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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 April 2017 (20.04.2017)

(10) International Publication Number
WO 2017/066573 A1

(51) International Patent Classification:
C12N 9/12 (2006.01) *C12N 15/62* (2006.01)

(21) International Application Number:
PCT/US2016/057044

(22) International Filing Date:
14 October 2016 (14.10.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/241,350 14 October 2015 (14.10.2015) US

(71) Applicant: **SHIRE HUMAN GENETIC THERAPIES, INC.** [US/US]; 300 Shire Way, Lexington, MA 02421 (US).

(72) Inventors: **STRACK-LOGUE, Bettina**; 300 Shire Way, Lexington, MA 02421 (US). **DEROSA, Frank**; 300 Shire Way, Lexington, MA 02421 (US). **HEARTLEIN, Michael**; 300 Shire Way, Lexington, MA 02421 (US).

(74) Agents: **CHEN, Fangli** et al.; Proskauer Rose LLP, One International Place, Boston, MA 02110 (US).

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

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

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

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

WO 2017/066573 A1

(54) Title: MODIFICATION OF RNA-RELATED ENZYMES FOR ENHANCED PRODUCTION

(57) Abstract: The present invention provides, among other things, methods and compositions for large-scale production of capped mRNA using SUMO- Guanylyl Transferase fusion protein.

MODIFICATION OF RNA-RELATED ENZYMES FOR ENHANCED PRODUCTION

RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application Serial No. 62/241,350, filed October 14, 2015, the disclosure of which is hereby incorporated by reference.

SEQUENCE LISTING

[0002] The present specification makes reference to a Sequence Listing (submitted electronically as a .txt file named "SL_SHR-1187WO" on October 14, 2016. The .txt file was generated October 14, 2016 and is 28,402 bytes in size. The entire contents of the Sequence Listing are herein incorporated by reference.

BACKGROUND

[0003] Messenger RNA ("mRNA") therapy is becoming an increasingly important approach for the treatment of a variety of diseases. Effective mRNA therapy requires effective delivery of the mRNA to the patient and efficient production of the protein encoded by the mRNA within the patient's body. To optimize mRNA delivery and protein production in vivo, a proper cap are typically required at the 5' end of the construct, which protects the mRNA from degradation and facilitates successful protein translation. Therefore, the large-scale production of enzymes capable of capping mRNA is particularly important for producing mRNA for therapeutic applications.

SUMMARY OF THE INVENTION

[0004] The present invention provides improved methods for effective production of enzymes capable of capping mRNA. The present invention is, in part, based on the surprising discovery that modifying a guanylyl transferase (GT) with a SUMO tag makes it possible to produce GT on the large scale needed for producing capped mRNA for therapeutic applications.

[0005] Thus, in one aspect, the present invention provides methods of producing a capped RNA or RNA analog oligonucleotide, wherein a fusion protein facilitates the steps of transferring and methylating a guanylyl molecule to the 5' end of the RNA or RNA analog oligonucleotide.

[0006] In some embodiments, the fusion protein comprises a guanylyl transferase and a small ubiquitin-like molecule (SUMO) protein. In some embodiments, the guanylyl transferase comprises SEQ ID NO: 6 and and SEQ ID NO: 7 and the SUMO protein comprises SEQ ID NO: 5. In some embodiments, the fusion protein comprises SEQ ID NO: 8 and SEQ ID NO: 7.

[0007] In some embodiments, the one end of the RNA or RNA analog oligonucleotide is the 5' end.

[0008] In some embodiments, the fusion protein has comparable phosphatase activity, guanylyl transferase activity and methylation activity relative to a wild-type guanylyl transferase protein.

[0009] In another aspect, the present invention provides fusion proteins, wherein a fusion protein comprises guanylyl transferase and a small ubiquitin-like molecule (SUMO) protein.

[0010] In some embodiments, the guanylyl transferase comprises SEQ ID NO: 6 and SEQ ID NO: 7 and the SUMO protein comprises SEQ ID NO: 5. In some embodiments, the guanylyl transferase comprises a large subunit and a small subunit. In some embodiments, the SUMO protein is covalently linked and co-expressed with the large subunit. In some embodiments, the fusion protein has comparable phosphatase activity, guanylyl transferase activity and methylation activity relative to a wild-type guanylyl transferase protein.

[0011] In another aspect, the present invention provides vectors encoding a fusion protein comprising guanylyl transferase protein and a small ubiquitin-like molecule (SUMO) protein.

[0012] In some embodiments, the vector comprises SEQ ID NO: 1 and SEQ ID NO: 2. In some embodiments, the vector comprises SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 3. In some embodiments, the vector comprises SEQ ID NO: 4 and SEQ ID NO: 3.

[0013] In another aspect, the present invention provides methods to produce a guanylyl transferase by fermentation, comprising: a) culturing in a fermentation medium a

microorganism that is transformed with at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding a guanylyl transferase that has an amino acid sequence that is at least 90% identical SEQ ID NO: 6 and SEQ ID NO: 7; and b) collecting a product produced from the step of culturing.

[0014] In some embodiments, the guanylyl transferase comprises a guanylyl transferase fusion protein. In some embodiments, the guanylyl transferase fusion protein has comparable phosphatase activity, guanylyl transferase activity and methylation activity relative to a wild-type guanylyl transferase protein. In some embodiments, the guanylyl transferase fusion protein comprises a small ubiquitin-like molecule (SUMO) protein. In some embodiments, the guanylyl transferase fusion protein comprises SEQ ID NO: 8.

[0015] In some embodiments, the SUMO protein is bound to the guanylyl transferase by a covalent link. In some embodiments, the covalent link is between the SUMO protein and a large subunit of the guanylyl transferase.

[0016] In some embodiments, the fermentation medium is selected from the group consisting of Terrific Broth, Cinnabar, 2xYT and LB. In some embodiments, the microorganism is a bacterium.

[0017] In some embodiments, the nucleic acid sequence encoding the guanylyl transferase is at least 90% identical to SEQ ID NO: 2 and SEQ ID NO: 3.

[0018] In some embodiments, the recombinant nucleic acid molecule further comprises a nucleic acid sequence encoding a small ubiquitin-like molecule (SUMO) protein. In some embodiments, the nucleic acid sequence encoding a small ubiquitin-like molecule (SUMO) protein is at least 90% identical to SEQ ID NO: 1.

[0019] In some embodiments, the product is a guanylyl transferase. In some embodiments, the product is a guanylyl transferase comprises a guanylyl transferase fusion protein. In some embodiments, the guanylyl transferase fusion protein further comprises a small ubiquitin-like molecule (SUMO) protein.

BRIEF DESCRIPTION OF THE DRAWING

[0020] The drawings are for illustration purposes and are in no way limiting.

[0021] **Figures 1A and 1B** are diagrams of exemplary mRNA capped structures present in various embodiments of the invention.

[0022] **Figure 2** demonstrates exemplary yield of soluble SUMO-GT protein produced by fermentation compared to that of GT protein produced via the shake flask method.

DEFINITIONS

[0023] In order for the present invention to be more readily understood, certain terms are first defined. Additional definitions for the following terms and other terms are set forth throughout the specification.

[0024] *Approximately*: As used herein, the term "approximately" or "about," as applied to one or more values of interest, refers to a value that is similar to a stated reference value. In certain embodiments, the term "approximately" or "about" refers to a range of values that fall within 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value).

[0025] *Batch culture*: As used herein, the term "batch culture" refers to a method of culturing cells in which all the components that will ultimately be used in culturing the cells, including the medium (see definition of "medium" below) as well as the cells themselves, are provided at the beginning of the culturing process. Thus, a batch culture typically refers to a culture allowed to progress from inoculation to conclusion without refeeding the cultured cells with fresh medium. A batch culture is typically stopped at some point and the cells and/or components in the medium are harvested and optionally purified.

[0026] *Biologically active*: As used herein, the phrase "biologically active" refers to a characteristic of any substance that has activity in a biological system (e.g., cell culture, organism, etc.). For instance, a substance that, when administered to an organism, has a biological effect on that organism, is considered to be biologically active. Biological activity can also be determined by in vitro assays (for example, in vitro enzymatic assays such as sulfate release assays). In particular embodiments, where a protein or polypeptide is biologically active, a portion of that protein or polypeptide that shares at least one biological activity of the protein or polypeptide is typically referred to as a "biologically active" portion. In some embodiments, a protein is produced and/or purified from a cell culture system, which displays biologically activity when administered to a subject. In some embodiments, a

protein requires further processing in order to become biologically active. In some embodiments, a protein requires posttranslational modification such as, but is not limited to, glycosylation (e.g., sialyation), farnysylation, cleavage, folding, formylglycine conversion and combinations thereof, in order to become biologically active. In some embodiments, a protein produced as a proform (i.e. immature form), may require additional modification to become biologically active.

[0027] *Bioreactor*: As used herein, the term "bioreactor" refers to a vessel used for the growth of a host cell culture. A bioreactor can be of any size so long as it is useful for the culturing of mammalian cells. Typically, a bioreactor will be at least 1 liter and may be 10, 100, 250, 500, 1000, 2500, 5000, 8000, 10,000, 12,0000 liters or more, or any volume in between. Internal conditions of a bioreactor, including, but not limited to pH, osmolarity, CO2 saturation, O2 saturation, temperature and combinations thereof, are typically controlled during the culturing period. A bioreactor can be composed of any material that suitable for holding cells in media under the culture conditions of the present invention, including glass, plastic or metal. In some embodiments, a bioreactor may be used for performing animal cell culture. In some embodiments, a bioreactor may be used for performing mammalian cell culture. In some embodiments, a bioreactor may be used with cells and/or cell lines derived from such organisms as, but not limited to, mammalian cell, insect cells, bacterial cells, yeast cells and human cells. In some embodiments, a bioreactor is used for large-scale cell culture production and is typically at least 100 liters and may be 200, 500, 1000, 2500, 5000, 8000, 10,000, 12,0000 liters or more, or any volume in between. One of ordinary skill in the art will be aware of and will be able to choose suitable bioreactors for use in practicing the present invention.

[0028] *Cell density*: As used herein, the term "cell density" refers to that number of cells present in a given volume of medium.

[0029] *Cell culture or culture*: As used herein, these terms refer to a cell population that is gown in a medium under conditions suitable to survival and/or growth of the cell population. As will be clear to those of ordinary skill in the art, these terms as used herein may refer to the combination comprising the cell population and the medium in which the population is grown.

[0030] *Cultivation*: As used herein, the term "cultivation" or grammatical equivalents refers to a process of maintaining cells under conditions favoring growth or survival. The

terms “cultivation” and “cell culture” or any synonyms are used inter-changeably in this application.

[0031] *Culture vessel*: As used herein, the term “culture vessel” refers to any container that can provide an aseptic environment for culturing cells. Exemplary culture vessels include, but are not limited to, glass, plastic, or metal containers.

[0032] *Expression*: As used herein, “expression” of a nucleic acid sequence refers to one or more of the following events: (1) production of an RNA template from a DNA sequence (e.g., by transcription); (2) processing of an RNA transcript (e.g., by splicing, editing, 5’ cap formation, and/or 3’ end formation); (3) translation of an RNA into a polypeptide or protein; and/or (4) post-translational modification of a polypeptide or protein.

[0033] *Fed-batch culture*: As used herein, the term “fed-batch culture” refers to a method of culturing cells in which additional components are provided to the culture at some time subsequent to the beginning of the culture process. The provided components typically comprise nutritional supplements for the cells which have been depleted during the culturing process. A fed-batch culture is typically stopped at some point and the cells and/or components in the medium are harvested and optionally purified.

[0034] *Homology*: As used herein, the term “homology” refers to the overall relatedness between polymeric molecules, *e.g.*, between nucleic acid molecules (*e.g.*, DNA molecules and/or RNA molecules) and/or between polypeptide molecules. In some embodiments, polymeric molecules are considered to be “homologous” to one another if their sequences are at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identical. In some embodiments, polymeric molecules are considered to be “homologous” to one another if their sequences are at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% similar.

[0035] *Identity*: As used herein, the term “identity” refers to the overall relatedness between polymeric molecules, *e.g.*, between nucleic acid molecules (*e.g.*, DNA molecules and/or RNA molecules) and/or between polypeptide molecules. Calculation of the percent identity of two nucleic acid sequences, for example, can be performed by aligning the two sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second nucleic acid sequences for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least

60%, at least 70%, at least 80%, at least 90%, at least 95%, or substantially 100% of the length of the reference sequence. The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which needs to be introduced for optimal alignment of the two sequences. The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For example, the percent identity between two nucleotide sequences can be determined using the algorithm of Meyers and Miller (CABIOS, 1989, 4: 11-17), which has been incorporated into the ALIGN program (version 2.0) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. The percent identity between two nucleotide sequences can, alternatively, be determined using the GAP program in the GCG software package using an NWSgapdna.CMP matrix. Various other sequence alignment programs are available and can be used to determine sequence identity such as, for example, Clustal.

[0036] *Integrated Viable Cell Density*: As used herein, the term "integrated viable cell density" refers to the average density of viable cells over the course of the culture multiplied by the amount of time the culture has run. Assuming the amount of polypeptide and/or protein produced is proportional to the number of viable cells present over the course of the culture, integrated viable cell density is a useful tool for estimating the amount of polypeptide and/or protein produced over the course of the culture.

[0037] *Isolated*: As used herein, the term "isolated" refers to a substance and/or entity that has been (1) separated from at least some of the components with which it was associated when initially produced (whether in nature and/or in an experimental setting), and/or (2) produced, prepared, and/or manufactured by the hand of man. Isolated substances and/or entities may be separated from about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more than about 99% of the other components with which they were initially associated. In some embodiments, isolated agents are about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more than about 99% pure. As used herein, a substance is "pure" if it is substantially free of other

components. As used herein, calculation of percent purity of isolated substances and/or entities should not include excipients (*e.g.*, buffer, solvent, water, *etc.*)

[0038] *Medium*: As used herein, the term "medium" refer to a solution containing nutrients which nourish growing cells. Typically, these solutions provide essential and non-essential amino acids, vitamins, energy sources, lipids, and trace elements required by the cell for minimal growth and/or survival. The solution may also contain components that enhance growth and/or survival above the minimal rate, including hormones and growth factors. In some embodiments, medium is formulated to a pH and salt concentration optimal for cell survival and proliferation. In some embodiments, medium may be a "chemically defined medium" – a serum-free media that contains no proteins, hydrolysates or components of unknown composition. In some embodiment, chemically defined medium is free of animal-derived components and all components within the medium have a known chemical structure. In some embodiments, medium may be a "serum based medium" – a medium that has been supplemented with animal derived components such as, but not limited to, fetal calf serum, horse serum, goat serum, donkey serum and/or combinations thereof.

[0039] *Nucleic acid*: As used herein, the term "nucleic acid," in its broadest sense, refers to a compound and/or substance that is or can be incorporated into an oligonucleotide chain. In some embodiments, a nucleic acid is a compound and/or substance that is or can be incorporated into an oligonucleotide chain via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (e.g., nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising individual nucleic acid residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably. In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA and/or cDNA. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, i.e., analogs having other than a phosphodiester backbone. For example, the so-called "peptide nucleic acids," which are known in the art and have peptide bonds instead of phosphodiester bonds in the backbone, are considered within the scope of the present invention. The term "nucleotide sequence encoding an amino acid sequence" includes all nucleotide sequences that are degenerate versions of each other and/or encode the same amino acid sequence. Nucleotide sequences that encode proteins and/or RNA may include introns. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, etc. Where appropriate,

e.g., in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, backbone modifications, etc. A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. The term "nucleic acid segment" is used herein to refer to a nucleic acid sequence that is a portion of a longer nucleic acid sequence. In many embodiments, a nucleic acid segment comprises at least 3, 4, 5, 6, 7, 8, 9, 10, or more residues. In some embodiments, a nucleic acid is or comprises natural nucleosides (e.g., adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (e.g., 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, C-5 propynyl-cytidine, C-5 propynyl-uridine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (e.g., methylated bases); intercalated bases; modified sugars (e.g., 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (e.g., phosphorothioates and 5'-N-phosphoramidite linkages). In some embodiments, the present invention is specifically directed to "unmodified nucleic acids," meaning nucleic acids (e.g., polynucleotides and residues, including nucleotides and/or nucleosides) that have not been chemically modified in order to facilitate or achieve delivery.

[0040] *Perfusion process*: As used herein, the term "perfusion process" refers to a method of culturing cells in which additional components are provided continuously or semi-continuously to the culture subsequent to the beginning of the culture process. The provided components typically comprise nutritional supplements for the cells which have been depleted during the culturing process. A portion of the cells and/or components in the medium are typically harvested on a continuous or semi-continuous basis and are optionally purified. Typically, a cell culture process involving a perfusion process is referred to as "perfusion culture." Typically, nutritional supplements are provided in a fresh medium during a perfusion process. In some embodiments, a fresh medium may be identical or similar to the base medium used in the cell culture process. In some embodiments, a fresh medium may be different than the base medium but containing desired nutritional supplements. In some embodiments, a fresh medium is a chemically-defined medium.

[0041] *Seeding*: As used herein, the term "seeding" refers to the process of providing a cell culture to a bioreactor or another vessel for large scale cell culture production. In some embodiments a "seed culture" is used, in which the cells have been propagated in a smaller cell culture vessel, i.e. Tissue-culture flask, Tissue-culture plate, Tissue-culture roller bottle, etc., prior to seeding. Alternatively, in some embodiments, the cells may have been frozen and thawed immediately prior to providing them to the bioreactor or vessel. The term refers to any number of cells, including a single cell.

[0042] *Subject:* As used herein, the term "subject" means any mammal, including humans. In certain embodiments of the present invention the subject is an adult, an adolescent or an infant. Also contemplated by the present invention are the administration of the pharmaceutical compositions and/or performance of the methods of treatment in-utero.

[0043] *Vector*: As used herein, "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it is associated. In some embodiment, vectors are capable of extra-chromosomal replication and/or expression of nucleic acids to which they are linked in a host cell such as a eukaryotic and/or prokaryotic cell. Vectors capable of directing the expression of operatively linked genes are referred to herein as "expression vectors."

[0044] *Viable cell density*: As used herein, the term "viable cell density" refers to the number of living cells per unit volume.

DETAILED DESCRIPTION

[0045] The present invention provides, among other things, methods and compositions for large-scale production of capped mRNA using SUMO- Guanylyl Transferase fusion protein.

[0046] Various aspects of the invention are described in further detail in the following subsections. The use of subsections is not meant to limit the invention. Each subsection may apply to any aspect of the invention. In this application, the use of "or" means "and/or" unless stated otherwise.

SUMO-Guanylyl Transferase Fusion Protein

Small Ubiquitin-like Modifier (SUMO)

[0047] As used herein, a SUMO tag is any protein or a portion of a protein that can substitute for at least partial activity of a SUMO protein.

[0048] SUMO proteins are small proteins that are covalently attached to and detached from other proteins in order to modify the functions of those proteins. The modification of a protein with a SUMO protein is a post-translational modification involved in various cellular processes such as nuclear-cytosolic transport, transcriptional regulation, apoptosis, protein stability, response to stress and progression through the cell cycle. There are at least 4 SUMO paralogs in vertebrates, designated SUMO-1, SUMO-2, SUMO-3, and SUMO-4. SUMO-2 and SUMO-3 are structurally and functionally very similar and are distinct from SUMO-1. The amino acid sequence (SEQ ID NO: 1) spans amino acids 3-92 of a typical wild-type or naturally occurring SUMO-3 protein is shown in Table 1. In addition, a codon optimized DNA sequence encoding the SUMO-3 protein is also provided in Table 1, as SEQ ID NO: 5.

Table 1. Small Ubiquitin-like Modifier

SUMO-3 Protein sequence	EEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAY CERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG **(SEQ ID NO: 1)**
SUMO-3 DNA sequence	GAAGAGAAACCGAAAGAGGGCGTTAAGACCGAGAATGACCAC ATTAACCTGAAGGTCGCTGGTCAAGATGGCAGCGTGGTGCAGT TTAAGATCAAGCGTCACACGCCGTTGAGCAAGCTGATGAAGGC TTACTGCGAGCGTCAGGGTCTGAGCATGCGTCAGATCCGCTTTC GTTTCGATGGCCAGCCGATCAATGAGACTGACACCCCAGCGCA ACTGG **(SEQ ID NO: 5)**

[0049] Thus, in some embodiments, a SUMO protein is a human SUMO-3 protein (SEQ ID NO: 1). In some embodiments, the SUMO protein may be another SUMO paralog, such as SUMO-1, SUMO-2 or SUMO-4. In some embodiments, a suitable replacement protein may be a homologue or an analogue of human SUMO-3 protein. For example, a homologue or an analogue of SUMO-3 protein may be a modified SUMO-3 protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to a wild-type or naturally-occurring SUMO-3 protein (e.g. SEQ ID NO: 1), while retaining substantial SUMO-3 protein activity. Thus, in some embodiments, an enzyme suitable for the present invention is substantially homologous to a wild-type or naturally-occurring

SUMO-3 protein (SEQ ID NO: 1). In some embodiments, an enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 1. In some embodiments, an enzyme suitable for the present invention is substantially identical to a wild-type or naturally-occurring SUMO-3 protein (SEQ ID NO: 1). In some embodiments, a protein suitable for the present invention contains a fragment or a portion of a SUMO protein. In some embodiments, the SUMO protein comprises human SUMO-1, human SUMO-2, human SUMO-3, any one of *Arabidopsis Zhalania* SUMO-l through SUMO-8, tomato SUMO, any one of *Xenopus laevis* SUMO-l through SUMO-3, *Drosophila melanogasler* Smt3, *Caenorhabdilis elegans* SMO-1, *Schizosaccharomyces pombe* Pmt3, malarial parasite *Plasmodium falciparum* SUMO, mold *Aspergillus nidulans* SUMO, an equivalent thereof, a homologue thereof, or a combination thereof.

[0050] In some embodiments, the SUMO protein is encoded by a nucleic acid derived from an organism selected from the group consisting of human, mouse, insect, plant, yeast, and other eukaryotic organisms. In some embodiments, the SUMO protein is encoded by a nucleic acid derived from an organism selected from the group consisting of *Homo sapiens*, *Arabidopsis Zhalania*, tomato, *Xenopus laevis*, *Drosophila melanogasler*, *Caenorhabdilis elegans*, *Schizosaccharomyces pombe*, *Plasmo diumfalciparum*, or *Aspergillus nidulans*. In some embodiments, a nucleic acid suitable for the present invention has an sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 5. In some embodiments, a nucleic acid suitable for the present invention is substantially identical to a nucleic acid encoding a wild-type or naturally-occurring SUMO-3 protein (SEQ ID NO: 5).

Guanylyl Transferase (GT)

[0051] As used herein, a GT protein is any protein or portion of a protein that can substitute for at least partial activity of naturally-occurring Guanylyl Transferase (GT) protein. As used herein, the terms “a GT protein” and “a GT enzyme” and grammatical equivalents are used interchangeably.

[0052] GT is an enzyme derived from the Vaccinia Virus system that facilitates the transfer and methylation of a guanylyl molecule to the 5’ end of a messenger RNA molecule. This process, known as mRNA capping, is highly regulated and important for the creation of stable and mature mRNA able to undergo translation during protein synthesis. The GT

enzyme comprises a heterodimer that includes a "large subunit" (D1, about 97 kDa) and a "small subunit (D12, about 33 kDa). GT provides three enzymatic functions: phosphatase activity (cleavage of the nascent 5' triphosphate of mRNA to a diphosphate), guanylyl transferase activity (incorporation of a GTP molecule to the 5' end of the mRNA moiety) and methylation activity (incorporation of a methyl group at the N^7 position of the guanylyl base). The amino acid sequence of the large subunit (SEQ ID NO: 6) and small subunit (SEQ ID NO: 7) of a typical wild-type or naturally occurring GT protein are shown in Table 2. In addition, codon optimized DNA sequences encoding the large and small subunits of GT are also provided in Table 2, as SEQ ID NO: 2 and SEQ ID NO: 3, respectively.

Table 2. Guanylyl Transferase

Large subunit (Protein sequence)	MDANVVSSSTIATYIDALAKNASELEQRSTAYEINNELELVFIKPPL ITLTNVVNISTIQESFIRFTVTNKEGVKIRTKIPLSKVHGLDVKNVQL VDAIDNIVWEKKSLVTENRLHKECLLRLSTEERHIFLDYKKYGSSI RLELVNLIQAKTKNFTIDFKLKYFLGSGAQSKSSLLHAINHPKSRPN TSLEIEFTPRDNETVPYDELIKELTTLSRHIFMASPENVILSPPINAPI KTFMLPKQDIVGLDLENLYAVTKTDGIPITIRVTSNGLYCYFTHLG YIIRYPVKRIIDSEVVVFGEAVKDKNWTVYLIKLIEPVNAINDRLEE SKYVESKLVDICDRIVFKSKKYEGPFTTTSEVVDMLSTYLPKQPEG VILFYSKGPKSNIDFKIKKENTIDQTANVVFRYMSSEPIIFGESSIFVE YKKFSNDKGFPKEYGSGKIVLYNGVNYLNNIYCLEYINTHNEVGI KSVVVPIKFIAEFLVNGEILKPRIDKTMKYINSEDYYGNQHNIIVEH LRDQSIKIGDIFNEDKLSDVGHQYANNDKFRLNPEVSYFTNKRTRG PLGILSNYVKTLLISMYCSKTFLDDSNKRKVLAIDFGNGADLEKYF YGEIALLVATDPDADAIARGNERYNKLNSGIKTKYYKFDYIQETIR SDTFVSSVREVFYFGKFNIIDWQFAIHYSFHPRHYATVMNNLSELT ASGGKVLITTMDGDKLSKLTDKKTFIIHKNLPSSENYMSVEKIADD RIVVYNPSTMSTPMTEYIIKKNDIVRVFNEYGFVLVDNVDFATIIER SKKFINGASTMEDRPSTRNFFELNRGAIKCEGLDVEDLLSYYVVY VFSKR **(SEQ ID NO: 6)**
Small subunit (Protein sequence)	MDEIVKNIREGTHVLLPFYETLPELNLSLGKSPLPSLEYGANYFLQI SRVNDLNRMPTDMLKLFTHDIMLPESDLDKVYEILKINSVKYYGR STKADAVVADLSARNKLFKRERDAIKSNNHLTENNLYISDYKMLT FDVFRPLFDFVNEKYCIIKLPTLFGRGVIDTMRIYCSLFKNVRLLKC VSDSWLKDSAIMVASDVCKKNLDLFMSHVKSVTKSSSWKDVNSV QFSILNNPVDTEFINKFLEFSNRVYEALYYVHSLLYSSMTSDSKSIE NKHQRRLVKLLL **(SEQ ID NO: 7)**
Large subunit (DNA sequence)	AGATGGAAGATGAAGATACCATCGACGTCTTTCAGCAACAGAC CGGTGGTATGGATGCTAACGTCGTTAGCAGCAGCACCATTGCG ACTTACATTGATGCACTGGCCAAAAACGCATCTGAGCTTGAGC AGCGCAGCACCGCCTACGAGATCAATAACGAATTGGAGCTGGT TTTCATTAAACCGCCGCTGATCACGCTGACGAACGTCGTGAAC ATTAGCACGATTCAAGAGAGCTTTATTCGTTTCACCGTTACCAA

	TAAAGAAGGCGTGAAGATCCGTACCAAGATTCCGCTGAGCAAA GTGCATGGTCTGGACGTGAAAAATGTGCAGCTGGTTGATGCGA TCGATAACATCGTGTGGGAGAAGAAATCTTTGGTCACGGAAAA TCGTCTGCACAAGGAATGTCTGCTGCGTCTGTCAACCGAAGAA CGCCACATCTTCCTGGACTACAAGAAGTATGGTTCCAGCATCCG TCTGGAACTGGTGAACCTGATTCAGGCAAAGACCAAGAACTTC ACCATTGACTTCAAACTGAAGTATTTCCTGGGCTCTGGTGCACA GAGCAAATCCAGCTTGTTGCACGCGATTAACCATCCGAAGAGC CGTCCGAATACGAGCCTGGAGATCGAATTCACGCCGCGTGATA ACGAAACCGTTCCGTACGATGAGCTGATTAAAGAACTGACGAC GTTGAGCCGCCACATCTTTATGGCCAGCCCGGAAAACGTGATC CTTAGCCCGCCTATCAATGCGCCGATTAAAACCTTTATGTTACC GAAACAAGACATTGTGGGTCTGGACCTGGAAAACCTGTACGCG GTCACCAAAACGGACGGCATTCCGATCACGATTCGTGTTACCA GCAATGGTCTGTACTGCTATTTCACTCATTTGGGCTATATCATT CGTTATCCGGTGAAACGCATCATTGATTCTGAGGTTGTCGTTTT CGGCGAAGCAGTCAAGGACAAGAATTGGACTGTGTACCTGATC AAATTGATTGAACCGGTTAACGCCATCAATGACCGCCTGGAAG AGTCGAAATATGTTGAAAGCAAACTGGTGGATATTTGTGATCG TATCGTGTTCAAGAGCAAGAAATATGAAGGCCCGTTCACCACG ACCAGCGAAGTTGTTGACATGCTGAGCACCTATCTGCCGAAAC AACCTGAGGGTGTGATTCTGTTTTACTCCAAGGGTCCGAAGAG CAACATTGATTTCAAAATCAAGAAAGAGAATACCATTGATCAG ACCGCCAACGTTGTGTTCCGCTATATGTCCAGCGAGCCTATCAT TTTCGGTGAGTCGAGCATCTTTGTTGAATACAAAAAGTTTAGCA ACGATAAGGGTTTTCCGAAAGAATACGGTTCCGGTAAGATTGT GTTGTACAACGGCGTCAATTATCTGAACAACATCTACTGTCTGG AGTACATCAATACCCATAACGAAGTTGGCATTAAGTCTGTTGTC GTCCCGATCAAATTCATCGCGGAGTTCCTGGTTAACGGTGAGAT TCTGAAGCCGCGTATTGATAAAACTATGAAATACATTAACTCC GAAGATTACTACGGTAATCAGCATAACATCATCGTCGAGCACT TGCGTGATCAAAGCATTAAGATCGGTGACATCTTTAACGAAGA TAAGCTGAGCGATGTAGGCCACCAGTATGCGAACAATGACAAA TTTCGCCTGAATCCGGAAGTCAGCTACTTTACGAATAAGCGCAC CCGTGGTCCACTGGGTATCCTGAGCAATTATGTTAAAACCCTGT TGATTTCCATGTACTGCTCCAAAACGTTCCTGGACGACAGCAAC AAGCGCAAAGTTCTGGCGATCGACTTCGGTAATGGTGCCGATC TGGAGAAGTACTTTTATGGTGAGATCGCATTGCTGGTTGCTACC GACCCGGATGCAGATGCGATCGCCCGTGGCAACGAGCGTTACA ATAAGCTGAATAGCGGTATCAAGACCAAATACTACAAATTCGA CTATATTCAAGAGACGATCCGCTCGGACACCTTTGTATCCAGCG TGCGTGAGGTGTTTTACTTCGGTAAATTCAACATCATTGACTGG CAATTCGCCATTCACTATAGCTTTCACCCACGCCACTATGCGAC GGTCATGAACAACCTGTCTGAGCTGACCGCGAGCGGCGGTAAA GTTCTGATCACCACGATGGACGGTGACAAGCTGTCTAAACTGA CCGACAAAAAGACCTTCATTATTCACAAAAATCTCCCGTCGAG CGAGAATTACATGTCCGTCGAAAAGATTGCGGACGACCGTATT GTTGTCTACAACCCGAGCACTATGTCGACCCCAATGACCGAGT ATATCATCAAAAAGAATGACATTGTGCGTGTCTTTAATGAATAC GGTTTTGTGCTGGTCGACAACGTCGATTTTGCGACCATCATCGA

	GAGAAGCAAGAAATTCATTAATGGCGCTTCTACGATGGAAGAT CGCCCGAGCACGCGTAACTTCTTTGAGCTGAATCGTGGCGCGA TTAAGTGCGAGGGCCTGGACGTCGAGGATCTGCTGTCGTATTA CGTGGTTTATGTGTTTAGCAAACGTTAATGA **(SEQ ID NO: 2)**
Small subunit (DNA sequence)	ATGGACGAAATTGTCAAGAATATCCGTGAAGGTACCCACGTTT TACTGCCATTCTACGAGACGCTGCCGGAACTGAACCTGAGCCT GGGTAAAAGCCCTCTGCCGAGCCTGGAGTATGGTGCGAACTAT TTTCTGCAGATTTCCCGTGTAAACGATTTGAACCGCATGCCGAC GGACATGCTGAAACTGTTCACCCACGACATCATGCTGCCGGAA TCTGATCTGGATAAAGTTTACGAGATCTTGAAAATCAATTCAGT GAAGTACTATGGCCGTAGCACCAAGGCCGATGCGGTGGTCGCA GACCTGAGCGCGCGTAACAAACTGTTTAAACGTGAACGTGACG CAATTAAGAGCAATAACCATCTGACCGAGAACAATTTGTACAT CAGCGACTACAAGATGTTGACTTTTGACGTGTTTCGTCCGCTGT TCGACTTTGTTAATGAGAAATACTGCATTATCAAGCTGCCGACG TTGTTTGGTCGCGGCGTCATTGATACGATGCGCATTTACTGCTC TCTCTTCAAGAATGTGCGCCTGCTGAAGTGTGTCTCCGACAGCT GGCTGAAAGATAGCGCTATTATGGTTGCGAGCGACGTGTGTAA AAAGAACCTGGATCTGTTCATGAGCCACGTGAAGAGCGTTACC AAAAGCAGCAGCTGGAAAGACGTTAACAGCGTCCAGTTCTCCA TTCTGAATAACCCGGTCGATACCGAGTTTATCAACAAGTTCCTT GAATTCAGCAATCGCGTTTATGAGGCCCTGTATTACGTTCATAG CCTGCTGTATAGCTCCATGACCTCTGATAGCAAATCGATCGAGA ATAAACACCAACGTCGTCTGGTGAAACTGCTGCTGTAATGA **(SEQ ID NO: 3)**

[0053] Thus, in some embodiments, a GT enzyme is a heterodimer comprising large and small subunits (SEQ ID NO: 6 and SEQ ID NO: 7, respectively). In some embodiments, the GT enzyme of the invention may be a homologue or analogue of one or the other of the GT large and small subunits. For example, a homologue or analogue of GT protein may be a modified GT protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to SEQ ID NO: 6 and/or SEQ ID NO: 7, while retaining substantial GT protein activity. Thus, in some embodiments, an enzyme suitable for the present invention is substantially homologous to the GT protein large and small subunits (SEQ ID NO: 6 and SEQ ID NO: 7). In some embodiments, an enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 6. In some embodiments, an enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 7. In some embodiments, an

enzyme suitable for the present invention is substantially identical to the large and small subunits of GT (SEQ ID NO: 6 and SEQ ID NO: 7). In some embodiments, an enzyme suitable for the present invention contains a fragment or a portion of a GT protein.

[0054] In some embodiments, the GT protein is encoded by a nucleic acid derived from an virus selected from the group consisting of Vaccinia virus, Rabbitpox virus, Cowpox virus, Taterapox virus, Monkeypox virus, Variola major virus, Camelpox virus, Ectromelia virus, Variola minor virus, Orthopox virus, Raccoonpox virus, Skunkpox virus, Volepox virus, Yoka pox virus, Swinepox virus, Yaba monkey tumor virus, Deerpox virus, Myxoma virus, Tanapox virus, Goatpox virus, Rabbit fibroma virus, Lumpy skin disease virus, Sheeppox virus, Eptesipox virus, Squirrelpox virus, Molluscum contagiosum virus, Cotia virus, Orf virus, Bovine popular stomatitis virus, Pseudocowpox virus, Canarypox virus, Pidgeonpox virus, Penguinpox virus, and Fowlpox virus. In some embodiments, nucleic acids suitable for the present invention have a sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 2. In some embodiments, nucleic acids suitable for the present invention have a sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 3. In some embodiments, nucleic acids suitable for the present invention are substantially identical to a nucleic acid encoding a GT protein (SEQ ID NO: 2 and SEQ ID NO: 3).

SUMO-GT Fusion

[0055] As used herein, a SUMO-GT fusion protein is any protein or portion of a protein that comprises a SUMO protein covalently linked to a Guanylyl Transferase (GT) protein, wherein the fusion protein can substitute for at least partial activity of naturally-occurring Guanylyl Transferase (GT) protein. As used herein, the terms "a SUMO-GT fusion protein" and "a SUMO-GT fusion enzyme" and grammatical equivalents are used interchangeably. An exemplary amino acid sequence of the fusion of SUMO and the GT large subunit (SEQ ID NO: 8) are shown in Table 3. In addition, an exemplary DNA sequence encoding the fusion of SUMO and the GT large subunit is also provided in Table 3, as SEQ ID NO: 4.

Table 3. SUMO-GT Fusion

SUMO-GT large subunit DNA construct with His	ATGGGCCATCATCATCACCATCACGGCAGCCTGCAAGAAGAGA AACCGAAAGAGGGCGTTAAGACCGAGAATGACCACATTAACCT GAAGGTCGCTGGTCAAGATGGCAGCGTGGTGCAGTTTAAGATC

tag and linker	AAGCGTCACACGCCGTTGAGCAAGCTGATGAAGGCTTACTGCG AGCGTCAGGGTCTGAGCATGCGTCAGATCCGCTTTCGTTTCGAT GGCCAGCCGATCAATGAGACTGACACCCCAGCGCAACTGGAGA TGGAAGATGAAGATACCATCGACGTCTTTCAGCAACAGACCGG TGGTATGGATGCTAACGTCGTTAGCAGCAGCACCATTGCGACTT ACATTGATGCACTGGCCAAAAACGCATCTGAGCTTGAGCAGCG CAGCACCGCCTACGAGATCAATAACGAATTGGAGCTGGTTTTC ATTAAACCGCCGCTGATCACGCTGACGAACGTCGTGAACATTA GCACGATTCAAGAGAGCTTTATTCGTTTCACCGTTACCAATAAA GAAGGCGTGAAGATCCGTACCAAGATTCCGCTGAGCAAAGTGC ATGGTCTGGACGTGAAAAATGTGCAGCTGGTTGATGCGATCGA TAACATCGTGTGGGAGAAGAAATCTTTGGTCACGGAAAATCGT CTGCACAAGGAATGTCTGCTGCGTCTGTCAACCGAAGAACGCC ACATCTTCCTGGACTACAAGAAGTATGGTTCCAGCATCCGTCTG GAACTGGTGAACCTGATTCAGGCAAAGACCAAGAACTTCACCA TTGACTTCAAACTGAAGTATTTCCTGGGCTCTGGTGCACAGAGC AAATCCAGCTTGTTGCACGCGATTAACCATCCGAAGAGCCGTC CGAATACGAGCCTGGAGATCGAATTCACGCCGCGTGATAACGA AACCGTTCCGTACGATGAGCTGATTAAAGAACTGACGACGTTG AGCCGCCACATCTTTATGGCCAGCCCGGAAAACGTGATCCTTA GCCCGCCTATCAATGCGCCGATTAAAACCTTTATGTTACCGAAA CAAGACATTGTGGGTCTGGACCTGGAAAACCTGTACGCGGTCA CCAAAACGGACGGCATTCCGATCACGATTCGTGTTACCAGCAA TGGTCTGTACTGCTATTTCACTCATTTGGGCTATATCATTCGTTA TCCGGTGAAACGCATCATTGATTCTGAGGTTGTCGTTTTCGGCG AAGCAGTCAAGGACAAGAATTGGACTGTGTACCTGATCAAATT GATTGAACCGGTTAACGCCATCAATGACCGCCTGGAAGAGTCG AAATATGTTGAAAGCAAACTGGTGGATATTTGTGATCGTATCGT GTTCAAGAGCAAGAAATATGAAGGCCCGTTCACCACGACCAGC GAAGTTGTTGACATGCTGAGCACCTATCTGCCGAAACAACCTG AGGGTGTGATTCTGTTTTACTCCAAGGGTCCGAAGAGCAACATT GATTTCAAAATCAAGAAAGAGAATACCATTGATCAGACCGCCA ACGTTGTGTTCCGCTATATGTCCAGCGAGCCTATCATTTTCGGT GAGTCGAGCATCTTTGTTGAATACAAAAAGTTTAGCAACGATA AGGGTTTTCCGAAAGAATACGGTTCCGGTAAGATTGTGTTGTAC AACGGCGTCAATTATCTGAACAACATCTACTGTCTGGAGTACAT CAATACCCATAACGAAGTTGGCATTAAGTCTGTTGTCGTCCCGA TCAAATTCATCGCGGAGTTCCTGGTTAACGGTGAGATTCTGAAG CCGCGTATTGATAAAACTATGAAATACATTAACTCCGAAGATT ACTACGGTAATCAGCATAACATCATCGTCGAGCACTTGCGTGA TCAAAGCATTAAGATCGGTGACATCTTTAACGAAGATAAGCTG AGCGATGTAGGCCACCAGTATGCGAACAATGACAAATTTCGCC TGAATCCGGAAGTCAGCTACTTTACGAATAAGCGCACCCGTGG TCCACTGGGTATCCTGAGCAATTATGTTAAAACCCTGTTGATTT CCATGTACTGCTCCAAAACGTTCCTGGACGACAGCAACAAGCG CAAAGTTCTGGCGATCGACTTCGGTAATGGTGCCGATCTGGAG AAGTACTTTTATGGTGAGATCGCATTGCTGGTTGCTACCGACCC GGATGCAGATGCGATCGCCCGTGGCAACGAGCGTTACAATAAG CTGAATAGCGGTATCAAGACCAAATACTACAAATTCGACTATA TTCAAGAGACGATCCGCTCGGACACCTTTGTATCCAGCGTGCGT

	GAGGTGTTTTACTTCGGTAAATTCAACATCATTGACTGGCAATT CGCCATTCACTATAGCTTTCACCCACGCCACTATGCGACGGTCA TGAACAACCTGTCTGAGCTGACCGCGAGCGGCGGTAAAGTTCT GATCACCACGATGGACGGTGACAAGCTGTCTAAACTGACCGAC AAAAAGACCTTCATTATTCACAAAAATCTCCCGTCGAGCGAGA ATTACATGTCCGTCGAAAAGATTGCGGACGACCGTATTGTTGTC TACAACCCGAGCACTATGTCGACCCCAATGACCGAGTATATCA TCAAAAAGAATGACATTGTGCGTGTCTTTAATGAATACGGTTTT GTGCTGGTCGACAACGTCGATTTTGCGACCATCATCGAGAGAA GCAAGAAATTCATTAATGGCGCTTCTACGATGGAAGATCGCCC GAGCACGCGTAACTTCTTTGAGCTGAATCGTGGCGCGATTAAG TGCGAGGGCCTGGACGTCGAGGATCTGCTGTCGTATTACGTGG TTTATGTGTTTAGCAAACGTTAATGA **(SEQ ID NO: 4)**
SUMO-GT large subunit protein with His tag and linker	MGHHHHHHGSLQEEKPKEGVKTENDHINLKVAGQDGSVVQFKIK RHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMED EDTIDVFQQQTGGMDANVVSSSTIATYIDALAKNASELEQRSTAY EINNELELVFIKPPLITLTNVVNISTIQESFIRFTVTNKEGVKIRTKIPL SKVHGLDVKNVQLVDAIDNIVWEKKSLVTENRLHKECLLRLSTEE RHIFLDYKKYGSSIRLELVNLIQAKTKNFTIDFKLKYFLGSGAQSKS SLLHAINHPKSRPNTSLEIEFTPRDNETVPYDELIKELTTLSRHIFMA SPENVILSPPINAPIKTFMLPKQDIVGLDLENLYAVTKTDGIPITIRV TSNGLYCYFTHLGYIIRYPVKRIIDSEVVVFGEAVKDKNWTVYLIK LIEPVNAINDRLEESKYVESKLVDICDRIVFKSKKYEGPFTTTSEVV DMLSTYLPKQPEGVILFYSKGPKSNIDFKIKKENTIDQTANVVFRY MSSEPIIFGESSIFVEYKKFSNDKGFPKEYGSGKIVLYNGVNYLNNI YCLEYINTHNEVGIKSVVVPIKFIAEFLVNGEILKPRIDKTMKYINSE DYYGNQHNIIVEHLRDQSIKIGDIFNEDKLSDVGHQYANNDKFRL NPEVSYFTNKRTRGPLGILSNYVKTLLISMYCSKTFLDDSNKRKVL AIDFGNGADLEKYFYGEIALLVATDPDADAIARGNERYNKLNSGI KTKYYKFDYIQETIRSDTFVSSVREVFYFGKFNIIDWQFAIHYSFHP RHYATVMNNLSELTASGGKVLITTMDGDKLSKLTDKKTFIIHKNL PSSENYMSVEKIADDRIVVYNPSTMSTPMTEYIIKKNDIVRVFNEY GFVLVDNVDFATIIERSKKFINGASTMEDRPSTRNFFELNRGAIKCE GLDVEDLLSYYVVYVFSKR **(SEQ ID NO: 8)**

[0056] In some embodiments, the SUMO-GT fusion protein comprises SEQ ID NO: 8. In some embodiments, the SUMO-GT fusion protein is a heterodimer comprising SEQ ID NO: 8 and SEQ ID NO: 7. In some embodiments, the GT enzyme of the invention may be a homologue or analogue of one or the other of the GT large and small subunits. For example, a homologue or analogue of the SUMO-GT fusion protein may be a modified SUMO-GT fusion protein containing one or more amino acid substitutions, deletions, and/or insertions as compared to SEQ ID NO: 8 and/or SEQ ID NO: 7, while retaining substantial GT protein activity. Thus, in some embodiments, a SUMO-GT fusion protein suitable for the present

invention is substantially homologous to the heterodimer comprising the GT small subunit (SEQ ID NO: 7) and the fusion of SUMO and the GT large subunit (SEQ ID NO: 8). In some embodiments, an enzyme suitable for the present invention has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 8 and SEQ ID NO: 7. In some embodiments, an enzyme suitable for the present invention is substantially identical to the heterodimer comprising the GT small subunit (SEQ ID NO: 7) and the fusion of SUMO and the GT large subunit (SEQ ID NO: 8). In some embodiments, an enzyme suitable for the present invention contains a fragment or a portion of a GT protein covalently bound to a SUMO protein.

Production of SUMO-GT Fusion Protein

Host Cells

[0057] As used herein, the term "host cells" refers to cells that can be used to produce a SUMO-GT fusion protein. In particular, host cells are suitable for producing a SUMO-GT fusion protein at a large scale. In some embodiments, host cells are able to produce SUMO-GT fusion protein in an amount of or greater than about 5 picogram/cell/day (e.g., greater than about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 picogram/cell/day). In some embodiments, host cells are able to produce SUMO-GT fusion protein in an amount ranging from about 5-100 picogram/cell/day (e.g., about 5-90 picogram/cell/day, about 5-80 picogram/cell/day, about 5-70 picogram/cell/day, about 5-60 picogram/cell/day, about 5-50 picogram/cell/day, about 5-40 picogram/cell/day, about 5-30 picogram/cell/day, about 10-90 picogram/cell/day, about 10-80 picogram/cell/day, about 10-70 picogram/cell/day, about 10-60 picogram/cell/day, about 10-50 picogram/cell/day, about 10-40 picogram/cell/day, about 10-30 picogram/cell/day, about 20-90 picogram/cell/day, about 20-80 picogram/cell/day, about 20-70 picogram/cell/day, about 20-60 picogram/cell/day, about 20-50 picogram/cell/day, about 20-40 picogram/cell/day, about 20-30 picogram/cell/day).

[0058] Suitable host cells can be derived from a variety of organisms, including, but not limited to, bacteria, yeast, insects, plants, birds (e.g., avian systems), amphibians, and mammals. In some embodiments, host cells are non-mammalian cells. Non-limiting

examples of non-mammalian host cells suitable for the present invention include cells and cell lines derived from *Escherichia coli*, *Salmonella typhimurium*, *Bacillus subtilis*, *Bacillus lichenifonnis*, *Bacteroides fragilis*, *Clostridia perfringens*, *Clostridia difficile* for bacteria; *Pichia pastoris*, *Pichia methanolica*, *Pichia angusta, Schizosacccharomyces pombe*, *Saccharomyces cerevisiae,* and *Yarrowia lipolytica* for yeast; *Sodoptera frugiperda*, *Trichoplusis ni*, *Drosophila melangoster* and *Manduca sexta* for insects; and and *Xenopus Laevis* from amphibian.

[0059] In some embodiments, host cells are mammalian cells. Any mammalian cell susceptible to cell culture, and to expression of polypeptides, may be utilized in accordance with the present invention as a host cell. Non-limiting examples of mammalian cells that may be used in accordance with the present invention include human embryonic kidney 293 cells (HEK293), HeLa cells; BALB/c mouse myeloma line (NSO/l, ECACC No: 85110503); human retinoblasts (PER.C6 (CruCell, Leiden, The Netherlands)); monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human fibrosarcomacell line (e.g., HT-1080); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham et al., J. Gen Virol., 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells +/-DHFR (CHO, Urlaub and Chasin, Proc. Natl. Acad. Sci. USA, 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod., 23:243-251 (1980)); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1 587); human cervical carcinoma cells (HeLa, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI cells (Mather et al., Annals N.Y. Acad. Sci., 383:44-68 (1982)); MRC 5 cells; FS4 cells; a human hepatoma line (Hep G2), human cell line CAP and AGE1.HN, and Glycotope's panel.

[0060] Additionally, any number of available hybridoma cell lines may be utilized in accordance with the present invention. One skilled in the art will appreciate that hybridoma cell lines might have different nutrition requirements and/or might require different culture conditions for optimal growth and polypeptide or protein expression, and will be able to modify conditions as needed.

Expression Vectors

[0061] Various nucleic acid constructs can be used to express SUMO-GT fusion protein described herein in host cells. A suitable vector construct typically includes, in addition to SUMO-GT fusion protein-encoding sequences (also referred to as SUMO-GT fusion transgene), regulatory sequences, gene control sequences, promoters, non-coding sequences and/or other appropriate sequences for expression of the protein and, optionally, for replication of the construct. Typically, the coding region is operably linked with one or more of these nucleic acid components.

[0062] "Regulatory sequences" typically refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, enhancers, 5' untranslated sequences, translation leader sequences, introns, and 3' untranslated sequences such as polyadenylation recognition sequences. Sometimes, "regulatory sequences" are also referred to as "gene control sequences."

[0063] "Promoter" typically refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence that can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions.

[0064] The "3' non-coding sequences" typically refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

[0065] The "translation leader sequence" or "5' non-coding sequences" typically refers to a nucleotide sequence located between the promoter sequence of a gene and the

coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

[0066] Typically, the term "operatively linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operatively linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operatively linked to regulatory sequences in sense or antisense orientation.

[0067] The coding region of a transgene may include one or more silent mutations to optimize codon usage for a particular cell type. For example, the codons of an SUMO-GT fusion transgene may be optimized for expression in a bacterial cell. In some embodiments, the codons of an SUMO-GT fusion transgene may be optimized for expression in an *E. coli* cell. In some embodiments, the codons of an SUMO-GT fusion transgene may be optimized for expression in a mammalian cell. In some embodiments, the codons of an SUMO-GT fusion transgene may be optimized for expression in a human cell.

[0068] Optionally, a construct may contain additional components such as one or more of the following: a splice site, an enhancer sequence, a selectable marker gene under the control of an appropriate promoter, an amplifiable marker gene under the control of an appropriate promoter, and a matrix attachment region (MAR) or other element known in the art that enhances expression of the region where it is inserted.

[0069] Once transfected or transduced into host cells, a suitable vector can express extrachromosomally (episomally) or integrate into the host cell's genome.

[0070] In some embodiments, a DNA construct that integrates into the cell's genome, it need include only the transgene nucleic acid sequences. In that case, the express of the transgene is typically controlled by the regulatory sequences at the integration site. Optionally, it can include additional various regulatory sequences described herein.

Culture Medium and Conditions

[0071] The term "medium" and "culture medium" as used herein refers to a general class of solution containing nutrients suitable for maintaining and/or growing cells *in vitro*. Typically, medium solutions provide, without limitation, essential and nonessential amino acids, vitamins, energy sources, lipids, and trace elements required by the cell for at least

minimal growth and/or survival. In other embodiments, the medium may contain an amino acid(s) derived from any source or method known in the art, including, but not limited to, an amino acid(s) derived either from single amino acid addition(s) or from a peptone or protein hydrolysate addition(s) (including animal or plant source(s)). Vitamins such as, but not limited to, Biotin, Pantothenate, Choline Chloride, Folic Acid, Myo-Inositol, Niacinamide, Pyridoxine, Riboflavin, Vitamin B12, Thiamine, Putrescine and/or combinations thereof. Salts such as, but not limited to, $CaCl_2$, KCl, $MgCl_2$, NaCl, Sodium Phosphate Monobasic, Sodium Phosphate Dibasic, Sodium Selenite, $CuSO_4$, $ZnCl_2$ and/or combinations thereof. Fatty acids such as, but not limited to, Arachidonic Acid, Linoleic Acid, Oleic Acid, Lauric Acid, Myristic Acid, as well as Methyl-beta-Cyclodextrin and/or combinations thereof). In some embodiments, medium comprises additional components such as glucose, glutamine, Na-pyruvate, insulin or ethanolamine, a protective agent such as Pluronic F68. In some embodiments, the medium may also contain components that enhance growth and/or survival above the minimal rate, including hormones and growth factors. Medium may also comprise one or more buffering agents. The buffering agents may be designed and/or selected to maintain the culture at a particular pH (*e.g.*, a physiological pH, (*e.g.*, pH 6.8 to pH 7.4)). A variety of buffers suitable for culturing cells are known in the art and may be used in the methods. Suitable buffers (*e.g.*, bicarbonate buffers, HEPES buffer, Good's buffers, *etc.*) are those that have the capacity and efficiency for maintaining physiological pH despite changes in carbon dioxide concentration associated with cellular respiration. The solution is preferably formulated to a pH and salt concentration optimal for cell survival and proliferation.

[0072] In some embodiments, medium may be a chemically defined medium. As used herein, the term "chemically-defined nutrient medium" refers to a medium of which substantially all of the chemical components are known. In some embodiments, a chemically defined nutrient medium is free of animal-derived components. In some cases, a chemically-defined medium comprises one or more proteins (e.g., protein growth factors or cytokines.) In some cases, a chemically-defined nutrient medium comprises one or more protein hydrolysates. In other cases, a chemically-defined nutrient medium is a protein-free media, i.e., a serum-free media that contains no proteins, hydrolysates or components of unknown composition.

[0073] Typically, a chemically defined medium can be prepared by combining various individual components such as, for example, essential and nonessential amino acids,

vitamins, energy sources, lipids, salts, buffering agents, and trace elements, at predetermined weight or molar percentages or ratios. Exemplary serum-free, in particular, chemically-defined media are described in US Pub. No. 2006/0148074, the disclosure of which is hereby incorporated by reference.

[0074] In some embodiments, a chemically defined medium suitable for the present invention is a commercially available medium such as, but not limited to, Terrific Broth, Cinnabar, 2xYT or LB. In some embodiments, a chemically defined medium suitable for the present invention is a mixture of one or more commercially available chemically defined mediums. In various embodiments, a suitable medium is a mixture of two, three, four, five, six, seven, eight, nine, ten, or more commercially available chemically defined media. In some embodiments, each individual commercially available chemically defined medium (e.g., such as those described herein) constitutes, by weight, 1%, 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or more, of the mixture. Ratios between each individual component medium may be determined by relative weight percentage present in the mixture. In some embodiments, protein expression is increased with the addition of IPTG to repress the promoter.

[0075] In some embodiments, a chemically defined medium may be supplemented by one or more animal derived components. Such animal derived components include, but are not limited to, fetal calf serum, horse serum, goat serum, donkey serum, human serum, and serum derived proteins such as albumins (e.g., bovine serum albumin or human serum albumin).

[0076] The present invention provides a method of producing SUMO-GT fusion protein at a large scale. Typical large-scale procedures for producing a fusion polypeptide of interest include batch cultures and fed-batch cultures. Batch culture processes traditionally comprise inoculating a large-scale production culture with a seed culture of a particular cell density, growing the cells under conditions (e.g., suitable culture medium, pH, and temperature) conducive to cell growth, viability, and/or productivity, harvesting the culture when the cells reach a specified cell density, and purifying the expressed polypeptide. Fed-batch culture procedures include an additional step or steps of supplementing the batch culture with nutrients and other components that are consumed during the growth of the cells. In some embodiments, a large-scale production method according to the present invention uses a fed-batch culture system.

Purification of Expressed SUMO-GT Fusion Protein

[0077] Various methods may be used to purify or isolate SUMO-GT fusion protein produced according to various methods described herein. In some embodiments, the expressed SUMO-GT fusion protein is secreted into the medium and thus cells and other solids may be removed, as by centrifugation or filtering for example, as a first step in the purification process. Alternatively or additionally, the expressed SUMO-GT fusion protein is bound to the surface of the host cell. In this embodiment, the host cells (for example, bacterials cells) expressing the polypeptide or protein are lysed for purification. Lysis of host cells (e.g., bacterials cells) can be achieved by any number of means well known to those of ordinary skill in the art, including physical disruption by glass beads and exposure to high pH conditions.

[0078] The SUMO-GT fusion protein may be isolated and purified by standard methods including, but not limited to, chromatography (e.g., ion exchange, affinity, size exclusion, and hydroxyapatite chromatography), gel filtration, centrifugation, or differential solubility, ethanol precipitation or by any other available technique for the purification of proteins (See, e.g., Scopes, Protein Purification Principles and Practice 2nd Edition, Springer-Verlag, New York, 1987; Higgins, S. J. and Hames, B. D. (eds.), Protein Expression: A Practical Approach, Oxford Univ Press, 1999; and Deutscher, M. P., Simon, M. I., Abelson, J. N. (eds.), Guide to Protein Purification: Methods in Enzymology (Methods in Enzymology Series, Vol 182), Academic Press, 1997, all incorporated herein by reference). For immunoaffinity chromatography in particular, the protein may be isolated by binding it to an affinity column comprising antibodies that were raised against that protein and were affixed to a stationary support. Protease inhibitors such as phenyl methyl sulfonyl fluoride (PMSF), leupeptin, pepstatin or aprotinin may be added at any or all stages in order to reduce or eliminate degradation of the polypeptide or protein during the purification process. Protease inhibitors are particularly desired when cells must be lysed in order to isolate and purify the expressed polypeptide or protein.

Solubility

[0079] Various methods may be used to determine the solubility of a protein in an expression system. In an exemplary method, bacteria are spun down and resuspended in a mild lysis buffer containing 1% IGEPAL and protease inhibitors. Lysis is supported by repeated freezing and thawing the bacteria. Soluble and insoluble fraction are separated by

centrifigation. To determine the total amount of recombinant protein the same volume of bacterial culture is spun down and lysed in the same amount of lysis buffer containing 1% IGEPAL and 0.1% SDS. Soluble and total protein are analyzed by SDS-PAGE, with western blotting if necessary. In some embodiments, the expression system is *E. coli*. In some embodiments, solubility of GT is improved when it has been produced as a fusion protein. In some embodiments, the fusion protein is a SUMO-GT fusion protein. In some embodiments, the SUMO-GT fusion protein has increased solubility compared to the non-fusion GT protein. In some embodiments, the increased solubility of the SUMO-GT fusion protein compared to the non-fusion GT protein is observed during shake flask production of the SUMO-GT fusion protein. In some embodiments, the increased solubility of the SUMO-GT fusion protein compared to the non-fusion GT protein is observed during fermentation production of the SUMO-GT fusion protein.

Use of SUMO-GT Fusion in mRNA Capping

Production of capped mRNAs

[0080] According to the present invention, a SUMO-GT fusion protein described herein may be used to produce capped mRNAs by *in vitro* transcription. Various in vitro transcription assays are available in the art and can be used to practice the present invention. For example, *in vitro* transcription was originally developed by Krieg and Melton (METHODS ENZYMOL., 1987, 155: 397-415) for the synthesis of RNA using an RNA phage polymerase. Typically these reactions include at least a phage RNA polymerase (T7, T3 or SP6), a DNA template containing a phage polymerase promoter, nucleotides (ATP, CTP, GTP and UTP), and a buffer containing a magnesium salt. RNA synthesis yields may be optimized by increasing nucleotide concentrations, adjusting magnesium concentrations and by including inorganic pyrophosphatase (U.S. Pat. No. 5,256,555; Gurevich, et al., ANAL. BIOCHEM. 195: 207-213 (1991); Sampson, J.R. and Uhlenbeck, O.C., PROC. NATL. ACAD. SCI. USA. 85, 1033-1037 (1988); Wyatt, J.R., et al., BIOTECHNIQUES, 11: 764-769 (1991)). The RNA synthesized in these reactions is usually characterized by a 5' terminal nucleotide that has a triphosphate at the 5' position of the ribose. Typically, depending on the RNA polymerase and promoter combination used, this nucleotide is a guanosine, although it can be an adenosine (see e.g., Coleman, T. M., et al., NUCLEIC ACIDS RES., 32: e14 (2004)). In these

reactions, all four nucleotides are typically included at equimolar concentrations and none of them is limiting.

[0081] Some embodiment of the invention are batch reactions—that is, all components are combined and then incubated at about 37 °C to promote the polymerization of the RNA until the reaction terminates. Typically, a batch reaction is used for convenience and to obtain as much RNA as needed from such reactions for their experiments. In some embodiments, a "fed-batch" system (see, e.g., JEFFREY A. KERN, BATCH AND FED-BATCH STRATEGIES FOR LARGE-SCALE PRODUCTION OF RNA BY IN VITRO TRANSACTION (University of Colorado) (1997)) is used to increase the efficiency of the *in vitro* transcription reaction. All components are combined, but then additional amounts of some of the reagents are added over time, such as the nucleotides and magnesium, to try to maintain constant reaction conditions. In addition, in some embodiments, the pH of the reaction may be held at 7.4 by monitoring it over time and adding KOH as needed.

[0082] To synthesize a capped RNA by *in vitro* transcription, a cap analog (e.g., N-7 methyl GpppG; i.e., m^7GpppG) is included in the transcription reaction. In some embodiments, the cap analog will be incorporated at the 5' terminus by the enzyme guanylyl transferase. In some embodiments, the guanylyl transferase is a fusion protein. In some embodiments, the guanylyl transferase fusion protein formed when a guanylyl transferase is covalently linked to a SUMO protein. In some embodiments, the cap analog will be incorporated only at the 5' terminus because it does not have a 5' triphosphate. In some embodiments using a T7, T3 and SP6 RNA polymerase, the +1 nucleotide of their respective promoters is usually a G residue and if both GTP and m^7GpppG are present in equal concentrations in the transcription reaction, then they each have an equal chance of being incorporated at the +1 position. In some embodiments, m^7GpppG is present in these reactions at several-fold higher concentrations than the GTP to increase the chances that a transcript will have a 5' cap. In some embodiments, a mMESSAGE mMACHINE® kit (Cat. #1344, Ambion, Inc.) is used according to manufacturer's instructions, where it is recommended that the cap to GTP ratio be 4:1 (6 mM: 1.5 mM). In some embodiments, as the ratio of the cap analog to GTP increases in the reaction, the ratio of capped to uncapped RNA increases proportionally. Considerations of capping efficiency must be balanced with considerations of yield. Increasing the ratio of cap analog to GTP in the transcription reaction produces lower yields of total RNA because the concentration of GTP becomes limiting when holding the total concentration of cap and GTP constant. Thus, the final RNA

yield is dependent on GTP concentration, which is necessary for the elongation of the transcript. The other nucleotides (ATP, CTP, UTP) are present in excess.

[0083] In particular embodiments, mRNA are synthesized by *in vitro* transcription from a plasmid DNA template encoding a gene of choice. In some embodiments, *in vitro* transcription includes addition of a 5' cap structure, Cap1 (FIG. 1B), which has a 2'-O-methyl residue at the 2' OH group of the ribose ring of base 1, by enzymatic conjugation of GTP via a guanylyl transferase. In some embodiments, *in vitro* transcription includes addition of a 5' cap structure, Cap0 (FIG. 1A), which lacks the 2'-O-methyl residue, by enzymatic conjugation of GTP via a guanylyl transferase. In some embodiments, *in vitro* transcription includes addition of a 5' cap of any of the cap structures disclosed herein by enzymatic conjugation of GTP via a guanylyl transferase.

Capping efficiency

[0084] The present invention significantly increases capping efficiency. In some embodiments, the use of a SUMO-GT fusion protein in an *in vitro* capping assay results in at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% capped mRNA. In some embodiments, the use of a SUMO-GT fusion protein in an *in vitro* capping assay results in substantially 100% capped mRNA. In some embodiments, the use of a SUMO-GT fusion protein in an *in vitro* capping assay results in increase of mRNA capping efficiency by at least about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 1-fold, 1.5-fold, 2-fold, 2.5-fold, 3-fold, 3.5-fold, 4-fold, 4.5-fold, or 5-fold as compared to a control assay using a non-fusion GT protein but under otherwise identical conditions.

[0085] In addition, the present invention permits large-scale production of capped mRNA with high efficiency. In some embodiments, capped mRNA is produced at a scale of or greater than 1 gram, 5 grams, 10 grams, 15 grams, 20 grams, 25 grams, 30 grams, 35 grams, 40 grams, 45 grams, 50 grams, 75 grams, 100 grams, 150 grams, 200 grams, 250 grams, 300 grams, 350 grams, 400 grams, 450 grams, 500 grams, 550 grams, 600 grams, 650 grams, 700 grams, 750 grams, 800 grams, 850 grams, 900 grams, 950 grams, 1 kg, 2.5 kg, 5 kg, 7.5 kg, 10 kg, 25 kg, 50 kg, 75 kg, or 100 kg per batch.

[0086] Methods of estimating capping efficiency are known in the art. For example, the T7 RNA polymerase can be incubated with a cap dinucleotide, all four ribonucleotide triphosphates, [α-^{32}P]GTP, and a short DNA template in which G is the first ribonucleotide specified after the promoter (see Grudzien, E. et al. "*Novel cap analogs for in vitro synthesis*

of mRNA with high translation efficiency", RNA, 10: 1479-1487 (2004)). Any nucleotide on the 5′ side of a G residue acquires a ^{32}P-labeled 3′-phosphate group after RNase T2 digestion by nearest-neighbor transfer. Anion exchange chromatography is then used to resolve labeled nucleoside 3′-monophosphates, resulting from internal positions in the RNA, from 5′-terminal products. 5′-terminal products are of two types. Uncapped RNAs yield labeled guanosine 5′-triphosphate 3′-monophosphate (p3Gp*; in which * indicates the labeled phosphate group). Capped RNAs yield various 5′-terminal structures, depending on the nature of the cap analog used (m^7Gp3Gp* and Gp3m^7Gp* when the cap analog is m^7Gp3G).

[0087] Improved methods of directly quantitating mRNA capping efficiency in a sample (e.g., a representative aliquot sample from an *in vitro* synthesis reaction) are provided in WO 2014/152673, which is incorporated herein by reference. Some embodiments comprise the use of a cap specific binding substance under conditions that permit the formation of a complex between the cap specific binding substance and the capped mRNA. The formation of a complex between the cap specific binding substance and the capped mRNA allows quantitative determination of the amount of the complex (i.e., capped mRNAs) relative to a positive control of capped products or negative control of uncapped products. In other words, binding indicates the amount of capped mRNA targets in the sample and the capping efficiency in a reaction from which the sample is derived. Thus, in some embodiments, the step of quantitatively determining the amount of the complex comprises performing an ELISA-type assay wherein the cap specific binding substance is an antibody or other protein that specifically binds an mRNA cap. Complex formation can be quantified by addition of a detection agent specific for the cap specific binding substance (e.g., a goat anti-mouse antibody that binds a mouse anti-m^7G antibody) and which produces a signal directly proportional to the amount of capped mRNA. Embodiments of the invention may be used to quantify the capping efficiency of a wide variety of RNA species, including *in vitro* transcribed mRNA, isolated eukaryotic mRNA, and viral RNA.

[0088] Additional improved methods of directly quantitating mRNA capping efficiency in a sample (e.g., a representative aliquot sample from an *in vitro* synthesis reaction) are provided in WO 2014/152659, which is incorporated herein by reference. Some embodiments of the invention comprise chromatographic methods of quantitating mRNA capping efficiency. These methods are based in part on the insights that the versatility of enzymatic manipulation can be used to increase the resolution of chromatographic separation of polynucleotides. Thus, by amplifying the power of chromatographic separation through

enzymatic manipulation, embodiments of the invention increase the efficiency, quality and throughput of quantitation. For example, not only can the chromatographic methods described herein quantitate capping efficiency, they can also provide information on the modification of the cap (e.g., methylation status at particular cap positions). Thus, embodiments of the invention can simultaneously quantitate capping efficiency and the efficiency of cap modification (e.g., methlylation efficiency). This quantification provides important characterization of an mRNA drug product that has significant impact on the protein production.

[0089] The invention will be more fully understood by reference to the following examples. They should not, however, be construed as limiting the scope of the invention. All literature citations are incorporated by reference.

EXAMPLES

Example 1: SUMO-GT Construct Design

[0090] A new construct incorporating a small ubiquitin-like modifier (SUMO) tag covalently linked and co-expressed with the large subunit faction of a guanylyl transferase (GT) heterodimer was synthesized.

Small ubiquitin-like modifier (SUMO) DNA:

GAAGAGAAACCGAAAGAGGGCGTTAAGACCGAGAATGACCACATTAACCTGAA
GGTCGCTGGTCAAGATGGCAGCGTGGTGCAGTTTAAGATCAAGCGTCACACGCC
GTTGAGCAAGCTGATGAAGGCTTACTGCGAGCGTCAGGGTCTGAGCATGCGTCA
GATCCGCTTTCGTTTCGATGGCCAGCCGATCAATGAGACTGACACCCCAGCGCAA
CTGG **(SEQ ID NO: 1)**

Guanylyl transferase (GT) large subunit DNA:

AGATGGAAGATGAAGATACCATCGACGTCTTTCAGCAACAGACCGGTGGTATGG
ATGCTAACGTCGTTAGCAGCAGCACCATTGCGACTTACATTGATGCACTGGCCAA
AAACGCATCTGAGCTTGAGCAGCGCAGCACCGCCTACGAGATCAATAACGAATT
GGAGCTGGTTTTCATTAAACCGCCGCTGATCACGCTGACGAACGTCGTGAACATT
AGCACGATTCAAGAGAGCTTTATTCGTTTCACCGTTACCAATAAAGAAGGCGTGA
AGATCCGTACCAAGATTCCGCTGAGCAAAGTGCATGGTCTGGACGTGAAAAATG
TGCAGCTGGTTGATGCGATCGATAACATCGTGTGGGAGAAGAAATCTTTGGTCAC
GGAAAATCGTCTGCACAAGGAATGTCTGCTGCGTCTGTCAACCGAAGAACGCCA
CATCTTCCTGGACTACAAGAAGTATGGTTCCAGCATCCGTCTGGAACTGGTGAAC
CTGATTCAGGCAAAGACCAAGAACTTCACCATTGACTTCAAACTGAAGTATTTCC
TGGGCTCTGGTGCACAGAGCAAATCCAGCTTGTTGCACGCGATTAACCATCCGAA

GAGCCGTCCGAATACGAGCCTGGAGATCGAATTCACGCCGCGTGATAACGAAAC
CGTTCCGTACGATGAGCTGATTAAAGAACTGACGACGTTGAGCCGCCACATCTTT
ATGGCCAGCCCGGAAAACGTGATCCTTAGCCCGCCTATCAATGCGCCGATTAAA
ACCTTTATGTTACCGAAACAAGACATTGTGGGTCTGGACCTGGAAAACCTGTACG
CGGTCACCAAAACGGACGGCATTCCGATCACGATTCGTGTTACCAGCAATGGTCT
GTACTGCTATTTCACTCATTTGGGCTATATCATTCGTTATCCGGTGAAACGCATCA
TTGATTCTGAGGTTGTCGTTTTCGGCGAAGCAGTCAAGGACAAGAATTGGACTGT
GTACCTGATCAAATTGATTGAACCGGTTAACGCCATCAATGACCGCCTGGAAGA
GTCGAAATATGTTGAAAGCAAACTGGTGGATATTTGTGATCGTATCGTGTTCAAG
AGCAAGAAATATGAAGGCCCGTTCACCACGACCAGCGAAGTTGTTGACATGCTG
AGCACCTATCTGCCGAAACAACCTGAGGGTGTGATTCTGTTTTACTCCAAGGGTC
CGAAGAGCAACATTGATTTCAAAATCAAGAAAGAGAATACCATTGATCAGACCG
CCAACGTTGTGTTCCGCTATATGTCCAGCGAGCCTATCATTTTCGGTGAGTCGAG
CATCTTTGTTGAATACAAAAAGTTTAGCAACGATAAGGGTTTTCCGAAAGAATAC
GGTTCCGGTAAGATTGTGTTGTACAACGGCGTCAATTATCTGAACAACATCTACT
GTCTGGAGTACATCAATACCCATAACGAAGTTGGCATTAAGTCTGTTGTCGTCCC
GATCAAATTCATCGCGGAGTTCCTGGTTAACGGTGAGATTCTGAAGCCGCGTATT
GATAAAACTATGAAATACATTAACTCCGAAGATTACTACGGTAATCAGCATAAC
ATCATCGTCGAGCACTTGCGTGATCAAAGCATTAAGATCGGTGACATCTTTAACG
AAGATAAGCTGAGCGATGTAGGCCACCAGTATGCGAACAATGACAAATTTCGCC
TGAATCCGGAAGTCAGCTACTTTACGAATAAGCGCACCCGTGGTCCACTGGGTAT
CCTGAGCAATTATGTTAAAACCCTGTTGATTTCCATGTACTGCTCCAAAACGTTCC
TGGACGACAGCAACAAGCGCAAAGTTCTGGCGATCGACTTCGGTAATGGTGCCG
ATCTGGAGAAGTACTTTTATGGTGAGATCGCATTGCTGGTTGCTACCGACCCGGA
TGCAGATGCGATCGCCCGTGGCAACGAGCGTTACAATAAGCTGAATAGCGGTAT
CAAGACCAAATACTACAAATTCGACTATATTCAAGAGACGATCCGCTCGGACAC
CTTTGTATCCAGCGTGCGTGAGGTGTTTTACTTCGGTAAATTCAACATCATTGACT
GGCAATTCGCCATTCACTATAGCTTTCACCCACGCCACTATGCGACGGTCATGAA
CAACCTGTCTGAGCTGACCGCGAGCGGCGGTAAAGTTCTGATCACCACGATGGA
CGGTGACAAGCTGTCTAAACTGACCGACAAAAAGACCTTCATTATTCACAAAAA
TCTCCCGTCGAGCGAGAATTACATGTCCGTCGAAAAGATTGCGGACGACCGTATT
GTTGTCTACAACCCGAGCACTATGTCGACCCCAATGACCGAGTATATCATCAAAA
AGAATGACATTGTGCGTGTCTTTAATGAATACGGTTTTGTGCTGGTCGACAACGT
CGATTTTGCGACCATCATCGAGAGAAGCAAGAAATTCATTAATGGCGCTTCTACG
ATGGAAGATCGCCCGAGCACGCGTAACTTCTTTGAGCTGAATCGTGGCGCGATTA
AGTGCGAGGGCCTGGACGTCGAGGATCTGCTGTCGTATTACGTGGTTTATGTGTT
TAGCAAACGTTAATGA **(SEQ ID NO: 2)**

Guanylyl transferase (GT) small subunit DNA:

ATGGACGAAATTGTCAAGAATATCCGTGAAGGTACCCACGTTTTACTGCCATTCT
ACGAGACGCTGCCGGAACTGAACCTGAGCCTGGGTAAAAGCCCTCTGCCGAGCC
TGGAGTATGGTGCGAACTATTTTCTGCAGATTTCCCGTGTAAACGATTTGAACCG
CATGCCGACGGACATGCTGAAACTGTTCACCCACGACATCATGCTGCCGGAATCT

GATCTGGATAAAGTTTACGAGATCTTGAAAATCAATTCAGTGAAGTACTATGGCC
GTAGCACCAAGGCCGATGCGGTGGTCGCAGACCTGAGCGCGCGTAACAAACTGT
TTAAACGTGAACGTGACGCAATTAAGAGCAATAACCATCTGACCGAGAACAATT
TGTACATCAGCGACTACAAGATGTTGACTTTTGACGTGTTTCGTCCGCTGTTCGAC
TTTGTTAATGAGAAATACTGCATTATCAAGCTGCCGACGTTGTTTGGTCGCGGCG
TCATTGATACGATGCGCATTTACTGCTCTCTCTTCAAGAATGTGCGCCTGCTGAA
GTGTGTCTCCGACAGCTGGCTGAAAGATAGCGCTATTATGGTTGCGAGCGACGTG
TGTAAAAAGAACCTGGATCTGTTCATGAGCCACGTGAAGAGCGTTACCAAAAGC
AGCAGCTGGAAAGACGTTAACAGCGTCCAGTTCTCCATTCTGAATAACCCGGTCG
ATACCGAGTTTATCAACAAGTTCCTTGAATTCAGCAATCGCGTTTATGAGGCCCT
GTATTACGTTCATAGCCTGCTGTATAGCTCCATGACCTCTGATAGCAAATCGATC
GAGAATAAACACCAACGTCGTCTGGTGAAACTGCTGCTGTAATGA **(SEQ ID NO: 3)**

SUMO-GT large subunit DNA construct with His tag and linker:

ATGGGCCATCATCATCACCATCACGGCAGCCTGCAAGAAGAGAAACCGAAAGAG
GGCGTTAAGACCGAGAATGACCACATTAACCTGAAGGTCGCTGGTCAAGATGGC
AGCGTGGTGCAGTTTAAGATCAAGCGTCACACGCCGTTGAGCAAGCTGATGAAG
GCTTACTGCGAGCGTCAGGGTCTGAGCATGCGTCAGATCCGCTTTCGTTTCGATG
GCCAGCCGATCAATGAGACTGACACCCCAGCGCAACTGGAGATGGAAGATGAAG
ATACCATCGACGTCTTTCAGCAACAGACCGGTGGTATGGATGCTAACGTCGTTAG
CAGCAGCACCATTGCGACTTACATTGATGCACTGGCCAAAAACGCATCTGAGCTT
GAGCAGCGCAGCACCGCCTACGAGATCAATAACGAATTGGAGCTGGTTTTCATT
AAACCGCCGCTGATCACGCTGACGAACGTCGTGAACATTAGCACGATTCAAGAG
AGCTTTATTCGTTTCACCGTTACCAATAAAGAAGGCGTGAAGATCCGTACCAAGA
TTCCGCTGAGCAAAGTGCATGGTCTGGACGTGAAAAATGTGCAGCTGGTTGATGC
GATCGATAACATCGTGTGGGAGAAGAAATCTTTGGTCACGGAAAATCGTCTGCA
CAAGGAATGTCTGCTGCGTCTGTCAACCGAAGAACGCCACATCTTCCTGGACTAC
AAGAAGTATGGTTCCAGCATCCGTCTGGAACTGGTGAACCTGATTCAGGCAAAG
ACCAAGAACTTCACCATTGACTTCAAACTGAAGTATTTCCTGGGCTCTGGTGCAC
AGAGCAAATCCAGCTTGTTGCACGCGATTAACCATCCGAAGAGCCGTCCGAATA
CGAGCCTGGAGATCGAATTCACGCCGCGTGATAACGAAACCGTTCCGTACGATG
AGCTGATTAAAGAACTGACGACGTTGAGCCGCCACATCTTTATGGCCAGCCCGG
AAAACGTGATCCTTAGCCCGCCTATCAATGCGCCGATTAAAACCTTTATGTTACC
GAAACAAGACATTGTGGGTCTGGACCTGGAAAACCTGTACGCGGTCACCAAAAC
GGACGGCATTCCGATCACGATTCGTGTTACCAGCAATGGTCTGTACTGCTATTTC
ACTCATTTGGGCTATATCATTCGTTATCCGGTGAAACGCATCATTGATTCTGAGGT
TGTCGTTTTCGGCGAAGCAGTCAAGGACAAGAATTGGACTGTGTACCTGATCAAA
TTGATTGAACCGGTTAACGCCATCAATGACCGCCTGGAAGAGTCGAAATATGTTG
AAAGCAAACTGGTGGATATTTGTGATCGTATCGTGTTCAAGAGCAAGAAATATG
AAGGCCCGTTCACCACGACCAGCGAAGTTGTTGACATGCTGAGCACCTATCTGCC
GAAACAACCTGAGGGTGTGATTCTGTTTTACTCCAAGGGTCCGAAGAGCAACATT
GATTTCAAAATCAAGAAAGAGAATACCATTGATCAGACCGCCAACGTTGTGTTCC

GCTATATGTCCAGCGAGCCTATCATTTTCGGTGAGTCGAGCATCTTTGTTGAATA
CAAAAAGTTTAGCAACGATAAGGGTTTTCCGAAAGAATACGGTTCCGGTAAGAT
TGTGTTGTACAACGGCGTCAATTATCTGAACAACATCTACTGTCTGGAGTACATC
AATACCCATAACGAAGTTGGCATTAAGTCTGTTGTCGTCCCGATCAAATTCATCG
CGGAGTTCCTGGTTAACGGTGAGATTCTGAAGCCGCGTATTGATAAAACTATGAA
ATACATTAACTCCGAAGATTACTACGGTAATCAGCATAACATCATCGTCGAGCAC
TTGCGTGATCAAAGCATTAAGATCGGTGACATCTTTAACGAAGATAAGCTGAGC
GATGTAGGCCACCAGTATGCGAACAATGACAAATTTCGCCTGAATCCGGAAGTC
AGCTACTTTACGAATAAGCGCACCCGTGGTCCACTGGGTATCCTGAGCAATTATG
TTAAAACCCTGTTGATTTCCATGTACTGCTCCAAAACGTTCCTGGACGACAGCAA
CAAGCGCAAAGTTCTGGCGATCGACTTCGGTAATGGTGCCGATCTGGAGAAGTA
CTTTTATGGTGAGATCGCATTGCTGGTTGCTACCGACCCGGATGCAGATGCGATC
GCCCGTGGCAACGAGCGTTACAATAAGCTGAATAGCGGTATCAAGACCAAATAC
TACAAATTCGACTATATTCAAGAGACGATCCGCTCGGACACCTTTGTATCCAGCG
TGCGTGAGGTGTTTTACTTCGGTAAATTCAACATCATTGACTGGCAATTCGCCATT
CACTATAGCTTTCACCCACGCCACTATGCGACGGTCATGAACAACCTGTCTGAGC
TGACCGCGAGCGGCGGTAAAGTTCTGATCACCACGATGGACGGTGACAAGCTGT
CTAAACTGACCGACAAAAAGACCTTCATTATTCACAAAAATCTCCCGTCGAGCGA
GAATTACATGTCCGTCGAAAAGATTGCGGACGACCGTATTGTTGTCTACAACCCG
AGCACTATGTCGACCCCAATGACCGAGTATATCATCAAAAAGAATGACATTGTG
CGTGTCTTTAATGAATACGGTTTTGTGCTGGTCGACAACGTCGATTTTGCGACCAT
CATCGAGAGAAGCAAGAAATTCATTAATGGCGCTTCTACGATGGAAGATCGCCC
GAGCACGCGTAACTTCTTTGAGCTGAATCGTGGCGCGATTAAGTGCGAGGGCCTG
GACGTCGAGGATCTGCTGTCGTATTACGTGGTTTATGTGTTTAGCAAACGTTAAT
GA **(SEQ ID NO: 4)**

Small ubiquitin-like modifier (SUMO) protein:

EEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIR
FRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG **(SEQ ID NO: 5)**

Guanylyl transferase (GT) large subunit protein:

MDANVVSSSTIATYIDALAKNASELEQRSTAYEINNELELVFIKPPLITLTNVVNISTIQ
ESFIRFTVTNKEGVKIRTKIPLSKVHGLDVKNVQLVDAIDNIVWEKKSLVTENRLHKE
CLLRLSTEERHIFLDYKKYGSSIRLELVNLIQAKTKNFTIDFKLKYFLGSGAQSKSSLL
HAINHPKSRPNTSLEIEFTPRDNETVPYDELIKELTTLSRHIFMASPENVILSPPINAPIK
TFMLPKQDIVGLDLENLYAVTKTDGIPITIRVTSNGLYCYFTHLGYIIRYPVKRIIDSEV
VVFGEAVKDKNWTVYLIKLIEPVNAINDRLEESKYVESKLVDICDRIVFKSKKYEGPF
TTTSEVVDMLSTYLPKQPEGVILFYSKGPKSNIDFKIKKENTIDQTANVVFRYMSSEPI
IFGESSIFVEYKKFSNDKGFPKEYGSGKIVLYNGVNYLNNIYCLEYINTHNEVGIKSVV
VPIKFIAEFLVNGEILKPRIDKTMKYINSEDYYGNQHNIIVEHLRDQSIKIGDIFNEDKL
SDVGHQYANNDKFRLNPEVSYFTNKRTRGPLGILSNYVKTLLISMYCSKTFLDDSNK
RKVLAIDFGNGADLEKYFYGEIALLVATDPDADAIARGNERYNKLNSGIKTKYYKFD

YIQETIRSDTFVSSVREVFYFGKFNIIDWQFAIHYSFHPRHYATVMNNLSELTASGGK
VLITTMDGDKLSKLTDKKTFIIHKNLPSSENYMSVEKIADDRIVVYNPSTMSTPMTEY
IIKKNDIVRVFNEYGFVLVDNVDFATIIERSKKFINGASTMEDRPSTRNFFELNRGAIK
CEGLDVEDLLSYYVVYVFSKR **(SEQ ID NO: 6)**

Guanylyl transferase (GT) small subunit protein:

MDEIVKNIREGTHVLLPFYETLPELNLSLGKSPLPSLEYGANYFLQISRVNDLNRMPT
DMLKLFTHDIMLPESDLDKVYEILKINSVKYYGRSTKADAVVADLSARNKLFKRERD
AIKSNNHLTENNLYISDYKMLTFDVFRPLFDFVNEKYCIIKLPTLFGRGVIDTMRIYCS
LFKNVRLLKCVSDSWLKDSAIMVASDVCKKNLDLFMSHVKSVTKSSSWKDVNSVQ
FSILNNPVDTEFINKFLEFSNRVYEALYYVHSLLYSSMTSDSKSIENKHQRRLVKLLL
(SEQ ID NO: 7)

SUMO-GT large subunit protein with His tag and linker:

MGHHHHHHGSLQEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKA
YCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGMDANVVSSST
IATYIDALAKNASELEQRSTAYEINNELELVFIKPPLITLTNVVNISTIQESFIRFTVTNK
EGVKIRTKIPLSKVHGLDVKNVQLVDAIDNIVWEKKSLVTENRLHKECLLRLSTEER
HIFLDYKKYGSSIRLELVNLIQAKTKNFTIDFKLKYFLGSGAQSKSSLLHAINHPKSRP
NTSLEIEFTPRDNETVPYDELIKELTTLSRHIFMASPENVILSPPINAPIKTFMLPKQDIV
GLDLENLYAVTKTDGIPITIRVTSNGLYCYFTHLGYIIRYPVKRIIDSEVVVFGEAVKD
KNWTVYLIKLIEPVNAINDRLEESKYVESKLVDICDRIVFKSKKYEGPFTTTSEVVDM
LSTYLPKQPEGVILFYSKGPKSNIDFKIKKENTIDQTANVVFRYMSSEPIIFGESSIFVEY
KKFSNDKGFPKEYGSGKIVLYNGVNYLNNIYCLEYINTHNEVGIKSVVVPIKFIAEFL
VNGEILKPRIDKTMKYINSEDYYGNQHNIIVEHLRDQSIKIGDIFNEDKLSDVGHQYA
NNDKFRLNPEVSYFTNKRTRGPLGILSNYVKTLLISMYCSKTFLDDSNKRKVLAIDFG
NGADLEKYFYGEIALLVATDPDADAIARGNERYNKLNSGIKTKYYKFDYIQETIRSDT
FVSSVREVFYFGKFNIIDWQFAIHYSFHPRHYATVMNNLSELTASGGKVLITTMDGD
KLSKLTDKKTFIIHKNLPSSENYMSVEKIADDRIVVYNPSTMSTPMTEYIIKKNDIVRV
FNEYGFVLVDNVDFATIIERSKKFINGASTMEDRPSTRNFFELNRGAIKCEGLDVEDL
LSYYVVYVFSKR **(SEQ ID NO: 8)**

Example 2: Production of SUMO-GT Protein

Shake Flask

[0091] Production of SUMO-GT fusion protein can be performed according to standard methods and procedures. For example, to test and compare expression of the GT and SUMO-GT fusion proteins, a single colony of the E. coli Rosetta strain (Novagen) containing each of the SUMO-eGFP plasmids was inoculated into 5 ml of Luria-Bertani (LB) media containing 100 μg/ml Kanamycin and 30 μg/ml chloramphenicol. This strain is

derived from the lambda DE3 lysogen strain and carries a chromosomal copy of the IPTG-inducible T7 RNA polymerase along with tRNAs on a pACYC-based plasmid. The cells were grown at 37 °C. overnight with shaking at 250 rpm. The next morning the overnight culture was transferred into 100 ml fresh medium to permit exponential growth. When the OD600 value reached -0.6-0.7, protein expression was induced by addition of 1 mM IPTG (isopropropyl-β-D-thiogalactopyranoside), followed by prolonged cultivation at either 37 °C for 3 hours or 20 °C. overnight (about 15 hours).

[0092] After the *E. coli* cells were harvested from LB medium (100 ml) by centrifugation (8,000xg for 10 min at 4 °C.), the cell pellets were suspended in 6 ml of lysis buffer (PBS containing 300 mM NaCl, 10 mM imidazole, 0.1% Triton XlOO and 1 mM PMSF, pH 8.0). The cells were lysed by sonication (at 50% output for 5x30 second pulses). The sonication was conducted with the tube jacketed in wet ice and 1 min intervals between the pulse cycles to prevent heating. After the lysates were incubated with DNase and RNase (each at 40 μg/ml) for 15 min to digest nucleic acids, they were centrifuged at 20,000 g for 30 min at 4 °C., and the supernatant (soluble protein fractions) was collected. The pellets was washed once with 6 ml of the lysis buffer to further extract the soluble fraction; the wash (6 ml) was combined with previous extract (6 ml) to make final volume of 12 ml for the soluble protein sample.

[0093] Insoluble protein samples were prepared from *E. coli* inclusion bodies. Briefly, after the extract containing soluble proteins were removed, the pellets containing inclusion bodies were suspended in the denaturing solubilization buffer (Novagen) that contained 50 mM CAPS (pH 11.0), 0.3% N-laurylsarcosine, and 1 mM DTT and incubated for 20 min at room temperature with shaking. The extract (insoluble protein fraction) was obtained by high-speed centrifugation (80,000xg for 20 min at 4 °C.).

[0094] For detection of expressed proteins using SDS-PAGE, 5μl of the samples prepared above were mixed with 3 μl of SDSPAGE sample buffer containing SDS and β-mercaptoethanol and were heated at 95 °C. for 5 min to facilitate denaturation and reduction of proteins. Proteins were visualized using 15% SDS-polyacrylamide gels with Tris-Glycine running buffer and Coomassie blue staining.

Fermentation

[0095] The substantial increase in the solubility of the final SUMO-GT complexed enzyme was also reproduced by fermentation. Fermentation was performed according to

standard methods and procedures. For example, fermentation methods for production of SUMO-GT fusion protein comprised cell lysis, Immobilized Metal Affinity Chromatography (IMAC), Cation Exchange Chromatography, Anion Exchange Chromatography, and Tangential Flow Filtration (TFF) formulation. Quality testing of the SUMO-GT fusion protein that resulted from fermentation comprised Reducing SDS PAGE to determine purity and identity, Reverse-Phase HPLC to determine purity, A280 measurement of concentration and Limulus amebocyte lysate (LAL) assay to test for endotoxin.

[0096] As shown in **Figure 2**, the yield of soluble SUMO-GT protein produced by fermentation is comparable to that of GT protein produced via the shake flask method.

EQUIVALENTS

[0097] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above Description, but rather is as set forth in the following claims:

CLAIMS

We claim:

1. A method of producing a capped RNA or RNA analog oligonucleotide, wherein a fusion protein facilitates the steps of transferring and methylating a guanylyl molecule to the 5' end of the RNA or RNA analog oligonucleotide.

2. The method of claim 1, wherein the fusion protein comprises a guanylyl transferase and a small ubiquitin-like molecule (SUMO) protein.

3. The method of claim 2, wherein the guanylyl transferase comprises SEQ ID NO: 6 and and SEQ ID NO: 7 and the SUMO protein comprises SEQ ID NO: 5.

4. The method of claim 2, wherein the fusion protein comprises SEQ ID NO: 8 and SEQ ID NO: 7.

5. The method of claim 1, wherein the fusion protein has comparable phosphatase activity, guanylyl transferase activity and methylation activity relative to a wild-type guanylyl transferase protein.

6. A fusion protein, wherein the fusion protein comprises guanylyl transferase and a small ubiquitin-like molecule (SUMO) protein.

7. The fusion protein of claim 6, wherein the guanylyl transferase comprises SEQ ID NO: 6 and SEQ ID NO: 7 and the SUMO protein comprises SEQ ID NO: 5.

8. The fusion protein of claim 6, wherein the guanylyl transferase comprises a large subunit and a small subunit.

9. The fusion protein of claim 8, wherein the SUMO protein is covalently linked and co-expressed with the large subunit.

10. The fusion protein of claim 6, wherein the fusion protein has comparable phosphatase activity, guanylyl transferase activity and methylation activity relative to a wild-type guanylyl transferase protein.

11. A vector encoding a fusion protein comprising guanylyl transferase protein and a small ubiquitin-like molecule (SUMO) protein.

12. The vector of claim 11, wherein the vector comprises SEQ ID NO: 1 and SEQ ID NO: 2.

13. The vector of claim 11, wherein the vector comprises SEQ ID NO: 1, SEQ ID NO: 2, and SEQ ID NO: 3.

14. The vector of claim 11, wherein the vector comprises SEQ ID NO: 4 and SEQ ID NO: 3.

15. A method to produce a guanylyl transferase by fermentation, comprising: a) culturing in a fermentation medium a microorganism that is transformed with at least one recombinant nucleic acid molecule comprising a nucleic acid sequence encoding a guanylyl transferase that has an amino acid sequence that is at least 90% identical SEQ ID NO: 6 and SEQ ID NO: 7; and b) collecting a product produced from the step of culturing.

16. The method of claim 15, wherein the guanylyl transferase comprises a guanylyl transferase fusion protein.

17. The method of claim 16, wherein the guanylyl transferase fusion protein has comparable phosphatase activity, guanylyl transferase activity and methylation activity relative to a wild-type guanylyl transferase protein.

18. The method of claim 16, wherein the guanylyl transferase fusion protein comprises a small ubiquitin-like molecule (SUMO) protein.

19. The method of claim 18, wherein the guanylyl transferase fusion protein comprises SEQ ID NO: 8.

20. The method of claim 18, wherein the SUMO protein is bound to the guanylyl transferase by a covalent link.

21. The method of claim 20, wherein the covalent link is between the SUMO protein and a large subunit of the guanylyl transferase.

22. The method of claim 15, wherein the fermentation medium is selected from the group consisting of Terrific Broth, Cinnabar, 2xYT and LB.

23. The method of claim 15, wherein the microorganism is a bacterium.

24. The method of claim 15, wherein the nucleic acid sequence encoding the guanylyl transferase is at least 90% identical to SEQ ID NO: 2 and SEQ ID NO: 3.

25. The method of claim 15, wherein the recombinant nucleic acid molecule further comprises a nucleic acid sequence encoding a small ubiquitin-like molecule (SUMO) protein.

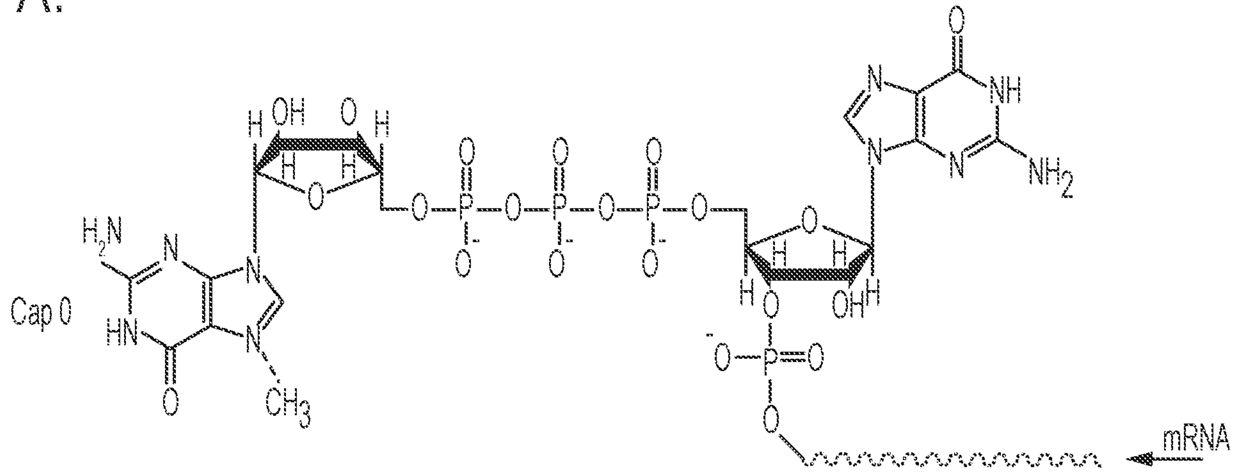
26. The method of claim 25, wherein the nucleic acid sequence encoding a small ubiquitin-like molecule (SUMO) protein is at least 90% identical to SEQ ID NO: 1.

27. The method of claim 15, wherein the product is a guanylyl transferase.

28. The method of claim 27, wherein the product is a guanylyl transferase comprising a guanylyl transferase fusion protein.

29. The method of claim 28, wherein the guanylyl transferase fusion protein further comprises a small ubiquitin-like molecule (SUMO) protein.

A.



B.

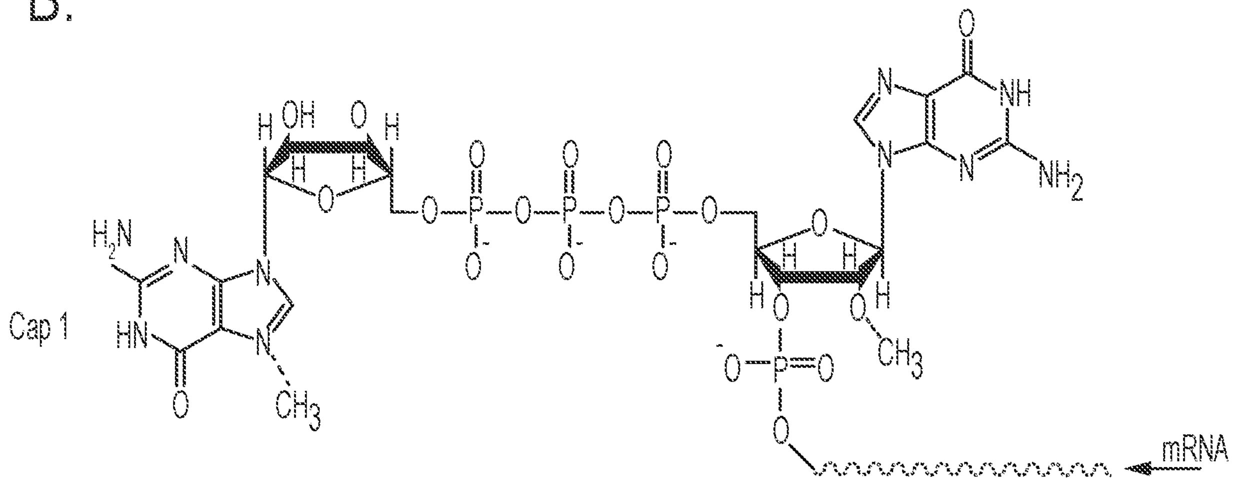


FIGURE 1

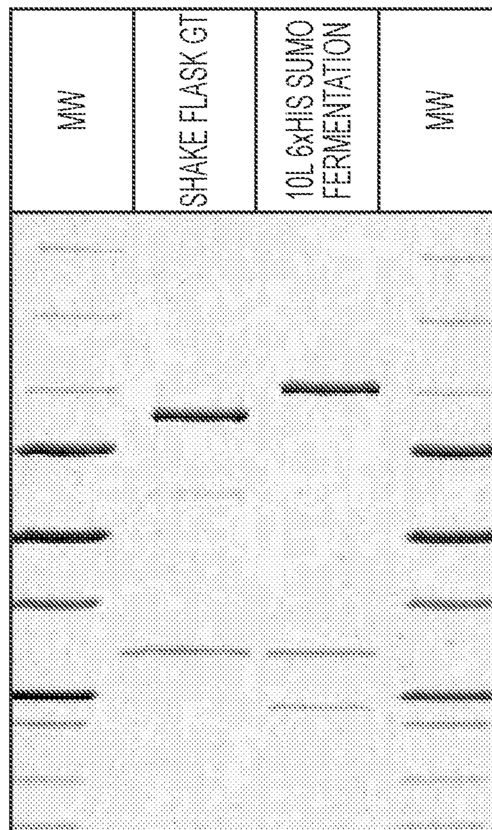


FIGURE 2

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/057044

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/12 C12N15/62
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, Sequence Search, WPI Data, EMBASE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02/090495 A2 (CORNELL RES FOUNDATION INC [US]) 14 November 2002 (2002-11-14) page 5, paragraph 7 - page 6, last paragraph; example 3	6,11

X	KYRIELEIS OTTO J P ET AL: "Crystal Structure of Vaccinia Virus mRNA Capping Enzyme Provides Insights into the Mechanism and Evolution of the Capping Apparatus", STRUCTURE, vol. 22, no. 3, 4 March 2014 (2014-03-04), pages 452-465, XP028625435, ISSN: 0969-2126, DOI: 10.1016/J.STR.2013.12.014	1,2,5,6, 8-11, 15-18, 20-25, 27-29
Y	the whole document	3,4,7, 12-14, 19,26
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 December 2016	05/01/2017

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Wiame, Ilse

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	-& OTTO J.P. KYRIELEIS ET AL: "Crystal Structure of Vaccinia Virus mRNA Capping Enzyme Provides Insights into the Mechanism and Evolution of the Capping Apparatus Supplemental Information", STRUCTURE, vol. 22, no. 3, 1 March 2014 (2014-03-01), pages 452-465, XP055329824, AMSTERDAM, NL ISSN: 0969-2126, DOI: 10.1016/j.str.2013.12.014 the whole document -----	
X	WO 2014/028429 A2 (MODERNA THERAPEUTICS INC [US]) 20 February 2014 (2014-02-20)	1,5, 15-17, 22-24, 27,28
	paragraphs [0010], [0016], [0021], [0022], [0032]; examples 19-21; table 29 -----	
Y	Anonymous: "SUMOpro-3 Gene Fusion Technology - Product Manual", , 28 January 2014 (2014-01-28), pages 1-10, XP055329912, Retrieved from the Internet: URL:https://www.lifesensors.com/wp-content/uploads/2015/10/1010K_1011_1010A_1016_Ecoli_SUMOpro3_Manual.pdf [retrieved on 2016-12-16] the whole document -----	3,4,7, 12-14, 19,26
A	GUO P ET AL: "INTERACTION AND MUTUAL STABILIZATION OF THE TWO SUBUNITS OF VACCINIA VIRUS MRNA CAPPING ENZYME COEXPRESSED IN ESCHERICHIA COLI", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, NATIONAL ACADEMY OF SCIENCES, US, vol. 87, no. 11, 1 June 1990 (1990-06-01), pages 4023-4027, XP000608138, ISSN: 0027-8424, DOI: 10.1073/PNAS.87.11.4023 the whole document -----	15-29

-/--

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	FRESCO L D ET AL: "ACTIVE SITE OF THE MRNA-CAPPING ENZYME GUANYLYLTRANSFERASE FROM SACCHAROMYCES CEREVISIAE: SIMILARITY TO THE NUCLEOTIDYL ATTACHMENT MOTIF OF DNA AND RNA LIGASES", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, NATIONAL ACADEMY OF SCIENCES, US, vol. 91, 1 July 1994 (1994-07-01), pages 6624-6628, XP002928212, ISSN: 0027-8424, DOI: 10.1073/PNAS.91.14.6624 page 6624, column 2, last paragraph -----	6

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2016/057044

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 02090495 A2	14-11-2002	AT 490968 T	15-12-2010
		AU 2002257243 A1	18-11-2002
		CA 2445049 A1	14-11-2002
		EP 1392717 A1	03-03-2004
		JP 2004532033 A	21-10-2004
		US 2003086918 A1	08-05-2003
		US 2005069988 A1	31-03-2005
		US 2009317868 A1	24-12-2009
		WO 02090495 A2	14-11-2002
WO 2014028429 A2	20-02-2014	EP 2885419 A2	24-06-2015
		US 2015211039 A1	30-07-2015
		WO 2014028429 A2	20-02-2014