

Description

Title of Invention: NON-NATURAL 5' UNTRANSLATED REGION AND USE THEREOF

Technical Field

The present disclosure relates to a non-natural 5' untranslated region and use thereof.

Background Art

Messenger RNA (mRNA) is frequently used as an alternative to plasmid DNA as a gene delivery molecule in the field of cancer immunotherapy and stem cell-based biomedical research. As a direct source of gene products, mRNA has several advantages including the lack of a requirement for nuclear entry, which poses a significant barrier to DNA delivery, especially in non-dividing cells. Also, by avoiding the expression of oncogenes caused by abnormal transcription and insertional mutagenesis, mRNA is less likely to integrate into the host genome.

For a given gene, untranslated regions (UTRs) of the gene including a 5'UTR and a 3'UTR are regions involved in regulating expression. The 5'UTR is a DNA regulatory region at the 5' end of all sequences encoding a protein, and is transcribed into mRNA but is not translated into a protein. The 5'UTR may include various regulatory elements, such as a 5'-cap structure, a G-quadruplex structure (G4), a stem-loop structure, and an internal ribosome entry site (IRES), and such regulatory elements play an important role in controlling translation initiation. The 3'UTR positioned downstream of the sequence encoding a protein is known to be involved in many regulatory processes including transcript cleavage, stability and polyadenylation, translation, and mRNA localization. The 3'UTR serves as a binding site for many regulatory proteins and small non-coding RNAs (e.g., microRNAs).

In general, current mRNA therapies are often not tissue-specific or rely on native or standard UTR sequences that provide a low level of protein expression. In this regard, there is a need for novel 5'UTRs and a combination thereof that can stabilize mRNA therapies and increase protein synthesis.

Disclosure of Invention

Technical Problem

An aspect provides a messenger RNA (mRNA) including a non-natural 5' untranslated region (5'UTR).

Another aspect provides DNA encoding the mRNA.

Another aspect provides a method of obtaining the mRNA, the method including transcribing the mRNA by using the DNA as a template.

Another aspect provides a method of obtaining a polypeptide, the method including translating the mRNA.

Another aspect provides a composition including the mRNA.

Another aspect provides a method of utilizing the mRNA, the method including introducing the mRNA to a host cell.

Solution to Problem

As used in the present specification, "5' untranslated region (5'-UTR)" refers to an mRNA region where a polypeptide is not encoded, directly upstream (i.e., 5') of the first codon, i.e., an initiation codon, of an mRNA transcript to be translated by ribosomes.

As used in the present specification, "3' untranslated region (3'-UTR)" refers to an mRNA region where a polypeptide is not encoded, directly downstream (i.e., 3') of a stop codon, i.e., a termination codon, of an mRNA transcript signaling termination of translation.

As used in the present specification, "open reading frame (ORF)" or "sequence encoding a polypeptide" refers to a continuous region of DNA that starts with an initiation codon, such as a methionine codon (ATG) and ends with a stop codon, such as TAA, TAG, or TGA, and encodes a polypeptide.

As used in the present specification, "polyadenylate sequence," "poly (A)," or "poly (A) tail" refers to an mRNA region including multiple continuous adenine nucleotides, downstream of the 3'-UTR, for example, directly downstream (i.e., 3') of the 3'-UTR. The poly (A) tail may include 1,000, 500, 400, or 300 or less adenine nucleotides. For example, the poly (A) tail may include 10 to 300 adenine nucleotides. For example, the poly (A) tail may include 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, or 300 nucleotides. The poly (A) tail may include, for example, 50 to 250 adenine nucleotides. The poly (A) tail in a living body, such as a cell or a subject, may serve to protect the mRNA from an attack of an enzyme, such as a cytoplasmic enzyme, and aid in termination of transcription, export of

the mRNA from the nucleus, and translation. A 3' poly (A) tail refers to a region (stretch) of the adenine nucleotides, and is generally added to mRNA that is transcribed during mRNA processing or to immature mRNA.

As used in the present specification, the term "operatively linked" refers to connection of nucleotide sequences on a single nucleic acid fragment so that a function is affected by another. For example, a promoter is operatively linked to a coding sequence (e.g., an ORF) when the promoter is capable of affecting the expression of the coding sequence (that is, when transcription of the coding sequence is regulated by the promoter). The coding sequence may be operatively linked to a regulatory sequence in a sense direction or an anti-sense direction. Unless otherwise stated in the present specification, each nucleic acid sequence included in the claimed mRNA or DNA is operatively linked to one another.

As used herein, unless a location of a nucleotide sequence is otherwise specified, a nucleotide sequence is connected or located from the 5'-end to the 3'-end.

As used in the present specification, the term "vector" or "nucleic acid construct" refers to an arbitrary nucleic acid capable of carrying a gene, an ORF, or a DNA fragment into a cell. The vector may be, for example, one which may be replicated in a cell. The vector may be a virus, a bacteriophage, a provirus, a plasmid, a phagemid, a transposon, or an artificial chromosome such as an yeast artificial chromosome (YAC), a bacterial artificial chromosome (BAC), a plant artificial chromosome (PLAC). etc.

The first aspect provides an mRNA including a non-natural 5' untranslated region (5'UTR),

wherein the 5'UTR does not form a secondary structure. As used in the present specification, the term "secondary structure" refers to binding between paired bases.

The mRNA including the non-natural 5'UTR may have a sequence of Formula 1:

Formula1: AGN_aGCCACC

wherein N is A, U, G, or C, each N is identical to or different from each other, and a indicates the number of repeats of N and is 42 or 62.

The mRNA including the 5'UTR may include a nucleotide sequence of any one of SEQ ID NOs: 1 to 9. The mRNA may consist of a nucleotide sequence of any one of SEQ ID NOs: 1 to 9. The mRNA may be an isolated nucleic acid, e.g., a transcript, including a

nucleotide sequence of any one of SEQ ID NOs: 1 to 9. The transcript may include a sequence encoding a target protein.

The mRNA may further include a sequence encoding a polypeptide. The sequence encoding a polypeptide may be operatively linked to the 5'UTR upstream of the 5'UTR.

The polypeptide may be an antigenic or therapeutic polypeptide. The antigenic polypeptide may be a viral antigenic polypeptide.

The therapeutic polypeptide may be any polypeptide having disease-treating activity, for example, anticancer activity. The polypeptide having anticancer activity may be a tumor suppressor, an immune stimulating protein, a radiation enhancing protein, or a chemotherapy enhancing protein. For example, the tumor suppressor may be involved in regulating the cell cycle of cancer cells themselves, and may accordingly have an anticancer activity. Immune cells involved in the immune stimulation may include B cells, NK cells, T cells, or mast cells. The polypeptide having anticancer activity may not have antigenicity in a host cell or an organism including the same. The polypeptide having anticancer activity may be an antibody, a hormone, a cytokine, an enzyme, a derivative thereof, or a combination thereof. Also, the polypeptide having anticancer activity may suppress proliferation of cancer cells. The polypeptide may be a recombinant one including a fusion protein.

The polypeptide may be a p53 protein.

The sequence encoding a polypeptide may be codon-optimized. The codon optimization may vary depending on a host cell to which the mRNA molecule is introduced. Also, the sequence encoding a polypeptide may include a degenerate sequence.

The mRNA may further include an expression regulatory sequence. The expression regulatory sequence may be involved in translational regulation or mRNA stability. The expression regulatory sequence may be a Kozak sequence, a sequence encoding the 3'UTR, or a poly (A) sequence.

The sequence encoding the 5'UTR, the sequence encoding a polypeptide, and the expression regulatory sequence may be transcribed in vitro or in vivo, and may be form a common transcript.

The sequence encoding the 3'UTR may be operatively linked to the 3'UTR upstream of the sequence encoding a polypeptide.

The sequence encoding the 3'UTR may be derived from an arbitrary gene. When the 3'UTR is combined with the 5'UTR, expression efficiency of RNA may be synergistically improved.

The poly (A) sequence may be operatively linked to the 3'UTR downstream of the 3'UTR.

The 5'UTR may have activity of increasing the translation yield, stability of the RNA, or a combination thereof, compared to a case where the native 5'UTR is used.

In an embodiment, the mRNA may include "5'-cap" such as Cap0, Cap1, or Cap2. The term "5'-cap" refers to a cap structure found at the 5'-end of an mRNA molecule. The 5'-cap is generally a 7-methylguanine ribonucleotide connected to the 5' position of the first nucleotide via 5' triphosphate in a 5'-to-3' chain. In the cap0, riboses at the first and second cap-proximal nucleotides of the mRNA may both include 2'-hydroxyl. In the cap1, riboses at the first and second cap-proximal nucleotides of the mRNA may include 2'-methoxy and 2'-hydroxyl, respectively. In the cap2, riboses at the first and second cap-proximal nucleotides of the mRNA may both include 2'-methoxy. The 7-methylguanine ribonucleotide may be further modified. For example, in the modified 7-methylguanine ribonucleotide, 7-methylguanine 3'-methoxy (7mG(3'OMe)) or 2'-hydroxyl and 3'-hydroxyl may each independently be substituted with a sulfonyl-containing group selected from the group consisting of a mesyl group, an esyl group, a triflyl group, a tresyl group, a tosyl group, a brosyl group, a nosyl group, and a dansyl group. The modified 7-methylguanine ribonucleotide substituted with the sulfonyl-containing group may be, for example, 7-methylguanine 3'-mesyl (7mG(3'OMs)). Most endogenous mRNAs in higher eukaryotes including mammalian mRNA such as human mRNA may have the cap1 or the cap2. The cap0 and other cap structures different from the cap1 and the cap2 may be immunogenic in mammals such as humans.

Such caps may be introduced co-transcriptionally. For example, the 5'-cap or a 5'-cap analog may be introduced to a transcript at initiation at which the 5'-cap or the 5'-cap analog is generated, while a DNA template is transcribed *in vitro* in the presence of the 5'-cap or the 5'-cap analog. Also, such caps may be introduced post-transcriptionally.

The 5'-cap or the 5'-cap analog may be introduced to RNA post-transcriptionally by a capping enzyme such as a capping enzyme from vaccinia virus. The 5'-end sequence with the 5'-end cap structure formed may include, for example, 7mG(5')ppp(5')ApG, m7G(3'OMe)(5')ppp(5')ApG, m7G(3'OMe)(5')ppp(5')(2'OMeA)pG, m7G(3'OMs)(5')ppp(5')(2'OMeA)pG, or m7G(5')ppp(5')(2'OMeA)pG.

In an embodiment, the mRNA may include a modified nucleotide. In an embodiment, the mRNA may have at least one U substituted with N1-methyl-pseudouridine. In the mRNA, the 5'-end sequence with the 5'-end cap structure formed may have a cap structure of m7(3'OMeG)(5')ppp(5')(2'OMeA)pG or a cap structure of m7(3'OMsG)(5')ppp(5')(2'OMeA)pG. In the mRNA, all Us may be substituted with N1-methyl-pseudouridine. In the mRNA, all Us may be substituted with N1-methyl-pseudouridine, and the 5'-end sequence with the 5'-end cap structure formed may have a cap structure of m7(3'OMeG)(5')ppp(5')(2'OMeA)pG or a cap structure of m7(3'OMsG)(5')ppp(5')(2'OMeA)pG.

The mRNA may include a nucleotide sequence of any one of SEQ ID NOs: 1 to 9, wherein at least one U of the sequence may be substituted with N1-methyl-pseudouridine. In the mRNA, all Us of the sequence may be substituted with N1-methyl-pseudouridine.

The second aspect provides a method of obtaining a polypeptide, the method including translating the mRNA including the sequence encoding a polypeptide.

The third aspect provides a composition including the mRNA as an active ingredient. The composition may be for delivering the mRNA or a product expressed thereby to a subject.

The fourth aspect provides a composition including the mRNA as an active ingredient, the mRNA including the mRNA or a product expressed thereby to induce a therapeutic response or an immune response to the polypeptide.

In the third aspect and the fourth aspect, the composition may be for providing immunogenicity or for treating a disease. The disease may differ according to the selected polypeptide. The disease may be, for example, infection caused by a virus or a foreign cell, a cancer, a metabolic disease, an inflammatory disease, a gastrointestinal disease,

an endocrine disease, sepsis, or an autoimmune disease. In the disease, the foreign cells may be prokaryotic or eukaryotic cells. The foreign cells may be bacteria.

The mRNA may include a sequence encoding a polypeptide. The polypeptide may be an antigenic or therapeutic polypeptide. The antigenic polypeptide may be a viral antigenic polypeptide. The polypeptide is as described in the first aspect.

Also, the mRNA may have anticancer activity by regulating the expression of other genes such as siRNA or miRNA.

The composition may be brought in contact with a cell, a tissue, or a subject in an "effective amount" of the mRNA.

The effective amount may be determined based on a target tissue, a target cell, a means of administration, physical properties of the mRNA (e.g., size of the mRNA and an amount of modified nucleoside), and other components of the composition. Generally, an effective amount of the composition may cause an induced or boosted immune response as a function of the antigen production in the cells.

The composition may be administered by intramuscular, subcutaneous, intradermal, intranasal, or pulmonary administration.

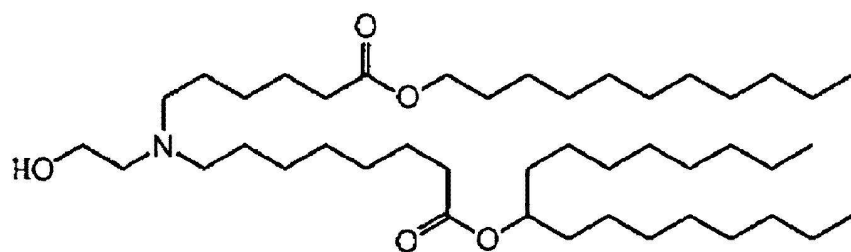
The composition may include at least one pharmaceutically acceptable carrier or an excipient. In the composition, the mRNA may be formulated with or in complex with the carrier or the excipient. The carrier or the excipient may be one known in the art.

Relative amounts of active ingredients, pharmaceutically acceptable carriers or excipients, and/or other additional components included in the composition may vary depending on the identity, size and/or conditions of the subject to be treated, and/or routes of administration. For example, the composition may include 0.1 % to 100 %, for example, 0.5 % to 50 %, 1.0 % to 30 %, 5.0 % to 80 %, or 80 % (w/w) or more of active ingredients.

The composition may be formulated as nanoparticles. The nanoparticles may be those known in the art for delivering a polynucleotide, such as an mRNA, into a cell. For example, the nanoparticles may be those known to be used in an RNA vaccine. The nanoparticles may be lipid nanoparticles (LNPs). The composition may be formulated as or bound to LNPs. The binding may include binding with LNPs or binding to the surface of LNPs. For example, the composition may be formulated inside a lipid polycation complex, or may be bound to a lipid polycation complex. The lipid polycation complex is

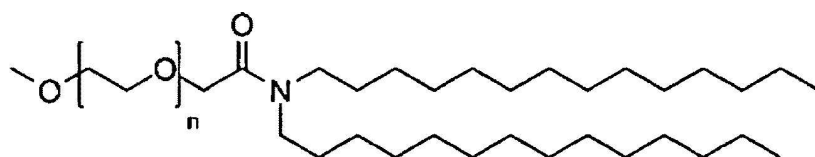
also known as a cationic lipid nanoparticle. The polycation may include MC3, Lipid 319, C12-200, 5A2-SC8, 306Oi10, Moderna Lipid 5, Acuitas A9, SM-102, ALC-0315, Arcturus Lipid 2,2 (8,8) 4C CH₃, Genevant CL1, or cationic polypeptides such as polylysine, polyornithine, and/or polyarginine. Also, the composition may be formulated with or bound to LNPs including 1,2 distearoyl-sn-glycero-3-phosphocholine (DSPC), dipalmitoylphosphatidylcholine (DPPC), sterol such as cholesterol, or a non-cationic lipid such as dioleoylphosphatidylethanolamine (DOPE). Also, the composition may be formulated with or bound to LNPs including a polyethyleneglycol (PEG)-lipid such as 1,2-dimyristoyl-rac-glycero-3-methoxypolyethylene glycol-2000 (PEG2000-DMG), 2-[(polyethylene glycol)-2000]-N,N-ditetradecylacetamide (ALC-0159), and polyethylene glycoldimethacrylate (PEG-DMA).

The LNP formulation may include cationic lipids, phospholipids, sterols such as cholesterol, PEG-lipids, or a combination thereof. The LNP formulation may include, for example, cationic lipids, phospholipids, sterols such as cholesterol, and PEG-lipids. Also, the LNPs may include PEG-modified lipids, non-cationic lipids, sterols, ionizable lipids, or a combination thereof. The LNPs may include 0.5 to 15 mol% of PEG-modified lipids, 5 to 25 mol% of non-cationic lipids, 25 to 55 mol% of sterols, and 20 to 60 mol% of ionizable lipids. The PEG-modified lipid may be PEG2000-DMG, the non-cationic lipid may be DSPC, the sterol may be cholesterol, and the ionizable cationic lipid may be Compound 1 having the following structure:

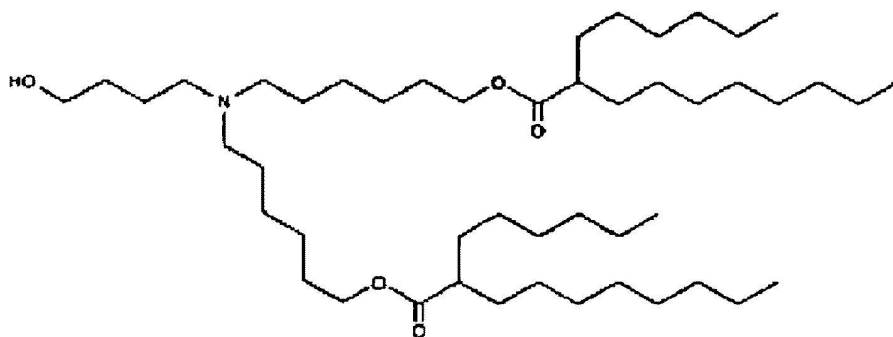


(Compound 1).

The PEG-modified lipid may be Compound 2 (ALC-0159) having the following structure, the non-cationic lipid may be DSPC, the sterol may be cholesterol, and the ionizable cationic lipid may be Compound 3 (ALC-0315) having the following structure:



(Compound 2); and



(Compound 3).

The fifth aspect provides a method of utilizing the mRNA, the method including introducing the mRNA to a host cell.

In the method, the host cell may be any cell derived from a subject. The subject may be a mammal including a human. The cell may be, for example, a stem cell, a reproductive cell or a somatic cell. In the method, the host cell may be an antigen-presenting cell including a dendritic cell, a monocyte, or a macrophage.

The method may include administering the mRNA into a subject. The administering may be parenteral or oral administration. The administering may be intramuscular, subcutaneous, intradermal, intranasal, or pulmonary administration. The subject may be a mammal including a human.

The method may be to immunize the subject against the polypeptide or to induce a therapeutic response to the therapeutic polypeptide. In addition, the method may treat a disease in a subject.

The sixth aspect provides DNA encoding the mRNA.

The DNA may further include a sequence for introducing a transcribable nucleotide sequence. The sequence for introducing a transcribable nucleotide sequence may be operatively linked to a 3'UTR upstream of a nucleotide encoding the 3'UTR. The sequence for introducing a transcribable nucleotide sequence may be a cloning site. The cloning site may be a multiple cloning site. The cloning site may include a recognition site for a restriction enzyme, a cleavage site, or a combination thereof.

The DNA may further include a promoter sequence. The promoter may be operatively linked to the 5'UTR upstream of the 5'UTR.

The nucleotide sequence encoding the 5'-UTR, the sequence encoding a polypeptide or the sequence for introducing a transcribable nucleotide, and the nucleotide sequence encoding the 3'UTR may be linked to one another to form a common transcript when transcription occurs in vitro or in vivo.

The DNA may be a vector. The vector may be an expression vector or a transcription vector, for the expression of a target protein or an mRNA.

The seventh aspect provides a method of obtaining an mRNA, the method including transcribing an mRNA by using the DNA as a template. The transcribing may include in vitro, ex vivo, or in vivo transcription. The in vivo transcription may include transcription occurring in a subject.

The term "in vitro transcription" or "RNA in vitro transcription" refers to a process of synthesizing RNA including an mRNA in a non-cell system, that is, in vitro. The cloning vector DNA including a plasmid DNA vector may be applied as a template for generating an RNA transcript. These cloning vectors are also generally known as transcription vectors. RNA may be obtained by DNA-dependent in vitro transcription with a suitable DNA template. The DNA template may be a linearized plasmid DNA template. The promoter controlling RNA in vitro transcription may be an arbitrary promoter for a DNA-dependent RNA polymerase. The DNA-dependent RNA polymerase may be a T7 RNA polymerase, a T3 RNA polymerase, a SP6 RNA polymerase, or a combination thereof. The DNA template for RNA in vitro transcription may be obtained by cloning a nucleic acid and introducing the same to a vector for RNA in vitro transcription. The DNA may include cDNA corresponding to each RNA to be transcribed in vitro. The vector for RNA in vitro transcription may be circular plasmid DNA. The cDNA may be obtained by reverse-transcription of an mRNA or by chemical synthesis.

The transcription vector may include a promoter, a nucleotide sequence encoding the 5'-UTR, a sequence encoding a polypeptide, such as an ORF, a nucleotide sequence encoding the 3'-UTR, and a poly (A) tail. The nucleotide sequence of the 5'-UTR may include at least one of nucleotide sequences of SEQ ID NOs: 1 to 9. The nucleotide sequence of the 3'UTR may include any known 3'UTR sequence. The transcription vector

may also include a replication origin, such as *E. coli* replication origin ColE1 ori, a selection marker gene, or a combination thereof. The promoter may be a T7 promoter. The selection marker may be an antibiotic resistance enzyme, such as a kanamycin resistance enzyme.

The in vitro transcription may include obtaining a transformed host cell by introducing the transcription vector to a host cell, for example, *E. coli*, and proliferating the host cell, which includes the plasmid DNA that becomes a template for an mRNA, by culturing the host cell. The in vitro transcription may include isolating the plasmid DNA from the proliferated host cell, for example, *E. coli*. The isolating of the plasmid DNA may include isolating a cell layer from the culture and isolating the plasmid DNA from the cell layer. The isolating of the cell layer may be performed by centrifugation, separation, precipitation, or a combination thereof. Isolation of plasmid DNA from cells may be performed by methods known in the art, such as an alkali extraction method, affinity chromatography, high performance liquid chromatography (HPLC), electrophoresis, or a combination thereof.

The obtaining of RNA may include producing an RNA transcript of the DNA by culturing the cell including the DNA.

Advantageous Effects of Invention

The mRNA according to an aspect is relatively stable and has a high production yield, and thus may be used to transcribe and translate an mRNA efficiently.

When a method of obtaining polypeptide according to another aspect is used, a polypeptide may be obtained efficiently.

The composition for delivering the mRNA or a product expressed thereby to a subject, including the mRNA, according to another aspect may be used to deliver the mRNA or a product expressed thereby to a subject

The composition for inducing a therapeutic response or an immune response to a polypeptide according to another aspect may be used to induce a therapeutic response or an immune response to a polypeptide in a subject.

The DNA according to another aspect may be used to obtain the mRNA.

According to the method of obtaining RNA in accordance with another aspect, RNA may be obtained efficiently.

Best Mode for Carrying out the Invention

Hereinafter, the present disclosure will be described in more detail through examples. However, these examples are intended to illustrate the present disclosure, and the scope of the present disclosure is not limited to these examples.

Example 1: Unstructured 5'UTR non-natural sequence increasing yield of genes

5'UTR sequences increasing a yield of genes were selected. A 5'UTR sequence is a non-natural sequence that does not exist in nature, and has activity of increasing a yield of genes, compared to the native sequences.

First, random non-natural sequences with a length of 50 bp or 70 bp and a mean free energy of 0 were selected, and sequences that did not form a secondary structure between bases were selected.

The selected mRNA sequences were designed to start with 'AG' for capping optimization to protect the 5'-end and end with 'GCCACC' which is a Kozak sequence for stable gene expression.

For the sequence selection, sequences having an excellent activity of increasing the yield of genes were first selected. Based on these sequences, some base sequences were substituted to improve the activity in increasing the yield of genes, and therefore 9 non-natural sequences were finally selected (Table 1).

【Table 1】

SEQ ID NO	5'UTR sequence
1	AGCCGUCUCCAUCGUUACAGUCACCGUUUUCGCAUUCUGUCUA GCCACC
2	AGAAUCCAGUCAACGCAUAUCCACGUAUACCCGAACACAUCACAG CCACC
3	AGCCCCCAACACUGACUUAUUCGCUUUAAGUUCCUCCAAACGAG CCACC
4	AGAGCAGAACGCGUAACCACACCAUCGCAACAUCCACUUGAUAG CCACC

5	AGACCAGACACCUAACAGACUAGACAUUUUACCGAAAUCCGCAUU AACCAUUCUAAACUCGUUAUAGCCACC
6	AGAUCACUAAACACCAUAUAACGCUAUCUCCUCCUUCUCCGACAU UCCAGUACUCAUUCCACAAGCCACC
7	AGUACCCACUCCAGAUAAACUGACAACCAUACGCGAACAACGACG AAUCGUUAUCACACGCAGAAGCCACC
8	AGCGAGAUAAACUUACAAGGAGAGAGAAGACGAGGAGAGAUAAAUU AGAGAGAAAACGAGUAAGAGCCACC
9	AGUAAACUAACCACUCCAUCACGUUAUAUUAUCCUCUUUCACACCU UCCAUAUUCUGAACUAACCGCCACC

Example 2: Preparation of template vector using non-natural 5'UTR sequence

One of the sequences encoding the non-natural 5'UTR nucleotide sequences including the nucleotide sequences of SEQ ID NOs: 1 to 9 obtained in Example 1, a gene encoding a polypeptide, and a sequence encoding a native CYP2E1 3'UTR nucleotide sequence were operatively linked to one another to prepare a vector. The polypeptide may be a wild-type p53 protein having an amino acid sequence of SEQ ID NO: 10.

Specifically, the vector is one in which a promoter, a nucleotide sequence encoding 5'-UTR, an open reading frame (ORF) encoding a polypeptide, a nucleotide sequence encoding 3'-UTR, and a poly (A) tail.

The vector may also include *E. coli* replication origin ColE1 ori and a selection marker gene. The promoter may be a T7 promoter. The selection marker may be a kanamycin resistance enzyme.

2-1. Preparation of template vector for transcription of p53 mRNA

A vector for producing a native p53 mRNA and/or a protein product thereof was prepared, the vector including one of nucleotides encoding the designed non-natural 5'UTR sequence, a nucleotide sequence, as an ORF, encoding a native p53 protein, a nucleotide sequence encoding a native CYP2E1 3'UTR, and a poly (A). This vector was then introduced to a microbial cell, and the resulting microbial cell was cultured to produce a cloning vector. Through in vitro transcription with the produced vector as a template, mRNA was produced.

Specifically, a vector pUC57-Kan^R, in which an antibiotic resistance gene was substituted with a kanamycin resistance gene, was produced from the pUC57-Amp^R vector (Addgene). The produced vector was treated and cleaved with restriction enzymes HindIII and EcoRI, and the cleaved vector was connected to a synthetically produced polynucleotide encoding a native p53 protein by using a ligase to insert a sequence encoding a native p53 protein. Since the inserted polynucleotide encoding a native p53 protein includes a nucleotide sequence encoding 5'UTR and a sequence encoding a native p53 protein, and restriction enzyme recognition sequences are at the 3'-end of the native p53 nucleotide sequence, the restriction enzyme recognition sequences were cleaved to introduce a nucleotide sequence encoding the native P450 2E1 3'UTR.

【Table 2】

SEQ ID NO	Gene sequence (DNA) of template vector for transcription of native p53 mRNA
11	<u>TAATACGACTCACTATA</u>
12	AGTGCTCCCCATCCAATAAAGTGTCCCTCTGTCCGAACTAAAGTGC ACTGTCCCAGCACC
13	TCCGCGAATTGAACGAGGCCCTGGAGTTGAAAGACGCTCAAGCCGGA AAGGAACCAGGTGGGTCTAGGGCCCACTCAAGCCATCTGAAGAGTAA GAAAGGCCAGTCCACTTCCAGGCACAAAAAGTTAATGTTCAAGACGG AGGGTCCTGACTCTGAC
14	GTGTGTGGAGGACACCCTGAACCCCCCGCTTTCAAACAAGTTTTCAAA TTGTTTGAGGTCAGGATTTCTCAAAGTATTCTTTCTTTGCATATGAG TATTTGAAAATAAATATTTTCCCAGAATATAAATAAATCATCACATGATT ATTTTAACTAT
15	AA AA AA

Table 2 shows a promoter (SEQ ID NO: 1) included in the prepared pUC57-Kan^R vector, a polynucleotide encoding 5'UTR (SEQ ID NO: 12), a polynucleotide encoding a native p53 protein (SEQ ID NO: 13), a polynucleotide encoding 3'UTR (SEQ ID NO: 4),

and a poly (A) sequence (SEQ ID NO: 15). In Table 2, the underlined nucleotides in SEQ ID NO: 11 represent a portion of a T7 promoter involved in the initiation of mRNA transcription. Also, the bolded nucleotides of SEQ ID NO: 12 represent a human complement component 3 (C3) native 5'UTR sequence. Also, SEQ ID NO: 13 is a nucleotide sequence downstream of the 5'UTR, and is a nucleotide sequence encoding a native p53 protein. The native p53 protein had an amino acid sequence of SEQ ID NO: 10. Also, SEQ ID NO: 14 is a native 3'UTR sequence of human cytochrome P450 2E1 gene.

2-2. Connection of non-natural 5'UTR sequence to nucleotide sequence encoding native p53 protein and cloning

To substitute the 5'UTR sequence of the prepared template vector with a nucleotide sequence encoding one non-natural 5'UTR sequence of SEQ ID NOs: 1 to 9 or a nucleotide sequence encoding a native C3 5'UTR sequence of SEQ ID NO: 16, a DNA fragment containing the nucleotide sequence encoding the 5'UTR was obtained through a PCR reaction with a primer set of a primer including the nucleotide sequence encoding the non-natural 5'UTR sequence and a primer including the sequence of the XhoI restriction enzyme recognition sequence of the native p53 protein, and with the template vector as a template.

Each of DNA fragments containing the nucleotide sequences encoding 5'-UTRs secured by PCR reaction, and the template vector prepared in Example 2-1 were treated and cleaved with restriction enzymes HindIII and XhoI, and the cleaved sequences were connected by using a ligase. As a result, a vector encoding the native p53 protein substituted with a nucleotide sequence encoding the 5'UTR of one of SEQ ID NOs: 1 to 9 was prepared.

Example 4: Production and yield of gene product using non-natural 5'UTR sequence

The vector prepared in Example 3, having a non-natural 5'UTR sequence-p53 protein-coding sequence-3'UTR sequence-poly (A) structure, was introduced into *E. coli*, and the *E. coli* was cultured to proliferate the *E. coli* including a plasmid DNA as a template for the mRNA. The proliferated *E. coli* was isolated from the culture by high-speed centrifugation, and the cell-derived template DNA was secured by using an alkali

extraction method. The secured template DNA was linearized by using the restriction enzyme recognition sequence located at the end of the poly (A) tail. mRNA having nucleotide sequences with a non-natural 5'UTR-p53 protein-coding sequence-3'UTR sequence-poly (A) structure was produced via in vitro transcription using the linearized cell-derived template DNA. Here, the non-natural 5'UTR has a sequence of any one of SEQ ID NOs: 1 to 9, the 3'UTR has a sequence of SEQ ID NO: 14 (native CYP2E1), and the poly(A) and the p53 protein-coding protein are as described above.

Specifically, to an aqueous solution containing 10 ng/μl of the linearized plasmid DNA, 4 mM of a capping reagent, 5 mM of ATP, 5 mM of CTP, 5 mM of GTP, 5 mM of N1-methylpseudouridine-5'-triphosphate, 20 mM of magnesium ion salt, 10 mM of DTT, 0.01 U/μL of inorganic pyrophosphatase (PPase), and 1 U/μL of RNAase inhibitor, 5 U/μL of T7 RNA polymerase was added, and the reaction mixture thus obtained was incubated at 37 °C for at least one hour, to proceed in vitro transcription. For use as the capping reagent, the cap analog compound m7G(3'OMs)pppA(2'OMe)pG of Example 8 described in 10-2023-0141482 published on October 10, 2023 was used. The capping reagent is a reagent for mRNA transcription and co-transcriptional capping, and forms a 5'-cap structure of m7G(3'OMs)pppA(2'OMe)pG. Here, m7G represents 7-methylguanosine, A represents adenosine, G represents guanosine, p is -P(=O)(OH)O-, Ms represents mesyl, and Me represents methyl. The reaction product was treated with a DNase, and mRNAs (SEQ ID NOs: 17 to 26) of Materials 1 to 10 were obtained by using a Monarch® RNA Cleanup Kit (New England Biolabs). A configuration of nucleotide sequences of the non-natural 5'UTR, p53 protein sequence, and 3'UTR included in the mRNAs and a relative synthesis percentage according to the UTR combination are as shown in Table 3. The mRNAs may further have a 5'-cap structure and a poly (A) sequence structure (SEQ ID NOs: 27 to 36). The relative synthesis yield represents a relative synthesis yield of the mRNA according to in vitro transcription, and was calculated based on a synthesis yield of Material 1 (100 %).

Relative synthesis yield (%) = (synthesis yield of each material / synthesis yield of Material 1) × 100

【Table 3】

Material No.	5'UTR sequence	ORF sequence	3'UTR sequence	Yield (mg/mL)	Relative synthesis yield (%)
1	Native C3 SEQ ID NO: 16	Native p53	Native CYP2E1	4.75	100.0
2	SEQ ID NO: 1	Native p53	Native CYP2E1	4.94	104.0
3	SEQ ID NO: 2	Native p53	Native CYP2E1	6.99	147.2
4	SEQ ID NO: 3	Native p53	Native CYP2E1	7.30	153.7
5	SEQ ID NO: 4	Native p53	Native CYP2E1	7.34	154.5
6	SEQ ID NO: 5	Native p53	Native CYP2E1	7.02	147.8
7	SEQ ID NO: 6	Native p53	Native CYP2E1	7.38	155.4
8	SEQ ID NO: 7	Native p53	Native CYP2E1	6.64	139.8
9	SEQ ID NO: 8	Native p53	Native CYP2E1	7.83	164.8
10	SEQ ID NO: 9	Native p53	Native CYP2E1	7.28	152.2

As shown in Table 3, it is confirmed that the structural changes according to the UTR sequences can lead to different transcription efficiency, which in turn leads to an improved synthesis yield of p53 mRNA through *in vitro* transcription (IVT) depending on the non-natural 5'UTR. Also, it was confirmed that the translation yield increased in the translation using Materials 2 to 10.

Claims

- [Claim 1] A messenger RNA (mRNA) comprising a non-natural 5' untranslated region (5'UTR),
wherein the 5'UTR does not form a secondary structure.
- [Claim 2] The mRNA of claim 1, wherein the mRNA comprising the non-natural 5'UTR has a sequence of Formula 1:
Formula1: AGN_aGCCACC
wherein N is A, U, G, or C, each N is identical to or different from each other, and a indicates the number of repeats of N and is 42 or 62.
- [Claim 3] The mRNA of claim 1, wherein the mRNA comprising the non-natural 5'UTR comprises a nucleotide sequence of any one of SEQ ID NOs: 1 to 9.
- [Claim 4] The mRNA of claim 1, wherein the mRNA further comprises a sequence encoding a polypeptide.
- [Claim 5] The mRNA of claim 4, wherein the mRNA further comprises a Kozak sequence, a 3'UTR sequence, or a poly (A) sequence.
- [Claim 6] The mRNA of claim 1, wherein the mRNA comprises a 5'-cap selected from Cap0, Cap1, and Cap2.
- [Claim 7] The mRNA of claim 1, wherein the mRNA has one or more modified nucleotides.
- [Claim 8] DNA encoding the mRNA of any one of claims 1 to 7.
- [Claim 9] A composition comprising, as an active ingredient, the mRNA or the DNA of any one of claims 1 to 8.