. Terms such as "prevention", "protection", "avoidance", "containment" and the like refer to prophylactic or preventive measures that prevent progression of an undiagnosed target pathological condition or disorder. Thus, a "subject in need thereof" may include a subject that already has a disease; a subject susceptible to having a disease; and a subject in need of disease prevention.

The term "prophylactically or therapeutically effective amount" refers to an amount sufficient to prevent or inhibit occurrence of a disease or symptom and/or slow, alleviate or delay development or severity of a disease or symptom. A prophylactically or therapeutically effective amount is affected by factors including, but not limited to, the rate and severity of development of a disease or symptom, the age, sex, body weight and physiological condition of the subject, the duration of treatment and the specific route of administration. A prophylactically or therapeutically effective amount may be administered in one or more doses. A prophylactically or therapeutically effective amount may be achieved by continuous or intermittent administration.

As used herein, the term "subject" describes an organism, for example a mammal, to which treatment using the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination, composition or pharmaceutical composition of the present invention can be provided. Preferably, the subject is a rodent or a human.

As used herein, the term "first polypeptide (chain)" and similarly used "second polypeptide (chain)", "third polypeptide (chain)", "fourth polypeptide (chain)", "fifth polypeptide (chain)", "sixth polypeptide (chain)" and the like are used only for distinguishing purposes, and are not intended to limit the order and linkage manner of these polypeptides (chains). The respective polypeptides (chains) themselves and among one another may be the same or different. If needed, two or more first polypeptides (chains) may also be comprised, which may be the same or different. The above terms used in different schemes may be the same or different.

As used herein, the term "first polynucleotide" and similarly used "second polynucleotide", "third polynucleotide", "fourth polynucleotide", "fifth polynucleotide", "sixth polynucleotide" and the like are used only for distinguishing purposes, and are not intended to limit the order and linkage manner of these polynucleotides. The respective polynucleotides themselves and among one another may be the same or different. If needed, two or more first polynucleotides may also be comprised, which may be the same or different. The above terms used in different schemes may be the same or different.

As used herein and unless otherwise specified, the term "lipid" refers to a group of organic compounds that include, but are not limited to, fatty acid esters and are generally characterized by being poorly soluble in water but soluble in many nonpolar organic solvents. Although lipids

generally have poor water solubility, certain classes of lipids (for example, lipids modified with polar groups, such as DMG-PEG2000) have limited water solubility and are soluble in water under certain conditions. Known types of lipids include biomolecules such as fatty acids, waxes, sterols, fat-soluble vitamins, monoglycerides, diglycerides, triglycerides and phospholipids. Lipids can be classified into at least three categories: (1) "simple lipids", including fats and oils, and waxes; (2) "complex lipids", including phospholipids and glycolipids (for example DMPE-PEG2000); and (3) "derived lipids", such as steroids. In addition, as used herein, lipids also include lipidoid compounds. The term "lipidoid compound", also abbreviated as "lipidoid", refers to a lipid-like compound (for example, an amphiphilic compound having lipid-like physical properties).

The term "lipid nanoparticle" or "LNP" refers to a particle having at least one nanometer (nm)-scale dimension (for example, 1 to 1,000 nm), which contains one or more types of lipid molecules. The LNP provided herein may further contain at least one non-lipid payload molecule (for example, one or more nucleic acid molecules). In some embodiments, the LNP comprises a non-lipid payload molecule partially or completely encapsulated within a lipid shell. Specifically, in some embodiments, wherein the payload is a negatively charged molecule (for example mRNA, circRNA or saRNA), and the lipid component of the LNP comprises at least one cationic lipid. Without being bound by theory, it is expected that the cationic lipid may interact with the negatively charged payload molecule and facilitate incorporation and/or encapsulation of the payload into the LNP during LNP formation. Other lipids that may form part of the LNP provided herein include, but are not limited to, neutral lipids and charged lipids, such as steroids, polymer-conjugated lipids and various zwitterionic lipids. In certain embodiments, the LNP according to the present disclosure comprises one or more of the lipids of series 01, 02, 03 and 04, for example one or more of the lipids according to formulae (01-I), (01-II), (02-I), (02-II), (03-I), (04-I) and (04-III) (and sub-formulae thereof) described herein.

The term "cationic lipid" refers to a lipid that is positively charged at any pH value or hydrogen ion activity of its environment, or that is capable of being positively charged in response to the pH value or hydrogen ion activity of its environment (for example, its intended environment of use). Thus, the term "cationic" encompasses "permanently cationic" and "cationizable". In certain embodiments, the positive charge in the cationic lipid is caused by the presence of a quaternary

nitrogen atom. In certain embodiments, the cationic lipid comprises a zwitterionic lipid, wherein the zwitterionic lipid is positively charged in its intended environment of use (for example, at physiological pH). In certain embodiments, the cationic lipid comprises one or more of the lipids of series 01, 02, 03 and 04, for example one or more of the lipids according to formulae (01-I), (01-II), (02-I), (02-II), (03-I), (04-I) and (04-III) (and sub-formulae thereof) described herein. The term "anionic lipid" refers to a lipid that is negatively charged at any pH value or hydrogen ion activity of its environment, or that is capable of being negatively charged in response to the pH value or hydrogen ion activity of its environment (for example, its intended environment of use). Exemplary anionic lipids include one or more phosphate groups that are negatively charged, for example, at physiological pH.

The term "polymer-conjugated lipid" refers to a molecule that comprises both a lipid moiety and a polymer moiety. An example of a polymer-conjugated lipid is a PEGylated lipid (PEG-lipid), wherein the polymer moiety comprises polyethylene glycol.

The term "neutral lipid" encompasses any lipid molecule that exists in an uncharged form or in a neutral zwitterionic form at a selected pH value or within a selected pH range. In some embodiments, the selected useful pH value or range corresponds to the pH conditions in the environment of the intended lipid use, such as physiological pH. As non-limiting examples, neutral lipids that may be used in combination with the present disclosure include, but are not limited to, phospholipid phosphatidylcholines, such as 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC); phosphatidylethanolamines, such as 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 2-((2,3-bis(oleoyloxy)propyl)dimethylammonio)ethyl hydrogen phosphate (DOCP); sphingomyelin (SM); ceramide. The neutral lipids provided herein may be synthetic or derived from natural sources or compounds (isolated or modified therefrom).

The term "steroid" refers to a lipid having cyclopentanoperhydrophenanthrene as the parent nucleus, including sterols and sterol derivatives. Steroid compounds do not contain bound fatty acids and are nonsaponifiable lipids. Herein, a steroid may serve as a structural lipid in the lipid

component of a nanoparticle composition. As non-limiting examples, steroids that may be used in combination with the present disclosure include, but are not limited to, cholesterol, coprostanol, sitosterol, ergosterol, campesterol, stigmasterol and mixtures thereof.

5 The term "charged lipid" encompasses any lipid molecule that exists in a positively charged or negatively charged form at a selected pH value or within a selected pH range. In some embodiments, the selected pH value or range corresponds to pH conditions in the environment of the intended lipid use, such as physiological pH. As non-limiting examples, charged lipids that may be used in combination with the present disclosure include, but are not limited to, phospholipid phosphatidylserine, phosphatidic acid, phosphatidylglycerol, phosphatidylinositol, 10 sterol hemisuccinate, dialkyl trimethylammonium-propane (for example DOTAP, DOTMA), dialkyl dimethylaminopropane, ethyl phosphocholine, dimethylaminoethane carbamoyl sterol (for example DC-Chol), sodium 1,2-dioleoyl-sn-glycero-3-phospho-L-serine (DOPS-Na), sodium 1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DOPG-Na) and sodium 1,2-dioleoyl-sn-glycero-3-phosphate (DOPA-Na). The charged lipids provided herein may be synthetic or derived 15 from natural sources or compounds (isolated or modified therefrom).

As used herein and unless otherwise specified, the term "alkyl" refers to a saturated straight-chain or branched hydrocarbon chain group consisting only of carbon and hydrogen atoms. In an embodiment, alkyl has, for example, 1 to 24 carbon atoms (C₁-C₂₄ alkyl), 4 to 20 carbon atoms (C₄-C₂₀ alkyl), 6 to 16 carbon atoms (C₆-C₁₆ alkyl), 6 to 9 carbon atoms (C₆-C₉ alkyl), 1 to 15 20 carbon atoms (C₁-C₁₅ alkyl), 1 to 12 carbon atoms (C₁-C₁₂ alkyl), 1 to 8 carbon atoms (C₁-C₈ alkyl) or 1 to 6 carbon atoms (C₁-C₆ alkyl), and is attached to the rest of the molecule through a single bond. Examples of alkyl include, but are not limited to, methyl, ethyl, n-propyl, 1-methylethyl (isopropyl), n-butyl, n-pentyl, 1,1-dimethylethyl (tert-butyl), 3-methylhexyl, 2-methylhexyl and the like. Unless otherwise specified, alkyl is optionally substituted.

25 As used herein and unless otherwise specified, the term "alkenyl" refers to a straight-chain or branched hydrocarbon chain group consisting only of carbon and hydrogen atoms, which contains one or more carbon-carbon double bonds. As understood by a person of ordinary skill in the art, the term "alkenyl" also encompasses groups having "cis" and "trans" configurations or, alternatively, "E" and "Z" configurations. In an embodiment, alkenyl has, for example, 2 to 24

carbon atoms (C₂-C₂₄ alkenyl), 4 to 20 carbon atoms (C₄-C₂₀ alkenyl), 6 to 16 carbon atoms (C₆-C₁₆ alkenyl), 6 to 9 carbon atoms (C₆-C₉ alkenyl), 2 to 15 carbon atoms (C₂-C₁₅ alkenyl), 2 to 12 carbon atoms (C₂-C₁₂ alkenyl), 2 to 8 carbon atoms (C₂-C₈ alkenyl) or 2 to 6 carbon atoms (C₂-C₆ alkenyl), and is attached to the rest of the molecule through a single bond. Examples of alkenyl include, but are not limited to, vinyl, prop-1-enyl, but-1-enyl, pent-1-enyl, pent-1,4-dienyl and the like. Unless otherwise specified, alkenyl is optionally substituted.

As used herein and unless otherwise specified, the term "alkynyl" refers to a straight-chain or branched hydrocarbon chain group consisting only of carbon and hydrogen atoms, which contains one or more carbon-carbon triple bonds. In an embodiment, alkynyl has, for example, 2 to 24 carbon atoms (C₂-C₂₄ alkynyl), 4 to 20 carbon atoms (C₄-C₂₀ alkynyl), 6 to 16 carbon atoms (C₆-C₁₆ alkynyl), 6 to 9 carbon atoms (C₆-C₉ alkynyl), 2 to 15 carbon atoms (C₂-C₁₅ alkynyl), 2 to 12 carbon atoms (C₂-C₁₂ alkynyl), 2 to 8 carbon atoms (C₂-C₈ alkynyl) or 2 to 6 carbon atoms (C₂-C₆ alkynyl), and is attached to the rest of the molecule through a single bond. Examples of alkynyl include, but are not limited to, ethynyl, propynyl, butynyl, pentynyl and the like. Unless otherwise specified, alkynyl is optionally substituted.

As used herein and unless otherwise specified, the term "alkylene" or "alkylene chain" refers to a straight-chain or branched divalent hydrocarbon chain that connects the rest of the molecule to a group, which consists only of carbon and hydrogen and is saturated. In an embodiment, alkylene has, for example, 1 to 24 carbon atoms (C₁-C₂₄ alkylene), 1 to 15 carbon atoms (C₁-C₁₅ alkylene), 1 to 12 carbon atoms (C₁-C₁₂ alkylene), 1 to 8 carbon atoms (C₁-C₈ alkylene), 1 to 6 carbon atoms (C₁-C₆ alkylene), 2 to 4 carbon atoms (C₂-C₄ alkylene), 1 to 2 carbon atoms (C₁-C₂ alkylene). Examples of alkylene include, but are not limited to, methylene, ethylene, propylene, n-butylene and the like. The alkylene chain is attached to the rest of the molecule through a single bond and is attached to the group through a single bond. The points of attachment of the alkylene chain to the rest of the molecule and to the group may be through one carbon or any two carbons within the chain. Unless otherwise specified, the alkylene chain is optionally substituted.

As used herein and unless otherwise specified, the term "alkenylene" refers to a straight-chain or branched divalent hydrocarbon chain that connects the rest of the molecule to a group, which consists only of carbon and hydrogen and contains one or more carbon-carbon double bonds. In

an embodiment, alkenylene has, for example, 2 to 24 carbon atoms (C₂-C₂₄ alkenylene), 2 to 15 carbon atoms (C₂-C₁₅ alkenylene), 2 to 12 carbon atoms (C₂-C₁₂ alkenylene), 2 to 8 carbon atoms (C₂-C₈ alkenylene), 2 to 6 carbon atoms (C₂-C₆ alkenylene) or 2 to 4 carbon atoms (C₂-C₄ alkenylene). Examples of alkenylene include, but are not limited to, vinylene, propenylene, n-butenylene and the like. Alkenylene is attached to the rest of the molecule through a single bond or double bond and is attached to the group through a single bond or double bond. The points of attachment of alkenylene to the rest of the molecule and to the group may be through one carbon or any two carbons within the chain. Unless otherwise specified, alkenylene is optionally substituted.

As used herein and unless otherwise specified, the term "cycloalkyl" refers to a non-aromatic saturated monocyclic or polycyclic hydrocarbon group consisting only of carbon and hydrogen atoms. Cycloalkyl may include fused or bridged ring systems. In an embodiment, cycloalkyl has, for example, 3 to 15 ring carbon atoms (C₃-C₁₅ cycloalkyl), 3 to 10 ring carbon atoms (C₃-C₁₀ cycloalkyl) or 3 to 8 ring carbon atoms (C₃-C₈ cycloalkyl). Cycloalkyl is attached to the rest of the molecule through a single bond. Examples of monocyclic cycloalkyl include, but are not limited to, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl and cyclooctyl. Examples of polycyclic cycloalkyl include, but are not limited to, adamantyl, norbornyl, decalanyl, 7,7-dimethyl-bicyclo[2.2.1]heptyl and the like. Unless otherwise specified, cycloalkyl is optionally substituted.

As used herein and unless otherwise specified, the term "cycloalkylene" is divalent cycloalkyl. Unless otherwise specified, cycloalkylene is optionally substituted.

As used herein and unless otherwise specified, the term "cycloalkenyl" refers to a non-aromatic monocyclic or polycyclic hydrocarbon group consisting only of carbon and hydrogen atoms and including one or more carbon-carbon double bonds. Cycloalkenyl may include fused or bridged ring systems. In an embodiment, cycloalkenyl has, for example, 3 to 15 ring carbon atoms (C₃-C₁₅ cycloalkenyl), 3 to 10 ring carbon atoms (C₃-C₁₀ cycloalkenyl) or 3 to 8 ring carbon atoms (C₃-C₈ cycloalkenyl). Cycloalkenyl is attached to the rest of the molecule through a single bond. Examples of monocyclic cycloalkenyl include, but are not limited to, cyclopropenyl, cyclobutenyl,

cyclopentenyl, cyclohexenyl, cycloheptenyl, cyclooctenyl and the like. Unless otherwise specified, cycloalkenyl is optionally substituted.

As used herein and unless otherwise specified, the term "cycloalkenylene" is divalent cycloalkenyl. Unless otherwise specified, cycloalkenylene is optionally substituted.

5 As used herein and unless otherwise specified, the term "heterocyclyl" refers to a non-aromatic group monocyclic or polycyclic moiety containing one or more (for example one, one or two, one to three, or one to four) heteroatoms independently selected from nitrogen, oxygen, phosphorus and sulfur. Heterocyclyl may be attached to the main structure at any heteroatom or carbon atom. Heterocyclyl may be a monocyclic, bicyclic, tricyclic, tetracyclic or other polycyclic system,
10 wherein the polycyclic system may be a fused, bridged or spiro system. A heterocyclyl polycyclic system may contain one or more heteroatoms in one or more rings. Heterocyclyl may be saturated or partially unsaturated. A saturated heterocycloalkyl may be referred to as "heterocycloalkyl." A partially unsaturated heterocycloalkyl may be referred to as "heterocycloalkenyl" when the heterocyclyl contains at least one double bond, or may be referred to as "heterocycloalkynyl" when
15 the heterocyclyl contains at least one triple bond. In an embodiment, heterocyclyl has, for example, 3 to 18 ring atoms (3-membered to 18-membered heterocyclyl), 4 to 18 ring atoms (4-membered to 18-membered heterocyclyl), 5 to 18 ring atoms (3-membered to 18-membered heterocyclyl), 4 to 8 ring atoms (4-membered to 8-membered heterocyclyl) or 5 to 8 ring atoms (5-membered to 8-membered heterocyclyl). When appearing herein, a numerical range, such as "3 to 18," refers to
20 each integer in the given range; for example, "3 to 18 ring atoms" means that the heterocyclyl may consist of 3 ring atoms, 4 ring atoms, 5 ring atoms, 6 ring atoms, 7 ring atoms, 8 ring atoms, 9 ring atoms, 10 ring atoms, etc. (up to and including 18 ring atoms). Examples of heterocyclyl include, but are not limited to, imidazolyl, imidazolidinyl, oxazolyl, oxazolidinyl, thiazolyl, thiazolidinyl, pyrazolidinyl, pyrazolyl, isoxazolidinyl, isoxazolyl, isothiazolidinyl, isothiazolyl, morpholinyl, pyrrolyl, pyrrolidinyl, furanyl, tetrahydrofuranyl, thienyl, pyridyl, piperidinyl, quinolinyl and
25 isoquinolinyl. Unless otherwise specified, heterocyclyl is optionally substituted.

As used herein and unless otherwise specified, the term "heterocyclylene" is divalent heterocyclyl. Unless otherwise specified, heterocyclylene is optionally substituted.

As used herein and unless otherwise specified, the term "aryl" refers to a monocyclic aromatic group and/or polycyclic monovalent aromatic group containing at least one aromatic hydrocarbon ring. In certain embodiments, aryl has 6 to 18 ring carbon atoms (C₆-C₁₈ aryl), 6 to 14 ring carbon atoms (C₆-C₁₄ aryl) or 6 to 10 ring carbon atoms (C₆-C₁₀ aryl). Examples of aryl include, but are not limited to, phenyl, naphthyl, fluorenyl, azulenyl, anthracenyl, phenanthrenyl, pyrenyl, biphenyl and terphenyl. The term "aryl" also refers to bicyclic, tricyclic or other polycyclic hydrocarbon rings, wherein at least one ring is an aromatic ring, and the other rings may be saturated, partially unsaturated or aromatic rings, for example dihydronaphthyl, indenyl, indanyl or tetrahydronaphthyl/tetralinyl. Unless otherwise specified, aryl is optionally substituted.

As used herein and unless otherwise specified, the term "arylene" is divalent aryl. Unless otherwise specified, arylene is optionally substituted.

As used herein and unless otherwise specified, the term "heteroaryl" refers to a monocyclic aromatic group and/or polycyclic aromatic group containing at least one aromatic ring, wherein at least one aromatic ring contains one or more (for example one, one or two, one to three, or one to four) heteroatoms independently selected from O, S and N. Heteroaryl may be attached to the main structure at any heteroatom or carbon atom. In certain embodiments, heteroaryl has 5 to 20, 5 to 15 or 5 to 10 ring atoms. The term "heteroaryl" also refers to bicyclic, tricyclic or other polycyclic rings, wherein at least one ring is an aromatic ring, and the other rings may be saturated, partially unsaturated or aromatic rings, wherein at least one aromatic ring contains one or more heteroatoms independently selected from O, S and N. Examples of monocyclic heteroaryl include, but are not limited to, pyrrolyl, pyrazolyl, pyrazolinyl, imidazolyl, oxazolyl, isoxazolyl, thiazolyl, thiadiazolyl, isothiazolyl, furanyl, thienyl, oxadiazolyl, pyridyl, pyrazinyl, pyrimidinyl, pyridazinyl and triazinyl. Examples of bicyclic heteroaryl include, but are not limited to, indolyl, benzothiazolyl, benzoxazolyl, benzothienyl, quinolinyl, tetrahydroisoquinolinyl, isoquinolinyl, benzimidazolyl, benzopyranyl, indolizinyl, benzofuranyl, isobenzofuranyl, chromonyl, coumarinyl, cinnolinyl, quinoxalinyl, indazolyl, purinyl, pyrrolopyridyl, furopyridyl, thienopyridyl, dihydroisoindolyl and tetrahydroquinolinyl. Examples of tricyclic heteroaryl include, but are not limited to, carbazolyl, benzoindolyl, phenanthrolinyl, acridinyl, phenanthridinyl and xanthenyl. Unless otherwise specified, heteroaryl is optionally substituted.

As used herein and unless otherwise specified, the term "heteroaryl" is divalent heteroaryl. Unless otherwise specified, heteroaryl is optionally substituted.

When a group described herein is referred to as being "substituted," it may be substituted with one or more of any suitable substituents. Illustrative examples of substituents include, but are not limited to, those visible in the exemplary compounds and embodiments provided herein, and: halogen atoms, such as F, Cl, Br or I; cyano; oxo (=O); hydroxy (-OH); alkyl; alkenyl; alkynyl; cycloalkyl; aryl; $-(C=O)OR'$; $-O(C=O)R'$; $-C(=O)R'$; $-OR'$; $-S(O)_xR'$; $-S-SR'$; $-C(=O)SR'$; $-SC(=O)R'$; $-NR'R'$; $-NR'C(=O)R'$; $-C(=O)NR'R'$; $-NR'C(=O)NR'R'$; $-OC(=O)NR'R'$; $-NR'C(=O)OR'$; $-NR'S(O)_xNR'R'$; $-NR'S(O)_xR'$; and $-S(O)_xNR'R'$, wherein: R' at each occurrence is independently H, C_1 - C_{16} alkyl or cycloalkyl, and x is 0, 1 or 2. In some embodiments, the substituent is C_1 - C_{12} alkyl. In other embodiments, the substituent is cycloalkyl. In other embodiments, the substituent is halo, such as fluoro. In other embodiments, the substituent is oxo. In other embodiments, the substituent is hydroxy. In other embodiments, the substituent is alkoxy ($-OR'$). In other embodiments, the substituent is carboxyl. In other embodiments, the substituent is amino ($-NR'R'$).

As used herein and unless otherwise specified, the term "optionally present" or "optionally" (for example optionally substituted) means that the subsequently described event or circumstance may or may not occur, and that the description includes instances in which the event or circumstance occurs and instances in which it does not occur. For example, "optionally substituted alkyl" means that alkyl may be substituted or may be unsubstituted, and that the description includes both substituted alkyl and alkyl having no substitution.

As used herein and unless otherwise specified, the term "prodrug" of a biologically active compound refers to a compound that can be converted into the biologically active compound under physiological conditions or by solvolysis. In an embodiment, the term "prodrug" refers to a pharmaceutically acceptable metabolic precursor of a biologically active compound. A prodrug may be inactive when administered to a subject in need thereof, but is converted in vivo into a biologically active compound. Prodrugs are generally rapidly transformed in vivo to yield the parent biologically active compound, for example by hydrolysis in blood. Prodrug compounds generally provide advantages of solubility, tissue compatibility or delayed release in a mammalian

organism (see Bundgard, H., Design of Prodrugs (1985), pp. 7-9, pp. 21-24 (Elsevier, Amsterdam)). A discussion of prodrugs is provided in Higuchi, T. et al., A.C.S. Symposium Series, Vol. 14; and Bioreversible Carriers in Drug Design, Edward B. Roche ed., American Pharmaceutical Association and Pergamon Press, 1987.

5 In an embodiment, the term "prodrug" is also intended to include any covalently bonded carriers that release the active compound in vivo when such prodrug is administered to a mammalian subject. Prodrugs of compounds may be prepared by modifying functional groups present in the compounds in such a manner that the modifications are cleaved, in routine manipulation or in vivo, to yield the parent compounds. Prodrugs include compounds in which
10 hydroxy, amino or mercapto groups are bonded to any group that, when the prodrug of the compound is administered to a mammalian subject, cleaves to form a free hydroxy group, a free amino group or a free mercapto group, respectively.

Examples of prodrugs include, but are not limited to, acetate, formate and benzoate derivatives of alcohol functional groups or amide derivatives of amine functional groups in the compounds
15 provided herein.

As used herein and unless otherwise specified, the term "pharmaceutically acceptable salt" includes both acid addition salts and base addition salts.

Examples of pharmaceutically acceptable acid addition salts include, but are not limited to, hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid and the like; and
20 organic acids, such as but not limited to acetic acid, 2,2-dichloroacetic acid, adipic acid, alginic acid, ascorbic acid, aspartic acid, benzenesulfonic acid, benzoic acid, 4-acetamidobenzoic acid, camphoric acid, camphor-10-sulfonic acid, capric acid, caproic acid, caprylic acid, carbonic acid, cinnamic acid, citric acid, cyclamic acid, dodecylsulfuric acid, ethane-1,2-disulfonic acid, ethanesulfonic acid, 2-hydroxyethanesulfonic acid, formic acid, fumaric acid, galactaric acid,
25 gentisic acid, glucoheptonic acid, gluconic acid, glucuronic acid, glutamic acid, glutaric acid, 2-oxoglutaric acid, glycerophosphoric acid, glycolic acid, hippuric acid, isobutyric acid, lactic acid, lactobionic acid, lauric acid, maleic acid, malic acid, malonic acid, mandelic acid, methanesulfonic acid, mucic acid, naphthalene-1,5-disulfonic acid, naphthalene-2-sulfonic acid, 1-hydroxy-2-naphthoic acid, nicotinic acid, oleic acid, orotic acid, oxalic acid, palmitic acid, pamoic acid,

propionic acid, pyroglutamic acid, pyruvic acid, salicylic acid, 4-aminosalicylic acid, sebacic acid, stearic acid, succinic acid, tartaric acid, thiocyanic acid, p-toluenesulfonic acid, trifluoroacetic acid, undecylenic acid and the like.

Examples of pharmaceutically acceptable base addition salts include, but are not limited to, salts prepared by adding inorganic bases or organic bases to free acid compounds. Salts derived from inorganic bases include, but are not limited to, sodium salts, potassium salts, lithium salts, ammonium salts, calcium salts, magnesium salts, iron salts, zinc salts, copper salts, manganese salts, aluminum salts and the like. In an embodiment, the inorganic salt is an ammonium salt, sodium salt, potassium salt, calcium salt and magnesium salt. Salts derived from organic bases include, but are not limited to, salts of: primary, secondary and tertiary amines; substituted amines, including naturally occurring substituted amines; cyclic amines and basic ion-exchange resins, such as ammonia, isopropylamine, trimethylamine, diethylamine, triethylamine, tripropylamine, diethanolamine, ethanolamine, deanol, 2-dimethylaminoethanol, 2-diethylaminoethanol, dicyclohexylamine, lysine, arginine, histidine, caffeine, procaine, hydrabamine, choline, betaine, benethamine, benzathine, ethylenediamine, glucosamine, methylglucosamine, theobromine, triethanolamine, tromethamine, purines, piperazine, piperidine, N-ethylpiperidine, polyamine resins and the like. In an embodiment, the organic base is isopropylamine, diethylamine, ethanolamine, trimethylamine, dicyclohexylamine, choline and caffeine.

The compounds provided herein may contain one or more asymmetric centers and therefore may give rise to enantiomers, diastereomers and other stereoisomeric forms, which may be defined in terms of absolute stereochemistry as (R)- or (S)- or, for amino acids, as (D)- or (L)-. Unless otherwise specified, the compounds provided herein are intended to include all such possible isomers, as well as their racemic and optically pure forms. Optically active (+) and (-), (R)- and (S)-, or (D)- and (L)-isomers may be prepared using chiral synthons or chiral reagents, or resolved using conventional techniques, for example chromatography and fractional crystallization. Conventional techniques for preparing/isolating individual enantiomers include chiral synthesis from a suitable optically pure precursor or resolution of the racemate (or racemate of a salt or derivative) using, for example, chiral high-pressure liquid chromatography (HPLC). When the compounds described herein contain olefinic double bonds or other centers of geometric

asymmetry, unless otherwise specified, the compounds are intended to include E and Z geometric isomers. Likewise, all tautomeric forms are also intended to be included.

As used herein and unless otherwise specified, the term "isomer" refers to different compounds having the same molecular formula. A "stereoisomer" is an isomer that differs only in the arrangement of atoms in space. An "atropisomer" is a stereoisomer resulting from hindered rotation around a single bond. An "enantiomer" is one of a pair of stereoisomers that are non-superimposable mirror images of each other. A mixture of any ratio of a pair of enantiomers may be referred to as a "racemic" mixture. A "diastereomer" is a stereoisomer having at least two asymmetric atoms but not being mirror images of each other.

A "stereoisomer" may also include E and Z isomers or mixtures thereof, as well as cis and trans isomers or mixtures thereof. In certain embodiments, the compounds described herein are isolated as E or Z isomers. In other embodiments, the compounds described herein are mixtures of E and Z isomers.

A "tautomer" refers to isomeric forms of a compound that are in equilibrium with each other. The concentration of the isomeric forms will depend on the environment in which the compound is located, and may vary according to, for example, whether the compound is a solid or is in an organic solution or aqueous solution.

It should also be noted that the compounds described herein may contain unnatural proportions of atomic isotopes at one or more atoms. For example, a compound may be radiolabeled with radioactive isotopes, such as tritium (^3H), iodine-125 (^{125}I), sulfur-35 (^{35}S) or carbon-14 (^{14}C), or may be isotope-enriched, such as deuterium (^2H), carbon-13 (^{13}C) or nitrogen-15 (^{15}N). As used herein, an "isotopologue" is an isotope-enriched compound. The term "isotope-enriched" refers to the isotopic composition of an atom being different from the natural isotopic composition of that atom. "Isotope-enriched" may also refer to a compound containing at least one atom whose isotopic composition is different from the natural isotopic composition of that atom. The term "isotopic composition" refers to the amount of each isotope present for a given atom. Radiolabeled and isotope-enriched compounds may be useful as therapeutic agents, for example cancer therapeutic agents; research reagents, for example binding assay reagents; and diagnostic agents, for example in vivo imaging agents. All isotopic variants of the compounds described herein,

whether radioactive or not, are intended to be encompassed within the scope of the embodiments provided herein. In some embodiments, isotopologues of the compounds described herein are provided, for example the isotopologues are deuterium-, carbon-13- and/or nitrogen-15-enriched. As used herein, "deuterated" means that at least one hydrogen (H) in a compound has been replaced
5 by deuterium (represented by D or ^2H), i.e., the compound is enriched in deuterium at least one position.

It should be noted that, if there is a discrepancy between a depicted structure and the name of that structure, the depicted structure shall control.

As used herein and unless otherwise specified, the term "pharmaceutically acceptable carrier,
10 diluent or excipient" includes, but is not limited to, any adjuvant, carrier, excipient, glidant, sweetening agent, diluent, preservative, dye/colorant, flavor enhancer, surfactant, wetting agent, dispersant, suspending agent, stabilizer, isotonic agent, solvent or emulsifier approved by the United States Food and Drug Administration as acceptable for use in humans or domestic animals.

The term "composition" is intended to encompass a product containing specified ingredients
15 (for example the mRNA molecule provided herein) in specified amounts that are optionally selected.

The names of nucleic acid constructs used herein (such as 0201, 0302, 0402, 0505, 0603, 0702, 0803, 0201-J0, 0301-J0, 0501-J0, 0701-J0, 0901-J0, 1101, 1201, 1301, 1401, 1501, 1601, 1701, 1801, 1901, 2001 and the like) are used merely to distinguish or identify nucleic acids, constructs,
20 sequences or products, and are not intended to indicate that such identifiers are features of the nucleic acid or nucleic acid combination, constructs, sequences or products of the present invention. A person skilled in the art should understand that, for example for purposes of distinction or identification, other nucleic acid or nucleic acid combination, constructs, sequences or products may likewise possibly use such identifiers, but such identifiers do not refer to the same
25 or equivalent nucleic acid or nucleic acid combination, constructs, sequences or products. Similarly, similar numbers or identifiers used in the examples are also merely for convenience of illustration, and the nucleic acid or nucleic acid combination, constructs, sequences or products of the present invention are defined by the features described in the appended claims.

Mpox virus

As used herein, the term "Mpox virus (Mpox Virus, MPXV)" belongs to the genus Orthopoxvirus of the family Poxviridae, together with variola virus and vaccinia virus, and MPXV has morphology similar to other known orthopoxviruses. MPXV is oval or brick-shaped and is enveloped by a lipoprotein outer membrane, and its genome is linear double-stranded DNA with a genome size of about 197 kb. MPXV can enter the host via oropharyngeal, nasopharyngeal or intradermal routes. The virus replicates at the site of infection and then spreads to regional lymph nodes. Although MPXV is a DNA virus, its life cycle occurs in the cytoplasm. The MPXV replication process produces two types of infectious viral particles: mature virions (MVs) and enveloped virions (EVs); wherein MVs enter cells by macropinocytosis or membrane fusion, whereas EVs enter by membrane fusion. After replication of viral DNA and assembly of viral proteins are completed in the cytoplasm, MVs are released by cell lysis, whereas EVs are released by exocytosis (see, for example, Kumar N et al., The 2022 outbreak and the pathobiology of the mpox virus. J Autoimmun. 2022 Jul;131:102855.).

As used herein, the terms "A29L protein" and "A29L peptide," "A35R protein" and "A35R peptide," "B2R protein" and "B2R peptide," "B6R protein" and "B6R peptide," "C15L protein" and "C15L peptide," "M1R protein" and "M1R peptide," "H3L protein" and "H3L peptide," and "E8L protein" and "E8L peptide" may each be used interchangeably, and they are all major antigen proteins (or antigen polypeptides) of mpox virus. The "A29L protein," "A35R protein," "B2R protein," "B6R protein," "M1R protein," "E8L protein," "H3L protein" and "C15L protein" of mpox virus are homologous to the "A27L protein," "A33R protein," "A56R protein," "B5R protein," "L1R protein," "D8L protein," "H3L protein" and "F9L protein" of vaccinia virus, respectively.

It has been reported that the "A29L protein" is a surface membrane fusion protein located in mature virions and can bind to cell-surface heparan sulfate; the "A35R protein" is an envelope protein located in enveloped virions and participates in formation of actin-containing microvilli and cell-to-cell spread of virions; the "B2R protein" is a membrane glycoprotein hemagglutinin located in enveloped virions and can prevent cell fusion; the "B6R protein" is a palmitoylated glycoprotein located in enveloped virions, is required for efficient cell spread, and participates in complement regulation; the "M1R protein" is a myristoylated surface membrane protein located

in mature virions and plays an important role in viral entry into a host; the "E8L protein" is a surface membrane adsorption protein located in mature virions and can bind to chondroitin sulfate; the "H3L protein" is a surface membrane protein located in mature virions and can bind to heparan sulfate and cell-surface proteoglycans; and the "C15L protein" is located in mature virions (see, 5 for example, Papukashvili D et al., Strategy of developing nucleic acid-based universal monkeypox vaccine candidates. Front Immunol. 2022 Oct 27;13:1050309.).

As used herein, the A29L protein is as set forth in NCBI accession number YP_010377135.1; the A35R protein is as set forth in NCBI accession number YP_010377142.1; the B2R protein is as set forth in NCBI accession number YP_010377158.1; the B6R protein is as set forth in NCBI 10 accession number NP_536594.1; the C15L protein is as set forth in NCBI accession number URF91574.1; the M1R protein is as set forth in NCBI accession number URF91574.1 or NP_536507.1; the H3L protein is as set forth in NCBI accession number UVB87333.1; and the E8L protein is as set forth in NCBI accession number YP_010377102.1.

Polypeptide

15 In one aspect, the present invention provides a polypeptide or polypeptide combination comprising one or more peptide fragments, wherein each peptide fragment is independently selected from peptide fragments of mpox virus A29L, A35R, B2R, B6R, C15L, M1R, and H3L proteins.

In some embodiments, the polypeptide or polypeptide combination comprises 1, 2, 3, 4, 5, 6, 20 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22 peptide fragments, wherein the peptide fragments are each independently selected from peptide fragments of mpox virus A29L, A35R, B2R, B6R, C15L, M1R, and H3L proteins.

In some embodiments, the polypeptide or polypeptide combination comprises one peptide fragment, wherein the peptide fragment is selected from peptide fragments of mpox virus A29L, 25 A35R, B2R, B6R, C15L, M1R, and H3L proteins; or the polypeptide or polypeptide combination comprises 2, 3, 4, 5, 6 or 7 peptide fragments each derived from a different protein, wherein the peptide fragments are selected from peptide fragments of mpox virus A29L, A35R, B2R, B6R, C15L, M1R, and H3L proteins.

In an embodiment, the polypeptide or polypeptide combination comprises three peptide fragments each derived from a different protein, wherein the peptide fragments are selected from peptide fragments of mpox virus A29L, A35R, B2R, B6R, C15L, M1R, and H3L proteins.

5 In a preferred embodiment, the polypeptide or polypeptide combination comprises six peptide fragments each derived from a different protein, wherein the peptide fragments are selected from peptide fragments of mpox virus A29L, A35R, B2R, B6R, C15L, M1R, and H3L proteins.

In some embodiments, the polypeptide or polypeptide combination further comprises one or more linkers, an IgE signal peptide and/or an HA polypeptide tag.

10 In an embodiment, the polypeptide or polypeptide combination comprises a peptide fragment of the A29L protein. In an embodiment, the polypeptide or polypeptide combination comprises a peptide fragment of the A35R protein. In an embodiment, the polypeptide or polypeptide combination comprises a peptide fragment of the B2R protein. In an embodiment, the polypeptide or polypeptide combination comprises a peptide fragment of the B6R protein. In an embodiment, the polypeptide or polypeptide combination comprises a peptide fragment of the C15L protein. In
15 an embodiment, the polypeptide or polypeptide combination comprises a peptide fragment of the M1R protein. In an embodiment, the polypeptide or polypeptide combination comprises a peptide fragment of the H3L protein. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, and C15L proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A29L, B6R, and M1R
20 proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of B6R, M1R, and H3L proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of B6R, M1R, and A35R proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A29L, B2R, and C15L proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide
25 fragments of B2R, H3L, and C15L proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of B2R, A35R, and B6R proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A29L, M1R, and C15L proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of H3L, M1R, and C15L proteins. In an embodiment, the polypeptide or polypeptide

combination comprises peptide fragments of A29L, A35R, B2R, B6R, C15L, and M1R proteins. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, B6R, C15L, M1R, and H3L proteins.

5 In an embodiment, the peptide fragment of the A29L protein comprises full-length A29L peptide. In an embodiment, the peptide fragment of the A29L protein comprises full-length A29L peptide, wherein the full-length A29L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 111 or 113.

10 In an embodiment, the peptide fragment of the A35R protein comprises full-length A35R peptide. In an embodiment, the full-length A35R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 114. In an embodiment, the peptide fragment of the A35R protein comprises an A35R extracellular segment. In an embodiment, the A35R extracellular segment comprises the amino acid sequence of SEQ ID NO: 115.

15 In an embodiment, the peptide fragment of the B2R protein comprises a B2R signal peptide. In an embodiment, the B2R signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 116. In an embodiment, the peptide fragment of the B2R protein comprises the full-length B2R peptide other than the B2R signal peptide portion. In an embodiment, the peptide fragment of the B2R protein comprises the full-length B2R peptide other than the B2R signal peptide portion, and the peptide fragment of the B2R protein comprises the amino acid sequence as set forth in SEQ ID NO:117. In an embodiment, the peptide fragment of the B2R protein comprises full-length
20 B2R peptide. In an embodiment, the full-length B2R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 118. In an embodiment, the peptide fragment of the B2R protein comprises a B2R extracellular segment. In an embodiment, the B2R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO:119. In an embodiment, the peptide fragment of the B2R protein comprises a B2R signal peptide and a B2R extracellular segment. In
25 an embodiment, the peptide fragment of the B2R protein comprises a B2R signal peptide and a B2R extracellular segment, and the peptide fragment of the B2R protein comprises the amino acid sequence of SEQ ID NO: 120.

In an embodiment, the peptide fragment of the B6R protein comprises a B6R signal peptide. In an embodiment, the B6R signal peptide comprises the amino acid sequence as set forth in SEQ

ID NO: 121. In an embodiment, the peptide fragment of the B6R protein comprises the full-length B6R peptide other than the B6R signal peptide portion. In an embodiment, the peptide fragment of the B6R protein comprises the full-length B6R peptide other than the B6R signal peptide portion, and the peptide fragment of the B6R protein comprises the amino acid sequence as set forth in
5 SEQ ID NO:122. In an embodiment, the peptide fragment of the B6R protein comprises full-length B6R peptide. In an embodiment, the full-length B6R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 123. In an embodiment, the peptide fragment of the B6R protein comprises a B6R extracellular segment. In an embodiment, the B6R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO:124. In an embodiment, the peptide
10 fragment of the B6R protein comprises a B6R signal peptide and a B6R extracellular segment. In an embodiment, the peptide fragment of the B6R protein comprises a B6R signal peptide and a B6R extracellular segment, and the peptide fragment of the B6R protein comprises the amino acid sequence of SEQ ID NO: 125.

In an embodiment, the peptide fragment of the C15L protein comprises full-length C15L
15 peptide. In an embodiment, the full-length C15L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 126. In an embodiment, the peptide fragment of the C15L protein comprises a C15L extracellular segment. In an embodiment, the C15L extracellular segment comprises the amino acid sequence of SEQ ID NO: 127.

In an embodiment, the peptide fragment of the M1R protein comprises full-length M1R peptide.
20 In an embodiment, the full-length M1R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 130. In an embodiment, the peptide fragment of the M1R protein comprises a M1R extracellular segment. In an embodiment, the M1R extracellular segment comprises the amino acid sequence of SEQ ID NO: 128.

In an embodiment, the peptide fragment of the H3L protein comprises full-length H3L peptide.
25 In an embodiment, the full-length H3L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 131. In an embodiment, the peptide fragment of the H3L protein comprises a H3L extracellular segment. In an embodiment, the H3L extracellular segment comprises the amino acid sequence of SEQ ID NO: 132.

In an embodiment, the polypeptide or polypeptide combination comprises the IgE signal peptide, the peptide fragment of A29L protein and the HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises the peptide fragment of A35R protein and the HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises the peptide fragment of B2R protein and the HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises the peptide fragment of B6R protein, the linker, and the HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises the IgE signal peptide, the linker, the HA polypeptide tag and the peptide fragment of C15L protein. In an embodiment, the polypeptide or polypeptide combination comprises the IgE signal peptide, the linker, the HA polypeptide tag and the peptide fragment of M1R protein. In an embodiment, the polypeptide or polypeptide combination comprises the IgE signal peptide, the linker, the HA polypeptide tag and the peptide fragment of H3L protein. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, C15L proteins, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A29L, B6R, M1R proteins, the IgE signal peptide, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of B6R, M1R, H3L proteins, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of B6R, M1R, A35R proteins, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A29L, B2R, C15L proteins, the IgE signal peptide, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of B2R, H3L, C15L proteins, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of B2R, A35R, B6R proteins, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A29L, M1R, C15L proteins, the IgE signal peptide, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of H3L, M1R, C15L proteins, the IgE signal peptide, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A29L, A35R, B2R, B6R, C15L and M1R proteins, the IgE signal peptide, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises the peptide fragments of A29L, A35R, B2R, B6R, C15L, and M1R

proteins, the IgE signal peptide, an HA polypeptide tag, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises the peptide fragments of A35R, B2R, B6R, C15L, M1R, and H3L proteins, the IgE signal peptide, an HA polypeptide tag, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, B6R, C15L, M1R, and H3L proteins, and the linker. In an embodiment, the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, B6R, C15L, M1R and H3L proteins, the IgE signal peptide, and the linker.

The linker may comprise an amino acid sequence of any length, particularly an amino acid sequence of 1-50, preferably 1-30, for example 1-10 amino acid residues. Exemplary linkers may include, but are not limited to, polyglycine (G), polyalanine (A), polyserine (S) or a combination thereof, for example GGAS, GGGs, GGGSG or (G4S)_n, wherein n is an integer of 1-30, preferably 1-10. The linker may also be a hinge region or a functional equivalent thereof. Other suitable linkers may be organic compounds or polymers generally suitable for use in pharmaceutical proteins, including but not limited to polyethylene glycol.

As used herein, the terms "HA polypeptide tag" and "HA tag protein" may be used interchangeably, which are derived from the hemagglutinin surface epitope of influenza virus, have little influence on the spatial structure of a foreign target protein, and are easily constructed as a tag protein fused to the N-terminus or C-terminus. "HA polypeptide tag" and "HA tag protein" are widely used for detection, separation and purification of proteins.

In an embodiment, the polypeptide or polypeptide combination comprises an IgE signal peptide. In an embodiment, the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134. In an embodiment, the polypeptide or polypeptide combination comprises an HA polypeptide tag. In an embodiment, the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 135 or 136. In an embodiment, the polypeptide or polypeptide combination comprises a linker. In an embodiment, the linker is a GS linker. In a preferred embodiment, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 137 or 138.

In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105,

106, 107, 108, 109 and 110, or an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to the amino acid sequence as set forth in one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a peptide fragment of the A29L protein, and an HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a peptide fragment of the A29L protein, and an HA polypeptide tag; the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134, the peptide fragment of the A29L protein comprises the amino acid sequence as set forth in SEQ ID NO: 111, and the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 135. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 89.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a peptide fragment of the A35R protein and an HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, full-length A35R peptide and an HA polypeptide tag, wherein the full-length A35R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 114, and the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 136. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 90.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a peptide fragment of the B2R protein and an HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B2R signal peptide, the full-length B2R peptide other than the signal peptide portion, and an HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B2R signal peptide, the full-length B2R peptide other than the signal peptide portion, and an HA polypeptide tag, wherein the B2R signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 116, the full-length B2R peptide other than the signal peptide portion comprises the amino acid sequence as set forth in SEQ ID

NO: 117, and the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 136. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, full-length B2R peptide and an HA polypeptide tag. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, full-length B2R peptide and an HA polypeptide tag, wherein the full-length B2R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 118, and the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 136. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 91.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a peptide fragment of the B6R protein, a linker, an HA polypeptide tag, and a peptide fragment of the B6R protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B6R signal peptide, a GS linker, an HA polypeptide tag, and the full-length B6R peptide other than the signal peptide portion. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B6R signal peptide, a GS linker, an HA polypeptide tag, and the full-length B6R peptide other than the signal peptide portion, wherein the B6R signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 121, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 137, the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 136, and the full-length B6R peptide other than the signal peptide portion comprises the amino acid sequence as set forth in SEQ ID NO: 122. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 92.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a linker, an HA polypeptide tag, and a peptide fragment of the C15L protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a GS linker, an HA polypeptide tag, and full-length C15L peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a GS linker, an HA polypeptide tag, and full-length C15L peptide, wherein the IgE signal peptide comprises the amino acid sequence as set

forth in SEQ ID NO: 134, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 137, the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 136, and the full-length C15L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 126. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 93.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a linker, an HA polypeptide tag, and a peptide fragment of the M1R protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a GS linker, an HA polypeptide tag, and an M1R extracellular segment. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a GS linker, an HA polypeptide tag, and an M1R extracellular segment, wherein the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 137, the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 136, and the M1R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 128. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 94.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a linker, an HA polypeptide tag, and a peptide fragment of the H3L protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a GS linker, an HA polypeptide tag, and full-length H3L peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a GS linker, an HA polypeptide tag, and full-length H3L peptide, wherein the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 137, the HA polypeptide tag comprises the amino acid sequence as set forth in SEQ ID NO: 136, and the full-length H3L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 131. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 95.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a peptide fragment of the A29L protein, a linker, a peptide fragment of the A35R protein, a linker, a peptide fragment of the B6R protein, a linker, a peptide fragment of the B2R protein, a linker, a peptide fragment of the C15L protein, a linker, and a peptide fragment of the M1R protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, full-length A29L peptide, a GS linker, an A35R extracellular segment, a GS linker, a B6R extracellular segment, a GS linker, a B2R extracellular segment, a GS linker, a C15L extracellular segment, a GS linker, and full-length M1R peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, full-length A29L peptide, a GS linker, an A35R extracellular segment, a GS linker, a B6R extracellular segment, a GS linker, a B2R extracellular segment, a GS linker, a C15L extracellular segment, a GS linker, and full-length M1R peptide; wherein the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134, the full-length A29L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 113, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the A35R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 115, the B6R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 124, the B2R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 119, the C15L extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 127, and the full-length M1R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 130. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 101.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a peptide fragment of the B2R protein, a linker, a peptide fragment of the A35R protein, a linker, and a peptide fragment of the C15L protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B2R signal peptide, a B2R extracellular segment, a GS linker, an A35R extracellular segment, a GS linker, and full-length C15L peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B2R signal peptide, a B2R extracellular segment, a GS linker,

an A35R extracellular segment, a GS linker, and full-length C15L peptide; wherein the B2R signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 116, the B2R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 119, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the A35R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 115, and the full-length C15L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 126. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 102.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a peptide fragment of the A29L protein, a linker, a peptide fragment of the M1R protein, a linker, and a peptide fragment of the B6R protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, full-length A29L peptide, a GS linker, an M1R extracellular segment, a GS linker, and the full-length B6R peptide other than the signal peptide portion. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, full-length A29L peptide, a GS linker, an M1R extracellular segment, a GS linker, and the full-length B6R peptide other than the signal peptide portion; wherein the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134, the full-length A29L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 113, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the M1R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 128, and the full-length B6R peptide other than the signal peptide portion comprises the amino acid sequence as set forth in SEQ ID NO: 122. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 103.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a peptide fragment of the B6R protein, a linker, a peptide fragment of the M1R protein, a linker, and a peptide fragment of the H3L protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B6R signal peptide, a B6R extracellular segment, a GS linker, an M1R extracellular segment, a GS linker, and full-length H3L

peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B6R signal peptide, a B6R extracellular segment, a GS linker, an M1R extracellular segment, a GS linker, and full-length H3L peptide; wherein the B6R signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 121, the B6R extracellular segment
5 comprises the amino acid sequence as set forth in SEQ ID NO: 124, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the M1R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 128, and the full-length H3L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 131. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ
10 ID NO: 104.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a peptide fragment of the B6R protein, a linker, a peptide fragment of the A35R protein, a linker, and a peptide fragment of the M1R protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B6R signal peptide, a
15 B6R extracellular segment, a GS linker, an A35R extracellular segment, a GS linker, and full-length M1R peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B6R signal peptide, a B6R extracellular segment, a GS linker, an A35R extracellular segment, a GS linker, and full-length M1R peptide; wherein the B6R signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 121, the B6R extracellular
20 segment comprises the amino acid sequence as set forth in SEQ ID NO: 124, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the A35R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 115, and the full-length M1R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 130. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as
25 set forth in SEQ ID NO: 105.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a peptide fragment of the A29L protein, a linker, a peptide fragment of the B2R protein, a linker, and a peptide fragment of the C15L protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-

terminus, an IgE signal peptide, full-length A29L peptide, a GS linker, a B2R extracellular segment, a GS linker, and full-length C15L peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, full-length A29L peptide, a GS linker, a B2R extracellular segment, a GS linker, and full-length C15L peptide; wherein the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134, the full-length A29L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 113, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the B2R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 119, and the full-length C15L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 126. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 106.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a peptide fragment of the B2R protein, a linker, a peptide fragment of the H3L protein, a linker, and a peptide fragment of the C15L protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B2R signal peptide, a B2R extracellular segment, a GS linker, an H3L extracellular segment, a GS linker, and full-length C15L peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B2R signal peptide, a B2R extracellular segment, a GS linker, an H3L extracellular segment, a GS linker, and full-length C15L peptide; wherein the B2R signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 116, the B2R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 119, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the H3L extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 132, and the full-length C15L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 126. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 107.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a peptide fragment of the B2R protein, a linker, a peptide fragment of the A35R protein, a linker, and a peptide fragment of the B6R protein. In an embodiment, the polypeptide or

polypeptide combination comprises, from N-terminus to C-terminus, a B2R signal peptide, a B2R extracellular segment, a GS linker, an A35R extracellular segment, a GS linker, and the full-length B6R peptide other than the signal peptide portion. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, a B2R signal peptide, a B2R
5 extracellular segment, a GS linker, an A35R extracellular segment, a GS linker, and the full-length B6R peptide other than the signal peptide portion; wherein the B2R signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 116, the B2R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 119, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the A35R extracellular segment comprises the amino
10 acid sequence as set forth in SEQ ID NO: 115, and the full-length B6R peptide other than the signal peptide portion comprises the amino acid sequence as set forth in SEQ ID NO: 122. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 108.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus
15 to C-terminus, an IgE signal peptide, a peptide fragment of the A29L protein, a linker, a peptide fragment of the M1R protein, a linker, and a peptide fragment of the C15L protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, full-length A29L peptide, a GS linker, an M1R extracellular segment, a GS linker, and full-length C15L peptide. In an embodiment, the polypeptide or
20 polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, full-length A29L peptide, a GS linker, an M1R extracellular segment, a GS linker, and full-length C15L peptide; wherein the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134, the full-length A29L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 113, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the
25 M1R extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 128, and the full-length C15L peptide comprises the amino acid sequence as set forth in SEQ ID NO: 126. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 109.

In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, a peptide fragment of the H3L protein, a linker, a peptide fragment of the C15L protein, a linker, and a peptide fragment of the M1R protein. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, an H3L extracellular segment, a GS linker, a C15L extracellular segment, a GS linker, and full-length M1R peptide. In an embodiment, the polypeptide or polypeptide combination comprises, from N-terminus to C-terminus, an IgE signal peptide, an H3L extracellular segment, a GS linker, a C15L extracellular segment, a GS linker, and full-length M1R peptide; wherein the IgE signal peptide comprises the amino acid sequence as set forth in SEQ ID NO: 134, the H3L extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 132, the GS linker comprises the amino acid sequence as set forth in SEQ ID NO: 138, the C15L extracellular segment comprises the amino acid sequence as set forth in SEQ ID NO: 127, and the full-length M1R peptide comprises the amino acid sequence as set forth in SEQ ID NO: 130. In an embodiment, the polypeptide or polypeptide combination comprises the amino acid sequence as set forth in SEQ ID NO: 110.

In an embodiment, the polypeptide or polypeptide combination comprises two or more amino acid sequences selected from SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110, or an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to the amino acid sequence as set forth in SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide, a second polypeptide, a third polypeptide, a fourth polypeptide, a fifth polypeptide and a sixth polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 89, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 90, the third polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 91, the fourth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 92, the fifth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 93, and the sixth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 94.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide, a second polypeptide, a third polypeptide, a fourth polypeptide, a fifth polypeptide and a sixth polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 90, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 91, the third polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 92, the fourth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 93, the fifth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 94, the sixth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 95.

10 In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 102, and the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 103.

15 In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 105, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 106.

20 In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 108, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 109.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 102, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 104.

25 In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 105, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 107.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 110, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 108.

5 The polypeptide or polypeptide combination of the present invention may be used as a polypeptide antigen for inducing a protective immune response against mpox virus infection in a subject.

Nucleic acid:

10 In one aspect, the present invention provides a nucleic acid or nucleic acid combination encoding the polypeptide or polypeptide combination described herein. In some embodiments, the nucleic acid or nucleic acid combination comprises a polynucleotide encoding the polypeptide or polypeptide combination described herein. The nucleic acid may be single-stranded or double-stranded. Nucleic acids include, but are not limited to, DNA, cDNA, RNA (e.g., mRNA, circRNA or saRNA), recombinantly produced and chemically synthesized nucleic acids. The nucleic acid
15 or nucleic acid combination of the present invention may be comprised in a vector. The nucleic acid of the present invention may include naturally occurring, synthetic and modified nucleotides. In some embodiments, the nucleic acid or nucleic acid combination of the present invention is used to express the polypeptide or polypeptide combination described herein in a cell to provide a polypeptide antigen. In some embodiments, the polypeptide antigen may induce an immune
20 response against mpox virus in a suitable subject.

 The nucleic acid may comprise one or more segments (nucleotide fragments) (for example, 1, 2, 3, 4, 5, 6, 7, 8 or more segments). The nucleic acid may comprise a segment encoding a polypeptide of interest (for example, a polypeptide and polypeptide antigen described herein). In particular embodiments, the nucleic acid may comprise a coding sequence of a polypeptide of
25 interest and regulatory sequences (including but not limited to transcriptional and translational regulatory sequences). In an embodiment, the regulatory sequence comprises one or more of the following: a promoter sequence, a 5' untranslated region (5' UTR) sequence, a 3' untranslated region (3' UTR) sequence, and a poly(A) sequence.

Coding sequence

As used herein, a "coding sequence" refers to a nucleotide sequence in a nucleic acid that can serve as a template for synthesis in a biological process of a nucleotide sequence (for example, tRNA and mRNA) having a defined nucleotide sequence or a defined amino acid sequence. The coding sequence may be a DNA sequence or an RNA sequence. If an mRNA corresponding to a DNA sequence (including a coding strand identical to the mRNA sequence and a template strand complementary thereto) is translated into a polypeptide in a biological process, the DNA sequence or mRNA sequence may be considered to encode the polypeptide.

As used herein, "codon" refers to a nucleotide sequence of three consecutive nucleotides in a nucleic acid (also known as a triplet codon), which encodes a specific amino acid. Synonymous codons (codons encoding the same amino acid) are used at different frequencies in different species, which is referred to as "codon preference". It is generally believed that, for a given species, a coding sequence using codons preferred by that species can have higher translation efficiency and accuracy in the expression system of that species. Therefore, the nucleic acid may be "codon-optimized", i.e., codons in the nucleic acid are changed to reflect codons preferred by a host cell, while preferably not altering the amino acid sequence encoded thereby. Those skilled in the art will understand that, due to the degeneracy of codons, the nucleic acid or nucleic acid combination of the present invention may comprise such a coding sequence that differs from the coding sequence described herein (for example, having about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the coding sequence described herein) but encodes the same amino acid sequence. In particular embodiments, the nucleic acid of the present invention comprises codons optimized for a host (for example, a subject, particularly a human) cell, such that the polypeptide of the present invention is optimally expressed in the host (for example, a subject, particularly a human). In particular embodiments, the RNA of the present invention comprises codons optimized for a host (for example, a subject, particularly a human) cell, such that the polypeptide of the present invention is optimally expressed in the host (for example, a subject, particularly a human).

In some embodiments, the nucleic acid or nucleic acid combination of the present invention comprises a coding sequence of a polypeptide as described herein. In some embodiments, the nucleic acid or nucleic acid combination of the present invention comprises a coding sequence of

a polypeptide antigen as described herein. In some embodiments, the nucleic acid or nucleic acid combination of the present invention comprises a nucleotide sequence complementary to the coding sequence described herein. In some embodiments, the coding sequence comprises a start codon at its 5' end and comprises a stop codon at its 3' end. In some embodiments, the coding sequence comprises an open reading frame (ORF) described herein.

In some embodiments, the coding sequence of the present invention encodes any polypeptide antigen of the present invention.

In an embodiment, the coding sequence of the present invention encodes a polypeptide antigen, wherein the polypeptide antigen comprises:

- (1) the amino acid sequence as set forth in one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110; (2) an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to the amino acid sequence as set forth in one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110; (3) an immunogenic fragment of the amino acid sequence as set forth in one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110; (4) an immunogenic fragment of an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to the amino acid sequence as set forth in one of SEQ ID NOs: 89, 90, 91, 92, 93, 94, 95, 101, 102, 103, 104, 105, 106, 107, 108, 109 and 110.

- In an embodiment, the coding sequence of the present invention encodes a polypeptide, wherein the polypeptide comprises the amino acid sequence of the polypeptide antigen as described above.

- In an embodiment, the coding sequence of the polypeptide or polypeptide antigen described herein comprises a nucleotide sequence, wherein the nucleotide sequence comprises: (1) the nucleotide sequence as set forth in one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22; (2) a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22; (3) the nucleotide sequence as set forth in one of SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44; or (4) a

nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination
5 described herein comprises a nucleotide sequence, wherein the nucleotide sequence comprises: (1) two or more nucleotide sequences selected from SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22; (2) two or more nucleotide sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to nucleotide sequences selected from SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22; (3) two or more nucleotide sequences
10 selected from SEQ ID NO: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44, or (4) two or more nucleotide sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to nucleotide sequences selected from SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination
15 comprises nucleotide sequences of a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 1 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 1; the nucleotide
20 sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 2 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 2; the nucleotide sequence of the third polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 3 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%
25 identity to the nucleotide sequence as set forth in SEQ ID NO: 3; the nucleotide sequence of the fourth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 4 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 4; the nucleotide sequence of the fifth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 5 or a

nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 5; the nucleotide sequence of the sixth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 6 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 6.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 23 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 23; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 24 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 24; the nucleotide sequence of the third polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 25 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 25; the nucleotide sequence of the fourth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 26 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 26; the nucleotide sequence of the fifth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 27 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 27; the nucleotide sequence of the sixth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 28 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 28.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein

the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 2 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 2; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 3 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 3; the nucleotide sequence of the third polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 4 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 4; the nucleotide sequence of the fourth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 5 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 5; the nucleotide sequence of the fifth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 6 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 6; the nucleotide sequence of the sixth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 7 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 7.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 24 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 24; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 25 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 25; the nucleotide sequence of the third polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 26 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%,

97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 26; the nucleotide sequence of the fourth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 27 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 27; the nucleotide sequence of the fifth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 28 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 28; the nucleotide sequence of the sixth polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 29 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 29.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 14 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 14; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 15 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 15.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 36 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 36; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 37 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 37.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in

SEQ ID NO: 17 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 17; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 18 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 18.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 39 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 39; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 40 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 40.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 20 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 20; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 21 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 21.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 42 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 42; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 43 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 43.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 14 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 14; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 16 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 16.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 36 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 36; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 38 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 38.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 17 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 17; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 19 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 19.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 39 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 39; the nucleotide

sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 41 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 41.

5 In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 22 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 22; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID
10 NO: 20 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 20.

In an embodiment, the coding sequence of the polypeptide or polypeptide combination comprises nucleotide sequences of a first polynucleotide and a second polynucleotide; wherein the nucleotide sequence of the first polynucleotide comprises: the nucleotide sequence as set forth in
15 SEQ ID NO: 44 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 44; the nucleotide sequence of the second polynucleotide comprises: the nucleotide sequence as set forth in SEQ ID NO: 42 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 42.

20 In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide, a second polypeptide, a third polypeptide, a fourth polypeptide, a fifth polypeptide and a sixth polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 89, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 90, the third polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 91, the
25 fourth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 92, the fifth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 93, and the sixth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 94; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide, the second polypeptide, the third polypeptide, the fourth polypeptide, the fifth polypeptide and the

sixth polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 1 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 1; the nucleotide sequence encoding the second polypeptide comprises: the
5 nucleotide sequence as set forth in SEQ ID NO: 2 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 2; the nucleotide sequence encoding the third polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 3 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth
10 in SEQ ID NO: 3; the nucleotide sequence encoding the fourth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 4 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 4; the nucleotide sequence encoding the fifth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 5 or a nucleotide sequence having at least 70%, 75%, 80%,
15 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 5; the nucleotide sequence encoding the sixth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 6 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 6.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide,
20 a second polypeptide, a third polypeptide, a fourth polypeptide, a fifth polypeptide and a sixth polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 89, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 90, the third polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 91, the fourth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 92, the fifth
25 polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 93, the sixth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 94; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide, the second polypeptide, the third polypeptide, the fourth polypeptide, the fifth polypeptide and the sixth polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the

nucleotide sequence as set forth in SEQ ID NO: 23 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 23; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 24 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 24; the nucleotide sequence encoding the third polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 25 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 25; the nucleotide sequence encoding the fourth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 26 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 26; the nucleotide sequence encoding the fifth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 27 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 27; the nucleotide sequence encoding the sixth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 28 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 28.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide, a second polypeptide, a third polypeptide, a fourth polypeptide, a fifth polypeptide and a sixth polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 90, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 91, the third polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 92, the fourth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 93, the fifth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 94, the sixth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 95; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide, the second polypeptide, the third polypeptide, the fourth polypeptide, the fifth polypeptide and the sixth polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the

nucleotide sequence as set forth in SEQ ID NO: 2 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 2; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 3 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 3; the nucleotide sequence encoding the third polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 4 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 4; the nucleotide sequence encoding the fourth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 5 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 5; the nucleotide sequence encoding the fifth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 6 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 6; the nucleotide sequence encoding the sixth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 7 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 7.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide, a second polypeptide, a third polypeptide, a fourth polypeptide, a fifth polypeptide and a sixth polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 90, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 91, the third polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 92, the fourth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 93, the fifth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 94, the sixth polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 95; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide, the second polypeptide, the third polypeptide, the fourth polypeptide, the fifth polypeptide and the sixth polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 24 or a nucleotide sequence having at least 70%,

75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 24; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 25 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 25; the nucleotide sequence encoding the third polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 26 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 26; the nucleotide sequence encoding the fourth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 27 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 27; the nucleotide sequence encoding the fifth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 28 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 28; the nucleotide sequence encoding the sixth polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 29 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 29.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 102, and the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 103; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 14 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 14; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 15 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 15.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 102, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 103; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 36 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 36; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 37 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 37.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 105, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 106; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 17 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 17; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 18 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 18.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 105, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 106; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 39 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 39; the nucleotide sequence encoding the second polypeptide comprises: the

nucleotide sequence as set forth in SEQ ID NO: 40 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 40.

5 In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 108, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 109; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 20
10 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 20; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 21 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 21.

15 In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 108, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 109; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence
20 encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 42 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 42; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 43 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%,
25 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 43.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 102, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 104; the coding sequence described herein comprises nucleotide sequences

encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 14 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 14; the nucleotide sequence
5 encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 16 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 16.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set
10 forth in SEQ ID NO: 102, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 104; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 36 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%
15 identity to the nucleotide sequence as set forth in SEQ ID NO: 36; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 38 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 38.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide
20 and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 105, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 107; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 17 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%
25 identity to the nucleotide sequence as set forth in SEQ ID NO: 17; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 19 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 19.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 105, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 107; the coding sequence described herein comprises nucleotide sequences
5 encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 39 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 39; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO:
10 41 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 41.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 110, the second polypeptide comprises the amino acid sequence as set forth
15 in SEQ ID NO: 108; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 22 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 22; the nucleotide sequence
20 encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 20 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 20.

In an embodiment, the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide comprises the amino acid sequence as set
25 forth in SEQ ID NO: 110, the second polypeptide comprises the amino acid sequence as set forth in SEQ ID NO: 108; the coding sequence described herein comprises nucleotide sequences encoding the first polypeptide and the second polypeptide; wherein the nucleotide sequence encoding the first polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 44 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%

identity to the nucleotide sequence as set forth in SEQ ID NO: 44; the nucleotide sequence encoding the second polypeptide comprises: the nucleotide sequence as set forth in SEQ ID NO: 42 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to the nucleotide sequence as set forth in SEQ ID NO: 42.

5 **RNA**

In some embodiments, the nucleic acid or nucleic acid combination of the present invention is RNA. In some embodiments, the polynucleotide of the present invention is RNA. As used herein, the definition of "RNA" encompasses single-stranded, double-stranded, linear and circular RNA. The RNA of the present invention may be non-amplifying RNA or self-amplifying RNA. The RNA
10 of the present invention may be RNA produced by chemical synthesis, recombinantly produced RNA, and RNA produced by in vitro transcription. In some embodiments, the RNA of the present invention is used for expressing the polypeptide of the present invention in a host cell. In some embodiments, the RNA of the present invention is used for expressing the polypeptide combination of the present invention in a host cell.

15 In some embodiments, the RNA of the present invention is single-stranded RNA. In some embodiments, the RNA of the present invention is RNA produced by in vitro transcription (IVT-RNA). IVT-RNA can be obtained by in vitro transcription using a DNA template by an RNA polymerase (for example, as described herein).

In some embodiments, the RNA of the present invention is messenger RNA (mRNA). In
20 general, mRNA may comprise a 5'-UTR sequence, a coding sequence of a polypeptide (for example, a polypeptide or polypeptide combination of the present invention), a 3'-UTR sequence and an optional poly(A) sequence. MRNA may be produced, for example, by in vitro transcription or chemical synthesis. In an embodiment, the mRNA of the present invention is obtained by in vitro transcription using a DNA template by an RNA polymerase (for example, T7 RNA
25 polymerase). In an embodiment, the mRNA of the present invention comprises (1) 5'-UTR, (2) a coding sequence, (3) 3'-UTR and (4) an optional poly(A) sequence. The 5'-UTR, coding sequence, 3'-UTR and poly(A) sequence are as described herein. In an embodiment, the mRNA of the present invention is a nucleoside-modified mRNA. In an embodiment, the mRNA of the present invention

comprises an optional 5' cap. In an embodiment, the mRNA of the present invention comprises a cap-proximal sequence.

In some embodiments, the RNA of the present invention is circular RNA (circRNA). As used herein, the term "circRNA" or "circular RNA" refers to RNA that forms a circular structure through covalent bonds. In general, circRNA is a single-stranded RNA, which may comprise a UTR (5'-UTR and/or 3'-UTR) and an open reading frame, but does not comprise a 5' cap and a poly(A) tail. CircRNA may comprise elements promoting cap-independent translation in the UTR, such as an internal ribosome entry site (IRES), and an optional polyadenylate (polyA) tract or polyadenylate interspersed with cytidylate (polyAC). CircRNA can be produced using enzymatic or chemical methods. In general, enzymatic methods may include, for example, end ligation of a linear precursor RNA catalyzed by T4 ligase and formation of circRNA using a ribozyme during IVT. Methods for preparing circRNA using ribozymes may include, for example, group I intron-based methods, group II intron-based methods and hairpin ribozyme-based methods (see, for example, Wesselhoeft R. A. et al., *Nature Communications*, (2018) 9: 2629; Petkovic, S. et al., *Nucleic Acids Research*, 2015, Vol. 43, No. 4, 2454-2465; Chen, XJ et al., *Front. Bioeng. Biotechnol.*, 30 November 2021; and Lambowitz, A.M. et al., *Cold Spring Harb Perspect Biol.* 2011 Aug; 3(8): a003616).

In some embodiments, the RNA of the present invention is self-amplifying RNA (saRNA). As used herein, the terms "self-replicating RNA", "self-amplifying RNA (saRNA)" and "replicon" may be used interchangeably to refer to RNA that can self-amplify and be translated. In general, saRNA may be derived from or comprise a self-replicating subgenome of a viral genome (for example, a positive-sense single-stranded RNA virus, such as an alphavirus, such as Venezuelan equine encephalitis virus (VEEV)), which includes viral genes encoding nonstructural proteins nsP1, nsP2, nsP3 and nsP4 that form RNA-dependent RNA polymerase (RdRP), while the viral genes encoding structural proteins are replaced by genes encoding a polypeptide of interest. SaRNA may comprise two open reading frames (ORFs) separated by a subgenomic promoter, namely a first ORF encoding nsP1-nsP4 followed at the 3' end by a second ORF encoding a polypeptide of interest (for example, a polypeptide or polypeptide combination described herein) under the drive of the subgenomic promoter. SaRNA may further comprise a 5' cap and a 5'-UTR

at the 5' end, and a 3'-UTR and a poly(A) sequence at the 3' end. In addition, the untranslated regions of saRNA may further comprise 5' and 3' conserved sequence elements (CSEs) required for self-amplification for recognition by RdRP. Similar to mRNA, saRNA may also be produced by in vitro transcription or chemical synthesis. For a detailed description of saRNA, see, for example, WO2016135675A1 and WO2021183564A1.

In some embodiments, the RNA of the present invention (for example, mRNA, circRNA and saRNA) comprises a coding sequence of a polypeptide as described herein. In some embodiments, the RNA of the present invention (for example, mRNA, circRNA and saRNA) comprises a coding sequence of a polypeptide antigen as described herein. In some embodiments, the RNA of the present invention (for example, mRNA, circRNA and saRNA) comprises a coding sequence of a polypeptide combination as described herein.

In some embodiments, the RNA of the present invention further comprises structural elements that help improve the stability and/or translation efficiency of RNA, including but not limited to a 5' cap, a cap-proximal sequence, 5'-UTR, 3'-UTR and a poly(A) sequence.

As used herein, the term "untranslated region (UTR)" generally refers to a region (non-coding region) in RNA (such as mRNA) that is not translated into an amino acid sequence, or a corresponding region in DNA. Generally, a UTR located at the 5' end (upstream) of an open reading frame (start codon) may be referred to as a 5' untranslated region, 5'-UTR; a UTR located at the 3' end (downstream) of an open reading frame (stop codon) may be referred to as a 3'-UTR. In the presence of a 5' cap, the 5'-UTR is located downstream of the 5' cap, for example, directly adjacent to the 5' cap. In particular embodiments, an optimized "Kozak sequence" may be included in the 5'-UTR, for example, at a position proximate to the start codon, to improve translation efficiency. In the presence of a poly(A) sequence, the 3'-UTR is located upstream of the poly(A) sequence, for example, directly adjacent to the poly(A) sequence.

In some embodiments, the RNA of the present invention comprises a 5'-UTR. In a preferred embodiment, the 5'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 139 or 148. In some embodiments, the RNA of the present invention comprises a 3'-UTR. In a preferred embodiment, the 3'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 140, 141 or 149. In some embodiments, the RNA of the present invention comprises a 5'-UTR and a 3'-

UTR. In an embodiment, the 5'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 139 or 148, and the 3'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 140, 141 or 149. In a specific embodiment, the 5'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 139, and the 3'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 140 or 141. In a specific embodiment, the 5'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 148, and the 3'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 149.

As used herein, the term "poly(A) sequence" or "poly(A) tail" refers to a nucleotide sequence comprising continuous or discontinuous adenylates. The poly(A) sequence is typically located at the 3' end of RNA, for example, at the 3' end (downstream) of the 3'-UTR. In some embodiments, the poly(A) sequence does not comprise nucleotides other than adenylates at its 3' end. The poly(A) sequence may be transcribed during preparation of IVT-RNA by a DNA-dependent RNA polymerase according to the coding sequence of a DNA template, or may be ligated to the free 3' end of IVT-RNA, for example the 3' end of the 3'-UTR, by a DNA-independent RNA polymerase (poly(A) polymerase).

In some embodiments, the RNA of the present invention comprises a poly(A) sequence. In an embodiment, the poly(A) sequence comprises continuous adenylates. In an embodiment, the poly(A) sequence may comprise at least 20, 30, 40, 50, 60, 70, 75, 80, 85, 95 or 100 and up to 120, 150, 180, 200, 300 adenylates. In an embodiment, the continuous adenylate sequence in the poly(A) sequence is interrupted by a sequence comprising U, C or G nucleotides. In an embodiment, the poly(A) sequence comprises at least 50 nucleotides. In an embodiment, the poly(A) sequence comprises at least 80 nucleotides. In an embodiment, the poly(A) sequence comprises at least 100 nucleotides. In some embodiments, the poly(A) sequence comprises about 70, 80, 90, 100, 120 or 150 nucleotides. In a preferred embodiment, the poly(A) sequence comprises 121 nucleotides. In a preferred embodiment, the poly(A) sequence comprises the nucleotide sequence as set forth in SEQ ID NO: 145.

As used herein, the term "5' cap" generally relates to an N7-methylguanosine structure linked to the 5' end of mRNA through a 5' to 5' triphosphate bond (also referred to as an "m7G cap" or "m7Gppp-"). The 5' cap may be co-transcriptionally added to RNA during in vitro transcription

(for example, using anti-reverse cap analog "ARCA"), or may be ligated to RNA after transcription using a capping enzyme. In some embodiments, RNA (for example, IVT-RNA) is modified into "Cap0 RNA" using a capping enzyme (for example, vaccinia virus capping enzyme). In some embodiments, additional methylation occurs at the ribose 2'-O position of the nucleotide immediately adjacent to the m⁷G cap in Cap0 RNA (for example, by 2'-O-methyltransferase),
5 producing "Cap1 RNA". In some embodiments, a cap analog is used to produce 5' cap-modified RNA. For a description of "cap analogs", see, for example, Contreas, R. et al. (1982). Nucl. Acids Res. 10, 6353-6363 and US7074596B2. Examples of cap analogs include, but are not limited to, N7-methylguanosine-5'-triphosphate-5'-guanosine (m⁷G(5')ppp(5')G), N7-methylguanosine-5'-triphosphate-5'-adenosine (m⁷G(5')ppp(5')A) and 3'-O-Me-m⁷G(5')ppp(5')G (ARCA).
10

The RNA of the present invention (for example, mRNA and saRNA) may be RNA comprising a Cap0 (no methylation occurs on the ribose of the nucleotide adjacent to m⁷G), Cap1 (methylation of the ribose of the nucleotide adjacent to m⁷G) or Cap2 (methylation of the ribose of the second nucleotide downstream of m⁷G) structure. In a preferred embodiment, the mRNA of the present invention is Cap1 mRNA. In a preferred embodiment, the mRNA of the present invention
15 comprises ⁷MeGpppG_{2'}OMe.

In some embodiments, the RNA of the present invention further comprises a cap-proximal sequence. In some embodiments, the 5' UTR comprises a cap-proximal sequence. In some embodiments, the 5' UTR does not comprise a cap-proximal sequence. In some embodiments, the cap-proximal sequence comprises a sequence adjacent to the 5' cap. In an embodiment, the cap-proximal sequence comprises the nucleotide sequence as set forth in SEQ ID NO: 147. In an embodiment, one or more residues in the cap-proximal sequence may be included in the RNA because they have been included in the cap entity (for example, a Cap1 structure, etc.); or, in an embodiment, at least some residues in the cap-proximal sequence may be enzymatically added
25 (for example, by a polymerase such as T7 polymerase).

In some embodiments, the nucleic acid or nucleic acid combination of the present invention comprises a polynucleotide encoding a polypeptide or polypeptide combination described herein.

In some embodiments, the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA, wherein said polynucleotide comprises the nucleotide sequence as set

forth in one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22. In some embodiments, the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA, wherein the polynucleotide comprises the nucleotide sequence as set forth in one of SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66, or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to nucleotide sequences as set forth in one of SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66.

10 In some embodiments, the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA, wherein said polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22, or two or more nucleotide sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity to nucleotide sequences selected from SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22. In some embodiments, the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA, wherein the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66, or two or more nucleotide sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to nucleotide sequences selected from SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66.

25 In an embodiment, the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 1, the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 2, the third polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 3, the fourth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 4, the fifth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 5, and the sixth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 6. In an embodiment, the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third

polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 45, the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 46, and the third polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 47, and the fourth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 48, and the fifth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 49, and the sixth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 50.

In an embodiment, the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 2, the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 3, the third polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 4, the fourth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 5, the fifth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 6, and the sixth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 7. In an embodiment, the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 46, the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 47, and the third polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 48, and the fourth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 49, and the fifth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 50, and the sixth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 51.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 14, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 15. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set

forth in SEQ ID NO: 58, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 59.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in
5 SEQ ID NO: 17, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 18. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 61, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 62.

10 In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 20, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 21. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set
15 forth in SEQ ID NO: 64, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 65.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 14, and the second polynucleotide comprises the nucleotide sequence as set forth in
20 SEQ ID NO: 16. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 58, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 60.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second
25 polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 17, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 19. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set

forth in SEQ ID NO: 61, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 63.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 22, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 20. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 66, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 64.

10 **Modified nucleotides**

In some embodiments, the nucleotides in the RNA (for example, mRNA) of the present invention may be naturally occurring nucleotides (for example, naturally occurring ribonucleotides) and modified nucleotides. The modified nucleotide may be, for example, a nucleotide not present in naturally occurring RNA, such as a non-standard nucleotide or a deoxynucleotide. Modification of a nucleotide may occur on a nucleoside, for example on the ribose moiety and/or nucleobase moiety. The modified nucleotide may be incorporated during transcription (for example, in vitro transcription), or may be added during chemical synthesis of RNA.

In some embodiments, RNA is modified by comprising one or more modified nucleosides. In some embodiments, the RNA (for example, mRNA) of the present invention is modified by comprising one or more modified nucleobases. In some embodiments, the modified nucleobase comprises modified cytosine, modified uracil, modified adenine, modified guanine or a combination thereof.

Exemplary nucleobases and nucleosides having modified uracil may include, but are not limited to: pseudouridine (ψ), pyridin-4-one ribonucleoside, 5-aza-uracil, 6-aza-uracil, 2-thio-5-aza-uracil, 2-thio-uracil (s2U), 4-thio-uracil (s4U), 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxy-uracil (ho5U), 5-aminoallyl-uracil, 5-halo-uracil (for example, 5-iodo-uracil or 5-bromo-uracil), 3-methyluracil (m3U), 5-methoxy-uracil (mo5U), uracil 5-oxyacetic acid (cmo5U), methyl uracil 5-oxyacetate (mcmo5U), 5-carboxymethyl-uracil (cm5U), 1-carboxymethyl-pseudouridine, 5-carboxyhydroxymethyl-uracil (chm5U), methyl 5-carboxyhydroxymethyl-uracil (mchm5U), 5-

methoxycarbonylmethyl-uracil (mcm5U), 5-methoxycarbonylmethyl-2-thio-uracil (mcm5s2U), 5-aminomethyl-2-thio-uracil (nm5s2U), 5-methylaminomethyl-uracil (mnm5U), 5-methylaminomethyl-2-thio-uracil (mnm5s2U), 5-methylaminomethyl-2-seleno-uracil (mnm5se2U), 5-carbamoylmethyl-uracil (ncm5U), 5-carboxymethylaminomethyl-uracil (cmnm5U), 5-carboxymethylaminomethyl-2-thio-uracil (cmnm5s2U), 5-propynyl-uracil, 1-propynyl-pseudouracil, 5-taurinomethyl-uracil (tm5U), 1-taurinomethyl-pseudouridine, 5-taurinomethyl-2-thio-uracil (tm5s2U), 1-taurinomethyl-4-thio-pseudouridine, 5-methyl-uracil (m5U, namely having the nucleobase deoxythymine), 1-methyl-pseudouridine (m1ψ), 1-ethyl-pseudouridine (Et1ψ), 5-methyl-2-thio-uracil (m5s2U), 1-methyl-4-thio-pseudouridine (m1s4ψ), 4-thio-1-methyl-pseudouridine, 3-methyl-pseudouridine (m3ψ), 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouracil (D), dihydropseudouridine, 5,6-dihydrouracil, 5-methyl-dihydrouracil (m5D), 2-thio-dihydrouracil, 2-thio-dihydropseudouridine, 2-methoxy-uracil, 2-methoxy-4-thio-uracil, 4-methoxy-pseudouridine, 4-methoxy-2-thio-pseudouridine, N1-methyl-pseudouridine, 3-(3-amino-3-carboxypropyl)uracil (acp3U), 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine (acp3ψ), 5-(isopentenylaminomethyl)uracil (m5U), 5-(isopentenylaminomethyl)-2-thio-uracil (m5s2U), 5,2'-O-dimethyl-uridine (m5Um), 2-thio-2'-O-methyl-uridine (s2Um), 5-methoxycarbonylmethyl-2'-O-methyl-uridine (mcm5Um), 5-carbamoylmethyl-2'-O-methyl-uridine (ncm5Um), 5-carboxymethylaminomethyl-2'-O-methyl-uridine (cmnm5Um), 3,2'-O-dimethyl-uridine (m3Um) and 5-(isopentenylaminomethyl)-2'-O-methyl-uridine (inm5Um), 1-thio-uracil, deoxythymidine, 5-(2-methoxycarbonyl-ethenyl)-uracil, 5-(carbamoyl-hydroxymethyl)-uracil, 5-carbamoylmethyl-2-thio-uracil, 5-carboxymethyl-2-thio-uracil, 5-cyanomethyl-uracil, 5-methoxy-2-thio-uracil and 5-[3-(1-E-propenylamino)]uracil.

Exemplary nucleobases and nucleosides having modified cytosine may include, but are not limited to: 5-azacytosine, 6-azacytosine, pseudoisocytidine, 3-methylcytosine (m3C), N4-acetylcytosine (ac4C), 5-formylcytosine (f5C), N4-methyl-cytosine (m4C), 5-methyl-cytosine (m5C), 5-halo-cytosine (for example, 5-iodo-cytosine), 5-hydroxymethyl-cytosine (hm5C), 1-methyl-pseudoisocytidine, pyrrolocytosine, pyrrolopseudoisocytidine, 2-thiocytosine (s2C), 2-thio-5-methylcytosine, 4-thiopseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-

methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-
 zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxy-cytosine,
 2-methoxy-5-methyl-cytosine, 4-methoxy-pseudoisocytidine, 4-methoxy-1-methyl-
 pseudoisocytidine, lysidine (k2C), 5,2'-O-dimethyl-cytidine (m5Cm), N4-acetyl-2'-O-methyl-
 5 cytidine (ac4Cm), N4,2'-O-dimethyl-cytidine (m4Cm), 5-formyl-2'-O-methyl-cytidine (fSCm),
 N4,N4,2'-O-trimethyl-cytidine (m42Cm), 1-thio-cytosine, 5-hydroxy-cytosine, 5-(3-
 azidopropyl)-cytosine and 5-(2-azidoethyl)-cytosine.

Exemplary nucleobases and nucleosides having modified adenine may include, but are not
 limited to: 2-amino-purine, 2,6-diaminopurine, 2-amino-6-halo-purine (for example, 2-amino-6-
 10 chloro-purine), 6-halo-purine (for example, 6-chloro-purine), 2-amino-6-methyl-purine, 8-azido-
 adenine, 7-deaza-adenine, 7-deaza-8-aza-adenine, 7-deaza-2-amino-purine, 7-deaza-8-aza-2-
 amino-purine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyl-adenine
 (m1A), 2-methyl-adenine (m2A), N6-methyl-adenine (m6A), 2-methylthio-N6-methyl-adenine
 (ms2m6A), N6-isopentenyl-adenine (i6A), 2-methylthio-N6-isopentenyl-adenine (ms2i6A), N6-
 15 (cis-hydroxyisopentenyl)adenine (io6A), 2-methylthio-N6-(cis-hydroxyisopentenyl)adenine
 (ms2io6A), N6-glycylcarbamoyl-adenine (g6A), N6-threonylcarbamoyl-adenine (t6A), N6-
 methyl-N6-threonylcarbamoyl-adenine (m6t6A), 2-methylthio-N6-threonylcarbamoyl-adenine
 (ms2g6A), N6,N6-dimethyl-adenine (m62A), N6-hydroxynorvalylcarbamoyl-adenine (hn6A), 2-
 methylthio-N6-hydroxynorvalylcarbamoyl-adenine (ms2hn6A), N6-acetyl-adenine (ac6A), 7-
 20 methyl-adenine, 2-methylthio-adenine, 2-methoxy-adenine, N6,2'-O-dimethyl-adenosine
 (m6Am), N6,N6,2'-O-trimethyl-adenosine (m62Am), 1,2'-O-dimethyl-adenosine (m1Am), 2-
 amino-N6-methyl-purine, 1-thio-adenine, 8-azido-adenine, N6-(19-amino-pentaoxanonadecyl)-
 adenine, 2,8-dimethyl-adenine, N6-formyl-adenine and N6-hydroxymethyl-adenine.

Exemplary nucleobases and nucleosides having modified guanine may include, but are not
 25 limited to: inosine (I), 1-methyl-inosine (m1I), wyosine (imG), methylwyosine (mimG), 4-
 demethyl-wyosine (imG-14), isowyosine (imG2), wybutosine (yW), peroxywybutosine (o2yW),
 hydroxywybutosine (OHyW), undermodified hydroxywybutosine (OHyW*), 7-deaza-guanine,
 queuosine (Q), epoxyqueuosine (oQ), galactosyl-queuosine (galQ), mannosyl-queuosine (manQ),
 7-cyano-7-deaza-guanine (preQO), 7-aminomethyl-7-deaza-guanine (preQ1), archaeosine (G+),

7-deaza-8-aza-guanine, 6-thio-guanine, 6-thio-7-deaza-guanine, 6-thio-7-deaza-8-aza-guanine, 7-methyl-guanine (m7G), 6-thio-7-methyl-guanine, 7-methyl-inosine, 6-methoxy-guanine, 1-methyl-guanine (m1G), N2-methyl-guanine (m2G), N2,N2-dimethyl-guanine (m22G), N2,7-dimethyl-guanine (m2,7G), N2,N2,7-dimethyl-guanine (m2,2,7G), 8-oxo-guanine, 7-methyl-8-oxo-guanine, 1-methyl-6-thio-guanine, N2-methyl-6-thio-guanine, N2,N2-dimethyl-6-thio-guanine, N2-methyl-2'-O-methyl-guanosine (m2Gm), N2,N2-dimethyl-2'-O-methyl-guanosine (m22Gm), 1-methyl-2'-O-methyl-guanosine (m1Gm), N2,7-dimethyl-2'-O-methyl-guanosine (m2,7Gm), 2'-O-methyl-inosine (Im), 1,2'-O-dimethyl-inosine (m1Im), 1-thio-guanine and O-6-methyl-guanine.

10 In some embodiments, the proportion of modified nucleobases in the RNA of the present invention is 10%-100%, that is, the RNA of the present invention may be modified by replacing 10%-100% of the nucleobases therein with modified nucleobases.

In some embodiments, the RNA (for example, mRNA) of the present invention is modified by replacing one or more uracils with modified uracil. In some embodiments, the modified uracil
15 comprises 1-methylpseudouracil, pseudouracil, 5-methyl-uracil or a combination thereof. In an embodiment, the modified uracil comprises pseudouracil. In an embodiment, the modified uracil comprises 5-methyl-uracil. In an embodiment, the modified uracil comprises 1-methylpseudouracil.

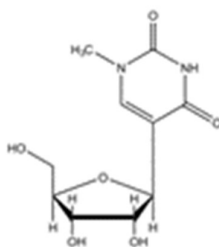
In an embodiment, RNA is modified by replacing at least one uracil with modified uracil. In
20 an embodiment, RNA is modified by replacing all uracils with modified uracil. In an embodiment, the proportion of modified uracil in RNA is 10%-100%, for example 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or 100%. In an embodiment, the proportion of modified uracil in RNA is 20%-100%. In an embodiment, 20%-100% of uracil in RNA is replaced with 1-methylpseudouracil. In a preferred embodiment, 100% of uracil in RNA is replaced with 1-
25 methylpseudouracil.

In some embodiments, the RNA (for example, mRNA) of the present invention is modified by replacing one or more uridines with modified uridine. In some embodiments, the modified uridine comprises 1-methylpseudouridine, pseudouridine, 5-methyl-uridine or a combination thereof. In an embodiment, the modified uridine comprises pseudouridine. In an embodiment, the modified

uridine comprises 5-methyl-uridine. In an embodiment, the modified uridine comprises 1-methylpseudouridine.

In an embodiment, RNA is modified by replacing at least one uridine with modified uridine. In an embodiment, RNA is modified by replacing all uridines with modified uridine. In an embodiment, the proportion of modified uridine in RNA is 10%-100%, for example 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or 100%. In an embodiment, the proportion of modified uridine in RNA is 20%-100%. In an embodiment, 20%-100% of uridine in RNA is replaced with 1-methylpseudouridine. In a preferred embodiment, 100% of uridine in RNA is replaced with 1-methylpseudouridine.

10 1-methyl-pseudouridine has the following structure:



In a specific embodiment, the RNA of the present invention comprises the nucleotide sequence as set forth in any one of SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66, and wherein 100% of uracil is replaced with 1-methylpseudouracil.

15 In a specific embodiment, the RNA of the present invention comprises the nucleotide sequence as set forth in any one of SEQ ID NO: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66, and wherein 100% of uridine is replaced with 1-methylpseudouridine.

In a specific embodiment, the mRNA of the present invention comprises the nucleotide sequence as set forth in any one of SEQ ID NO: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66, and wherein 100% of uracil is replaced with 1-methylpseudouracil.

In a specific embodiment, the mRNA of the present invention comprises the nucleotide sequence as set forth in any one of SEQ ID NO: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66, and wherein 100% of uridine is replaced with 1-methylpseudouridine.

DNA

In some embodiments, the nucleic acid or nucleic acid combination of the present invention is DNA. In some embodiments, the polynucleotide of the present invention is DNA. Such DNA may be, for example, a DNA template for in vitro transcription of the RNA of the present invention or DNA for expressing a polypeptide or polypeptide antigen in a host cell. The DNA may be double-
5 stranded, single-stranded, linear and circular DNA.

The DNA template may be provided in a suitable transcription vector. Generally, the DNA template may be a double-stranded complex comprising a nucleotide sequence identical to the coding sequence described herein (coding strand) and a nucleotide sequence complementary to the coding sequence described herein (template strand). As known to those skilled in the art, the DNA
10 template may comprise a promoter, DNA sequences corresponding to a 5'-UTR, a coding sequence, a 3'-UTR and an optionally present poly(A) sequence. The promoter may be a promoter usable by a suitable RNA polymerase (particularly a DNA-dependent RNA polymerase) known to those skilled in the art, including but not limited to promoters for SP6, T3 and T7 RNA polymerases. In some embodiments, the 5'-UTR, coding sequence, 3'-UTR and poly(A) sequence in the DNA
15 template are the corresponding sequences comprised in the RNA described herein or are complementary thereto. The DNA may be provided in a plasmid vector (for example, a circular plasmid vector).

In some embodiments, the DNA of the present invention comprises a coding sequence of a polypeptide described herein. In some embodiments, the DNA of the present invention comprises
20 a coding sequence of a polypeptide antigen described herein. In some embodiments, the DNA of the present invention comprises a coding sequence of a polypeptide combination described herein.

In some embodiments, the DNA of the present invention comprises a DNA sequence corresponding to a 5'-UTR sequence. In a preferred embodiment, the DNA sequence corresponding to the 5'-UTR sequence comprises the nucleotide sequence as set forth in SEQ ID
25 NO: 142 or 150. In some embodiments, the DNA of the present invention comprises a DNA sequence corresponding to a 3'-UTR sequence. In a preferred embodiment, the DNA sequence corresponding to the 3'-UTR sequence comprises the nucleotide sequence as set forth in SEQ ID NO: 143, 144 or 151. In some embodiments, the DNA of the present invention comprises DNA sequences corresponding to the 5'-UTR and 3'-UTR sequences. In an embodiment, the DNA

sequence corresponding to the 5'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 142 or 150, and the DNA sequence corresponding to the 3'-UTR sequence comprises the nucleotide sequence as set forth in SEQ ID NO: 143, 144 or 151. In a specific embodiment, the DNA sequence corresponding to the 5'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 142, and the DNA sequence corresponding to the 3'-UTR sequence comprises the nucleotide sequence as set forth in SEQ ID NO: 143 or 144. In a specific embodiment, the DNA sequence corresponding to the 5'-UTR comprises the nucleotide sequence as set forth in SEQ ID NO: 150, and the DNA sequence corresponding to the 3'-UTR sequence comprises the nucleotide sequence as set forth in SEQ ID NO: 151.

10 In some embodiments, the DNA of the present invention comprises a DNA sequence corresponding to a poly(A) sequence. In an embodiment, the DNA sequence corresponding to the poly(A) sequence comprises consecutive deoxyadenylates. In an embodiment, the DNA sequence corresponding to the poly(A) sequence may comprise at least 20, 30, 40, 50, 60, 70, 75, 80, 85, 95 or 100 and up to 120, 150, 180, 200 or 300 deoxyadenylates. In an embodiment, the DNA sequence corresponding to the poly(A) sequence comprises at least 50 nucleotides. In an embodiment, the DNA sequence corresponding to the poly(A) sequence comprises at least 80 nucleotides. In an embodiment, the DNA sequence corresponding to the poly(A) sequence comprises at least 100 nucleotides. In some embodiments, the DNA sequence corresponding to the poly(A) sequence comprises about 70, 80, 90, 100, 120 or 150 nucleotides. In a preferred embodiment, the DNA sequence corresponding to the poly(A) sequence comprises 121 nucleotides. In a preferred embodiment, the DNA sequence corresponding to the poly(A) sequence comprises the nucleotide sequence as set forth in SEQ ID NO: 146.

25 In some embodiments, the polynucleotide is DNA, and the polynucleotide comprises the nucleotide sequence as set forth in one of SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44 or a nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44. In some embodiments, the polynucleotide is DNA, and the polynucleotide comprises the nucleotide sequence as set forth in one of SEQ ID NOs: 67, 68, 69, 70, 71, 72, 73, 79, 80, 81, 82, 83, 84, 85, 86, 87 and 88, or a

nucleotide sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to nucleotide sequences as set forth in one of SEQ ID NOs: 67, 68, 69, 70, 71, 72, 73, 79, 80, 81, 82, 83, 84, 85, 86, 87 and 88.

In some embodiments, the polynucleotide is DNA, and the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44, or two or more nucleotide sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity to nucleotide sequences selected from SEQ ID NOs: 23, 24, 25, 26, 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43 and 44. In some embodiments, the polynucleotide is DNA, and the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 67, 68, 69, 70, 71, 72, 73, 79, 80, 81, 82, 83, 84, 85, 86, 87 and 88, or two or more nucleotide sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to nucleotide sequences selected from SEQ ID NOs: 67, 68, 69, 70, 71, 72, 73, 79, 80, 81, 82, 83, 84, 85, 86, 87 and 88.

In an embodiment, the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 23, the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 24, the third polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 25, the fourth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 26, the fifth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 27, and the sixth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 28. In an embodiment, the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 67, the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 68, and the third polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 69, and the fourth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 70, and the fifth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 71, and the sixth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 72.

In an embodiment, the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 24, the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 25, the third polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 26, the fourth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 27, the fifth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 28, and the sixth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 29. In an embodiment, the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 68, the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 69, and the third polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 70, and the fourth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 71, and the fifth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 72, and the sixth polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 73.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 36, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 37. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 80, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 81.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 39, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 40. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set

forth in SEQ ID NO: 83, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 84.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in
5 SEQ ID NO: 42, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 43. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 86, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 87.

10 In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 36, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 38. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set
15 forth in SEQ ID NO: 80, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 82.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 39, and the second polynucleotide comprises the nucleotide sequence as set forth in
20 SEQ ID NO: 41. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 83, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 85.

In an embodiment, the polynucleotide comprises a first polynucleotide and a second
25 polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 44, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 42. In an embodiment, the polynucleotide comprises a first polynucleotide and a second polynucleotide; wherein the first polynucleotide comprises the nucleotide sequence as set

forth in SEQ ID NO: 88, and the second polynucleotide comprises the nucleotide sequence as set forth in SEQ ID NO: 86.

Vector and host cell

In still another aspect, the present invention further provides an expression vector comprising
5 the nucleic acid or nucleic acid combination of the present invention. The expression vector may further comprise additional polynucleotide sequences, such as regulatory sequences and antibiotic resistance genes. The nucleic acid or nucleic acid combination of the present invention may be present in one or more expression vectors.

The present invention further provides a host cell comprising the nucleic acid or nucleic acid
10 combination or expression vector of the present invention. Various methods known in the art may be used to introduce the nucleic acid or nucleic acid combination or expression vector of the present invention into a suitable host cell. Such methods include, but are not limited to, liposome transfection, electroporation, viral transduction, calcium phosphate transfection and the like. In a preferred embodiment, the host cell is used for expressing the polypeptide, polypeptide antigen or
15 polypeptide combination of the present invention. Examples of host cells include, but are not limited to, prokaryotic cells (for example bacteria, such as *Escherichia coli*) and eukaryotic cells (for example yeast, insect cells and mammalian cells). Mammalian host cells suitable for expression include, but are not limited to, human cervical carcinoma cells (HeLa cells), human embryonic kidney cells (HEK cells, for example HEK293 cells), Chinese hamster ovary (CHO)
20 cells and other mammalian cells suitable for expression.

Composition

In one aspect, the nucleic acid or nucleic acid combination described herein is formulated for in vitro and in vivo delivery. In one aspect, the present invention provides a composition comprising the nucleic acid or nucleic acid combination of the present invention. In particular, in
25 some embodiments, the nucleic acid or nucleic acid combination is formulated as a lipid-containing composition. In some embodiments, the composition comprises a lipid encapsulating the nucleic acid or nucleic acid combination. In some embodiments, the lipid-containing composition forms lipid nanoparticles that enclose the nucleic acid or nucleic acid combination within a lipid shell. In some embodiments, the lipid shell protects the nucleic acid or nucleic acid

combination from degradation. In some embodiments, the lipid nanoparticle further facilitates transport of the enclosed nucleic acid or nucleic acid combination into intracellular compartments and/or tissues to exert an intended therapeutic function. In certain embodiments, when present in a lipid nanoparticle, the nucleic acid resists degradation by nucleases in an aqueous solution. Lipid nanoparticles comprising nucleic acids and methods for preparing the same are known in the art, such as those disclosed, for example, in U.S. Patent Publication No. 2004/0142025, U.S. Patent Publication No. 2007/0042031, PCT Publication No. WO 2017/004143, PCT Publication No. WO 2015/199952, PCT Publication No. WO 2013/016058, PCT Publication No. WO 2013/086373 and PCT Publication No. WO 2023/098842, the entire disclosure of each of which is incorporated herein by reference in its entirety for all purposes.

Nanoparticle compositions that can be used in conjunction with the present disclosure include, for example, lipid nanoparticles (LNPs), nanolipoprotein particles, liposomes, lipid vesicles and lipoplexes. In some embodiments, the nanoparticle composition is a vesicle comprising one or more lipid bilayers. In some embodiments, the nanoparticle composition comprises two or more concentric bilayers separated by aqueous compartments. The lipid bilayers may be functionalized and/or crosslinked to one another. The lipid bilayers may comprise one or more ligands, proteins or channels.

In some embodiments, the composition comprises at least a first lipid. In some embodiments, the composition comprises at least a first lipid, wherein the first lipid is a cationic lipid. In some embodiments, the composition comprises at least a first lipid (for example a cationic lipid), wherein the first lipid comprises one or more of the compounds of lipid series 01, 02, 03 and 04 as described herein, for example one or more of the compounds according to formulae (01-I), (01-II), (02-I), (02-II), (03-I), (04-I) and (04-III) (and sub-formulae thereof). In some embodiments, the first lipid comprises one or more of the compounds listed in Tables 01-1, 02-1, 03-1 and 04-1. In some embodiments, the composition further comprises a second lipid, wherein the second lipid is a polymer-conjugated lipid. In some embodiments, the second lipid comprises one or more compound of formula 05-I. In some embodiments, the composition is formulated as a lipid nanoparticle that encapsulates the nucleic acid or nucleic acid combination in a lipid.

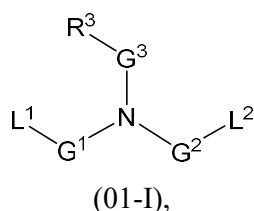
In some embodiments, the nanoparticle composition comprises a lipid component, wherein the lipid component comprises at least one lipid, such as one or more of the compounds of lipid series 01, 02, 03, 04 and 05 as described herein, for example one or more of the compounds according to formulae (01-I), (01-II), (02-I), (02-II), (03-I), (04-I), (04-III) and (05-I) (and sub-formulae thereof). For example, in some embodiments, the nanoparticle composition may comprise a lipid component comprising one of the compounds provided herein. The nanoparticle composition may further comprise one or more other lipid or non-lipid components, and exemplary other lipid or non-lipid components may be found, for example, in PCT Publication No. WO 2023/098842.

Cationic lipid

Cationic lipids include the following lipid series 01-04 (and sub-formulae thereof).

1) Lipid series 01

In an embodiment, the cationic lipid comprises a compound of formula (01-I):



or one or more of a pharmaceutically acceptable salt, prodrug or stereoisomer thereof, wherein:

G^1 and G^2 are each independently a bond, C_2 - C_{12} alkylene or C_2 - C_{12} alkenylene, wherein one or more $-CH_2-$ in the alkylene or alkenylene is optionally replaced by $-O-$;

L^1 is $-OC(=O)R^1$, $-C(=O)OR^1$, $-OC(=O)OR^1$, $-C(=O)R^1$, $-OR^1$, $-S(O)_xR^1$, $-S-SR^1$, $-C(=O)SR^1$, $-SC(=O)R^1$, $-NR^aC(=O)R^1$, $-C(=O)NR^bR^c$, $-NR^aC(=O)NR^bR^c$, $-OC(=O)NR^bR^c$, $-NR^aC(=O)OR^1$, $-SC(=S)R^1$, $-C(=S)SR^1$, $-C(=S)R^1$, $-CH(OH)R^1$, $-P(=O)(OR^b)(OR^c)$, $-(C_6-C_{10} \text{ arylene})-R^1$, $-(6\text{- to } 10\text{-membered heteroarylene})-R^1$ or R^1 ;

L^2 is $-OC(=O)R^2$, $-C(=O)OR^2$, $-OC(=O)OR^2$, $-C(=O)R^2$, $-OR^2$, $-S(O)_xR^2$, $-S-SR^2$, $-C(=O)SR^2$, $-SC(=O)R^2$, $-NR^dC(=O)R^2$, $-C(=O)NR^eR^f$, $-NR^dC(=O)NR^eR^f$, $-OC(=O)NR^eR^f$, $-NR^dC(=O)OR^2$, $-SC(=S)R^2$, $-C(=S)SR^2$, $-C(=S)R^2$, $-CH(OH)R^2$, $-P(=O)(OR^e)(OR^f)$, $-(C_6-C_{10} \text{ arylene})-R^2$, $-(6\text{- to } 10\text{-membered heteroarylene})-R^2$ or R^2 ;

R^1 and R^2 are each independently C_6 - C_{32} alkyl or C_6 - C_{32} alkenyl;

R^a , R^b , R^d and R^e are each independently H, C_1 - C_{24} alkyl or C_2 - C_{24} alkenyl;

R^c and R^f are each independently C_1 - C_{32} alkyl or C_2 - C_{32} alkenyl;

G^3 is C_2 - C_{24} alkylene, C_2 - C_{24} alkenylene, C_3 - C_8 cycloalkylene or C_3 - C_8 cycloalkenylene;

R^3 is $-N(R^4)R^5$;

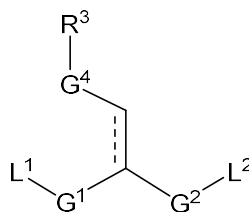
R^4 is C_3 - C_8 cycloalkyl, C_3 - C_8 cycloalkenyl, 4- to 8-membered heterocyclyl or C_6 - C_{10} aryl; or
5 a portion of R^4 , G^3 or G^3 , together with the nitrogen to which they are attached, forms a cyclic moiety;

R^5 is C_1 - C_{12} alkyl or C_3 - C_8 cycloalkyl; or R^4 and R^5 , together with the nitrogen to which they are attached, form a cyclic moiety;

X is 0, 1 or 2; and

10 each alkyl, alkenyl, cycloalkyl, cycloalkenyl, heterocyclyl, aryl, alkylene, alkenylene, cycloalkylene, cycloalkenylene, arylene, heteroarylene and cyclic moiety is independently optionally substituted.

In an embodiment, the cationic lipid comprises a compound of formula (01-II):



15 or one or more of a pharmaceutically acceptable salt, prodrug or stereoisomer thereof, wherein:

----- Is a single bond or a double bond;

G^1 and G^2 are each independently a bond, C_2 - C_{12} alkylene or C_2 - C_{12} alkenylene, wherein one or more $-CH_2-$ in the alkylene or alkenylene is optionally replaced by $-O-$;

20 L^1 is $-OC(=O)R^1$, $-C(=O)OR^1$, $-OC(=O)OR^1$, $-C(=O)R^1$, $-OR^1$, $-S(O)_xR^1$, $-S-SR^1$, $-C(=O)SR^1$, $-SC(=O)R^1$, $-NR^aC(=O)R^1$, $-C(=O)NR^bR^c$, $-NR^aC(=O)NR^bR^c$, $-OC(=O)NR^bR^c$, $-NR^aC(=O)OR^1$, $-SC(=S)R^1$, $-C(=S)SR^1$, $-C(=S)R^1$, $-CH(OH)R^1$, $-P(=O)(OR^b)(OR^c)$, $-(C_6-C_{10} \text{ arylene})-R^1$, $-(6- \text{ to } 10\text{-membered heteroarylene})-R^1$ or R^1 ;

L^2 is $-OC(=O)R^2$, $-C(=O)OR^2$, $-OC(=O)OR^2$, $-C(=O)R^2$, $-OR^2$, $-S(O)_xR^2$, $-S-SR^2$, $-C(=O)SR^2$,
25 $-SC(=O)R^2$, $-NR^dC(=O)R^2$, $-C(=O)NR^eR^f$, $-NR^dC(=O)NR^eR^f$, $-OC(=O)NR^eR^f$, $-NR^dC(=O)OR^2$, -

SC(=S)R², -C(=S)SR², -C(=S)R², -CH(OH)R², -P(=O)(OR^e)(OR^f), -(C₆-C₁₀ arylene)-R², -(6- to 10-membered heteroarylene)-R² or R²;

R¹ and R² are each independently C₆-C₃₂ alkyl or C₆-C₃₂ alkenyl;

R^a, R^b, R^d and R^e are each independently H, C₁-C₂₄ alkyl or C₂-C₂₄ alkenyl;

5 R^c and R^f are each independently C₁-C₃₂ alkyl or C₂-C₃₂ alkenyl;

G⁴ is a bond, C₁-C₂₃ alkylene, C₂-C₂₃ alkenylene, C₃-C₈ cycloalkylene or C₃-C₈ cycloalkenylene;

R³ is -N(R⁴)R⁵;

10 R⁴ is C₁-C₁₂ alkyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl, 4- to 8-membered heterocyclyl or C₆-C₁₀ aryl; or a portion of R⁴, G³ or G³, together with the nitrogen to which they are attached, forms a cyclic moiety;

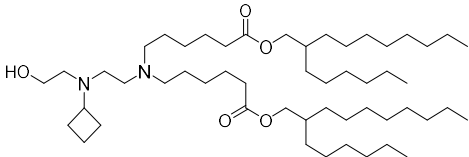
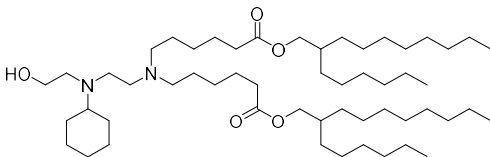
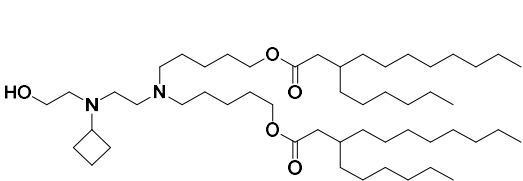
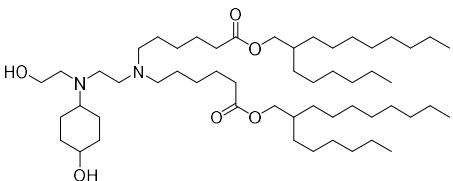
R⁵ is C₁-C₁₂ alkyl or C₃-C₈ cycloalkyl; or R⁴ and R⁵, together with the nitrogen to which they are attached, form a cyclic moiety;

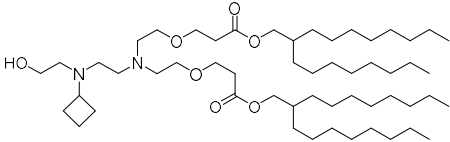
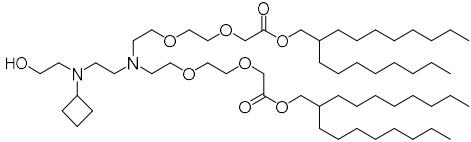
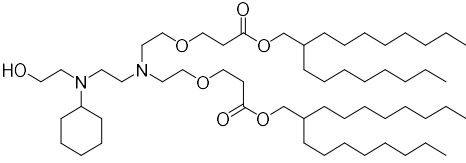
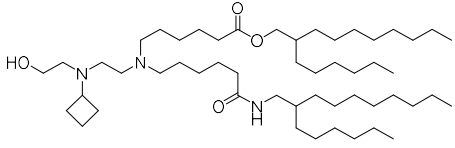
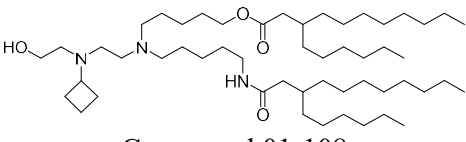
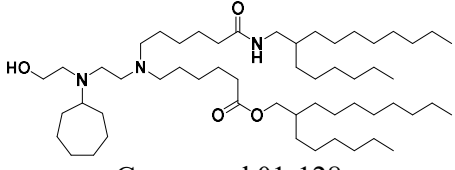
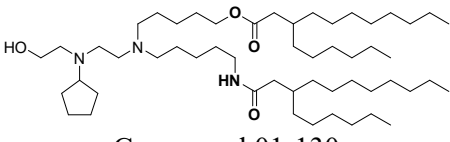
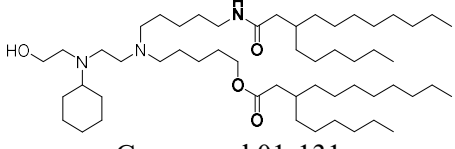
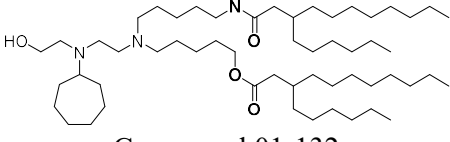
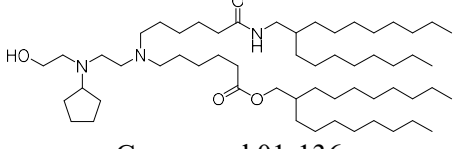
X is 0, 1 or 2; and

15 each alkyl, alkenyl, cycloalkyl, cycloalkenyl, heterocyclyl, aryl, alkylene, alkenylene, cycloalkylene, cycloalkenylene, arylene, heteroarylene and cyclic moiety is independently optionally substituted.

In an embodiment, the cationic lipid comprises a compound in Table 01-1, or one or more of a pharmaceutically acceptable salt, prodrug or stereoisomer thereof.

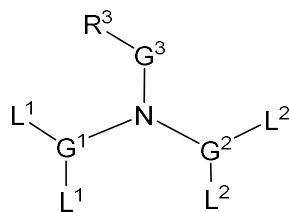
20 **Table 01-1.**

 Compound 01-1	 Compound 01-2
 Compound 01-20	 Compound 01-21

 Compound 01-102	 Compound 01-103
 Compound 01-104	 Compound 01-106
 Compound 01-108	 Compound 01-128
 Compound 01-130	 Compound 01-131
 Compound 01-132	 Compound 01-136

2) Lipid series 02

In an embodiment, the cationic lipid comprises a compound of formula (02-I):



(02-I)

5 or one or more of a pharmaceutically acceptable salt, prodrug or stereoisomer thereof, wherein:

G^1 and G^2 are each independently C_2 - C_{12} alkylene or C_2 - C_{12} alkenylene, wherein in G^1 and G^2 , one or more $-CH_2-$ is optionally replaced by $-O-$, $-C(=O)O-$ or $-OC(=O)-$;

each L^1 is independently $-OC(=O)R^1$, $-C(=O)OR^1$, $-OC(=O)OR^1$, $-C(=O)R^1$, $-OR^1$, $-S(O)_xR^1$, $-S-SR^1$, $-C(=O)SR^1$, $-SC(=O)R^1$, $-NR^aC(=O)R^1$, $-C(=O)NR^bR^c$, $-NR^aC(=O)NR^bR^c$, -

OC(=O)NR^bR^c, -NR^aC(=O)OR¹, -SC(=S)R¹, -C(=S)SR¹, -C(=S)R¹, -CH(OH)R¹, -
P(=O)(OR^b)(OR^c), -NR^aP(=O)(OR^b)(OR^c);

each L² is independently -OC(=O)R², -C(=O)OR², -OC(=O)OR², -C(=O)R², -OR², -S(O)_xR²,
-S-SR², -C(=O)SR², -SC(=O)R², -NR^dC(=O)R², -C(=O)NR^eR^f, -NR^dC(=O)NR^eR^f, -
5 OC(=O)NR^eR^f, -NR^dC(=O)OR², -SC(=S)R², -C(=S)SR², -C(=S)R², -CH(OH)R², -
P(=O)(OR^c)(OR^f), -NR^dP(=O)(OR^e)(OR^f);

R¹ and R² are each independently C₆-C₂₄ alkyl or C₆-C₂₄ alkenyl;

R^a, R^b, R^d and R^e are each independently H, C₁-C₂₄ alkyl or C₂-C₂₄ alkenyl;

R^c and R^f are each independently C₁-C₂₄ alkyl or C₂-C₂₄ alkenyl;

10 G³ is C₂-C₁₂ alkylene or C₂-C₁₂ alkenylene, wherein part or all of the alkylene or alkenylene is
optionally replaced by C₃-C₈ cycloalkylene or C₃-C₈ cycloalkenylene;

R³ is -N(R⁴)R⁵, -OR⁶ or -SR⁶;

R⁴ is C₁-C₁₂ alkyl, C₂-C₁₂ alkenyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl, C₆-C₁₀ aryl or 4- to
8-membered heterocycloalkyl;

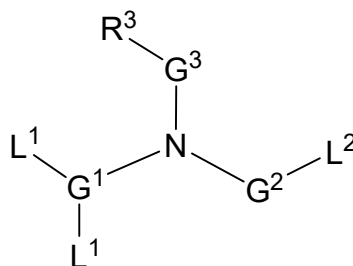
15 R⁵ is H, C₁-C₁₂ alkyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl, C₆-C₁₀ aryl or 4- to 8-membered
heterocycloalkyl;

R⁶ is hydrogen, C₁-C₁₂ alkyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl or C₆-C₁₀ aryl;

X is 0, 1 or 2; and

each alkyl, alkenyl, cycloalkyl, cycloalkenyl, heterocycloalkyl, aryl, alkylene, alkenylene,
20 cycloalkylene and cycloalkenylene is independently optionally substituted.

In an embodiment, the cationic lipid is a compound of formula (02-II):



(02-II),

or a pharmaceutically acceptable salt, prodrug or stereoisomer thereof, wherein:

G¹ and G² are each independently C₂-C₁₂ alkylene or C₂-C₁₂ alkenylene, wherein in G¹ and G², one or more -CH₂- is optionally replaced by -O-, -C(=O)O- or -OC(=O)-;

each L¹ is independently -OC(=O)R¹, -C(=O)OR¹, -OC(=O)OR¹, -C(=O)R¹, -OR¹, -S(O)_xR¹, -S-SR¹, -C(=O)SR¹, -SC(=O)R¹, -NR^aC(=O)R¹, -C(=O)NR^bR^c, -NR^aC(=O)NR^bR^c, -OC(=O)NR^bR^c, -NR^aC(=O)OR¹, -SC(=S)R¹, -C(=S)SR¹, -C(=S)R¹, -CH(OH)R¹, -P(=O)(OR^b)(OR^c), -NR^aP(=O)(OR^b)(OR^c);

each L² is independently -OC(=O)R², -C(=O)OR², -OC(=O)OR², -C(=O)R², -OR², -S(O)_xR², -S-SR², -C(=O)SR², -SC(=O)R², -NR^dC(=O)R², -C(=O)NR^eR^f, -NR^dC(=O)NR^eR^f, -OC(=O)NR^eR^f, -NR^dC(=O)OR², -SC(=S)R², -C(=S)SR², -C(=S)R², -CH(OH)R², -P(=O)(OR^e)(OR^f), -NR^dP(=O)(OR^e)(OR^f);

R¹ and R² are each independently C₆-C₂₄ alkyl or C₆-C₂₄ alkenyl;

R^a, R^b, R^d and R^e are each independently H, C₁-C₂₄ alkyl or C₂-C₂₄ alkenyl;

R^c and R^f are each independently C₁-C₂₄ alkyl or C₂-C₂₄ alkenyl;

G³ is C₂-C₁₂ alkylene or C₂-C₁₂ alkenylene, wherein part or all of the alkylene or alkenylene is optionally replaced by C₃-C₈ cycloalkylene or C₃-C₈ cycloalkenylene;

R³ is -N(R⁴)R⁵, -OR⁶ or -SR⁶;

R⁴ is C₁-C₁₂ alkyl, C₂-C₁₂ alkenyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl, C₆-C₁₀ aryl or 4- to 8-membered heterocycloalkyl;

R⁵ is H, C₁-C₁₂ alkyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl, C₆-C₁₀ aryl or 4- to 8-membered heterocycloalkyl;

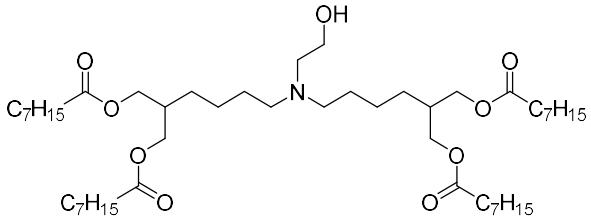
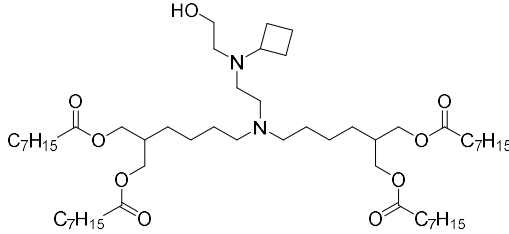
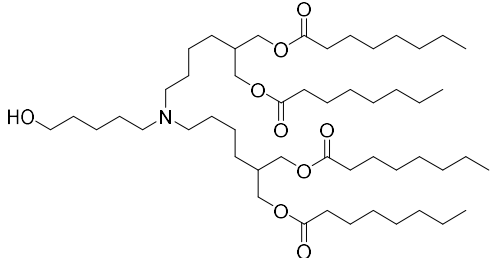
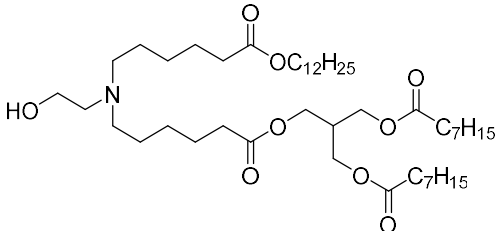
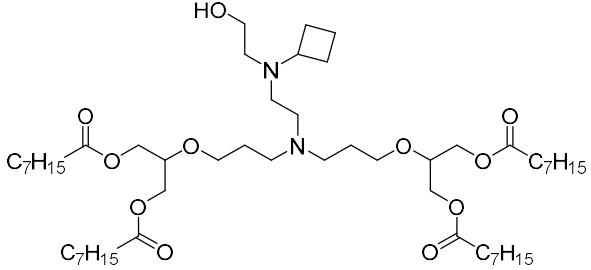
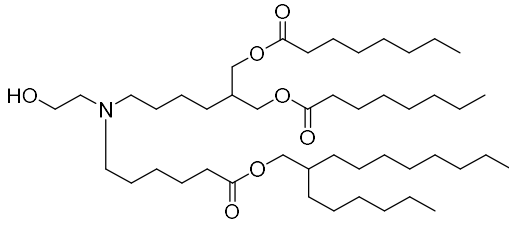
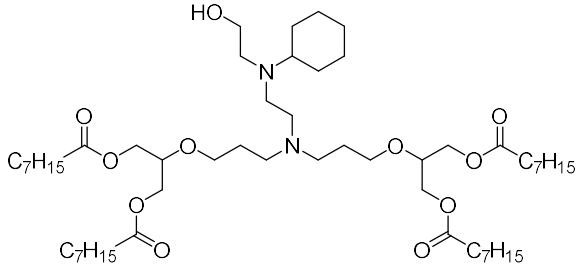
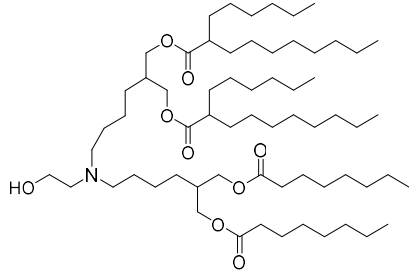
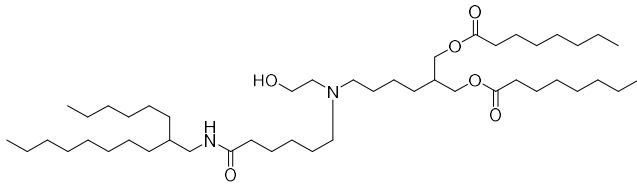
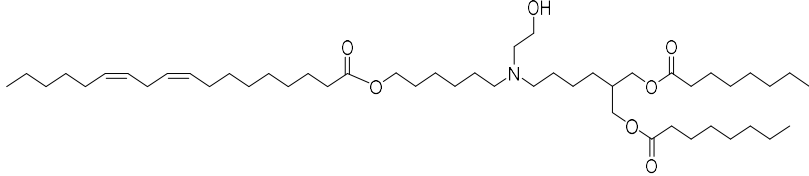
R⁶ is hydrogen, C₁-C₁₂ alkyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl or C₆-C₁₀ aryl;

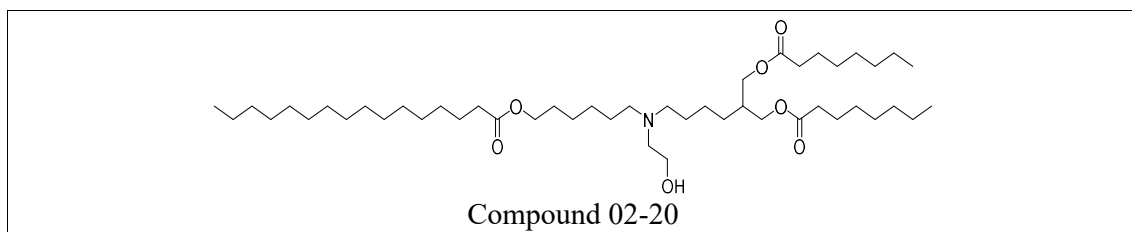
X is 0, 1 or 2; and

each alkyl, alkenyl, cycloalkyl, cycloalkenyl, heterocycloalkyl, aryl, alkylene, alkenylene, cycloalkylene and cycloalkenylene is independently optionally substituted.

In an embodiment, the cationic lipid comprises a compound in Table 02-1, or one or more of a pharmaceutically acceptable salt, prodrug or stereoisomer thereof.

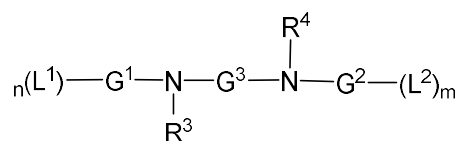
Table 02-1.

 Compound 02-1	 Compound 02-2
 Compound 02-3	 Compound 02-4
 Compound 02-9	 Compound 02-10
 Compound 02-14	 Compound 02-12
 Compound 02-11	
 Compound 02-22	



3) Lipid series 03

In an embodiment, the cationic lipid comprises a compound of formula (03-I):



(03-I),

5 or a pharmaceutically acceptable salt, prodrug or stereoisomer thereof, wherein:

G^1 and G^2 are each independently a bond, C_2 - C_{12} alkylene or C_2 - C_{12} alkenylene, wherein in G^1 and G^2 , one or more $-CH_2-$ is optionally replaced by $-O-$;

each L^1 is independently $-OC(=O)R^1$, $-C(=O)OR^1$, $-OC(=O)OR^1$, $-C(=O)R^1$, $-OR^1$, $-S(O)_xR^1$, $-S-SR^1$, $-C(=O)SR^1$, $-SC(=O)R^1$, $-NR^aC(=O)R^1$, $-C(=O)NR^bR^c$, $-NR^aC(=O)NR^bR^c$, $-OC(=O)NR^bR^c$, $-NR^aC(=O)OR^1$, $-SC(=S)R^1$, $-C(=S)SR^1$, $-C(=S)R^1$, $-CH(OH)R^1$, $-P(=O)(OR^b)(OR^c)$, $-NR^aP(=O)(OR^b)(OR^c)$, $-(C_6-C_{10} \text{ arylene})-R^1$, $-(6- \text{ to } 10\text{-membered heteroarylene})-R^1$, $-(4- \text{ to } 8\text{-membered heterocyclylene})-R^1$ or R^1 ;

each L^2 is independently $-OC(=O)R^2$, $-C(=O)OR^2$, $-OC(=O)OR^2$, $-C(=O)R^2$, $-OR^2$, $-S(O)_xR^2$, $-S-SR^2$, $-C(=O)SR^2$, $-SC(=O)R^2$, $-NR^dC(=O)R^2$, $-C(=O)NR^eR^f$, $-NR^dC(=O)NR^eR^f$, $-OC(=O)NR^eR^f$, $-NR^dC(=O)OR^2$, $-SC(=S)R^2$, $-C(=S)SR^2$, $-C(=S)R^2$, $-CH(OH)R^2$, $-P(=O)(OR^e)(OR^f)$, $-NR^dP(=O)(OR^e)(OR^f)$, $-(C_6-C_{10} \text{ arylene})-R^2$, $-(6- \text{ to } 10\text{-membered heteroarylene})-R^2$, $-(4- \text{ to } 8\text{-membered heterocyclylene})-R^2$ or R^2 ;

R^1 and R^2 are each independently C_6 - C_{24} alkyl or C_6 - C_{24} alkenyl;

R^a , R^b , R^d and R^e are each independently H, C_1 - C_{24} alkyl or C_2 - C_{24} alkenyl;

20 R^c and R^f are each independently C_1 - C_{24} alkyl or C_2 - C_{24} alkenyl;

G^3 is C_2 - C_{12} alkylene or C_2 - C_{12} alkenylene, wherein part or all of the alkylene or alkenylene is optionally replaced by C_3 - C_8 cycloalkylene, C_3 - C_8 cycloalkenylene, C_3 - C_8 cycloalkynylene, 4- to 8-membered heterocyclylene, C_6 - C_{10} arylene or 5- to 10-membered heteroarylene;

R^3 is hydrogen, C_1 - C_{12} alkyl, C_2 - C_{12} alkenyl, C_2 - C_{12} alkynyl, C_3 - C_8 cycloalkyl, C_3 - C_8 cycloalkenyl, C_3 - C_8 cycloalkynyl, 4- to 8-membered heterocyclyl, C_6 - C_{10} aryl or 5- to 10-membered heteroaryl; or a portion of R^3 , G^1 or G^1 , together with the nitrogen to which they are attached, forms a cyclic moiety; or a portion of R^3 , G^3 or G^3 , together with the nitrogen to which they are attached, forms a cyclic moiety;

R^4 is C_1 - C_{12} alkyl or C_3 - C_8 cycloalkyl;

X is 0, 1 or 2;

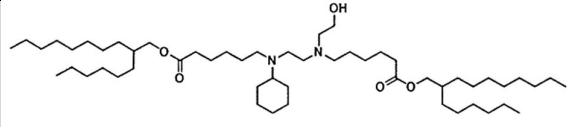
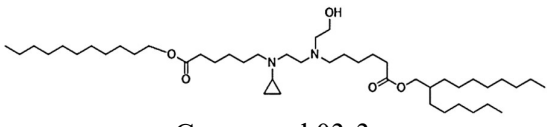
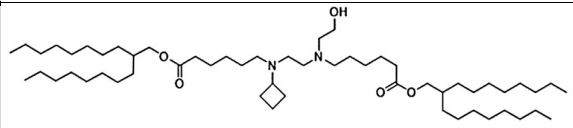
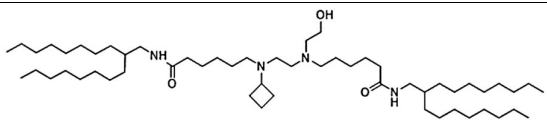
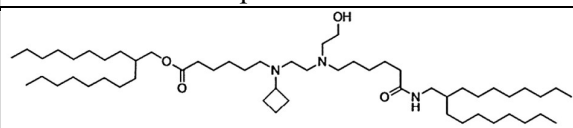
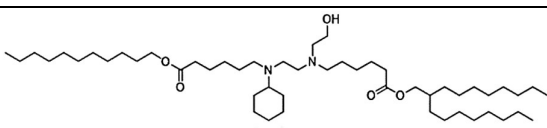
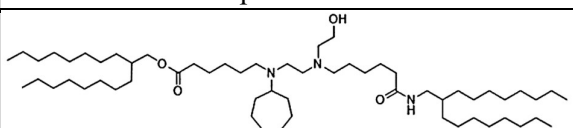
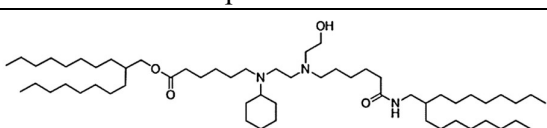
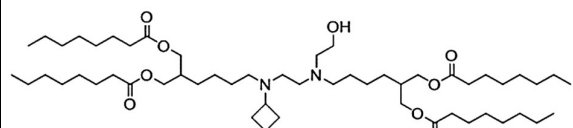
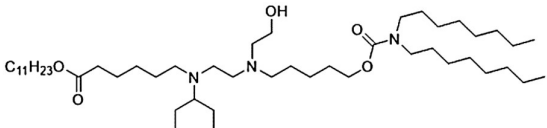
N is 1 or 2;

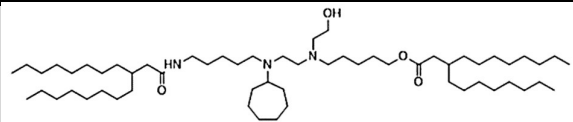
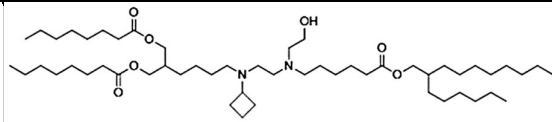
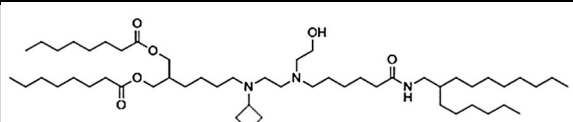
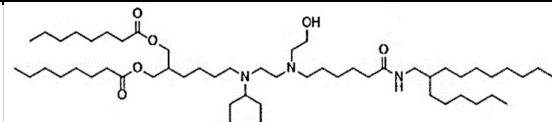
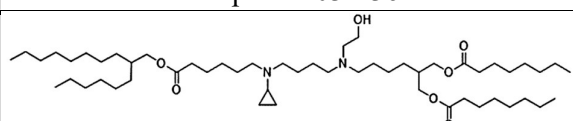
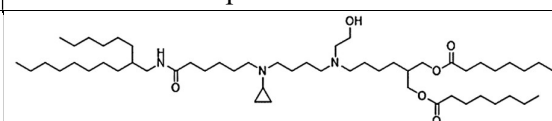
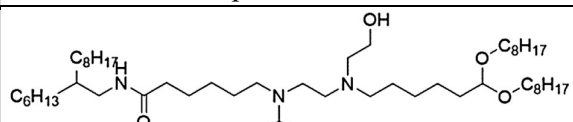
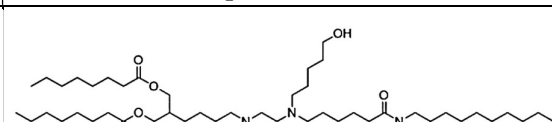
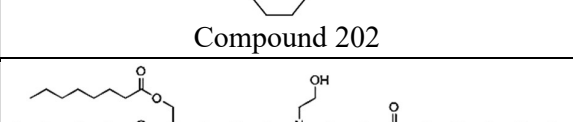
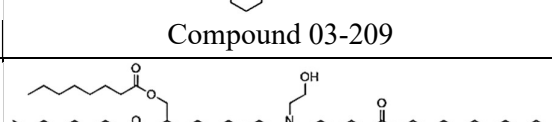
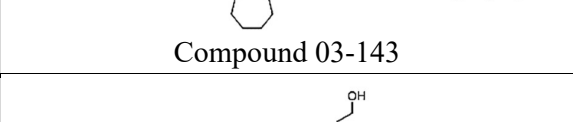
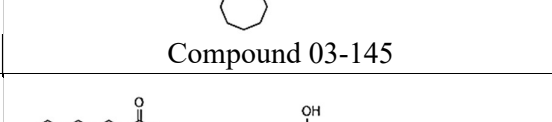
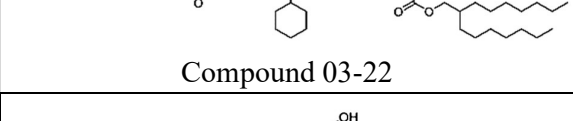
M is 1 or 2; and

each alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, cycloalkynyl, heterocyclyl, aryl, heteroaryl, alkylene, alkenylene, cycloalkylene, cycloalkenylene, cycloalkynylene, heterocyclylene, arylene, heteroarylene and cyclic moiety is independently optionally substituted.

In an embodiment, the cationic lipid comprises a compound in Table 03-1, or one or more of a pharmaceutically acceptable salt, prodrug or stereoisomer thereof.

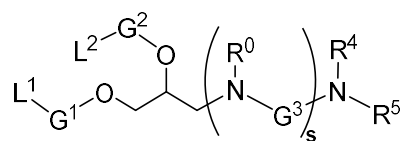
Table 03-1.

4) Lipid series 04

In an embodiment, the cationic lipid comprises a compound of formula (04-I):



(04-I),

5 or one or more of a pharmaceutically acceptable salt, prodrug or stereoisomer thereof, wherein:

G¹ and G² are each independently a bond, C₂-C₁₂ alkylene or C₂-C₁₂ alkenylene;

L^1 is $-OC(=O)R^1$, $-C(=O)OR^1$, $-OC(=O)OR^1$, $-C(=O)R^1$, $-OR^1$, $-S(O)_xR^1$, $-S-SR^1$, $-C(=O)SR^1$, $-SC(=O)R^1$, $-NR^aC(=O)R^1$, $-C(=O)NR^bR^c$, $-NR^aC(=O)NR^bR^c$, $-OC(=O)NR^bR^c$, $-NR^aC(=O)OR^1$, $-SC(=S)R^1$, $-C(=S)SR^1$, $-C(=S)R^1$, $-CH(OH)R^1$, $-P(=O)(OR^b)(OR^c)$, $-(C_6-C_{10} \text{ arylene})-R^1$, $-(6\text{- to } 10\text{-membered heteroarylene})-R^1$ or R^1 ;

5 L^2 is $-OC(=O)R^2$, $-C(=O)OR^2$, $-OC(=O)OR^2$, $-C(=O)R^2$, $-OR^2$, $-S(O)_xR^2$, $-S-SR^2$, $-C(=O)SR^2$, $-SC(=O)R^2$, $-NR^dC(=O)R^2$, $-C(=O)NR^eR^f$, $-NR^dC(=O)NR^eR^f$, $-OC(=O)NR^eR^f$, $-NR^dC(=O)OR^2$, $-SC(=S)R^2$, $-C(=S)SR^2$, $-C(=S)R^2$, $-CH(OH)R^2$, $-P(=O)(OR^e)(OR^f)$, $-(C_6-C_{10} \text{ arylene})-R^2$, $-(6\text{- to } 10\text{-membered heteroarylene})-R^2$ or R^2 ;

R^1 and R^2 are each independently C_5-C_{32} alkyl or C_5-C_{32} alkenyl;

10 R^a , R^b , R^d and R^e are each independently H, C_1-C_{24} alkyl or C_2-C_{24} alkenyl;

R^c and R^f are each independently C_1-C_{32} alkyl or C_2-C_{32} alkenyl;

R^0 is C_1-C_{12} alkyl, C_2-C_{12} alkenyl, C_3-C_8 cycloalkyl, C_3-C_8 cycloalkenyl, C_6-C_{10} aryl or 4- to 8-membered heterocycloalkyl;

G^3 is C_2-C_{12} alkylene or C_2-C_{12} alkenylene;

15 R^4 is C_1-C_{12} alkyl, C_2-C_{12} alkenyl, C_3-C_8 cycloalkyl, C_3-C_8 cycloalkenyl, C_6-C_{10} aryl or 4- to 8-membered heterocycloalkyl;

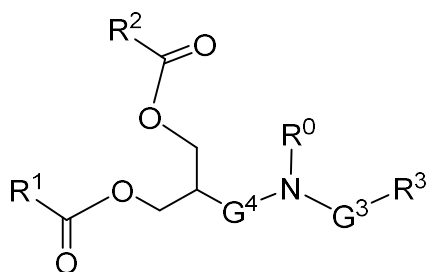
R^5 is C_1-C_{12} alkyl, C_3-C_8 cycloalkyl, C_3-C_8 cycloalkenyl, C_6-C_{10} aryl or 4- to 8-membered heterocycloalkyl;

X is 0, 1 or 2;

20 S is 0 or 1; and

each alkyl, alkenyl, cycloalkyl, cycloalkenyl, heterocycloalkyl, aryl, alkylene, alkenylene, arylene and heteroarylene is independently optionally substituted.

In an embodiment, the cationic lipid is a compound of formula (04-III):



(04-III),

or a pharmaceutically acceptable salt, prodrug or stereoisomer thereof, wherein:

R¹ and R² are each independently C₅-C₃₂ alkyl or C₅-C₃₂ alkenyl;

R⁰ is C₁-C₁₂ alkyl, C₂-C₁₂ alkenyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl, C₆-C₁₀ aryl or 4- to 8-membered heterocycloalkyl;

5 G³ is C₂-C₁₂ alkylene or C₂-C₁₂ alkenylene;

G⁴ is C₂-C₁₂ alkylene or C₂-C₁₂ alkenylene;

R³ is -N(R⁴)R⁵ or -OR⁶;

R⁴ is C₁-C₁₂ alkyl, C₂-C₁₂ alkenyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl, C₆-C₁₀ aryl or 4- to 8-membered heterocycloalkyl;

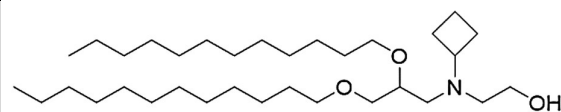
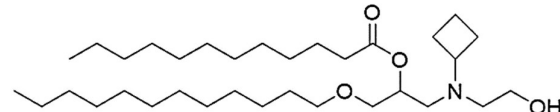
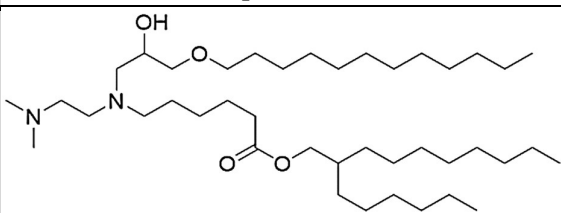
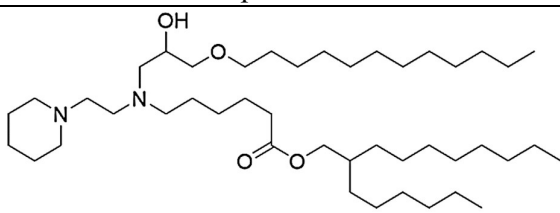
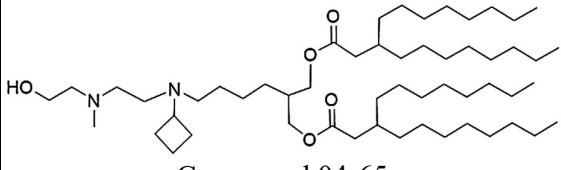
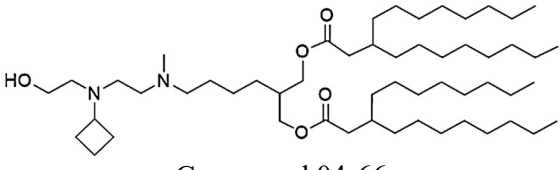
10 R⁵ is C₁-C₁₂ alkyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl, C₆-C₁₀ aryl or 4- to 8-membered heterocycloalkyl; or R⁴ and R⁵, together with the nitrogen to which they are attached, form a cyclic moiety;

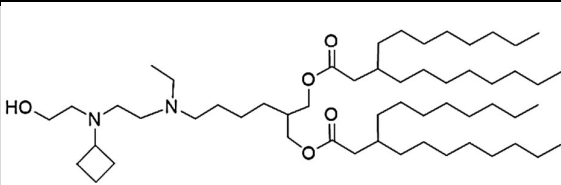
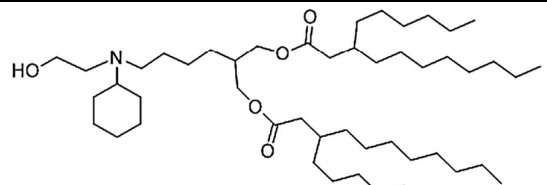
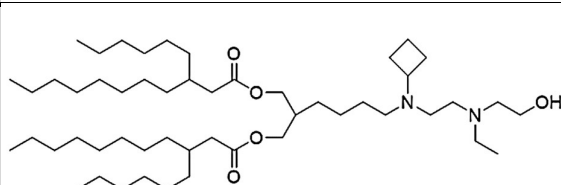
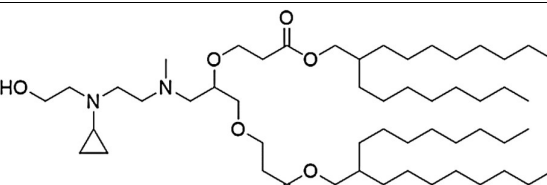
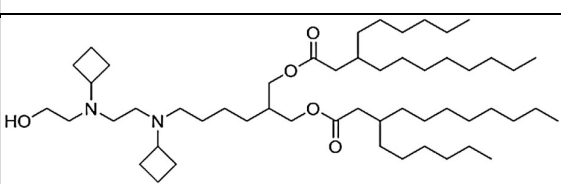
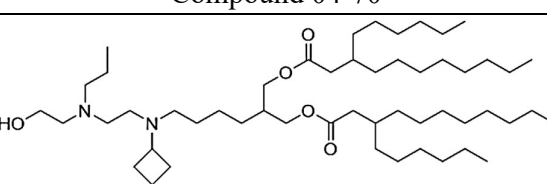
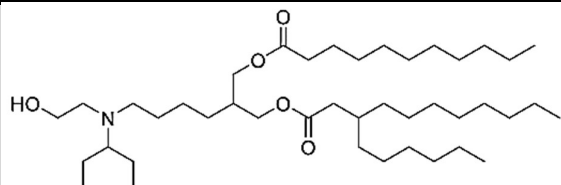
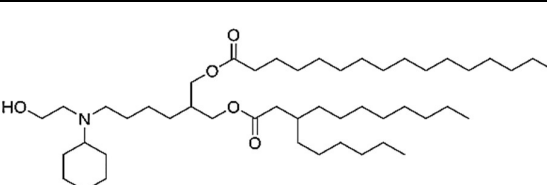
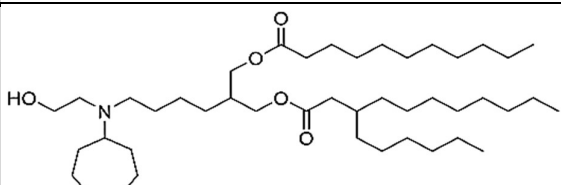
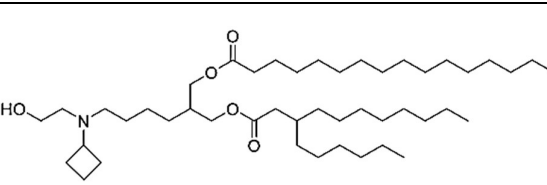
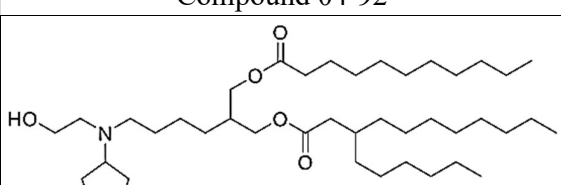
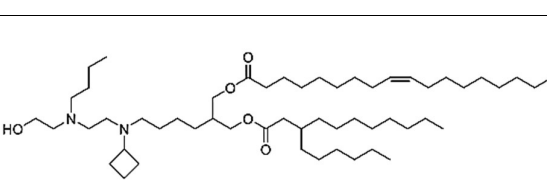
R⁶ is hydrogen, C₁-C₁₂ alkyl, C₃-C₈ cycloalkyl, C₃-C₈ cycloalkenyl or C₆-C₁₀ aryl; and

each alkyl, alkenyl, cycloalkyl, cycloalkenyl, heterocycloalkyl, aryl, alkylene, alkenylene and
15 cyclic moiety is independently optionally substituted.

In an embodiment, the cationic lipid comprises a compound in Table 04-1, or a pharmaceutically acceptable salt, prodrug or stereoisomer thereof.

Table 04-1.

 Compound 04-1	 Compound 04-2
 Compound 04-7	 Compound 04-8
 Compound 04-65	 Compound 04-66

 Compound 04-67	 Compound 04-68
 Compound 04-69	 Compound 04-70
 Compound 04-79	 Compound 04-80
 Compound 04-85	 Compound 04-86
 Compound 04-92	 Compound 04-98
 Compound 04-84	 Compound 04-139

In a preferred embodiment, the cationic lipid comprises compound 01-1.

It is to be understood that any embodiment of the compounds provided herein as set forth above, and any specific substituent and/or variable of the compounds provided herein as set forth above, may be independently combined with other embodiments and/or substituents and/or variables of the compounds to form embodiments not specifically set forth above. Furthermore, where a list of

substituents and/or variables is recited for any particular group or variable, it is to be understood that each individual substituent and/or variable may be deleted from the particular embodiment and/or technical solution, and that the remaining list of substituents and/or variables will be deemed to be within the scope of the embodiments provided herein.

5 It is to be understood that in the present specification, combinations of substituents and/or variables of the depicted formulas are permissible only when such contributions result in stable compounds.

Other ionizable lipids

As described herein, in some embodiments, the nanoparticle composition provided herein
10 comprises, in addition to one or more lipids of series 01, 02, 03 and 04, for example one or more lipids according to formula (01-I), (01-II), (02-I), (02-II), (03-I), (04-I) and (04-III) (and subformulas thereof), one or more charged or ionizable lipids. Without being bound by theory, it is contemplated that certain charged or zwitterionic lipid components of the nanoparticle composition are similar to lipid components in cell membranes, thereby improving cellular uptake
15 of the nanoparticles. Exemplary other charged lipids, ionizable lipids or cationic lipids that may form part of the nanoparticle composition of the present invention can be found, for example, in PCT Publication No. WO 2023/098842, the contents of which are incorporated herein by reference in their entirety.

In addition, in some embodiments, the charged or ionizable lipid that may form part of the
20 nanoparticle composition of the present invention is a lipid comprising a cyclic amine group. Additional cationic lipids suitable for the formulations and methods disclosed herein include those cationic lipids described in WO2015199952A1, WO2016176330A1 and WO2015011633A1, the entire contents of which are incorporated herein by reference in their entirety. In addition, in some embodiments, the charged or ionizable lipid that may form part of the nanoparticle composition of
25 the present invention is a lipid comprising a cyclic amine group. Additional cationic lipids suitable for the formulations and methods disclosed herein include those cationic lipids described in WO2015199952A1, WO2016176330A1 and WO2015011633A1, the entire contents of which are incorporated herein by reference in their entirety.

Polymer-conjugated lipid

In some embodiments, the lipid component of the nanoparticle composition may comprise one or more polymer-conjugated lipids, such as PEGylated lipid (PEG lipid). Without being bound by theory, it is contemplated that the polymer-conjugated lipid component in the nanoparticle composition may improve colloidal stability and/or reduce protein adsorption of the nanoparticles.

5 Exemplary polymer-conjugated lipids that may be used in conjunction with the present disclosure include, but are not limited to, PEG-modified phosphatidylethanolamine, PEG-modified phosphatidic acid, PEG-modified ceramide, PEG-modified dialkylamine, PEG-modified diacylglycerol, PEG-modified dialkylglycerol and mixtures thereof. For example, the PEG lipid may be PEG-c-DOMG, PEG-DMG, PEG-DLPE, PEG-DMPE, PEG-DPPC, PEG-DSPE,
10 ceramide-PEG2000 or Chol-PEG2000.

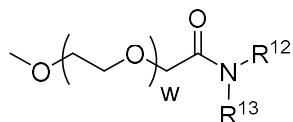
In an embodiment, the polymer-conjugated lipid is a PEGylated lipid. For example, some embodiments include PEGylated diacylglycerol (PEG-DAG), such as 1-(monomethoxy-polyethylene glycol)-2,3-dimyristoylglycerol (PEG-DMG); PEGylated phosphatidylethanolamine (PEG-PE); PEG succinate diacylglycerol (PEG-S-DAG), such as 4-O-(2',3'-
15 di(tetradecanoyloxy)propyl-1-O-(ω -methoxy(polyethoxy)ethyl) succinate (PEG-S-DMG); PEGylated ceramide (PEG-cer); or PEG dialkoxypropyl carbamate, such as ω -methoxy(polyethoxy)ethyl-N-(2,3-di(tetradecyloxy)propyl) carbamate or 2,3-di(tetradecyloxy)propyl-N-(ω -methoxy)(polyethoxy)ethyl carbamate.

In some embodiments, the polymer-conjugated lipid (preferably, PEGylated lipid) is present at
20 a concentration in the range of 0.5 mol% to 5 mol%. In an embodiment, the polymer-conjugated lipid is present at a concentration in the range of 1.0 mol% to 2.5 mol%. In an embodiment, the polymer-conjugated lipid is present at a concentration in the range of about 1.2 mol% to 2 mol%. In an embodiment, the polymer-conjugated lipid is present at a concentration of about 1.7 mol%. In an embodiment, the polymer-conjugated lipid is present at a concentration of about 1.5 mol%.

25 In an embodiment, the molar ratio of the cationic lipid to the polymer-conjugated lipid is in the range of about 35:1 to about 25:1. In an embodiment, the molar ratio of the cationic lipid to the polymer-conjugated lipid is in the range of about 100:1 to about 20:1.

In an embodiment, the molar ratio of the cationic lipid to the polymer-conjugated lipid is in the range of about 35:1 to about 25:1. In an embodiment, the molar ratio of the cationic lipid to the polymer-conjugated lipid is in the range of about 100:1 to about 20:1.

In an embodiment, the PEGylated lipid has the following formula:



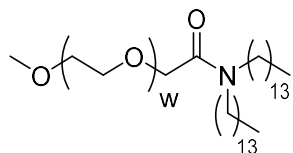
or a pharmaceutically acceptable salt, tautomer or stereoisomer thereof, wherein:

R^{12} and R^{13} are each independently a linear or branched, saturated or unsaturated alkyl chain containing 10 to 30 carbon atoms, wherein the alkyl chain is optionally interrupted by one or more ester bonds; and

w has an average value in the range of 30 to 60.

In an embodiment, R^{12} and R^{13} are each independently a linear saturated alkyl chain containing 12 to 16 carbon atoms. In other embodiments, the average value of w is in the range of 42 to 55, for example, the average value of w is 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54 or 55. In some specific embodiments, the average value of w is about 49.

In an embodiment, the PEGylated lipid has the following formula:

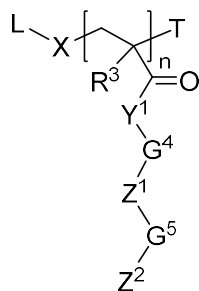


wherein the average value of w is about 49.

The polymer-conjugated lipids also include one or more of the following compounds according to lipid series 05 (and subformulas thereof).

5) Lipid series 05

In an embodiment, provided herein is a compound of formula (05-I):



(05-I),

or one or more of a pharmaceutically acceptable salt or stereoisomer thereof, wherein:

L is a lipid;

5 X is a linker;

each R^3 is independently H or $\text{C}_1\text{-C}_6$ alkyl;

each Y^1 is independently a bond, O, S or NR^a ;

each G^4 is independently a bond or $\text{C}_1\text{-C}_{12}$ alkylene, wherein one or more $-\text{CH}_2-$ are each independently optionally replaced by $-\text{O}-$, $-\text{S}-$ or $-\text{NR}^a-$;

10 each G^5 is independently a bond or $\text{C}_1\text{-C}_{12}$ alkylene, wherein one or more $-\text{CH}_2-$ are each independently optionally replaced by $-\text{O}-$, $-\text{S}-$ or $-\text{NR}^a-$;

each R^a is independently H, $\text{C}_1\text{-C}_{12}$ alkyl or $\text{C}_2\text{-C}_{12}$ alkenyl;

one of Z^1 and Z^2 is a positively charged moiety, and the other of Z^1 and Z^2 is a negatively charged moiety;

15 N is an integer from 2 to 100;

T is hydrogen, halogen, alkyl, alkenyl, $-\text{OR}''$, $-\text{SR}''$, $-\text{COOR}''$, $-\text{OCOR}''$, $-\text{NR}''\text{R}''$, $-\text{N}^+(\text{R}'')_3$, $-\text{P}^+(\text{R}'')_3$, $-\text{S-C(=S)-S-R}''$, $-\text{S-C(=S)-O-R}''$, $-\text{S-C(=S)-NR}''\text{R}''$, $-\text{S-C(=S)-aryl}$, cyano, azido, aryl, heteroaryl or a targeting group, wherein R'' at each occurrence is independently hydrogen or alkyl; and

20 each alkyl, alkenyl, alkylene, aryl and heteroaryl is independently optionally substituted. It is to be understood that any embodiment of the compounds provided herein as set forth above, and any specific substituent and/or variable of the compounds provided herein as set forth above, may be independently combined with other embodiments and/or substituents and/or variables of the compounds to form embodiments not specifically set forth above. Furthermore, where a list of
25 substituents and/or variables is recited for any particular group or variable, it is to be understood

that each individual substituent and/or variable may be deleted from the particular embodiment and/or claim, and that the remaining list of substituents and/or variables will be deemed to be within the scope of the embodiments provided herein.

It is to be understood that in the present specification, combinations of substituents and/or variables of the depicted formulas are permissible only when such contributions result in stable compounds.

Structural lipid

In some embodiments, the lipid component of the nanoparticle composition may comprise one or more structural lipids. Without being bound by theory, it is contemplated that the structural lipid may stabilize the amphiphilic structure of the nanoparticles, such as, but not limited to, the lipid bilayer structure of the nanoparticles. Exemplary structural lipids that may be used in conjunction with the present disclosure include, but are not limited to, cholesterol, coprostanol, sitosterol, ergosterol, campesterol, stigmasterol, brassicasterol, tomatine, tomatidine, ursolic acid, α -tocopherol and mixtures thereof. In certain embodiments, the structural lipid is cholesterol. In some embodiments, the structural lipid includes cholesterol and a corticosteroid (such as prednisolone, dexamethasone, prednisone and hydrocortisone) or a combination thereof.

In some embodiments, the structural lipid is present at a concentration in the range of 20 mol% to 50 mol%. In some embodiments, the structural lipid is present at a concentration in the range of 30 mol% to 50 mol%. Preferably, the structural lipid is present at a concentration in the range of 32 mol% to 46 mol%. More preferably, the structural lipid is present at a concentration in the range of 34 mol% to 44 mol%. Further more preferably, the structural lipid is present at a concentration in the range of 36 mol% to 42 mol%.

In some embodiments, the structural lipid is present at a concentration of about 30 mol%, 31 mol%, 32 mol%, 33 mol%, 34 mol%, 35 mol%, 36 mol%, 37 mol%, 38 mol%, 38.5 mol%, 39 mol%, 40 mol%, 41 mol%, 42 mol%, 43 mol%, 44 mol%, 45 mol%, 46 mol%, 47 mol%, 48 mol%, 49 mol% or 50 mol%.

In an embodiment, the lipid nanoparticles provided herein comprise a steroid or steroid analogue. In an embodiment, the steroid or steroid analogue is cholesterol.

In an embodiment, the steroid is present at a concentration in the range of 39 mol% to 49 mol%, 40 mol% to 46 mol%, 40 mol% to 44 mol%, 40 mol% to 42 mol%, 42 mol% to 44 mol% or 44 mol% to 46 mol%. In an embodiment, the steroid is present at a concentration of 40 mol%, 41 mol%, 42 mol%, 43 mol%, 44 mol%, 45 mol% or 46 mol%.

5 In an embodiment, the molar ratio of the cationic lipid to the steroid is in the range of 1.0:0.9 to 1.0:1.2, or 1.0:1.0 to 1.0:1.2. In an embodiment, the molar ratio of the cationic lipid to cholesterol is in the range of about 5:1 to 1:1. In an embodiment, the steroid is present at a concentration in the range of 32 mol% to 40 mol% steroid.

10 In an embodiment, the molar ratio of the cationic lipid to the steroid is in the range of 1.0:0.9 to 1.0:1.2, or 1.0:1.0 to 1.0:1.2. In an embodiment, the molar ratio of the cationic lipid to cholesterol is in the range of about 5:1 to 1:1. In an embodiment, the steroid is present at a concentration in the range of 32 mol% to 40 mol% steroid.

Phospholipid

In some embodiments, the lipid component of the nanoparticle composition may comprise one or more phospholipids, such as one or more (poly)unsaturated lipids. Without being bound by theory, it is contemplated that the phospholipids may assemble into one or more lipid bilayer structures. Exemplary phospholipids that may form part of the nanoparticle composition of the present invention include, but are not limited to, 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,2-dilinoeoyl-sn-glycero-3-phosphocholine (DLPC), 1,2-dimyristoyl-sn-glycero-phosphocholine (DMPC), and the like. Exemplary phospholipids can be found, for example, in PCT Publication No. WO 2023/098842, the contents of which are incorporated herein by reference in their entirety. In certain embodiments, the nanoparticle composition comprises DSPC. In certain embodiments, the nanoparticle composition comprises DOPE. In some embodiments, the nanoparticle composition comprises both DSPC and DOPE.

15
20
25

Additional exemplary phospholipids include, for example, dipalmitoylphosphatidylglycerol (DPPG), palmitoylloleoylphosphatidylethanolamine (POPE), dioleoylphosphatidylethanolamine 4-(N-maleimidomethyl)-cyclohexane-1-carboxylate (DOPE-mal), dipalmitoylphosphatidylethanolamine (DPPE), dimyristoylphosphoethanolamine (DMPE),

distearoyl-phosphatidylethanolamine (DSPE), 16-O-monomethyl PE, 16-O-dimethyl PE, 18-1-trans PE, 1-stearoyl-2-oleoyl-phosphatidylethanolamine (SOPE), and 1,2-dielaioyl-sn-glycero-3-phosphoethanolamine (transDOPE). In an embodiment, the neutral lipid is 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC). In an embodiment, the neutral lipid is selected from DSPC, DPPC, DMPC, DOPC, POPC, DOPE and SM. In an embodiment, the phospholipid is 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC). In an embodiment, the phospholipid is selected from DSPC, DPPC, DMPC, DOPC, POPC, DOPE and SM.

In an embodiment, the phospholipid is phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidic acid (PA) or phosphatidylglycerol (PG).

In addition, the phospholipids that may form part of the nanoparticle composition of the present invention also include those phospholipids described in WO2017/112865, the entire contents of which are incorporated herein by reference in their entirety.

In some embodiments, the phospholipid is present at a concentration in the range of 5 mol% to 40 mol%. In some embodiments, the phospholipid is present at a concentration in the range of 5 mol% to 15 mol%. Preferably, the phospholipid is present at a concentration in the range of 7 mol% to 13 mol%. More preferably, the phospholipid is present at a concentration in the range of 9 mol% to 11 mol%.

In some embodiments, the phospholipid is present at a concentration of about 5 mol%, 6 mol%, 7 mol%, 8 mol%, 9 mol%, 9.5 mol%, 10 mol%, 10.5 mol%, 11 mol%, 12 mol%, 13 mol%, 14 mol% or 15 mol%.

Formulation

According to the present disclosure, the nanoparticle composition described herein may comprise at least one lipid component and one or more additional components, such as a nucleic acid or nucleic acid combination (for example, a therapeutic nucleic acid or nucleic acid combination described herein). The nanoparticle composition may be designed for one or more particular applications or targets. The components of the nanoparticle composition may be selected based on a particular application or target, and/or based on the efficacy, toxicity, cost, ease of use, availability or other characteristics of one or more components. Similarly, a particular formulation

of the nanoparticle composition may be selected for a particular application or target according to, for example, the efficacy and toxicity of a particular combination of each component.

The lipid component of the nanoparticle composition may comprise, for example, one or more of the lipids of series 01, 02, 03 and 04 described herein, for example one or more lipids according to formula (01-I), (01-II), (02-I), (02-II), (03-I), (04-I) and (04-III) (and subformulas thereof), phospholipids (such as unsaturated lipids, for example DOPE or DSPC), PEG lipid and structural lipid. The components of the lipid component may be provided at particular fractions.

In an embodiment, provided herein is a nanoparticle composition, wherein the nanoparticle composition comprises the cationic or ionizable lipid compound provided herein, a nucleic acid or nucleic acid combination, and one or more other lipids. In an embodiment, the cationic or ionizable lipid compound comprises one or more of the compounds of series 01, 02, 03 and 04 as described herein, for example one or more compounds according to formula (01-I), (01-II), (02-I), (02-II), (03-I), (04-I) and (04-III) (and subformulas thereof), and optionally one or more other ionizable lipid compounds. In an embodiment, the one or more other lipids are selected from neutral lipids, steroids and polymer-conjugated lipid. In an embodiment, the nucleic acid or nucleic acid combination is encapsulated within the lipid nanoparticles or associated with the lipid nanoparticles.

In an embodiment, provided herein is a nanoparticle composition (lipid nanoparticle), wherein the nanoparticle composition comprises:

- i) cationic lipid between 40 mol% and 50 mol%;
- ii) neutral lipid;
- iii) steroid;
- iv) polymer-conjugated lipid; and
- v) a nucleic acid or nucleic acid combination.

As used herein, "mol%" refers to the mole percentage of a component relative to the total number of moles of all lipid components in the LNP (i.e., the total number of moles of the cationic lipid, neutral lipid, steroid and polymer-conjugated lipid). Unless specifically indicated, the sum of the amounts (mol%) of the lipids in the composition is 100 mol%, i.e., the sum of the amounts (mol%) of the cationic lipid, phospholipid, steroid and PEG-conjugated lipid is 100 mol%.

In an embodiment, the lipid nanoparticle comprises 20 mol% to 65 mol% cationic lipid. More preferably, the lipid nanoparticle comprises 25 mol% to 60 mol% cationic lipid. Further more preferably, the lipid nanoparticle comprises 30 mol% to 55 mol% cationic lipid. Further more preferably, the lipid nanoparticle comprises 35 mol% to 55 mol% cationic lipid. Further more preferably, the lipid nanoparticle comprises 40 mol% to 52 mol% cationic lipid.

In an embodiment, the lipid nanoparticle comprises 41 mol% to 49 mol%, 41 mol% to 48 mol%, 42 mol% to 48 mol%, 43 mol% to 48 mol%, 44 mol% to 48 mol%, 45 mol% to 48 mol%, 46 mol% to 48 mol% or 47.2 mol% to 47.8 mol% cationic lipid. In an embodiment, the lipid nanoparticle comprises about 47.0 mol%, 47.1 mol%, 47.2 mol%, 47.3 mol%, 47.4 mol%, 47.5 mol%, 47.6 mol%, 47.7 mol%, 47.8 mol%, 47.9 mol% or 48.0 mol% cationic lipid.

In an embodiment, the neutral lipid is present at a concentration in the range of 5 mol% to 15 mol%, 7 mol% to 13 mol% or 9 mol% to 11 mol%. In an embodiment, the neutral lipid is present at a concentration of about 9.5 mol%, 10 mol% or 10.5 mol%. In an embodiment, the molar ratio of the cationic lipid to the neutral lipid is in the range of about 4.1:1.0 to about 4.9:1.0, about 4.5:1.0 to about 4.8:1.0, or about 4.7:1.0 to 4.8:1.0.

In an embodiment, the steroid is present at a concentration in the range of 39 mol% to 49 mol%, 40 mol% to 46 mol%, 40 mol% to 44 mol%, 40 mol% to 42 mol%, 42 mol% to 44 mol% or 44 mol% to 46 mol%. In an embodiment, the steroid is present at a concentration of 40 mol%, 41 mol%, 42 mol%, 43 mol%, 44 mol%, 45 mol% or 46 mol%. In an embodiment, the molar ratio of the cationic lipid to the steroid is in the range of 1.0:0.9 to 1.0:1.2 or 1.0:1.0 to 1.0:1.2. In an embodiment, the steroid is cholesterol.

In an embodiment, the nucleic acid or nucleic acid combination:lipid ratio in the LNP (i.e., N/P, wherein N represents the number of moles of the cationic lipid, and P represents the number of moles of phosphate present as part of the nucleic acid backbone) is in the range of 2:1 to 30:1, for example 3:1 to 22:1. In an embodiment, the N/P is in the range of 6:1 to 20:1 or 2:1 to 12:1. Exemplary N/P ranges include about 3:1, about 6:1, about 12:1 and about 22:1.

In an embodiment, provided herein is a lipid nanoparticle, wherein the lipid nanoparticle comprises:

- i) a cationic lipid having an effective pKa greater than 6.0;

- ii) 5 mol% to 15 mol% neutral lipid;
- iii) 1 mol% to 15 mol% anionic lipid;
- iv) 30 mol% to 45 mol% steroid;
- v) polymer-conjugated lipid; and

5 vi) a nucleic acid or nucleic acid combination,

wherein mol% is determined based on the total number of moles of lipids present in the lipid nanoparticle.

In an embodiment, the composition provided herein comprises:

(a) (i) 20 mol% to 65 mol% cationic lipid; (ii) 5 mol% to 40 mol% phospholipid; (iii) 20 mol%
10 to 50 mol% steroid; and (iv) polymer-conjugated lipid;

(b) (i) 40 mol% to 55 mol% cationic lipid; (ii) 5 mol% to 15 mol% phospholipid; (iii) 35 mol%
to 50 mol% steroid; and (iv) 2 mol% to 10 mol% polymer-conjugated lipid;

or

(c) (i) 45 mol% to 55 mol% cationic lipid; (ii) 6 mol% to 10 mol% phospholipid; (iii) 40 mol%
15 to 48 mol% steroid; and (iv) 1 mol% to 2.5 mol% polymer-conjugated lipid.

In an embodiment, the cationic lipid may be any of a variety of lipid species that carries a net positive charge at a selected pH, such as physiological pH. Exemplary cationic lipids are described below. In an embodiment, the cationic lipid has a pKa greater than 6.25. In an embodiment, the cationic lipid has a pKa greater than 6.5. In an embodiment, the cationic lipid has a pKa greater
20 than 6.1, greater than 6.2, greater than 6.3, greater than 6.35, greater than 6.4, greater than 6.45, greater than 6.55, greater than 6.6, greater than 6.65 or greater than 6.7.

In an embodiment, the lipid nanoparticle comprises 40 mol% to 45 mol% cationic lipid. In an embodiment, the lipid nanoparticle comprises 45 mol% to 50 mol% cationic lipid.

In an embodiment, the molar ratio of the cationic lipid to the neutral lipid is in the range of
25 about 2:1 to about 8:1. In an embodiment, the lipid nanoparticle comprises 5 mol% to 10 mol% neutral lipid.

Exemplary anionic lipids include, but are not limited to, phospholipid phosphatidylglycerol, dioleoyl phospholipid phosphatidylglycerol (DOPG), dipalmitoyl phospholipid phosphatidylglycerol (DPPG) or 1,2-distearoyl-sn-glycero-3-phosphate-(1'-rac-glycerol) (DSPG).

In an embodiment, the lipid nanoparticle comprises 1 mol% to 10 mol% anionic lipid. In an embodiment, the lipid nanoparticle comprises 1 mol% to 5 mol% anionic lipid. In an embodiment, the lipid nanoparticle comprises 1 mol% to 9 mol%, 1 mol% to 8 mol%, 1 mol% to 7 mol% or 1 mol% to 6 mol% anionic lipid. In an embodiment, the molar ratio of the anionic lipid to the neutral lipid is in the range of 1:1 to 1:10.

In an embodiment, the steroid is cholesterol. In an embodiment, the molar ratio of the cationic lipid to cholesterol is in the range of about 5:1 to 1:1. In an embodiment, the lipid nanoparticle comprises 32 mol% to 40 mol% steroid.

In an embodiment, the sum of the mol% of the neutral lipid and the mol% of the anionic lipid is in the range of 5 mol% to 15 mol%. In an embodiment, wherein the sum of the mol% of the neutral lipid and the mol% of the anionic lipid is in the range of 7 mol% to 12 mol%.

In an embodiment, the molar ratio of the anionic lipid to the neutral lipid is in the range of 1:1 to 1:10. In an embodiment, the sum of the mol% of the neutral lipid and the mol% of the steroid is in the range of 35 mol% to 45 mol%.

In an embodiment, the lipid nanoparticle comprises:

- i) 45 mol% to 55 mol% cationic lipid;
- ii) 5 mol% to 10 mol% neutral lipid;
- iii) 1 mol% to 5 mol% anionic lipid; and
- iv) 32 mol% to 40 mol% steroid.

In an embodiment, the lipid nanoparticle comprises 1.0 mol% to 2.5 mol% conjugated lipid. In an embodiment, the polymer-conjugated lipid is present at a concentration of about 1.5 mol%.

In an embodiment, the neutral lipid is present at a concentration in the range of 5 mol% to 15 mol%, 7 mol% to 13 mol% or 9 mol% to 11 mol%. In an embodiment, the neutral lipid is present at a concentration of about 9.5 mol%, 10 mol% or 10.5 mol%. In an embodiment, the molar ratio of the cationic lipid to the neutral lipid is in the range of about 4.1:1.0 to about 4.9:1.0, about 4.5:1.0 to about 4.8:1.0, or about 4.7:1.0 to 4.8:1.0.

In an embodiment, the steroid is cholesterol. In some embodiments, the steroid is present at a concentration in the range of 39 mol% to 49 mol%, 40 mol% to 46 mol%, 40 mol% to 44 mol%, 40 mol% to 42 mol%, 42 mol% to 44 mol% or 44 mol% to 46 mol%. In an embodiment, the steroid

is present at a concentration of 40 mol%, 41 mol%, 42 mol%, 43 mol%, 44 mol%, 45 mol% or 46 mol%. In certain embodiments, the molar ratio of the cationic lipid to the steroid is in the range of 1.0:0.9 to 1.0:1.2 or 1.0:1.0 to 1.0:1.2.

In an embodiment, the molar ratio of the cationic lipid to the steroid is in the range of 5:1 to 1:1.

In an embodiment, the lipid nanoparticle comprises 1.0 mol% to 2.5 mol% conjugated lipid. In an embodiment, the polymer-conjugated lipid is present at a concentration of about 1.5 mol%.

In an embodiment, the molar ratio of the cationic lipid to the polymer-conjugated lipid is in the range of about 100:1 to about 20:1. In an embodiment, the molar ratio of the cationic lipid to the polymer-conjugated lipid is in the range of about 35:1 to about 25:1.

In an embodiment, the molar ratio of the cationic lipid to the polymer-conjugated lipid is in the range of about 100:1 to about 20:1. In an embodiment, the molar ratio of the cationic lipid to the polymer-conjugated lipid is in the range of about 35:1 to about 25:1.

In some embodiments, the lipid nanoparticle has an average diameter in the range of 50 nm to 180 nm. In some embodiments, the lipid nanoparticle has an average diameter in the range of 50 nm to 150 nm. In some embodiments, the lipid nanoparticle has an average diameter in the range of 50 nm to 120 nm. In an embodiment, the lipid nanoparticle has an average diameter in the range of 50 nm to 100 nm or 60 nm to 85 nm.

In an embodiment, the composition comprises the cationic lipid provided herein, DSPC, cholesterol and PEG-lipid, and a nucleic acid or nucleic acid combination. In an embodiment, the cationic lipid provided herein, DSPC, cholesterol and PEG-lipid are in a molar ratio of 50:10:38.5:1.5.

In a preferred embodiment, the cationic lipid comprises compound 01-1.

In a preferred embodiment, the PEG-lipid comprises DMG-PEG.

In a preferred embodiment, the lipid nanoparticle comprises 30-55 mol% compound 01-1, 0.5-5 mol% DMG-PEG (e.g., DMG-PEG2000), 5-40 mol% DSPC and 20-50 mol% cholesterol.

In a most preferred embodiment, the composition comprises 50 mol% compound 01-1, 1.5 mol% DMG-PEG, 10 mol% DSPC and 38.5 mol% cholesterol.

The nanoparticle composition may be designed for one or more particular applications or targets. For example, the nanoparticle composition may be designed for delivering a nucleic acid or nucleic acid combination, such as RNA or RNA combination, to particular cells, tissues, organs or systems, or groups thereof, in a mammal. The physicochemical properties of the nanoparticle composition may be altered to increase selectivity for a particular bodily target. For example, the particle size may be adjusted based on the fenestration sizes of different organs. The nucleic acid or nucleic acid combination comprised in the nanoparticle composition may also be selected based on one or more desired delivery targets. For example, the nucleic acid or nucleic acid combination may be selected for a particular indication, condition, disease or disorder, and/or for delivery to particular cells, tissues, organs or systems, or groups thereof (e.g., local or specific delivery). In certain embodiments, the nanoparticle composition may comprise a nucleic acid or nucleic acid combination (particularly mRNA) encoding a target polypeptide, which is capable of being translated in a cell to produce the target polypeptide. Such a composition may be designed for specific delivery to a particular organ. In certain embodiments, the composition may be designed for specific delivery to the liver of a mammal.

The amount of the nucleic acid or nucleic acid combination in the nanoparticle composition may depend on the size, composition, desired target and/or application, or other properties of the nanoparticle composition, as well as the properties of the nucleic acid or nucleic acid combination. For example, the amount of RNA that may be used in the nanoparticle composition may depend on the size, sequence and other features of the RNA. The relative amounts of the nucleic acid or nucleic acid combination and other components (e.g., lipids) in the nanoparticle composition may also vary. In some embodiments, the wt/wt ratio of the lipid component to the nucleic acid or nucleic acid combination in the nanoparticle composition may be about 5:1 to about 60:1, such as about 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 22:1, 25:1, 30:1, 35:1, 40:1, 45:1, 50:1 and 60:1. For example, the wt/wt ratio of the lipid component to the nucleic acid or nucleic acid combination may be about 10:1 to about 40:1. In certain embodiments, the wt/wt ratio is about 20:1. The amount of the nucleic acid or nucleic acid combination in the nanoparticle composition may be measured, for example, using absorption spectroscopy (e.g., ultraviolet-visible spectroscopy).

In some embodiments, the nanoparticle composition comprises one or more RNAs, and the one or more RNAs, lipids and amounts thereof may be selected to provide a particular N:P ratio. The N:P ratio of the composition refers to the molar ratio of nitrogen atoms in one or more lipids to the number of phosphate groups in the RNA. In some embodiments, a lower N:P ratio is selected.

5 The one or more RNAs, lipids and amounts thereof may be selected to provide an N:P ratio of about 2:1 to about 30:1, such as 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 12:1, 14:1, 16:1, 18:1, 20:1, 22:1, 24:1, 26:1, 28:1 or 30:1. In certain embodiments, the N:P ratio may be about 2:1 to about 8:1. In other embodiments, the N:P ratio is about 5:1 to about 8:1. For example, the N:P ratio may be about 5.0:1, about 5.5:1, about 5.67:1, about 6.0:1, about 6.5:1 or about 7.0:1. For example,

10 the N:P ratio may be about 5.67:1.

Other characterization methods and various features of the nanoparticle composition, such as particle size, polydispersity index and zeta potential, may be found, for example, in PCT Publication No. WO 2023/098842, the contents of which are incorporated herein by reference in their entirety.

15 The encapsulation efficiency of the nucleic acid or nucleic acid combination describes the amount of the nucleic acid or nucleic acid combination encapsulated by, or otherwise associated with, the nanoparticle composition after preparation relative to the initial amount provided. A higher encapsulation efficiency (e.g., close to 100%) is desirable. Encapsulation efficiency may be measured, for example, by comparing the amount of the nucleic acid or nucleic acid combination

20 in a solution containing the nanoparticle composition before and after disruption of the nanoparticle composition with one or more organic solvents or detergents. Fluorescence may be used to measure the amount of free nucleic acid or nucleic acid combination (e.g., RNA) in a solution. For the nanoparticle composition described herein, the encapsulation efficiency of the nucleic acid or nucleic acid combination may be at least 50%, for example 50%, 55%, 60%, 65%,

25 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%. In some embodiments, the encapsulation efficiency may be at least 80%. In certain embodiments, the encapsulation efficiency may be at least 90%.

The nanoparticle composition may optionally comprise one or more coatings. For example, the nanoparticle composition may be formulated as a capsule, film or tablet having a coating. A

capsule, film or tablet comprising the composition described herein may have any useful size, tensile strength, hardness or density.

In one aspect, the present invention further provides a vaccine preparation (also referred to as a "vaccine agent") comprising the nucleic acid or nucleic acid combination described herein.

5 In some embodiments, the vaccine preparation comprises the composition described herein.

In some embodiments, the vaccine preparation comprises compound 01-1, DMG-PEG, DSPC and cholesterol; wherein the nucleic acid or nucleic acid combination encodes the polypeptide or polypeptide combination described herein. In an embodiment, the nucleic acid or nucleic acid combination comprises the nucleotide sequence as set forth in any one of SEQ ID NO: 1, 2, 3, 4,
10 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 or SEQ ID NO: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66.

In some embodiments, the vaccine preparation comprises compound 01-1, DMG-PEG, DSPC and cholesterol; wherein the nucleic acid or nucleic acid combination encodes the polypeptide or polypeptide combination described herein. In an embodiment, the nucleic acid or nucleic acid
15 combination comprises two or more nucleotide sequences selected from SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 or SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66.

In an embodiment, the vaccine preparation comprises a nucleic acid or nucleic acid combination encoding the polypeptide or polypeptide combination described herein and a lipid
20 encapsulating the nucleic acid or nucleic acid combination, the lipid comprises compound 01-1, DMG-PEG, DSPC and cholesterol, and the nucleic acid or nucleic acid combination comprises the nucleotide sequence as set forth in any one of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 or SEQ ID NO: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66.

25 In an embodiment, the vaccine preparation comprises a nucleic acid or nucleic acid combination encoding the polypeptide or polypeptide combination described herein and a lipid encapsulating the nucleic acid or nucleic acid combination, the lipid comprises compound 01-1, DMG-PEG, DSPC and cholesterol, and the nucleic acid or nucleic acid combination comprises two or more nucleotide sequences selected from SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16,

17, 18, 19, 20, 21 and 22 or SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66.

In an embodiment, the vaccine preparation comprises a nucleic acid or nucleic acid combination encoding the polypeptide or polypeptide combination described herein and a lipid encapsulating the nucleic acid or nucleic acid combination, the lipid comprises 50 mol% compound 01-1, 1.5 mol% DMG-PEG, 10 mol% DSPC and 38.5 mol% cholesterol, and the nucleic acid or nucleic acid combination comprises the nucleotide sequence as set forth in any one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 or SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66.

In an embodiment, the vaccine preparation comprises a nucleic acid or nucleic acid combination encoding the polypeptide or polypeptide combination described herein and a lipid encapsulating the nucleic acid or nucleic acid combination, the lipid comprises 50 mol% compound 01-1, 1.5 mol% DMG-PEG, 10 mol% DSPC and 38.5 mol% cholesterol, and the nucleic acid or nucleic acid combination comprises two or more nucleotide sequences selected from SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 13, 14, 15, 16, 17, 18, 19, 20, 21 and 22 or SEQ ID NOs: 45, 46, 47, 48, 49, 50, 51, 57, 58, 59, 60, 61, 62, 63, 64, 65 and 66.

Pharmaceutical composition/kit

According to the present disclosure, the nanoparticle composition may be formulated, in whole or in part, as a pharmaceutical composition. The pharmaceutical composition may comprise one or more nanoparticle compositions. For example, the pharmaceutical composition may comprise one or more nanoparticle compositions comprising one or more different nucleic acid or nucleic acid combination. The pharmaceutical composition may further comprise one or more pharmaceutically acceptable carriers, diluents, excipients or auxiliary components, such as those described herein. General guidance regarding the formulation and manufacture of pharmaceutical compositions and agents may be obtained, for example, in Remington, The Science and Practice of Pharmacy, 21st ed., A. R. Gennaro; Lippincott, Williams & Wilkins, Baltimore, Md., 2006. Conventional carriers, diluents, excipients and auxiliary components may be used in any pharmaceutical composition, unless any conventional carrier, diluent, excipient or auxiliary component is incompatible with one or more components of the nanoparticle composition. A

carrier, diluent, excipient or auxiliary component is incompatible with a component of the nanoparticle composition if the combination of the carrier, diluent, excipient or auxiliary component with the component of the nanoparticle composition results in any undesired biological effect or other deleterious effect.

5 In some embodiments, one or more carriers, diluents, excipients or auxiliary components may constitute greater than 50% of the total mass or volume of the pharmaceutical composition comprising the nanoparticle composition. For example, the one or more carriers, diluents, excipients or auxiliary components may constitute 50%, 60%, 70%, 80%, 90% or a higher percentage of the pharmaceutical composition. In some embodiments, the pharmaceutically
10 acceptable carrier, diluent or excipient is at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% pure. In some embodiments, the carrier, diluent or excipient is approved for human and veterinary use. In some embodiments, the carrier, diluent or excipient is approved by the United States Food and Drug Administration. In some embodiments, the carrier, diluent or excipient is of pharmaceutical grade. In some embodiments, the carrier, diluent or excipient
15 complies with the standards of the United States Pharmacopeia (USP), European Pharmacopoeia (EP), British Pharmacopoeia and/or International Pharmacopoeia.

 The relative amounts of the one or more nanoparticle compositions, the one or more pharmaceutically acceptable carriers, diluents, excipients and/or any additional ingredients in the pharmaceutical composition according to the present disclosure will vary depending on the identity,
20 size and/or condition of the subject being treated and further depending on the route of administration of the composition. For example, the pharmaceutical composition may comprise between 0.1% and 100% (wt/wt) of one or more nanoparticle compositions.

 In certain embodiments, the nanoparticle composition and/or pharmaceutical composition of the present disclosure is stored and/or transported under refrigerated or frozen conditions, and the
25 refrigerated or frozen storage and/or transportation conditions may be found, for example, in PCT Publication No. WO 2023/098842, the contents of which are incorporated herein by reference in their entirety.

 In certain embodiments, the pharmaceutical composition of the present disclosure comprises the nanoparticle composition disclosed herein and a pharmaceutically acceptable carrier, diluent

or excipient selected from one or more of: Tris, acetate (e.g., sodium acetate), citrate (e.g., sodium citrate), physiological saline, PBS and sucrose. In certain embodiments, the pharmaceutical composition of the present disclosure has a pH value between about 7 and 8 (e.g., 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9 or 8.0, or between 7.5 and 8 or between 7 and 7.8). For example, the pharmaceutical composition of the present disclosure comprises the nanoparticle composition disclosed herein, Tris, physiological saline and sucrose, and has a pH value of about 7.5-8, which is suitable for storage and/or transportation, for example, at about -20°C. For example, the pharmaceutical composition of the present disclosure comprises the nanoparticle composition disclosed herein and PBS, and has a pH value of about 7-7.8, which is suitable for storage and/or transportation, for example, at about 4°C or a lower temperature. In the context of the present disclosure, "stability," "stabilization" and "stable" refer to resistance of the nanoparticle composition and/or pharmaceutical composition disclosed herein to chemical or physical changes (e.g., degradation, particle size change, aggregation, change in encapsulation, etc.) under given manufacturing, preparation, transport, storage and/or use conditions, for example when stress, such as shear force, freeze/thaw stress, etc., is applied.

The nanoparticle composition and/or pharmaceutical composition comprising one or more nanoparticle compositions may be administered to any patient or subject, including a patient or subject who may benefit from delivery of a nucleic acid or nucleic acid combination to one or more particular cells, tissues, organs or systems, or groups thereof. Although the descriptions provided herein regarding nanoparticle compositions and pharmaceutical compositions comprising nanoparticle compositions are primarily directed to compositions suitable for administration to humans, those skilled in the art will understand that such compositions are generally suitable for administration to any other mammal. Modifications of compositions suitable for administration to humans to render the compositions suitable for administration to various animals are well known, and an ordinarily skilled veterinary pharmacologist can design and/or perform such modifications with only ordinary experimentation, if any. Subjects to whom the composition is contemplated to be administered include, but are not limited to, humans, other primates and other mammals, including commercially relevant mammals, such as cattle, pigs, horses, sheep, cats, dogs, mice and/or rats.

A pharmaceutical composition comprising one or more nanoparticle compositions may be prepared by any method known or hereafter developed in the field of pharmacology. In general, such methods of preparation include combining the active ingredient with a carrier, diluent, excipient and/or one or more other auxiliary components, and then, if desired or necessary, dividing, shaping and/or packaging the product into a desired single-dose or multi-dose unit.

The pharmaceutical composition according to the present disclosure may be prepared, packaged and/or sold in bulk, as a single unit dose and/or as multiple single unit doses. As used herein, a "unit dose" is a discrete amount of a pharmaceutical composition comprising a predetermined amount of an active ingredient (e.g., a nanoparticle composition). The amount of the active ingredient is generally equal to the dose of the active ingredient to be administered to the subject and/or a convenient fraction of this dose, such as one-half or one-third of this dose.

The pharmaceutical composition may be prepared in various forms suitable for various routes and methods of administration. For example, the pharmaceutical composition may be prepared as liquid dosage forms (e.g., emulsions, microemulsions, nanoemulsions, solutions, suspensions, syrups and elixirs), injectable forms, solid dosage forms (e.g., capsules, tablets, pills, powders and granules), dosage forms for topical and/or transdermal administration (e.g., ointments, pastes, creams, lotions, gels, powders, solutions, sprays, inhalants and patches), suspensions, powders and other forms.

Liquid dosage forms for oral and parenteral administration include, but are not limited to, pharmaceutically acceptable emulsions, microemulsions, nanoemulsions, solutions, suspensions, syrups and/or elixirs. In addition to the active ingredient, the liquid dosage form may further comprise inert diluents commonly used in the art, for example water or other solvents, solubilizing agents and emulsifiers, such as ethanol, isopropanol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butanediol, dimethylformamide, oils (especially cottonseed oil, peanut oil, corn oil, germ oil, olive oil, castor oil and sesame oil), glycerol, tetrahydrofurfuryl alcohol, polyethylene glycol and sorbitan fatty acid esters, and mixtures thereof. In addition to inert diluents, the oral composition may also comprise additional therapeutic agents and/or prophylactic agents, and additional agents, such as wetting agents, emulsifying agents and suspending agents, sweetening agents, flavoring agents and/or perfuming agents. In certain

embodiments for parenteral administration, the composition is mixed with a solubilizing agent, such as Cremophor™, alcohols, oils, modified oils, glycols, polysorbates, cyclodextrins, polymers and/or combinations thereof.

Injectable preparations, for example sterile injectable aqueous or oleaginous suspensions, may be formulated according to known techniques using suitable dispersing agents, wetting agents and/or suspending agents. The sterile injectable preparation may be a sterile injectable solution, suspension and/or emulsion in a non-toxic parenterally acceptable diluent and/or solvent, for example a solution in 1,3-butanediol. Acceptable vehicles and solvents that may be used include water, Ringer's solution (U.S.P.) and isotonic sodium chloride solution. Sterile fixed oils are conventionally used as solvents or suspending media. For this purpose, any bland fixed oil may be used, including synthetic monoglycerides or diglycerides. Fatty acids such as oleic acid may be used in the preparation of injectables.

Injectable preparations may be sterilized, for example, by filtration through a bacteria-retaining filter, and/or by incorporating sterilizing agents in the form of sterile solid compositions, which may be dissolved or dispersed in sterile water or other sterile injectable medium before use.

The present invention provides methods for delivering a nucleic acid or nucleic acid combination to a mammalian cell or organ, producing an antigen polypeptide in a mammalian cell, and preventing and/or treating a disease or disorder in a mammal in need thereof, the methods comprising administering to the mammal a nanoparticle composition comprising the nucleic acid or nucleic acid combination and/or contacting the mammalian cell with the nanoparticle composition.

The polypeptide/polypeptide combination, nucleic acid/nucleic acid combination or composition of the present disclosure may also be provided in the form of a kit. Accordingly, the present disclosure also relates to a kit comprising the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination or composition described herein. The kit may further comprise instructions for use or comprise a suitable container, as needed. In certain embodiments, the kit further comprises a device for administration. The kit generally includes a label indicating the intended use and/or method of use of the contents of the kit. The term "label" includes any written or recorded material provided on or with the kit or otherwise accompanying the kit.

Therapeutic application

The present invention provides the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination (particularly RNA), composition, vaccine preparation, pharmaceutical composition or kit of the present invention for use in preventing and/or treating mpox virus infection in a subject.

The present invention provides use of the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination (particularly RNA), composition, vaccine preparation, pharmaceutical composition or kit of the present invention in the manufacture of a medicament for preventing and/or treating mpox virus infection in a subject.

The present invention provides use of the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination (particularly RNA), composition, vaccine preparation, pharmaceutical composition or kit of the present invention in the manufacture of a vaccine for preventing and/or treating mpox virus infection in a subject.

The present invention provides a method for preventing and/or treating mpox virus infection in a subject in need thereof, the method comprising administering to the subject in need thereof a therapeutically effective amount of the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination (particularly RNA), composition, vaccine preparation, pharmaceutical composition or kit of the present invention. In an embodiment, the method comprises administering a therapeutically effective amount of a composition comprising the nucleic acid or nucleic acid combination of the present invention. In an embodiment, the method comprises administering a therapeutically effective amount of a composition comprising the RNA or RNA combination of the present invention, particularly a composition comprising the LNP described herein. In an embodiment, the method comprises administering a therapeutically effective amount of a composition comprising the mRNA or mRNA combination of the present invention, particularly a composition comprising the LNP described herein.

In some embodiments, the therapeutically effective amount is provided in one or more administrations. In some embodiments, the therapeutically effective amount is provided in two administrations. In some embodiments, the therapeutically effective amount is provided in three administrations. In some embodiments, the therapeutically effective amount is provided in four

administrations. In some embodiments, the therapeutically effective amount is provided in five administrations. In some embodiments, the therapeutically effective amount is provided in six administrations. Preferably, in some embodiments, the therapeutically effective amount is provided in two administrations.

5 In some embodiments, the polypeptide or polypeptide combination, nucleic acid or nucleic acid combination (particularly RNA), composition or pharmaceutical composition of the present invention may be administered to a subject by any method known to those skilled in the art, for example parenterally, orally, transmucosally, transdermally, intramuscularly, intravenously, intradermally, subcutaneously or intraperitoneally. Preferably, the composition of the present
10 invention is administered by intramuscular injection.

Beneficial effects

The polypeptide or polypeptide combination, nucleic acid or nucleic acid combination (particularly RNA), composition, pharmaceutical composition or kit of the present invention may exhibit excellent effects, including, for example but not limited to: (1) high in vitro expression
15 efficiency; (2) inducing in a subject a high level of binding antibodies against mpox virus antigen proteins; (3) inducing in a subject a high level of vaccinia virus cross-neutralizing antibody; and (4) inducing in a subject a high level of cellular immune response.

Examples

The present invention is further described with reference to the following examples. It should
20 be understood that these examples are provided merely by way of illustration and do not limit the present invention. The following materials and instruments are all commercially available or prepared according to methods well known in the art. The following experiments were all performed according to the manufacturer's instructions or according to methods and procedures well known in the art.

25 Example 1 Preparation of mRNA

1.1. Design and Linearization of DNA Template Plasmids

DNA open reading frame (ORF) sequences of constructs respectively encoding different mpox virus antigen polypeptide monomers (including antigen polypeptide monomers in which the polypeptide of interest is A29L, A35R, B2R, B6R, C15L, M1R or H3L) or antigen polypeptide

multimers (including antigen polypeptide multimers in which the polypeptide of interest is A29L+A35R+B6R+B2R+C15L+M1R, B2R+A35R+C15L, A29L+M1R+B6R, B6R+M1R+H3L, B6R+A35R+M1R, A29L+B2R+C15L, B2R+H3L+C15L, B2R+A35R+B6R, A29L+M1R+C15L or H3L+C15L+ M1R) were designed, and the constructs were respectively named 0201, 0302, 0402, 0505, 0603, 0702, 0803, 1101, 1201, 1301, 1401, 1501, 1601, 1701, 1801, 1901, 2001; wherein information on the DNA open reading frame (ORF) sequences of the constructs and the amino acid sequences encoded thereby is shown in Table 2 and Table 1, respectively.

Table 1. Amino acid sequence encoded by the respective constructs

Construct Name	Polypeptide of interest	Encoding one antigen polypeptide (antigen polypeptide monomer)	Encoding multiple antigen polypeptides (antigen polypeptide multimer)	Encoded full-length polypeptide	Amino acid sequence of the encoded full-length polypeptide (SEQ ID NO:)
0201	A29L	Full-length A29L peptide (SEQ ID NO: 111)	/	IgE signal peptide (SEQ ID NO: 134)+full-length A29L peptide (SEQ ID NO: 111)+HA polypeptide tag (SEQ ID NO: 135)	89
0302	A35R	Full-length A35R peptide (SEQ ID NO: 114)	/	Full-length A35R peptide (SEQ ID NO: 114)+HA polypeptide tag (SEQ ID NO: 136)	90
0402	B2R	B2R signal peptide (SEQ ID NO: 116), full-length B2R peptide (excluding the signal peptide) (SEQ ID NO: 117)	/	B2R signal peptide (SEQ ID NO: 116)+full-length B2R peptide (excluding the signal peptide) (SEQ ID NO: 117)+ HA polypeptide tag (SEQ ID NO: 136)	91

0505	B6R	B6R signal peptide (SEQ ID NO: 121), full-length B6R peptide (excluding the signal peptide) (SEQ ID NO: 122)	/	B6R signal peptide (SEQ ID NO: 121)+GS linker (SEQ ID NO: 137)+HA polypeptide tag (SEQ ID NO: 136)+full-length B6R peptide (excluding the signal peptide) (SEQ ID NO: 122)	92
0603	C15L	Full-length C15L peptide (SEQ ID NO: 126)	/	IgE signal peptide (SEQ ID NO: 134)+GS linker (SEQ ID NO: 137)+HA polypeptide tag (SEQ ID NO: 136)+full-length C15L peptide (SEQ ID NO: 126)	93
0702	M1R	M1R extracellular segment (SEQ ID NO: 128)	/	IgE signal peptide (SEQ ID NO: 134)+GS linker (SEQ ID NO: 137)+HA polypeptide tag (SEQ ID NO: 136)+M1R extracellular segment (SEQ ID NO: 128)	94
0803	H3L	Full-length H3L peptide (SEQ ID NO: 131)	/	IgE signal peptide (SEQ ID NO: 134)+GS linker (SEQ ID NO: 137)+HA polypeptide tag (SEQ ID NO: 136)+full-length H3L peptide (SEQ ID NO: 131)	95
0201-J0	A29L	Full-length A29L peptide (SEQ ID NO: 112)	/	IgE signal peptide (SEQ ID NO: 134)+full-length A29L peptide (SEQ ID NO: 112)	96

0301-J0	A35R	Full-length A35R peptide (SEQ ID NO: 114)	/	IgE signal peptide (SEQ ID NO: 134)+full-length A35R peptide (SEQ ID NO: 114)	97
0501-J0	B6R	B6R signal peptide (SEQ ID NO: 121), full-length B6R peptide (excluding the signal peptide) (SEQ ID NO: 122)	/	IgE signal peptide (SEQ ID NO: 134)+B6R signal peptide (SEQ ID NO: 121)+full-length B6R peptide (excluding the signal peptide) (SEQ ID NO: 122)	98
0701-J0	M1R	Full-length M1R peptide (SEQ ID NO: 129)	/	IgE signal peptide (SEQ ID NO: 134)+full-length M1R peptide (SEQ ID NO: 129)	99
0901-J0	E8L	Full-length E8L peptide (SEQ ID NO: 133)	/	IgE signal peptide (SEQ ID NO: 134)+full-length E8L peptide (SEQ ID NO: 133)	100

1101	A29L+A3 5R+B6R+ B2R+ C15L+M1 R	/	Full-length A29L peptide (SEQ ID NO: 113), A35R extracellular segment (SEQ ID NO: 115), B6R extracellular segment (SEQ ID NO: 124), B2R extracellular segment (SEQ ID NO: 119), C15L extracellular segment (SEQ ID NO: 127), full-length M1R peptide (SEQ ID NO: 130)	IgE signal peptide (SEQ ID NO: 134)+full-length A29L peptide (SEQ ID NO: 113)+GS linker (SEQ ID NO: 138)+ A35R extracellular segment (SEQ ID NO: 115)+GS linker (SEQ ID NO: 138)+B6R extracellular segment (SEQ ID NO: 124)+GS linker (SEQ ID NO: 138)+B2R extracellular segment (SEQ ID NO: 119)+GS linker (SEQ ID NO: 138)+full-length M1R peptide (SEQ ID NO: 130)	101
1201	B2R+A35 R+C15L	/	B2R signal peptide (SEQ ID NO: 116), B2R extracellular segment (SEQ ID NO: 119), A35R extracellular segment (SEQ ID NO: 115), full-length C15L peptide (SEQ ID NO: 126)	B2R signal peptide (SEQ ID NO: 116)+B2R extracellular segment (SEQ ID NO: 119)+GS linker (SEQ ID NO: 138)+ A35R extracellular segment (SEQ ID NO: 115)+GS linker (SEQ ID NO: 138)+ C15L full-length peptide (SEQ ID NO: 126)	102

1301	A29L+M1R+B6R	/	Full-length A29L peptide (SEQ ID NO: 113), M1R extracellular segment (SEQ ID NO: 128), full-length B6R peptide (excluding the signal peptide) (SEQ ID NO: 122)	IgE signal peptide (SEQ ID NO: 134)+full-length A29L peptide (SEQ ID NO: 113)+GS linker (SEQ ID NO: 138)+ M1R extracellular segment (SEQ ID NO: 128)+GS linker (SEQ ID NO: 138)+ full-length B6R peptide (excluding the signal peptide) (SEQ ID NO: 122)	103
1401	B6R+M1R+H3L	/	B6R signal peptide (SEQ ID NO: 121), B6R extracellular segment (SEQ ID NO: 124), M1R extracellular segment (SEQ ID NO: 128), full-length H3L peptide (SEQ ID NO: 131)	B6R signal peptide (SEQ ID NO: 121)+B6R extracellular segment (SEQ ID NO: 124)+GS linker (SEQ ID NO: 138)+ M1R extracellular segment (SEQ ID NO: 128)+GS linker (SEQ ID NO: 138)+ full-length H3L peptide (SEQ ID NO: 131)	104
1501	B6R+A35R+M1R	/	B6R signal peptide (SEQ ID NO: 121), B6R extracellular segment (SEQ ID NO: 124), A35R extracellular segment (SEQ ID NO: 115), full-length M1R peptide (SEQ ID NO: 130)	B6R signal peptide (SEQ ID NO: 121)+B6R extracellular segment (SEQ ID NO: 124)+GS linker (SEQ ID NO: 138)+ A35R extracellular segment (SEQ ID NO: 115)+GS linker (SEQ ID NO: 138)+ full-length M1R peptide (SEQ ID NO: 130)	105

1601	A29L+B2R+C15L	/	Full-length A29L peptide (SEQ ID NO: 113), B2R extracellular segment (SEQ ID NO: 119), full-length C15L peptide (SEQ ID NO: 126)	IgE signal peptide (SEQ ID NO: 134)+full-length A29L peptide (SEQ ID NO: 113)+GS linker (SEQ ID NO: 138)+B2R extracellular segment (SEQ ID NO: 119)+GS linker (SEQ ID NO: 138)+ full-length C15L peptide (SEQ ID NO: 126)	106
1701	B2R+H3L+C15L	/	B2R signal peptide (SEQ ID NO: 116), B2R extracellular segment (SEQ ID NO: 119), H3L extracellular segment (SEQ ID NO: 132), full-length C15L peptide (SEQ ID NO: 126)	B2R signal peptide (SEQ ID NO: 116)+B2R extracellular segment (SEQ ID NO: 119)+GS linker (SEQ ID NO: 138)+H3L extracellular segment (SEQ ID NO: 132)+GS linker (SEQ ID NO: 138)+ full-length C15L peptide (SEQ ID NO: 126)	107
1801	B2R+A35R+B6R	/	B2R signal peptide (SEQ ID NO: 116), B2R extracellular segment (SEQ ID NO: 119), A35R extracellular segment (SEQ ID NO: 115), full-length B6R peptide (excluding the signal peptide) (SEQ ID NO: 122)	B2R signal peptide (SEQ ID NO: 116)+B2R extracellular segment (SEQ ID NO: 119)+GS linker (SEQ ID NO: 138)+A35R extracellular segment (SEQ ID NO: 115)+GS linker (SEQ ID NO: 138)+ full-length B6R peptide (excluding the signal peptide) (SEQ ID NO: 122)	108

1901	A29L+M1 R+C15L	/	Full-length A29L peptide (SEQ ID NO: 113), M1R extracellular segment (SEQ ID NO: 128), full- length C15L peptide (SEQ ID NO: 126)	IgE signal peptide (SEQ ID NO: 134)+full-length A29L peptide (SEQ ID NO: 113)+GS linker (SEQ ID NO: 138)+ M1R extracellular segment (SEQ ID NO: 128)+GS linker (SEQ ID NO: 138)+ full-length C15L peptide (SEQ ID NO: 126)	109
2001	H3L+C15 L+ M1R	/	H3L extracellular segment (SEQ ID NO: 132), C15L extracellular segment (SEQ ID NO: 127), full- length M1R peptide (SEQ ID NO: 130)	IgE signal peptide (SEQ ID NO: 134)+ H3L extracellular segment (SEQ ID NO: 132)+GS linker (SEQ ID NO: 138)+ C15L extracellular segment (SEQ ID NO: 127)+GS linker (SEQ ID NO: 138)+ full-length M1R peptide (SEQ ID NO: 130)	110

Table 2. Nucleic acid sequence of each construct (SEQ ID NO:)

Construct Name	RNA(ORF)	MRNA	DNA(ORF)	DNA
0201	1	45	23	67
0302	2	46	24	68
0402	3	47	25	69
0505	4	48	26	70
0603	5	49	27	71
0702	6	50	28	72
0803	7	51	29	73
0201-J0	8	52	30	74
0301-J0	9	53	31	75
0501-J0	10	54	32	76
0701-J0	11	55	33	77
0901-J0	12	56	34	78
1101	13	57	35	79
1201	14	58	36	80
1301	15	59	37	81
1401	16	60	38	82
1501	17	61	39	83
1601	18	62	40	84
1701	19	63	41	85
1801	20	64	42	86
1901	21	65	43	87
2001	22	66	44	88

The 5'-UTR sequence (the corresponding DNA sequence is as set forth in SEQ ID NO: 142 or 150), DNA ORF and 3'-UTR sequence (the corresponding DNA sequence is as set forth in SEQ ID NOs: 143, 144 or 151) were inserted into plasmid pJ241 for IVT (internally developed, containing a kanamycin resistance gene, a T7 promoter sequence, a poly(A) tag (the corresponding DNA sequence is as set forth in SEQ ID NO: 146) and a unique type IIS restriction site downstream of the poly(A) sequence) for IVT vector construction.

Linearization was performed by digestion with the type IIS restriction enzyme Esp3I/BsmBI (Thermo, FD0294) or Bsa I (Kactus, GMP-BSA-EE101-11). Specifically, each 10 µg of plasmid was mixed with 10 U of restriction enzyme and incubated at 37°C for 4 hours to ensure complete linearization. The reaction was terminated by adding 1/10 volume of 3 M sodium acetate (pH 5.5) and 2.5 volumes of ethanol, thoroughly mixed, and cooled at -20°C for 1 hour. Centrifugation was performed at 4°C and 13800 g for 15 minutes to precipitate the linearized DNA, which was washed twice with 70% ethanol and resuspended in nuclease-free H₂O.

1.2. In vitro transcription of mRNA from a DNA template

Table 3 shows a 20 µL in vitro transcription reaction system, and a reaction mixture was prepared according to this reaction system for in vitro transcription. UTP was modified with 1-methylpseudouridine.

Table 3. 20 µL in vitro transcription reaction system

Reagent		Supplier	Catalog No.	Amount
Nuclease-free H ₂ O		Invitrogen	10977-015	Bring up to 20 µL
RNase inhibitor (40U/µl)		Kactus	GMP-RNI-ME101-12	0.5 µL
RNTP Mix	A	Wuhan Glycogene	MR-1001	8 µL (final 10 mM each)
	U	Wuhan Glycogene	MR-1004	
	N1-Me-pUTP	Wuhan Glycogene	MR-3002	
	C	Wuhan Glycogene	MR-1003	
	G	Wuhan Glycogene	MR-1002	
10× IVT reaction buffer		Kactus	GMP-T7P-EE101-A	2 µL
1M MgCl ₂		Invitrogen	AM9530G	0.8 µL
0.1M DTT		Sigma	43816-10ML	2 µL
100 U/mL inorganic pyrophosphatase		Kactus	GMP-PYR-YE101-12	0.8 µL
100mM NaCl		Invitrogen	AM9760G	1 µL
T7 RNA polymerase (50U/µL)		Kactus	GMP-T7P-EE101-12	2 µL

Linearized DNA				1 µg
----------------	--	--	--	------

In vitro transcription: the reaction mixture was incubated at 37°C for 6 hours, and then 1 µL of DNase I (RNase-free, 1 U/µL, Kactus, GMP-DNI-EE001-12) was added to remove the DNA template, followed by incubation at 37°C for 30 minutes. The synthesized RNA was purified by adding 0.5 volumes of 7.5 M LiCl and 50 mM EDTA, incubating at -20°C for 45 minutes, and then centrifuging at 4°C and 13800 g for 15 minutes to precipitate mRNA. The supernatant was removed, and the pellet was rinsed twice with 500 µL of ice-cold 70% ethanol. After rinsing, the mRNA was resuspended in nuclease-free H₂O, adjusted to a concentration of 1 mg/mL, and stored at -20°C.

MRNA capping: each 10 µg of uncapped mRNA was heated at 65°C for 10 minutes, placed on ice for 5 minutes, then mixed with 10 U of vaccinia virus capping enzyme (Kactus, GMP-VCS-VE101-13), 50 U of mRNA cap 2'-O-methyltransferase (Kactus, GMP-MEH-VE101-13), 0.2 mM SAM, 0.5 mM GTP and 1 U of RNase inhibitor, and incubated at 37°C for 60 minutes to generate a Cap 1 modified structure. As described above, the modified mRNA was precipitated by LiCl, and the RNA was resuspended in nuclease-free H₂O and stored at -20°C.

HPLC purification: RNA (data not shown) was purified by high performance liquid chromatography (HPLC) using a C4 column (5 µm) (10 mm×250 mm column). Buffer A contained 0.1 M triethylammonium acetate (TEAA) (pH = 7.0), and Buffer B contained 0.1 M TEAA (pH = 7.0) and 25% acetonitrile.

The mRNA sequences shown in Table 2 were obtained by the above in vitro transcription, capping and purification processes.

In addition, constructs respectively encoding A29L, A35R, B6R, M1R and E8L antigen polypeptides were prepared by the above method and were respectively named 0201-J0, 0301-J0, 0501-J0, 0701-J0 and 0901-J0. Information on the nucleic acid sequences of the constructs and the amino acid sequences encoded thereby is shown in Table 2 and Table 1, respectively.

Example 2 In vitro transfection

In this example, the mRNAs prepared in Example 1 and encoding different mpox virus antigen polypeptide monomers or antigen polypeptide multimers were transfected into HEK293T cells using Lipofectamine 2000 transfection reagent, and the in vitro expression of different mRNA

constructs was evaluated by flow cytometry. The specific in vitro transfection method and flow cytometry detection method were as follows.

In vitro transfection: HEK293T cells (Cell Bank of the Chinese Academy of Sciences, Shanghai, SCSP-502) were seeded in a 24-well cell culture plate at 1.5×10^5 cells/well. The cells were cultured overnight at 37°C and 5% CO₂, and transfection was performed after the cell confluence reached 70%-90%. An exemplary method for transfection was as follows: 1 µg of mRNA was added to 25 µL of serum-free Opti-MEM medium and mixed well by pipetting to serve as mixture A. Separately, another 25 µL of serum-free Opti-MEM medium was taken, 1 µL of Lipofectamine 2000 transfection reagent was added thereto, and after thorough mixing, it served as mixture B. Mixture A was added to mixture B, gently mixed, and allowed to stand at room temperature for 5 minutes. Finally, the mixed solution was added dropwise to HEK293T cells and cultured at 37°C, 5% CO₂ for 18-24 hours. Each mRNA was transfected in two replicate wells. In addition, a mixed solution containing only Lipofectamine 2000 and no mRNA was prepared to transfect HEK293T cells under the same conditions and cultured under the same conditions, thereby serving as a blank control.

Sample collection: the cell culture supernatant was aspirated and discarded, HEK293T cells were gently pipetted off with 200 µL of staining buffer (PBS containing 2% FBS), resuspended and mixed, and transferred to a 96-well plate. For cells transfected with mRNA molecules encoding antigen polypeptide multimers, the amount of cells in one well of each 24-well plate should be equally divided into the corresponding number of wells in a 96-well plate; for example, cells transfected with 1101 mRNA (encoding 6 antigen polypeptides) should be equally divided into 6 wells in the 96-well plate, and cells transfected with 1201, 1301, 1401, 1501, 1601, 1701, 1801, 1901 or 2001 mRNA (each respectively encoding 3 antigen polypeptides) should be equally divided into 3 wells in the 96-well plate. After the cells were transferred, they were centrifuged at 300 g, the supernatant was discarded, and the cells were washed twice with 300 µL of staining buffer.

Staining: corresponding primary and secondary antibodies were selected according to the antigen polypeptide encoded by each mRNA for staining, so as to detect whether the mRNA could

successfully express the antigen polypeptide in vitro. Table 4-1 shows the staining antibodies used.

Table 4-2 shows the names, suppliers and catalog numbers of the staining antibodies used.

Table 4-1. Staining antibodies used for flow cytometry

Construct name	Encoded polypeptide	Staining primary antibody	Primary antibody concentration	Staining secondary antibody	Secondary antibody dilution
0201	A29L	Murine anti-A29L	5 µg/mL	Anti-murine-FITC	1:100
0302	A35R	Human anti-A35R	2 µg/mL	Anti-human-PE	1:50
0402	B2R	Rabbit anti-B2R	5 µg/mL	Anti-rabbit-FITC	1:100
0505	B6R	Human anti-B6R	2 µg/mL	Anti-human-PE	1:50
0603	C15L	Rabbit anti-C15L	5 µg/mL	Anti-rabbit-FITC	1:100
0702	M1R	Murine anti-M1R	5 µg/mL	Anti-murine-FITC	1:100
0803	H3L	Murine anti-H3L	5 µg/mL	Anti-murine-FITC	1:100
1101	A29L	Murine anti-A29L	5 µg/mL	Anti-murine-FITC	1:100
	A35R	Human anti-A35R	2 µg/mL	Anti-human-PE	1:50
	B6R	Human anti-B6R	2 µg/mL	Anti-human-PE	1:50
	B2R	Rabbit anti-B2R	5 µg/mL	Anti-rabbit-FITC	1:100
	C15L	Rabbit anti-C15L	5 µg/mL	Anti-rabbit-FITC	1:100
	M1R	Murine anti-M1R	5 µg/mL	Anti-murine-FITC	1:100
1201	B2R	Rabbit anti-B2R	5 µg/mL	Anti-rabbit-FITC	1:100
	A35R	Human anti-A35R	2 µg/mL	Anti-human-PE	1:50
	C15L	Rabbit anti-C15L	5 µg/mL	Anti-rabbit-FITC	1:100
1301	A29L	Murine anti-A29L	5 µg/mL	Anti-murine-FITC	1:100
	M1R	Murine anti-M1R	5 µg/mL	Anti-murine-FITC	1:100
	B6R	Human anti-B6R	2 µg/mL	Anti-human-PE	1:50
1401	B6R	Human anti-B6R	2 µg/mL	Anti-human-PE	1:50
	M1R	Murine anti-M1R	5 µg/mL	Anti-murine-FITC	1:100

	H3L	Murine anti-H3L	5 µg/mL	Anti-murine-FITC	1:100
1501	B6R	Human anti-B6R	2 µg/mL	Anti-human-PE	1:50
	A35R	Human anti-A35R	2 µg/mL	Anti-human-PE	1:50
	M1R	Murine anti-M1R	5 µg/mL	Anti-murine-FITC	1:100
1601	A29L	Murine anti-A29L	5 µg/mL	Anti-murine-FITC	1:100
	B2R	Rabbit anti-B2R	5 µg/mL	Anti-rabbit-FITC	1:100
	C15L	Rabbit anti-C15L	5 µg/mL	Anti-rabbit-FITC	1:100
1701	B2R	Rabbit anti-B2R	5 µg/mL	Anti-rabbit-FITC	1:100
	H3L	Murine anti-H3L	5 µg/mL	Anti-murine-FITC	1:100
	C15L	Rabbit anti-C15L	5 µg/mL	Anti-rabbit-FITC	1:100
1801	B2R	Rabbit anti-B2R	5 µg/mL	Anti-rabbit-FITC	1:100
	A35R	Human anti-A35R	2 µg/mL	Anti-human-PE	1:50
	B6R	Human anti-B6R	2 µg/mL	Anti-human-PE	1:50
1901	A29L	Murine anti-A29L	5 µg/mL	Anti-murine-FITC	1:100
	M1R	Murine anti-M1R	5 µg/mL	Anti-murine-FITC	1:100
	C15L	Rabbit anti-C15L	5 µg/mL	Anti-rabbit-FITC	1:100
2001	H3L	Murine anti-H3L	5 µg/mL	Anti-murine-FITC	1:100
	C15L	Rabbit anti-C15L	5 µg/mL	Anti-rabbit-FITC	1:100
	M1R	Murine anti-M1R	5 µg/mL	Anti-murine-FITC	1:100
Blank control	—	Murine anti-A29L	5 µg/mL	Anti-murine-FITC	1:100
		Human anti-A35R	2 µg/mL	Anti-human-PE	1:50
		Rabbit anti-B2R	5 µg/mL	Anti-rabbit-FITC	1:100
		Human anti-B6R	2 µg/mL	Anti-human-PE	1:50
		Rabbit anti-C15L	5 µg/mL	Anti-rabbit-FITC	1:100
		Murine anti-M1R	5 µg/mL	Anti-murine-FITC	1:100

		Murine anti-H3L	5 µg/mL	Anti-murine-FITC	1:100
--	--	-----------------	---------	------------------	-------

Table 4-2. Antibody names, suppliers and catalog numbers of staining antibodies used for flow cytometry

Staining primary antibody	Antibody name	Manufacturer	Catalog No.
Murine anti-A29L	Anti-mpox virus/MPXV A29L antibody (SAA0286)	Antibody System	RVV13401
Human anti-A35R	Anti-MPXV A35R monoclonal antibody	Novoprotein	NC083
Rabbit anti-B2R	Anti-mpox virus/MPXV B2R/HA/hemagglutinin polyclonal antibody	Antibody System	PVV12501
Human anti-B6R	Anti-MPXV B6R monoclonal antibody	Novoprotein	NC082
Rabbit anti-C15L	Rabbit anti-MPXV C15L polyclonal antibody	Tsingke Biotechnology	NA (custom product)
Murine anti-M1R	Anti-mpox virus/MPXV M1R antibody (SAA0283)	Antibody System	RVV13301
Murine anti-H3L	Anti-mpox virus/MPXV H3L antibody (SAA0282)	Antibody System	RVV12701
Staining secondary antibody	Antibody name	Manufacturer	Catalog No.
Anti-human-PE	PE anti-human IgG Fc	Biolegend	410708
Anti-murine-FITC	Fluorescein (FITC)-conjugated affinity-purified goat anti-mouse IgG (subclasses 1+2a+2b+3), reacting specifically with Fc ⁺ fragments (absorbed with human, bovine and rabbit serum proteins)	Jacson	155-095-164
Anti-rabbit-FITC	FITC donkey anti-rabbit IgG	Biolegend	406403

After the primary antibody was diluted to the working concentration with staining buffer, the cells were resuspended, 50 µL/well, and stained at 4°C in the dark for 30 minutes. After incubation, the cells were centrifuged at 300 g, the primary antibody dilution was discarded, the cells were washed three times with staining buffer, and the supernatant was discarded. The secondary antibody was then diluted to the working concentration with staining buffer, and the cells were resuspended, 50 µL/well, and stained at 4°C in the dark for 30 minutes. After incubation, the cells were centrifuged at 300 g, the secondary antibody dilution was discarded, the cells were washed three times with staining buffer, and the supernatant was discarded. Finally, the cells were

resuspended with staining buffer at 200 μ L/well, and the samples were detected using a flow cytometer (BD biosciences, LSRFortessa).

Detection of in vitro expression by flow cytometry: the blank control group cells were used as the baseline for gating, and the ratio of the number of HEK293T cells successfully expressing the target polypeptide to the total number of cells was detected.

The flow cytometry results are shown in Table 5 (data are shown as the mean value of the expression-positive rates of two replicate wells). Except for 0201, 1101 and 1301, the remaining mRNAs could successfully express the target polypeptides in vitro. Among them, 0302, 1201 and 1401 had relatively high in vitro expression levels; 0402, 0505, 0603, 0702, 0803, 1601, 1701 and 1901 had moderate in vitro expression levels, and 1501, 1801 and 2001 had relatively low in vitro expression levels. Although successful expression of the target polypeptides of 0201 and 1101 was not detected in this example, in subsequent experiments, after the mRNAs were encapsulated in liposomes to prepare mRNA vaccines and mice were immunized, specific binding antibodies against the corresponding antigen polypeptides could be detected in the sera of the immunized mice, suggesting that failure to detect expression of the target polypeptides in this example does not necessarily mean that the mRNA molecules are not expressed, but may be a false negative caused by insufficient sensitivity of the detection method or other reasons. For 1301, expression of the target protein was not detected at the in vitro level, and specific binding antibodies against the antigen protein were also not detected in the sera of immunized mice, suggesting that 1301 cannot successfully express the target protein.

Table 5. In vitro expression results

Construct name	Encoded polypeptide	A29L	A35R	B2R	B6R	C15L	M1R	H3L
0201	A29L	0.1	—	—	—	—	—	—
0302	A35R	—	95.3	—	—	—	—	—
0402	B2R	—	—	35.9	—	—	—	—
0505	B6R	—	—	—	49.3	—	—	—
0603	C15L	—	—	—	—	28.6	—	—
0702	M1R	—	—	—	—	—	27.7	—
0803	H3L	—	—	—	—	—	—	17.2
1101	A29L-A35R-B6R-B2R-C15L-M1R	0.1	0.6	1.3	0.3	0.3	0.2	—
1201	B2R-A35R-C15L	—	60.2	13.3	—	17.9	—	—
1301	A29L-M1R-B6R	0.1	—	—	0.3	—	0.2	—
1401	B6R-M1R-H3L	—	—	—	90.3	—	33.1	6.5

1501	B6R-A35R-M1R	—	7.1	—	4.2	—	1.5	—
1601	A29L-B2R-C15L	4.4	—	23.9	—	36.3	—	—
1701	B2R-H3L-C15L	—	—	18.3	—	28.3	—	5.0
1801	B2R-A35R-B6R	—	2.5	2.1	2.3	—	—	—
1901	A29L-M1R-C15L	3.2	—	—	—	19.3	11.7	—
2001	H3L-C15L-M1R	—	—	—	—	1.6	1.3	0.6
Blank control group	—	0.2	0.8	1.7	0.3	0.3	0.3	0.3

"—" indicates not applicable or not detected

Example 3 Preparation of mRNA-LNP formulations

3.1. Experimental materials

The lipids of series 01 were synthesized with reference to the method described in patent WO2021204175; the lipids of series 02 were synthesized with reference to the method described in patent WO 2023/138611; the lipids of series 03 were synthesized with reference to the method described in patent WO2022152109; and the lipids of series 04 were synthesized with reference to the method described in patent WO2022247755.

DSPC, cholesterol, and mPEG2000-DMG (i.e., DMG-PEG 2000) were purchased through conventional commercial channels.

3.2. Exemplary method for preparing lipid nanoparticles

Briefly, cationic lipid (such as Compound 01-1), DSPC, cholesterol and PEGylated lipid were dissolved in ethanol at a molar ratio of 50:10:38.5:1.5, and the mRNAs prepared in Example 1 were diluted in 10 to 50 mM citrate buffer (pH = 4) according to the individual or combined modes shown in Table 6. The ethanol lipid solution and the mRNA aqueous solution were mixed at a volume ratio of 1:3 by using a microfluidic device (self-developed equipment) with a total flow rate in the range of 9-30 mL/min to prepare LNPs at a total lipid-to-mRNA weight ratio of about 10:1 to 30:1. Ethanol was removed by dialysis and replaced with DPBS. Finally, the lipid nanoparticles were filtered through a 0.2 μ m sterile filter. The finally obtained mRNA-LNP formulations (lipid nanoparticles comprising mRNA) were named mRNA-LNP formulations G1, G2, G3, G4, G5, G6, G7, G8, G9 and G10, respectively. The antigen polypeptide encoded by G10 was identical to the antigen polypeptide encoded by the mRNA vaccine in Zhang RR et al., Rational development of multicomponent mRNA vaccine candidates against mpox. The mRNA vaccine in Emerg Microbes Infect. 2023 Dec;12(1):2192815. encoded the same antigen

polypeptide; therefore, G10 was used as the reference product of the mRNA vaccine in subsequent experiments.

Table 6. Composition of each mRNA-LNP formulation

Formulation name	MRNA or mRNA combination	Remarks
G1	0201 + 0302 + 0402 + 0505 + 0603 + 0702	Each component maintained an equimolar ratio with the antigen polypeptides encoded by groups G4-G6
G2	0302 + 0402 + 0505 + 0603 + 0702 + 0803	Each component maintained an equimolar ratio with the antigen polypeptides encoded by groups G7-G9
G3	1101	—
G4	1201 + 1301	Mixed at an equal mass ratio
G5	1501 + 1601	Mixed at an equal mass ratio
G6	1801 + 1901	Mixed at an equal mass ratio
G7	1201 + 1401	Mixed at an equal mass ratio
G8	1501 + 1701	Mixed at an equal mass ratio
G9	2001 + 1801	Mixed at an equal mass ratio
G10	0201-J0 + 0301-J0 + 0501-J0 + 0701-J0 + 0901-J0	Mixed at an equal mass ratio

Example 4 Evaluation of immunogenicity of the candidate vaccine

5 In this example, immunogenicity assessment was performed on the mRNA-LNP formulations (vaccines) G1, G2, G3, G4, G5, G6, G7, G8, G9 and G10 prepared in Example 3.

In this example, 55 female BALB/c mice aged 6-8 weeks (Zhejiang Vital River) were randomly divided into 11 groups according to body weight (groups G1, G2, G3, G4, G5, G6, G7, G8, G9 and G10 named after the administered formulations, and group G11 administered PBS as placebo),
10 with 5 mice in each group, and according to the administration design in Table 7, a total dose of 2 µg of mpox virus mRNA vaccine or placebo was injected into each mouse by intramuscular injection, 50 µL/mouse. At week 3 after the initial immunization, i.e., day 21 after the initial immunization, one booster immunization at an equal dose was performed. Orbital blood collection was performed on the test mice on day 28 after the initial immunization. Serum was separated,
15 inactivated at 56°C for 30 min, and frozen at -80°C for later use.

Table 7. Administration design table

Group	Contained mRNA or mRNA combination	Total dose (μg/mouse)	Dose of each component (μg/mouse)
G1	0201 + 0302 + 0402 + 0505 + 0603 + 0702	2	0.2+0.3+0.4+0.4+0.4+0.3
G2	0302 + 0402 + 0505 + 0603 + 0702 + 0803	2	0.26+0.37+0.38+0.30+0.28+0.4
G3	1101	2	2
G4	1201 + 1301	2	1
G5	1501 + 1601	2	1
G6	1801 + 1901	2	1
G7	1201 + 1401	2	1
G8	1501 + 1701	2	1
G9	2001 + 1801	2	1
G10	0201-J0 + 0301-J0 + 0501-J0 + 0701-J0 + 0901-J0	2	0.4
G11	PBS	—	—

4.1 Detection of vaccine-induced serum binding antibody titers

The above-collected mouse sera were used to detect vaccine-induced serum binding antibody titers, and the specific detection method was as follows.

- 5 Coating: each monkeypox antigen protein (A29L antigen protein: Novoprotein, DRA207; A35R antigen protein: Novoprotein, DRA209; B2R antigen protein: Antibody System, EVV12501; B6R antigen protein: Novoprotein, DRA211; C15L antigen protein: customized protein from Tsingke Biotechnology; M1R antigen protein: Novoprotein, DRA210; H3L antigen protein: Antibody System, EVV12701; E8L antigen protein: Antibody System, YVV13201) was diluted to
- 10 1 μg/mL with 1×PBS buffer and added to a 96-well ELISA plate, 100 μL/well, sealed with a plate sealing film, and coated overnight at 4°C. After coating was completed, the liquid in the wells was discarded, and the plate was washed 3 times with PBST (300 μL/well).

- Blocking: 5% BSA was added into the wells to block nonspecific binding sites, 100 μL/well, sealed with a plate sealing film, and incubated at 37°C for 1 h. After blocking was completed, the
- 15 liquid in the wells was discarded, and the plate was washed 3 times with PBST (300 μL/well).

Addition of samples to be tested: the heat-inactivated sera of the test mice were subjected to three-fold serial dilution into 4 concentration gradients starting from an appropriate dilution with 5% BSA, and the diluted sera were added to the blocked ELISA plate, 100 μ L/well, sealed with a plate sealing film, and incubated at 37°C for 1 h. After incubation, the liquid in the wells was discarded, and the plate was washed 5 times with PBST (300 μ L/well).

Incubation with secondary antibody: Peroxidase-conjugated Affini Pure Donkey Anti-Mouse IgG (H+L) (Jackson, 715-035-151) was diluted 50,000-fold with 5% BSA and added to the ELISA plate, 100 μ L/well, sealed with a plate sealing film, and incubated at 37°C for 1 h. After incubation, the liquid in the wells was discarded, and the plate was washed 5 times with PBST (300 μ L/well).

Color development and termination: TMB substrate solution was added into the wells, 100 μ L/well, and incubated at room temperature protected from light for 5 min. After incubation was completed, ELISA stop solution (Solarbio, C1058) was added into the wells, 100 μ L/well, mixed thoroughly, and the absorbance value at a wavelength of 450 nm was read using a multifunctional microplate reader (Molecular Devices, ID5).

Result determination: 2.1 times the average absorbance value of mouse sera in the PBS group (G11) was used as the cut-off value, the highest dilution with an OD₄₅₀ reading greater than the cut-off value was selected, and the binding antibody titer in the sera of mice in each group was calculated according to the following formula: serum binding antibody titer (IgG titer) = OD reading at 450 nm/cut-off value $\times$ serum dilution factor. For groups in which the OD₄₅₀ reading was still lower than the cut-off value under a 1:100 dilution condition, the serum binding antibody titer was manually defined as half of the starting dilution, i.e., 50, which meant that no specific binding antibody titer against the coated antigen protein was detected in the mouse sera.

The test results are shown in Table 8. Except that mice in groups G4 and G8 did not produce specific binding antibodies against some antigens, mice in the other groups all produced specific binding antibodies against the antigen proteins encoded by the inoculated vaccines.

In the in vitro expression test (Example 2), successful expression of the A29L, M1R and B6R antigen polypeptides encoded by the 1301 mRNA molecule was not detected; likewise, production of specific binding antibodies against the A29L, M1R and B6R proteins was also not detected in the sera of mice in the G4 group immunized with the 1301 component, and it was reasonably

speculated that this was due to non-expression of 1301 mRNA. However, although successful expression of the 0201 and 1101 mRNA molecules was not detected in the in vitro expression test (Example 2), binding antibody titers against the respective antigen polypeptides encoded by 0201 or 1101 could be detected in the sera of mice immunized with the 0201 or 1101 component, suggesting that the 0201 and 1101 mRNA molecules were successfully expressed but at low expression levels below the detection limit level of the test method.

In addition, among all groups, the M1R antigen-specific binding antibody titer in the sera of mice in groups G6 and G7 was the highest. Moreover, the sera of mice in group G7 had the highest level of B6R binding antibody titer.

Table 8. Serum binding antibody titer of test mice

Group	Specific binding antibody titers against antigen polypeptides in the sera of test mice							
	A29L	A35R	B2R	B6R	C15L	M1R	H3L	E8L
G1	1897	400830	401541	51322	220048	385276	—	—
G2	—	416877	519825	26757	208430	458519	76099	—
G3	3979	13011	19583	6763	117417	13453	—	—
G4	50	42885	528079	50	313817	50	—	—
G5	29001	324	763742	69573	995947	41525	—	—
G6	12910	45044	349956	248	628080	1111940	—	—
G7	—	17427	302889	681074	133933	856499	50905	—
G8	—	50	887373	4936	758921	35196	113083	—
G9	—	10022	300280	549	359878	47304	13636	—
G10	28187	114777	—	121940	—	606369	—	114123
G11	50	50	50	50	50	50	50	50

"—" indicates not detected (because the mRNA contained in the vaccine does not encode the antigen polypeptide)

4.2 Detection of vaccine-induced serum neutralizing antibody titers

The collected mouse sera were used to detect vaccine-induced serum neutralizing antibody titers, and the specific detection method was as follows.

Cell preparation: Vero cells (Cell Bank of the Chinese Academy of Sciences, SCSP-520) were seeded into a 96-well cell culture plate at 2×10^4 cells/well, and the cells were cultured overnight for 8-16 hours at 37°C and 5% CO₂ using maintenance medium (DMEM containing 2% FBS). When the cell confluence reached 90%-95% on the next day, they were used for detection of neutralizing antibody.

Virus neutralization: several other sterile 96-well plates were taken, and the heat-inactivated serum samples were subjected to three-fold serial dilution into 6 concentration gradients starting from an appropriate dilution with maintenance medium; vaccinia virus (Elstree strain) (ATCC, VR-1549) was then diluted to 2000 PFU/mL with maintenance medium, and subsequently mixed with the diluted serum samples at equal volumes. Two replicate wells were set for each serum dilution, virus infection wells without serum samples were set as a virus control group (12 replicate wells), maintenance medium wells were set as a cell control group (12 replicate wells), and all of the above mixtures or dilutions were incubated at 37°C for 1 hour.

Cell infection: the cell culture supernatant was aspirated and discarded, and 100 µl of the incubated virus/serum mixture, virus dilution or maintenance medium alone was added to a 96-well plate containing Vero cells, followed by infection at 37°C for 1 hour. During infection, the cell culture plate was shaken once every 15 minutes to allow the virus to infect the cells uniformly. After infection was completed, the supernatant was aspirated and discarded, and semi-solid medium pre-warmed to 37°C (DMEM medium containing 1.8% sodium carboxymethylcellulose and 2% FBS) was added into the wells, 200 µL/well, so that the cells were cultured at 37°C and 5% CO₂ for 15-17 hours.

Fixation and permeabilization: after culture was completed, the cell supernatant was aspirated and discarded, the cells were washed once with PBS buffer (200 µL/well), and 4% tissue fixative was then added to the cells, 50 µL/well, and allowed to act at room temperature for 30 minutes. Subsequently, the medium of the cells was changed to 0.25% Triton X, 50 µL/well, and the cells were permeabilized at room temperature for 30 minutes. The liquid in the wells was discarded, and the cells were washed 3 times with PBST (300 µL/well).

Blocking: 50 µL of 10% BSA was added into the wells to block nonspecific binding sites, followed by incubation at 37°C for 1 h. After blocking was completed, the liquid in the wells was discarded, and the cells were washed 5 times with PBST (300 µL/well).

Antibody incubation: an HRP-conjugated anti-vaccinia virus antibody (Antibodies-Online, ABIN288495) was diluted 2000-fold with 5% BSA and added to the 96-well plate, 100 µL/well, followed by incubation at 37°C for 1 h. After incubation was completed, the liquid in the wells was discarded, and the cells were washed 5 times with PBST (300 µL/well).

Color development and reading: according to the instructions for use, TrueBlue™ peroxidase substrate (KPL, 5510-0030) was added to the 96-well plate at 50 µL/well, and incubated at room temperature protected from light for 15 minutes to allow viral plaques to develop color. After color development was completed, the color development solution was aspirated and discarded, and the 96-well plate was placed at room temperature overnight to dry. During this process, attention was paid to protection from light. On the next day, spot counting was performed using a fluorescent enzyme-linked immunospot analyzer (CTL).

Result calculation: the plaque area of a test well was taken as A_x , the mean plaque area of the virus control wells was A_v , and the mean plaque area of the cell control wells was A_c . The serum neutralization inhibition rate (%) was calculated according to the following formula: serum neutralization inhibition rate (%) = $1 - (A_x - A_c) / (A_v - A_c)$. The calculated serum neutralization inhibition rates and their corresponding concentration logarithm (Log) values were imported into GraphPad Prism 9.0 software to calculate the neutralizing antibody titer NT_{50} in the serum sample, i.e., the reciprocal of the serum dilution required to reduce plaque production by 50%. Whether there was a significant difference between groups was calculated by one-way analysis of variance (One-way ANOVA).

The test results are shown in FIG. 1. For mRNA vaccines with fusion expression, different vaccine design modes had different effects on the immunogenicity of the candidate vaccine. For example, the mRNA vaccines immunized in mice of groups G1, G3, G4, G5 and G6 (G1, G3, G4, G5 and G6 mRNA vaccines) all targeted 6 antigen polypeptides (A29L, A35R, B2R, B6R, C15L and M1R), wherein the G1 mRNA vaccine comprised 6 separately expressed mRNA molecules. In comparison, the fusion expression design mode (1501 + 1601) adopted by the G5 mRNA vaccine did not significantly affect the immunogenicity of the candidate vaccine, whereas the fusion expression design mode (1801 + 1901) adopted by the G6 mRNA vaccine could significantly enhance the ability of the candidate vaccine to induce neutralizing antibody.

Similarly, the mRNA vaccines immunized in mice of groups G2, G7, G8 and G9 (G2, G7, G8 and G9 mRNA vaccines) also all targeted 6 antigen polypeptides (A35R, B2R, B6R, C15L, M1R and H3L), wherein the fusion expression design mode (1201 + 1401) adopted by the G7 mRNA

vaccine could significantly enhance the ability of the candidate vaccine to induce neutralizing antibody.

In addition, among the mRNA vaccines G1-G10, the G6 and G7 candidate vaccines induced the highest serum neutralizing antibody titers, and their neutralization levels were significantly higher than those of the multivalent vaccine G1 and/or G2 prepared by co-encapsulation of separately expressed mRNA molecules, as well as the reference vaccine G10. Moreover, reducing the number of mRNAs required for preparing the vaccine through the fusion expression design greatly reduced the workload and difficulty required during vaccine production and release.

4.3 Detection of vaccine-induced cellular immune response

On day 42 after the initial immunization, the test mice were euthanized, mouse spleen tissues were collected, ground to prepare a single-cell suspension, and then the vaccine-induced cellular immune response was detected. The specific detection method was as follows.

Preparation of splenocyte suspension and counting of viable cells: an appropriate amount of 6-well plates was prepared, 3 mL of pre-cooled 1640 medium containing 2% FBS was added to each well, and a sterile cell strainer was placed in each well. The freshly collected spleen tissue was placed on the cell strainer, gently ground with the plunger of a 5 mL syringe, and filtered through the cell strainer to remove fat and other tissues, thereby forming a splenocyte suspension. The splenocyte suspension was transferred to a 15 mL centrifuge tube and centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded. The splenocytes were resuspended with 2 mL of 1× red blood cell lysis buffer and allowed to act at room temperature for 3-5 min, and then 4 mL of PBS buffer containing 10% FBS was immediately added and mixed thoroughly by inversion to terminate the lysis reaction. The splenocytes after red blood cell lysis were filtered through a cell strainer into another new 15 mL centrifuge tube to remove cell clumps, followed by centrifugation at 4°C and 300 g for 5 min, and the supernatant was completely removed. The cells were resuspended with 1 mL of pre-cooled 1640 medium containing 10% FBS and gently pipetted to mix thoroughly. Then 10 µL of cells was taken and mixed 1:1 with an equal volume of AO/PI staining solution (Count Star, RE010212), and the viable cell concentration was calculated. The cell concentration was adjusted to 1.5×10^7 viable cells/mL, and then 0.1 mL of cells was added to

a U-bottom 96-well cell culture plate, i.e., 1.5×10^6 viable cells per well, with two replicate wells for each sample.

Preparation of antigen stimulant working solution, negative control medium working solution and cell activation stimulant working solution (positive control):

5 1) antigen stimulant working solution: CD28 antibody (Biolegend, 122002), CD49d antibody (BD Biosciences, 553313) and monkeypox antigen pooled peptide library (GenScript) were added to pre-cooled 1640 medium containing 10% FBS so that the final concentrations of the CD28 antibody, the CD49d antibody and each individual monkeypox antigen peptide pool were all 4 $\mu\text{g/mL}$.

10 2) Negative control medium working solution: CD28 antibody and CD49d antibody were added to pre-cooled 1640 medium containing 10% FBS so that the final concentrations of both were 4 $\mu\text{g/mL}$.

 3) Cell activation stimulant working solution: a 500 $\times$ cell activation mixture (Cell activation cocktail, Biolegend, 423302) was diluted 250-fold with pre-cooled 1640 medium containing 10%
15 FBS so that its concentration became 2 $\times$.

 Stimulation: for the two replicate wells of each sample, 100 μL of antigen stimulant working solution was added to one replicate well so that the final concentrations of the CD28 antibody, the CD49d antibody and each individual monkeypox antigen peptide pool were all 2 $\mu\text{g/mL}$; 100 μL of negative control medium working solution was added to the other replicate well so that the final
20 concentrations of CD28 antibody and CD49d were both 2 $\mu\text{g/mL}$; for positive control wells, 100 μL of cell activation stimulant working solution was added to each well so that the working concentration of the Cell activation cocktail in each well was 1 $\times$. The cell samples were stimulated at 37°C and 5% CO_2 for 2 h.

 Blocking: a 1000 $\times$ Brefeldin A blocking agent (Biolegend, 420601) was diluted 50-fold with
25 pre-cooled 1640 medium containing 10% FBS, i.e., diluted to 20 $\times$, and then 10.5 μL of 20 $\times$ Brefeldin A was added to the 200 μL cell suspension in each well so that its working concentration was 1 $\times$, followed by gentle pipetting to mix thoroughly. The cell samples added with the blocking agent were placed at 37°C and 5% CO_2 for blocking for 4 h. After blocking was completed, the cells were placed in a centrifuge and centrifuged at 4°C and 300 g for 5 min, and the supernatant

was discarded. The cells were washed twice with PBS buffer, 200 μ L/well, centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded.

Live/dead dye staining: the live/dead cell dye Zombie green (Biolegend, 423112) was diluted 1000-fold with PBS buffer and then added to the cells to resuspend the cells, 50 μ L/well, followed by incubation at 4°C protected from light for 30 min. After incubation was completed, the cells were placed in a centrifuge and centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded. The cells were washed three times with flow staining buffer (PBS buffer containing 2% FBS), 200 μ L/well, centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded.

Surface staining: surface staining antibodies, anti-mouse CD45 (Biolegend, 103138), anti-mouse CD3 (Biolegend, 100216), anti-mouse CD4 (Biolegend, 100434) and anti-mouse CD8a (Biolegend, 100714), were sequentially added to the flow staining buffer so that the final concentration of each antibody was 2 μ g/mL. 50 μ L of the surface staining antibody mixture was added to the wells to resuspend the cells, followed by incubation at 4°C protected from light for 30 min. After incubation was completed, the cells were placed in a centrifuge and centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded. The cells were washed three times with flow staining buffer, 200 μ L/well, centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded.

Fixation, permeabilization and intracellular staining: 50 μ L of fixation solution in the fixation/permeabilization kit (BD Biosciences, 554714) was added to the cells to resuspend the cells, followed by incubation at 4°C protected from light for 20 min. After fixation was completed, the cells were placed in a centrifuge and centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded. The 10 $\times$ permeabilization solution in the fixation/permeabilization kit was diluted to 1 $\times$ with double-distilled water, and then the cells were washed three times with 1 $\times$ permeabilization solution, 200 μ L/well, centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded. Intracellular staining antibodies, anti-mouse IFN- γ (Biolegend, 505808), anti-mouse TNF- α (Biolegend, 506328), anti-mouse IL-2 (Biolegend, 503810) and anti-mouse IL-4 (Biolegend, 504118), were sequentially added to 1 $\times$ permeabilization solution so that the final concentration of each antibody was 2 μ g/mL. 50 μ L of the intracellular staining antibody mixture was added to the wells to resuspend the cells, followed by incubation at 4°C protected from light

for 30 min. After incubation was completed, the cells were placed in a centrifuge and centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded. The cells were washed three times with 1× permeabilization solution, 200 µL/well, centrifuged at 4°C and 300 g for 5 min, and the supernatant was discarded. The cells were resuspended with 200 µL of flow staining buffer, and the samples were detected using a flow cytometer (BD Biosciences, LSRFortessa).

The test results are shown in FIG. 2. Compared with the multivalent vaccine G1 prepared by co-encapsulation of separately expressed mRNA molecules, the mRNA vaccine G6 with fusion expression design induced a higher level of cellular immune response. For the mRNA vaccine expressing another antigen polypeptide combination, the mRNA vaccine G7 with fusion expression induced a level of cellular immune response comparable to that induced by the multivalent vaccine G2 prepared by co-encapsulation of separately expressed mRNA molecules (data not shown). In addition, all mRNA vaccines prepared in Example 3 induced a Th1-type immune response (partial data not shown).

Sequence Listing

SEQ ID NO:	Description	Sequence
1	0201 ORF RNA	AUGGACUGGACCUGGAUUCUGUUCCUGGUGGCGGCCGCCA CUAGGGUGCACAGCGACGGGACCCUGUUCCCAGGGGACGA CGACCUGGCCAUCCCCGCUACCGAGUUCUUCAGCACCAAG GCUGCCAAGAACCCCGAGACUAAGCGGGAGGCCAUCGUGA AGGCCUACGGGGACGAUAAUGAGGAGACGCUGAAGCAGC GUCUCACCAAUCUGGAGAAGAAGAUUACUAAUAUCACCA CCAAGUUCGAGCAGAUUCGAGAAGUGCUGUAAGAGGAAUG ACGAGGUGCUGUUUCGGCUGGAGAACCACGCCGAGACACU GCGGGCUGCCAUGAUCUCUCUUGCCAAGAAGAUCGAUGU GCAGACCGGCAGGCACCCUUAUUACCCUACGACGUGCCC GAUUAUGCCGAGUGA
2	0302 ORF RNA	AUGAUGACUCCCGAGAAUGACGAGGAGCAGACCAGCGUG UUCUCCGCUACCGUGUACGGGGACAAGAUCAGGGGAAG AAUAAGAGGAAGCGGGUGAUCGGCCUGUGCAUCAGGAUC AGCAUGGUGAUCAGCCUGCUGAGCAUGAUCACCAUGUCU GCAUUCUGAUCGUGCGGCUGAAUCAGUGCAUGAGCGCCA ACAAGGCCGCCAUCACUGAUUCAGCCGUGGCCGUGGCCGC CGCUUCCUCUACUCACAGGAAGGUGGUGUCCUCCACCACU CAGUACGAUCACAAGGAGAGCUGCAAUGGCCUGUAUUAC CAGGGCUCUUGCUACAUCUGCACAGCGAUUACAAGAGCU

			UUGAGGAUGCCAAGGCCAAUUGCGCCGCCGAGAGCAGCAC UCUGCCUAACAAGAGCGACGUGCUGACUACUUGGCUGAUC GACUAUGUGGAGGACACCUGGGGGAGCGAUGGGAACCCC AUCACCAAGACUACCAGCGACUAUCAGGACAGCGACGUGA GCCAGGAGGUGCGGAAGUACUUCUGCACCUACCCUUACGA CGUGCCUGACUACGCCUGA
3	0402 RNA	ORF	AUGACCCAGCUGCCUAUCCUGCUGCUGCUGAUCAGCCUGG UGUACGCUACCCCUAGUCCCCAGACUCCAAGAAGAUCGG GGACGACGCCACUAUCAGCUGCAGCAGGAAUAACACCAAC UACUACGUGGUGAUGAGCGCCUGGUUAUAAGGAGCCUAAC AGCAUCAUCCUGCUGGCCGCCAAGAGCGAUGUGCUGUACU UCGACAACUAUACCAAGGACAAGAUCUCUUACGAUAGCCC CUACGAUGACCUGGUGACUACCAUCACCAUCAAGAGUCUG ACCGCCGGGGAUGCCGGGACCUAUAUCUGUGCGUUUUUU AUGACCUCUACAACAAACGACACAGAUAAAGGUGGACUAC GAGGAGUAUAGCAUCGAGCUGAUCGUGAACACCGACAGC GAGAGUACCAUCGACAUAUCCUGAGCGGGAGUACUCCCG AGACUAUCAGCGAGAAGCCAGAGGACAUCGACAAUAGCA ACUGCAGCAGCGUGUUCGAGAUCACUACUCCUGAGCCUAU UACAGAUAAACGUGGAUGAUCACACCGACACCGUGACCUAC ACCAGCGACAGCAUCAACACCGUGAAUGCUAGCAGCGGUG AGAGUACCACGGAUGAGAUCCCCGAGCCCAUCACGGAUAA GGAGGAGGAUCACACCGUGACCGACACGGUGAGUUACACC ACCGUGUCUACCUCUCAGGGAUUGUGACCACAAAGUCUA CCACCGAUGAUGCCGAUCUGUAUGACACAUAACAAUGACAA CGAUACCGUGCCGCCACCACCGUGGGCGGCAGCACAAACA UCCAUAUAGUAACUACAAGACUAAGGAUUUCGUGGAGAUC UUUGGGAUCACAACCCUGAUUAUCCUGAGCGCCGUGGCGA UCUUCUGUAUCACCUACUAUAUCUGUAAUAAGCAUCCCAG GAAGUAUAAGACCGAGAACAAGGUGUAUCCGUACGACGU GCCGGACUACGCCUGA
4	0505 RNA	ORF	AUGAAGACCAUCUCCGUGGUGACCCUGCUGUGCGUGCUGC CAGCAGUGGUGUACAGUGGCGGCGGGGGAGCUACCCCU ACGACGUGCCCGACUACGCCACCUGUACGGUGCCAACCAU GAACAACGCUAAGCUGACCUCCACAGAGACGAGUUUUA UGACAAGCAGAAGGUGACCUUCACCUGCGACAGUGGCUAC CACAGCUUGGACCCCAACGCUGUGUGCGAGACAGACAAGU GGAAGUACGAGAACCCGUGCAAGAAGAUGUGCACCGUGA GUGACUACGUCAGCGAGCUGUACGACAAGCCGCUGUACGA GGUCAACUCCACGAUGACACUCUCUUGCAACGGGGAGACA AAGUACUUCAGGUGCGAGGAGAAGAACGGGAUAUACCAGC UGGAAUGAUACCGUGACAUGCCCCAAUGCCGAGUGCCAGC CCCUGCAGCUCGAGCAUGGGAGCUGUCAGCCCGUGAAGGA

			GAAGUACUCCUUCGGGGAGUACAUGACCAUCAAUUGCGA CGUGGGGUACGAGGUGAUCGGGGUGAGCUACAUCUCUUG CACCGCCAAUAGCUGGAACGUGAUUCCCAGCUGUCAGCAG AAGUGCGACAUCCCUUCUCUGUCCAACGGGCUGAUCAGCG GGAGCACGUUCAGCAUUGGCGGGGUGAUCCACCUGAGCU GCAAGAGUGGGUUCACCCUGACCGGGUCCCCGAGCAGCAC CUGCAUCGAUGGGAAGUGGAACCCGAUCCUGCCAACUUGC GUCCGGUCCAACGAGGAGUUCGACCCCGUGGACGACGGAC CAGACGACGAAACCGAUCUGAGCAAGUUGAGCAAGGACG UGGUGCAGUACGAGCAGGAGAUCGAGAGCCUGGAGGCCA CCUACCACAUCAUCAUCAUGGCACUGACCAUCAUGGGCGU GAUAUUCUUGAUCAGCAUCAUUGUGCUGGUGUGUUCUUG CGACAAGAACAACGACCAGUACAAGUUUCACAAGCUGCUG CCCUGA
5	0603 RNA	ORF	AUGGACUGGACCUGGAUUCUGUUCUGGUGGGCGGCCGU ACUCGGGUGCAUAGCGGCGGCGGGGGUCCUACCCCUACG ACGUGCCCGACUAUGCCGCCGAGACAAAGGAGUUUAAGAC CCUGUACAACCUGUUUAUCGACAGCUAUCUGCAGAAGCUG GCUCAGCACUCUAUUCCACCAAUGUGACAUGC GCCAUUC ACAUUGGUGAGGUGAUCGGCCAGUUCAAGAACUGCGCCC UGAGGAUACGAAUAAGUGCAUGAGCAACAGCAGGCUGA GCUUCACCCUGAUGGUGGAGUCCUUCAUCGAGGUGAUA GCCUGCUGAGCGAGAAGGACAGGAGGGCCAUCGCCGAGG AGAUCGGCAUCGACCUGGACGAUGUGCCCAGCGUGGUGA GCAAGCUGGAGAAGAACUGCAACGCAUAUGCUGAGGUGA ACAACAUCAUCGAUAUUCAGAAGCUGAAUAUCGGUGAGU GUUCCGCUCCCCCUGGGCAGCAUAUGCUGUUGCAGAU CGU GAACACGGGCAGUGCCGAGGCCAACUGCGGGCUGCAGACC AUCGUGAAGUCACUGAAUAAGAUCUACGUGCCACCCAUCA UCGAGAACCGGCUGCCCUACUACGACCCGUGGUUCCUGGU GGGGGUGGCCAUUAUUCUGGUGAUCUUCACGGUGGCCAU CUGCAGCAUCCGCAGGAACCUGGCCCUGAAGUACAGGUAC GGGACCUUCCUGUACGUCUGA
6	0702 RNA	ORF	AUGGACUGGACCUGGAUUCUGUUCUGGUGGCUGCCGCCA CCAGGGUGCACAGUGGCGGAGGAGGCAGCUACCCCUACGA CGUCCCCGACUACGCUGGGGCUGCUGCCUCCAUCCAGACC ACUGUGAACACCCUGAGUGAGCGGAUCAGCAGCAAGCUG GAGCAGGAGGCCAACGCCUCUGCUCAGACCAAGUGUGAUA UAGAGAUUGGCAAUUUCUAUAUCAGGCAGAACCACGGGU GCAACAUCACCGUGAAGAAUAUGUGCUCCGCUGACGCCGA CGCUCAGCUGGACGCCGUGCUGAGCGCGGCCACGGAGACA UAUUCCGGUCUGACCCCUAGAGCAGAAGGCGUAUGUGCCUG CCAUGUUCACAGCUGCACUGAACAUCACAGACCAGCGUGAA

		CACCGUGGUGAGGGACUUCGAGAACUACGUGAAGCAGAC AUGCAACAGCAGCGCGGUGGUGGAUAAUAAGCUCAAGAU CCAGAAUGUGAUCAUUGACGAGUGUUAUGGGGCCCCUGG GAGCCCCACUAACCUCGAGUUUAUUAUACGGGCAGCAGC AAGGGCAACUGCGCCAUCAAGGCCCUGAUGCAGCUGACAA CCAAGGCCACCACCCAGAUUGCCCCCAGGCAGGUGGCCGG UUGA
7	0803 ORF RNA	AUGGACUGGACCUGGAUUCUGUUUCUGGUGGCCGCCGCCA CACGGGUGCAUUCCGGCGGCGGCGGCAGCUACCCAUACGA UGUGCCUGACUACGCCGCCGCCGUGAAGACACCCGUGAUC GUGGUGCCCGUGAUUGACCGGCCCCCUAGCGAGACAUUCC CCAAUGUGCACGAGCACAUCAACGACCAGAAGUUCGACGA CGUGAAGGACAACGAGGUGAUGCAGGAGAAGAGGGACGU GGUGAUUGUGAACGACGAUCCUGACCACUACAAGGACUA CGUGUUCAUCCAGUGGACAGGCGGGAACAUCAGGGACGA CGACAAGUAUACCCACUUCUUCUCCGGCUUCUGUAACACC AUGUGUACAGAGGAGACAAAGCGGAACAUUGCUCGGCAC CUGGCCCUGUGGGACAGCAAGUUCUUCACCGAGCUGGAGA ACAAGAACGUGGAGUAUGUGGUGAUCAUCGAGAACGACA ACGUCAUCGAGGACAUCACCUUUCUGCGCCCUGUGCUGAA GGCCAUCCACGAUAAGAAGAUCGACAUCCUGCAGAUGC GG GAGAUCAUCACCGGGAUAAGGUGAAGACCGAGCUGGUG AUCGACAAGGAUCACGCCAUCUUCACCUACACCGGCGGCU AUGAUGUGAGCCUGAGCGCCUACAUAUCCGGGUGACCAC CGCCCUGAAUAUCGUGGAUGAGAUCAUCAAGUCCGGCGG GCUGAGCAGCGGCUUCUACUUCGAGAUCCGAGAAUCGAG AACGAGAUGAAGAUAACAGGCAGAUCAUGGACAACAGC GCCAAGUACGUGGAGCACGACCCUCGGCUGGUGGCCGAGC ACCGGUUCGAGACUAUGAAGCCCAACUUCUGGUCCAGAAU CGGGACUGUGGCCGCCAAGCGGUACCCCGGCGUGAUGUAC ACAUUCACCACCCACUGAUCAGCUUCUUGGCCUGUUCG ACAUCAACGUGAUCGGCCUGAUCGUGAUCCUGUUCAUCAU GUUCAUGCUGAUCUUAACGUGAAGUCCAAGCUGCUGUG GUUCCUGACCGGCACCUUCGUGACUGCCUUCAUCUGA
8	0201-J0 ORF RNA	AUGGACUGGACCUGGAUUCUCUUCUUGGUGGCAGCAGCC ACGCGAGUCCACUCCAUGGACGGAACUCUUUUCCCCGGAG AUGACGAUCUUGCAAUUCAGCAACUGAAUUUUUCUCUA CAAAGGCUGCUAAAAAUCCAGAGACUAAACGCGAAGCAA UUGUUAAGCCUAUGGAGACGAUAAUGAGGAAACUCUCA AACAACGGCUAACUAAUUGGAAAAAAGAUUACUAAUA UAACAACAAAGUUUGAACAAAUAGAAAAGUGUUGUAAAC GCAACGAUGAAGUUCUAUUUAGGUUGGAAAAUCACGCUG

		AAACUCUAAGAGCGGCUAUGAUUAUCUCUGGCUAAAAAGA UUGAUGUACAGACUGGACGGCAUCCAUAUGAGUAA
9	0301-J0 ORF RNA	AUGGACUGGACCUGGAUUCUCUUCUUGGUGGCAGCAGCC ACGCGAGUCCACUCCAUGAUGACACCAGAAAACGACGAAG AGCAGACAUCUGUGUUCUCCGCUACUGUUUACGGAGACA AAAUUCAGGGAAAGAAUAAACGCAAACGCGUGAUUGGUC UAUGUAUUAGAAUAUCUAUGGUUAUUUCACUACUAUCUA UGAUUACCAUGUCCGCGUUUCUCAUAGUGCGCCUAAAUCA AUGCAUGUCUGCUAACAAGGCUGCUAUUACUGACUCCGCU GUUGCCGUUGCUGCGGCAUCAUCUACUCAUAGAAAGGUU GUGUCUAGCACUACACAAUAUGAUCACAAAGAAAGCUGU AAUGGUUUUAUUUACCAGGGUUCUUGUUUAUUUAUACA UCAGACUAUAAGUCAUUCGAGGAUGCUAAAGCAAACUGC GCUGCGGAAUCAUCAACACUACCCAAUAAAUCCGAUGUCU UGACUACCUGGCUCAUUGAUUAUGUUGAGGAUACAUGGG GAUCUGAUGGUAAUCCAUAUACAAAAACUACAUCCGAUU AUCAAGAUUCUGAUGUAUCACAAGAAGUUAGAAAGUAUU UUUGUACAUA
10	0501-J0 ORF RNA	AUGGACUGGACCUGGAUUCUCUUCUUGGUGGCAGCAGCC ACGCGAGUCCACUCCAUGAAAACGAUUUCCGUUGUUACGU UGUUUAUGCGUACUACCUGCUGUUGUUUAUUCAACAUGUA CUGUACCCACUAUGAAUAACGCUAAAUUAACGUCUACCGA AACAUUCGUUUAUGAUAAACAGAAAGUUACGUUUACAUG UGAUUCAGGAUAUCAUUCUUGGAUCCAAUUGCUGUCUG CGAAACAGAUAAAUUGGAAUACGAAAAUCCAUGCAAGAA AAUGUGCACAGUUUCUGAUUAUGUCUCUGAACUAUAUGA UAAGCCAUUAUACGAAGUGAAUUCCACCAUGACACUAAG UUGCAACGGUGAAACAAAUAUUUUCGUUGUGAAGAAAA AAAUGGAAAUACUUCUUGGAAUGAUACUGUCACGUGUCC UAAUGCGGAAUGUCAACCUCUCAAUUAAGAACACGGAUC GUGUCAACCAGUUAAGAAAAAUACUCAUUUGGGGAAUA UAUGACUAUCAACUGUGAUGUUGGAUAUGAGGUUAUUGG UGUUUCGUUAUAAGUUGUACGGCUAAUUCUUGGAAUGU UAUCCAUAUGUCAACAAAAAUGUGAUUAUACCGUCCCU AUCUAAUGGAUUAUUUCCGGAUCUACAUUUUCUAUCGG UGGCGUUAUACAUCUUAAGUUGUAAAAGUGGUUUUACACU AACGGGGUCCCCAUCAUCCACAUGUAUCGACGGUAAAUGG AAUCCCAUACUCCCAACAUGUGUACGAUCUAACGAAGAAU UUGAUCCAGUGGAUGAUGGUCCCGACGAUGAGACAGAUC UGAGCAAACUCUCGAAAGACGUUGUACAAUAUGAACAAAG AAAUAGAAUCGUUAGAAGCAACUUAUCAUAUAUAUCAUAA UGGCGUUGACAAUUAUGGGUGUCAUAUUUCUAAUCCU

		UUAUAGUAUUAGUUUGUUCCUGUGACAAAAUAAUGACC AAUAUAAGUUCCAUAUUUUGCUACCGUGA
11	0701-J0 ORF RNA	AUGGACUGGACCUGGAUUCUCUUCUUGGUGGCAGCAGCC ACGCGAGUCCACUCCAUGGGUGCCGCAGCAAGCAUACAGA CGACUGUGAAUACACUCAGUGAACGUUUCGUCUAAA UAGAACAAGAAGCGAACGCUAGUGCUCAAACAAAAUGUG AUAUAGAAAUCGGAAAUUUUUAUAUCCGACAAAACCAUG GAUGUAACAUCACUGUUAAAAUAUGUGCUCUGCGGACG CGGAUGCUCAGUUGGAUGCUGUGUUAUCAGCCGCUACAG AAACAUUAUAGUGGAUUAACACCGGAACAAAAGCAUACG UACCAGCUAUGUUUACUGCUGCGUUAAACAUUCAGACGA GUGUAAACACUGUUGUUAAGAGAUUUUGAAAAUUAUGUGA AACAGACUUGUAAUUCUAGCGCUGUUGUCGAUAACAAU UAAAGAUACAAAACGUAAUUAUAGAUGAAUGUUACGGAG CCCCAGGAUCUCCAACAAAUUUGGAUUUAUUAAUACAG GAUCUAGCAAAGGAAAUUGUGCCAUAUAGGCGUUGAUGC AAUUGACUACUAAGGCCACUACUCAAUAGCACCUCAGACA AGUUGCUGGUACAGGAGUUCAGUUUUAUAUGAUUGUUUAU CGGUGUUUAUAAUUAUUGGCAGCGUUGUUUAUGUACUAGC CAAGCGUAUGCUGUUCACAUCCACCAAUGAUAAAAUCA ACUUAUUUUAGCCAUAUAGGAAAACGUCCAUUGGACUAC UUACAUGGACACAUCUUUAGAACUUCUCCGAUGAUUAU UGCUACCACGGAUAUACAAAACUGA
12	0901-J0 ORF RNA	AUGGACUGGACCUGGAUUCUCUUCUUGGUGGCAGCAGCC ACGCGAGUCCACUCCAUGCCUCAACAACUUCUCCUAUUA AUUAAGAAACUAAAAAGCAAUUUCUGACGCGCGAUUGA AGACGUUAGACAUACAUAUAAUGAGUCGAAACCAACCA CUAUCCAGAAUACUGGAAAACUAGUAAGGAUUAUUUUUA AAGGAGGAUAUAUAAUGUGGAGGGUUUCUCCCCAAUGAAU AUGUGUUAUCAACAAUACAUAUAUUAUUGGGGAAAGGAAG ACGAUUAUGGAUCCAUCACUUGAUAGAUGUAUACAAU ACUCUGGAGAGAUUAAUCUGGUUCAUUGGAAUAAGAAAA AAUAUAGUUCUUAUGAAGAGGCAAAAAACACGAUGAUG GAAUUAUCAUUAUUGCUAUAUUCUUAACAAGUAUCAGAU AUAAAAAUGUAUAUUUCAAAGAUAGUCAAUCAAUUGG AUUCCAUAAGAUCCGCUAAUAUGUCUGCGCCAUUUGAUU CAGUAUUUUUAUCUAGACAAUUUGCUACCUAGUACGUUGG AUUAUUUUACAUAUCUAGGAACAACUAUCAACCACUCUG CAGACGCAGCAUGGAUAAUUUUUCCAACACCAUAAACA UUCAUUCUGAUCAACUACUAAAUCAGAACACUAUUGU CGUCGUCUAAUCAUGAGGGUAAACCUCAUUAUAUAACAG AGAAUUAUAGAAAUCCGUAAAUUGAACGACGACACGC AAGUAUAUUAUUCUGGGGAGAUUAUACGAGCUGCAACCA

			CCUCACCAGUUCGAGAGAACUAUUUUUAUGAAAUGGUUGU CUGAUUUUGAGAGAGGCAUGUUUUUCAUAUUUAUCAAUAU AUAUCGAAGGGAUAAAACAUUCGCAAUAUUGCCAUAG UAUUCGUAUUUAUACUCACCGCUAUUCUCUUUCUUAUGA GUCAACGAUAUUCGCGAGAAAAACAAAACUAG
13	1101 RNA	ORF	AUGGACUGGACGUGGAUUCUCUCCUGGUGGCAGCCGCCA CUCGUGUGCACUCCGAUGGGACCUUGUUCCCUGGGGACGA CGACCUGGCCAUCCCCGCUACGGAGUUCUUCAGCACCAAG GCUGCCAAGAACCCGGAGACUAAGCGGGAGGCCAUCGUCA AGGCCUACGGAGAUGACAACGAGGAGACGCUGAAGCAGC GGCUGACCAACCUGGAGAAGAAGAUACGAACAUCACGAC CAAGUUCGAGCAGAUUGAGAAGUGCUGCAAGCGGAACGA CGAGGUGUUGUUCGCUUGGAGAACCACGCAGAGACUCU GCGGGCAGCCAUGAUCAGCUUGGCGAAGAAGAUCGACGU GCAGACUGGACGCCACCCCUAUGAGGGAGGAGGCGGCAGC GGCGGCGGCGGCAGCCGGCUGAAUCAGUGCAUGAGCGCCA ACAAGGCCGCCAUCACUGAUUCAGCCGUCGCCGUCGUCG CGCCUCCUCCACUCAUAGGAAGGUGGUGUCCAGUACGACA CAGUACGACCAUAAGGAGAGCUGUAAUGGGGCUGUACUAC CAGGGGUCGUGCUACAUCUGCACAGCGACUACAAGUCGU UUGAGGAUGCGAAGGCGAACUGCGCAGCCGAGAGCAGCA CGUUGCCUAACAAGAGCGACGUGCUGACUACUUGGCUGA UUGACUAUGUCGAGGAUACCUGGGGCUUGGAUGGGAACC CCAUCACGAAGACCACGUCUGACUACCAGGACUCCGAUGU GUCUCAGGAGGUGCGGAAGUACUUCUGUACUGGCGGCGG CGGUUCUGGUGGGGGCGGCAGCACCGUCCCCACC AUGAACAACGCCAAGCUGACCAGUACAGAGACUCCUUA ACGAUAAGCAGAAGGUGACAUUUACAUGCGAUUCCGGCU ACCACUCGUUGGACCCCAACGCUGUCUGUGAGACAGACAA GUGGAAGUACGAGAAUCCAUGUAAGAAGAUGUGCACGGU GUCCGACUAUGUUUCUGAGCUGUACGACAAGCCCCUUUAC GAGGUGAACAGCACCAUGACUCUCAGCUGCAACGGCGAGA CUAAGUACUUUCGGUGUGAGGAGAAGAACGGUAAUACGU CCUGGAACGAUACCGUUACUUGCCCCAACGCCGAGUGCCA GCCGUUGCAGCUGGAGCAUGGUAGCUGUCAGCCCGUGAA GGAGAAGUAUAGUUUCGGCGAGUACAUGACCAUUAUUG UGAUGUGGGUUACGAGGUGAUCGGGGUGAGUUACAUCUC UUGCACUGCCAAUAGCUGGAACGUGAUUCCAGCUGUCAG CAGAAGUGUGACAUAACCUUCUCUGUCCAACGGGCUGAUA GCGGGAGCACGUUCAGCAUUGGCGGGGUGAUCCACCUCAG CUGCAAGAGCGGGUUCACGUUGACCGGGUCGCCUCGUCG ACCUGCAUCGACGGGAAGUGGAAUCCCAUUCUGCCCACGU GCGUGCGGUCGAACGAGGAGUUCGACCCGGUCGACGAUG

		GACCCGAUGACGAGACUGAUCUGUCUAAGUUGUCGAAGG ACGUCGUUCAGUACGAGCAGGAGAUCGAGUCCCUGGAGG CCACGUACCACGGCGGUGGGCGGGAGCGGGGGAGGUGGCA GCAGUCCCCAGACUCCAAGAAGAUCCGGGGACGACGCCAC CAUCUCCUGCUCCCCGCAACAACACUAAUACUACGUGGUU AUGUCCGCCUGGUACAAGGAGCCCAACUCGAUCAUCCUGC UCGCUGCUAAGAGCGACGUCCUGUACUUCGACAACUAUAC CAAGGACAAGAUCAGCUACGACUCACCCUACGAUGACCUC GUGACCACCAUCACAAUUAAGAGUCUGACCGCCGGGGACG CCGGCACAUAUUAUUGCGCCUUCUUAUGACGAGUACCAC CAAUGAUACCGACAAGGUGGAUUAUGAGGAGUACUCGAU CGAGCUUAUCGUGAACACGGACAGUGAGUCGACCAUCGAC AUCAUCCUGUCCGGUAGCACCCUGAGACAAUCUCCGAGA AGCCGGAGGACAUCGACAAUAGCAAUUGCAGUCCGUCU UCGAGAUCACGACCCUGAGCCCAUUACCGACAACGUGGA UGAUCACACCGACACGGUGACGUACACCAGCGACUCUAU AACACCGUCAUUGCUUCCAGUGGGGAGUCCACCACUGACG AGAUUCCGGAGCCCAUCACCGACAAGGAGGAGGAUCACAC CGUGACCGACACGGUGAGUUACACCACCGUGUCUACCUC UCUGGUUAUUGUGACUACCAAGAGCACACGGACGACGCCG AUCUGUACGACACGUACAAUGACAACGACACCGUGCCGCC CACCACCGUGGGCGGGUCCACGACGUCCAUAUAGUAACUAC AAGACUAAGGACUUCGUGGGGGGCGGGCGGUUCCGGUGGC GGGGGCAGCGCUGAAACCAAGGAGUUAAGACGCUGUAC AACUUGUUAUCGACUCUUAACUGCAGAAGUUGGCCCAGC ACUCCAUCCCUACCAAUGUCACUUGUGCCAUCCACAUUGG AGAGGUGAUCGGCCAGUUAAGAACUGCGCCCUGAGGAU CACCAACAAGUGCAUGAGCAAUAGCCGCCUGUCCUUCACG CUCAUGGUGGAGUCCUUAUCGAGGUGAUCUCCUGCUGA GCGAGAAGGACAGGCGGGCUAUUGCUGAGGAGAUCGGUA UUGACCUCGACGACGUGCCGUCCGUCGUCUCCAAGCUGGA GAAGAACUGCAACGCAUAUGCUGAGGUGAACAACAUAU CGAUUAUCAGAAGCUGAAUAUCGGUGAGUGUCCGCUC CCCUGGGCAGCAUAUGCUGUUGCAGAUUGUCAAUACCGG GAGUGCUGAGGCCAACUGCGGGUUGCAGACCAUCGUGAA GAGUCUGAACAAGAUUUACGUGCCCCGAUCAUCGAGAAC CGCCUGCCCUACUACGACCCCGGUGGGGGCGGGUCUGGAG GAGGCGGCAGCGGCGCUGCCGCCUCCAUCAGACCACGGU GAUAACGUUGUCCGAGCGGAUUUCGAGCAAGCUGGAGCA GGAGGCCAACGCCUCUGCUCAGACCAAGUGUGAUUAUAGA GAUUGGCAAUUUCUAUAUCAGGCAGAACCACGGGUGCAA CAUCACCGUGAAGAAUAUGUGCUCCGCUGACGCCGACGCU CAGCUGGACGCCGUGCUGAGCGCGGCCACGGAGACAUAU
--	--	---

			CCGGUCUGACCCCUGAGCAGAAGGCGUAUGUGCCUGCCAU GUUCACAGCUGCUCUGAACAUCCAGACGUCCGUGAAUACC GUGGUGAGGGACUUCGAGAACUACGUGAAGCAGACGUGU AAUAGCAGCGCUGUCGUCGAUAAUAAGCUGAAGAUAUCCAG AACGUUAUUAUCGACGAGUGCUACGGCGCCCCCGGCUCCC CGACCAACCUGGAGUUCAUUAUACCGGUUCCUCCAAGGG GAACUGCGCUAUUAAGGCGCUGAUGCAGCUGACUACGAA GGCGACGACCCAGAUCGCGCCUCGCCAGGUGGCGGGGACC GGGGUGCAGUUCUACAUGAUCGUGAUCGGUGUGAUAUC CUGGCAGCUCUCUUUAUGUACUACGCCAAGCGCAUGCUGU UCACCAGCACCAACGACAAGAUUAAGCUGAUCCUCGCCAA UAAGGAGAACGUCCACUGGACCACCUACAUGGACACAUUC UUCCGCACCAGCCCUAUGAUUAUCGCCACCACUGACAUUC AGAACUGA
14	1201 RNA	ORF	AUGACCCAGCUGCCUAUCCUGCUGCUGCUGAUCAGUCUGG UGUACGCUACACCGUCUCCCCAGACUCCAAGAAGAUCGG GGACGACGCGACGAUCAGCUGCAGCAGGAAUAACACCAAC UACUACGUGGUGAUGAGCGCCUGGUACAAGGAGCCGAAU AGCAUCAUUCUGCUGGGCCGCAAGUCUGACGUGCUGUACU UCGACAAUUAACACAAAGGACAAGAUCUCCUACGACAGUCC UUAUGAUGAUCUGGUGACCACGAUCACCAUUAAGUCUCU GACUGCUGGGGAUGCCGGUACGUUAUUCUGCGCCUUCUUC AUGACGUCCACGACCAAUGACACUGAUAAGGUGGGAUUAC GAGGAGUACUCCAUCGAGCUGAUCGUGAACACCGACAGCG AGUCUACCAUCGACAUCAUCCUGUCCGGGAGCACCCCCGA GACUAUCAGUGAGAAGCCUGAGGAUAUUGACAACAGCAA UUGCAGUAGUGUCUUCGAGAUCACCACCCCGGAGCCUAUC ACUGACAACGUGGACGAUCACACCGACACGGUGACGUACA CCAGUGAUAGCAUCAACACCGUGAAUGCUAGCUCCGGGGA GUCCACGACCGAUGAGAUCCCGGAGCCUAUACCCGAUAAG GAGGAGGAUCACACCGUGACCGACACGGUGAGUUACACCA CCGUGUCUACCUCUCCGGGAUCGUGACUACGAAGAGCAC UACUGACGAUGCUGAUCUGUACGAUACCUACAACGAUAA CGAUACCGUGCCGCCACCACCGUGGGCGGCAGCACCACC AGCAUCAGCAACUAUAAGACUAAGGACUUCGUGGGGGGC GGGGGGAGCGGCGGCGGCGGCAGCCGGCUGAAUCAGUGC AUGAGCGCCAACAAGGCCGCCAUCACUGAUUCAGCCGUGG CUGUCGCCGCCGCCAGCUCCACCCACCGGAAGGUGGUGUC CAGCACCAACCAGUACGAUCACAAGGAGUCCUGCAAUGGC CUGUAUUACCAGGGCUCUUGCUACAUCCUGCACAGCGAUU ACAAGAGCUUUGAGGAUGCCAAGGCCAAUUGUGCAGCCG AGAGCAGCACGUGCCUAACAAGAGCGACGUGCUGACUAC UUGGCUGAUCGACUAUGUGGAGGACACCUGGGGGAGUGA

		UGGGAACCCCAUCACUAAGACCACGUCCGACUAUCAGGAC AGUGACGUGAGUCAGGAGGUGCGGAAGUACUUCUGCACC GGCGGCGGGGGCAGCGGCGGGGGCGGCAGCGCUGAGACA AAGGAGUUCAAGACCCUUUACAACUUGUUCAUCGAUAGU UAUCUGCAGAAGCUGGCCCAGCACAGCAUCCCCACCAUG UGACAUGCGCCAUUCACAUUGGUGAGGUGAUUGGCCAGU UCAAGAACUGCGCCCUGCGGAUAACUAACAAGUGUAUGU CUAUUCGCCUGUCCUUCACGCUCAUGGUGGAGUCCUU CAUCGAGGUGAUCUCCCUGCUGAGCGAGAAGGACAGGCG GGCUAUCGCCGAGGAGAUCCGGCAUCGACCUGGAUGACGUC CCCAGCGUGGUGUCCAAGCUGGAGAAGAACUGCAACGCAU AUGCUGAGGUGAACAACAUAUCGAUAUUCAGAAGCUGA AUAUCGGUGAGUGUCCCGCUCCCCUGGGCAGCAUAUGCU GUUGCAGAUCGUCAAUACCGGGUCUGCCGAGGCCAACUGC GGGCUGCAGACCAUCGUGAAGUCACUGAAUAAGAUCUAC GUGCCACCCAUCAUCGAGAACCAGGUGCCCCUACUACGACC CGUGGUUCCUGGUGGGGGUGGCCAUUAUUCUGGUGAUCU UCACGGUGGCCAUCUGCAGCAUCCGCAGGAACCUGGCCCU GAAGUACCGGUACGGUACCUUCCUCUACGUGUGA
15	1301 RNA	ORF AUGGACUGGACCUGGAUUCUGUCCUGGUGGCGGCCGCCA CAAGGGUGCACUCCGAUGGGACCUUGUUCCCUGGGGACGA CGACCUGGCCAUCCCCGCUACGGAGUUCUUCAGCACCAAG GCUGCCAAGAACCCGGAGACUAAGCGGGAGGCCAUCGUCA AGGCCUACGGAGAUGACAAUGAGGAGACGCUGAAGCAGC GUCUCACCAACCUUGAGAAGAAGAUACCAACAUCACUAC GAAGUUUGAGCAGAUUGAGAAGUGCUGCAAGCGCAAUGA UGAGGUGCUGUUUCGGCUGGAGAACCACGCCGAGACUCU GCGGGCUGCCAUGAUCAGCCUCGCCAAGAAGAUCGACGUG CAGACCGGAAGGCACCCCUAUGAGGGGGGAGGAGGGAGC GGCGGCGGCGGGAGCGGGGCCGCGGCCAGUAUCCAGACCA CCGUGAACACUCUGUCUGAGCGUAUCAGCAGCAAGCUGGA GCAGGAGGCCAACGCCUCUGCUCAGACCAAGUGUGAUUA GAGAUUGGCAAUUUCUAUAUCAGGCAGAACCACGGGUGC AACAUACCGUGAAGAAUAUGUGCUCCGCUGACGCCGACG CUCAGCUGGACGCCGUGCUGAGCGCGGCCACGGAGACUA UUCCGGUCUGACCCCUGAGCAGAAGGCGUAUGUGCCUGCC AUGUUCACAGCUGCGCUGAAUAUCCAGACGAGUGUUAAC ACGGUGGUGCGGGACUUUGAGAACUACGUGAAGCAGACG UGCAACAGCUCAGCGGUGGUGGAUAAUAAGCUCAAGAUC CAGAAUGUGAUCAUUGACGAGUGUUAUGGGGCCCCUGGG AGCCCCACUAACCUCGAGUUUAUUAUACGGGCAGCAGCA AGGGCAACUGCGCUAUAAGGCGCUGAUGCAGCUGACAAC GAAGGCCACCACCAGAUUGCUCGCGGCAGGUGGCAGGA

			GGGGGAGGGGGGUCCGGCGGGGGAGGCUCGACUUGCACG GUGCCGACGAUGAAUAACGCCAAGCUGACCUCCACCGAGA CGUCGUUCAACGACAAGCAGAAGGUGACGUUACCUGCGA CAGCGGCUAUCACAGCUUGGACCCCAACGCCGUCUGCGAG ACGGACAAGUGGAAGUACGAGAACCCGUGCAAGAAGAUG UGCACGGUGUCCGACUAUGUGUCCGAGCUGUACGAUAAG CCGCUGUACGAGGUGAACAGCACCAUGACGCUGUCGUGCA ACGGCGAGACAAAGUACUUCCGGUGCGAGGAGAAGAACG GGAACACCAGCUGGAACGAUACAGUGACCUGUCCCAACGC AGAGUGCCAGCCUCUGCAGUUGGAGCACGGGUCCUGUCAG CCCGUUAAGGAGAAGUACAGCUUUGGCGAGUACAUGACG AUUAACUGCGACGUCGGUACGAGGUGAUCCGGCGUGAGU UACAUCUCAUGUACCGCCAACAGCUGGAAUGUCAUACCUA GCUGCCAGCAGAAGUGUGACAUUCCGUCUCUCUCCAACGG GCUGAUCAGCGGGUCCACGUUCUCUAUCGGCGGCGUUAU CAUCUGUCCUGCAAGUCGGGCUUCACCCUGACCGGGUCCC CCUCUUCCACUUGCAUCGACGGCAAGUGGAACCCCAUCCU GCCCACCUGCGUGCGGAGCAACGAGGAGUUCGACCCCGUG GAUGACGGGCCUGAUGAUGAGACUGAUCUCUCUAAGCUG UCUAAGGACGUGGUCCAGUACGAGCAGGAGAUCGAGUCG CUGGAGGCCACCUACCACAUCAUCAUUAUGGCCUCACCA UCAUGGGGGUGAUCUUCUUGAUCAGCAUCAUCGUGCUGG UGUGUUCUUGCGACAAGAACAACGACCAGUACAAGUUUC ACAAGCUGCUGCCUUGA
16	1401 RNA	ORF	AUGAAGACCAUCUCCGUGGUGACCCUGCUGUGCGUGCUGC CAGCAGUGGUGUACAGCACCUGUACGGUGCCGACGAUGA AUAACGCCAAGCUGACCUCACCGAGACGUCGUUCAACGA CAAGCAGAAGGUGACGUUACCCUGCGACAGCGGCUAUCAC AGCUUGGACCCCAACGCCGUCUGCGAGACGGACAAGUGGA AGUACGAGAACCCGUGCAAGAAGAUGUGCACGGUGUCCG ACUAUGUGUCCGAGCUGUACGAUAAGCCGUGUACGAGG UGAACAGCACCAUGACGCUGUCGUGCAACGGCGAGACAAA GUACUUCCGGUGCGAGGAGAAGAACGGGAACACCAGCUG GAACGAUACAGUGACCUGUCCCAACGCAGAGUGCCAGCCU CUGCAGUUGGAGCACGGGUCCUGUCAGCCCGUUAAGGAG AAGUACAGCUUUGGCGAGUACAUGACGAUUAACUGCGAC GUCGGUACGAGGUGAUCCGGCGUGAGUUACAUCUCAUGU ACCGCCAACAGCUGGAAUGUCAUACCUAGCUGCCAGCAGA AGUGUGACAUUCCGUCUCUCUCCAACGGGCUGAUCAGCGG GUCCACGUUCUCUAUCGGCGGCGUUAUUCAUCUGUCUUGC AAGUCCGGCUUCACGUUGACCGGGUCGCCCUCGUCGACCU GCAUCGACGGGAAGUGGAAUCCCAUUCUGCCCACGUGCGU GCGGUCGAACGAGGAGUUCGACCCGGUCGACGAUGGGCCG

		GACGACGAGACAGAUCUGUCAAGUUGAGCAAGGAUGUG GUUCAGUAUGAGCAGGAGAUCGAGAGCCUGGAGGCCACG UACCACGGGGGAGGGGGGAGCGGCGGCGGGAGCGGG GCCGCGGCCAGUAUCCAGACCACCGUGAACACUCUGUCUG AGCGUAUCAGCAGCAAGCUGGAGCAGGAGGCCAACGCCUC UGCUCAGACCAAGUGUGAUUAUAGAGAUUGGCAAUUUCUA UAUCAGGCAGAACCACGGGUGCAACAUCACCGUGAAGAA UAUGUGCUCCGCUGACGCCGACGCUCAGCUGGACGCCGUG CUGAGCGCGGCCACGGAGACAUAUUCGGUCUGACCCUG AGCAGAAGGCGUAUGUGCCUGCCAUGUUCACAGCUGCGCU GAAUAUCCAGACGAGUGUUAACACGGUGGUGCGGGACUU UGAGAACUACGUGAAGCAGACGUGCAACUCCAGCGCCGUG GUGGAUAAUAAGCUCAAGAUCAGAAUGUGAUCAUUGAC GAGUGUUAUGGGGCCCCUGGGAGCCCCACUAACCUCGAGU UUAUUAUACCGGUUCCUCCAAGGGGAACUGCGCCAUCAA GGCGCUGAUGCAGUUGACAACGAAGGCGACGACCCAGAUC GCGCCUCGGCAGGUGGCCGGCGGGUGGGCGGGUCAGGCG GCGGGGGGAGCGCCGCCGUGAAGACGCCGGUGAUCGUGG UGCCCGUGAUCGACCGGCCUCCGAGCGAGACGUUCCCCAA CGUGCAUGAGCAUAUCAACGAUCAGAAGUUCGAUGACGU UAAGGACAACGAGGUGAUGCAGGAGAAGCGGGACGUGGU GAUCGUGAAUGACGACCCGGAUCACUACAAGGACUACGU GUUUAUCCAGUGGACCGGAGGCAACAUCGGGAUGACGA CAAGUACACCCAUUUUUUCUCCGGCUUCUGUAACACCAUG UGCACAGAGGAGACGAAGCGGAACAUCGCUCGCCACCUGG CACUGUGGGACUCCAAGUUCUUCACAGAGCUGGAGAACAA AGAAUGUGGAGUACGUCGUCAUCAUCGAGAAUGAUAAUG UGAUUGAGGACAUUACAUUUCUCAGGCCGGUGCUCAAGG CCAUCCAUGACAAGAAGAUCGACAUCCUGCAGAU GCGGGA GAUCAUCACCGGUAAUAAGGUGAAGACUGAGCUGGUCAU UGACAAGGACCACGCCAUCUUCACCUAUACCGGUGGCUAC GACGUCUCGCUGAGCGCGUAUAUCAUACGCGUGACCACCG CUCUUAUAUCGUCGACGAGAUCAUUAAGAGCGGUGGUC UGUCCUCGGGCUUCUACUUCGAGAUAGCCCGGAUCGAGAA CGAGAUGAAGAUCAACCGCCAGAUCAUGGACAACUCAGCG AAGUACGUCGAGCAUGAUCCCCGCCUGGUGGCCGAGCACC GGUUUGAGACAAUGAAGCCCAACUUCUGGUCCCGGAUAG GCACCGUGGCCGCCAAGCGGUAUCCGGGCGUCAUGUACAC GUUCACCACCCCGCUCAUCUCAUUCUUCGGCCUGUUCGAC AUUAACGUCAUCGGCCUGAUCGUGAUACUGUUCAUCAUG UUCAUGCUGAUCUUAACGUGAAGAGCAAGUUGCUGUGG UUCUGACAGGAACCUUCGUGACUGCCUUAUUUGA
--	--	--

17	1501 RNA	ORF	AUGAAGACCAUCUCGGUGGUGACACUGCUGUGCGUGUUG CCGGCGGUUGUUUAUAGUACGUGCACCGUGCCGACUAUG AACAAACGCCAAGUUGACGUCUACUGAGACUUCUUUCAAU GACAAGCAGAAGGUGACGUUCACCUGCGACAGCGGCUAUC ACAGCUUGGACCCCAACGCCGUCUGCGAGACGGACAAGUG GAAGUACGAGAACCCGUGCAAGAAGAUGUGCACGGUGUC CGACUAUGUGUCCGAGCUGUACGAUAAGCCGCUGUACGA GGUGAACAGCACCAUGACGCUGUCCUGCAACGGCGAGACA AAGUACUCCGUUGCGAGGAGAAGAACGGCAACACCAGC UGGAAUGACACCGUGACCUGUCCGAACGCCGAGUGCCAGC CCCUGCAGUUGGAGCAUUGGCUCCUGCCAGCCCGUGAAGGA GAAGUACUCCUUCGGGGGAGUACAUGACUAUCAACUGCGA CGUGGGCUACGAGGUGAUCGGGGGUGUCUUUAUAUCAGCUG CACCGCCAAUAGCUGGAACGUGAUUCCAGCUGUCAGCAG AAGUGUGACAUACCUUCUCUGUCCAACGGGCUGAUCAGCG GGAGCACGUUCAGCAUUGGCGGGGUGAUCCACCUGAGUU GCAAGUCUGGCUUCACGCUGACCGGAUCUCCUAGCUCGAC CUGCAUCGACGGGAAGUGGAACCCCAUCCUGCCAACUUGC GUCCGGUCCAACGAGGAGUUCGACCCCGUGGACGACGGAC CAGACGACGAAACCGAUCUGAGCAAGUUGAGCAAGGAUG UCGUGCAGUACGAGCAGGAGAUUCGAGUCCUGGAGGCGA CCUAUCACGGUGGGGGAGGGAGCGGCGGCGGCGGCUCACG GCUGAAUCAGUGCAUGAGCGCCAACAAGGCCGCCAUCACU GAUUCAGCCGUGGCCGUCGCCGCCGCUUCCUCCACCCACC GGAAGGUCGUCUCCAGCACCAACCCAGUAUGAUCACAAGGA GUCCUGCAACGGGCUGUACUAUCAGGGCAGUUGCUACAUC CUGCACAGCGACUACAAGUCGUUUGAGGAUGCUAAGGCC AACUGCGCCGCCGAAAGCUCCACGCUGCCGAAUAAGUCGG AUGUCCUGACCACGUGGCUCAUCGACUACGUGGAGGACAC CUGGGGCUCGGAUGGGAACCCCAUCACGAAGACCACCAGC GACUAUCAGGACAGCGAUGUGAGCCAGGAGGUGCGUAAG UACUUCUGCACCGGGGGCGGAGGGGUCUGGAGGAGGCGGC AGCGGCGCUGCCGCCUCCAUCCAGACCACUGUGAACACCC UGUCCGAGCGCAUUUCGUCCAAGCUGGAGCAGGAGGCCAA CGCCUCCGCCCAGACGAAGUGCGACAUCGAGAUAGGGAAU UUCUACAUCCGCCAGAACCACGGGUGCAACAUCACCGUGA AGAAUAUGUGCUCGCGUGACGCCGACGCUCAGCUGGACGC CGUGCUGAGCGCGGCCACGGAGACAUAUUCGGGUCUGACC CCGGAGCAGAAGGCUUACGUGCCUGCCAUGUUUACGGCGG CCCUGAAUAUACAGACUUCUGUGAAUACGGUGGUGCGGG ACUUCGAGAAUACGUGAAGCAGACUUGCAACUCCAGCGC CGUGGUGGAUAAUAAGCUCAAGAUAUCCAGAAUGUGAUCAU UGACGAGUGUUAUGGGGGCCCCUGGGAGCCCCACUAACCUC
----	-------------	-----	---

		GAGUUUAUUAUACCGGUUCCUCCAAGGGGAACUGCGCC AUCAAGGCGCUGAUGCAGCUGACUACUAAGGCCACCACCC AGAUCGCCCCACGUCAGGUGGCUGGCACCGGCGUUCAGUU CUACAUGAUCGUGAUCGGGGUGAUCAUUCUGGCUGCGCU GUUCAUGUACUACGCCAAGCGCAUGCUCUUUACUUCGACC AACGAUAAGAUCAAACUUAUCUUGGCGAAUAAGGAGAAU GUGCAUUGGACGACGUACAUGGACACGUUCUCCGGACG UCUCCUAUGAUCAUUGCGACCACCGAUAUUCAGAACUGA
18	1601 ORF RNA	AUGGACUGGACAUGGAUUCUGUCCUGGUGGCCGCCGCCA CCAGGGUGCACAGCGACGGCACCCUGUUCCCAGGCGACGA CGACCUGGCCAUCCCCGCUACCGAGUUCUUCAGCACCAAG GCUGCCAAGAACCCCGAGACUAAGCGGGAGGCCAUCGUGA AGGCCUACGGGGAUGAUAACGAGGAGACUCUGAAGCAGC GGCUGACCAACCUGGAGAAGAAGAUUACCAACAUUACCAC CAAGUUCGAGCAGAUUCGAGAAGUGCUGCAAGCGGAACGA UGAGGUGCUGUUUCGGCUGGAGAACCACGCCGAGACACU GCGCGCCGCCAUGAUCAGCCUGGCCAAGAAGAUUCGACGUG CAGACCGGCCGCCACCCCUACGAGGGGGGCGGCGGCUCUG GGGGCGGCGGGAGCAGCCCCAGACUCCAAGAAGAUUCG GGAUGAUGCCACCAUCAGCUGCUCUCCCGGAACAACACCAAC UACUACGUGGUGAUGUCCGCCUGGUACAAGGAGCCUAAC AGCAUCAUCCUGCUGGCCGCCAAGAGCGAUGUGCUGUACU UUGACAAUUAACCAAGGAUAAGAUCAGCUACGACAGCCC CUACGACGAUCUGGUGACUACCAUCACCAUCAAGAGCCUG ACAGCCGGCGACGCAGGCACCUACAUCUGCGCCUUCUUCA UGACCAGCACCAACCAUGACACCGACAAGGUGGAUUACGA GGAGUACUCCAUCGAGCUGAUCGUGAAUACUGACUCAGA GUCUACCAUCGACAUAUCCUGUCCGGGAGCACCCCCGAG ACUAUCAGUGAGAAGCCUGAGGACAUCGACAACAGCAAC UGCAGCAGUGUGUUCGAGAUCACCACCCAGAGCCUAUCA CUGACAACGUGGAUGAUCACACCGACACUGUGACAUACAC CAGUGAUAGCAUCAACACCGUGAAUGCUAGCUCUGGGGA GAGCACCAUGAUGAGAUCCCCGAGCCUAUACCCGACAAG GAGGAGGAUCACACCGUGACCGACACUGUGAGUUACACCA CAGUGUCUACCUCCUCUGGGAUCGUGACCACCAAGAGCAC CACCGACGAUGCCGACCUGUACGAUACAUAUAGAUAAAC GAUACAGUGCCUCCCACCACCGUGGGAGGCAGCACCA GCAUUCUAAUUAACAAGACCAAGGACUUCGUGGGGGGCG GGGGGAGCGGGGGCGGGGGGAGCGCCGAAACCAAGGAGU UCAAGACCCUGUACAACCUGUUCAUUCGAUAGCUACCUGCA GAAGCUGGCCCCAGCACAGCAUCCCCACCAUUGUGACAUGC GCCAUUCACAUUGGGGAGGUGAUCGGCCAGUUCAAGAAC UGCGCCCUGAGGAUCACCAUAAGUGCAUGAGCAAUAGCC

		GCCUGUCCUUCACUCUGAUGGUGGAGUCCUUCAUCGAGGU GAUCUCCCUGCUGUCAGAGAAGGACAGGGCGGGCUAUUGCC GAGGAGAUCGGGAUCGACCUGGACGACGUGCCCAGCGUG GUGUCCAAGCUGGAGAAGAACUGCAACGCCUACGCCGAGG UGAACAAUCAUCAUGGAUUAUUCAGAAGCUGAAUAUCGGUG AGUGUUCCGCCCCUCCCGGCCAGCACAUUGCUGCUGCAGAU CGUGAAUACAGGGAGCGCUGAGGCCAACUGCGGGCUCCAG ACCAUCGUGAAGAGCCUGAAUAAGAUAUACGUGCCUCCU AUCAUCGAGAAUCGGCUGCCCUACUACGACCCCUGGUUCC UGGUGGGCGUGGCCAUUAUCCUGGUGAUCUUCACCGUGG CCAUCUGUAGCAUCAGGCGGAACCUGGCCCUGAAGUACAG GUACGGGACCUUCCUGUACGUGUGA
19	1701 RNA	ORF AUGACCCAGCUGCCCAUCCUGCUGCUGCUGAUCAGCCUGG UGUACGCCACUCCCAGUCCUCAGACCUCCAAGAAGAUUGG CGACGAUGCCACCAUCAGCUGCUCCCGGAUAUAUACCAAC UACUAUGUGGGUGAUGAGCGCCUGGUACAAGGAGCCUAAC AGCAUCAUCCUGCUGGGCCGCAAGAGCGAUGUGCUGUACU UCGAUAACUACACCAAGGACAAGAUCAGCUACGAUAGCCC CUACGACGAUCUGGUGACUACCAUCACCAUCAAGAGUCUG ACUGCCGGCGACGCCGGGACUUAACAUCUGCGCCUUCUUCA UGACCAGCACCAACCAUUGACACCGACAAGGUGGACUACGA GGAGUACAGCAUCGAGCUGAUCGUGAACACCGACAGCGA GAGCACCAUCGACAUAUCCUGUCUGGCAGCACCCCCGAG ACUAUCAGCGAGAAGCCUGAGGACAUCGAUAACAGCAAC UGUUCCAGCGUUUUCGAGAUCACCACCCCGAGCCUAUCA CCGAUAACGUGGACGAUCACACCGACACCGUGACCUACAC CUCCGAUAGCAUCAACACCGUGAAUGCCAGCAGCGGGGAG AGUACUACAGAUGAGAUCCCCGAGCCUAUACCGGAUAAGG AGGAGGAUCACACCGUGACCGACACUGUGAGUUACACCAC AGUGUCUACCUCUCCGGCAUCGUGACUACCAAGAGUACC ACCGACGAUGCCGAUCUGUACGAUACCUACAACGAUAACG ACACAGUGCCUCCCACCACCGUGGGAGGCAGCACACCAG CAUCAGCAACUACAAGACCAAGGAUUUUGUGGGCGGCGG GGGCUCCGGCGGCGGAGGCUCCGCCGCCGUGAAGACCCCC GUGAUUGUGGUGCCUGUGAUCGAUCGGCCCCCUAGCGAG ACAUUUCCCAACGUGCACGAGCACAUAACGAUCAGAAGU UCGAUGACGUGAAGGACAACGAGGUGAUGCAGGAGAAGC GGGACGUGGUGAUCGUGAACGACGAUCCUGACCACUACA AGGACUACGUGUUAUCCAGUGGACCGGGGGCAACAUCCG GGAUGACGACAAGUACACCCACUUCUUCUCCGGCUUCUGU AACACCAUGUGUACAGAGGAGACAAAGCGGAACAUUGCU CGGCACCUGGCCUGUGGGACAGCAAGUUCUUCACUGAGC UGGAGAACAAGAACGUGGAGUAUGUGGUGAUCUUCGAGA

		ACGACAAUGUGAUCGAGGACAUUACCUUCCUCAGGCCUGU GCUGAAGGCCAUCCACGAUAAGAAGAUCGACAUCUGCAG AUGCGGGAGAUAUCACCGGGAAUAAGGUGAAGACCGAG CUGGUGAUCGACAAGGAUCACGCCAUCUUCACCUACACCG GCGGCUAUGAUGUGAGCCUGAGCGCUUACAUAUCCGGG UGACCACCGCCCUGAAUAUCGUGGAUGAGAUAUUAAGU CCGGGGGGGCUGUCCAGCGGGUUCUACUUCGAGAUUGC GGAUCGAGAACGAGAUGAAGAUAACCGGCAGAUCAUGG AUAUUCUGCCAAGUACGUGGAGCACGACCCUCGGCUGGU GGCUGAGCAUCGCUUCGAGACUAUGAAGCCCAACUUCUG AGCAGGAUUGGGACCGUGGCCGCCAAGCGGUACCCAGGGG UGAUGUACACCUUUACCACACCCCUGAUCAGCGGGCGGG GGGCAGCGGCGGGGGCGGCAGCGCUGAGACAAAGGAGUU CAAGACCCUGUACAACCUGUUUAUCGACAGCUAUCUGCAG AAGCUGGCUCAGCACUCUAUUCCCACCAUGUGACAUGCG CCAUUCACAUUGGGGAGGUGAUCGGCCAGUUCAAGAACU GCGCCCUGAGGAUCACCAUAAGUGCAUGAGCAACAGCAG GCUGAGCUUCACCCUGAUGGUGGAGUCCUUCAUCGAGGU GAUCAGCCUGCUGAGCGAGAAGGACAGGAGGGCCAUCGCC GAGGAGAUCGGCAUCGACCUGGACGAUGUGCCCAGCGUG GUGAGCAAGCUGGAGAAGAAUUGUAACGCCUACGCCGAG GUGAACAAUAUCAUCGAUAUUCAGAAGCUGAAUAUCGGG GAGUGCAGUGCUCUUUUUGGGCAGCACAUUGCUGCUGCAG UUGUGAACACCGGCAGCGCCGAGGCCAAUUGCGGGCUGCA GACCAUUGUGAAGUCACUGAAUAAGAUCUACGUGCCACCC AUAUCGAGAACCGGCUGCCAUACUACGAUCCUGGUUCC UGGUGGGGGUGGGCCAUAUUCUGGUGAUCUUCACAGUGG CCAUCUGCAGCAUCAGGCGCAACCUGGCUCUGAAGUACAG GUACGGGACCUUCCUGUACGUGUGA
20	1801 RNA	ORF AUGACCCAGCUGCCUAUCCUGCUGCUGCUGAUCAGCCUGG UGUACGCUACCCCUAGUCCCCAGACUCCAAGAAGAUCGG GGACGACGCCACGAUCAGCUGCAGCAGGAUAACACCAAC UACUACGUGGUGAUGAGCGCCUGGUACAAGGAGCCAAAU AGCAUCAUUCUGCUGGCCGCCAAGUCUGACGUGCUGUACU UCGACAAUUAACAAAGGACAAGAUCUCCUACGACAGUCC UUAUGAUGAUCUGGUGACCACGAUCACCAUUAAGUCUCU GACUGCUGGGGAUGCCGGUACCUAUAUCUGUGCGUUCUU UAUGACCUCUACAACGAACGACACAGAUAAAGGUGGACUA CGAGGAGUAUAGCAUCGAGCUGAUCGUGAACACCGACAG CGAGAGCACAAUUGACAUAUCCUCAGCGGGAGCACCCCC GAGACAAUCUCUGAGAAGCCCCGAGGACAUUGAUAAUAGC AAUUGUAGCUCCGUGUUCGAGAUUACUACUCCCGAGCCCA UUACCGAUAAUGUGGAUGAUAUACAGAUACGGUGACCU

			ACACCAGCGACAGCAUCAACACCGUGAAUGCUAGCAGCGG UGAGAGCACCACCGAUGAGAUCCCGGAGCCUAUACCCGAU AAGGAGGAGGAUCACACCGUGACCGACACGGUGAGUUAC ACCACCGUGUCUACCUCUCCGGGAUCGUGACUACCAAGA GCACCACGGAUGACGCCGAUCUGUAUGACACCUACAACGA UAAUGAUACCGUGCCGCCACCACCGUGGGCGGCAGCACC ACCAGCAUCAGCAACUAUAAGACUAAGGACUUCGUGGGG GGCGGGGGGAGCGGCGGCGGCGGCAGCCGGCUGAAUCAG UGCAUGAGCGCCAACAAGGCCGCCAUCACUGAUUCAGCCG UGGCUGUCGCCGCCGCCAGCUCCACCCACCGGAAGGUGGU GUCCAGCACCAACCAGUACGAUCACAAGGAGUCCUGCAAU GGCCUGUAUUACCAGGGCUCUUGCUACAUCUGCACAGCG AUUACAAGAGCUUUGAGGAUGCCAAGGCCAAUUGUGCAG CCGAGAGCAGCACGCUGCCUAACAAGAGCGACGUGCUGAC UACUUGGCUGAUCGACUAUGUGGAGGACACCUGGGGCAG UGAUGGGAACCCCAUCACUAAGACCACCUCUGACUACCAG GACUCCGACGUGAGUCAGGAGGUGAGGAAGUACUUCUGU ACGGGCGGCGGCGGCAGCGGCGGCGGCGGGUCCACCUGCA CCGUGCCCACCAUGAACAACGCCAAGCUGACUUCUACCGA GACUUCUUUCAACGACAAGCAGAAGGUGACCUUCACCUGC GACAGCGGCUAUCAUUCCUCGACCCUAACGCCGUGUGCG AGACAGACAAGUGGAAGUAUGAGAACCCGUGCAAGAAGA UGUGUACGGUGUCCGACUACGUCUCUGAGCUGUACGACA AGCCUCUGUACGAGGUCAACUCCACUAUGACUCUGUCUUG CAACGGCGAGACAAAGUACUUUAGGUGCGAGGAGAAGAA CGGCAAUACCAGCUGGAACGACACCGUGACGUGCCCCAAC GCCGAGUGUCAGCCUCUCCAGCUGGAGCACGGCAGCUGUC AGCCUGUGAAGGAGAAGUACAGCUUCGGCGAGUACAUGA CCAUCAACUGUGAUGUGGGGUACGAGGUGAUCGGGGUGA GCUAUAUCAGCUGCACCGCCAAUAGCUGGAACGUGAUUCC CAGCUGUCAGCAGAAGUGCGACAUCCCUUCUCUGUCCAAC GGGCUGAUCAGCGGGAGCACGUUCAGCAUUGGCGGGGUG AUCCACCUGAGCUGCAAGUCAGGGUUCACCCUGACUGGCU CACCCAGCAGCACCUCAUCGAUGGCAAGUGGAACCCCAU UCUGCCCACCUGUGUGAGAAGCAACGAGGAGUUCGAUCCU GUGGACGACGGACCAGACGACGAAACCGAUCUGUCCAAGC UGUCCAAGGAUGUGGUGCAGUACGAGCAGGAGAUUCGAGA GCCUGGAGGCCACCUACCACAUCAUCAUCAUGGCACUGAC CAUCAUGGGCGUGAUAUUCUUGAUCAGCAUCAUUGUGCU GGUGUGUUCUUGCGACAAGAACAACGACCAGUACAAGUU UCACAAGCUGCUGCCCUGA
21	1901 RNA	ORF	AUGGACUGGACCUGGAUUCUGUCCUGGUGGCGGCCGCCA CAAGGGUGCACUCCGAUGGGACCUUGUUCCUGGGGACGA

			CGACCUGGCCAUCCCCGCUACGGAGUUCUUCAGCACCAAG GCUGCCAAGAACCCGGAGACUAAGCGGGAGGCCAUCGUCA AGGCCUACGGAGAUGACAAUGAGGAGACGCUGAAGCAGC GUCUCACCAACCUUGAGAAGAAGAUACCAACAUCACUAC GAAGUUUGAGCAGAUUGAGAAGUGCUGCAAGCGCAAUGA UGAGGUGCUGUUUCGGCUGGAGAACCACGCCGAGACUCU GCGGGCUGCCAUGAUCAGCCUCGCCAAGAAGAUCGACGUG CAGACCGGAAGGCACCCCUAUGAGGGGGGAGGAGGGAGC GGCGGCGGGAGCGGGGCCGCGGCCAGUAUCCAGACCA CCGUGAACACUCUGUCUGAGCGUAUCAGCAGCAAGCUGGA GCAGGAGGCCAACGCCUCUGCUCAGACCAAGUGUGAUUA GAGAUUGGCAAUUUCUAUAUCAGGCAGAACACGGGUGC AACAUACCGUGAAGAAUAUGUGCUCGCGUGACGCCGACG CUCAGCUGGACGCCGUGCUGAGCGCGGCCACGGAGACAU UCCGGUCUGACCCCUAGCAGAAAGGCGUAUGUGCCUGCC AUGUUCACAGCUGCGCUGAAUAUCCAGACGAGUGUUAAC ACGGUGGUGCGGGACUUUGAGAACUACGUGAAGCAGACG UGCAACUCCAGCGCCGUGGUGGAUAAUAAGCUCAAGAUC AGAAUGUGAUCAUUGACGAGUGUUAUGGGGCCCCUGGA GCCCCACUAACCUCGAGUUUAUUAUACCGGUUCCUCAA GGGGAACUGCGCCAUCAAGGCGCUGAUGCAGUUGACAAC GAAGGCCACCACCAGAUCGCCCCCAGGCAGGUUGCCGGG GGCGGCGGGGGGAGCGGUGGCGGCGGCAGCGCUGAGACA AAGGAGUUCAAGACCCUGUACAACUUGUUAUAGAUAUCU UAUCUUCAGAAGUUGGCCCAGCACUCCAUCCCUACCAAUG UCACUUGUGCCAUCCACAUUGGAGAGGUGAUCGGCCAGU UCAAGAACUGCGCCCUGAGGAUCACCAACAAGUGCAUGAG CAAUAGCCGCCUGUCCUUCACGCUCAUGGUGGAGUCCUUC AUCGAGGUGAUCUCCUGCUGAGCGAGAAGGACAGGCGG GCUAUUGCUGAGGAGAUCGGUAUUGACCUCGACGACGUG CCGUCCGUCGUCUCCAAGCUGGAGAAGAACUGCAACGCAU AUGCUGAGGUGAACAACAUAUCGAUAUUCAGAAGCUGA AUAUCGGUGAGUGUCCGCUCCCCCUGGGCAGCAUAUGCU GUUGCAGAUUGUCAAUACCGGGAGUGCUGAGGCCAACUG CGGGCUGCAGACCAUCGUGAAGUCACUGAAUAAGAUCUA CGUGCCACCAUCAUCGAGAACAGGCUGCCGUACUACGAC CCGUGGUUCUUGGUGGGGGUGGCCAUUAUUCUGGUGAUC UUCACGGUGGCCAUCUGCAGCAUCCGCCGGAACCUGGCUC UGAAGUACCGGUACGGGACCUUCCUGUACGUGUGA
22	2001 RNA	ORF	AUGGAUUGGACCUGGAUUCUGUCCUCGUUGCCGCCGCCA CCAGGGUGCAUAGUGCGGCAGUGAAGACUCCCGUUAUCG UCGUCCCGGUGAUUGACCGCCCGCCUCGGAGACGUUCCC CAACGUGCAUGAGCACAUCAACGAUCAGAAGUUCGACGA

	UGUCAAGGACAAUGAGGUCAUGCAGGAGAAGCGGGACGU GGUGAUCGUCAACGACGAUCCCGAUCACUACAAGGAUUAC GUCUUCAUCCAGUGGACGGGCGGCAACAUCCGGGACGACG AUAAGUACACGCAUUUCUUCUCCGGCUUCUGUAACACCAU GUGUACAGAGGAGACGAAGCGGAACAUCGCCAGGCAUCU GGCGCUGUGGGACUCGAAGUUCUUCACCGAGCUGGAGAA UAAGAACGUGGAGUACGUGGUGAUCaucGAGAACGACAA CGUCAUCGAGGACAUCACGUUUCUGCGGCCGGUGCUGAAG GCCAUCCACGACAAGAAGAUCGACAUCCUGCAGAUGC GGGAGAUAUCACGGGUAAUAAGGUGAAGACCGAGCUGG UCAUUGACAAGGACCACGCGAUAUUCACCUACACUGGGGG GAUACGACGUGUCCUUGUCCGCCUAUAUUAUCCGGGUG ACCACGCUCUGAACAUCGUGGAUGAGAUCAUCAAGUCGG GAGGCCUCAGCUCUCCGGCUUCUACUUCGAGAUCCGCCG GAUCGAGAACGAGAUGAAGAUCAACCGCCAGAUCAUGG ACAACAGCGCCAAGUAUGUCGAGCACGACCCUCGGCUG GUGGCAGAGCACAGAUUCGAGACUAUGAAGCCCAACUUC UGGAGUAGGAUCGGCACUGUCGCCGCCAAGCGCUACCC AGGUGUGAUGUACACCUUCACCACCCCGCUGAUCAGCG GGGGGGGGGGUGAGCGGCGGAGACAAAGGAGUUCAAG ACCCGUGUACAACUUGUUCAUCGACUCCUACCUCAGAA GUUGGCUAGCACUCGAUCCCCACCAUGUGACAUGC GCCAUUCACAUUGGUGAGGUGAUCGGGCAGUUCAAGA ACUGUGCUCUGCGGAUUACCAACAAGUGUAUGAGUAA UCCCCGCCUGUCCUUCACGCUCAUGGUGGAGUCCUUC AUCGAGGUGAUCUCCUGCUGAGCGGAGAAGGACAGG CGGGCCAUCGCCGAGGAGAUCCGGCAUCGACCUGGAC GAUGUUCGUCUGUCUCCAAGCUGGAGAAGAACUGCA AACGCAUAUGCUGAGGUGAACAACAUCAUCGAUAUUC AGAAGCUGAAUAUCGGUGAGUGUCCGCUCCCCUGGG CAGCAUAUGCUGUUGCAGAUCGUGAAUACCGGGAGU GCGGAGGCAGAACUGCGGGCUCCAGACCAUCGUGA AGAGCCUGAACAAGAUCUACGUCCCGCCUAUCAUCG AGAAUCGGCUGCCGUACUAUGACCCUGGUGGCGGCGG CAGCGGGGGGGGGGUCGGGGGGCCGUCGUUCCAUC CAGACCACGUGAACACUCUGUCUGAGCGUAUCAGCAG CAAGCUGGAGCAGGAGGCCAACGCCUCUGCUCAGACCA AGUGUGAUUAAGAUUGGCAAUUUCUAUAUCAGGCAG AACCACGGGUGCAACAUCACCGUGAAGAAUAUGUGC UCCGCUGACGCCGACGCUAGCUGGACGCGCGGCCAC GAGACAUAUUCCGGUCUGACCCUGAGCAGAAGGCGU AUGUGCCUGCCAUUUUCACAGCUGCGCUGAAUAUCC AGACGAGUGUUAACACGGUGGUGCGGGAAUUUCGAG AAUUAUGUGAAGCAGACCUGUAACAGCUCGGCUGU UGUGGACAACAAGCUCAAGAUC
--	---

			CAGAAUGUGAUCAUUGACGAGUGUUACGGGGCCCCCGGC UCCCCACGAAUCUGGAGUUCAUCAACACGGGCAGCAGCA AGGGCAACUGCGCCAUUAAGGCGCUGAUGCAGCUGACCAC UAAGGCCACAACCCAGAUCGCCCCUAGGCAGGUGGCCGGG ACUGGCGUGCAGUUCUAUAUGAUAGUGAUUUGGCGUGAUC AUCCUGGCCGCCUGUUCAUGUACUACGCCAAGCGCAUGC UCUUUACUUCGACCAACGAUAAGAUAACUUUAUCUUGG CGAAUAAGGAGAAUGUGCAUUGGACGACGUACAUGGACA CCUUCUUUCGCACAUCACCUAUGAUCAUUGCCACGACUGA CAUUCAGAACUGA
23	0201 DNA	ORF	ATGGACTGGACCTGGATTCTGTTCTTGGTGGCGGCCGCCACT AGGGTGCACAGCGACGGGACCCTGTTCCAGGGGACGACG ACCTGGCCATCCCCGCTACCGAGTTCTTCAGCACCAAGGCT GCCAAGAACCCCGAGACTAAGCGGGAGGCCATCGTGAAGG CCTACGGGGACGATAATGAGGAGACGCTGAAGCAGCGTCTC ACCAATCTGGAGAAGAAGATTACTAATATCACCACCAAGTTC GAGCAGATCGAGAAGTGCTGTAAGAGGAATGACGAGGTGCT GTTTCGGCTGGAGAACCACGCCGAGACACTGCGGGCTGCCA TGATCTCTCTTGCCAAGAAGATCGATGTGCAGACCGGCAGG CACCTTATTACCCTTACGACGTGCCCCGATTATGCCGAGTGA
24	0302 DNA	ORF	ATGATGACTCCCGAGAATGACGAGGAGCAGACCAGCGTGTT CTCCGCTACCGTGTACGGGGACAAGATCCAGGGGAAGAATA AGAGGAAGCGGGTGATCGGCCTGTGCATCAGGATCAGCATG GTGATCAGCCTGCTGAGCATGATCACCATGTCTGCATTCTG ATCGTGCGGCTGAATCAGTGCATGAGCGCCAACAAGGCCGC CATCACTGATTACGCCGTGGCCGTGGCCGCCGCTTCTCTAC TCACAGGAAGGTGGTGTCTCCACCACTCAGTACGATCACA AGGAGAGCTGCAATGGCCTGTATTACCAGGGCTCTTGCTACA TCCTGCACAGCGATTACAAGAGCTTTGAGGATGCCAAGGCC AATTGCGCCGCCGAGAGCAGCACTCTGCCTAACAAGAGCGA CGTGCTGACTACTTGGCTGATCGACTATGTGGAGGACACCTG GGGGAGCGATGGGAACCCCATCACCAAGACTACCAGCGACT ATCAGGACAGCGACGTGAGCCAGGAGGTGCGGAAGTACTTC TGCACCTACCCTTACGACGTGCCTGACTACGCCTGA
25	0402 DNA	ORF	ATGACCCAGCTGCCTATCCTGCTGCTGCTGATCAGCCTGGTG TACGCTACCCCTAGTCCCCAGACTTCCAAGAAGATCGGGGA CGACGCCACTATCAGCTGCAGCAGGAATAACACCAACTACT ACGTGGTGATGAGCGCCTGGTATAAGGAGCCTAACAGCATC ATCCTGCTGGCCGCCAAGAGCGATGTGCTGTACTTCGACAA CTATACCAAGGACAAGATCTCTTACGATAGCCCCTACGATGA CCTGGTGACTACCATCACCATCAAGAGTCTGACCGCCGGGG ATGCCGGGACCTATATCTGTGCGTTTTTTATGACCTCTACAAC AAACGACACAGATAAGGTGGACTACGAGGAGTATAGCATCG

			AGCTGATCGTGAACACCGACAGCGAGAGTACCATCGACATC ATCCTGAGCGGGAGTACTCCCGAGACTATCAGCGAGAAGCC AGAGGACATCGACAATAGCAACTGCAGCAGCGTGTTTCGAGA TCACTACTCCTGAGCCTATTACAGATAACGTGGATGATCACA CCGACACCGTGACCTACACCAGCGACAGCATCAACACCGTG AATGCTAGCAGCGGTGAGAGTACCACGGATGAGATCCCCGA GCCCATCACGGATAAGGAGGAGGATCACACCGTGACCGACA CGGTGAGTTACACCACCGTGTCTACCTCCTCAGGGATTGTGA CCACAAAGTCTACCACCGATGATGCCGATCTGTATGACACAT ACAATGACAACGATACCGTGCCGCCCCACCACCGTGGGCGGC AGCACAACATCCATTAGTAATAAGACTAAGGATTTCGTG GAGATCTTTGGGATCACAACCCTGATTATCCTGAGCGCCGTG GCGATCTTCTGTATCACCTACTATATCTGTAATAAGCATCCCA GGAAGTATAAGACCGAGAACAAGGTGTATCCGTACGACGTG CCGGACTACGCCTGA
26	0505 DNA	ORF	ATGAAGACCATCTCCGTGGTGACCCTGCTGTGCGTGCTGCCA GCAGTGGTGTACAGTGGCGGCGGGGGGAGCTACCCCTACGA CGTGCCCGACTACGCCACCTGTACGGTGCCAACCATGAACA ACGCTAAGCTGACCTCCACAGAGACGAGTTTTAATGACAAG CAGAAGGTGACCTTCACCTGCGACAGTGGCTACCACAGCTT GGACCCCAACGCTGTGTGCGAGACAGACAAGTGGAAGTAC GAGAACCCGTGCAAGAAGATGTGCACCGTGAGTGACTACGT CAGCGAGCTGTACGACAAGCCGCTGTACGAGGTCAACTCCA CGATGACACTCTCTTGCAACGGGGAGACAAAGTACTTCAGG TGCGAGGAGAAGAACGGGAATACCAGCTGGAATGATACCGT GACATGCCCCAATGCCGAGTGCCAGCCCCTGCAGCTCGAGC ATGGGAGCTGTCAGCCCGTGAAGGAGAAGTACTCCTTCGGG GAGTACATGACCATCAATTGCGACGTGGGGTACGAGGTGATC GGGGTGAGCTACATCTCTTGACCCGCCAATAGCTGGAACGT GATTCCCAGCTGTCAGCAGAAGTGCGACATCCCTTCTCTGTC CAACGGGCTGATCAGCGGGAGCACGTTTCAGCATTGGCGGGG TGATCCACCTGAGCTGCAAGAGTGGGTTCACCCTGACCGGG TCCCCGAGCAGCACCTGCATCGATGGGAAGTGGAACCCGAT CCTGCCAACTTGCGTCCGGTCCAACGAGGAGTTCGACCCCCG TGGACGACGGACCAGACGACGAAACCGATCTGAGCAAGTT GAGCAAGGACGTGGTGCAGTACGAGCAGGAGATCGAGAGC CTGGAGGCCACCTACCACATCATCATGCGACTGACCATC ATGGGCGTGATATTCTTGATCAGCATCATTGTGCTGGTGTGTT CTTGCGACAAGAACAACGACCAGTACAAGTTTCACAAGCTG CTGCCCTGA
27	0603 DNA	ORF	ATGGACTGGACCTGGATTCTGTTCTGCTGGTGGCGGCCGCTACT CGGGTGCATAGCGGCGGCGGGGGGTCTACCCCTACGACGT GCCCCGACTATGCCGCCGAGACAAAGGAGTTTAAGACCCTGT

			ACAACCTGTTTATCGACAGCTATCTGCAGAAGCTGGCTCAGC ACTCTATTCCCACCAATGTGACATGCGCCATTACATTGGTG AGGTGATCGGCCAGTTCAAGAACTGCGCCCTGAGGATCACG AATAAGTGCATGAGCAACAGCAGGCTGAGCTTCACCCTGAT GGTGGAGTCCTTCATCGAGGTGATCAGCCTGCTGAGCGAGA AGGACAGGAGGGCCATCGCCGAGGAGATCGGCATCGACCTG GACGATGTGCCCAGCGTGGTGAGCAAGCTGGAGAAGAAGT GCAACGCATATGCTGAGGTGAACAACATCATCGATATTGAGA AGCTGAATATCGGTGAGTGTTCCGCTCCCCCTGGGCAGCATA TGCTGTTGCAGATCGTGAACACGGGCAGTGCCGAGGCCAAC TGCGGGCTGCAGACCATCGTGAAGTCACTGAATAAGATCTAC GTGCCACCCATCATCGAGAACCGGCTGCCCTACTACGACCCG TGGTTCCTGGTGGGGGTGGCCATTATTCTGGTGATCTTCACG GTGGCCATCTGCAGCATCCGCAGGAACCTGGCCCTGAAGTA CAGGTACGGGACCTTCCTGTACGTCTGA
28	0702 DNA	ORF	ATGGACTGGACCTGGATTCTGTTCTGGTGGCTGCCGCCACC AGGGTGACAGTGGCGGAGGAGGCAGCTACCCCTACGACGT CCCCGACTACGCTGGGGCTGCTGCCTCCATCCAGACCACTGT GAACACCCTGAGTGAGCGGATCAGCAGCAAGCTGGAGCAG GAGGCCAACGCCTCTGCTCAGACCAAGTGTGATATAGAGATT GGCAATTTCTATATCAGGCAGAACCACGGGTGCAACATCACC GTGAAGAATATGTGCTCCGCTGACGCCGACGCTCAGCTGGA CGCCGTGCTGAGCGCGGCCACGGAGACATATTCCGGTCTGA CCCCTGAGCAGAAGGCGTATGTGCCTGCCATGTTACAGCT GCACTGAACATCCAGACCAGCGTGAACACCGTGGTGAGGG ACTTCGAGAACTACGTGAAGCAGACATGCAACAGCAGCGCG GTGGTGGATAATAAGCTCAAGATCCAGAATGTGATCATTGAC GAGTGTTATGGGGCCCCCTGGGAGCCCCACTAACCTCGAGTTT ATTAATACGGGCAGCAGCAAGGGCAACTGCGCCATCAAGGC CCTGATGCAGCTGACAACCAAGGCCACCACCCAGATTGCCC CCAGGCAGGTGGCCGTTGA
29	0803 DNA	ORF	ATGGACTGGACCTGGATTCTGTTTCTGGTGGCCGCCGCCACA CGGGTGCAATCCGGCGGCGGCGGCAGCTACCCATACGATGT GCCTGACTACGCCGCCCGCTGAAGACACCCGTGATCGTGG TGCCCGTGATTGACCGGCCCCCTAGCGAGACATTCCCCAATG TGCACGAGCACATCAACGACCAGAAGTTCGACGACGTGAA GGACAACGAGGTGATGCAGGAGAAGAGGGGACGTGGTGATT GTGAACGACGATCCTGACCACTACAAGGACTACGTGTTTCATC CAGTGGACAGGCGGGAACATCAGGGACGACGACAAGTATAC CCACTTCTTCTCCGGCTTCTGTAACACCATGTGTACAGAGGA GACAAAGCGGAACATTGCTCGGCACCTGGCCCTGTGGGACA GCAAGTTCTTCACCGAGCTGGAGAACAAGAACGTGGAGTAT GTGGTGATCATCGAGAACGACAACGTCATCGAGGACATCAC

		CTTTCTGCGCCCTGTGCTGAAGGCCATCCACGATAAGAAGAT CGACATCCTGCAGATGCGGGAGATCATCACCGGGAATAAGG TGAAGACCGAGCTGGTGATCGACAAGGATCACGCCATCTTC ACCTACACCGGCGGCTATGATGTGAGCCTGAGCGCCTACATC ATCCGGGTGACCACCGCCCTGAATATCGTGATGAGATCATC AAGTCCGGCGGGCTGAGCAGCGGCTTCTACTTCGAGATCGC CAGAATCGAGAACGAGATGAAGATCAACAGGCAGATCATGG ACAACAGCGCCAAGTACGTGGAGCACGACCCTCGGCTGGTG GCCGAGCACCGGTTTCGAGACTATGAAGCCCAACTTCTGGTC CAGAATCGGGACTGTGGCCGCCAAGCGGTACCCCGGCGTGA TGTACACATTCACCACCCCACTGATCAGCTTCTTTGGCCTGT TCGACATCAACGTGATCGGCCTGATCGTGATCCTGTTCATCAT GTTTCATGCTGATCTTCAACGTGAAGTCCAAGCTGCTGTGGTT CCTGACCGGCACCTTCGTGACTGCCTTCATCTGA
30	0201-J0 ORF DNA	ATGGACTGGACCTGGATTCTCTTCTTGGTGGCAGCAGCCACG CGAGTCCACTCCATGGACGGAACCTCTTTTCCCCGGAGATGA CGATCTTGCAATTCCAGCAACTGAATTTTTCTCTACAAAGGC TGCTAAAAATCCAGAGACTAAACGCGAAGCAATTGTAAAG CCTATGGAGACGATAATGAGGAACTCTCAAACAACGGCTA ACTAATTTGGAAAAAAGATTACTAATATAACAACAAAGTTT GAACAAATAGAAAAGTGTTGTAAACGCAACGATGAAGTTCT ATTTAGGTTGGAAAATCACGCTGAACTCTAAGAGCGGCTAT GATATCTCTGGCTAAAAAGATTGATGTACAGACTGGACGGCA TCCATATGAGTAA
31	0301-J0 ORF DNA	ATGGACTGGACCTGGATTCTCTTCTTGGTGGCAGCAGCCACG CGAGTCCACTCCATGATGACACCAGAAAACGACGAAGAGCA GACATCTGTGTTCTCCGCTACTGTTTACGGAGACAAAATTCA GGGAAAGAATAAACGCAACGCGTGATTGGTCTATGTATTAG AATATCTATGGTTATTTCACTACTATCTATGATTACCATGTCCG CGTTTCTCATAGTGCGCCTAAATCAATGCATGTCTGCTAACA AGGCTGCTATTACTGACTCCGCTGTTGCCGTTGCTGCGGCAT CATCTACTCATAGAAAGGTTGTGTCTAGCACTACACAATATG ATCACAAAGAAAGCTGTAATGGTTTATATTACCAGGGTTCTT GTTATATATTACATTCAGACTATAAGTCATTTCGAGGATGCTAA AGCAAACCTGCGCTGCGGAATCATCAACACTACCCAATAAATC CGATGTCTTGACTACCTGGCTCATTGATTATGTTGAGGATACA TGGGGATCTGATGGTAATCCAATTACAAAACTACATCCGAT TATCAAGATTCTGATGTATCACAAGAAGTTAGAAAGTATTTTT GTACATAA
32	0501-J0 ORF DNA	ATGGACTGGACCTGGATTCTCTTCTTGGTGGCAGCAGCCACG CGAGTCCACTCCATGAAAACGATTTCCGTTGTTACGTTGTTA TGCGTACTACCTGCTGTTGTTTATTCAACATGTACTGTACCCA CTATGAATAACGCTAAATTAACGTCTACCGAAACATCGTTTAA

		TGATAAACAGAAAGTTACGTTTACATGTGATTCAGGATATCAT TCTTTGGATCCAAATGCTGTCTGCGAAACAGATAAATGGAAA TACGAAAATCCATGCAAGAAAATGTGCACAGTTTCTGATTAT GTCTCTGAACTATATGATAAGCCATTATACGAAGTGAATTCCA CCATGACACTAAGTTGCAACGGTGAAACAAAATATTTTCGTT GTGAAGAAAAAAATGGAAATACTTCTTGGAATGATACTGTCA CGTGTCCCTAATGCGGAATGTCAACCTCTTCAATTAGAACACG GATCGTGTCAACCAGTTAAAGAAAAATACTCATTTGGGGAAT ATATGACTATCAACTGTGATGTTGGATATGAGGTTATTGGTGT TTCGTATATAAGTTGTACGGCTAATTCTTGGAATGTTATTCCAT CATGTCAACAAAAATGTGATATACCGTCCCTATCTAATGGATT AATTTCCGGATCTACATTTTCTATCGGTGGCGTTATACATCTTA GTTGTAAAAGTGGTTTTTACACTAACGGGGTCCCCATCATCCA CATGTATCGACGGTAAATGGAATCCCATACTCCCAACATGTG TACGATCTAACGAAGAATTTGATCCAGTGGATGATGGTCCCG ACGATGAGACAGATCTGAGCAAACCTCTCGAAAGACGTTGTA CAATATGAACAAGAAATAGAATCGTTAGAAGCAACTTATCAT ATAATCATAATGGCGTTGACAATTATGGGTGTCATATTTCTAAT CTCCATTATAGTATTAGTTTGTTCCTGTGACAAAAATAATGAC CAATATAAGTTCCATAAATTGCTACCGTGA
33	0701-J0 ORF DNA	ATGGACTGGACCTGGATTCTCTTCTTGTTGGCAGCAGCCACG CGAGTCCACTCCATGGGTGCCGCAGCAAGCATAACAGACGAC TGTGAATACACTCAGTGAACGTATCTCGTCTAAATTAGAACA AGAAGCGAACGCTAGTGCTCAAACAAAATGTGATATAGAAA TCGGAAATTTTTATATCCGACAAAACCATGGATGTAACATCA CTGTTAAAAATATGTGCTCTGCGGACGCGGATGCTCAGTTGG ATGCTGTGTTATCAGCCGCTACAGAAACATATAGTGGATTAA CACCGGAACAAAAGCATACTGACGCTATGTTTACTGCTG CGTTAAACATTCAGACGAGTGTAACACTGTTGTTAGAGATT TTGAAAATTATGTGAAACAGACTTGTAATTCTAGCGCTGTTG TCGATAACAAATTAAAGATACAAAACGTAATTATAGATGAAT GTTACGGAGCCCCAGGATCTCCAACAAATTTGGAATTTATTA ATACAGGATCTAGCAAAGGAAATTGTGCCATTAAGGCGTTGA TGCAATTGACTACTAAGGCCACTACTCAAATAGCACCTAGAC AAGTTGCTGGTACAGGAGTTCAGTTTTATATGATTGTTATCGG TGTATAATATTGGCAGCGTTGTTTATGTACTATGCCAAGCGT ATGCTGTTACATCCACCAATGATAAAATCAAACCTATTTTAG CCAATAAGGAAAACGTCCATTGGACTACTTACATGGACACAT TCTTTAGAACTTCTCCGATGATTATTGCTACCACGGATATACA AAACTGA
34	0901-J0 ORF DNA	ATGGACTGGACCTGGATTCTCTTCTTGTTGGCAGCAGCCACG CGAGTCCACTCCATGCCTCAACAACCTATCTCCTATTAATATAG AAACTAAAAAAGCAATTTCTGACGCGCGATTGAAGACGTTA

		GACATACATTATAATGAGTCGAAACCAACCACTATCCAGAAT ACTGGAAAAC TAGTAAGGATTAATTTTAAAGGAGGATATATA AGTGGAGGGTTTCTCCCAATGAATATGTGTTATCAACAATA CATATATATTGGGGAAAGGAAGACGATTATGGATCCAATCACT TGATAGATGTATACAAATACTCTGGAGAGATTAATCTGGTTCA TTGGAATAAGAAAAAATATAGTTCTTATGAAGAGGCAAAAA AACACGATGATGGAATTATCATTATTGCTATATTCTTACAAGTA TCAGATCATAAAAATGTATATTTTCAAAAGATAGTCAATCAAT TGGATTCCATTAGATCCGCTAATATGTCTGCGCCATTTGATT AGTATTTTATCTAGACAATTTGCTACCTAGTACGTTGGATTATT TTACATATCTAGGAACAACCTATCAACCACTCTGCAGACGCAG CATGGATAATTTTCCAACACCAATAAACATTCTGATCA ACTATCTAAATTCAGAACACTATTGTCGTCGCTAATCATGAG GGTAAACCTCATTATATAACAGAGAATTATAGAAATCCGTATA AATTGAACGACGACACGCAAGTATATTATTCTGGGGAGATTA TACGAGCTGCAACCACCTCACCAGTTCGAGAGAACTATTTTA TGAAATGGTTGTCTGATTTGAGAGAGGCATGTTTTTCATATTA TCAAAAATATATCGAAGGGAATAAAACATTTCGCAATTATTGC CATAGTATTCGTATTTATACTCACCCTATTCTCTTCTTATGA GTCAACGATATTTCGCGAGAAAAACAAAAC TAG
35	1101 DNA	ORF ATGGACTGGACGTGGATTCTCTTCCTGGTGGCAGCCGCCACT CGTGTGCACTCCGATGGGACCTTGTTCCCTGGGGACGACGA CCTGGCCATCCCCGCTACGGAGTTCTTCAGCACCAAGGCTG CCAAGAACCCGGAGACTAAGCGGGAGGCCATCGTCAAGGC CTACGGAGATGACAACGAGGAGACGCTGAAGCAGCGGCTG ACCAACCTGGAGAAGAAGATCACGAACATCACGACCAAGTT CGAGCAGATTGAGAAGTGCTGCAAGCGGAACGACGAGGTG TTGTTCCGCTTGGAGAACCACGCAGAGACTCTGCGGGCAGC CATGATCAGCTTGGCGAAGAAGATCGACGTGCAGACTGGAC GCCACCCCTATGAGGGAGGAGGCGGCAGCGGCGGGCGGCGG CAGCCGGCTGAATCAGTGCATGAGCGCCAACAAGGCCGCCA TCACTGATTCAAGCCGTCGCCGTCGCTGCCGCCTCCTCCACTC ATAGGAAGGTGGTGTCCAGTACGACACAGTACGACCATAAG GAGAGCTGTAATGGGCTGTACTACCAGGGGTCGTGCTACATC CTGCACAGCGACTACAAGTCGTTTGAGGATGCGAAGGCGAA CTGCGCAGCCGAGAGCAGCACGTTGCCTAACAAGAGCGAC GTGCTGACTACTTGGCTGATTGACTATGTCGAGGATACCTGG GGCTCGGATGGGAACCCCATCACGAAGACCACGTCTGACTA CCAGGACTCCGATGTGTCTCAGGAGGTGCGGAAGTACTTCT GTACTGGCGGCGGCGGTTCTGGTGGGGGCGGCAGCACCTGC ACCGTCCCCACCATGAACAACGCCAAGCTGACCAGTACAGA GACTTCCTTCAACGATAAGCAGAAGGTGACATTTACATGCGA TTCCGGCTACCACTCGTTGGACCCCAACGCTGTCTGTGAGA

		CAGACAAGTGGAAGTACGAGAATCCATGTAAGAAGATGTGC ACGGTGTCCGACTATGTTTCTGAGCTGTACGACAAGCCCCTT TACGAGGTGAACAGCACCATGACTCTCAGCTGCAACGGCGA GACTAAGTACTTTTCGGTGTGAGGAGAAGAACGGTAATACGT CCTGGAACGATACCGTTACTTGCCCCAACGCCGAGTGCCAG CCGTTGCAGCTGGAGCATGGTAGCTGTCAGCCCGTGAAGGA GAAGTATAGTTTTCGGCGAGTACATGACCATTAATTGTGATGT GGGTTACGAGGTCATCGGGGTGAGTTACATCTCTTGCACTGC CAATAGCTGGAACGTGATTCCCAGCTGTCAGCAGAAGTGTG ACATACCTTCTCTGTCCAACGGGGCTGATCAGCGGGAGCACG TTCAGCATTGGCGGGGTGATCCACCTCAGCTGCAAGAGCGG GTTACGTTGACCGGGTCGCCCTCGTCGACCTGCATCGACG GGAAGTGGAATCCCATTCTGCCCACGTGCGTGCGGTGCAAC GAGGAGTTCGACCCGGTCGACGATGGACCCGATGACGAGAC TGATCTGTCTAAGTTGTGCAAGGACGTCGTTACGTACGAGCA GGAGATCGAGTCCCTGGAGGCCACGTACCACGGCGGTGGCG GGAGCGGGGGAGGTGGCAGCAGTCCCCAGACTTCCAAGAA GATCGGGGACGACGCCACCATCTCCTGCTCCCGCAACAACA CTAACTACTACGTGGTTATGTCCGCCTGGTACAAGGAGCCCA ACTCGATCATCCTGCTCGCTGCTAAGAGCGACGTCCTGTACT TCGACAACCTATACCAAGGACAAGATCAGCTACGACTCACCC TACGATGACCTCGTGACCACCATCACAATTAAGAGTCTGACC GCCGGGGACGCCGGCACATACATTGCGCCTTCTTCATGACG AGTACCACCAATGATACCGACAAGGTGGATTATGAGGAGTAC TCGATCGAGCTTATCGTGAACACGGACAGTGAGTCGACCAT CGACATCATCCTGTCCGGTAGCACCCCTGAGACAATCTCCGA GAAGCCGGAGGACATCGACAATAGCAATTGCAGTTCCGTCT TCGAGATCACGACCCCTGAGCCCATTACCGACAACGTGGAT GATCACACCGACACGGTGACGTACACCAGCGACTCTATTAA CACCGTCAATGCTTCCAGTGGGGAGTCCACCCTGACGAGA TTCCGGAGCCCATACCGACAAGGAGGAGGATCACACCGTG ACCGACACGGTGAGTTACACCACCGTGTCTACCTCCTCTGGT ATTGTGACTACCAAGAGCACCACGGACGACGCCGATCTGTA CGACACGTACAATGACAACGACACCGTGCCGCCACCACCG TGGGCGGGTCCACGACGTCCATTAGTAACTACAAGACTAAG GACTTCGTGGGGGGCGGCGGTTCCGGTGGCGGGGGCAGCG CTGAAACCAAGGAGTTCAAGACGCTGTACAACCTGTTTCATC GACTCTTACCTGCAGAAGTTGGCCCAGCACTCCATCCCTACC AATGTCACTTGTGCCATCCACATTGGAGAGGTGATCGGCCAG TTCAAGAACTGCGCCCTGAGGATCACCAACAAGTGCATGAG CAATAGCCGCCTGTCCTTACGCTCATGGTGGAGTCCTTCAT CGAGGTGATCTCCCTGCTGAGCGAGAAGGACAGGCGGGCTA TTGCTGAGGAGATCGGTATTGACCTCGACGACGTGCCGTCC
--	--	---

			GTCGTCTCCAAGCTGGAGAAGAACTGCAACGCATATGCTGA GGTGAACAACATCATCGATATTCAGAAGCTGAATATCGGTGA GTGTTCCGCTCCCCCTGGGCAGCATATGCTGTTGCAGATTGT CAATACCGGGAGTGCTGAGGCCAACTGCGGGTTGCAGACCA TCGTGAAGAGTCTGAACAAGATTTACGTGCCCCCGATCATCG AGAACCGCCTGCCCTACTACGACCCCGGTGGGGGCGGGTCT GGAGGAGGCGGCAGCGGCGCTGCCGCCTCCATCCAGACCAC GGTGAATACGTTGTCCGAGCGGATTTGAGCAAGCTGGAGC AGGAGGCCAACGCCTCTGCTCAGACCAAGTGTGATATAGAG ATTGGCAATTTCTATATCAGGCAGAACCACGGGTGCAACATC ACCGTGAAGAATATGTGCTCCGCTGACGCCGACGCTCAGCT GGACGCCGTGCTGAGCGCGGCCACGGAGACATATTCCGGTC TGACCCCTGAGCAGAAGGCGTATGTGCCTGCCATGTTTACA GCTGCTCTGAACATCCAGACGTCCGTGAATACCGTGGTGAG GGACTTCGAGAACTACGTGAAGCAGACGTGTAATAGCAGCG CTGTGCTCGATAATAAGCTGAAGATCCAGAACGTTATTATCG ACGAGTGCTACGGCGCCCCCGGCTCCCCGACCAACCTGGAG TTCATTAATACCGGTTTCTCCAAGGGGAACTGCGCTATTAAG GCGCTGATGCAGCTGACTACGAAGGCGACGACCCAGATCGC GCCTCGCCAGGTGGCGGGGACCGGGGTGCAGTTCTACATGA TCGTGATCGGTGTGATCATCCTGGCAGCTCTCTTTATGTACTA CGCCAAGCGCATGCTGTTTACCAGCACCAACGACAAGATTA AGCTGATCCTCGCCAATAAGGAGAACGTCCACTGGACCACC TACATGGACACATTCTTCCGCACCAGCCCTATGATTATCGCCA CCACTGACATTCAGAACTGA
36	1201 DNA	ORF	ATGACCCAGCTGCCTATCCTGCTGCTGCTGATCAGTCTGGTG TACGCTACACCGTCTCCCCAGACTTCCAAGAAGATCGGGGA CGACGCGACGATCAGCTGCAGCAGGAATAACACCAACTACT ACGTGGTGATGAGCGCCTGGTACAAGGAGCCGAATAGCATC ATTCTGCTGGCCGCCAAGTCTGACGTGCTGTACTTCGACAAT TACACAAAGGACAAGATCTCCTACGACAGTCCTTATGATGAT CTGGTGACCACGATCACCATTAAGTCTCTGACTGCTGGGGAT GCCGGTACGTATATCTGCGCCTTCTTCATGACGTCCACGACC AATGACACTGATAAGGTGGATTACGAGGAGTACTCCATCGAG CTGATCGTGAACACCGACAGCGAGTCTACCATCGACATCATC CTGTCCGGGAGCACCCCCGAGACTATCAGTGAGAAGCCTGA GGATATTGACAACAGCAATTGCAGTAGTGTCTTCGAGATCAC CACCCCGGAGCCTTACTGACAACGTGGACGATCACACCG ACACGGTGACGTACACCAGTGATAGCATCAACACCGTGAAT GCTAGCTCCGGGGAGTCCACGACCGATGAGATCCCGGAGCC TATCACCGATAAGGAGGAGGATCACACCGTGACCGACACGG TGAGTTACACCACCGTGTCTACCTCCTCCGGGATCGTGACTA CGAAGAGCACTACTGACGATGCTGATCTGTACGATACCTACA

		ACGATAACGATACCGTGCCGCCCACCACCGTGGGCGGCAGC ACCACCAGCATCAGCAACTATAAGACTAAGGACTTCGTGGG GGGCGGGGGGAGCGGCGGCGGCAGCCGGCTGAATCAG TGCATGAGCGCCAACAAGGCCGCCATCACTGATTCAGCCGT GGCTGTGCGCCGCCAGCTCCACCCACCGGAAGGTGGTGT CCAGCACCACCCAGTACGATCACAAGGAGTCCTGCAATGGC CTGTATTACCAGGGCTCTTGCTACATCCTGCACAGCGATTAC AAGAGCTTTGAGGATGCCAAGGCCAATTGTGCAGCCGAGAG CAGCACGCTGCCTAACAAGAGCGACGTGCTGACTACTTGGC TGATCGACTATGTGGAGGACACCTGGGGGAGTGATGGGAAC CCCATCACTAAGACCACGTCCGACTATCAGGACAGTGACGT GAGTCAGGAGGTGCGGAAGTACTTCTGCACCGGCGGCGGG GGCAGCGGCGGGGGCGGCAGCGCTGAGACAAAGGAGTTCA AGACCCTTTACAACCTTGTTTCATCGATAGTTATCTGCAGAAGC TGGCCCAGCACAGCATCCCCACCAATGTGACATGCGCCATTC ACATTGGTGAGGTGATTGGCCAGTTCAAGAACTGCGCCCTG CGGATAACTAACAAGTGTATGTCTAATTCCCGCCTGTCCTTCA CGCTCATGGTGGAGTCCTTCATCGAGGTGATCTCCCTGCTGA GCGAGAAGGACAGGCGGGCTATCGCCGAGGAGATCGGCATC GACCTGGATGACGTCCCCAGCGTGGTGTCCAAGCTGGAGAA GAACTGCAACGCATATGCTGAGGTGAACAACATCATCGATAT TCAGAAGCTGAATATCGGTGAGTGTTCCGCTCCCCCTGGGCA GCATATGCTGTTGCAGATCGTCAATACCGGGTCTGCCGAGGC CAACTGCGGGCTGCAGACCATCGTGAAGTCACTGAATAAGA TCTACGTGCCACCCATCATCGAGAACCGGCTGCCCTACTACG ACCCGTGGTTCCTGGTGGGGGTGGCCATTATTCTGGTGATCT TCACGGTGGCCATCTGCAGCATCCGCAGGAACCTGGCCCTG AAGTACCGGTACGGTACCTTCCTCTACGTGTGA
37	1301 DNA	ORF ATGGACTGGACCTGGATTCTGTTCCCTGGTGGCGGCCGCCAC AAGGGTGCACCTCCGATGGGACCTTGTTCCCTGGGGACGACG ACCTGGCCATCCCCGCTACGGAGTTCTTCAGCACCAAGGCT GCCAAGAACCCGGAGACTAAGCGGGAGGCCATCGTCAAGG CCTACGGAGATGACAATGAGGAGACGCTGAAGCAGCGTCTC ACCAACCTTGAGAAGAAGATCACCAACATCACTACGAAGTT TGAGCAGATTGAGAAGTGCTGCAAGCGCAATGATGAGGTGC TGTTTCGGCTGGAGAACCACGCCGAGACTCTGCGGGCTGCC ATGATCAGCCTCGCCAAGAAGATCGACGTGCAGACCGGAAG GCACCCCTATGAGGGGGGAGGAGGGAGCGGCGGCGGCGGG AGCGGGGCCGCGGCCAGTATCCAGACCACCGTGAACACTCT GTCTGAGCGTATCAGCAGCAAGCTGGAGCAGGAGGCCAAC GCCTCTGCTCAGACCAAGTGTGATATAGAGATTGGCAATTTC TATATCAGGCAGAACCACGGGTGCAACATACCGTGAAGAA TATGTGCTCCGCTGACGCCGACGCTCAGCTGGACGCCGTGC

			TGAGCGCGGCCACGGAGACATATTCCGGTCTGACCCCTGAG CAGAAGGCGTATGTGCCTGCCATGTTACAGCTGCGCTGAAT ATCCAGACGAGTGTTAACACGGTGGTGCGGGACTTTGAGAA CTACGTGAAGCAGACGTGCAACAGCTCAGCGGTGGTGGATA ATAAGCTCAAGATCCAGAATGTGATCATTGACGAGTGTTATG GGGCCCCTGGGAGCCCCACTAACCTCGAGTTTATTAATACGG GCAGCAGCAAGGGCAACTGCGCTATCAAGGCGCTGATGCAG CTGACAACGAAGGCCACCACCAGATTGCTCCGCGGCAGGT GGCAGGAGGGGGAGGGGGGTCCGGCGGGGGAGGCTCGACT TGCACGGTGCCGACGATGAATAACGCCAAGCTGACCTCCAC CGAGACGTCGTTCAACGACAAGCAGAAGGTGACGTTACCT GCGACAGCGGCTATCACAGCTTGGACCCCAACGCCGCTGC GAGACGGACAAGTGGAAGTACGAGAACCCGTGCAAGAAGA TGTGCACGGTGTCCGACTATGTGTCCGAGCTGTACGATAAGC CGCTGTACGAGGTGAACAGCACCATGACGCTGTCGTGCAAC GGCGAGACAAAGTACTTCCGGTGCGAGGAGAAGAACGGGA ACACCAGCTGGAACGATACAGTGACCTGTCCCAACGCAGAG TGCCAGCCTCTGCAGTTGGAGCACGGGTCTGTGAGCCCGT TAAGGAGAAGTACAGCTTTGGCGAGTACATGACGATTAAC GCGACGTCGGTTACGAGGTGATCGGCGTGAGTTACATCTCAT GTACCGCCAACAGCTGGAATGTCATACCTAGCTGCCAGCAG AAGTGTGACATTCCGTCTCTCTCCAACGGGCTGATCAGCGG GTCCACGTTCTCTATCGGCGGCGTTATTCATCTGTCTGCAAG TCGGGCTTCACCCTGACCGGGTCCCCCTCTTCCACTTGATC GACGGCAAGTGGAACCCCATCCTGCCCACCTGCGTGCGGAG CAACGAGGAGTTCGACCCCGTGGATGACGGGCTGATGATG AGACTGATCTCTCTAAGCTGTCTAAGGACGTGGTCCAGTACG AGCAGGAGATCGAGTCGCTGGAGGCCACCTACCACATCATC ATCATGGCCCTCACCATCATGGGGGTGATCTTCTTGATCAGC ATCATCGTGCTGGTGTGTTCTTGCGACAAGAACAACGACCA GTACAAGTTTCACAAGCTGCTGCCTTGA
38	1401 DNA	ORF	ATGAAGACCATCTCCGTGGTGACCCTGCTGTGCGTGCTGCCA GCAGTGGTGTACAGCACCTGTACGGTGCCGACGATGAATAA CGCCAAGCTGACCTCCACCGAGACGTCGTTCAACGACAAGC AGAAGGTGACGTTACCTGCGACAGCGGCTATCACAGCTTG GACCCCAACGCCGTCTGCGAGACGGACAAGTGGAAGTACG AGAACCCGTGCAAGAAGATGTGCACGGTGTCCGACTATGTG TCCGAGCTGTACGATAAGCCGCTGTACGAGGTGAACAGCAC CATGACGCTGTCGTGCAACGGCGAGACAAAGTACTTCCGGT GCGAGGAGAAGAACGGGAACACCAGCTGGAACGATACAGT GACCTGTCCCAACGCAGAGTGCCAGCCTCTGCAGTTGGAGC ACGGGTCCTGTGAGCCCGTTAAGGAGAAGTACAGCTTTGGC GAGTACATGACGATTAACCTGCGACGTCGGTTACGAGGTGATC

	GGCGTGAGTTACATCTCATGTACCGCCAACAGCTGGAATGTC ATACCTAGCTGCCAGCAGAAGTGTGACATTCGGTCTCTCTCC AACGGGCTGATCAGCGGGTCCACGTTCTCTATCGGCGGCGTT ATTCATCTGTCTTGCAAGTCCGGCTTCACGTTGACCGGGTCG CCCTCGTCGACCTGCATCGACGGGAAGTGGAATCCCATTCTG CCCACGTGCGTGCGGTGCAACGAGGAGTTCGACCCGGTTCGA CGATGGGCCGGACGACGAGACAGATCTGTCAAAGTTGAGCA AGGATGTGGTTCAGTATGAGCAGGAGATCGAGAGCCTGGAG GCCACGTACCACGGGGGAGGGGGGAGCGGGCGGCGGGGGA GCGGGGCCGCGGCCAGTATCCAGACCACCGTGAACACTCTG TCTGAGCGTATCAGCAGCAAGCTGGAGCAGGAGGCCAACGC CTCTGCTCAGACCAAGTGTGATATAGAGATTGGCAATTTCTAT ATCAGGCAGAACCACGGGTGCAACATCACCGTGAAGAATAT GTGCTCCGCTGACGCCGACGCTCAGCTGGACGCCGTGCTGA GCGCGGCCACGGAGACATATTCCGGTCTGACCCCTGAGCAG AAGGCGTATGTGCCTGCCATGTTACAGCTGCGCTGAATATC CAGACGAGTGTTAACACGGTGGTGCGGGACTTTGAGAACTA CGTGAAGCAGACGTGCAACTCCAGCGCCGTGGTGGATAATA AGCTCAAGATCCAGAATGTGATCATTGACGAGTGTTATGGGG CCCCTGGGAGCCCCACTAACCTCGAGTTTATTAATACCGGTT CCTCCAAGGGGAAC TGCGCCATCAAGGCGCTGATGCAGTTG ACAACGAAGGCGACGACCCAGATCGCGCCTCGGCAGGTGG CCGGCGGCGGTGGCGGGTCAGGCGGCGGGGGGAGCGCCGC CGTGAAGACGCCGGTGATCGTGGTGCCCGTGATCGACCGGC CTCCGAGCGAGACGTTCCCCAACGTGCATGAGCATATCAAC GATCAGAAGTTCGATGACGTTAAGGACAACGAGGTGATGCA GGAGAAGCGGGACGTGGTGATCGTGAATGACGACCCGGATC ACTACAAGGACTACGTGTTTATCCAGTGGACCGGAGGCAAC ATCCGGGATGACGACAAGTACACCCATTTTTTCTCCGGCTTC TGTAACACCATGTGCACAGAGGAGACGAAGCGGAACATCGC TCGCCACCTGGCACTGTGGGACTCCAAGTTCTTCACAGAGC TGGAGAACAAGAATGTGGAGTACGTCGTCATCATCGAGAAT GATAATGTGATTGAGGACATTACATTTCTCAGGCCGGTGCTC AAGGCCATCCATGACAAGAAGATCGACATCCTGCAGATGCG GGAGATCATCACCGGTAATAAGGTGAAGACTGAGCTGGTCA TTGACAAGGACCACGCCATCTTCACCTATACCGGTGGCTACG ACGTCTCGCTGAGCGCGTATATCATACGCGTGACCACCGCTC TTAATATCGTCGACGAGATCATTAAAGAGCGGTGGTCTGTCCT CGGGCTTCTACTTCGAGATAGCCCGGATCGAGAACGAGATG AAGATCAACCGCCAGATCATGGACAACTCAGCGAAGTACGT CGAGCATGATCCCCGCCTGGTGGCCGAGCACCGGTTTGAGA CAATGAAGCCCAACTTCTGGTCCCGGATAGGCACCGTGGCC GCCAAGCGGTATCCGGGCGTCATGTACACGTTACCACCCCG
--	---

		CTCATCTCATTCTTCGGCCTGTTTCGACATTAACGTCATCGGCC TGATCGTGATACTGTTTCATCATGTTTCATGCTGATCTTCAACGT GAAGAGCAAGTTGCTGTGGTTCCTGACAGGAACCTTCGTGA CTGCCTTCATTTGA
39	1501 DNA	ORF ATGAAGACCATCTCGGTGGTGACACTGCTGTGCGTGTGGCC GGCGGTTGTTTATAGTACGTGCACCGTGCCGACTATGAACAA CGCCAAGTTGACGTCTACTGAGACTTCTTTCAATGACAAGC AGAAGGTGACGTTACCTGCGACAGCGGCTATCACAGCTTG GACCCCAACGCCGTCTGCGAGACGGACAAGTGGAAGTACG AGAACCCGTGCAAGAAGATGTGCACGGTGTCCGACTATGTG TCCGAGCTGTACGATAAGCCGCTGTACGAGGTGAACAGCAC CATGACGCTGTCCTGCAACGGCGAGACAAAGTACTTCCGTT GCGAGGAGAAGAACGGCAACACCAGCTGGAATGACACCGT GACCTGTCCGAACGCCGAGTGCCAGCCCCTGCAGTTGGAGC ATGGCTCCTGCCAGCCCGTGAAGGAGAAGTACTCCTTCGGG GAGTACATGACTATCAACTGCGACGTGGGCTACGAGGTGATC GGGGTGTCTTATATCAGCTGCACCGCCAATAGCTGGAACGTG ATTCCCAGCTGTCAGCAGAAGTGTGACATACCTTCTCTGTCC AACGGGCTGATCAGCGGGAGCACGTTTCAGCATTGGCGGGGT GATCCACCTGAGTTGCAAGTCTGGCTTCACGCTGACCGGATC TCCTAGCTCGACCTGCATCGACGGGAAGTGGAACCCCATCC TGCCAACTTGCGTCCGGTCCAACGAGGAGTTCGACCCCGTG GACGACGGACCAGACGACGAAACCGATCTGAGCAAGTTGA GCAAGGATGTCGTGCAGTACGAGCAGGAGATCGAGTCCCTG GAGGCGACCTATCACGGTGGGGGAGGGAGCGGCGGCGGCG GCTCACGGCTGAATCAGTGCATGAGCGCCAACAAGGCCGCC ATCACTGATTCAGCCGTGGCCGTCGCCGCCGCTTCCTCCACC CACCGGAAGGTCGTCTCCAGCACCAACCCAGTATGATCACA GGAGTCCTGCAACGGGCTGTACTATCAGGGCAGTTGCTACAT CCTGCACAGCGACTACAAGTCGTTTGAGGATGCTAAGGCCA ACTGCGCCGCCGAAAGCTCCACGCTGCCGAATAAGTCGGAT GTCCTGACCACGTGGCTCATCGACTACGTGGAGGACACCTG GGGCTCGGATGGGAACCCCATCACGAAGACCACCAGCGACT ATCAGGACAGCGATGTGAGCCAGGAGGTGCGTAAGTACTTC TGCACCGGGGGCGGAGGGTCTGGAGGAGGCGGCAGCGGCG CTGCCGCTCCATCCAGACCACTGTGAACACCCTGTCCGAG CGCATTTTCGTCCAAGCTGGAGCAGGAGGCCAACGCCTCCGC CCAGACGAAGTGCGACATCGAGATAGGGAATTTCTACATCCG CCAGAACCACGGGTGCAACATCACCGTGAAGAATATGTGCT CCGCTGACGCCGACGCTCAGCTGGACGCCGTGCTGAGCGCG GCCACGGAGACATATTCCGGTCTGACCCCGGAGCAGAAGGC TTACGTGCCTGCCATGTTTACGGCGGCCCTGAATATACAGAC TTCTGTGAATACGGTGGTGCGGGACTTCGAGAATTACGTGAA

			GCAGACTTGCAACTCCAGCGCCGTGGTGGATAATAAGCTCA AGATCCAGAATGTGATCATTGACGAGTGTTATGGGGCCCCTG GGAGCCCCACTAACCTCGAGTTTATTAATACCGGTTCTCCA AGGGGAACTGCGCCATCAAGGCGCTGATGCAGCTGACTACT AAGGCCACCACCCAGATCGCCCCACGTCAGGTGGCTGGCAC CGGCGTTCAGTTCTACATGATCGTGATCGGGGTGATCATTCT GGCTGCGCTGTTTCATGTACTACGCCAAGCGCATGCTCTTTAC TTCGACCAACGATAAGATCAAACCTTATCTTGGCGAATAAGGA GAATGTGCATTGGACGACGTACATGGACACGTTCTTCCGGAC GTCTCCTATGATCATTGCGACCACCGATATTCAGAACTGA
40	1601 DNA	ORF	ATGGACTGGACATGGATTCTGTTCTTGGTGGCCGCCGCCACC AGGGTGCACAGCGACGGCACCCCTGTTCCAGGCGACGACG ACCTGGCCATCCCCGCTACCGAGTTCTTCAGCACCAAGGCT GCCAAGAACCCCGAGACTAAGCGGGAGGCCATCGTGAAGG CCTACGGGGATGATAACGAGGAGACTCTGAAGCAGCGGCTG ACCAACCTGGAGAAGAAGATTACCAACATTACCACCAAGTT CGAGCAGATCGAGAAGTGCTGCAAGCGGAACGATGAGGTG CTGTTTCGGCTGGAGAACCACGCCGAGACACTGCGCGCCGC CATGATCAGCCTGGCCAAGAAGATCGACGTGCAGACCGGCC GCCACCCCTACGAGGGGGGGCGGCGGCTCTGGGGGCGGCGG GAGCAGCCCCCAGACTTCCAAGAAGATCGGGGATGATGCCA CCATCAGCTGCTCCCGGAACAACACCAACTACTACGTGGTG ATGTCCGCCTGGTACAAGGAGCCTAACAGCATCATCCTGCTG GCCGCCAAGAGCGATGTGCTGTACTTTGACAATTACACCAA GGATAAGATCAGCTACGACAGCCCCTACGACGATCTGGTGA CTACCATCACCATCAAGAGCCTGACAGCCGGCGACGCAGGC ACCTACATCTGCGCCTTCTTCATGACCAGCACCAACCAATGAC ACCGACAAGGTGGATTACGAGGAGTACTCCATCGAGCTGAT CGTGAATACTGACTCAGAGTCTACCATCGACATTATCCTGTC CGGGAGCACCCCCGAGACTATCAGTGAGAAGCCTGAGGAC ATCGACAACAGCAACTGCAGCAGTGTGTTTCGAGATCACCAC CCCAGAGCCTATCACTGACAACGTGGATGATCACACCGACA CTGTGACATACACAGTGATAGCATCAACACCGTGAATGCTA GCTCTGGGGAGAGCACCACTGATGAGATCCCCGAGCCTATC ACCGACAAGGAGGAGGATCACACCGTGACCGACACTGTGA GTTACACCACAGTGTCTACCTCCTCTGGGATCGTGACCACCA AGAGCACCAACCGACGATGCCGACCTGTACGATACATAAATG ATAACGATACAGTGCCTCCCACCACCGTGGGAGGCAGCACC ACCAGCATTTCTAATTACAAGACCAAGGACTTCGTGGGGGG CGGGGGGAGCGGGGGCGGGGGGAGCGCCGAAACCAAGGA GTTCAAGACCCTGTACAACCTGTTTCATCGATAGCTACCTGCA GAAGCTGGCCCAGCACAGCATTCCCACCAATGTGACATGCG CCATTACATTGGGGAGGTGATCGGCCAGTTCAAGAACTGC

			GCCCTGAGGATCACCAATAAGTGCATGAGCAATAGCCGCCTG TCCTTCACTCTGATGGTGGAGTCCTTCATCGAGGTGATCTCC CTGCTGTCAGAGAAGGACAGGCGGGCTATTGCCGAGGAGAT CGGGATCGACCTGGACGACGTGCCCAGCGTGGTGTCCAAGC TGGAGAAGAACTGCAACGCCTACGCCGAGGTGAACAACATC ATCGATATTCAGAAGCTGAATATCGGTGAGTGTTCCGCCCT CCCGGCCAGCACATGCTGCTGCAGATCGTGAATACAGGGAG CGCTGAGGCCAACTGCGGGCTCCAGACCATCGTGAAGAGCC TGAATAAGATTTACGTGCCTCCTATCATCGAGAATCGGCTGC CCTACTACGACCCCTGGTTCCTGGTGGGCGTGGCCATTATCC TGGTGATCTTCACCGTGGCCATCTGTAGCATCAGGCGGAACC TGGCCCTGAAGTACAGGTACGGGACCTTCCTGTACGTGTGA
41	1701 DNA	ORF	ATGACCCAGCTGCCCATCCTGCTGCTGCTGATCAGCCTGGTG TACGCCACTCCCAGTCCTCAGACCTCCAAGAAGATTGGCGA CGATGCCACCATCAGCTGCTCCCGGAATAATACCAACTACTA TGTGGTGATGAGCGCCTGGTACAAGGAGCCTAACAGCATCA TCCTGCTGGCCGCCAAGAGCGATGTGCTGTACTTCGATAACT ACACCAAGGACAAGATCAGCTACGATAGCCCCTACGACGAT CTGGTGACTACCATCACCATCAAGAGTCTGACTGCCGGCGA CGCCGGGACTTACATCTGCGCCTTCTTCATGACCAGCACCAC CAATGACACCGACAAGGTGGACTACGAGGAGTACAGCATCG AGCTGATCGTGAACACCGACAGCGAGAGCACCATCGACATC ATCCTGTCTGGCAGCACCCCCGAGACTATCAGCGAGAAGCC TGAGGACATCGATAACAGCAACTGTTCCAGCGTTTTTCGAGAT CACCACCCCCGAGCCTATCACCGATAACGTGGACGATCACAC CGACACCGTGACCTACACCTCCGATAGCATCAACACCGTGA ATGCCAGCAGCGGGGAGAGTACTACAGATGAGATCCCCGAG CCTATCACCGATAAGGAGGAGGATCACACCGTGACCGACAC TGTGAGTTACACCACAGTGTCTACCTCCTCCGGCATCGTGAC TACCAAGAGTACCACCGACGATGCCGATCTGTACGATACCTA CAACGATAACGACACAGTGCCTCCCACCACCGTGGGAGGCA GCACCACCAGCATCAGCAACTACAAGACCAAGGATTTTGTG GGCGGCGGGGGCTCCGGCGGCGGAGGCTCCGCCGCCGTGA AGACCCCCGTGATTGTGGTGCCTGTGATCGATCGGCCCCCTA GCGAGACATTTCCCAACGTGCACGAGCACATCAACGATCAG AAGTTCGATGACGTGAAGGACAACGAGGTGATGCAGGAGA AGCGGGACGTGGTGATCGTGAACGACGATCCTGACCACTAC AAGGACTACGTGTTTCATCCAGTGGACCGGGGGCAACATCCG GGATGACGACAAGTACACCCACTTCTTCTCCGGCTTCTGTAA CACCATGTGTACAGAGGAGACAAAGCGGAACATTGCTCGGC ACCTGGCCCTGTGGGACAGCAAGTTCTTCACTGAGCTGGAG AACAAGAACGTGGAGTATGTGGTGATCATCGAGAACGACAA TGTGATCGAGGACATTACCTTCCTCAGGCCTGTGCTGAAGGC

			CATCCACGATAAGAAGATCGACATCCTGCAGATGCGGGAGAT CATCACCGGGAATAAGGTGAAGACCGAGCTGGTGATCGACA AGGATCACGCCATCTTCACCTACACCGGCGGCTATGATGTGA GCCTGAGCGCTTACATCATCCGGGTGACCACCGCCCTGAATA TCGTGGATGAGATCATTAAGTCCGGGGGGGCTGTCCAGCGGG TTCTACTTCGAGATTGCTAGGATCGAGAACGAGATGAAGATC AACCGGCAGATCATGGATAATTCTGCCAAGTACGTGGAGCAC GACCCTCGGCTGGTGGCTGAGCATCGCTTCGAGACTATGAA GCCCAACTTCTGGAGCAGGATTGGGACCGTGGCCGCCAAGC GGTACCCAGGGGTGATGTACACCTTTACCACACCCCTGATCA GCGGCGGCGGGGGCAGCGGCGGGGGCGGCAGCGCTGAGAC AAAGGAGTTCAAGACCCTGTACAACCTGTTTATCGACAGCT ATCTGCAGAAGCTGGCTCAGCACTCTATTCCCACCAATGTGA CATGCGCCATTACATTGGGGAGGTGATCGGCCAGTTCAAG AACTGCGCCCTGAGGATCACCAATAAGTGCATGAGCAACAG CAGGCTGAGCTTCACCCTGATGGTGGAGTCCTTCATCGAGGT GATCAGCCTGCTGAGCGAGAAGGACAGGAGGGCCATCGCC GAGGAGATCGGCATCGACCTGGACGATGTGCCCAGCGTGGT GAGCAAGCTGGAGAAGAATTGTAACGCCTACGCCGAGGTGA ACAATATCATCGATATTCAGAAGCTGAATATCGGGGAGTGCA GTGCTCCCCCTGGGCAGCACATGCTGCTGCAGATTGTGAAC ACCGGCAGCGCCGAGGCCAATTGCGGGCTGCAGACCATTGT GAAGTCACTGAATAAGATCTACGTGCCACCCATCATCGAGAA CCGGCTGCCATACTACGATCCCTGGTTCCTGGTGGGGGTGGC CATTATTCTGGTGATCTTCACAGTGGCCATCTGCAGCATCAG GCGCAACCTGGCTCTGAAGTACAGGTACGGGACCTTCCTGT ACGTGTGA
42	1801 DNA	ORF	ATGACCCAGCTGCCTATCCTGCTGCTGCTGATCAGCCTGGTG TACGCTACCCCTAGTCCCCAGACTTCCAAGAAGATCGGGGA CGACGCCACGATCAGCTGCAGCAGGAATAACACCAACTACT ACGTGGTGATGAGCGCCTGGTACAAGGAGCCAAATAGCATC ATTCTGCTGGCCGCCAAGTCTGACGTGCTGTACTTCGACAAT TACACAAAGGACAAGATCTCCTACGACAGTCCTTATGATGAT CTGGTGACCACGATCACCATTAAGTCTCTGACTGCTGGGGAT GCCGGTACCTATATCTGTGCGTTCTTTATGACCTCTACAACGA ACGACACAGATAAGGTGGACTACGAGGAGTATAGCATCGAG CTGATCGTGAACACCGACAGCGAGAGCACAATTGACATTAT CCTCAGCGGGAGCACCCCCGAGACAATCTCTGAGAAGCCCG AGGACATTGATAATAGCAATTGTAGCTCCGTGTTTCGAGATTA CTACTCCCGAGCCCATTACCGATAATGTGGATGATCATACAGA TACGGTGACCTACACCAGCGACAGCATCAACACCGTGAATG CTAGCAGCGGTGAGAGCACCAACCGATGAGATCCCGGAGCCT ATCACCGATAAGGAGGAGGATCACACCGTGACCGACACGGT

		GAGTTACACCACCGTGTCTACCTCCTCCGGGATCGTGACTAC CAAGAGCACCACGGATGACGCCGATCTGTATGACACCTACA ACGATAATGATACCGTGCCGCCACCACCGTGGGCGGCAGC ACCACCAGCATCAGCAACTATAAGACTAAGGACTTCGTGGG GGGCGGGGGGAGCGGCGGCGGCGGCAGCCGGCTGAATCAG TGCATGAGCGCCAACAAGGCCGCCATCACTGATTGAGCCGT GGCTGTGCGCCGCCGAGCTCCACCCACCGGAAGGTGGTGT CCAGCACCACCCAGTACGATCACAAGGAGTCCTGCAATGGC CTGTATTACCAGGGCTCTTGCTACATCCTGCACAGCGATTAC AAGAGCTTTGAGGATGCCAAGGCCAATTGTGCAGCCGAGAG CAGCACGCTGCCTAACAAGAGCGACGTGCTGACTACTTGGC TGATCGACTATGTGGAGGACACCTGGGGCAGTGATGGGAAC CCCATCACTAAGACCACCTCTGACTACCAGGACTCCGACGT GAGTCAGGAGGTGAGGAAGTACTTCTGTACGGGCGGCGGCG GCAGCGGCGGCGGCGGGTCCACCTGCACCGTGCCCAACATG AACACGCCAAGCTGACTTCTACCGAGACTTCTTTCAACGA CAAGCAGAAGGTGACCTTCACCTGCGACAGCGGCTATCATT CCCTCGACCCTAACGCCGTGTGCGAGACAGACAAGTGGAAG TATGAGAACCCGTGCAAGAAGATGTGTACGGTGTCCGACTA CGTCTCTGAGCTGTACGACAAGCCTCTGTACGAGGTCAACT CCACTATGACTCTGTCTTGCAACGGCGAGACAAAGTACTTTA GGTGCGAGGAGAAGAACGGCAATACCAGCTGGAACGACAC CGTGACGTGCCCCAACGCCGAGTGTGAGCCTCTCCAGCTGG AGCACGGCAGCTGTGAGCCTGTGAAGGAGAAGTACAGCTTC GGCGAGTACATGACCATCAACTGTGATGTGGGGTACGAGGT GATCGGGGTGAGCTATATCAGCTGCACCGCCAATAGCTGGAA CGTGATTCCCAGCTGTGAGCAGAAGTGCGACATCCCTTCTCT GTCCAACGGGCTGATCAGCGGGAGCACGTTGAGCATTGGCG GGGTGATCCACCTGAGCTGCAAGTCAGGGTTCACCCTGACT GGCTCACCCAGCAGCACCTGCATCGATGGCAAGTGGAACCC CATTCTGCCCACCTGTGTGAGAAGCAACGAGGAGTTCGATC CTGTGGACGACGGACCAGACGACGAAACCGATCTGTCCAAG CTGTCCAAGGATGTGGTGCAGTACGAGCAGGAGATCGAGAG CCTGGAGGCCACCTACCACATCATCATGCGACTGACCAT CATGGGCGTGATATTCTTGATCAGCATCATTGTGCTGGTGTGT TCTTGCGACAAGAACAACGACCAGTACAAGTTTCACAAGCT GCTGCCCTGA
43	1901 DNA	ORF ATGGACTGGACCTGGATTCTGTTCCCTGGTGGCGGCCGCCAC AAGGGTGACTIONCCGATGGGACCTTGTTCCCTGGGGACGACG ACCTGGCCATCCCCGCTACGGAGTTCTTCAGCACCAAGGCT GCCAAGAACCCGGAGACTAAGCGGGAGGCCATCGTCAAGG CCTACGGAGATGACAATGAGGAGACGCTGAAGCAGCGTCTC ACCAACCTTGAGAAGAAGATCACCAACATCACTACGAAGTT

			TGAGCAGATTGAGAAGTGCTGCAAGCGCAATGATGAGGTGC TGTTTCGGCTGGAGAACCACGCCGAGACTCTGCGGGCTGCC ATGATCAGCCTCGCCAAGAAGATCGACGTGCAGACCGGAAG GCACCCCTATGAGGGGGGAGGAGGGAGCGGCGGCGGCGGG AGCGGGGCCGCGGCCAGTATCCAGACCACCGTGAACACTCT GTCTGAGCGTATCAGCAGCAAGCTGGAGCAGGAGGCCAAC GCCTCTGCTCAGACCAAGTGTGATATAGAGATTGGCAATTTC TATATCAGGCAGAACCACGGGTGCAACATCACCGTGAAGAA TATGTGCTCCGCTGACGCCGACGCTCAGCTGGACGCCGTGC TGAGCGCGGCCACGGAGACATATTCCGGTCTGACCCCTGAG CAGAAGGCGTATGTGCCTGCCATGTTACAGCTGCGCTGAAT ATCCAGACGAGTGTTAACACGGTGGTGCGGGACTTTGAGAA CTACGTGAAGCAGACGTGCAACTCCAGCGCCGTGGTGGATA ATAAGCTCAAGATCCAGAATGTGATCATTGACGAGTGTTATG GGCCCCCTGGGAGCCCCACTAACCTCGAGTTTATTAATACCG GTTCCCTCAAGGGGAAGTGCGCCATCAAGGCGCTGATGCAG TTGACAACGAAGGCCACCACCCAGATCGCCCCCAGGCAGGT TGCCGGGGGCGGCGGGGGGAGCGGTGGCGGCGGCAGCGCT GAGACAAAGGAGTTCAAGACCCTGTACAACTTGTTCATAGA TTCTTATCTTCAGAAGTTGGCCCAGCACTCCATCCCTACCAAT GTCACCTGTGCCATCCACATTGGAGAGGTGATCGGCCAGTTC AAGAACTGCGCCCTGAGGATCACCAACAAGTGCATGAGCAA TAGCCGCCTGTCCTTCACGCTCATGGTGGAGTCCTTCATCGA GGTGATCTCCCTGCTGAGCGAGAAGGACAGGCGGGCTATTG CTGAGGAGATCGGTATTGACCTCGACGACGTGCCGTCCGTC GTCTCCAAGCTGGAGAAGAACTGCAACGCATATGCTGAGGT GAACAACATCATCGATATTCAGAAGCTGAATATCGGTGAGTG TTCCGCTCCCCCTGGGCAGCATATGCTGTTGCAGATTGTCAA TACCGGGAGTGCTGAGGCCAACTGCGGGCTGCAGACCATCG TGAAGTCACTGAATAAGATCTACGTGCCACCCATCATCGAGA ACAGGCTGCCGTACTACGACCCGTGGTTCTTGGTGGGGGTG GCCATTATTCTGGTGATCTTCACGGTGGCCATCTGCAGCATCC GCCGGAACCTGGCTCTGAAGTACCGGTACGGGACCTTCCTG TACGTGTGA
44	2001 DNA	ORF	ATGGATTGGACCTGGATTCTGTTCCCTCGTTGCCGCCGCCACC AGGGTGCATAGTGCGGCAGTGAAGACTCCCGTTATCGTCGTC CCGGTGATTGACCGCCCGCCCTCGGAGACGTTCCCCAACGT GCATGAGCACATCAACGATCAGAAGTTCGACGATGTCAAGG ACAATGAGGTCATGCAGGAGAAGCGGGACGTGGTGATCGTC AACGACGATCCCGATCACTACAAGGATTACGTCTTCATCCAG TGGACGGGCGGCAACATCCGGGACGACGATAAGTACACGCA TTTCTTCTCCGGCTTCTGTAACACCATGTGTACAGAGGAGAC GAAGCGGAACATCGCCAGGCATCTGGCGCTGTGGGACTCGA

	AGTTCTTCACCGAGCTGGAGAATAAGAACGTGGAGTACGTG GTGATCATCGAGAACGACAACGTCATCGAGGACATCACGTT TCTGCGGCCGGTGCTGAAGGCCATCCACGACAAGAAGATCG ACATCCTGCAGATGCGGGAGATCATCACGGGTAATAAGGTGA AGACCGAGCTGGTCATTGACAAGGACCACGCGATATTACC TAACTGGGGGATACGACGTGTCCTTGTCCGCCTATATTATCC GGGTGACCACCGCTCTGAACATCGTGGATGAGATCATCAAG TCGGGAGGCCTCAGCTCCGGCTTCTACTTCGAGATCGCCCC GATCGAGAACGAGATGAAGATCAACCGCCAGATCATGGACA ACAGCGCCAAGTATGTCGAGCACGACCCTCGGCTGGTGGCA GAGCACAGATTCGAGACTATGAAGCCCAACTTCTGGAGTAG GATCGGCACTGTCGCCGCCAAGCGCTACCCAGGTGTGATGT ACACCTTCACCACCCCGCTGATCAGCGGGGGGGGGGGTAGC GGCGGCGGCGGCAGTGCCGAGACAAAGGAGTTCAAGACCC TGTACAACTTGTTTCATCGACTCCTACCTCCAGAAGTTGGCTC AGCACTCGATCCCCACCAATGTGACATGCGCCATTACATTG GTGAGGTGATCGGGCAGTTCAAGAACTGTGCTCTGCGGATT ACCAACAAGTGTATGAGTAATCCCGCCTGTCCTTCACGCTC ATGGTGGAGTCCTTCATCGAGGTGATCTCCCTGCTGAGCGAG AAGGACAGGCGGGCCATCGCCGAGGAGATCGGCATCGACCT GGACGATGTTCCGTCTGTCGTCTCCAAGCTGGAGAAGAACT GCAACGCATATGCTGAGGTGAACAACATCATCGATATTCAGA AGCTGAATATCGGTGAGTGTTCCGCTCCCCCTGGGCAGCATA TGCTGTTGCAGATCGTGAATACCGGGAGTGCGGAGGCGAAC TGCGGGCTCCAGACCATCGTGAAGAGCCTGAACAAGATCTA CGTCCCGCCTATCATCGAGAATCGGCTGCCGTACTATGACCC TGGTGGCGGCGGCAGCGGGGGGGGGGGGGTGGGGGGCCGCT GCTTCCATCCAGACCACCGTGAACACTCTGTCTGAGCGTATC AGCAGCAAGCTGGAGCAGGAGGCCAACGCCTCTGCTCAGA CCAAGTGTGATATAGAGATTGGCAATTTCTATATCAGGCAGA ACCACGGGTGCAACATACCGTGAAGAATATGTGCTCCGCT GACGCCGACGCTCAGCTGGACGCCGTGCTGAGCGCGGCCAC GGAGACATATTCGGTCTGACCCCTGAGCAGAAGGCGTATGT GCCTGCCATGTTACAGCTGCGCTGAATATCCAGACGAGTGT TAACACGGTGGTGCGGGATTCGAGAATTATGTGAAGCAGA CCTGTAACAGCTCGGCTGTTGTGGACAACAAGCTCAAGATC CAGAATGTGATCATTGACGAGTGTTACGGGGCCCCCGGCTCC CCCACGAATCTGGAGTTCATCAACACGGGCAGCAGCAAGGG CAACTGCGCCATTAAGGCGCTGATGCAGCTGACCACTAAGG CCACAACCCAGATCGCCCCTAGGCAGGTGGCCGGGACTGGC GTGCAGTTCTATATGATAGTGATTGGCGTGATCATCCTGGCCG CCCTGTTTCATGTACTACGCCAAGCGCATGCTCTTTACTTCGA CCAACGATAAGATCAAACCTTATCTTGGCGAATAAGGAGAATG
--	---

		TGCATTGGACGACGTACATGGACACCTTCTTTTCGCACATCAC CTATGATCATTGCCACGACTGACATTCAGAACTGA
45	0201 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUUGCAGAAGCUC GAAUAAACGCUCAACUUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUGUUCUGGUGGCGGCCGCCACUAGGGUGCA CAGCGACGGGACCCUGUUCCAGGGGACGACGACCUGGCC AUCCCCGCUACCGAGUUCUUCAGCACCAAGGCUGCCAAGA ACCCCGAGACUAAGCGGGAGGCCAUCGUGAAGGCCUACGG GGACGAUAAUGAGGAGACGCUGAAGCAGCGUCUCACCAA UCUGGAGAAGAAGAUUACUAAUAUACACCACCAAGUUCGA GCAGAU CGAGAAGUGCUGUAAGAGGAAUGACGAGGUGCU GUUUCGGCUGGAGAACCACGCCGAGACACUGCGGGCUGCC AUGAUCUCUCUUGCCAAGAAGAUCGAUGUGCAGACCGGC AGGCACCCUUAUUACCCUACGACGUGCCCGAUUAUGCCG AGUGAU AAGAAUUCUAGGCUGGAGCCUCGGUGGCCAUGC UUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUUC CUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGAGU GGGCGGCACUAGUAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAAAAAAAAAA
46	0302 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUUGCAGAAGCUC GAAUAAACGCUCAACUUUGGCGCCGCCACCAUGAUGACUC CCGAGAAUGACGAGGAGCAGACCAGCGUGUUCUCCGCUAC CGUGUACGGGGACAAGA UCCAGGGGAAGAAUAAGAGGAA GCGGGUGAUCGGCCUGUGCAUCAGGAUCAGCAUGGUGAU CAGCCUGCUGAGCAUGAUCACCAUGUCUGCAUUCUGAUC GUGCGGCUGAAUCAGUGCAUGAGCGCCAACAAGGCCGCCA UCACUGAUUCAGCCGUGGCCGUGGCCGCCGCUUCCUCUAC UCACAGGAAGGUGGUGUCCUCCACCACUCAGUACGAUCAC AAGGAGAGCUGCAAUGGCCUGUAUUACCAGGGCUCUUGC UACAUCCUGCACAGCGAUUACAAGAGCUUUGAGGAUGCC AAGGCCAAUUGCGCCGCCGAGAGCAGCACUCUGCCUAACA AGAGCGACGUGCUGACUACUUGGCUGAUCGACUAUGUGG AGGACACCUGGGGGAGCGAUGGGAACCCCAUCACCAAGAC UACCAGCGACUAUCAGGACAGCGACGUGAGCCAGGAGGU GCGGAAGUACUUCUGCACCUACCCUACGACGUGCCUGAC UACGCCUGAU AAGAAUUCUAGGCUGGAGCCUCGGUGGCC AUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCC CCUUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUC UGAGUGGGCGGCACUAGUAAAAAAAAAAAAAAAAAAAAAAAA AA

		AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAA
47	0402 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUUGGCGCCGCCACCAUGACCCAGC UGCCUAUCCUGCUGCUGCUGAUCAGCCUGGUGUACGCUAC CCCUAGUCCCCAGACUUCCAAGAAGAUCGGGGACGACGCC ACUAUCAGCUGCAGCAGGAAUAACACCAACUACUACGUGG UGAUGAGCGCCUGGUUAUAAGGAGCCUAAACAGCAUCAUCC UGCUGGCCGCCAAGAGCGAUGUGCUGUACUUCGACAACUA UACCAAGGACAAGAUCUCUUACGAUAGCCCCUACGAUGAC CUGGUGACUACCAUCACCAUCAAGAGUCUGACCGCCGGGG AUGCCGGGACCUAUUAUCUGUGCGUUUUUUAUGACCUCUA CAACAAACGACACAGAUAAAGGUGGACUACGAGGAGUAUA GCAUCGAGCUGAUCGUGAACACCGACAGCGAGAGUACCAU CGACAUCAUCCUGAGCGGGAGUACUCCCGAGACUAUCAGC GAGAAGCCAGAGGACAUCGACAAUAGCAACUGCAGCAGC GUGUUCGAGAUCACUACUCCUGAGCCUAUUACAGAUAAAC GUGGAUGAUCACACCGACACCGUGACCUACACCAGCGACA GCAUCAACACCGUGAAUGCUAGCAGCGGUGAGAGUACCAC GGAUGAGAUGCCCGAGCCCAUCACGGAUAAGGAGGAGGA UCACACCGUGACCGACACGGUGAGUUACACCACCGUGUCU ACCUCCUCAGGGAUUGUGACCACAAAGUCUACCACCGAUG AUGCCGAUCUGUAUGACACAUACAAUGACAACGAUACCG UGCCGCCCACCACCGUGGGCGGCAGCACAAUCAUUAAG UAACUACAAGACUAAGGAUUUCGUGGAGAUCUUUGGGAU CACAACCCUGAUUAUCCUGAGCGCCGUGGCGAUCUUCUGU AUCACCUACUAUAUCUGUAAUAAGCAUCCAGGAAGUAU AAGACCGAGAACAAGGUGUAUCCGUACGACGUGCCGGAC UACGCCUGAUAAGAAUUCUAGGCUGGAGCCUCGGUGGCC AUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCC CCUUCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUC UGAGUGGGCGGCACUAGUAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAA
48	0505 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUUGGCGCCGCCACCAUGAAGACCA UCUCCGUGGUGACCCUGCUGUGCGUGCUGCCAGCAGUGGU GUACAGUGGCGGCGGGGGGAGCUACCCCUACGACGUGCCC GACUACGCCACCUGUACGGUGCCAACCAUGAACAAACGCUA AGCUGACCUCCACAGAGACGAGUUUUAUGACAAGCAGA AGGUGACCUUCACCUGCGACAGUGGCUACCACAGCUUGGA CCCCAACGCUGUGUGCGAGACAGACAAGUGGAAGUACGA

		GAACCCGUGCAAGAAGAUGUGCACCGUGAGUGACUACGU CAGCGAGCUGUACGACAAGCCGCUGUACGAGGUCAACUCC ACGAUGACACUCUCUUGCAACGGGGAGACAAAGUACUUC AGGUGCGAGGAGAAGAACGGGAUACCAGCUGGAAUGAU ACCGUGACAUGCCCCAAUGCCGAGUGCCAGCCCCUGCAGC UCGAGCAUGGGAGCUGUCAGCCCGUGAAGGAGAAGUACU CCUUCGGGGAGUACAUGACCAUCAAUUGCGACGUGGGGU ACGAGGUGAUCGGGGUGAGCUACAUCUCUUGCACCGCCAA UAGCUGGAACGUGAUUCCAGCUGUCAGCAGAAGUGCGA CAUCCCUUCUCUGUCCAACGGGCUGAUCAGCGGGAGCACG UUCAGCAUUGGCGGGGUGAUCCACCUGAGCUGCAAGAGU GGGUUCACCCUGACCGGGUCCCCGAGCAGCACCUGCAUCG AUGGGAAGUGGAACCCGAUCCUGCCAACUUGCGUCCGGUC CAACGAGGAGUUCGACCCCGUGGACGACGGACCAGACGAC GAAACCGAUCUGAGCAAGUUGAGCAAGGACGUGGUGCAG UACGAGCAGGAGAUCGAGAGCCUGGAGGCCACCUACCACA UCAUCAUCAUGGCACUGACCAUCAUGGGCGUGAUUUCU UGAUCAGCAUCAUUGUGCUGGUGUGUUCUUGCGACAAGA ACAACGACCAGUACAAGUUUCACAAGCUGCUGCCCUGAUA AGAAUUCUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCC CCUUGGGCCUCCCCCAGCCCCUCCUCCCCUUCUGCACCC GUACCCCGUGGUCUUUGAAUAAAGUCUGAGUGGGCGGC ACUAGUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAA
49	0603 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUGUCCUGGUGGCGGCCGCUACUCGGGUGCA UAGCGGCGGCGGGGGGUCCUACCCCUACGACGUGCCCGAC UAUGCCGCCGAGACAAAGGAGUUUAAGACCCUGUACAACC UGUUUAUCGACAGCUAUCUGCAGAAGCUGGCUCAGCACUC UAUUCCACCAAUGUGACAUGCGCCAUUCACAUUGGUGAG GUGAUCGGCCAGUUCAAGAACUGCGCCCUGAGGAUCACGA AUAAGUGCAUGAGCAACAGCAGGCUGAGCUUACCCUGA UGGUGGAGUCCUUCAUCGAGGUGAUCAGCCUGCUGAGCG AGAAGGACAGGAGGGCCAUCGCCGAGGAGAUCGGCAUCG ACCUGGACGAUGUGCCCAGCGUGGUGAGCAAGCUGGAGA AGAACUGCAACGCAUAUGCUGAGGUGAACAACAUAUCG AUAUUCAGAAGCUGAAUAUCGGUGAGUGUCCGCUCCCC UGGGCAGCAUAUGCUGUUGCAGAUUGUGAACACGGGCAG UGCCGAGGCCAACUGCGGGCUGCAGACCAUCGUGAAGUCA CUGAAUAAGAUCUACGUGCCACCCAUCAUCGAGAACCGGC

		UGCCCUACUACGACCCGUGGUUCCUGGUGGGGGUGGCCAU UAUUCUGGUGAUCUUCACGGUGGCCAUCUGCAGCAUCCGC AGGAACCUGGCCCUGAAGUACAGGUACGGGACCUUCCUGU ACGUCUGAUAAAGAAUUCUAGGCUGGAGCCUCGGUGGCCA UGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCC UUCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUG AGUGGGCGGCACUAGUAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAAAAAAAAAAAAAAAA
50	0702 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUGUUCUGGUGGCUGCCGCCACCAGGGUGCA CAGUGGCGGAGGAGGCAGCUACCCCUACGACGUCCCCGAC UACGCUGGGGCUGCUGCCUCCAUCCAGACCACUGUGAACA CCCUGAGUGAGCGGAUCAGCAGCAAGCUGGAGCAGGAGG CCAACGCCUCUGCUCAGACCAAGUGUGAUUAUAGAGAUUG GCAAUUUCUAUAUCAGGCAGAACCACGGGUGCAACAUCAC CGUGAAGAAUAUGUGCUCCGCUGACGCCGACGCUCAGCUG GACGCCGUGCUGAGCGCGGCCACGGAGACAUAUUCCGGUC UGACCCCUAGCAGAAAGGCGUAUGUGCCUGCCAUGUUCAC AGCUGCACUGAACAUCCAGACCAGCGUGAACACCGUGGUG AGGGACUUCGAGAACUACGUGAAGCAGACAUGCAACAGC AGCGCGGUGGUGGAUAAUAAGCUCAAGAUCAGAAUGUG AUCAUUGACGAGUGUUAUGGGGCCCCUGGGAGCCCCACUA ACCUCGAGUUUAUUAUACGGGCAGCAGCAAGGGCAACU GCGCCAUCAAGGCCCUGAUGCAGCUGACAACCAAGGCCAC CACCCAGAUUGCCCCCAGGCAGGUGGCCGGUUGAUAAAGAA UUCUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCU GGGCCUCCCCCAGCCCCUCCUCCCCUCCUGCACCCGUAC CCCCUGGUCUUUGAAUAAAGUCUGAGUGGGCGGCACUA GUAAA AA AA AAAAAA
51	0803 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUGUUCUGGUGGCCGCCGCCACACGGGUGCA UUCCGGCGGGCGGGCGGCAGCUACCCAUACGAUGUGCCUGAC UACGCCGCCGCCGUGAAGACACCCGUGAUCGUGGUGCCCG UGAUUGACCGGCCCCCUAGCGAGACAUUCCCCAAUGUGCA CGAGCACAUCAACGACCAGAAGUUCGACGACGUGAAGGAC AACGAGGUGAUGCAGGAGAAGAGGGACGUGGUGAUUGUG

		AACGACGAUCCUGACCACUACAAGGACUACGUGUUCAUCC AGUGGACAGGCGGGAACAUCAGGGACGACGACAAGUAUA CCCACUUCUUCUCCGGCUUCUGUAACACCAUGUGUACAGA GGAGACAAAGCGGAACAUUGCUCGGCACCUGGCCCUGUGG GACAGCAAGUUCUUCACCGAGCUGGAGAACAAGAACGUG GAGUAUGUGGUGAUCAUUCGAGAACGACAACGUCAUCGAG GACAUCACCUUUCUGCGCCCUGUGCUGAAGGCCAUCCACG AUAAGAAGAUUCGACAUCCUGCAGAUUCGCGGAGAUCAUA CCGGGAUAUAGGUGAAGACCGAGCUGGUGAUCGACAAGG AUCACGCCAUCUUCACCUACACCGGCGGCUAUGAUGUGAG CCUGAGCGCCUACAUCAUCCGGGUGACCACCGCCCUGAAU AUCGUGGAUGAGAUAUCAAGUCCGGCGGGCUGAGCAGC GGCUUCUACUUCGAGAUCCGAGAAUCGAGAACGAGAUG AAGAUAACAGGCAGAUCAUGGACAACAGCGCCAAGUAC GUGGAGCACGACCCUCGGCUGGUGGCCGAGCACCAGGUUCG AGACUAUGAAGCCCAACUUCUGGUCCAGAAUCGGGACUG UGGCCGCCAAGCGGUACCCCGGCGUGAUGUACACAUUCAC CACCACUGAUCAGCUUCUUGGCCUGUUCGACAUAAC GUGAUCGGCCUGAUCGUGAUCCUGUUAUCAUGUUAUG CUGAUCUUAACGUGAAGUCCAAGCUGCUGUGGUUCCUG ACCGGCACCUUCGUGACUGCCUUAUCUGAUAAAGAAUUCU AGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGC CUCCCCCAGCCCCUCCUCCCCUUCUGCACCCGUACCCCC GUGGUCUUUGAAUAAAGUCUGAGUGGGCGGCACUAGUAA AA AA AA AA
52	0201-J0 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCU GAUUAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUCUUCUUGGUGGCAGCAGCCACGCGAGUCCA CUCCAUGGACGGAACUCUUUUUCCCCGGAGAUGACGAUCU GCAAUCCAGCAACUGAAUUUUUCUCUACAAAGGCUGCU AAAAUCCAGAGACUAAACGCGAAGCAAUUGUUAAGCC UAUGGAGACGAUAAUGAGGAAACUCUCAACAACGGCUA ACUAAUUUGGAAAAAAGAUUACUAAUAUAACAACAAG UUUGAACAAAUAGAAAAGUGUUGUAAACGCAACGAUGAA GUUCUAUUUAGGUUGGAAAAUCACGCUGAAACUCUAAGA GCGGCUAUGAUUUCUCUGGCUAAAAAGAUUGAUGUACAG ACUGGACGGCAUCCAUAUGAGUAAGAAUUCUAGGCUGGA GCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCC AGCCCCUCCUCCCCUUCUGCACCCGUACCCCCGUGGUCU UUGAAUAAAGUCUGAGUGGGCGGCACUAGUAAAAAAAAA

		AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
53	0301-J0 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUC GAAUAAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUCUUCUUGGUGGCAGCAGCCACGCGAGUCCA CUCCAUGAUGACACCAGAAAACGACGAAGAGCAGACAUCU GUGUUCUCCGCUACUGUUUACGGAGACAAAUUCAGGGA AAGAAUAAACGCAAACGCGUGAUUGGUCUAUGUAUUAGA AUAUCUAUGGUUAUUUCACUACUAUCUAUGAUUACCAUG UCCGCGUUUCUCAUAGUGCGCCUAAAUCA AUGC AUGUCUG CUAACAAGGCUGCUAUUACUGACUCCGCGUGUUGCCGUUGC UGCGGCAUCAUCUACUCAUAGAAAGGUUGUGUCUAGCAC UACACAAUAUGAUCACAAAGAAAGCUGUAAUGGUUUUA UUACCAGGGUUCUUGUUUAUAUUAUACAUCAGACUAUAA GUCAUUCGAGGAUGCUAAAAGCAAACUGCGCUGCGGAAUC AUCAACACUACCCAAUAAAUCCGAUGUCUUGACUACCUGG CUCAUUGAUUAUGUUGAGGAUACAUGGGGAUCUGAUGGU AAUCCAAUUAACAAAACUACAUCCGAUUAUCAAGAUUCU GAUGUAUCACAAGAAGUUAGAAAGUAUUUUUGUACAUA GAAUUCUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCC CUUGGGCCUCCCCCAGCCCCUCCUCCCUUCCUGCACCCG UACCCCGUGGUCUUUGAAUAAAGUCUGAGUGGGCGGCA CUAGUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAA
54	0501-J0 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUC GAAUAAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUCUUCUUGGUGGCAGCAGCCACGCGAGUCCA CUCCAUGAAAACGAUUUCCGUUGUUACGUUGUUAUGCGU ACUACCUGCUGUUGUUUAUUCAACAUGUACUGUACCCACU AUGAAUAAACGCUAAAUUAAACGUCUACCGAAACAUCGUUU AAUGAUAAACAGAAAGUUACGUUUACAUGUGAUUCAGGA UAUCAUUCUUGGAUCCAAAUGCUGUCUGCGAAACAGAU AAAUGGAAAUACGAAAAUCCAUGCAAGAAAAUGUGCACA GUUUCUGAUUAUGUCUCUGAACUAUAUGAUAAAGCCAUUA UACGAAGUGAAUUCCACCAUGACACUAAGUUGCAACGGU GAAACAAAUAUUUUCGUUGUGAAGAAAAAAUGGAAAU ACUUCUUGGAAUGAUACUGUCACGUGUCCUAAUGCGGAA UGUCAACCUCUCAAUUAGAACACGGAUCGUGUCAACCAG UUAAGAAAAAUACUCAUUUGGGGAUAUAUGACUAUCA ACUGUGAUGUUGGAUAUGAGGUUAUUGGUGUUUCGUUA

		UAAGUUGUACGGCUAAUUCUUGGAAUGUUAUUCCAUCAU GUCAACAAAAAUGUGAUUAUACCGUCCCUAUCUAAUGGAU UAAUUUCCGGAUCUACAUUUUCUAUCGGUGGCGUUAUAC AUCUUAAGUUGUAAAAGUGGUUUUACACUAACGGGGUCCC CAUCAUCCACAUGUAUCGACGGUAAAUGGAAUCCCAUACU CCCAACAUUGUGUACGAUCUAACGAAGAAUUUGAUCCAGU GGAUGAUGGUCCCGACGAUGAGACAGAUCUGAGCAAACU CUCGAAAGACGUUGUACAAUAUGAACAAGAAAUAGAAUC GUUAGAAGCAACUUAUCAUAUAAUCAUAAUGGCGUUGAC AAUUAUGGGUGUCAUAUUUCUAAUCUCCAUAUAGUAUU AGUUUGUUCCUGUGACAAAAUAAUGACCAAUAUAAGUU CCAUAAAUUGCACCGUGAGAAUUCUAGGCUGGAGCCUC GGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCC CUCCUCCCCUCCUGCACCCGUACCCCCGUGGUCUUUGAA UAAAGUCUGAGUGGGCGGCACUAGUAAAAAAAAAAAAAA AA AA AA
55	0701-J0 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCU GAAUAAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUCUUCUUGGUGGCAGCAGCCACGCGAGUCCA CUCCAUGGGUGCCGCAGCAAGCAUACAGACGACUGUGAAU ACACUCAGUGAACGUUUCGUCUAAAUUAGAACAAGAA GCGAACGCUAGUGCUCUAAACAAAUGUGAUUAUAGAAUC GGAAAUUUUUAUAUCCGACAAAACCAUGGAUGUAACAUC ACUGUAAAAAAAAUUGUGCUCUGCGGACGCGGAUGCUCAG UUGGAUGCUGUGUUAUCAGCCGCUACAGAAACAUUAAGU GGAUUAACACCGGAACAAAAAGCAUACGUACCAGCUAUG UUUACUGCUGCGUUAACAUCAGACGAGUGUAAACACU GUUGUUAGAGAUUUUGAAAAUUAUGUGAAACAGACUUGU AAUUCUAGCGCUGUUGUCGAUAACAAAUUAAAGAUACAA AACGUAAUUAUAGAUGAAUGUUACGGAGCCCCAGGAUCU CCAACAAAUUUGGAAUUUAUUAUACAGGAUCUAGCAAA GGAAAUUGUGCCAUAUAGGCGUUGAUGCAAUUGACUACU AAGGCCACUACUCAAUAGCACCUGAGACAAGUUGCUGGU ACAGGAGUUCAGUUUUUAUUGAUUGUUAUCGGUGUUAUA AUUUGGCAGCGUUGUUUAUGUACUAGCCAAGCGUAUG CUGUUCACAUCCACCAUGAUAAAUCAAACUUAUUUUA GCCAAUAAGGAAAACGUCCAUUGGACUACUUAUUGGAC ACAUUCUUUAGAACUUCUCCGAUGAUUAUUGCUACCACG GAUAUACAAAACUGAGAAUUCUAGGCUGGAGCCUCGGUG GCCAUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCC UCCCCUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAA

		GUCUGAGUGGGCGGCACUAGUAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
56	0901-J0 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUCUUCUUGGUGGCAGCAGCCACGCGAGUCCA CUCCAUGCCUCAACAACUAUCUCCUAUUAAUUAUAGAAACU AAAAAAGCAAUUUCUGACGCGCGAUUGAAGACGUUAGAC AUACAUAUAAUGAGUCGAAACCAACCACUAUCCAGAAU ACUGGAAAACUAGUAAGGAUUAUUUUAAAGGAGGAUUAU AUAAGUGGAGGGUUCUCCCCAAUGAAUAUGUGUUAUCA ACAAUACAUAUAUAUUGGGGAAAGGAAGACGAUUAUGGA UCCAAUCACUUGAUAGAUGUAUACAAAUACUCUGGAGAG AUUAAUCUGGUUCAUUGGAAUAAGAAAAAAUAUAGUUCU UAUGAAGAGGCCAAAAAACACGAUGAUGGAAUUAUCAUU AUUGCUAUAUUCUUAACAAGUAUCAGAUCAUAAAAUGUA UAUUUCAAAGAUAGUCAAUCAAUUGGAUUCCAUUAGA UCCGCUAAUAUGUCUGCGCCAUUUGAUUCAGUAUUUUUAU CUAGACAAUUUGCUACCUAGUACGUUGGAUUAUUUUACA UAUCUAGGAACAACUAUCAACCACUCUGCAGACGCAGCAU GGAUAAUUUUUCCAACACCAUAAACAUUCAUUCUGAUC AACUAUCUAAAUUCAGAACACUAUUGUCGUCGUCUAAUC AUGAGGGUAAACCUCAUUAUAUAACAGAGAAUUAUAGAA AUCCGUAAAUUGAACGACGACACGCAAGUAUAUUAUU CUGGGGAGAUUAUACGAGCUGCAACCACCUCACCAGUUCG AGAGAACUAUUUUUAUGAAAUGGUUGUCUGAUUUUGAGAGA GGCAUGUUUUUCAUAUUAUCAAAAAUAUAUCGAAGGGAA UAAAACAUUCGCAAUUAUUGCCAUAGUAUUCGUAUUUAU ACUCACCGCUAUUCUCUUCUUAUGAGUCAACGAUAUUCG CGAGAAAAACAAAACUAGGAAUUCUAGGCUGGAGCCUCG GUGGCCAUGCUCUUGCCCCUUGGGCCUCCCCCAGCCCC UCCUCCCCUCCUGCACCCGUACCCCGUGGUCUUUGAAU AAAGUCUGAGUGGGCGGCACUAGUAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
57	1101 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CGUGGAUUCUCUUCUGGUGGCAGCCGCCACUCGUGUGCA CUCCGAUGGGACCUUGUUCCCUUGGGGACGACGACCUGGCC AUCCCCGCUACGGAGUUCUUCAGCACCAAGGCUGCCAAGA ACCCGGAGACUAAGCGGGAGGCCAUCGUCAAGGCCUACGG

	AGAUGACAACGAGGAGACGCUGAAGCAGCGGCUGACCAA CCUGGAGAAGAAGAUCACGAACAUCACGACCAAGUUCGA GCAGAUUGAGAAGUGCUGCAAGCGGAACGACGAGGUGUU GUUCCGCUUGGAGAACCACGCAGAGACUCUGCGGGCAGCC AUGAUCAGCUUGGCGAAGAAGAUCGACGUGCAGACUGGA CGCCACCCCUAUGAGGGGAGGAGGCGGCAGCGGCGGCGGCG GCAGCCGGCUGAAUCAGUGCAUGAGCGCCAACAAGGCCGC CAUCACUGAUUCAGCCGUCGCCGUCGCGGCCGCCUCCUCC ACUCAUAGGAAGGUGGUGUCCAGUACGACACAGUACGAC CAUAAGGAGAGCUGUAAUGGGCUGUACUACCAGGGGUCG UGC UACA UCCUGCACAGCGACUACAAGUCGUUUGAGGAU GCGAAGGCGAACUGCGCAGCCGAGAGCAGCACGUUGCCUA ACAAGAGCGACGUGCUGACUACUUGGCUGAUUGACUAUG UCGAGGAUACCUGGGGCU CGGAUGGGAACCCCAUCACGAA GACCACGUCUGACUACCAGGACUCCGAUGUGUCUCAGGAG GUGCGGAAGUACUUCUGUACUGGCGGCGGCGGUUCUGGU GGGGGCGGCAGCACCUGCACC GUCCCCACCAUGAACAACG CCAAGCUGACCAGUACAGAGACU UCCUUAACGAUAAGCA GAAGGUGACA UUUACAUGCGAUUCCGGCUACCACUCGUU GGACCCCAACGCUGUCUGUGAGACAGACAAGUGGAAGUA CGAGAAUCCAUGUAAGAAGAUGUGCACGGUGUCCGACUA UGUUUCUGAGCUGUACGACAAGCCCCUUUACGAGGUGAA CAGCACCAUGACUCUCAGCUGCAACGGCGAGACUAAGUAC UUUCGGUGUGAGGAGAAGAACGGUAAUACGUCCUGGAAC GAUACCGUUA CUUGCCCCAACGCCGAGUGCCAGCCGUUGC AGCUGGAGCAUGGUAGCUGUCAGCCCGUGAAGGAGAAGU AUAGUUUCGGCGAGUACAUGACCAUUA AUUGUGAUGUGG GUUACGAGGUCAUCGGGGUGAGUUACAUCUCUUGCACUG CCAAUAGCUGGAACGUGAUUCCCAGCUGUCAGCAGAAGU GUGACAUACCUUCUCUGUCCAACGGGCUGAUCAGCGGGAG CACGUUCAGCAUUGGCGGGGUGAUCCACCUCAGCUGCAAG AGCGGGUUCACGUUGACCGGGUCGCCUCGUCGACCUGCA UCGACGGGAAGUGGAAUCCCAUUCUGCCCACGUGCGUGCG GUCGAACGAGGAGUUCGACCCGGUCGACGAUGGACCCGAU GACGAGACUGAUCUGUCUAAGUUGUCGAAGGACGUCGUU CAGUACGAGCAGGAGAUCGAGUCCCUGGAGGCCACGUACC ACGGCGGUGGGCGGGAGCGGGGGAGGUGGCAGCAGUCCCC AGACUUCCAAGAAGAUCGGGGACGACGCCACCAUCUCCUG CUCCCGCAACAACACUAACUACUACGUGGUUAUGUCCGCC UGGUACAAGGAGCCCAACUCGAUCAUCCUGCUCGCUGCUA AGAGCGACGUCCUGUACUUCGACAACUAUACCAAGGACAA GAUCAGCUACGACUCACCCUACGAUGACCUCGUGACCACC AUCACAAUUAAGAGUCUGACCGCCGGGGACGCCGGCACAU
--	--

	ACAUUUGCGCCUUCUUCAUGACGAGUACCACCAAUGAUAC CGACAAGGUGGAUUAUGAGGAGUACUCGAUCGAGCUUUAU CGUGAACACGGACAGUGAGUCGACCAUCGACAUCAUCCUG UCCGGUAGCACCCCUGAGACAAUCUCCGAGAAGCCGGAGG ACAUCGACAAUAGCAAUUGCAGUUCCGUCUUCGAGAUCAC GACCCCUGAGCCCAUUAACCGACAACGUGGAUGAUCACACC GACACGGUGACGUACACCAGCGACUCUAUUAACACCGUCA AUGCUUCCAGUGGGGAGUCCACCACUGACGAGAUUCCGGA GCCCAUACCCGACAAGGAGGAGGAUCACACCCGUGACCGAC ACGGUGAGUUACACCACCGUGUCUACCUCUCUGGUAUUG UGACUACCAAGAGCACCACGGACGACGCCGAUCUGUACGA CACGUACAAUGACAACGACACCCGUGCCGCCACCACCGUG GGCGGGUCCACGACGUCCAUAUAGUAACUACAAGACUAAG GACUUCGUGGGGGGCGGCGGUUCCGGUGGCGGGGGCAGC GCUGAAACCAAGGAGUUAAGACGCUGUACAACUUGUUC AUCGACUCUUAACUGCAGAAGUUGGCCCAGCACUCCAUC CUACCAUUGUCACUUGUGCCAUCCACAUUGGAGAGGUGA UCGGCCAGUUAAGAACUGCGCCCUGAGGAUCACCAACAA GUGCAUGAGCAAUAGCCGCCUGUCCUUCACGCUCAUGGUG GAGUCCUUCAUCGAGGUGAUCUCCUGCUGAGCGAGAAG GACAGGCGGGCUAUUGCUGAGGAGAUCGGUAUUGACCUC GACGACGUGCCGUCCGUCGUCUCCAAGCUGGAGAAGAACU GCAACGCAUAUGCUGAGGUGAACAACAUAUCAUAUUC AGAAGCUGAAUAUCGGUGAGUGUCCGCUCUUUUUGGGCA GCAUAUGCUGUUGCAGAUUGUCAAUACCGGGAGUGCUGA GGCCAACUGCGGGUUGCAGACCAUCGUGAAGAGUCUGAA CAAGAUUUACGUGCCCCCGAUCAUCGAGAACC GCCUGCCC UACUACGACCCCGUGGGGGGCGGGUCUGGAGGAGGCGGC AGCGGCGCUGCCGCCUCCAUCCAGACCACGGUGAAUACGU UGUCCGAGCGGAUUUCGAGCAAGCUGGAGCAGGAGGCCA ACGCCUCUGCUCAGACCAAGUGUGAUUAUAGAGAUUGGCA AUUUCUAUAUCAGGCAGAACCACGGGUGCAACAUCACCGU GAAGAAUAUGUGCUCCGCUGACGCCGACGCUCAGCUGGAC GCCGUGCUGAGCGCGGCCACGGAGACAUUAUCCGGUCUGA CCCCUGAGCAGAAGGCGUAUGUGCCUGCCAUGUUCACAGC UGCUCUGAACAUCACAGACGUCCGUGAAUACCGUGGUGAG GGACUUCGAGAACUACGUGAAGCAGACGUGUAAUAGCAG CGCUGUCGUCGAUAAUAAGCUGAAGAUAUCCAGAACGUUAU UAUCGACGAGUGCUACGGCGCCCCCGGCUCUUCCGACCAAC CUGGAGUUAUAAUACCGGUUCCUCCAAGGGGAACUGC GCUAUUAAGGCGCUGAUGCAGCUGACUACGAAGGCGACG ACCCAGAU CGCGCCUCGCCAGGUGGCGGGGACCGGGGUGC AGUUCUACAUGAUCGUGAUCGGUGUGAUAUCCUGGCAG
--	--

		CUCUCUUUAUGUACUACGCCAAGCGCAUGCUGUUCACCAG CACCAACGACAAGAUUAAGCUGAUCCUCGCCAAUAAGGAG AACGUCCACUGGACCACCUACAUGGACACAUUCUUCCGCA CCAGCCCUAUGAUUAUCGCCACCACUGACAUCAGAACUG AUAAGAAUUCUAGGCUGGAGCCUCGGUGGCCAUGCUCUCU UGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUCCUGC ACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGAGUGGGC GGCACUAGUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAAA
58	1201 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCU GAAUAAACGCUCAACUUGGCGCCGCCACCAUGACCCAGC UGCCUAUCCUGCUGCUGCUGAUCAGUCUGGUGUACGCUAC ACCGUCUCCCCAGACUCCAAGAAGAUCCGGGACGACGCG ACGAUCAGCUGCAGCAGGAAUAAACACCAACUACUACGUGG UGAUGAGCGCCUGGUACAAGGAGCCGAAUAGCAUCAUUC UGCUGGCCGCCAAGUCUGACGUGCUGUACUUCGACAAUUA CACAAAGGACAAGAUCUCCUACGACAGUCCUUAUGAUGA UCUGGUGACCACGAUCACCAUUAAGUCUCUGACUGCUGGG GAUGCCGGUACGUUAUUCUGCGCCUUCUUAUGACGUCCA CGACCAAUGACACUGAUAAAGGUGGAUUACGAGGAGUACU CCAUCGAGCUGAUCGUGAACACCGACAGCGAGUCUACCAU CGACAUCAUCCUGUCCGGGAGCACCCCCGAGACUUAUCAGU GAGAAGCCUGAGGAUAUUGACAACAGCAAUUGCAGUAGU GUCUUCGAGAUACCAACCCCGGAGCCUAUCACUGACAACG UGGACGAUCACACCGACACGGUGACGUACACCAGUGAUAG CAUCAACACCGUGAAUGCUAGCUCCGGGGAGUCCACGACC GAUGAGAUCCCGGAGCCUAUCACCGAUAAAGGAGGAGGAU CACACCGUGACCGACACGGUGAGUUACACCACCGUGUCUA CCUCCUCCGGGAUCGUGACUACGAAGAGCACUACUGACGA UGCUGAUCUGUACGAUACCUACAACGAUAACGAUACCGU GCCGCCACCACCGUGGGCGGCAGCACACCAGCAUCAGC AACUAUAAGACUAAGGACUUCGUGGGGGGCGGGGGGAGC GGCGGCGGCGGCAGCCGGCUGAAUCAGUGCAUGAGCGCCA ACAAGGCCGCCAUCACUGAUUCAGCCGUGGCUGUCGCCGC CGCCAGCUCCACCCACCGGAAGGUGGUGUCCAGCACACC CAGUACGAUCACAAGGAGUCCUGCAAUGGCCUGUAUUACC AGGGCUCUUGCUACAUCUGGCACAGCGAUUACAAGAGCUU UGAGGAUGCCAAGGCCAAUUGUGCAGCCGAGAGCAGCAC GCUGCCUAACAAGAGCGACGUGCUGACUACUUGGCUGAUC GACUAUGUGGAGGACACCUGGGGGAGUGAUGGGAACCCC AUCACUAAGACCACGUCCGACUAUCAGGACAGUGACGUGA

		GUCAGGAGGUGCGGAAGUACUUCUGCACCGGCGGGCGGGG GCAGCGGCGGGGGCGGCAGCGCUGAGACAAAGGAGUUCA AGACCCUUUACAACUUGUUCAUCGAUAGUUAUCUGCAGA AGCUGGCCCAGCACAGCAUCCCCACCAAUGUGACAUGCGC CAUUCACAUUGGUGAGGUGAUUGGCCAGUUCAAGAACUG CGCCCUGCGGAUAACUAACAAGUGUAUGUCUAAUUCCCGC CUGUCCUUCACGCUCAUGGUGGAGUCCUUCAUCGAGGUGA UCUCCCUGCUGAGCGAGAAGGACAGGCGGGCUAUCGCCGA GGAGAUCGGCAUCGACCUGGAUGACGUCCCCAGCGUGGUG UCCAAGCUGGAGAAGAACUGCAACGCAUAUGCUGAGGUG AACAAUCAUCAUGGAUAUUCAGAAGCUGAAUAUCGGUGAG UGUUCCGCUCCCCCUGGGCAGCAUAUGCUGUUGCAGAUUCG UCAAUACCGGGUCUGCCGAGGCCAACUGCGGGCUGCAGAC CAUCGUGAAGUCACUGAAUAAGAUCUACGUGCCACCCAUC AUCGAGAACCGGCUGCCCUACUACGACCCGUGGUUCCUGG UGGGGGUGGCCAUUAUUCUGGUGAUCUUCACGGUGGCCA UCUGCAGCAUCCGCAGGAACCUGGCCCUAGAAGUACCGGUA CGGUACCUUCCUCUACGUGUGAUAAGAAUUCUAGGCUGG AGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCC CAGCCCCUCCUCCCCUUCUGCACCCGUACCCCCGUGGUC UUUGAAUAAAGUCUGAGUGGGCGGCACUAGUAAAAAAAAA AAA AAA AAA
59	1301 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUGGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUGUUCUGGUGGCGGCCGCCACAAGGGUGCA CUCCGAUGGGACCUUGUUCCUGGGGACGACGACCUGGCC AUCCCCGCUACGGAGUUCUUCAGCACCAAGGCUGCCAAGA ACCCGGAGACUAAGCGGGAGGCCAUCGUCAAGGCCUACGG AGAUGACAAUGAGGAGACGCUGAAGCAGCGUCUCACCAA CCUUGAGAAGAAGAUACCAACAUCACUACGAAGUUUGA GCAGAUUGAGAAGUGCUGCAAGCGCAAUGAUGAGGUGCU GUUUCGGCUGGAGAACCACGCCGAGACUCUGCGGGCUGCC AUGAUCAGCCUCGCCAAGAAGAUCGACGUGCAGACCGGAA GGCACCCCUAUGAGGGGGGAGGAGGGAGCGGCGGCGGCG GGAGCGGGGCCGCGGCCAGUAUCCAGACCACCGUGAACAC UCUGUCUGAGCGUAUCAGCAGCAAGCUGGAGCAGGAGGC CAACGCCUCUGCUCAGACCAAGUGUGAUUAUAGAGAUUGG CAAUUCUAUAUCAGGCAGAACCACGGGUGCAACAUCACC GUGAAGAAUAUGUGCUCCGCUGACGCCGACGCUCAGCUGG ACGCCGUGCUGAGCGCGGCCACGGAGACAUAUUCCGGUCU GACCCUGAGCAGAAGGCGUAUGUGCCUGCCAUGUUCACA

		GCUGCGCUGAAUAUCCAGACGAGUGUUAACACGGUGGUG CGGGACUUUGAGAACUACGUGAAGCAGACGUGCAACAGC UCAGCGGUGGUGGAUAAUAAGCUCAAGAUCAGAAUGUG AUCAUUGACGAGUGUUAUGGGGCCCCUGGGAGCCCCACUA ACCUCGAGUUUAUUAAUACGGGCAGCAGCAAGGGCAACU GCGCUAUCAAGGCGCUGAUGCAGCUGACAACGAAGGCCAC CACCCAGAUUGCUCGCGGCAGGUGGCAGGAGGGGGAGG GGGGUCCGGCGGGGGAGGCUCGACUUGCACGGUGCCGACG AUGAAUAACGCCAAGCUGACCUCACCGAGACGUCGUUCA ACGACAAGCAGAAGGUGACGUUACCUGCGACAGCGGCUA UCACAGCUUGGACCCCAACGCCGUCUGCGAGACGGACAAG UGGAAGUACGAGAACCCGUGCAAGAAGAUGUGCACGGUG UCCGACUAUGUGUCCGAGCUGUACGAUAAGCCGCUGUACG AGGUGAACAGCACCAUGACGCUGUCGUGCAACGGCGAGAC AAAGUACUCCGGUGCGAGGAGAAGAACGGGAACACCAG CUGGAACGAUACAGUGACCUGUCCCAACGCAGAGUGCCAG CCUCUGCAGUUGGAGCACGGGUCCUGUCAGCCCGUUAAGG AGAAGUACAGCUUUGGCGAGUACAUGACGAUUAACUGCG ACGUCGGUUACGAGGUGAUCGGCGUGAGUUACAUCUCAU GUACCGCCAACAGCUGGAAUGUCAUACCUAGCUGCCAGCA GAAGUGUGACAUCCGUCUCUCUCCAACGGGCUGAUCAGC GGGUCCACGUUCUCUAUCGGCGGGCGUUAUUCAUCUGUCCU GCAAGUCGGGCUUCACCCUGACCGGGUCCCCCUCUUCCAC UUGCAUCGACGGCAAGUGGAACCCCAUCCUGCCCACCUGC GUGCGGAGCAACGAGGAGUUCGACCCCGUGGAUGACGGG CCUGAUGAUGAGACUGAUCUCUCUAAGCUGUCUAAGGAC GUGGUCCAGUACGAGCAGGAGAUCGAGUCGCUGGAGGCC ACCUACCACAUCAUCAUCAUGGCCUCACCAUCAUGGGGG UGAUCUUCUUGAUCAGCAUCAUCGUGCUGGUGUGUUCUU GCGACAAGAACAACGACCAGUACAAGUUUCACAAGCUGCU GCCUUGAUAAGAAUUCUAGGCUGGAGCCUCGGUGGCCAU GCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCU UCCUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGA GUGGGCGGCACUAGUAAAAAAAAAAAAAAAAAAAAAAAAA AAA AAA AAAAAAAAAAAAAAAAAAAAAAAAA
60	1401 mRNA	AUCCAAGCUUAGCUUGCUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUUGGCGCCGCCACCAUGAAGACCA UCUCCGUGGUGACCCUGCUGUGCGUGCUGCCAGCAGUGGU GUACAGCACCUGUACGGUGCCGACGAUGAAUAACGCCAAG CUGACCUCACCGAGACGUCGUUCAACGACAAGCAGAAGG UGACGUUCACCUGCGACAGCGGCUAUCACAGCUUGGACCC

	CAACGCCGUCUGCGAGACGGACAAGUGGAAGUACGAGAA CCCGUGCAAGAAGAUGUGCACGGUGUCCGACUAUGUGUCC GAGCUGUACGAUAAGCCGCUGUACGAGGUGAACAGCACC AUGACGCUGUCGUGCAACGGCGAGACAAAGUACUUCGG UGCGAGGAGAAGAACGGGAACACCAGCUGGAACGAUACA GUGACCUGUCCCAACGCAGAGUGCCAGCCUCUGCAGUUGG AGCACGGGUCCUGUCAGCCCGUUAAGGAGAAGUACAGCU UUGGCGAGUACAUGACGAUUAACUGCGACGUCGGUACG AGGUGAUCGGCGUGAGUUACAUCUCAUGUACCGCCAACA GCUGGAAUGUCAUACCUAGCUGCCAGCAGAAGUGUGACA UUCGUCUCUCUCCAACGGGCUGAUCAGCGGGUCCACGUU CUCUAUCGGCGGGCGUUAUUCUAUCUGUCUUGCAAGUCCGGC UUCACGUUGACCGGGUCGCCCUCGUCGACCUGCAUCGACG GGAAGUGGAAUCCCAUUCUGCCCACGUGCGUGCGGUCGAA CGAGGAGUUCGACCCGGUCGACGAUGGGCCGGACGACGAG ACAGAUCUGUCAAGUUGAGCAAGGAUGUGGUUCAGUAU GAGCAGGAGAUCGAGAGCCUGGAGGCCACGUACCACGGG GGAGGGGGGAGCGGCGGGCGGGAGCGGGGCCGCGGCC AGUAUCCAGACCACCGUGAACACUCUGUCUGAGCGUAUCA GCAGCAAGCUGGAGCAGGAGGCCAACGCCUCUGCUCAGAC CAAGUGUGAUUAUAGAGAUUGGCAAUUUCUAUAUCAGGCA GAACCACGGGUGCAACAUCACCGUGAAGAAUAUGUGCUCC GCUGACGCCGACGCUCAGCUGGACGCCGUGCUGAGCGCGG CCACGGAGACAUAUUCCGGUCUGACCCUGAGCAGAAGGC GUAUGUGCCUGCCAUGUUCACAGCUGCGCUGAAUAUCCAG ACGAGUGUUAACACGGUGGUGCGGGACUUGAGAACUAC GUGAAGCAGACGUGCAACUCCAGCGCCGUGGUGGAUAAU AAGCUCAAGAUCAGAAUGUGAUCAUUGACGAGUGUUAU GGGGCCCCUGGGAGCCCCACUAACCUCGAGUUUAUAAUA CCGGUUCCUCCAAGGGGAACUGCGCCAUAAGGCGCUGAU GCAGUUGACAACGAAGGCGACGACCCAGAUCCGCGCCUCGG CAGGUGGCCGCGCGGCGGUGGCGGGUCAGGCGGCGGGGGG AGCGCCGCCGUGAAGACGCCGGUGAUCGUGGUGCCCGUGA UCGACCGGCCUCCGAGCGAGACGUUCCCCAACGUGCAUGA GCAUAUCAACGAUCAGAAGUUCGAUGACGUUAAGGACAA CGAGGUGAUGCAGGAGAAGCGGGACGUGGUGAUCGUGAA UGACGACCCGGAUCACUACAAGGACUACGUGUUUAUCCAG UGGACCGGAGGCAACAUCGGGAUGACGACAAGUACACCC AUUUUUUCUCCGGCUUCUGUAACACCAUGUGCACAGAGG AGACGAAGCGGAACAUCGCUCGCCACCUGGCACUGUGGGA CUCCAAGUUCUUCACAGAGCUGGAGAACAAGAAUGUGGA GUACGUCGUCAUCAUCGAGAAUGAUAAUGUGAUUGAGGA CAUUACAUAUUCUCAGGCCGGUGCUCAAGGCCAUCCAUGAC
--	--

		AAGAAGAUCGACAUCCUGCAGAUGC GGGAGAUCaucacc GGUAAUAAGGUGAAGACUGAGCUGGUCAUUGACAAGGAC CACGCCAUUCUACCUAUACCGGUGGCUACGACGUCUCGC UGAGCGCGUAUAUCAUACGCGUGACCACCGCUCUUAUAU CGUCGACGAGAUCAUUAAGAGCGGUGGUCUGUCCUCGGG CUUCUACUUCGAGAUAGCCCGGAUCGAGAACGAGAUGAA GAUCAACCGCCAGAUCAUGGACAACUCAGCGAAGUACGUC GAGCAUGAUCCCCGCCUGGUGGCCGAGCACCGGUUUGAGA CAAUGAAGCCCAACUUCUGGUCCCGGAUAGGCACCGUGGC CGCCAAGCGGUAUCCGGGCGUCAUGUACACGUUACCACC CCGCUCAUCUCAUUCUUCGGCCUGUUCGACAUAACGUCA UCGGCCUGAUCGUGAUACUGUUCAUCAUGUUCAUGCUGA UCUUCAACGUGAAGAGCAAGUUGCUGUGGUUCCUGACAG GAACCUUCGUGACUGCCUUCAUUUGAUAAAGAAUUCUAGG CUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUC CCCCAGCCCCUCCUCCCUUCCUGCACCCGUACCCCCGUG GUCUUUGAAUAAAGUCUGAGUGGGCGGCACUAGUAAAAA AA AA AA
61	1501 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCU GAAUAAACGCUCAACUUGGCGCCGCCACCAUGAAGACCA UCUCGGUGGUGACACUGCUGUGCGUGUUGCCGGCGGUUG UUUAUAGUACGUGCACCGUGCCGACUAUGAACAACGCCAA GUUGACGUCUACUGAGACUUCUUAUGACAAGCAGAA GGUGACGUUACACUGCGACAGCGGCUAUCACAGCUUGGAC CCCAACGCCGUCUGCGAGACGGACAAGUGGAAGUACGAGA ACCCGUGCAAGAAGAUGUGCACGGUGUCCGACUAUGUGU CCGAGCUGUACGAUAAGCCGCUGUACGAGGUGAACAGCAC CAUGACGCUGUCCUGCAACGGCGAGACAAAGUACUUCCGU UGCGAGGAGAAGAACGGCAACACCAGCUGGAAUGACACC GUGACCUGUCCGAACGCCGAGUGCCAGCCCCUGCAGUUGG AGCAUGGCUCCUGCCAGCCCGUGAAGGAGAAGUACUCCU CGGGGAGUACAUGACUAUCAACUGCGACGUGGGCUACGA GGUGAUCGGGGUGUCUUAUAUCAGCUGCACCGCCAAUAG CUGGAACGUGAUUCCCAGCUGUCAGCAGAAGUGUGACAU ACCUUCUCUGUCCAACGGGCUGAUCAGCGGGAGCACGUUC AGCAUUGGCGGGGUGAUCCACCUGAGUUGCAAGUCUGGC UUCACGCUGACCGGAUCUCCUAGCUCGACCUGCAUCGACG GGAAGUGGAACCCCAUCCUGCCAACUUGCGUCCGGUCCAA CGAGGAGUUCGACCCCGUGGACGACGGACCAGACGACGAA ACCGAUCUGAGCAAGUUGAGCAAGGAUGUCGUGCAGUAC GAGCAGGAGAUCGAGUCCUGGAGGCGACCUAUCACGGU

		GGGGGAGGGAGCGGCGGCGGCGGCGCUCACGGCUGAAUCAG UGCAUGAGCGCCAACAAGGCCGCCAUCACUGAUUCAGCCG UGGCCGUCGCCGCCGCUUCCUCCACCCACCGGAAGGUCGU CUCCAGCACCAACCAGUAUGAUCACAAGGAGUCCUGCAAC GGGCUGUACUAUCAGGGCAGUUGCUACAUCUGCACAGCG ACUACAAGUCGUUUGAGGAUGCUAAGGCCAACUGCGCCGC CGAAAGCUCCACGCUGCCGAAUAAGUCGGAUGUCCUGACC ACGUGGCUCAUCGACUACGUGGAGGACACCUGGGGCUCGG AUGGGAACCCCAUCACGAAGACCACCAGCGACUAUCAGGA CAGCGAUGUGAGCCAGGAGGUGCGUAAGUACUUCUGCAC CGGGGGCGGAGGGUCUGGAGGAGGCGGCAGCGGCGCUGC CGCCUCCAUCCAGACCACUGUGAACACCCUGUCCGAGCGC AUUUCGUCCAAGCUGGAGCAGGAGGCCAACGCCUCCGCC AGACGAAGUGCGACAUCGAGAUAGGGAAUUCUACAUC GCCAGAACCACGGGUGCAACAUCACCGUGAAGAAUAUGU GCUCCGCUGACGCCGACGCUCAGCUGGACGCCGUGCUGAG CGCGGCCACGGAGACAUAUCCGGUCUGACCCCGGAGCAG AAGGCUUACGUGCCUGCCAUGUUUACGGCGGCCUGAAUA UACAGACUUCUGUGAAUACGGUGGUGCGGGACUUCGAGA AUUACGUGAAGCAGACUUGCAACUCCAGCGCCGUGGUGG AUAAUAAGCUCAAGAUCAGAAUGUGAUCAUUGACGAGU GUUAUGGGGCCCCUGGGAGCCCCACUAACCUCGAGUUUAU UAAUACCGGUUCCUCCAAGGGGAACUGCGCCAUAAGGCG CUGAUGCAGCUGACUACUAAGGCCACCACCCAGAUCGCC CACGUCAGGUGGCUGGCACCGGCGUUCAGUUCUACAUGAU CGUGAUCGGGGUGAUCAUUCUGGCUGCGCUGUUC AUGUA CUACGCCAAGCGCAUGCUCUUUACUUCGACCAACGAUAAG AUCAAACUUAUCUUGGCGAAUAAGGAGAAUGUGCAUUGG ACGACGUACAUGGACACGUUCUCCGGACGUCUCCUAUGA UCAUUGCGACCACCGAUAUUCAGAACUGAUAAAGAAUUCU AGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGC CUCCCCCAGCCCCUCCUCCCCUCCUGCACCCGUACCCCC GUGGUCUUUGAAUAAAGUCUGAGUGGGCGGCACUAGUAA AA AA AA AA
62	1601 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAUUAACGCUCAACUUGGCGCCGCCACCAUGGACUGGA CAUGGAUUCUGUCCUGGUGGCCGCCGCCACCAGGGUGCA CAGCGACGGCACCCUGUUCCAGGCGACGACGACCUGGCC AUCCCCGCUACCGAGUUCUUCAGCACCAAGGCUGCCAAGA ACCCCGAGACUAAGCGGGAGGCCAUCGUGAAGGCCUACGG

		GGAUGAUAAACGAGGAGACUCUGAAGCAGCGGCUGACCAA CCUGGAGAAGAAGAUUACCAACAUAUACCACCAAGUUCGA GCAGAUUCGAGAAGUGCUGCAAGCGGAACGAUGAGGUGCU GUUUCGGCUGGAGAACCACGCCGAGACACUGCGCGCCGCC AUGAUCAGCCUGGCCAAGAAGAUCGACGUGCAGACCGGCC GCCACCCCUACGAGGGGGGCGGCGGCUCUGGGGGCGGCGG GAGCAGCCCCCAGACUCCAAGAAGAUCGGGGAUGAUGCC ACCAUCAGCUGCUCCCGGAACAACACCAACUACUACGUGG UGAUGUCCGCCUGGUACAAGGAGCCUAAACAGCAUCAUCCU GCUGGCCGCCAAGAGCGAUGUGCUGUACUUUGACAAUUA CACCAAGGAUAAGAUCAGCUACGACAGCCCCUACGACGAU CUGGUGACUACCAUCACCAUCAAGAGCCUGACAGCCGGCG ACGCAGGCACCUACAUCUGCGCCUUCUUAUGACCAGCAC CACCAUUGACACCGACAAGGUGGAUUACGAGGAGUACUCC AUCGAGCUGAUCGUGAAUACUGACUCAGAGUCUACCAUC GACAUUAUCCUGUCCGGGAGCACCCCCGAGACUAUCAGUG AGAAGCCUGAGGACAUCGACAACAGCAACUGCAGCAGUG UGUUCGAGAUCACCACCCCAGAGCCUAUCACUGACAACGU GGAUGAUCACACCGACACUGUGACAUACACCAGUGAUAGC AUAACACCGUGAAUGCUAGCUCUGGGGAGAGCACCACUG AUGAGAUCCCCGAGCCUAUCACCGACAAGGAGGAGGAUCA CACCGUGACCGACACUGUGAGUUACACCACAGUGUCUACC UCCUCUGGGAUCGUGACCACCAAGAGCACCACCGACGAUG CCGACCUGUACGAUACAUAACGAUAACGAUACAGUGCC UCCCACCACCGUGGGAGGCAGCACCACCAGCAUUUCUAAU UACAAGACCAAGGACUUCGUGGGGGGCGGGGGGAGCGGG GGCGGGGGGAGCGCCGAAACCAAGGAGUUAAGACCCUG UACAACCUGUUAUCGAUAGCUACCUGCAGAAGCUGGCCC AGCACAGCAUUCCCACCAAUGUGACAUGCGCCAUUCACAU UGGGGAGGUGAUCGGCCAGUUAAGAACUGCGCCCUGAG GAUCACCAAUAAGUGCAUGAGCAAUAGCCGCCUGUCCUUC ACUCUGAUGGUGGAGUCCUUAUCGAGGUGAUCUCCUGC UGUCAGAGAAGGACAGGCGGGCUAUUGCCGAGGAGAUUCG GGAUCGACCUGGACGACGUGCCCAGCGUGGUGUCCAAGCU GGAGAAGAACUGCAACGCCUACGCCGAGGUGAACAACAUC AUCGAUAUUCAGAAGCUGAAUAUCGGUGAGUGUUCGCC CCUCCCGGCCAGCACAUGCUGCUGCAGAUCGUGAAUACAG GGAGCGCUGAGGCCAACUGCGGGCUCCAGACCAUCGUGAA GAGCCUGAAUAAGAUAUACGUGCCUCCUAUCAUCGAGAA UCGGCUGCCCUACUACGACCCCUGGUUCCUGGUGGGCGUG GCCAUUAUCCUGGUGAUCUUCACCGUGGCCAUCUGUAGCA UCAGGCGGAACCUGGCCUGAAGUACAGGUACGGGACCUU CCUGUACGUGUGAUAAGAAUUCUAGGCUGGAGCCUCGGU
--	--	--

		GGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUC CUCCCCUUCUGACCCCGUACCCCGUGGUCUUUGAAUAA AGUCUGAGUGGGCGGCACUAGUAAAAAAAAAAAAAAAAA AAA AAA AAA
63	1701 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUUGGCGCCGCCACCAUGACCCAGC UGCCCAUCCUGCUGCUGCUGAUCAGCCUGGUGUACGCCAC UCCAGUCCUCAGACCUGCAAGAAGAUUGGCGACGAUGCC ACCAUCAGCUGCUCCCGGAUAAUACCAACUACUAUGUGG UGAUGAGCGCCUGGUACAAGGAGCCUAAACAGCAUCAUCCU GCUGGCCGCCAAGAGCGAUGUGCUGUACUUCGAUAACUAC ACCAAGGACAAGAUCAGCUACGAUAGCCCCUACGACGAUC UGGUGACUACCAUCACCAUCAAGAGUCUGACUGCCGGCGA CGCCGGGACUUACAUCUGCGCCUUCUUCAUGACCAGCACC ACCAUGACACCGACAAGGUGGACUACGAGGAGUACAGC AUCGAGCUGAUCGUGAACACCGACAGCGAGAGCACCAUCG ACAUCAUCCUGUCUGGCGAGCACCCCCGAGACUAUCAGCGA GAAGCCUGAGGACAUCGAUAACAGCAACUGUUCAGCGU UUUCGAGAUCACCACCCCGAGCCUAUCACCGAUAAACGUG GACGAUCACACCGACACCGUGACCUACACCUCCGAUAGCA UCAACACCGUGAAUGCCAGCAGCGGGGAGAGUACUACAG AUGAGAUCCCCGAGCCUAUCACCGAUAAAGGAGGAGGAUC ACACCGUGACCGACACUGUGAGUUACACCACAGUGUCUAC CUCCUCCGGCAUCGUGACUACCAAGAGUACCACCGACGAU GCCGAUCUGUACGAUACCUACAACGAUAACGACACAGUGC CUCCCACCACCGUGGGAGGCAGCACACCAGCAUCAGCAA CUACAAGACCAAGGAUUUUGUGGGCGGCGGGGGCUCCGG CGGCGGAGGCUCGCGCCGCCGUGAAGACCCCGUGAUUGUG GUGCCUGUGAUCGAUCGGCCCCUAGCGAGACAUUCCCA ACGUGCACGAGCACAUCAACGAUCAGAAGUUCGAUGACG UGAAGGACAACGAGGUGAUGCAGGAGAAGCGGGACGUGG UGAUCGUGAACGACGAUCCUGACCACUACAAGGACUACGU GUUCAUCCAGUGGACCGGGGGCAACAUCCGGGAUGACGAC AAGUACACCCACUUCUUCUCCGGCUUCUGUAACACCAUGU GUACAGAGGAGACAAAGCGGAACAUUGCUCGGCACCUGG CCCUGUGGGACAGCAAGUUCUUCACUGAGCUGGAGAACA AGAACGUGGAGUAUGUGGUGAUAUCGAGAACGACAAUG UGAUCGAGGACAUUACCUUCCUCAGGCCUGUGCUGAAGGC CAUCCACGAUAAGAAGAUCGACAUCUGCAGAUGCAGGGA GAUCAUCACCGGGAUAAGGUGAAGACCGAGCUGGUGAU CGACAAGGAUCACGCCAUCUUCACCUACACCGGCGGCUAU

		GAUGUGAGCCUGAGCGCUUACAUCAUCCGGGUGACCACCG CCCUGAAUAUCGUGGAUGAGAUAUUAAGUCCGGGGGGC UGUCCAGCGGGUUCUACUUCGAGAUUGCUGGAUCGAGA ACGAGAUGAAGAUCAACCGGCAGAUCAUGGAUAAUUCUG CCAAGUACGUGGAGCACGACCCUCGGCUGGUGGCUGAGCA UCGCUUCGAGACUAUGAAGCCCAACUUCUGGAGCAGGAU UGGGACCGUGGCCGCCAAGCGGUACCCAGGGGUGAUGUAC ACCUUUACCACACCCCUGAUCAGCGGCGGCGGGGGCAGCG GCGGGGGCGGCAGCGCUGAGACAAAGGAGUUCAAGACCC UGUACAACCUGUUUAUCGACAGCUAUCUGCAGAAGCUGG CUCAGCACUCUAUUCCACCA AUGUGACAUGCGCCA UUCA CAUUGGGGAGGUGAUCGGCCAGUUCAAGAACUGCGCCCU GAGGAUCACCAUAAGUGCAUGAGCAACAGCAGGCUGAG CUUCACCCUGAUGGUGGAGUCCUUAUCGAGGUGAUCAGC CUGCUGAGCGAGAAGGACAGGAGGGCCAUCGCCGAGGAG AUCGGCAUCGACCUGGACGAUGUGCCCAGCGUGGUGAGCA AGCUGGAGAAGAAUUGUAACGCCUACGCCGAGGUGAACA AUAUCAUCGAUAUUCAGAAGCUGAAUAUCGGGGAGUGCA GUGCUCUUUUUGGGCAGCACAUUGCUGCUGCAGAUUGUGAA CACCGGCAGCGCCGAGGCCAAUUGCGGGCUGCAGACCAU GUGAAGUCACUGAAUAAGAUCUACGUGCCACCAUCAUCG AGAACCGGCUGCCAUAUACGAUCCUGGUUCCUGGUGGG GGUGGCCAUUAUUCUGGUGAUCUUCACAGUGGCCAUUCUG CAGCAUCAGGCGCAACCUGGCUCUGAAGUACAGGUACGGG ACCUUCCUGUACGUGUGAUAAGAAUUCUAGGCUGGAGCC UCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUUUUUUUAGC CCCUCCUUUUUUCUGCACCCGUACCCCGUGGUCUUUG AAUAAAGUCUGAGUGGGCGGCACUAGUAAAAAAAAAAAAA AAA AAA AAA
64	1801 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUGGGCGCCGCCACCAUGACCCAGC UGCCUAUCCUGCUGCUGCUGAUCAGCCUGGUGUACGCUAC CCCUAGUCCCCAGACUUCCAAGAAGAUCCGGGGACGACGCC ACGAUCAGCUGCAGCAGGAAUAACACCAACUACUACGUGG UGAUGAGCGCCUGGUACAAGGAGCCAAAUAUGCAUCAUUC UGCUGGCCGCCAAGUCUGACGUGCUGUACUUCGACAAUUA CACAAAGGACAAGAUCUCCUACGACAGUCCUUAUGAUGA UCUGGUGACCACGAUCACCAUUAAGUCUCUGACUGCUGGG GAUGCCGGUACCUAUAUCUGUGCGUUCUUUAUGACCUCU ACAACGAACGACACAGAUAAAGGUGGACUACGAGGAGUAU AGCAUCGAGCUGAUCGUGAACACCGACAGCGAGAGCACAA

	UUGACAUAUUAUCCUCAGCGGGAGCACCCCCGAGACAAUCUC UGAGAAGCCCCGAGGACAUAUUAUAGCAAUUGUAGCUC CGUGUUCGAGAUUACUACUCCCGAGCCCAUUAACCGAUAAU GUGGAUGAUCAUACAGAUACGGUGACCUACACCAGCGAC AGCAUCAACACCGUGAAUGCUAGCAGCGGUGAGAGCACCA CCGAUGAGAUCCCGGAGCCUAUCACCGAUAAAGGAGGAGG AUCACACCGUGACCGACACGGUGAGUUACACCACCGUGUC UACCUCCUCCGGGAUCGUGACUACCAAGAGCACCACGGAU GACGCCGAUCUGUAUGACACCUACAACGAUAAUGAUACCG UGCCGCCCACCACCGUGGGCGGCAGCACCAACAGCAUCAG CAACUAUAAGACUAAGGACUUCGUGGGGGGCGGGGGGAG CGGCGGGCGGGCGGCAGCCGGCUGAAUCAGUGCAUGAGCGCC AACAAGGCCGCCAUACUGAUUCAGCCGUGGCUGUCGCCG CCGCCAGCUCCACCCACCGGAAGGUGGUGUCCAGCACCA CCAGUACGAUCACAAGGAGUCCUGCAAUGGCCUGUAUUAC CAGGGCUCUUGCUACAUCUGCACAGCGAUUACAAGAGCU UUGAGGAUGCCAAGGCCAAUUGUGCAGCCGAGAGCAGCA CGCUGCCUAACAAGAGCGACGUGCUGACUACUUGGCUGAU CGACUAUGUGGAGGACACCUGGGGCAGUGAUGGGAACCC CAUCACUAAGACCACCUCUGACUACCAGGACUCCGACGUG AGUCAGGAGGUGAGGAAGUACUUCUGUACGGGCGGGCGGC GGCAGCGGGCGGGCGGGUCCACCUGCACCGUGCCCACCA UGAACAACGCCAAGCUGACUUCUACCGAGACUUCUUUCAA CGACAAGCAGAAGGUGACCUUCACCUGCGACAGCGGCUAU CAUUCCCUCGACCCUAACGCCGUGUGCGAGACAGACAAGU GGAAGUAUGAGAACCCGUGCAAGAAGAUGUGUACGGUGU CCGACUACGUCUCUGAGCUGUACGACAAGCCUCUGUACGA GGUCAACUCCACUAUGACUCUGUCUUGCAACGGCGAGACA AAGUACUUUAGGUGCGAGGAGAAGAACGGCAAUACCAGC UGGAACGACACCGUGACGUGCCCCAACGCCGAGUGUCAGC CUCUCCAGCUGGAGCACGGCAGCUGUCAGCCUGUGAAGGA GAAGUACAGCUUCGGCGAGUACAUGACCAUCAACUGUGA UGUGGGGUACGAGGUGAUCGGGGUGAGCUAUUUCAGCUG CACCGCCAAUAGCUGGAACGUGAUUCCCAGCUGUCAGCAG AAGUGCGACAUCCCUUCUCUGUCCAACGGGCUGAUCAGCG GGAGCACGUUCAGCAUUGGCGGGGUGAUCCACCUGAGCU GCAAGUCAGGGUUCACCCUGACUGGCUCACCCAGCAGCAC CUGCAUCGAUGGCAAGUGGAACCCCAUUCUGCCCACCUGU GUGAGAAGCAACGAGGAGUUCGAUCCUGUGGACGACGGA CCAGACGACGAAACCGAUCUGUCCAAGCUGUCCAAGGAUG UGGUGCAGUACGAGCAGGAGAUCGAGAGCCUGGAGGCCA CCUACCACAUCAUCAUCAUGGCACUGACCAUCAUGGGCGU GAUAUUCUUGAUCAGCAUCAUUGUGCUGGUGUGUUCUUG
--	--

		CGACAAGAACAACGACCAGUACAAGUUUCACAAGCUGCUG CCCUGAUAAGAAUUCUAGGCUGGAGCCUCGGUGGCCAUGC UUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUUC CUGCACCCGUACCCCGUGGUCUUUGAAUAAAGUCUGAGU GGGCGGCACUAGUAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAAAAAAAAAA
65	1901 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUUGGCGCCGCCACCAUGGACUGGA CCUGGAUUCUGUCCUGGUGGCGGCCGCCACAAGGGUGCA CUCCGAUGGGACCUUGUUCCUGGGGACGACGACCUGGCC AUCCCCGCUACGGAGUUCUUCAGACCAAGGCUGCCAAGA ACCCGGAGACUAAGCGGGAGGCCAUCGUCAAGGCCUACGG AGAUGACAAUGAGGAGACGCUGAAGCAGCGUCUCACCAA CCUUGAGAAGAAGAUACCAACAUCACUACGAAGUUUGA GCAGAUUGAGAAGUGCUGCAAGCGCAAUGAUGAGGUGCU GUUUCGGCUGGAGAACCACGCCGAGACUCUGCGGGCUGCC AUGAUCAGCCUCGCCAAGAAGAUCGACGUGCAGACCGGAA GGCACCCCUAUGAGGGGGGAGGAGGGAGCGGCGGCGGCG GGAGCGGGGCCGCGGCCAGUAUCCAGACCACCGUGAACAC UCUGUCUGAGCGUAUCAGCAGCAAGCUGGAGCAGGAGGC CAACGCCUCUGCUCAGACCAAGUGUGAUUAUAGAGAUUGG CAAUUUCUAUAUCAGGCAGAACCACGGGUGCAACAUCACC GUGAAGAAUAUGUGCUCGCGUGACGCCGACGCUCAGCUGG ACGCCGUGCUGAGCGCGGCCACGGAGACAUAUUCCGGUCU GACCCCUAGAGCAGAAGGCGUAUGUGCCUGCCAUGUUCACA GCUGCGCUGAAUAUCCAGACGAGUGUUAACACGGUGGUG CGGGACUUUGAGAACUACGUGAAGCAGACGUGCAACUCC AGCGCCGUGGUGGAUAAUAAGCUCAAGAUCAGAAUGUG AUCAUUGACGAGUGUUAUGGGGCCCCUGGGAGCCCCACUA ACCUCGAGUUUAUUAUACCGGUUCCUCCAAGGGGAACU GCGCCAUCAAGGCGCUGAUGCAGUUGACAACGAAGGCCAC CACCCAGAUCGCCCCCAGGCAGGUUGCCGGGGGCGGCGGG GGGAGCGGUGGCGGCGGCAGCGCUGAGACAAAGGAGUUC AAGACCCUGUACAACUUGUUCUAGAUUCUUAUCUUCAG AAGUUGGCCAGCACUCCAUCCCUACCAAUGUCACUUGUG CCAUCCACAUUGGAGAGGUGAUCGGCCAGUUCAAGAACU GCGCCCUGAGGAUCACCAACAAGUGCAUGAGCAAUAGCCG CCUGUCCUUCACGCUCAUGGUGGAGUCCUUCAUCGAGGUG AUCUCCUGCUGAGCGAGAAGGACAGGCGGGCUAUUGC GAGGAGAUCCGUAUUGACCUCGACGACGUGCCGUCGUCG UCUCCAAGCUGGAGAAGAACUGCAACGCAUAUGCUGAGG

		UGAACAACAUCAUCGAUAUUCAGAAGCUGAAUAUCGGUG AGUGUUCCGCUCCCCUGGGCAGCAUAUGCUGUUGCAGAU UGUCAAUACCGGGAGUGCUGAGGCCAACUGCGGGCUGCA GACCAUCGUGAAGUCACUGAAUAAGAUCUACGUGCCACCC AUCAUCGAGAACAGGCUGCCGUACUACGACCCGUGGUUCU UGGUGGGGGUGGCCAUUAUUCUGGUGAUCUUCACGGUGG CCAUCUGCAGCAUCCGCCGGAACCUGGCUCUGAAGUACCG GUACGGGACCUUCCUGUACGUGUGAUAAAGAAUUCUAGGC UGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCC CCCCAGCCCCUCCUCCCCUCCUGCACCCGUACCCCGUGG UCUUUGAAUAAAGUCUGAGUGGGCGGCACUAGUAAAAAA AA AA AA
66	2001 mRNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUC GAAUAAACGCUCAACUUGGCGCCGCCACCAUGGAUUGGA CCUGGAUUCUGUUCUUCGUUGCCGCCGCCACCAGGGUGCA UAGUGCGGCAGUGAAGACUCCCGUUAUCGUCGUCCCGGUG AUUGACCGCCCGCCUUCGGAGACGUUCCCCAACGUGCAUG AGCACAUCAACGAUCAGAAGUUCGACGAUGUCAAGGACA AUGAGGUCAUGCAGGAGAAGCGGGACGUGGUGAUCGUCA ACGACGAUCCCGAUCACUACAAGGAUUACGUCUUCAUCCA GUGGACGGGCGGCAACAUCGCGGACGACGAUAAGUACAC GCAUUUCUUCUCCGGCUUCUGUAAACACCAUGUGUACAGAG GAGACGAAGCGGAACAUCGCCAGGCAUCUGGCGCUGUGG GACUCGAAGUUCUUCACCGAGCUGGAGAAUAAGAACGUG GAGUACGUGGUGAUCAUCGAGAACGACAACGUCAUCGAG GACAUCACGUUUCUGCGGCCGGUGCUGAAGGCCAUCCACG ACAAGAAGAUCGACAUCCUGCAGAUGCGGGAGAUCAUCA CGGGUAAUAAGGUGAAGACCGAGCUGGUCAUUGACAAGG ACCACGCGAUAUUCACCUACACUGGGGGGAUACGACGUGUC CUUGUCCGCCUAUAUUAUCCGGGUGACCACCGCUCUGAAC AUCGUGGAUGAGAUCAUCAAGUCGGGAGGCCUCAGCUCC GGCUUCUACUUCGAGAUCGCCCCGGAUCGAGAACGAGAUG AAGAUCAACCGCCAGAUCAUGGACAACAGCGCCAAGUAUG UCGAGCACGACCCUCGGCUGGUGGCAGAGCACAGAUUCGA GACUAUGAAGCCCAACUUCUGGAGUAGGAUCGGCACUGU CGCCGCCAAGCGCUACCCAGGUGUGAUGUACACCUUACC ACCCCGCUGAUCAGCGGGGGGGGGGUAGCGGCGGCGGCG GCAGUGCCGAGACAAAGGAGUUCAAGACCCUGUACAACU UGUUCAUCGACUCCUACCUCAGAAAGUUGGCUCAGCACUC GAUCCCCACCAAUGUGACAUGCGCCAUUCACAUUGGUGAG GUGAUCGGGCAGUUCAAGAACUGUGCUCUGCGGAUUACC

		AACAAAGUGUAUGAGUAAUUCCCGCCUGUCCUUCACGCUCA UGGUGGAGUCCUUCAUCGAGGUGAUCUCCUGCUGAGCG AGAAGGACAGGGCGGGCCAUCGCCGAGGAGAUCCGCAUCG ACCUGGACGAUGUCCGUCUGUCGUCUCCAAGCUGGAGAA GAACUGCAACGCAUAUGCUGAGGUGAACAACAUCUAUCGA UAUUCAGAAGCUGAAUAUCGGUGAGUGUCCGCUCCCCCU GGGCAGCAUAUGCUGUUGCAGAUUCGUGAAUACCGGGAGU GCGGAGGCGAACUGCGGGCUCCAGACCAUCGUGAAGAGCC UGAACAAAGAUCUACGUCCCGCCUAUCAUCGAGAAUCGGCU GCCGUACUAUGACCCUGGUGGCGGCGGCAGCGGGGGGGG GGGGUCGGGGGCCGCGUCGUCCAUCCAGACCACCGUGAAC ACUCUGUCUGAGCGUAUCAGCAGCAAGCUGGAGCAGGAG GCCAACGCCUCUGCUCAGACCAAGUGUGAUUAUAGAGAUU GGCAAUUUCUAUAUCAGGCAGAACCACGGGUGCAACAUC ACCGUGAAGAAUAUGUGCUCCGCUGACGCCGACGCUCAGC UGGACGCCGUGCUGAGCGCGGCCACGGAGACAUAUUCCGG UCUGACCCUGAGCAGAAGGCGUAUGUGCCUGCCAUGUUC ACAGCUGCGCUGAAUAUCCAGACGAGUGUUAACACGGUG GUGCGGGAUUUCGAGAAUUAUGUGAAGCAGACCUGUAAC AGCUCGGCUGUUGUGGACAACAAGCUCAAGAUAUCCAGAAU GUGAUCAUUGACGAGUGUUACGGGGCCCCCGGCUCCCCCA CGAAUCUGGAGUUCAUCAACACGGGCAGCAGCAAGGGCA ACUGCGCCAUUAAGGCGCUGAUGCAGCUGACCACUAAGGC CACAACCCAGAUCGCCCCUAGGCAGGUGGCCGGGACUGGC GUGCAGUUCUAUAUGAUAGUGAUUUGGCGUGAUCAUCCUG GCCGCCCUGUUCAUGUACUACGCCAAGCGCAUGCUCUUUA CUUCGACCAACGAUAAGAUAACUUAUCUUGGCGAAUA AGGAGAAUGUGCAUUGGACGACGUACAUGGACACCUUCU UUCGCACAUCACCUAUGAUCAUUGCCACGACUGACAUCU GAACUGAUAAAGAAUUCUAGGCUGGAGCCUCGGUGGCCAU GCUUCUUGCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCU UCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA GUGGGCGGCACUAGUAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AA
67	0201 DNA	ATCCAAGCTTAGCTTGCTTGTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTGTTCTTGGTGGCGGCCGCCACTAGGGTGCACAGCG ACGGGACCCTGTTCCCAGGGGACGACGACCTGGCCATCCCC GCTACCGAGTTCTTCAGCACCAAGGCTGCCAAGAACCCCGA GACTAAGCGGGAGGCCATCGTGAAGGCCTACGGGGACGAT AATGAGGAGACGCTGAAGCAGCGTCTCACCAATCTGGAGA

		AGAAGATTACTAATATCACCACCAAGTTCGAGCAGATCGAG AAGTGCTGTAAGAGGAATGACGAGGTGCTGTTTCGGCTGGA GAACCACGCCGAGACACTGCGGGCTGCCATGATCTCTCTTG CCAAGAAGATCGATGTGCAGACCGGCAGGCACCCTTATTAC CCTTACGACGTGCCCCGATTATGCCGAGTGATAAGAATTCTA GGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCCTTGGGCCTC CCCCCAGCCCCCTCCTCCCCCTTCCTGCACCCGTACCCCCGTGG TCTTTGAATAAAAGTCTGAGTGGGCGGCACTAGTAAAAAAAAA AA AA AA
68	0302 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGATGACTCCCGA GAATGACGAGGAGCAGACCAGCGTGTTCTCCGCTACCGTGT ACGGGGACAAGATCCAGGGGAAGAATAAGAGGAAGCGGGT GATCGGCCTGTGCATCAGGATCAGCATGGTGATCAGCCTGC TGAGCATGATCACCATGTCTGCATTCCTGATCGTGCGGCTGA ATCAGTGCATGAGCGCCAACAAGGCCGCCATCACTGATTCA GCCGTGGCCGTGGCCGCCGCTTCCTCTACTCACAGGAAGGT GGTGTCTCTCCACCACTCAGTACGATCACAAGGAGAGCTGCA ATGGCCTGTATTACCAGGGCTCTTGCTACATCCTGCACAGC GATTACAAGAGCTTTGAGGATGCCAAGGCCAATTGCGCCGC CGAGAGCAGCACTCTGCCTAACAAGAGCGACGTGCTGACTA CTTGGCTGATCGACTATGTGGAGGACACCTGGGGGAGCGAT GGGAACCCCATCACCAAGACTACCAGCGACTATCAGGACA GCGACGTGAGCCAGGAGGTGCGGAAGTACTTCTGCACCTAC CCTTACGACGTGCCTGACTACGCCTGATAAGAATTCTAGGC TGGAGCCTCGGTGGCCATGCTTCTTGCCCCCTTGGGCCTCCCC CCAGCCCCCTCCTCCCCCTTCCTGCACCCGTACCCCCGTGGTCT TTGAATAAAAGTCTGAGTGGGCGGCACTAGTAAAAAAAAAAAA AA AA AA
69	0402 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGACCCAGCTGCC TATCCTGCTGCTGCTGATCAGCCTGGTGTACGCTACCCCTAG CCCCAGACTTCCAAGAAGATCGGGGACGACGCCACTATCAG CTGCAGCAGGAATAACACCAACTACTACGTGGTGATGAGCG CCTGGTATAAGGAGCCTAACAGCATCATCCTGCTGGCCGCC AAGAGCGATGTGCTGTACTTCGACAACTATACCAAGGACAA GATCTCTTACGATAGCCCCTACGATGACCTGGTGACTACCA TCACCATCAAGAGTCTGACCGCCGGGGATGCCGGGACCTAT ATCTGTGCGTTTTTTTATGACCTCTACAACAAACGACACAGA

		TAAGGTGGACTACGAGGAGTATAGCATCGAGCTGATCGTGA ACACCGACAGCGAGAGTACCATCGACATCATCCTGAGCGG GAGTACTCCCGAGACTATCAGCGAGAAGCCAGAGGACATC GACAATAGCAACTGCAGCAGCGTGTTTCGAGATCACTACTCC TGAGCCTATTACAGATAACGTGGATGATCACACCGACACCGT GACCTACACCAGCGACAGCATCAACACCGTGAATGCTAGCA GCGGTGAGAGTACCACGGATGAGATCCCCGAGCCCATCAC GGATAAGGAGGAGGATCACACCGTGACCGACACGGTGAGT TACACCACCGTGTCTACCTCCTCAGGGATTGTGACCACAAAG TCTACCACCGATGATGCCGATCTGTATGACACATACAATGAC AACGATACCGTGCCGCCACCACCGTGGGCGGCAGCACAA CATCCATTAGTAACTACAAGACTAAGGATTTTCGTGGAGATC TTTGGGATCACAACCCTGATTATCCTGAGCGCCGTGGCGATC TTCTGTATCACCTACTATATCTGTAATAAGCATCCCAGGAAG ATAAGACCGAGAACAAAGGTGTATCCGTACGACGTGCCGGA CTACGCCTGATAAGAATTCTAGGCTGGAGCCTCGGTGGCCA TGCTTCTTGCCCCTTGGGCCTCCCCCAGCCCCTCCTCCCCT TCCTGCACCCGTACCCCGTGGTCTTTGAATAAAGTCTGAGT GGGCGGCACTAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AA AA
70	0505 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGAAGACCATCTCC GTGGTGACCCTGCTGTGCGTGCTGCCAGCAGTGGTGTACAG TGGCGGCGGGGGGAGCTACCCCTACGACGTGCCCGACTACG CCACCTGTACGGTGCCAACCATGAACAACGCTAAGCTGACC TCCACAGAGACGAGTTTTAATGACAAGCAGAAGGTGACCTT CACCTGCGACAGTGGCTACCACAGCTTGGACCCCAACGCTG TGTGCGAGACAGACAAGTGGAAGTACGAGAACCCGTGCAA GAAGATGTGCACCGTGAGTGACTACGTCAGCGAGCTGTACG ACAAGCCGCTGTACGAGGTCAACTCCACGATGACACTCTCT TGCAACGGGGAGACAAAGTACTTCAGGTGCGAGGAGAAGA ACGGGAATACCAGCTGGAATGATACCGTGACATGCCCCAAT GCCGAGTGCCAGCCCCTGCAGCTCGAGCATGGGAGCTGTCA GCCCCGTGAAGGAGAAGTACTCCTTCGGGGAGTACATGACCA TCAATTGCGACGTGGGGTACGAGGTGATCGGGGTGAGCTAC ATCTCTTGCAACCGCCAATAGCTGGAACGTGATTCCCAGCTG TCAGCAGAAGTGCGACATCCCTTCTGTCCAACGGGCTGAT CAGCGGGAGCACGTTTCAGCATTGGCGGGGTGATCCACCTGA GCTGCAAGAGTGGGTTACCCCTGACCGGGTCCCCGAGCAGC ACCTGCATCGATGGGAAGTGGAACCCGATCCTGCCAACTTG CGTCCGGTCCAACGAGGAGTTCGACCCCGTGGACGACGGAC

		CAGACGACGAAACCGATCTGAGCAAGTTGAGCAAGGACGT GGTGCAGTACGAGCAGGAGATCGAGAGCCTGGAGGCCACC TACCACATCATCATCATGGCACTGACCATCATGGGCGTGATA TCTTGATCAGCATCATTGTGCTGGTGTGTTCTTGCGACAAGA ACAACGACCAGTACAAGTTTCACAAGCTGCTGCCCTGATAA GAATTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCT TGGGCCTCCCCCAGCCCCCTCCTCCCCCTTCCTGCACCCGTAC CCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCACTAGT AA AA AA AAAA
71	0603 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTGTTCTTGGTGGCGGCCGCTACTCGGGTGCATAGCG GCGGCGGGGGGTCTACCCCTACGACGTGCCCCGACTATGCC GCCGAGACAAAGGAGTTTAAGACCCTGTACAACCTGTTTAT CGACAGCTATCTGCAGAAGCTGGCTCAGCACTCTATTCCCA CCAATGTGACATGCGCCATTACATTGGTGAGGTGATCGGC CAGTTCAAGAACTGCGCCCTGAGGATCACGAATAAGTGCAT GAGCAACAGCAGGCTGAGCTTCACCCTGATGGTGGAGTCCT TCATCGAGGTGATCAGCCTGCTGAGCGAGAAGGACAGGAGG GCCATCGCCGAGGAGATCGGCATCGACCTGGACGATGTGCC CAGCGTGGTGAGCAAGCTGGAGAAGAACTGCAACGCATAT GCTGAGGTGAACAACATCATCGATATTCAGAAGCTGAATAT CGGTGAGTGTTCCGCTCCCCCTGGGCAGCATATGCTGTTGC AGATCGTGAACACGGGCAGTGCCGAGGCCAACTGCGGGCT GCAGACCATCGTGAAGTCACTGAATAAGATCTACGTGCCAC CCATCATCGAGAACCGGCTGCCCTACTACGACCCGTGGTTC CTGGTGGGGGTGGCCATTATTCTGGTGATCTTCACGGTGGC CATCTGCAGCATCCGCAGGAACCTGGCCCTGAAGTACAGGT ACGGGACCTTCCTGTACGTCTGATAAGAATTCTAGGCTGGA GCCTCGGTGGCCATGCTTCTTGCCCCCTGGGCCTCCCCCAG CCCCTCCTCCCCCTTCCTGCACCCGTACCCCCGTGGTCTTTGA ATAAAGTCTGAGTGGGCGGCACTAGTAAAAAAAAAAAAAAAA AA AA AA
72	0702 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTGTTCTTGGTGGCTGCCGCCACCAGGGTGCACAGTG GCGGAGGAGGCAGTACCCCTACGACGTCCCCGACTACGCT GGGGCTGCTGCCTCCATCCAGACCACTGTGAACACCCTGAG

		TGAGCGGATCAGCAGCAAGCTGGAGCAGGAGGCCAACGCC TCTGCTCAGACCAAGTGTGATATAGAGATTGGCAATTTCTAT ATCAGGCAGAACCACGGGTGCAACATCACCGTGAAGAATA TGTGCTCCGCTGACGCCGACGCTCAGCTGGACGCCGTGCTG AGCGCGGCCACGGAGACATATTCCGGTCTGACCCCTGAGCA GAAGGCGTATGTGCCTGCCATGTTACAGCTGCACTGAACA TCCAGACCAGCGTGAACACCGTGGTGAGGGACTTCGAGAA CTACGTGAAGCAGACATGCAACAGCAGCGCGGTGGTGGAT AATAAGCTCAAGATCCAGAATGTGATCATTGACGAGTGTTA TGGGGCCCCTGGGAGCCCCACTAACCTCGAGTTTATTAATA GGGCAGCAGCAAGGGCAACTGCGCCATCAAGGCCCTGATG CAGCTGACAACCAAGGCCACCACCCAGATTGCCCCCAGGCA GGTGGCCGGTTGATAAGAATTCTAGGCTGGAGCCTCGGTGG CCATGCTTCTTGCCCCCTTGGGCCTCCCCCAGCCCCCTCCTCC CCTTCCTGCACCCGTACCCCCGTGGTCTTTGAATAAAGTCTG AGTGGGCGGCACTAGTAAAAAAAAAAAAAAAAAAAAAAAAA AAA AAA AAA
73	0803 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTGTTTCTGGTGGCCGCCGCCACACGGGTGCATTCCG GCGGCGGCGGCAGCTACCCATACGATGTGCCTGACTACGCC GCCGCCGTGAAGACACCCGTGATCGTGGTGCCCGTGATTGA CCGGCCCCCTAGCGAGACATTCCCCAATGTGCACGAGCACA TCAACGACCAGAAGTTCGACGACGTGAAGGACAACGAGGT GATGCAGGAGAAGAGGGACGTGGTGATTGTGAACGACGAT CCTGACCACTACAAGGACTACGTGTTTCATCCAGTGGACAGG CGGGAACATCAGGGACGACGACAAGTATACCCACTTCTTCT CCGGCTTCTGTAACACCATGTGTACAGAGGAGACAAAGCGG AACATTGCTCGGCACCTGGCCCTGTGGGACAGCAAGTTCTT CACCGAGCTGGAGAACAAGAACGTGGAGTATGTGGTGATC ATCGAGAACGACAACGTTCATCGAGGACATCACCTTTCTGCG CCCTGTGCTGAAGGCCATCCACGATAAGAAGATCGACATCC TGCAGATGCGGGAGATCATCACCGGGAATAAGGTGAAGACC GAGCTGGTGATCGACAAGGATCACGCCATCTTCACCTACAC CGGCGGCTATGATGTGAGCCTGAGCGCCTACATCATCCGGG TGACCACCGCCCTGAATATCGTGGATGAGATCATCAAGTCCC GCGGGCTGAGCAGCGGCTTCTACTTCGAGATCGCCAGAATC GAGAACGAGATGAAGATCAACAGGCAGATCATGGACAACA GCGCCAAGTACGTGGAGCACGACCCTCGGCTGGTGGCCGA GCACCGGTTTCGAGACTATGAAGCCCAACTTCTGGTCCAGAA TCGGGACTGTGGCCGCCAAGCGGTACCCCGGCGTGATGTAC

		ACATTCACCACCCCACTGATCAGCTTCTTTGGCCTGTTCGAC ATCAACGTGATCGGCCTGATCGTGATCCTGTTTCATCATGTTT ATGCTGATCTTCAACGTGAAGTCCAAGCTGCTGTGGTTCTT GACCGGCACCTTCGTGACTGCCTTCATCTGATAAGAATTCT AGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCC TCCCCCAGCCCCTCCTCCCCTTCCTGCACCCGTACCCCCGT GGTCTTTGAATAAAGTCTGAGTGGGCGGCACTAGTAAAAAA AA AA AA
74	0201-J0 mRNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTCTTCTTGGTGGCAGCAGCCACGCGAGTCCACTCCA TGGACGGAACCTCTTTTCCCCGGAGATGACGATCTTGCAATTC CAGCAACTGAATTTTCTCTACAAAGGCTGCTAAAAATCCA GAGACTAAACGCGAAGCAATTGTTAAAGCCTATGGAGACG ATAATGAGGAACTCTCAAACAACGGCTAACTAATTTGGAA AAAAAGATTACTAATATAACAACAAAGTTTGAACAAATAG AAAAGTGTTGTAAACGCAACGATGAAGTTCTATTTAGGTTG GAAATCACGCTGAAACTCTAAGAGCGGCTATGATATCTCT GGCTAAAAAGATTGATGTACAGACTGGACGGCATCCATATG AGTAAGAATTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTT GCCCCTTGGGCCTCCCCCAGCCCCTCCTCCCCTTCCTGCAC CCGTACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGC ACTAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAA
75	0301-J0 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTCTTCTTGGTGGCAGCAGCCACGCGAGTCCACTCCA TGATGACACCAGAAAACGACGAAGAGCAGACATCTGTGTTC TCCGCTACTGTTTACGGAGACAAAATTCAGGGAAAGAATAA ACGCAAACGCGTGATTGGTCTATGTATTAGAATATCTATGG TTATTTCACTACTATCTATGATTACCATGTCCGCGTTTCTCAT GTGCGCCTAAATCAATGCATGTCTGCTAACAAGGCTGCTAT TACTGACTCCGCTGTTGCCGTTGCTGCGGCATCATCTACTCA TAGAAAGGTTGTGTCTAGCACTACACAATATGATCACAAAG/ AAGCTGTAATGGTTTATATTACCAGGGTTCTTGTTATATATT ACATTCAGACTATAAGTCATTCGAGGATGCTAAAGCAAAC GCGCTGCGGAATCATCAACACTACCCAATAAATCCGATGTC TTGACTACCTGGCTCATTGATTATGTTGAGGATACATGGGGA CTGATGGTAATCCAATTACAAAACTACATCCGATTATCAA

		GATTCTGATGTATCACAAGAAGTTAGAAAGTATTTTTGTAC ATAAGAATTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTTG CCCCTTGGGCCTCCCCCAGCCCCTCCTCCCCTTCCTGCACC CGTACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCA CTAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAA
76	0501-J0 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTCTTCTTGGTGGCAGCAGCCACGCGAGTCCACTCCA TGAAAACGATTTCCGTTGTTACGTTGTTATGCGTACTACCTG TGTTGTTTATTCAACATGTACTGTACCCACTATGAATAACGC AAATTAACGTCTACCGAAACATCGTTTAATGATAAACAGAA AGTTACGTTTACATGTGATTGAGGATATCATTCTTTGGATCC AAATGCTGTCTGCGAAACAGATAAATGGAAATACGAAAT CCATGCAAGAAAATGTGCACAGTTTCTGATTATGTCTCTGA ACTATATGATAAGCCATTATACGAAGTGAATTCCACCATGA CACTAAGTTGCAACGGTGAAACAAAATATTTTCGTTGTGAA GAAAAAATGGAAATACTTCTTGGAATGATACTGTCACGTG TCCTAATGCGGAATGTCAACCTCTTCAATTAGAACACGGATC GTGTCAACCAGTTAAAGAAAAATACTCATTGTTGGGAATATA TGACTATCAACTGTGATGTTGGATATGAGGTTATTGGTGTTC GTATATAAGTTGTACGGCTAATTCTTGGAATGTTATTCCATC ATGTCAACAAAAATGTGATATACCGTCCCTATCTAATGGAT TAATTTCCGGATCTACATTTTCTATCGGTGGCGTTATACATCT AGTTGTAAAAGTGGTTTACACTAACGGGGTCCCCATCATC CACATGTATCGACGGTAAATGGAATCCCATACTCCCAACAT GTGTACGATCTAACGAAGAATTTGATCCAGTGGATGATGGT CCCGACGATGAGACAGATCTGAGCAAACCTCTCGAAAGACG TTGTACAATATGAACAAGAAATAGAATCGTTAGAAGCAACT ATCATATAATCATAATGGCGTTGACAATTATGGGTGTCATA TTTCTAATCTCCATTATAGTATTAGTTTGTTCCTGTGACAAAA ATAATGACCAATATAAGTTCCATAAATTGCTACCGTGAGAA TTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGG GCCTCCCCCAGCCCCTCCTCCCCTTCCTGCACCCGTACCCC CGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCACTAGTAAA AA AA AA A
77	0701-J0 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG

		GATTCTCTTCTTGGTGGCAGCAGCCACGCGAGTCCACTCCA TGGGTGCCGCAGCAAGCATAACAGACGACTGTGAATACACTC AGTGAACGTATCTCGTCTAAATTAGAACAAGAAGCGAACGC TAGTGCTCAAACAAAATGTGATATAGAAATCGGAAATTTTTA ATCCGACAAAACCATGGATGTAACATCACTGTTAAAAATAT GTGCTCTGCGGACGCGGATGCTCAGTTGGATGCTGTGTTAT CAGCCGCTACAGAAACATATAGTGGATTAACACCGGAACA AAAAGCATACGTACCAGCTATGTTTACTGCTGCGTTAAACA TTCAGACGAGTGTA AACACTGTTGTTAGAGATTTTGAAAATT ATGTGAAACAGACTTGTAATTCTAGCGCTGTTGTCGATAAC AAATTAAAGATACAAAACGTAATTATAGATGAATGTTACGG AGCCCCAGGATCTCCAACAAATTTGGAATTTATTAATACAG GATCTAGCAAAGGAAATTGTGCCATTAAGGCGTTGATGCAA TTGACTACTAAGGCCACTACTCAAATAGCACCTAGACAAGTT GCTGGTACAGGAGTTCAGTTTTATATGATTGTTATCGGTGTT ATAATATTGGCAGCGTTGTTTATGTACTATGCCAAGCGTAT GCTGTTACATCCACCAATGATAAAATCAAACCTATTTTAG CCAATAAGGAAAACGTCCATTGGACTACTTACATGGACACA TTCTTTAGAACTTCTCCGATGATTATTGCTACCACGGATATA AAAAGTGAAGATTCTAGGCTGGAGCCTCGGTGGCCATGCTT CTTGCCCCCTTGGGCCTCCCCCAGCCCCTCCTCCCCTTCCTG CACCCGTACCCCGTGGTCTTTGAATAAAGTCTGAGTGGGC GGCACTAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAA AAA AAAAAAAAAAAAA
78	0901-J0 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTCTTCTTGGTGGCAGCAGCCACGCGAGTCCACTCCA TGCCTCAACAACTATCTCCTATTAATATAGAACTAAAAAAC CAATTTCTGACGCGCGATTGAAGACGTTAGACATACATTAT AATGAGTCGAAACCAACCACTATCCAGAATACTGGAAAAC AGTAAGGATTAATTTTAAAGGAGGATATATAAGTGGAGGGT TTCTCCCCAATGAATATGTGTTATCAACAATACATATATATC GGAAAGGAAGACGATTATGGATCCAATCACTTGATAGATGT ATACAAATACTCTGGAGAGATTAATCTGGTTCATTGGAATA AGAAAAATATAGTTCTTATGAAGAGGCAAAAAAACACGA TGATGGAATTATCATTATTGCTATATTCTTACAAGTATCAGAT ATAAAAATGTATATTTTCAAAAGATAGTCAATCAATTGGAT TCCATTAGATCCGCTAATATGTCTGCGCCATTTGATTCAGTAT TTATCTAGACAATTTGCTACCTAGTACGTTGGATTATTTTACA ATCTAGGAACAACATCAACCACTCTGCAGACGCAGCATGG ATAATTTTCCAACACCAATAAACATTCATTCTGATCAACTA

		TCTAAATTCAGAACACTATTGTCGTCGTCTAATCATGAGGGT AACCTCATTATATAACAGAGAATTATAGAAATCCGTATAAA TTGAACGACGACACGCAAGTATATTATTCTGGGGAGATTATA CGAGCTGCAACCACCTCACCAGTTCGAGAGAACTATTTTAT GAAATGGTTGTCTGATTTGAGAGAGGCATGTTTTTCATATTA TCAAAAATATATCGAAGGGAATAAAACATTTCGAATTATTGC CATAGTATTTCGTATTTATACTCACCGCTATTCTCTTTCTTATG AGTCAACGATATTCGCGAGAAAAACAAAACCTAGGAATTCT AGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCC TCCCCCAGCCCCCTCCTCCCCTTCCTGCACCCGTACCCCCGT GGTCTTTGAATAAAGTCTGAGTGGGCGGCACTAGTAAAAAA AA AA AA
79	1101 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACGTG GATTCTCTTCCTGGTGGCAGCCGCCACTCGTGTGCACTCCGA TGGGACCTTGTTCCCTGGGGACGACGACCTGGCCATCCCCG CTACGGAGTTCTTCAGCACCAAGGCTGCCAAGAACCCGGAG ACTAAGCGGGAGGCCATCGTCAAGGCCTACGGAGATGACA ACGAGGAGACGCTGAAGCAGCGGCTGACCAACCTGGAGAA GAAGATCACGAACATCACGACCAAGTTCGAGCAGATTGAG AAGTGCTGCAAGCGGAACGACGAGGTGTTGTTCCGCTTGGA GAACCACGCAGAGACTCTGCGGGCAGCCATGATCAGCTTG CGAAGAAGATCGACGTGCAGACTGGACGCCACCCCTATGA GGGAGGAGGCGGCAGCGGCGGCGGCAGCCGGCTGAAT CAGTGCATGAGCGCCAACAAGGCCGCCATCACTGATTCAGC CGTCGCCGTCGCTGCCGCCTCCTCCACTCATAGGAAGGTGG TGTCCAGTACGACACAGTACGACCATAAGGAGAGCTGTAAT GGGCTGTACTACCAGGGGTCGTGCTACATCCTGCACAGCGA CTACAAGTCGTTTGAGGATGCGAAGGCGAACTGCGCAGCCG AGAGCAGCACGTTGCCTAACAAGAGCGACGTGCTGACTACT TGGCTGATTGACTATGTCGAGGATACCTGGGGCTCGGATGG AACCCCATCACGAAGACCACGTCTGACTACCAGGACTCCGA TGTGTCTCAGGAGGTGCGGAAGTACTTCTGTACTGGCGGCG GCGGTTCTGGTGGGGGCGGCAGCACCTGCACCGTCCCCACC ATGAACAACGCCAAGCTGACCAGTACAGAGACTTCCTTCAA CGATAAGCAGAAGGTGACATTTACATGCGATTCCGGCTACC ACTCGTTGGACCCCAACGCTGTCTGTGAGACAGACAAGTGG AAGTACGAGAATCCATGTAAGAAGATGTGCACGGTGTCCG ACTATGTTTCTGAGCTGTACGACAAGCCCCTTTACGAGGTG AACAGCACCATGACTCTCAGCTGCAACGGCGAGACTAAGTA CTTTCGGTGTGAGGAGAAGAACGGTAATACGTCCTGGAACG

	ATACCGTTACTTGCCCCAACGCCGAGTGCCAGCCGTTGCAG CTGGAGCATGGTAGCTGTCAGCCCGTGAAGGAGAAGTATA GTTTCGGCGAGTACATGACCATTAATTGTGATGTGGGTAC GAGGTCATCGGGGTGAGTTACATCTCTTGCACTGCCAATAG CTGGAACGTGATTCCCAGCTGTCAGCAGAAGTGTGACATAC CTTCTCTGTCCAACGGGCTGATCAGCGGGAGCACGTTTCAGC ATTGGCGGGGTGATCCACCTCAGCTGCAAGAGCGGGTTCAC GTTGACCGGGTCGCCCTCGTCGACCTGCATCGACGGGAAGT GGAATCCCATTCTGCCCACGTGCGTGCGGTGCAACGAGGAG TTCGACCCGGTCGACGATGGACCCGATGACGAGACTGATCT GTCTAAGTTGTGCAAGGACGTGTTTCAGTACGAGCAGGAGA TCGAGTCCCTGGAGGCCACGTACCACGGCGGTGGCGGGAGC GGGGGAGGTGGCAGCAGTCCCCAGACTTCCAAGAAGATCG GGGACGACGCCACCATCTCCTGCTCCCGCAACAACACTAAC TACTACGTGGTTATGTCCGCCTGGTACAAGGAGCCCAACTCG ATCATCCTGCTCGCTGCTAAGAGCGACGTCTGTACTTCGA CAACTATACCAAGGACAAGATCAGCTACGACTCACCTACG ATGACCTCGTGACCACCATCACAATTAAGAGTCTGACCGCC GGGGACGCCGGCACATACATTTGCGCCTTCTTCATGACGAG TACCACCAATGATACCGACAAGGTGGATTATGAGGAGTACTC GATCGAGCTTATCGTGAACACGGACAGTGAGTCGACCATCG ACATCATCCTGTCCGGTAGCACCCCTGAGACAATCTCCGAG AAGCCGGAGGACATCGACAATAGCAATTGCAGTTCCGTCTT CGAGATCACGACCCCTGAGCCCATTACCGACAACGTGGATG ATCACACCGACACGGTGACGTACACCAGCGACTCTATTAAC ACCGTCAATGCTTCCAGTGGGGAGTCCACCACTGACGAGAT TCCGGAGCCCATCACCGACAAGGAGGAGGATCACACCGTGA CCGACACGGTGAGTTACACCACCGTGTCTACCTCCTCTGGT ATTGTGACTACCAAGAGCACCCACGGACGACGCCGATCTGTA CGACACGTACAATGACAACGACACCGTGCCGCCACCACCG TGGGCGGGTCCACGACGTCCATTAGTAACTACAAGACTAAG GACTTCGTGGGGGGCGGGCGGTTCGGTGCGGGGGGCAGCG CTGAAACCAAGGAGTTCAAGACGCTGTACAACTTGTTTCATC GACTCTTACCTGCAGAAGTTGGCCCAGCACTCCATCCCTAC CAATGTCACTTGTGCCATCCACATTGGAGAGGTGATCGGCC AGTTCAAGAACTGCGCCCTGAGGATCACCAACAAGTGCATG AGCAATAGCCGCCTGTCCTTACGCTCATGGTGAGTCCCTT CATCGAGGTGATCTCCCTGCTGAGCGAGAAGGACAGGCGG GCTATTGCTGAGGAGATCGGTATTGACCTCGACGACGTGCC GTCCGTCGTCTCCAAGCTGGAGAAGAAGTCAACGCATATG CTGAGGTGAACAACATCATCGATATTCAGAAGCTGAATATC GGTGAGTGTTCGCTCCCCCTGGGCAGCATATGCTGTTGCA GATTGTCAATACCGGGAGTGCTGAGGCCAACTGCGGGTTCG
--	--

		AGACCATCGTGAAGAGTCTGAACAAGATTTACGTGCCCCCG ATCATCGAGAACCGCCTGCCCTACTACGACCCCGGTGGGGG CGGGTCTGGAGGAGGCGGCAGCGCGCTGCCGCCTCCATCC AGACCACGGTGAATACGTTGTCCGAGCGGATTTTCGAGCAAG CTGGAGCAGGAGGCCAACGCCTCTGCTCAGACCAAGTGTGA TATAGAGATTGGCAATTTCTATATCAGGCAGAACCACGGGT CAACATCACCGTGAAGAATATGTGCTCCGCTGACGCCGACG CTCAGCTGGACGCCGTGCTGAGCGCGGCCACGGAGACATAT TCCGGTCTGACCCCTGAGCAGAAGGCGTATGTGCCTGCCAT GTTACAGCTGCTCTGAACATCCAGACGTCCGTGAATACCG TGGTGAGGGACTTCGAGAACTACGTGAAGCAGACGTGTAAT AGCAGCGCTGTCGTCGATAATAAGCTGAAGATCCAGAACGT TATTATCGACGAGTGCTACGGCGCCCCCGGCTCCCCGACCAA CCTGGAGTTCATTAATACCGGTTCCCTCCAAGGGGAACTGCG CTATTAAGGCGCTGATGCAGCTGACTACGAAGGCGACGACC CAGATCGCGCCTCGCCAGGTGGCGGGGACCGGGGTGCAGTT CTACATGATCGTGATCGGTGTGATCATCCTGGCAGCTCTCTT TATGTACTACGCCAAGCGCATGCTGTTACCAGCACCAACGA CAAGATTAAGCTGATCCTCGCCAATAAGGAGAACGTCCACT GGACCACCTACATGGACACATTCTTCCGCACCAGCCCTATG ATTATCGCCACCACTGACATTCAGAACTGATAAGAATTCTA GGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTC CCCCCAGCCCCCTCCTCCCCCTTCTGCACCCGTACCCCCGTGG TCTTTGAATAAAAGTCTGAGTGGGCGGCACTAGTAAAAAAA AA AA AA
80	1201 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGACCCAGCTGCC TATCCTGCTGCTGCTGATCAGTCTGGTGTACGCTACACCGTC TCCCCAGACTTCCAAGAAGATCGGGGACGACGCGACGATCA GCTGCAGCAGGAATAACACCAACTACTACGTGGTGATGAGC GCCTGGTACAAGGAGCCGAATAGCATCATTCTGCTGGCCGC CAAGTCTGACGTGCTGTACTTCGACAATTACACAAAGGACA AGATCTCCTACGACAGTCCTTATGATGATCTGGTGACCACG ATCACCATTAAGTCTCTGACTGCTGGGGATGCCGGTACGTA TATCTGCGCCTTCTTCATGACGTCCACGACCAATGACACTGA TAAGGTGGATTACGAGGAGTACTCCATCGAGCTGATCGTGA CACCGACAGCGAGTCTACCATCGACATCATCCTGTCCGGGA GCACCCCCGAGACTATCAGTGAGAAGCCTGAGGATATTGAC AACAGCAATTGCAGTAGTGTCTTCGAGATCACCACCCCGGA GCCTATCACTGACAACGTGGACGATCACACCGACACGGTGA CGTACACCAGTGATAGCATCAACACCGTGAATGCTAGCTCC

		GGGGAGTCCACGACCGATGAGATCCCGGAGCCTATCACCG ATAAGGAGGAGGATCACACCGTGACCGACACGGTGAGTTA CACCACCGTGTCTACCTCCTCCGGGATCGTGACTACGAAGA GCACTACTGACGATGCTGATCTGTACGATACCTACAACGAT AACGATACCGTGCCGCCCACACCGTGGGCGGCAGCACCAC CAGCATCAGCAACTATAAGACTAAGGACTTCGTGGGGGGC GGGGGGAGCGGCGGCGGCGGCAGCCGGCTGAATCAGTGCA TGAGCGCCAACAAGGCCGCCATCACTGATTCAGCCGTGGCT GTCGCCGCCGCCAGCTCCACCCACCGGAAGGTGGTGTCCAG CACCACCCAGTACGATCACAAGGAGTCCTGCAATGGCCTGT ATTACCAGGGCTCTTGCTACATCCTGCACAGCGATTACAAG AGCTTTGAGGATGCCAAGGCCAATTGTGCAGCCGAGAGCA GCACGCTGCCTAACAAGAGCGACGTGCTGACTACTTGCTG ATCGACTATGTGGAGGACACCTGGGGGAGTGATGGGAACC CCATCACTAAGACCACGTCCGACTATCAGGACAGTGACGTG AGTCAGGAGGTGCGGAAGTACTTCTGCACCGGCGGCGGGG GCAGCGGCGGGGGCGGCAGCGCTGAGACAAAGGAGTTCAA GACCCTTTACAACCTGTTCATCGATAGTTATCTGCAGAAGCT GGCCCAGCACAGCATCCCCACCAATGTGACATGCGCCATTC ACATTGGTGAGGTGATTGGCCAGTTCAAGAACTGCGCCCTG CGGATAACTAACAAGTGTATGTCTAATCCCCGCCTGTCCTTC ACGCTCATGGTGGAGTCCTTCATCGAGGTGATCTCCCTGCT GAGCGAGAAGGACAGGCGGGCTATCGCCGAGGAGATCGGC ATCGACCTGGATGACGTCCCCAGCGTGGTGTCCAAGCTGGA GAAGAACTGCAACGCATATGCTGAGGTGAACAACATCATC GATATTCAGAAGCTGAATATCGGTGAGTGTTCCGCTCCCCC TGGGCAGCATATGCTGTTGCAGATCGTCAATACCGGGTCTGC CGAGGCCAACTGCGGGCTGCAGACCATCGTGAAGTCACTGA ATAAGATCTACGTGCCACCCATCATCGAGAACCGGCTGCCC TACTACGACCCGTGGTTCCTGGTGGGGGTGGCCATTATTCTG GTGATCTTCACGGTGGCCATCTGCAGCATCCGCAGGAACCT GGCCCTGAAGTACCGGTACGGTACCTTCCTCTACGTGTGAT AAGAATTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCC CCTTGGGCCTCCCCCAGCCCCCTCCTCCCCTTCCTGCACCCG TACCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCACTA GTAAA AAA AAA AAAAAA
81	1301 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTGTTCCCTGGTGGCGGCCGCCACAAGGGTGCACCTCG ATGGGACCTTGTTCCCTGGGGACGACGACCTGGCCATCCCC

		GCTACGGAGTTCTTCAGCACCAAGGCTGCCAAGAACCCGGA GACTAAGCGGGAGGCCATCGTCAAGGCCTACGGAGATGAC AATGAGGAGACGCTGAAGCAGCGTCTCACCAACCTTGAGA AGAAGATCACCAACATCACTACGAAGTTTGAGCAGATTGAG AAGTGCTGCAAGCGCAATGATGAGGTGCTGTTTCGGCTGGA GAACCACGCCGAGACTCTGCGGGCTGCCATGATCAGCCTCG CCAAGAAGATCGACGTGCAGACCGGAAGGCACCCCTATGA GGGGGGAGGAGGGAGCGGCGGCGGGAGCGGGGGCCGC GGCCAGTATCCAGACCACCGTGAACACTCTGTCTGAGCGTA TCAGCAGCAAGCTGGAGCAGGAGGCCAACGCCTCTGCTCA GACCAAGTGTGATATAGAGATTGGCAATTTCTATATCAGGC AGAACCACGGGTGCAACATCACCGTGAAGAATATGTGCTCC GCTGACGCCGACGCTCAGCTGGACGCCGTGCTGAGCGCGGC CACGGAGACATATTCCGGTCTGACCCCTGAGCAGAAGGCGT ATGTGCCTGCCATGTTACAGCTGCGCTGAATATCCAGACG AGTGTTAACACGGTGGTGCGGGACTTTGAGAACTACGTGAA GCAGACGTGCAACAGCTCAGCGGTGGTGGATAATAAGCTC AAGATCCAGAATGTGATCATTGACGAGTGTTATGGGGCCCC TGGGAGCCCCACTAACCTCGAGTTTATTAATACGGGCAGCAC CAAGGGCAACTGCGCTATCAAGGCGCTGATGCAGCTGACA ACGAAGGCCACCACCCAGATTGCTCCGCGGCAGGTGGCAG GAGGGGGAGGGGGGTCCGGCGGGGGAGGCTCGACTTGAC GGTGCCGACGATGAATAACGCCAAGCTGACCTCCACCGAG ACGTCGTTCAACGACAAGCAGAAGGTGACGTTACCTGCGA CAGCGGCTATCACAGCTTGGACCCCAACGCCGTCTGCGAGA CGGACAAGTGGAAGTACGAGAACCCGTGCAAGAAGATGTG CACGGTGTCCGACTATGTGTCCGAGCTGTACGATAAGCCGC TGTACGAGGTGAACAGCACCATGACGCTGTCGTGCAACGGC GAGACAAAGTACTTCCGGTGCGAGGAGAAGAACGGGAACA CCAGCTGGAACGATACAGTGACCTGTCCCAACGCAGAGTGC CAGCCTCTGCAGTTGGAGCACGGGTCTGTGACCCCGTTAA GGAGAAGTACAGCTTTGGCGAGTACATGACGATTAAGTGC ACGTCGGTTACGAGGTGATCGGCGTGAGTTACATCTCATGT ACCGCCAACAGCTGGAATGTCATACCTAGCTGCCAGCAGAA GTGTGACATTCCGTCTCTCTCCAACGGGCTGATCAGCGGGT CCACGTTCTCTATCGGCGGCGTTATTCATCTGTCCTGCAAGT CGGGCTTCACCCTGACCGGGTCCCCCTCTTCCACTTGATCG ACGGCAAGTGGAACCCCATCCTGCCCACCTGCGTGCGGAGC AACGAGGAGTTCGACCCCGTGGATGACGGGCCTGATGATG AGACTGATCTCTCTAAGCTGTCTAAGGACGTGGTCCAGTAC GAGCAGGAGATCGAGTCGCTGGAGGCCACCTACCACATCAT CATCATGGCCCTCACCATCATGGGGGTGATCTTCTTGATCA GCATCATCGTGCTGGTGTGTTCTTGCGACAAGAACAACGAC
--	--	--

		CAGTACAAGTTTCACAAGCTGCTGCCTTGATAAGAATTCTA GGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTC CCCCAGCCCCTCCTCCCCTTCTTGCACCCGTACCCCCGTGG TCTTTGAATAAAGTCTGAGTGGGCGGCACTAGTAAAAAAA AA AA AA
82	1401 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGAAGACCATCTCC GTGGTGACCCTGCTGTGCGTGCTGCCAGCAGTGGTGTACAG CACCTGTACGGTGCCGACGATGAATAACGCCAAGCTGACCT CCACCGAGACGTCGTTCAACGACAAGCAGAAGGTGACGTTCC ACCTGCGACAGCGGCTATCACAGCTTGGACCCCAACGCCGT CTGCGAGACGGACAAGTGGAAGTACGAGAACCCGTGCAAG AAGATGTGCACGGTGTCCGACTATGTGTCCGAGCTGTACGA TAAGCCGCTGTACGAGGTGAACAGCACCATGACGCTGTCTG GCAACGGCGAGACAAAGTACTTCCGGTGCGAGGAGAAGAA CGGGAACACCAGCTGGAACGATACAGTGACCTGTCCCAAC GCAGAGTGCCAGCCTCTGCAGTTGGAGCACGGGTCTCTGTCA GCCCCTTAAGGAGAAGTACAGCTTTGGCGAGTACATGACGA TTAACTGCGACGTCGGTTACGAGGTGATCGGCGTGAGTTAC ATCTCATGTACCGCCAACAGCTGGAATGTCATACCTAGCTG CCAGCAGAAGTGTGACATTCCGTCTCTCTCCAACGGGCTGA TCAGCGGGTCCACGTTCTCTATCGGCGGCGTTATTCTCTGTG TTGCAAGTCCGGCTTCACGTTGACCGGGTCGCCCTCGTCGA CCTGCATCGACGGGAAGTGGAATCCCATTCTGCCCACGTGC GTGCGGTTCGAACGAGGAGTTCGACCCGGTCGACGATGGGC CGGACGACGAGACAGATCTGTCAAAGTTGAGCAAGGATGT GGTTCAGTATGAGCAGGAGATCGAGAGCCTGGAGGCCACG TACCACGGGGGAGGGGGGAGCGGCGGCGGGAGCGGGG CCGCGGCCAGTATCCAGACCACCGTGAACACTCTGTCTGAG CGTATCAGCAGCAAGCTGGAGCAGGAGGCCAACGCCTCTG CTCAGACCAAGTGTGATATAGAGATTGGCAATTTCTATATC AGGCAGAACCACGGGTGCAACATCACCGTGAAGAATATGT GCTCCGCTGACGCCGACGCTCAGCTGGACGCCGTGCTGAGC GCGGCCACGGAGACATATTCCGGTCTGACCCCTGAGCAGAA GGCGTATGTGCCTGCCATGTTACAGCTGCGCTGAATATCC AGACGAGTGTTAACACGGTGGTGCGGGACTTTGAGAACTAC GTGAAGCAGACGTGCAACTCCAGCGCCGTGGTGGATAATA AGCTCAAGATCCAGAATGTGATCATTGACGAGTGTTATGGG GCCCCTGGGAGCCCCACTAACCTCGAGTTTATTAATACCGG TTCCTCCAAGGGGAAGTGCGCCATCAAGGCGCTGATGCAGT TGACAACGAAGGCGACGACCCAGATCGCGCCTCGGCAGGT

		GGCCGGCGGCGGTTGGCGGGTCAGGCGGCGGGGGGAGCGCC GCCGTGAAGACGCCGGTGATCGTGGTGCCCGTGATCGACCG GCCTCCGAGCGAGACGTTCCCCAACGTGCATGAGCATATCA ACGATCAGAAGTTCGATGACGTTAAGGACAACGAGGTGAT GCAGGAGAAGCGGGACGTGGTGATCGTGAATGACGACCCG GATCACTACAAGGACTACGTGTTTATCCAGTGGACCGGAGG CAACATCCGGGATGACGACAAGTACACCCATTTTTTCTCCG GCTTCTGTAAACCATGTGCACAGAGGAGACGAAGCGGAA CATCGCTCGCCACCTGGCACTGTGGGACTCCAAGTTCTTCA CAGAGCTGGAGAACAAGAATGTGGAGTACGTCGTCATCATC GAGAATGATAATGTGATTGAGGACATTACATTTCTCAGGCC GGTGCTCAAGGCCATCCATGACAAGAAGATCGACATCCTGC AGATGCGGGAGATCATCACCGGTAATAAGGTGAAGACTGA GCTGGTCATTGACAAGGACCACGCCATCTTCACCTATAACCG GTGGCTACGACGTCTCGCTGAGCGCGTATATCATACGCGTG ACCACCGCTCTTAATATCGTCGACGAGATCATTAAAGAGCGG TGGTCTGTCTCGGGCTTCTACTTCGAGATAGCCCGGATCGA GAACGAGATGAAGATCAACCGCCAGATCATGGACAACCTCA GCGAAGTACGTCGAGCATGATCCCCGCCTGGTGGCCGAGCA CCGGTTTGAGACAATGAAGCCCAACTTCTGGTCCCGGATAG GCACCGTGGCCGCCAAGCGGTATCCGGGCGTCATGTACACG TTCACCACCCCGCTCATCTCATTCTTCGGCCTGTTTCGACATTA ACGTCATCGGCCTGATCGTGATACTGTTTCATCATGTTTCATGC TGATCTTCAACGTGAAGAGCAAGTTGCTGTGGTTTCTGACA GGAACCTTCGTGACTGCCTTCATTTGATAAGAATTCTAGGCT GGAGCCTCGGTGGCCATGCTTCTTGCCCCCTGGGCCTCCCCC CAGCCCCCTCCTCCCCCTTCTGCACCCGTACCCCCGTGGTCTT TGAATAAAGTCTGAGTGGGCGGCACTAGTAAAAAAAAAAAAA AA AA AA
83	1501 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGAAGACCATCTC GGTGGTGACACTGCTGTGCGTGTTGCCGGCGGTTGTTTATA GTACGTGCACCGTGCCGACTATGAACAACGCCAAGTTGACG TCTACTGAGACTTCTTTCAATGACAAGCAGAAGGTGACGTT CACCTGCGACAGCGGCTATCACAGCTTGGAACCCCAACGCCG TCTGCGAGACGGACAAGTGGAAGTACGAGAACCCGTGCAA GAAGATGTGCACGGTGTCCGACTATGTGTCCGAGCTGTACG ATAAGCCGCTGTACGAGGTGAACAGCACCATGACGCTGTCC TGCAACGGCGAGACAAAGTACTTCCGTTGCGAGGAGAAGA ACGGCAACACCAGCTGGAATGACACCGTGACCTGTCCGAAC GCCGAGTGCCAGCCCCTGCAGTTGGAGCATGGCTCCTGCCA

	GCCCCGTGAAGGAGAAGTACTCCTTCGGGGAGTACATGACTA TCAACTGCGACGTGGGCTACGAGGTGATCGGGGTGTCTTATA TCAGCTGCACCGCCAATAGCTGGAACGTGATTCCCAGCTGTC AGCAGAAGTGTGACATACCTTCTCTGTCCAACGGGCTGATC AGCGGGAGCACGTTTCAGCATTGGCGGGGTGATCCACCTGAG TTGCAAGTCTGGCTTCACGCTGACCGGATCTCCTAGCTCGAC CTGCATCGACGGGAAGTGGAACCCCATCCTGCCAACTTGCG TCCGGTCCAACGAGGAGTTCGACCCCGTGGACGACGGACCA GACGACGAAACCGATCTGAGCAAGTTGAGCAAGGATGTCTG TGCAGTACGAGCAGGAGATCGAGTCCCTGGAGGCGACCTAT CACGGTGGGGGAGGGAGCGGCGGCGGCGGCTCACGGCTGA ATCAGTGCATGAGCGCCAACAAGGCCGCCATCACTGATTCA GCCGTGGCCGTGCGCGCCGCTTCCTCCACCCACCGGAAGGT CGTCTCCAGCACCAACCCAGTATGATCACAAGGAGTCCTGCA ACGGGCTGTACTATCAGGGCAGTTGCTACATCCTGCACAGC GACTACAAGTCGTTTGAGGATGCTAAGGCCAACTGCGCCGC CGAAAGCTCCACGCTGCCGAATAAGTCGGATGTCCTGACCA CGTGGCTCATCGACTACGTGGAGGACACCTGGGGCTCGGAT GGGAACCCCATCACGAAGACCACCAGCGACTATCAGGACA GCGATGTGAGCCAGGAGGTGCGTAAGTACTTCTGCACCGGG GGCGGAGGGTCTGGAGGAGGCGGCAGCGGCGCTGCCGCCT CCATCCAGACCACTGTGAACACCCTGTCCGAGCGCATTTCG TCCAAGCTGGAGCAGGAGGCCAACGCCTCCGCCCAGACGA AGTGCGACATCGAGATAGGGAATTTCTACATCCGCCAGAAC CACGGGTGCAACATCACCGTGAAGAATATGTGCTCCGCTGA CGCCGACGCTCAGCTGGACGCCGTGCTGAGCGCGGCCACGG AGACATATTCCGGTCTGACCCCGGAGCAGAAGGCTTACGTG CCTGCCATGTTTACGGCGGCCCTGAATATACAGACTTCTGT GAATACGGTGGTGCGGGACTTCGAGAATTACGTGAAGCAG ACTTGCAACTCCAGCGCCGTGGTGGATAATAAGCTCAAGAT CCAGAATGTGATCATTGACGAGTGTTATGGGGCCCCTGGGA GCCCCACTAACCTCGAGTTTATTAATACCGGTTCTCTCAAG GGGAACTGCGCCATCAAGGCGCTGATGCAGCTGACTACTAA GGCCACCACCCAGATCGCCCCACGTCAGGTGGCTGGCACCG GCGTTCAGTTCTACATGATCGTGATCGGGGTGATCATTCTG GCTGCGCTGTTTCATGTACTACGCCAAGCGCATGCTCTTTACT TCGACCAACGATAAGATCAAACCTTATCTTGGCGAATAAGGA AATGTGCATTGGACGACGTACATGGACACGTTCTTCCGGAC GTCTCCTATGATCATTGCGACCACCGATATTCAGAACTGAT AAGAATTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCC CCTTGGGCCTCCCCCAGCCCCCTCCTCCCCTTCTGCACCCG TACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCACTA GTAA
--	---

		AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAA
84	1601 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACATG GATTCTGTTCTTGGTGGCCGCCGCCACCAGGGTGCACAGCG ACGGCACCCCTGTTCCCAGGCGACGACGACCTGGCCATCCCC GCTACCGAGTTCTTCAGCACCAAGGCTGCCAAGAACCCCGA GACTAAGCGGGAGGCCATCGTGAAGGCCTACGGGGATGAT AACGAGGAGACTCTGAAGCAGCGGCTGACCAACCTGGAGA AGAAGATTACCAACATTACCACCAAGTTCGAGCAGATCGAG AAGTGCTGCAAGCGGAACGATGAGGTGCTGTTTCGGCTGGA GAACCACGCCGAGACACTGCGCGCCGCCATGATCAGCCTGG CCAAGAAGATCGACGTGCAGACCGGCCGCCACCCCTACGA GGGGGGCGGCGGCTCTGGGGGCGGCGGGAGCAGCCCCCAG ACTTCCAAGAAGATCGGGGATGATGCCACCATCAGCTGCTC CCGGAACAACACCAACTACTACGTGGTGATGTCCGCCTGGT ACAAGGAGCCTAACAGCATCATCCTGCTGGCCGCCAAGAGC GATGTGCTGTACTTTGACAATTACACCAAGGATAAGATCAG CTACGACAGCCCCTACGACGATCTGGTGACTACCATCACCA TCAAGAGCCTGACAGCCGGCGACGCAGGCACCTACATCTGC GCCTTCTTCATGACCAGCACCACCAATGACACCGACAAGGT GGATTACGAGGAGTACTCCATCGAGCTGATCGTGAATACTG ACTCAGAGTCTACCATCGACATTATCCTGTCCGGGAGCACC CCCGAGACTATCAGTGAGAAGCCTGAGGACATCGACAACA GCAACTGCAGCAGTGTGTTTCGAGATCACCACCCAGAGCCT ATCACTGACAACGTGGATGATCACACCGACACTGTGACATA CACCAGTGATAGCATCAACACCGTGAATGCTAGCTCTGGGG AGAGCACCCTGATGAGATCCCCGAGCCTATCACCGACAAG GAGGAGGATCACACCGTGACCGACACTGTGAGTTACACCAC AGTGTCTACCTCCTCTGGGATCGTGACCACCAAGAGCACCA CCGACGATGCCGACCTGTACGATACATAACAATGATAACGAT ACAGTGCCTCCCACCACCGTGGGAGGCAGCACCACCAGCAT TTCTAATTACAAGACCAAGGACTTCGTGGGGGGCGGGGGGA GCGGGGGCGGGGGGAGCGCCGAAACCAAGGAGTTCAAGAC CCTGTACAACCTGTTTCATCGATAGCTACCTGCAGAAGCTGG CCCAGCACAGCATTCCCACCAATGTGACATGCGCCATTAC ATTGGGGAGGTGATCGGCCAGTTCAAGAACTGCGCCCTGAG GATCACCAATAAGTGCATGAGCAATAGCCGCCTGTCCTTCA CTCTGATGGTGGAGTCCTTCATCGAGGTGATCTCCCTGCTGT CAGAGAAGGACAGGCGGGGCTATTGCCGAGGAGATCGGGAT CGACCTGGACGACGTGCCAGCGTGGTGTCCAAGCTGGAGA AGAACTGCAACGCCTACGCCGAGGTGAACAACATCATCGAT

		ATTCAGAAGCTGAATATCGGTGAGTGTTCGCCCCCTCCCGG CCAGCACATGCTGCTGCAGATCGTGAATACAGGGAGCGCTG AGGCCAACTGCGGGCTCCAGACCATCGTGAAGAGCCTGAAT AAGATTTACGTGCCTCCTATCATCGAGAATCGGCTGCCCTA CTACGACCCCTGGTTCCTGGTGGGCGTGGCCATTATCCTGGT GATCTTCACCGTGGCCATCTGTAGCATCAGGCGGAACCTGG CCCTGAAGTACAGGTACGGGACCTTCCTGTACGTGTGATAA GAATTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCCT TGGGCCTCCCCCAGCCCCCTCCTCCCCTTCCTGCACCCGTAC CCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCACTAGT AA AA AA AAAA
85	1701 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGACCCAGCTGCC CATCCTGCTGCTGCTGATCAGCCTGGTGTACGCCACTCCCA GTCCTCAGACCTCCAAGAAGATTGGCGACGATGCCACCATC AGCTGCTCCCGGAATAATACCAACTACTATGTGGTGTATGAG CGCCTGGTACAAGGAGCCTAACAGCATCATCCTGCTGGCCG CCAAGAGCGATGTGCTGTACTTCGATAACTACACCAAGGAC AAGATCAGCTACGATAGCCCCTACGACGATCTGGTGACTAC CATCACCATCAAGAGTCTGACTGCCGGCGACGCCGGGACTT ACATCTGCGCCTTCTTCATGACCAGCACCACCAATGACACC GACAAGGTGGACTACGAGGAGTACAGCATCGAGCTGATCG TGAACACCGACAGCGAGAGCACCATCGACATCATCCTGTCT GGCAGCACCCCCGAGACTATCAGCGAGAAGCCTGAGGACA TCGATAACAGCAACTGTTCCAGCGTTTTTCGAGATCACCACCC CCGAGCCTATCACCGATAACGTGGACGATCACACCGACACC GTGACCTACACCTCCGATAGCATCAACACCGTGAATGCCAG CAGCGGGGAGAGTACTACAGATGAGATCCCCGAGCCTATC ACCGATAAGGAGGAGGATCACACCGTGACCGACACTGTGA GTTACACCACAGTGTCTACCTCCTCCGGCATCGTGACTACC AAGAGTACCACCGACGATGCCGATCTGTACGATACCTACAA CGATAACGACACAGTGCCTCCCACCACCGTGGGAGGCAGC ACCACCAGCATCAGCAACTACAAGACCAAGGATTTTGTGGG CGGCGGGGGCTCCGGCGGCGGAGGCTCCGCCGCCGTGAAG ACCCCCGTGATTGTGGTGCCTGTGATCGATCGGCCCCCTAG CGAGACATTTCCCAACGTGCACGAGCACATCAACGATCAGA AGTTCGATGACGTGAAGGACAACGAGGTGATGCAGGAGAA GCGGGACGTGGTGTATCGTGAACGACGATCCTGACCACTACA AGGACTACGTGTTTCATCCAGTGGACCGGGGGCAACATCCGG GATGACGACAAGTACACCCACTTCTTCTCCGGCTTCTGTAA

		CACCATGTGTACAGAGGAGACAAAGCGGAACATTGCTCGG CACCTGGCCCTGTGGGACAGCAAGTTCTTCACTGAGCTGGA GAACAAGAACGTGGAGTATGTGGTGATCATCGAGAACGAC AATGTGATCGAGGACATTACCTTCCTCAGGCCTGTGCTGAA GGCCATCCACGATAAGAAGATCGACATCCTGCAGATGCGG GAGATCATCACCGGGAATAAGGTGAAGACCGAGCTGGTGA TCGACAAGGATCACGCCATCTTCACCTACACCGGCGGCTATC ATGTGAGCCTGAGCGCTTACATCATCCGGGTGACCACCGCC CTGAATATCGTGGATGAGATCATTAAAGTCCGGGGGGCTGTC CAGCGGGTTCTACTTCGAGATTGCTAGGATCGAGAACGAGA TGAAGATCAACCGGCAGATCATGGATAATTCTGCCAAGTACC TGGAGCACGACCCTCGGCTGGTGGCTGAGCATCGCTTCGAG ACTATGAAGCCCAACTTCTGGAGCAGGATTGGGACCGTGGC CGCCAAGCGGTACCCAGGGGTGATGTACACCTTTACCACAC CCCTGATCAGCGGCGGCGGGGGCAGCGGCGGGGGCGGCAG CGCTGAGACAAAGGAGTTCAAGACCCTGTACAACCTGTTTA TCGACAGCTATCTGCAGAAGCTGGCTCAGCACTCTATTCCCA CCAATGTGACATGCGCCATTCACATTGGGGAGGTGATCGGC CAGTTCAAGAACTGCGCCCTGAGGATCACCAATAAGTGCAT GAGCAACAGCAGGCTGAGCTTCACCCTGATGGTGGAGTCCT TCATCGAGGTGATCAGCCTGCTGAGCGAGAAGGACAGGAGG GCCATCGCCGAGGAGATCGGCATCGACCTGGACGATGTGCC CAGCGTGGTGAGCAAGCTGGAGAAGAATTGTAACGCCTAC GCCGAGGTGAACAATATCATCGATATTCAGAAGCTGAATAT CGGGGAGTGCAGTGCTCCCCCTGGGCAGCACATGCTGCTGC AGATTGTGAACACCGGCAGCGCCGAGGCCAATTGCGGGCT GCAGACCATTGTGAAGTCACTGAATAAGATCTACGTGCCAC CCATCATCGAGAACCGGCTGCCATACTACGATCCCTGGTTC CTGGTGGGGGTGGCCATTATTCTGGTGATCTTCACAGTGGC CATCTGCAGCATCAGGCGCAACCTGGCTCTGAAGTACAGGT ACGGGACCTTCCTGTACGTGTGATAAGAATTCTAGGCTGGA GCCTCGGTGGCCATGCTTCTTGCCCCCTGGGCCTCCCCCAG CCCCTCCTCCCCCTTCCTGCACCCGTACCCCGTGGTCTTTGA ATAAAGTCTGAGTGGGCGGCACTAGTAAAAAAAAAAAAAAAA AA AA AA
86	1801 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGACCCAGCTGCC TATCCTGCTGCTGCTGATCAGCCTGGTGTACGCTACCCCTAG CCCCAGACTTCCAAGAAGATCGGGGACGACGCCACGATCA GCTGCAGCAGGAATAACACCAACTACTACGTGGTGTATGAGC GCCTGGTACAAGGAGCCAAATAGCATCATTCTGCTGGCCGC

	CAAGTCTGACGTGCTGTACTTCGACAATTACACAAAGGACA AGATCTCCTACGACAGTCCTTATGATGATCTGGTGACCACG ATCACCATTAAGTCTCTGACTGCTGGGGATGCCGGTACCTA TATCTGTGCGTTCTTTATGACCTCTACAACGAACGACACAGA TAAGGTGGACTACGAGGAGTATAGCATCGAGCTGATCGTGA ACACCGACAGCGAGAGCACAATTGACATTATCCTCAGCGGG AGCACCCCCGAGACAATCTCTGAGAAGCCCCGAGGACATTG ATAATAGCAATTGTAGCTCCGTGTTCGAGATTACTACTCCC GAGCCCATTACCGATAATGTGGATGATCATAACAGATACGGT GACCTACACCAGCGACAGCATCAACACCGTGAATGCTAGCA GCGGTGAGAGCACCACCGATGAGATCCCGGAGCCTATCACC GATAAGGAGGAGGATCACACCGTGACCGACACGGTGAGTT ACACCACCGTGTCTACCTCCTCCGGGATCGTGACTACCAAG AGCACACGGATGACGCCGATCTGTATGACACCTACAACGA TAATGATACCGTGCCGCCACCACCGTGGGCGGCAGCACCA CCAGCATCAGCAACTATAAGACTAAGGACTTCGTGGGGGGC GGGGGGAGCGGCGGCGGCGGCAGCCGGCTGAATCAGTGCA TGAGCGCCAACAAGGCCGCCATCACTGATTCAGCCGTGGCT GTCGCCGCCGCCAGCTCCACCCACCGGAAGGTGGTGTCCAG CACCACCCAGTACGATCACAAGGAGTCCTGCAATGGCCTGT ATTACCAGGGCTCTTGCTACATCCTGCACAGCGATTACAAG AGCTTTGAGGATGCCAAGGCCAATTGTGCAGCCGAGAGCA GCACGCTGCCTAACAAGAGCGACGTGCTGACTACTTGGCTG ATCGACTATGTGGAGGACACCTGGGGCAGTGATGGGAACC CCATCACTAAGACCACCTCTGACTACCAGGACTCCGACGTG AGTCAGGAGGTGAGGAAGTACTTCTGTACGGGCGGCGGCG GCAGCGGCGGCGGCGGGTCCACCTGCACCGTGCCACCATG AACACGCCAAGCTGACTTCTACCGAGACTTCTTTCAACGA CAAGCAGAAGGTGACCTTCACCTGCGACAGCGGCTATCATT CCCTCGACCCCTAACGCCGTGTGCGAGACAGACAAGTGGA GTATGAGAACCCGTGCAAGAAGATGTGTACGGTGTCCGACT ACGTCTCTGAGCTGTACGACAAGCCTCTGTACGAGGTCAAC TCCACTATGACTCTGTCTTGCAACGGCGAGACAAAGTACTTT AGGTGCGAGGAGAAGAACGGCAATACCAGCTGGAACGACA CCGTGACGTGCCCCAACGCCGAGTGTGAGCCTCTCCAGCTG GAGCACGGCAGCTGTCAGCCTGTGAAGGAGAAGTACAGCT TCGGCGAGTACATGACCATCAACTGTGATGTGGGGTACGAG GTGATCGGGGTGAGCTATATCAGCTGCACCGCCAATAGCTG GAACGTGATTCCCAGCTGTCAGCAGAAGTGCGACATCCCTT CTCTGTCCAACGGGCTGATCAGCGGGAGCACGTTTCAGCATT GGCGGGGTGATCCACCTGAGCTGCAAGTCAGGGTTCACCT GACTGGCTCACCCAGCAGCACCTGCATCGATGGCAAGTGGA ACCCATTCTGCCACCTGTGTGAGAAGCAACGAGGAGTTC
--	---

		GATCCTGTGGACGACGGACCAGACGACGAAACCGATCTGTC CAAGCTGTCCAAGGATGTGGTGCAGTACGAGCAGGAGATC GAGAGCCTGGAGGCCACCTACCACATCATCATCATGGCACT GACCATCATGGGCGTGATATTCTTGATCAGCATCATTGTGCT GGTGTGTTCTTGCGACAAGAACAACGACCAGTACAAGTTTC ACAAGCTGCTGCCCTGATAAGAATTCTAGGCTGGAGCCTCG GTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCAGCCCCCTC CTCCCCTTCCTGCACCCGTACCCCGTGGTCTTTGAATAAAG TCTGAGTGGGCGGCACTAGTAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
87	1901 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGACTGGACCTG GATTCTGTTCCCTGGTGGCGGCCGCCACAAGGGTGCCTCCG ATGGGACCTTGTTCCCTGGGGACGACGACCTGGCCATCCCC GCTACGGAGTTCTTCAGCACCAAGGCTGCCAAGAACCCGGA GACTAAGCGGGAGGCCATCGTCAAGGCCTACGGAGATGAC AATGAGGAGACGCTGAAGCAGCGTCTCACCAACCTTGAGA AGAAGATCACCAACATCACTACGAAGTTTGAGCAGATTGAG AAGTGCTGCAAGCGCAATGATGAGGTGCTGTTTCGGCTGGA GAACCACGCCGAGACTCTGCGGGCTGCCATGATCAGCCTCG CCAAGAAGATCGACGTGCAGACCGGAAGGCACCCCTATGA GGGGGGAGGAGGGAGCGGCGGCGGGAGCGGGGCGGC GGCCAGTATCCAGACCACCGTGAACACTCTGTCTGAGCGTA TCAGCAGCAAGCTGGAGCAGGAGGCCAACGCCTCTGCTCA GACCAAGTGTGATATAGAGATTGGCAATTTCTATATCAGGC AGAACCACGGGTGCAACATCACCGTGAAGAATATGTGCTCC GCTGACGCCGACGCTCAGCTGGACGCCGTGCTGAGCGCGGC CACGGAGACATATTCCGGTCTGACCCCTGAGCAGAAGGCGT ATGTGCCTGCCATGTTACAGCTGCGCTGAATATCCAGACG AGTGTTAACACGGTGGTGCGGGACTTTGAGAACTACGTGAA GCAGACGTGCAACTCCAGCGCCGTGGTGGATAATAAGCTCA AGATCCAGAATGTGATCATTGACGAGTGTTATGGGGCCCCT GGGAGCCCCACTAACCTCGAGTTTATTAATACCGGTTCCCTC CAAGGGGAACTGCGCCATCAAGGCGCTGATGCAGTTGACA ACGAAGGCCACCACCCAGATCGCCCCAGGCAGGTTGCCG GGGGCGGCGGGGGGAGCGGTGGCGGCGGCAGCGCTGAGAC AAAGGAGTTCAAGACCCTGTACAACTTGTTTCATAGATTCTT ATCTTCAGAAGTTGGCCCAGCACTCCATCCCTACCAATGTC ACTTGTGCCATCCACATTGGAGAGGTGATCGGCCAGTTCAA GAACTGCGCCCTGAGGATCACCAACAAGTGCATGAGCAAT AGCCGCCTGTCCTTCACGCTCATGGTGGAGTCCTTCATCGA

		GGTGATCTCCCTGCTGAGCGAGAAGGACAGGCGGGCTATTG CTGAGGAGATCGGTATTGACCTCGACGACGTGCCGTCCGTC GTCTCCAAGCTGGAGAAGAACTGCAACGCATATGCTGAGGT GAACAACATCATCGATATTCAGAAGCTGAATATCGGTGAGT GTTCCGCTCCCCCTGGGCAGCATATGCTGTTGCAGATTGTCA ATACCGGGAGTGCTGAGGCCAACTGCGGGCTGCAGACCATC GTGAAGTCACTGAATAAGATCTACGTGCCACCCATCATCGA GAACAGGCTGCCGTACTACGACCCGTGGTTCTTGGTGGGGG TGGCCATTATTCTGGTGATCTTCACGGTGGCCATCTGCAGCA CCGCCGGAACCTGGCTCTGAAGTACCGGTACGGGACCTTCC TGTACGTGTGATAAGAATTCTAGGCTGGAGCCTCGGTGGCCA TGCTTCTTGCCCCCTTGGGCCTCCCCCAGCCCCCTCCTCCCCT TCCTGCACCCGTACCCCCGTGGTCTTTGAATAAAGTCTGAGT GGGCGGCACTAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AA
88	2001 DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACCATGGATTGGACCTG GATTCTGTTCCCTCGTTGCCGCCGCCACCAGGGTGCATAGTG CGGCAGTGAAGACTCCCGTTATCGTCGTCCCGGTGATTGAC CGCCCGCCCTCGGAGACGTTCCCCAACGTGCATGAGCACAT CAACGATCAGAAGTTCGACGATGTCAAGGACAATGAGGTC ATGCAGGAGAAGCGGGACGTGGTGATCGTCAACGACGATC CCGATCACTACAAGGATTACGTCTTCATCCAGTGGACGGGC GGCAACATCCGGGACGACGATAAGTACACGCATTTCTTCTC CGGCTTCTGTAACACCATGTGTACAGAGGAGACGAAGCGG AACATCGCCAGGCATCTGGCGCTGTGGGACTCGAAGTTCTT CACCGAGCTGGAGAATAAGAACGTGGAGTACGTGGTGATC ATCGAGAACGACAACGTCATCGAGGACATCACGTTTCTGCG GCCGGTGCTGAAGGCCATCCACGACAAGAAGATCGACATC CTGCAGATGCGGGAGATCATCACGGGTAATAAGGTGAAGA CCGAGCTGGTCATTGACAAGGACCACGCGATATTCACCTAC ACTGGGGGATACGACGTGTCCTTGTCCGCCTATATTATCCG GGTGACCACCGCTCTGAACATCGTGGATGAGATCATCAAGT CGGGAGGCCTCAGCTCCGGCTTCTACTTCGAGATCGCCCGG ATCGAGAACGAGATGAAGATCAACCGCCAGATCATGGACA ACAGCGCCAAGTATGTTCGAGCACGACCCTCGGCTGGTGGCA GAGCACAGATTCGAGACTATGAAGCCCAACTTCTGGAGTAG GATCGGCACTGTCGCCGCCAAGCGCTACCCAGGTGTGATGT ACACCTTCACCACCCCGCTGATCAGCGGGGGGGGGGGTAGC GGCGGCGGCGGCAGTGCCGAGACAAAGGAGTTCAAGACCC TGTACAACTTGTTTCATCGACTCCTACCTCCAGAAGTTGGCTC

		AGCACTCGATCCCCACCAATGTGACATGCGCCATTACATT GGTGAGGTGATCGGGCAGTTCAAGAACTGTGCTCTGCGGAT TACCAACAAGTGTATGAGTAATTCCCGCCTGTCCTTCACGCT CATGGTGGAGTCCTTCATCGAGGTGATCTCCCTGCTGAGCG AGAAGGACAGGCGGGGCCATCGCCGAGGAGATCGGCATCGA CCTGGACGATGTTCCGTCTGTCGTCTCCAAGCTGGAGAAGA ACTGCAACGCATATGCTGAGGTGAACAACATCATCGATATT CAGAAGCTGAATATCGGTGAGTGTTCCGCTCCCCCTGGGCA GCATATGCTGTTGCAGATCGTGAATAACGGGAGTGCGGAGG CGAACTGCGGGCTCCAGACCATCGTGAAGAGCCTGAACAA GATCTACGTCCCGCCTATCATCGAGAATCGGCTGCCGTACT ATGACCCTGGTGGCGGGCGGCAGCGGGGGGGGGGGGGTTCGGG GGCCGCTGCTTCCATCCAGACCACCGTGAACACTCTGTCTG AGCGTATCAGCAGCAAGCTGGAGCAGGAGGCCAACGCCTC TGCTCAGACCAAGTGTGATATAGAGATTGGCAATTTCTATAT AGGCAGAACCACGGGTGCAACATCACCGTGAAGAATATGT GCTCCGCTGACGCCGACGCTCAGCTGGACGCCGTGCTGAGC GCGGCCACGGAGACATATTCCGGTCTGACCCCTGAGCAGAA GGCGTATGTGCCTGCCATGTTACAGCTGCGCTGAATATCC AGACGAGTGTTAACACGGTGGTGCGGGATTTCGAGAATTAT GTGAAGCAGACCTGTAACAGCTCGGCTGTTGTGGACAACAA GCTCAAGATCCAGAATGTGATCATTGACGAGTGTTACGGGG CCCCCGGCTCCCCCACGAATCTGGAGTTCATCAACACGGGC AGCAGCAAGGGCAACTGCGCCATTAAGGCGCTGATGCAGC TGACCACTAAGGCCACAACCCAGATCGCCCCTAGGCAGGTG GCCGGGACTGGCGTGCAGTTCTATATGATAGTGATTGGCGT GATCATCCTGGCCGCCCTGTTTCATGTACTACGCCAAGCGCA TGCTCTTTACTTCGACCAACGATAAGATCAAACCTTATCTTGG CGAATAAGGAGAATGTGCATTGGACGACGTACATGGACAC CTTCTTTTCGCACATCACCTATGATCATTGCCACGACTGACAT TCAGAACTGATAAGAATTCTAGGCTGGAGCCTCGGTGGCCA GCTTCTTGCCCCTTGGGCCTCCCCCAGCCCCTCCTCCCCTT CCTGCACCCGTACCCCGTGGTCTTTGAATAAAGTCTGAGT GGGCGGCACTAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAA AA AA AAAAAAAAAAAAAAAAAAAAAAAA
89	Amino acid sequence encoded by 0201	MDWTWILFLVAAATRVHSDGTLFPGDDDLAIPATEFFSTKAA KNPETKREAIVKAYGDDNEETLKQRLTNLEKKITNITTKFEQIE KCKKRNDEVLFRLNHAETLRAAMISLAKKIDVQTGRHPYYP YDVPDYAE
90	Amino acid sequence	MMTPENDEEQTSVFSATVYGDKIQGKNKRKRVLGLCIRISMVI SLLSMITMSAFLIVRLNQCMSANKAAITDSAVAVAAASSTHRK

	encoded by 0302	VVSSTTQYDHKESCNGLYYQGSCYILHSDYKSFEDAKANCAA ESSTLPNKSDVLTWTLIDYVEDTWGSDGNPITKTTSDYQSDV SQEVRKYFCTYPYDVPDYA
91	Amino acid sequence encoded by 0402	MTQLPILLLLISLVYATPSPQTSKKIGDDATISCSRNNNTNYYVV MSAWYKEPNSSIILAAKSDVLYFDNYTKDKISYDSPYDDLVT ITIKSLTAGDAGTYICAFFMTSTTNDTDKVDYEEYSIELIVNTDS ESTIDIILSGSTPETISEKPEDIDNSNCSSVFEITTEPITDNVDDH TDTVTYTSDSINTVNASSGESTTDEIPEPITDKEEDHTVTDTVSY TTVSTSSGIVTTKSTTDDADLYDTYNDNDTVPTTVGGSTTSIS NYKTKDFVEIFGITTLILSAVAIFCITYYICNKHPRKYKTENKV YPYDVPDYA
92	Amino acid sequence encoded by 0505	MKTISVVTLCLVLPVAVVYSGGGGSYPYDVPDYATCTVPTMNN AKLTSTETSFNDKQKVFTFCDSGYHSLDPNAV CETDKWKYEN PCKKMCTVSDYVSELYDKPLYEVNSTMTLSCNGETKYFRCEE KNGNTSWNDTVTCPNAECQPLQLEHGSCQPVEKYSFGEYMT INCDVGYEVIGVSYISCTANSWNVIPSCQQKCDIPSLSNGLISGS TFSIGGVIIHLSCSGFTLTGSPSSTCIDGKWNIPPTCVRNNEEF DPVDDGPDDETDL SKLSKD VVQYEQEIESLEATYHIIIMALTIM GVIFLISIIVLVCSCDKNNDQYKFHKLLP
93	Amino acid sequence encoded by 0603	MDWTWILFLVAAATRVHSGGGGSYPYDVPDYAAETKEFKTL YNLFIDSYLQKLAQHSIPTNVTC AIHIGEVIGQFKNCALRITNKC MSNSRLSFTLMVESFIEVISLLSEKDRRAIAEEIGIDLDDVPSVV SKLEKNCNAYAEVNNIIDIQKLNIGECSAPPGQHMLLQIVNTGS AEANCGLQTIVKSLNKIYVPIIENRLPYDPWFLVGVAIILVIF TVAICSIRRNLALKYRYGTFLYV
94	Amino acid sequence encoded by 0702	MDWTWILFLVAAATRVHSGGGGSYPYDVPDYAGAAASIQT VNTLSERISSKLEQEANASAQTKCDIEIGNFYIRQNHGCNITVK NMCSADADAQLDAVL SAATETYSGLTPEQKAYVPAMFTAAL NIQTSVNTVVRDFENYVKQTCNSSAVVDNKLKIQNVIIDECY GAPGSPTNLEFINTGSSKGNCAIKALMQLTTKATTQIAPRQVAG
95	Amino acid sequence encoded by 0803	MDWTWILFLVAAATRVHSGGGGSYPYDVPDYAAAVKTPVIV VPVIDRPPSETFPNVHEHINDQKFDDVKDNEVMQEKRDVVIVN DDPDHYKDYVFIQWTGGNIRDDDKYTHFFSGFCNTMCTEETK RNIARHLALWDSKFFTELENKNVEYVVIENDNVIEDITFLRPV LKAIHDKKIDILQMREIITGNKVKTTELVIDKDHAIFTYTG GYDV SLSA YIIRVTTALNIVDEIHKSGGLSSGFYFEIARIENEMKINRQI MDNSAKYVEHDPRLVAEHRFETMKPNFWSRIGTVAAKRYPG VMYTFTTPLISFFGLFDINVIGLIVILFIMFMLIFNVKSKLLWFLT GTFVTAFI
96	Amino acid sequence encoded by 0201-J0	MDWTWILFLVAAATRVHSM DGTLPFGDDDLAIPATEFFSTKA AKNPETKREAIVKAYGDDNEETLKQRLTNLEKKITNITTKFEQI EKCCKRNDEVLFRLNHAETLRAAMISLAKKIDVQTGRHPYE

97	Amino acid sequence encoded by 0301-J0	MDWTWILFLVAAATRVHSMMPENDEEQTSVFSATVYGDKI QGKNKRKRVIGLCIRISMVISLLSMITMSAFLIVRLNQCMSANK AAITDSAVAVAAASSTHRKVVSSTTQYDHKESCNGLYYQGSC YILHSDYKSFEDAKANCAAESSTLPNKSDVLTWWLIDYVEDTW GSDGNPITKTTSQDYQSDVSQEVRYFCT
98	Amino acid sequence encoded by 0501-J0	MDWTWILFLVAAATRVHSMKTISVVTLLCVLPVVYSTCTVP TMNNAKLTTSTETSFNDKQKVTFCTDSGYHSLDPNAVCECTDKW KYENPCKKMCTVSDYVSELYDKPLYEVNSTMTLSCNGETKYF RCEEKNGNTSWNDTVTCPNAECQPLQLEHGSCQPVKEKYSFG EYMTINCDVGYEVIGVSYISCTANSWNVIPSCQQKCDIPSLNGL LISGSTFSIGGVIHLSCKSGFTLTGSPSSTCIDGKWNPIPTCVRS NEEFDPVDDGPDDDETDLKSLKSDVVQYEQEIESLEATYHIIIMA LTMGVIFLISHIVLCSCDKNNDQYKFHKLPL
99	Amino acid sequence encoded by 0701-J0	MDWTWILFLVAAATRVHSMGAAASIQTTVNTLSERISSKLEQE ANASQTKCDIEIGNFYIRQNHGCNITVKNMCSADADAQLDA VLSAATETYSGLTPEQKAYVPAMFTAALNIQTSVNTVVRDFEN YVKQTCNSSAVVDNKLKIQNVIIDECYGAPGSPTNLEFINTGSS KGNCAIKALMQLTTKATTQIAPRQVAGTGVQFYMIVIGVILA ALFMYYAKRMLFTSTNDKIKLILANKENVHWTTYMDTFFRTS PMIATTDIQN
100	Amino acid sequence encoded by 0901-J0	MDWTWILFLVAAATRVHSMPPQLSPINIETKKAISDARLKTLD IHYNESKPTTIQNTGKLVRINFKGGYISGGFLPNEYVLSTIHIYW GKEDDYGSNHLIDVYKYSGEINLVHWNKKKYSSYEEAKKHD DGIHIAIFLQVSDHKNVYFQKIVNQLDSIRSANMSAPFDSVFYL DNLLPSTLDYFTYLGTTINHSADAAWIIFPTPINIHSQLSKFRT LLSSSNHEGKPHYITENYRNPYKLNDDETQVYYSGEIIRAATTSP VRENYFMKWLSDLREACFSYYQKYIEGNKTFIHAIVFVILTA ILFLMSQRYRSREKQN
101	Amino acid sequence encoded by 1101	MDWTWILFLVAAATRVHSDGTLFPGDDDLAIPATEFFSTKAA KNPETKREAIKAYGDDNEETLKQRLTNLEKKITNITTKFEQIE KCKKRNDVFLFRLENHAETLRAAMISLAKKIDVQTGRHPYEG GGGSGGGGSRLNQCMSANKAAITDSAVAVAAASSTHRKVVS TTQYDHKESCNGLYYQGSCYILHSDYKSFEDAKANCAAESST LPNKSDVLTWWLIDYVEDTWGSDGNPITKTTSQDYQSDVSQEV RYKYFCTGGGGSGGGGSTCTVPTMNNNAKLTTSTETSFNDKQKV TFTCDSGYHSLDPNAVCECTDKWKYENPCKKMCTVSDYVSEL YDKPLYEVNSTMTLSCNGETKYFRCEEKNGNTSWNDTVTCPN AECQPLQLEHGSCQPVKEKYSFGYMTINCDVGYEVIGVSYIS CTANSWNVIPSCQQKCDIPSLNGLISGSTFSIGGVIHLSCKSGF TLTGSPSSTCIDGKWNPIPTCVRSNEEFDPVDDGPDDDETDLK LSKDVVQYEQEIESLEATYHGGGGSGGGGSSPQTSKKIGDDAT ISCSRNTNYYVVMASAWYKEPNSIILLAAKSDVLYFDNYTKD KISYDSPYDDLVTITIKSLTAGDAGTYICAFFMTSTTNDTDKV

		DYEEYSIELIVNTDSESTIDIILSGSTPETISEKPEDIDNSNCSSVF EITTPEPITDNVDDHTDTVTYTSDSINTVNASSGESTTDEIPEPIT DKEEDHTVTDTVSYTTVSTSSGIVTTKSTTDDADLYDTYNDND TVPPTTVGGSTTSISNYKTKDFVGGGSGGGGSAETKEFKTLY NLFIDSYLQKLAQHISIPTNVTCAIHIGEVIQFKNCALRITNKC MSNSRSLFTLMVESFIEVISLLSEKDRRAIAEEIGIDLDDVPSVV SKLEKNCNAYAEVNNIIDIQKLNIGECSAPPGQHMLLQIVNTGS AEANCGLQTIVKSLNKIYVPPIIENRLPYYPGGGGSGGGGSGA AASIQTTVNTLSERISSKLEQEANASAQTKCDIEIGNFYIRQNHG CNITVKNMCSADADAQLDAVLSAATETYSGLTPEQKAYVPA MFTAALNIQTSVNTVVRDFENYVKQTCNSSAVVDNKLKIQNV IIDECYGAPGSPTNLEFINTGSSKGNCAIKALMQLTTKATTQIAP RQVAGTGVQFYMIVIGVILALFMYAKRMLFTSTNDKIKLIL ANKENVHWTTYMDTFFRTSPMIATTDIQN
102	Amino acid sequence encoded by 1201	MTQLPILLLLISLVYATPSPQTSKIGDDATISCSRNNNTNYYVV MSAWYKEPNSIILLAAKSDVLYFDNYTKDKISYDSPYDDLVT ITIKSLTAGDAGTYICAFMTSTTNDTDKVDYEEYSIELIVNTDS ESTIDIILSGSTPETISEKPEDIDNSNCSSVFEITTPEPITDNVDDH TDTVTYTSDSINTVNASSGESTTDEIPEPITDKEEDHTVTDTVSY TTVSTSSGIVTTKSTTDDADLYDTYNDNDTVPPTTVGGSTTSIS NYKTKDFVGGGSGGGGSRNLNQCMSANKAAITDSAVAVAAA SSTHRKVVSSTTQYDHKESCNGLYYQGSCYILHSDYKSFEDAK ANCAAESSTLPNKSDVLTTWLIDYVEDTWGSDGNPITKTTSDY QSDVVSQEVRYFCTGGGGSGGGGSAETKEFKTLYNLFIDSYL QKLAQHISIPTNVTCAIHIGEVIQFKNCALRITNKCMSNSRSLF TLMVESFIEVISLLSEKDRRAIAEEIGIDLDDVPSVVSLEKNCN AYAEVNNIIDIQKLNIGECSAPPGQHMLLQIVNTGSAEANCGL QTIVKSLNKIYVPPIIENRLPYYPWFLVGVAIILVIFTVAICSIR RNLALKYRYGTFLYV
103	Amino acid sequence encoded by 1301	MDWTWILFLVAAATRVHSDGTLFPGDDDLAIPATEFFSTKAA KNPETKREAIVKAYGDDNEETLKQRLTNLEKKITNITTKFEQIE KCKKRNDVLFRLNHAETLRAAMISLAKKIDVQTGRHPYEG GGGSGGGGSGAAASIQTTVNTLSERISSKLEQEANASAQTKCD IEIGNFYIRQNHGCNITVKNMCSADADAQLDAVLSAATETYSG LTPEQKAYVPAMFTAALNIQTSVNTVVRDFENYVKQTCNSSA VVDNKLKIQNVIIIDECYGAPGSPTNLEFINTGSSKGNCAIKALM QLTTKATTQIAPRQVAGGGGGSGGGGSTCTVPTMNNAKLTST ETSFNDKQKVTFCTDSGYHSLDPNAVCECTDKWKYENPCKKM CTVSDYVSELYDKPLYEVNSTMTLSCNGETKYFRCEEKNGNT SWNDTVTCPNAECQPLQLEHGSCQPVKEKYSFGEYMTINCDV GYEVIGVSYISCTANSWNVIPSCQQKCDIPSLNGLISGSTFSIG GVIHLSCKSGFTLTGSPSSTCIDGKWNPILPTCVRSNEEFD PVD

		DGPDDDETDL SKLSKDVVQYEQEIESLEATYHIIIMALTIMGVIFL ISIIVLVCSCDKNNDQYKFHKLLP
104	Amino acid sequence encoded by 1401	MKTISVVTL LCVLP AVVYSTCTVPTMNNAKLTSTETSFN DKQK VTFTCDSGYHSLDPNAV CETDKWKYENPCKKMCTVSDYVSE LYDKPLYEVNSTMTLSCNGETKYFRCEEKNGNTSWNDTVTCP NAECQPLQLEHGSCQPVKEKYSFGEYMTINCDVGYEVIGVSYI SCTANSWNVIPSCQQKCDIPSLSNGLISGSTFSIGGVIHLSCKSG FTLTGSPSSTCIDGKWNPI LPTCVR SNEEFDPVDDGPDDDETDL S KLSKDVVQYEQEIESLEATYHGGGGSGGGGSGAAASIQT TVN TLSERISSKLEQEANAS AQT KC DIEIGNFYIRQNHGCNITVKNM CSADADAQLDAVL SAATETYSGLTPEQKAYVPAMFTAALNIQ TSVNTVVRDFENYVKQTCNSSAVVDNKLKIQNVIIDECY GAPG SPTNLEFINTGSSKGNCAIKALMQLTTKATTQIAPRQVAGGGG GSGGGGSA AVKTPVIVVPVIDRPPSETFPNVHEHINDQKFDDV KDNEVMQEKR D VIVNDDPDHYKDYVFIQWTGGNIRDDDKY THFFSGFCNTMCTEETKRNIARHLALWDSKFFTELENKNVEYV VIIENDNVIEDITFLRPVLKAIHDKKIDILQMREITGNKVKTEL V IDKDHAIFTYTGGYDVSL SAYIIRVTTALNIVDEI IKS GGLSSGF YFEIARIENEMKINRQIMDNSAKYVEHDPRLVAEHRFETMKPN FWSRIGTVAAKRYPGVMYTF TPLISFFGLFDINVIGLIVILFIMF MLIFNVKSKLLWFLTGTFTAFI
105	Amino acid sequence encoded by 1501	MKTISVVTL LCVLP AVVYSTCTVPTMNNAKLTSTETSFN DKQK VTFTCDSGYHSLDPNAV CETDKWKYENPCKKMCTVSDYVSE LYDKPLYEVNSTMTLSCNGETKYFRCEEKNGNTSWNDTVTCP NAECQPLQLEHGSCQPVKEKYSFGEYMTINCDVGYEVIGVSYI SCTANSWNVIPSCQQKCDIPSLSNGLISGSTFSIGGVIHLSCKSG FTLTGSPSSTCIDGKWNPI LPTCVR SNEEFDPVDDGPDDDETDL S KLSKDVVQYEQEIESLEATYHGGGGSGGGGSR LNQCMSANKA AITDSAVAVAAASSTHRKVVSSTTQYDHKESCNGLYYQGSCYI LHSDYKSFEDAKANCAAESSTLPNKSDVLT TWLIDYVEDTWG SDGNPITKTTS DYQSDVSQEV RKYFCTGGGGSGGGGSGAAA SIQT TVNTLSERISSKLEQEANAS AQT KC DIEIGNFYIRQNHGCN ITVKNMCSADADAQLDAVL SAATETYSGLTPEQKAYVPAMFT AALNIQTSVNTVVRDFENYVKQTCNSSAVVDNKLKIQNVIIDE CYGAPGSPTNLEFINTGSSKGNCAIKALMQLTTKATTQIAPRQ VAGTGVQFYMIVIGV IILAALFMY YAKRMLFTSTNDKIKLILA NKENVHWT TYMDTFFRTSPMIATTDIQN
106	Amino acid sequence encoded by 1601	MDWTWILFLVAAATRVHSDGTLFPGDDDLAIPATEFFSTKAA KNPETKREAI VKAYGDDNEETLKQRLTNLEKKITNITTKFEQIE KCCKRNDEVLFRL ENHAETLRAAMISLAKKIDVQTGRHPYEG GGGSGGGGSSPQTSKKIGDDATISCSRNTNYYVVM SAWYKE PNSIILLAAKSDVLYFDNYTKDKISYDSPYDDLVT TITIKSLTAG DAGTYICAFFMTSTTNDTDKVDYEEYSIELIVNTDSESTIDIILS

		GSTPETISEKPEDIDNSNCSSVFEITTPEPITDNVDDHTDVTYTS DSINTVNASSGESTTDEIPEPITDKEEDHTVTDTVSYTTVSTSSG IVTTKSTTDDADLYDTYNDNDTVPPTTVGGSTTSISNYKTKDF VGGGSGGGGSAETKEFKTLYNLFIDSYLQKLAQHSIPTNVTC AIHIGEVIQFKNCALRITNKCMSNSRLSFTLMVESFIEVISLLSE KDRRAIAEEIGIDLDDVPSVVSKEKNCNAYA EVNNIIDIQKLN IGECSAPPGQHMLLQIVNTGSAEANCGLQTIVKSLNKIYVPPHIE NRLPYYPWFLVGVAIILVIFTVAICSIRRNALALKYRYGTFLYV
107	Amino acid sequence encoded by 1701	MTQLPILLLLISLVYATPSPQTSKKIGDDATISCSRNNNTNYYVV MSAWYKEPNSIILLAAKSDVLYFDNYTKDKISYDSPYDDLVT ITIKSLTAGDAGTYICAFFMTSTTNDTDKVDYEEYSIELIVNTDS ESTIDIILSGSTPETISEKPEDIDNSNCSSVFEITTPEPITDNVDDH TDTVTYTSDSINTVNASSGESTTDEIPEPITDKEEDHTVTDTVSY TTVSTSSGIVTTKSTTDDADLYDTYNDNDTVPPTTVGGSTTSIS NYKTKDFVGGGSGGGGSAAVKTPVIVVPVIDRPPSETFPNVH EHINDQKFDDVKDNEVMQEKRDVVIVNDDPDHYKDYVFIQW TGGNIRDDDKYTHFFSGFCNTMCTEETKRNIARHLALWDSKFF TELENKNVEYVVIENDNVIEDITFLRPVLKAIHDKKIDILQMRE IITGNKVKTELVIDKDHAIFTYTGGYDVLSAYIIRVTTALNIVD EIIKSGGLSSGFYFEIARIENEMKINRQIMDNSAKYVEHDPRLV AEHRFETMKPNFWSRIGTVA AKRYPGVMYTFTTPLISGGGSG GGGSAETKEFKTLYNLFIDSYLQKLAQHSIPTNVTC AIHIGEVIQ QFKNCALRITNKCMSNSRLSFTLMVESFIEVISLLSEKDRRAIAE EIGIDLDDVPSVVSKEKNCNAYA EVNNIIDIQKLNIGECSAPPG QHMLLQIVNTGSAEANCGLQTIVKSLNKIYVPPHIEENRLPYYP WFLVGVAIILVIFTVAICSIRRNALALKYRYGTFLYV
108	Amino acid sequence encoded by 1801	MTQLPILLLLISLVYATPSPQTSKKIGDDATISCSRNNNTNYYVV MSAWYKEPNSIILLAAKSDVLYFDNYTKDKISYDSPYDDLVT ITIKSLTAGDAGTYICAFFMTSTTNDTDKVDYEEYSIELIVNTDS ESTIDIILSGSTPETISEKPEDIDNSNCSSVFEITTPEPITDNVDDH TDTVTYTSDSINTVNASSGESTTDEIPEPITDKEEDHTVTDTVSY TTVSTSSGIVTTKSTTDDADLYDTYNDNDTVPPTTVGGSTTSIS NYKTKDFVGGGSGGGGSRNLNQCMSANKAAITDS AVAVAAA SSTHRKVVSSTTQYDHKESCNGLYYQGSCYILHSDYKSFEDAK ANCAAESSTLPNKSDVLT TWLIDYVEDTWGSDGNPITKTTSDY QSDVSQEV RKYFCTGGGGSGGGGSTCTVPTMNNAKLTSTET SFNDKQKVTFCTDSGYHSLDPNAV CETDKWKYENPCKKMCT VSDYVSELYDKPLYEVNSTMTLSCNGETKYFRCEEKNGNTSW NDTVTCPNAECQPLQLEHGSCQPVKEKYSFGEYMTINCDVGY EVIGVSYISCTANSWNVIPSCQKQCDIPSLSNGLISGSTFSIGGVI HLSCKSGFTLTGSPSSTCIDGKWNPILPTCVRSNEEFDPVDDGP DDETDLSKLSKDVVQYEQEIESLEATYHIIIMALTIMGVIFLISII VLVCSCDKNNDQYKFHKLLP

109	Amino acid sequence encoded by 1901	MDWTWILFLVAAATRVHSDGTLFPGDDDLAIPATEFFSTKAA KNPETKREAIVKAYGDDNEETLKQRLTNLEKKITNITTKFEQIE KCCKRNDEVLFRLNHAETLRAAMISLAKKIDVQTGRHPYEG GGGSGGGGSGAAASIQTTVNTLSERISSKLEQEANASAQTKCD IEIGNFYIRQNHGCNITVKNMCSADADAQLDAVLSAATETYS LTPEQKAYVPAMFTAALNIQTSVNTVVRDFENYVKQTCNSSA VVDNKLKIQNVIIDECYGAPGSPTNLEFINTGSSKGNCAIKALM QLTTKATTQIAPRQVAGGGGGSGGGGSAETKEFKTLYNLFIDS YLQKLAQHSIPTNVTCAIHIGEVIQFKNCALRITNKCMSNSRL SFTLMVESFIEVISLLSEKDRRAIAEEIGIDLDDVPSVVSLEKN CNAYAEVNNIIDIQKLNIGECSAPPGQHMLLQIVNTGSAEANC GLQTIVKSLNKIYVPIIENRLPYYPWFLVGVAILVIFTVAICS IRRNLALKYRYGTFLYV
110	Amino acid sequence encoded by 2001	MDWTWILFLVAAATRVHSAAVKTPVIVVPVIDRPPSETFPNVH EHINDQKFDDVDNEVMQEKRDVVIVNDDPDHYKDYVFIQW TGGNIRDDDKYTHFFSGFCNTMCTEETKRNIARHLALWDSKFF TELENKNVEYVVIENDNVIEDITFLRPVLKAIHDKKIDILQMRE IITGNKVKTTELVIDKDHAIFTYTGGYDVSLSAIIRVTTALNIVD EIKSGGLSSGFYFEIARIENEMKINRQIMDNSAKYVEHDPRLV AEHRFETMKPNFWSRIGTVA AKRYPGV MYTFTTPLISGGGSG GGGSAETKEFKTLYNLFIDSYLQKLAQHSIPTNVTCAIHIGEVI QFKNCALRITNKCMSNSRLSFTLMVESFIEVISLLSEKDRRAIAE EIGIDLDDVPSVVSLEKNCNAYAEVNNIIDIQKLNIGECSAPPG QHMLLQIVNTGSAEANCGLQTIVKSLNKIYVPIIENRLPYYP GGGSGGGGSGAAASIQTTVNTLSERISSKLEQEANASAQTKC DIEIGNFYIRQNHGCNITVKNMCSADADAQLDAVLSAATETYS GLTPEQKAYVPAMFTAALNIQTSVNTVVRDFENYVKQTCNSS AVVDNKLKIQNVIIDECYGAPGSPTNLEFINTGSSKGNCAIKAL MQLTTKATTQIAPRQVAGTGVQFYMIVIGVILAALFMYAKR MLFTSTNDKIKLILANKENVHWTYMDTFFRTSPMIATTDIQN
111	A29L peptide-1	DGTLFPGDDDLAIPATEFFSTKAAKNPETKREAIVKAYGDDNE ETLKQRLTNLEKKITNITTKFEQIEKCCKRNDEVLFRLNHAET LRAAMISLAKKIDVQTGRHPY
112	A29L peptide-2	MDGTLFPGDDDLAIPATEFFSTKAAKNPETKREAIVKAYGDDN EETLKQRLTNLEKKITNITTKFEQIEKCCKRNDEVLFRLNHAET TLRAAMISLAKKIDVQTGRHPYE
113	A29L peptide-3	DGTLFPGDDDLAIPATEFFSTKAAKNPETKREAIVKAYGDDNE ETLKQRLTNLEKKITNITTKFEQIEKCCKRNDEVLFRLNHAET LRAAMISLAKKIDVQTGRHPYE
114	A35R peptide	MMTPENDEEQTSVFSATVYGDKIQGKNKRKRVLGICIRISMVI SLLSMITMSAFLIVRLNQCMSANKAAITDSAVAVAAASSTHRK VVSSTTQYDHKESCNGLYYQGSCYILHSDYKSFEDAKANCAA

		ESSTLPNKSDVLTTWLIDYVEDTWGSDGNPITKTTSDYQSDV SQEVRKYFCT
115	A35R extracellular segment	RLNQCMSANKAAITDSAVAVAAASSTHRKVVSSTTQYDHKES CNGLYYQGSCYILHSDYKSFEDAKANCAAESSTLPNKSDVLTT WLIDYVEDTWGSDGNPITKTTSDYQSDVSQEVRKYFCT
116	B2R signal peptide	MTQLPILLLLISLVYATP
117	Full-length B2R peptide (excluding the signal peptide)	SPQTSKKIGDDATISCSRNNNTNYYVVMASAWYKEPNSIILAAK SDVLYFDNYTKDKISYDSPYDDLVTITIKSLTAGDAGTYICAF FMTSTTNDTDKVDYEEYSIELIVNTDSESTIDIILSGSTPETISEK PEDIDNSNCSSVFEITTPEPITDNVDDHTDTVTYTSDSINTVNAS SGESTTDEIPEPITDKEEDHTVTDTVSYTTVSTSSGIVTTKSTTD DADLYDTYNDNDTVPPTTVGGSTTSISNYKTKDFVEIFGITTLII LSAVAIFCITYYICNKHPRKYKTENKV
118	Full-length B2R peptide	MTQLPILLLLISLVYATPSPQTSKKIGDDATISCSRNNNTNYYVV MSAWYKEPNSIILAAKSDVLYFDNYTKDKISYDSPYDDLVT ITIKSLTAGDAGTYICAFFMTSTTNDTDKVDYEEYSIELIVNTDS ESTIDIILSGSTPETISEKPEDIDNSNCSSVFEITTPEPITDNVDDH TDTVTYTSDSINTVNASSGESTTDEIPEPITDKEEDHTVTDTVSY TTVSTSSGIVTTKSTTDDADLYDTYNDNDTVPPTTVGGSTTSIS NYKTKDFVEIFGITTLIILSAVAIFCITYYICNKHPRKYKTENKV
119	B2R extracellular segment	SPQTSKKIGDDATISCSRNNNTNYYVVMASAWYKEPNSIILAAK SDVLYFDNYTKDKISYDSPYDDLVTITIKSLTAGDAGTYICAF FMTSTTNDTDKVDYEEYSIELIVNTDSESTIDIILSGSTPETISEK PEDIDNSNCSSVFEITTPEPITDNVDDHTDTVTYTSDSINTVNAS SGESTTDEIPEPITDKEEDHTVTDTVSYTTVSTSSGIVTTKSTTD DADLYDTYNDNDTVPPTTVGGSTTSISNYKTKDFV
120	B2R signal peptide + extracellular segment	MTQLPILLLLISLVYATPSPQTSKKIGDDATISCSRNNNTNYYVV MSAWYKEPNSIILAAKSDVLYFDNYTKDKISYDSPYDDLVT ITIKSLTAGDAGTYICAFFMTSTTNDTDKVDYEEYSIELIVNTDS ESTIDIILSGSTPETISEKPEDIDNSNCSSVFEITTPEPITDNVDDH TDTVTYTSDSINTVNASSGESTTDEIPEPITDKEEDHTVTDTVSY TTVSTSSGIVTTKSTTDDADLYDTYNDNDTVPPTTVGGSTTSIS NYKTKDFV
121	B6R signal peptide	MKTISVVTLCLVLPVVYS
122	Full-length B6R peptide (excluding the signal peptide)	TCTVPTMNNAKLTSTETSFNDKQKVTFCTDSGYHSLDPNAV ETDKWKYENPCKKMCTVSDYVSELYDKPLYEVNSTMTLSCN GETKYFRCEEKNGNTSWNDTVTCPNAECQPLQLEHGSCQPVK EKYSFGEYMTINCVDGYEVIGVSYISCTANSWNVIPSCQKCDI PSLSNGLISGSTFSIGGVHLSCKSGFTLTGSPSSTCIDGKWN PILPTCVRSNEEFDPVDDGPDDDETDLKLSKDVVQYEQEIESLEAT YHIIIMALTIMGVIFLISIIVLVCSCDKNNDQYKFHKLPL

123	Full-length B6R peptide	MKTISVVTLCLVLPVAVYSTCTVPTMNNAKLTSTETSFNDKQKVTFTCDSGYHSLDPNAVCE ETDKWKYENPCKKMCTVSDYVSELYDKPLYEVNSTMTLSCNGETKYFRCEEKNGNTSWNDTVTCP NAECQPLQLEHGSCQPVKEKYSFGEYMTINCDVGYEVIGVSYISCTANSWNVIPSCQKCDIPSLNGLISGSTFSIGGVIHLSCKSG FTLTGSPSSTCIDGKWNPIPTCVRSNEEFDPVDDGPDDDETDL SKLSDVQYEQEIESLEATYHIIIMALTIMGVIFLISIIVLVCSCDKNNDQYKFHLLP
124	B6R extracellular segment	TCTVPTMNNAKLTSTETSFNDKQKVTFTCDSGYHSLDPNAVCE ETDKWKYENPCKKMCTVSDYVSELYDKPLYEVNSTMTLSCN GETKYFRCEEKNGNTSWNDTVTCPNAECQPLQLEHGSCQPVK EKYSFGEYMTINCDVGYEVIGVSYISCTANSWNVIPSCQKCDI PSLSNGLISGSTFSIGGVIHLSCKSGFTLTGSPSSTCIDGKWNPI PTCVRSNEEFDPVDDGPDDDETDL SKLSDVQYEQEIESLEATYH
125	B6R signal peptide + extracellular peptide	MKTISVVTLCLVLPVAVYSTCTVPTMNNAKLTSTETSFNDKQKVTFTCDSGYHSLDPNAVCE ETDKWKYENPCKKMCTVSDYVSELYDKPLYEVNSTMTLSCNGETKYFRCEEKNGNTSWNDTVTCP NAECQPLQLEHGSCQPVKEKYSFGEYMTINCDVGYEVIGVSYISCTANSWNVIPSCQKCDIPSLNGLISGSTFSIGGVIHLSCKSG FTLTGSPSSTCIDGKWNPIPTCVRSNEEFDPVDDGPDDDETDL SKLSDVQYEQEIESLEATYH
126	Full-length C15L peptide	AETKEFKTLYNLFIDSYLQKLAQHSIPTNVTC CAIHIGE VIGQFKN CALRITNKC MNSRLSFTLMVESFIEVISLLSEKDRRAIAEEIGID LDDVPSVVS KLEKNCNAYAEVN NIIDIQKLNIGEC SAPPGQHM LLQIVNTG SAEANCGLQTIVK SLNKIYVPPIIEN RLPYYDPWFL VGVAIILVIFTVA ICSI RRNLALKYRYGTFLYV
127	C15L extracellular segment	AETKEFKTLYNLFIDSYLQKLAQHSIPTNVTC CAIHIGE VIGQFKN CALRITNKC MNSRLSFTLMVESFIEVISLLSEKDRRAIAEEIGID LDDVPSVVS KLEKNCNAYAEVN NIIDIQKLNIGEC SAPPGQHM LLQIVNTG SAEANCGLQTIVK SLNKIYVPPIIEN RLPYYDP
128	M1R extracellular segment	GAAASIQTTVNTLSE RISSKLEQEANASAQTK CDIEIGNFYIRQNHGCNITV KNMCSADADAQLDAVL SAATETYSGLTPEQKAYVP AMFTAALNIQTSVNTV VRDFENYVKQTCNSSAV VDNKLKIQNVIIDECY GAPGSP TNLEFINTGSSKGNCAIKAL MQLTTKATTQIAPRQVAG
129	Full-length M1R peptide-1	MGAAASIQTTVNTLSE RISSKLEQEANASAQTK CDIEIGNFYIRQNHGCNITV KNMCSADADAQLDAVL SAATETYSGLTPEQKAYVP AMFTAALNIQTSVNTV VRDFENYVKQTCNSSAV VDNKLKIQNVIIDECY GAPGSP TNLEFINTGSSKGNCAIKAL MQLTTKATTQIAPRQVAG TG VQFYMIVIGVILAALFMY YAKRMLFTSTNDKIKLILAN KENVHWT TYMDTFFRTSPMIATTDIQN

130	Full-length M1R peptide-2	GAAASIQTTVNTLSEISSKLEQEANASAQTKCDIEIGNFYIRQN HGCNITVKNMCSADADAQLDAVLSAATETYSGLTPEQKAYVP AMFTAALNIQTSVNTVVRDFENYVKQTCNSSAVVDNKLKIQN VIIDECYGAPGSPTNLEFINTGSSKGNCAlKALMQLTTKATTQI APRQVAGTGVQFYMIVIGVILAAALFMYAKRMLFTSTNDKIK LILANKENVHWTTYMDTFFRTSPMIIATTDIQN
131	Full-length H3L peptide	AAVKTPVIVVPVIDRPPSETFPNVHEHINDQKFDDVKDNEVMQ EKRDVVIVNDDPDHYKDYVFIQWTGGNIRDDDKYTHFFSGFC NTMCTEETKRNIARHLALWDSKFFTELENKNVEYVVIENDNV IEDITFLRPVLKAIHDKKIDILQMREIITGNKVKTELVIDKDHAIF TYTGGYDVSL SAYIIRVTTALNIVDEIHKSGGLSSGFYFEIARIEN EMKINRQIMDNSAKYVEHDPRLVAEHRFETMKPNFWSRIGTV AAKRYPGVMYTFTTPLISFFGLFDINVIGLIVILFIMFMLIFNVKS KLLWFLTGTFTVAFI
132	H3L extracellular segment	AAVKTPVIVVPVIDRPPSETFPNVHEHINDQKFDDVKDNEVMQ EKRDVVIVNDDPDHYKDYVFIQWTGGNIRDDDKYTHFFSGFC NTMCTEETKRNIARHLALWDSKFFTELENKNVEYVVIENDNV IEDITFLRPVLKAIHDKKIDILQMREIITGNKVKTELVIDKDHAIF TYTGGYDVSL SAYIIRVTTALNIVDEIHKSGGLSSGFYFEIARIEN EMKINRQIMDNSAKYVEHDPRLVAEHRFETMKPNFWSRIGTV AAKRYPGVMYTFTTPLIS
133	Full-length E8L peptide	MPQQLSPINIETKKAISDARLKTLDIHYNESKPTTIQNTGKLVRI NFKGGYISGGFLPNEYVLSTIHIYWGKEDDYGSNHLIDVYKYS GEINLVHWNKKKYSSYEEAKKHDDGIIIIAIFLQVSDHKNVYFQ KIVNQLDSIRSANMSAPFDSVFYLDNLLPSTLDYFTYLGT TINH SADAAWIIFTPINIHSQLSKFRTLLSSSNHEGKPHYITENYRN PYKLNDDTQVYYSGEIIRAATTSPVRENYFMKWLSDLREACFS YYQKYIEGNKTFIIAIVFVILTALFLMSQRYRSREKQN
134	IgE signal peptide	MDWTWILFLVAAATRVHS
135	HA polypeptide tag-1	YPYDVPDYAE
136	HA polypeptide tag-2	YPYDVPDYA
137	GS linker-1	GGGGS
138	GS linker-2	GGGSGGGGS
139	5'UTR RNA	AUCCAAGCUUAGCUUGCUUGUUCUUUUUGCAGAAGCUCA GAAUAAACGCUCAACUUUGGCGCCGCCACC
140	3'UTR RNA-1	UAAGAAUUCUAGGCUGGAGCCUCGGUGGCCAUGCUUCUU GCCCCUUGGGCCUCCCCCAGCCCCUCCUCCCCUCCUGCA

		CCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGAGUGGGCG GCACUAGU
141	3'UTR RNA- 2	GAAUUCUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCC CUUGGGCCUCCCCCAGCCCCUCCUCCCCUCCUGCACCCG UACCCCCGUGGUCUUUGAAUAAAGUCUGAGUGGGCGGCA CUAGU
142	5'UTR DNA	ATCCAAGCTTAGCTTGCTTGTTCTTTTTGCAGAAGCTCAGAA TAAACGCTCAACTTTGGCGCCGCCACC
143	3'UTR DNA-1	TAAGAATTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCC CCTTGGGCCTCCCCCAGCCCCCTCCTCCCCTTCCTGCACCCG TACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCACTA GT
144	3'UTR DNA-2	GAATTCTAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCT TGGGCCTCCCCCAGCCCCCTCCTCCCCTTCCTGCACCCGTAC CCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGCACTAGT
145	PolyA tail RNA	AA AA AA AAAA
146	PolyA tail DNA	AA AA AA AAAA
147	Cap- proximal sequence	GGGGGGG
148	5'UTR RNA- 2	CUUGUUCUUUUUGCAGAAGCUCAGAAUAAACGCUCAACU UUGGC
149	3'UTR RNA- 3	UAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGG CCUCCCCCAGCCCCUCCUCCCCUCCUGCACCCGUACCCC CGUGGUCUUUGAAUAAAGUCUGAGUGGGCGGC
150	5'UTR DNA-2	CTTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTG GC
151	3'UTR DNA-3	TAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCC TCCCCCAGCCCCCTCCTCCCCTTCCTGCACCCGTACCCCCGT GGTCTTTGAATAAAGTCTGAGTGGGCGGC

CLAIMS

1. A polypeptide or polypeptide combination, comprising one or more peptide fragments, wherein the peptide fragments are each independently selected from peptide fragments of mpox virus A29L, A35R, B2R, B6R, C15L, M1R, and H3L proteins;
- 5 optionally, the polypeptide or polypeptide combination further comprises one or more linkers, an IgE signal peptide and/or an HA polypeptide tag.
2. The polypeptide or polypeptide combination according to claim 1, wherein
- (1) the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, B6R, C15L, M1R, and H3L proteins;
- 10 (2) the polypeptide or polypeptide combination comprises peptide fragments of A29L, A35R, B2R, B6R, C15L, and M1R proteins;
- (3) the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, and C15L proteins;
- (4) the polypeptide or polypeptide combination comprises peptide fragments of B6R, M1R, and H3L proteins;
- 15 (5) the polypeptide or polypeptide combination comprises peptide fragments of B2R, A35R, and B6R proteins;
- (6) the polypeptide or polypeptide combination comprises peptide fragments of A29L, M1R, and C15L proteins;
- 20 (7) the polypeptide or polypeptide combination comprises peptide fragments of A29L, B6R, and M1R proteins;
- (8) the polypeptide or polypeptide combination comprises peptide fragments of B6R, M1R, and A35R proteins;
- (9) the polypeptide or polypeptide combination comprises peptide fragments of A29L, B2R, and C15L proteins;
- 25 (10) the polypeptide or polypeptide combination comprises peptide fragments of B2R, H3L, and C15L proteins;
- (11) the polypeptide or polypeptide combination comprises peptide fragments of H3L, M1R, and C15L proteins;

- (12) the polypeptide or polypeptide combination comprises a peptide fragment of A29L protein;
(13) the polypeptide or polypeptide combination comprises a peptide fragment of A35R protein;
(14) the polypeptide or polypeptide combination comprises a peptide fragment of B2R protein;
(15) the polypeptide or polypeptide combination comprises a peptide fragment of B6R protein;
5 (16) the polypeptide or polypeptide combination comprises a peptide fragment of C15L protein;
(17) the polypeptide or polypeptide combination comprises a peptide fragment of M1R protein;

or

(18) the polypeptide or polypeptide combination comprises a peptide fragment of H3L protein.

3. The polypeptide according to claim 1 or 2, wherein

10 the peptide fragment of A29L protein comprises an amino acid sequence as set forth in SEQ
ID NO: 111 or 113;

the peptide fragment of A35R protein comprises an amino acid sequence as set forth in SEQ
ID NO: 114 or 115;

15 the peptide fragment of B2R protein comprises an amino acid sequence as set forth in SEQ ID
NO: 116, 117, 118, 119 or 120;

the peptide fragment of B6R protein comprises an amino acid sequence as set forth in SEQ ID
NO: 121, 122, 123, 124 or 125;

the peptide fragment of C15L protein comprises an amino acid sequence as set forth in SEQ
ID NO: 126 or 127;

20 the peptide fragment of M1R protein comprises an amino acid sequence as set forth in SEQ
ID NO: 128 or 130; or

the peptide fragment of H3L protein comprises an amino acid sequence as set forth in SEQ ID
NO: 131 or 132.

4. The polypeptide or polypeptide combination according to any one of claims 1-3, wherein

25 (1) the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R,
B6R, C15L, M1R, and H3L proteins, and a linker;

(2) the polypeptide or polypeptide combination comprises peptide fragments of A29L, A35R,
B2R, B6R, C15L, and M1R proteins, an IgE signal peptide, and a linker;

- (3) the polypeptide or polypeptide combination comprises peptide fragments of A29L, A35R, B2R, B6R, C15L, and M1R proteins, an IgE signal peptide, an HA polypeptide tag, and a linker;
- (4) the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, B6R, C15L, M1R, and H3L proteins, an IgE signal peptide, an HA polypeptide tag, and a linker;
- 5 (5) the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, B6R, C15L, M1R, and H3L proteins, an IgE signal peptide, and a linker;
- (6) the polypeptide or polypeptide combination comprises peptide fragments of A35R, B2R, and C15L proteins, and a linker;
- (7) the polypeptide or polypeptide combination comprises peptide fragments of B6R, M1R, 10 and H3L proteins, and a linker;
- (8) the polypeptide or polypeptide combination comprises peptide fragments of B2R, A35R, and B6R proteins, and a linker;
- (9) the polypeptide or polypeptide combination comprises peptide fragments of A29L, M1R, and C15L proteins, an IgE signal peptide, and a linker;
- 15 (10) the polypeptide or polypeptide combination comprises peptide fragments of A29L, B6R, and M1R proteins, an IgE signal peptide, and a linker;
- (11) the polypeptide or polypeptide combination comprises peptide fragments of B6R, M1R, and A35R proteins, and a linker;
- (12) the polypeptide or polypeptide combination comprises peptide fragments of A29L, B2R, 20 and C15L proteins, an IgE signal peptide, and a linker;
- (13) the polypeptide or polypeptide combination comprises peptide fragments of B2R, H3L, and C15L proteins, and a linker;
- (14) the polypeptide or polypeptide combination comprises peptide fragments of H3L, M1R, and C15L proteins, an IgE signal peptide, and a linker;
- 25 (15) the polypeptide or polypeptide combination comprises an IgE signal peptide, a peptide fragment of A29L protein and an HA polypeptide tag;
- (16) the polypeptide or polypeptide combination comprises a peptide fragment of A35R protein and an HA polypeptide tag;

- (17) the polypeptide or polypeptide combination comprises a peptide fragment of B2R protein and an HA polypeptide tag;
- (18) the polypeptide or polypeptide combination comprises a peptide fragment of B6R protein, a linker, and an HA polypeptide tag;
- 5 (19) the polypeptide or polypeptide combination comprises an IgE signal peptide, a linker, an HA polypeptide tag and a peptide fragment of C15L protein;
- (20) the polypeptide or polypeptide combination comprises an IgE signal peptide, a linker, an HA polypeptide tag and a peptide fragment of M1R protein; or
- (21) the polypeptide or polypeptide combination comprises an IgE signal peptide, a linker, an
10 HA polypeptide tag and a peptide fragment of H3L protein.
5. The polypeptide or polypeptide combination according to any one of claims 1-4, wherein the IgE signal peptide comprises an amino acid sequence as set forth in SEQ ID NO: 134; the HA polypeptide tag comprises an amino acid sequence as set forth in SEQ ID NO: 135 or 136; and/or
- 15 the linker is a GS linker, preferably, the GS linker comprises an amino acid sequence as set forth in SEQ ID NO: 137 or 138.
6. The polypeptide or polypeptide combination according to any one of claims 1-5, wherein the polypeptide or polypeptide combination comprises an amino acid sequence as set forth in one of SEQ ID NOs: 102, 104, 108, 109, 89, 90, 91, 92, 93, 94, 95, 101, 103, 105, 106, 107 and
20 110, or an amino acid sequence having at least 80% identity to the amino acid sequence as set forth in one of SEQ ID NOs: 102, 104, 108, 109, 89, 90, 91, 92, 93, 94, 95, 101, 103, 105, 106, 107 and 110.
7. The polypeptide or polypeptide combination according to any one of claims 1-5, wherein the polypeptide or polypeptide combination comprises two or more amino acid sequences
25 selected from SEQ ID NOs: 102, 104, 108, 109, 89, 90, 91, 92, 93, 94, 95, 101, 103, 105, 106, 107 and 110, or two or more amino acid sequences having at least 80% identity to amino acid sequences selected from SEQ ID NOs: 102, 104, 108, 109, 89, 90, 91, 92, 93, 94, 95, 101, 103, 105, 106, 107 and 110.
8. The polypeptide or polypeptide combination according to claim 7, wherein

- (1) the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide and the second polypeptide respectively comprise amino acid sequences as set forth in SEQ ID NOs: 102 and 104;
- 5 (2) the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide and the second polypeptide respectively comprise amino acid sequences as set forth in SEQ ID NOs: 108 and 109;
- 10 (3) the polypeptide or polypeptide combination comprises a first polypeptide, a second polypeptide, a third polypeptide, a fourth polypeptide, a fifth polypeptide and a sixth polypeptide, wherein the first polypeptide, the second polypeptide, the third polypeptide, the fourth polypeptide, the fifth polypeptide and the sixth polypeptide respectively comprise amino acid sequences as set forth in SEQ ID NOs: 89, 90, 91, 92, 93 and 94;
- 15 (4) the polypeptide or polypeptide combination comprises a first polypeptide, a second polypeptide, a third polypeptide, a fourth polypeptide, a fifth polypeptide and a sixth polypeptide, wherein the first polypeptide, the second polypeptide, the third polypeptide, the fourth polypeptide, the fifth polypeptide and the sixth polypeptide respectively comprise amino acid sequences as set forth in SEQ ID NOs: 90, 91, 92, 93, 94 and 95;
- 20 (5) the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide and the second polypeptide respectively comprise amino acid sequences as set forth in SEQ ID NOs: 102 and 103;
- (6) the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide and the second polypeptide respectively comprise amino acid sequences as set forth in SEQ ID NOs: 105 and 106;
- 25 (7) the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide and the second polypeptide respectively comprise amino acid sequences as set forth in SEQ ID NOs: 105 and 107; or
- (8) the polypeptide or polypeptide combination comprises a first polypeptide and a second polypeptide, wherein the first polypeptide and the second polypeptide respectively comprise amino acid sequences as set forth in SEQ ID NOs: 110 and 108.

9. A nucleic acid or nucleic acid combination, comprising a polynucleotide encoding the polypeptide or polypeptide combination according to any one of claims 1-8.

10. The nucleic acid or nucleic acid combination according to claim 9,

wherein the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably

5 mRNA,

the polynucleotide comprises a nucleotide sequence as set forth in one of SEQ ID NOs: 14, 16, 20, 21, 1, 2, 3, 4, 5, 6, 7, 13, 15, 17, 18, 19 and 22, or a nucleotide sequence having at least 80% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 14, 16, 20, 21, 1, 2, 3, 4, 5, 6, 7, 13, 15, 17, 18, 19 and 22; or

10 wherein the polynucleotide is DNA,

the polynucleotide comprises a nucleotide sequence as set forth in one of SEQ ID NOs: 36, 38, 42, 43, 23, 24, 25, 26, 27, 28, 29, 35, 37, 39, 40, 41 and 44, or a nucleotide sequence having at least 80% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 36, 38, 42, 43, 23, 24, 25, 26, 27, 28, 29, 35, 37, 39, 40, 41 and 44.

15 11. The nucleic acid or nucleic acid combination according to claim 9,

wherein the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA,

the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 14, 16, 20, 21, 1, 2, 3, 4, 5, 6, 7, 13, 15, 17, 18, 19 and 22, or two or more nucleotide sequences
20 having at least 80% identity to nucleotide sequences selected from SEQ ID NOs: 14, 16, 20, 21, 1, 2, 3, 4, 5, 6, 7, 13, 15, 17, 18, 19 and 22; or

wherein the polynucleotide is DNA,

the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs: 36, 38, 42, 43, 23, 24, 25, 26, 27, 28, 29, 35, 37, 39, 40, 41 and 44, or two or more nucleotide
25 sequences having at least 80% identity to nucleotide sequences selected from SEQ ID NOs: 36, 38, 42, 43, 23, 24, 25, 26, 27, 28, 29, 35, 37, 39, 40, 41 and 44.

12. The nucleic acid or nucleic acid combination according to claim 11, wherein

wherein the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA,

(1) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 14 and 16;

5 (2) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 20 and 21;

(3) the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide, wherein the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth
10 polynucleotide, the fifth polynucleotide and the sixth polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 1, 2, 3, 4, 5 and 6;

(4) the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide, wherein the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth
15 polynucleotide, the fifth polynucleotide and the sixth polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 2, 3, 4, 5, 6 and 7;

(5) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 14 and 15;

20 (6) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 17 and 18;

(7) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences
25 as set forth in SEQ ID NOs: 17 and 19; or

(8) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 22 and 20;

or

wherein the polynucleotide is DNA,

(1') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 36 and 38;

5 (2') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 42 and 43;

(3') the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide, wherein
10 the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth polynucleotide, the fifth polynucleotide and the sixth polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 23, 24, 25, 26, 27 and 28;

(4') the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide, wherein
15 the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth polynucleotide, the fifth polynucleotide and the sixth polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 24, 25, 26, 27, 28 and 29;

(5') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences
20 as set forth in SEQ ID NOs: 36 and 37;

(6') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 39 and 40;

(7') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein
25 the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 39 and 41; or

(8') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 44 and 42.

13. The nucleic acid or nucleic acid combination according to any one of claims 9-12,
wherein the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably mRNA,
the polynucleotide further comprises a 5'-UTR sequence; preferably, the 5'-UTR sequence
5 comprises a nucleotide sequence as set forth in SEQ ID NO: 139 or 148;
the polynucleotide further comprises a 3'-UTR sequence; preferably, the 3'-UTR sequence
comprises a nucleotide sequence as set forth in SEQ ID NO: 140, 141, or 149; and/or
the polynucleotide further comprises a poly(A) sequence; preferably, the poly(A) sequence
comprises a nucleotide sequence as set forth in SEQ ID NO: 145; or
10 wherein the polynucleotide is DNA,
the polynucleotide further comprises a DNA sequence corresponding to the 5'-UTR sequence;
preferably, the DNA sequence corresponding to the 5'-UTR sequence comprises a nucleotide
sequence as set forth in SEQ ID NO: 142 or 150;
the polynucleotide further comprises a DNA sequence corresponding to the 3'-UTR sequence;
15 preferably, the DNA sequence corresponding to the 3'-UTR sequence comprises a nucleotide
sequence as set forth in SEQ ID NO: 143, 144, or 151; and/or
the polynucleotide further comprises a DNA sequence corresponding to the poly(A) sequence;
preferably, the DNA sequence corresponding to the poly(A) sequence comprises a nucleotide
sequence as set forth in SEQ ID NO: 146.
- 20 14. The nucleic acid or nucleic acid combination according to any one of claims 9-13,
wherein the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably
mRNA,
the polynucleotide comprises a nucleotide sequence as set forth in one of SEQ ID NOs: 58, 60,
64, 65, 45, 46, 47, 48, 49, 50, 51, 57, 59, 61, 62, 63 and 66, or a nucleotide sequence having at
25 least 80% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 58, 60, 64, 65,
45, 46, 47, 48, 49, 50, 51, 57, 59, 61, 62, 63 and 66; or
wherein the polynucleotide is DNA,
the polynucleotide comprises a nucleotide sequence as set forth in one of SEQ ID NOs: 80, 82,
86, 87, 67, 68, 69, 70, 71, 72, 73, 79, 81, 83, 84, 85 and 88, or a nucleotide sequence having at

least 80% identity to the nucleotide sequence as set forth in one of SEQ ID NOs: 80, 82, 86, 87, 67, 68, 69, 70, 71, 72, 73, 79, 81, 83, 84, 85 and 88.

15. The nucleic acid or nucleic acid combination according to any one of claims 9-13,
wherein the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably
5 mRNA,

the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs:
58, 60, 64, 65, 45, 46, 47, 48, 49, 50, 51, 57, 59, 61, 62, 63 and 66, or two or more nucleotide
sequences having at least 80% identity to nucleotide sequences selected from SEQ ID NOs: 58,
60, 64, 65, 45, 46, 47, 48, 49, 50, 51, 57, 59, 61, 62, 63 and 66; or

10 wherein the polynucleotide is DNA,

the polynucleotide comprises two or more nucleotide sequences selected from SEQ ID NOs:
80, 82, 86, 87, 67, 68, 69, 70, 71, 72, 73, 79, 81, 83, 84, 85 and 88, or two or more nucleotide
sequences having at least 80% identity to nucleotide sequences selected from SEQ ID NOs: 80,
82, 86, 87, 67, 68, 69, 70, 71, 72, 73, 79, 81, 83, 84, 85 and 88.

15 16. The nucleic acid or nucleic acid combination according to claim 15, wherein
wherein the polynucleotide is RNA, preferably mRNA, circRNA or saRNA, more preferably
mRNA,

(1) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein
the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences
20 as set forth in SEQ ID NOs: 58 and 60;

(2) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein
the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences
as set forth in SEQ ID NOs: 64 and 65;

(3) the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third
25 polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide, wherein
the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth
polynucleotide, the fifth polynucleotide and the sixth polynucleotide respectively comprise
nucleotide sequences as set forth in SEQ ID NOs: 45, 46, 47, 48, 49 and 50;

(4) the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide, wherein the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth polynucleotide, the fifth polynucleotide and the sixth polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 46, 47, 48, 49, 50 and 51;

(5) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 58 and 59;

(6) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 61 and 62;

(7) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 61 and 63; or

(8) the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 66 and 64;

or

wherein the polynucleotide is DNA,

(1') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 80 and 82;

(2') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 86 and 87;

(3') the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide, wherein the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth

polynucleotide, the fifth polynucleotide and the sixth polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 67, 68, 69, 70, 71 and 72;

(4') the polynucleotide comprises a first polynucleotide, a second polynucleotide, a third polynucleotide, a fourth polynucleotide, a fifth polynucleotide and a sixth polynucleotide, wherein
5 the first polynucleotide, the second polynucleotide, the third polynucleotide, the fourth polynucleotide, the fifth polynucleotide and the sixth polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 68, 69, 70, 71, 72 and 73;

(5') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences
10 as set forth in SEQ ID NOs: 80 and 81;

(6') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 83 and 84;

(7') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein
15 the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 83 and 85; or

(8') the polynucleotide comprises a first polynucleotide and a second polynucleotide, wherein the first polynucleotide and the second polynucleotide respectively comprise nucleotide sequences as set forth in SEQ ID NOs: 88 and 86.

20 17. A composition, comprising the nucleic acid or nucleic acid combination according to any one of claims 9-16.

18. The composition according to claim 17, comprising a lipid encapsulating the nucleic acid or nucleic acid combination.

19. The composition according to claim 17 or 18, comprising at least a first lipid,
25 wherein the first lipid comprises one or more compounds of formula 01-I or 01-II, preferably the first lipid comprises one or more compounds listed in Table 01-1;

wherein the first lipid comprises one or more compounds of formula 02-I or 02-II, preferably the first lipid comprises one or more compounds listed in Table 02-1;

wherein the first lipid comprises one or more compounds of formula 03-I, preferably the first lipid comprises one or more compounds listed in Table 03-1; or

wherein the first lipid comprises one or more compounds of formula 04-I or 04-III, preferably the first lipid comprises one or more compounds listed in Table 04-1;

5 more preferably, the first lipid comprises compound 01-1.

20. The composition according to claim 19, further comprising a second lipid,

wherein the second lipid is a polymer-conjugated lipid; preferably, the second lipid comprises one or more compounds of formula 05-I.

21. The composition according to any one of claims 17-20, comprising
10 the first lipid, the second lipid, a phospholipid and a steroid.

22. The composition according to any one of claims 17-21, wherein the composition is formulated as a lipid nanoparticle that encapsulates the nucleic acid or nucleic acid combination in a lipid.

23. A pharmaceutical composition or kit, comprising the polypeptide or polypeptide
15 combination according to any one of claims 1-8, the nucleic acid or nucleic acid combination according to any one of claims 9-16, or the composition according to any one of claims 17-22.

24. An expression vector, comprising the nucleic acid or nucleic acid combination according to any one of claims 9-16.

25. A host cell, comprising the nucleic acid or nucleic acid combination according to any one
20 of claims 9-16 or the expression vector according to claim 24.

26. Use of the polypeptide or polypeptide combination according to any one of claims 1-8, the nucleic acid or nucleic acid combination according to any one of claims 9-16, the composition according to any one of claims 17-22, or the pharmaceutical composition or kit according to claim 23 in the preparation of a vaccine for preventing and/or treating mpox virus infection in a subject.

27. The use according to claim 26, wherein the subject is a mammal; preferably, the subject is
25 a human.