



(51) International Patent Classification:

C12Q 1/6876 (2018.01) C12Q 1/6869 (2018.01)
A61K 31/7115 (2006.01)

(21) International Application Number:

PCT/AU2024/051208

(22) International Filing Date:

14 November 2024 (14.11.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

2023903657 14 November 2023 (14.11.2023) AU

(71) Applicant: **THE UNIVERSITY OF QUEENSLAND**
[AU/AU]; St Lucia, Queensland 4072 (AU).

(72) Inventors: **MERCER, Timothy Robert**; Building 1,
Room 212, 11 Salisbury Road, Ipswich, Queensland 4035
(AU). **CHEETHAM, Seth William**; Building 1, Room
212, 11 Salisbury Road, Ipswich, Queensland 4035 (AU).
GUNTER, Helen Mary; Building 1, Room 212, 11 Salis-
bury Road, Ipswich, Queensland 4035 (AU).

(74) Agent: **SPRUSON & FERGUSON**; GPO Box 3898, Syd-
ney, New South Wales 2001 (AU).

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

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, CV,
GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE,
SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))
- in black and white; the international application as filed
contained color or greyscale and is available for download
from PATENTSCOPE

(54) Title: METHOD FOR DETECTING CAPPED RNA MOLECULES

(57) Abstract: The present disclosure relates to methods of identifying and/or quantifying a capped RNA molecule in vitro transcribed from a DNA template by an RNA polymerase in a sample comprising a mixture of capped RNA molecules and uncapped RNA molecules; wherein the capped RNA molecule has been capped with a capped primer having the general form m7GpppN1[N2]m[N3]n. In one embodiment, the method may involve identifying and/or quantifying at least the 5' terminal nucleoside of an RNA molecule from the sample.

METHOD FOR DETECTING CAPPED RNA MOLECULES

CROSS REFERENCE TO RELATED APPLICATION

[0001] The present application claims priority from Australian Provisional Application No. 2023903657, filed on 14 November 2023, the content of which is incorporated by reference herein in its entirety.

CROSS REFERENCE TO SEQUENCE LISTING

[0002] The present application contains a sequence listing which has been submitted electronically as an XML document in the ST.26 format and is incorporated by reference herein in its entirety.

FIELD OF THE DISCLOSURE

[0003] The present disclosure relates to methods for identifying capped and uncapped RNA molecules in a sample. In one form, it relates to methods of quantifying capped and/or uncapped mRNA molecules in a sample.

BACKGROUND OF THE DISCLOSURE

[0004] Any discussion of the prior art throughout the specification should in no way be considered as an admission that such prior art is widely known or forms part of the common general knowledge in the field.

[0005] Synthetic RNA compositions such as mRNA vaccines and therapies utilise at least some of the characteristics of endogenous mRNA to enable their translation by ribosomes within cells. Synthetic mRNA molecules may encompass a 5' cap, typically a 7-methyl-guanosine cap (5' m7G cap) or an analogue thereof, which may assist with ribosome recognition and subsequent translation. The 5' cap may also have additional functions, such as regulating the nuclear export of mRNA molecules and preventing degradation of mRNA molecules by exonucleases.

[0006] mRNA vaccines are typically rigorously analysed to ensure their quality, effectiveness and safety. This includes confirming the presence of the 5' cap. Uncapped mRNA molecules, even if otherwise intact, cannot typically be effectively bound by ribosomes and translated into a functional protein product. The presence of uncapped mRNA molecules within a synthetic mRNA pharmaceutical composition may reduce the activity and function of the composition.

[0007] Moreover, uncapped mRNAs activate RIG-I, a cytoplasmic antiviral innate immune sensor, leading to an interferon response (Hornung *et al.* 2006). Reverse Phase Ion-Pair Chromatography (IP-RP-HPLC) may be used to verify the presence of a 5' cap. Alternatively, Liquid Chromatography-Mass Spectrometry (LC-MS) can also be used for this purpose (Galloway *et al.*, 2020; Beverly *et al.*, 2016). However, both HPLC and MS-based methods are time-consuming and expensive. Furthermore, IP-RP-HPLC and LC-MS provide an inadequate quantitative assessment of the fraction of 5' capped mRNA molecules in a mixed mRNA sample.

[0008] In this context, there is a need for a method of identifying capped and uncapped RNA molecules in a sample. Further, there is a need for a method of quantifying capped and/or uncapped RNA molecules in a sample.

SUMMARY OF THE DISCLOSURE

[0009] In an aspect, the present disclosure provides a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template by an RNA polymerase in a sample comprising a mixture of capped RNA molecules and uncapped RNA molecules;

wherein the capped RNA molecule has been capped with a capped primer having the general form $m7GpppN^1[N^2]_m[N^3]_n$; wherein

m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate;

N^1 , N^2 , and each instance of N^3 are independently any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

a) identifying at least the 5' terminal nucleoside of an RNA molecule from the sample;

wherein:

identifying the presence of N^1 as the 5' terminal nucleoside identifies a capped RNA molecule; and/or

identifying the absence of N^1 as the 5' terminal nucleoside identifies an uncapped RNA molecule.

[0010] In an aspect, the present invention provides a method of identifying capped RNA molecules *in vitro* transcribed from a DNA template by an RNA polymerase in a sample comprising a mixture of capped RNA molecules and uncapped RNA molecules;

wherein the capped RNA molecules have been capped with a capped primer having the general form $m7GpppN^1[N^2]_m[N^3]_n$; wherein

m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate;

N^1 , N^2 , and each instance of N^3 are independently any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

a) identifying at least the 5' terminal nucleoside of a plurality of RNA molecules from the sample by:

identifying the presence of N¹ as the 5' terminal nucleoside and then identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +1 position of the DNA template, which identifies capped RNA molecules; and/or
identifying the absence of N¹ as the 5' terminal nucleoside and then identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +2 position of the DNA template, which identifies uncapped RNA molecules,

wherein:

the nucleoside at the +1 position of the DNA template comprises adenosine and N¹ comprises adenosine;

the nucleoside at the +1 position of the DNA template comprises cytidine and N¹ comprises cytidine; or

the nucleoside at the +1 position of the DNA template comprises thymidine and N¹ is uridine; and

b) quantifying the presence of N¹ and thereby quantifying the capped RNA molecules in the sample.

[0011] In an embodiment, the present disclosure provides a method of an aspect, wherein the nucleoside at the +1 position of the DNA template is adenosine and N¹ is selected from the group consisting of adenosine, N²-methyladenosine, or N⁶-methyladenosine, and the nucleoside at the +2 position of the DNA template comprises guanosine and N² comprises guanosine.

[0012] In an embodiment, N¹ is not a preferred initiation nucleoside for the RNA polymerase.

[0013] In an embodiment, step a) further comprises identifying the length of the RNA molecule, wherein the uncapped RNA molecule is x nucleotides long, and x is any suitable integer; and wherein identifying an RNA molecule that is at least x+1 nucleotides long identifies a capped RNA molecule; and identifying an RNA molecule that is x nucleotides long identifies an uncapped RNA molecule.

[0014] In an embodiment, the quantifying in step b) comprises:

determining the absolute amount of the capped RNA molecules in the sample;

determining the abundance of the capped RNA molecules relative to uncapped RNA molecules in the sample; and/or

determining the abundance of the capped RNA molecules relative to total RNA molecules in the sample.

[0015] In an embodiment, at least one of N¹, N², and each instance of N³ are independently any modified or unnatural nucleoside, and step a) further comprises identifying a modified or unnatural nucleoside, wherein:

identifying the presence of a modified or unnatural nucleoside identifies a capped RNA molecule; and

identifying the absence of a modified or unnatural nucleoside identifies an uncapped RNA molecule;

optionally wherein the modified or unnatural nucleoside is selected from N2-methyladenosine or N6-methyladenosine.

[0016] In an embodiment, N¹ is a modified or unnatural nucleoside, and step a) further comprises identifying the modified or unnatural nucleoside, wherein:

identifying the presence of the modified or unnatural nucleoside as the 5' terminal nucleoside identifies a capped RNA molecule; and

identifying the absence of the modified or unnatural nucleoside as the 5' terminal nucleoside identifies an uncapped RNA molecule.

[0017] In an embodiment, m = 1 and n = 0.

[0018] In an embodiment, the capped primer is selected from the group consisting of 5'm7GpppAN, m7Gpppm2AN, m7Gpppm2Am2N, m7Gpppm6AN, and m7Gpppm6Am2N, m7G(5')ppp(5')G, m7(3')-O-Me-G(5')ppp(5')G, m7G(5')ppp(5')(2'OMeA)pG and m7G(5')ppp(5')(2'OMeA)pU, m7(3'OMeG)(5')ppp(5')m6(2'OMeA)pG, m7Gpppm6A mN, m7GpppN mN, and 7mG3OMe5'ppp5'N.

[0019] In an embodiment, the RNA molecules have been transcribed by T7 RNA polymerase.

[0020] In an embodiment, the DNA template comprises:

a sequence selected from the group consisting of TAATACGACTCACTATA (SEQ ID NO: 23), TAATACGACTCACTATAAG (SEQ ID NO: 15) , and TAATACGACTCACTATAAT (SEQ ID NO: 24) wherein the RNA molecules have been transcribed by T7 RNA polymerase;

a sequence AATTAACCCTCACTATA (SEQ ID NO: 27) wherein the RNA molecules have been transcribed by T3 RNA polymerase; or

a sequence ATTTAGGTGACACTATA (SEQ ID NO: 28) wherein the RNA molecules have been transcribed by Sp6 RNA polymerase.

[0021] In an embodiment, the identifying of step a) further comprises

reverse transcribing an RNA molecule from the sample to form a complementary DNA (cDNA) molecule, and

identifying at least a terminal nucleoside of the cDNA molecule wherein identifying the terminal nucleoside of the cDNA molecule identifies the at least 5' terminal nucleoside of the RNA molecule.

[0022] In an embodiment, the identifying in step a) comprises identifying at least one to three 5' nucleosides of the RNA molecule, identifying at least one to ten 5' nucleosides of the RNA molecule, or identifying essentially all of the nucleosides of the RNA molecule.

[0023] In an embodiment, the identifying comprises a method selected from the group consisting of Reverse Transcription polymerase chain reaction (PCR) (RT-PCR), Real-Time Quantitative RT-PCR (qRT-PCR), TaqMan qPCR, Quantitative PCR (qPCR), digital PCR (dPCR), digital RT-PCR (dRT-PCR), RT-Droplet Digital PCR (RT-ddPCR), Nested PCR, Multiplex PCR, Touchdown PCR, Hot start PCR, and High-fidelity PCR), RT-PCR followed by High-Resolution Melt (HRM) Analysis, Illumina sequencing, Ion Torrent sequencing, PacBio sequencing, Nanopore sequencing, Sanger sequencing and RNA-seq sequencing; direct RNA sequencing; sequence-by-synthesis; ligation or probe hybridisation methods; RNA-sequencing, RNA ligation, cDNA sequencing, isothermal amplification methods (such as LAMP), RT-Loop-Mediated Isothermal Amplification (RT-LAMP), and cDNA synthesis with a specific template switching primer or hybridization with a specific oligonucleoside.

[0024] In an aspect, the present disclosure a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules;

wherein the capped RNA molecule has been capped with a capped primer having the general form $m7GpppN^1[N^2]_m[N^3]_n$; wherein

m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate;

N^1 is a modified or unnatural nucleoside, and N^2 and each instance of N^3 are independently any natural, modified or unnatural nucleoside; or N^2 is a modified or unnatural nucleoside, and N^1 and each instance of N^3 are independently any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

a) identifying whether at least the 5' terminal nucleoside of an RNA molecule from the sample is a natural, modified or unnatural nucleoside;

wherein:

identifying the presence of a modified or unnatural nucleoside as the 5' terminal nucleoside identifies a capped RNA molecule; and

identifying the absence of a modified or unnatural nucleoside as the 5' terminal nucleoside identifies an uncapped RNA molecule.

[0025] In an aspect, the present disclosure provides a method of quantifying a capped RNA molecule *in vitro* transcribed from a DNA template in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules;

wherein the capped RNA molecule has been capped with a capped primer having the general form $m7GpppN^1[N^2]_m[N^3]_n$; wherein

m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate;

N¹ is a modified or unnatural nucleoside and N² and each instance of N³ are independently any natural, modified or unnatural nucleoside, or

N² is a modified or unnatural nucleoside and N¹ and each instance N³ are independently any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

identifying whether at least the 5' terminal nucleoside of a plurality of RNA molecules from the sample is a natural, modified or unnatural nucleoside;

wherein:

identifying the presence of a modified or unnatural nucleoside as the 5' terminal nucleoside identifies a capped RNA molecule; and

identifying the absence of a modified or unnatural nucleoside as the 5' terminal nucleoside identifies an uncapped RNA molecule;

quantifying the presence of a modified or unnatural nucleoside as the 5' terminal nucleoside and thereby quantifying the capped RNA molecules in the sample.

[0026] In an aspect, the present disclosure provides a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules;

wherein the capped RNA molecule has been capped with a capped primer having the general form m7GpppN¹[N²]_m[N³]_n; wherein

m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate;

N¹ is any natural, modified or unnatural nucleoside wherein N¹ is not a preferred initiation nucleoside for the RNA polymerase;

N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; and

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

identifying the length of the RNA molecule, wherein:

the uncapped RNA molecule is x nucleotides long, and

x is any suitable integer; and

wherein:

identifying an RNA molecule that is at least $x+1$ nucleotides long identifies a capped RNA molecule; and

identifying an RNA molecule that is x nucleotides long identifies an uncapped RNA molecule.

[0027] In an aspect, the present disclosure provides a pharmaceutical composition comprising a capped RNA molecule identified by the method of any one of the preceding aspects.

[0028] In an aspect, the present disclosure provides a multivalent pharmaceutical composition comprising a plurality of populations of capped RNA molecule identified by the method of a method of the disclosure.

BRIEF DESCRIPTION OF DRAWINGS

[0029] **Figure 1** provides a schematic diagram illustrating (A) an example of a capped primer having the form $m7GpppAmG$, where Am is N^1 and G is N^2 ; (B) a DNA template comprising a native T7 promoter sequence (*italics*), an unmodified first transcriptional start site (TSS; underlined), showing the DNA sequence corresponding to the transcribed nucleotides (**bold**), where the arrow indicates the expected actual transcription start site; (C) a DNA template comprising a T7 promoter sequence (*italics*) having a modified first TSS (underlined) to enhance initiation of *in vitro* transcription with the capped primer of (A), showing the nucleotides of the DNA sequence corresponding to the transcribed nucleotides (**bold**) when transcription is initiated with a capped primer having the form $m7GpppAmG$, where the arrow indicates the expected actual transcription start site; (D) a DNA template comprising a T7 promoter sequence (*italics*) having a modified TSS (underlined) to enhance initiation of *in vitro* transcription with the capped primer of (A), showing the DNA sequence corresponding to the transcribed nucleotides (**bold**) when transcription is initiated with a guanosine triphosphate (rather than the capped primer), where the arrow indicates the expected actual transcription start site. The shaded box in (B), (C) and (D) highlights the position of two nucleotides that may be varied to correspond with the nucleosides in the capped primer. These nucleotides may act as alternative transcriptional start sites (TSS) depending on whether transcription is initiated with a capped primer (i.e., first transcription start site) or by a guanosine triphosphate (second transcription start site). The first terminal transcription start site corresponds to the 5' terminal +1 nucleoside of the capped RNA molecule (E); and the second terminal transcription start site corresponds to the 5' terminal +2 nucleoside of the capped RNA molecule (E). (E) shows a capped RNA molecule transcribed from the DNA template at (C), where transcription was initiated at the first TSS. The 5' nucleoside of the capped RNA molecule is the methylated adenosine provided by the capped primer, shown as the +1 position, and the next downstream nucleotide is guanosine from the capped primer, shown as the +2 position (boxed). (F) shows an uncapped RNA molecule transcribed from the DNA template at (D), where transcription was initiated at the second TSS. The 5' nucleoside is a guanosine which has been incorporated from a GTP during

transcription (boxed). This corresponds to the +2 position of the capped RNA molecule (shown in parentheses) and the second transcription start site.

[0030] Figure 2 provides a schematic diagram illustrating the key features of a synthetic mRNA polynucleotide molecule.

[0031] Figure 3 provides a schematic diagram illustrating an *in vivo* transcription and enzymatic capping process using T7 polymerase with (A) showing a DNA template having a native T7 promoter sequence (underlined) indicating the transcription start site (TSS) is a guanosine (G; bold) at the first transcription start site (TSS) (i.e. +1 position) of the DNA template; and (B) a mRNA molecule transcribed from the DNA template having a 5' m7G cap1 showing a methylated guanosine (boxed) following transcription and enzymatic capping.

[0032] Figure 4 provides a schematic diagram illustrating an *in vitro* co-transcriptional capping process using T7 polymerase and a capped primer (e.g., CleanCap1 AG reagent (m7Gpppm2AmG) with (A) a DNA template having an modified T7 promoter sequence (underlined) indicating an adenosine (A; bold) at the first transcription start site (TSS) (i.e. +1 position); and mRNA molecules following transcription and enzymatic capping from the DNA template having either a (B) CleanCap1 AG reagent (m7Gpppm2AmG) showing that the 5' terminal nucleotide (i.e., +1 with respect to the DNA template) is 2'OMeA (adenosine methylated at the 2'O position; bold), (C) CleanCap2 AG reagent (m7Gpppm2Am2G) showing that the 5' terminal nucleotide (+1 with respect to the DNA template) is 2'OMeA (adenosine methylated at the 2'O position; bold) and the next nucleotide downstream (+2 with respect to the DNA template) is 2'OMeG (guanosine methylated at the 2'O position), or (D) CleanCap M6 reagent (m7Gpppm6Am2G) showing that the 5' terminal nucleotide (+1 with respect to the DNA template) is 2'OMeA and N⁶ methylated adenosine (bold) and the next nucleotide downstream (i.e. +2 with respect to the DNA template) is guanosine. Detection of an adenosine as the 5' transcribed nucleoside and/or detection of methylation indicates incorporation of the synthetic 5' cap analogue.

[0033] Figure 5 provides a schematic diagram illustrating an *in vitro* transcription using T7 polymerase in the absence of capping with (A) a DNA template having an artificial T7 promoter sequence (underlined) indicating that transcription starts at the second transcription start site at the +2 guanosine (G) position (bold); and (B) mRNA molecules transcribed from the DNA template without a 5' cap analogue and showing that the 5' terminal nucleotide (+2 with respect to the DNA template) is unmethylated guanosine (bold). Detection of a guanosine as the 5' transcribed nucleoside and/or absence of methylation indicates absence of the synthetic 5' cap analogue.

[0034] Figure 6 provides (A) a schematic diagram illustrating a mRNA vaccine sample having capped and uncapped mRNA molecules following synthesis in the presence of the CleanCap reagent AG having the form m7GpppAmG, wherein (upper panel) transcription initiates at the

first TSS nucleotide (i.e. A (+1) nucleotide; bold) for capped RNA molecules or (lower panel) transcription initiates at the second TSS (i.e. G (+2) nucleotide; bold) for the uncapped RNA molecules; (B) the same molecules following library preparation incorporating a library adaptor; and (C) a histogram showing next generation sequencing results measuring the 5' terminal nucleotide showing (upper panel) alignment of capped mRNA molecules encoding green fluorescent protein starting with A(+1) and (lower panel) uncapped molecules starting with G(+2).

[0035] Figure 7 provides histograms of sequenced mRNA molecules read alignments across an artificial T7 promoter generated with (upper panel) co-transcriptionally capped mRNA molecules with a synthetic 5' cap analogue (CleanCapAG) with alignments starting at A(+1); (middle panel) uncapped mRNA molecules (in the absence of synthetic 5' cap analogue and no co-transcriptional capping) with alignments starting at G(+2); and (lower panel) co-transcriptionally capped mRNA molecules following post-transcriptional enzymatic de-capping with alignments starting at A(+1).

[0036] Figure 8 provides histograms of sequenced mRNA molecules read alignments across the native T7 promoter generated with (upper panel) uncapped mRNA molecules with alignments starting at G(+1); (middle panel) post-transcriptionally capped mRNA molecules using Faustovirus capping enzyme with alignments starting at G(+1); and (lower panel) post-transcriptionally capped mRNA molecules followed by post-transcriptional enzymatic de-capping with alignments starting at G(+1).

[0037] Figure 9 provides a schematic diagram illustrating the promoter region of the DNA template and transcription start site during *in vitro* transcription with T7 RNA polymerase and (A) the native T7 promoter sequence having a GG sequence at the transcription start site indicating that transcription initiates at the first nucleoside of the GG sequence (i.e. at a guanosine), (B) a modified T7 promoter sequence having an AG sequence at the transcription start site, when transcription initiates with the capped primer having the general format m7GpppAG, indicating that transcription initiates at the first nucleoside of the AG sequence (adenosine) for capped RNA molecules, or (C) modified T7 promoter sequence having an AG sequence at the transcription start site, when transcription initiates without the capped primer having the general format m7GpppAG, indicating that transcription initiates at the second nucleotide of the AG sequence (guanosine) for uncapped RNA molecules.

[0038] Figure 10 provides a histogram showing the fraction of mRNA molecules in a synthetic mRNA encoding the green fluorescent protein (GFP) of 1186 nucleotides (grey) in length for an uncapped mRNA, and 1187 nucleotide in length for a capped mRNA as measured using next-generation sequencing, with sequenced reads aligned to reference plasmid sequence.

[0039] **Figure 11** provides a scatter plot showing a direct correlation between the proportion of RNA molecules having an adenosine as the 5' terminal nucleoside and the proportion of capped molecules in a sample, as measured using next generation sequencing.

[0040] **Figure 12** provides the design of TaqMan probes to detect the 5' nucleotide of RNA molecules. Taqman probes were designed to be complementary to a region flanking the TSO, the TSS and the 5' UTR of the RNA molecule sequence. Two probes were designed, incorporating either (A) the first TSS (i.e. A (+1)) as well as the second TSS (i.e. G (+2)) in the case of capped mRNA, or (B) the second TSS (i.e. G (+2)) but not the first TSS (i.e. A (+1)) for uncapped mRNA. The capped mRNA detecting probe incorporated the HEX fluorophore, and the uncapped mRNA detecting probe incorporated the FAM fluorophore.

[0041] **Figure 13** provides the design of primers for SYBR green assay to detect the 5' nucleotide of RNA molecules. Primers were designed to be complementary to a region flanking the TSO and the TSS of the RNA molecule sequence. The 3' terminal nucleotide of the Forward primers was complementary to TSS of the RNA, terminating in either (A) the first TSS (i.e. A (+1)) in the case of synthetic Cap-analogue incorporation, or (B) the second TSS (i.e. G (+2)), but not the first TSS (i.e. A (+1)) for uncapped mRNA.

[0042] **Figure 14** provides results of the qRT-PCR cap detection assay. (A) shows the delta Ct values derived from the SYBR capping analysis normalised against the delta Ct values from the cap independent control primers. Results are presented as a ratio between each sample and the results from the 100% capped sample. (B) shows the scatter plot of delta Ct analysis of Taqman capping assay. The delta Ct values derived from the Taqman capping analysis were normalised against delta Cts from the cap independent control primers. Results are presented as a ratio between each sample and the results from the 100% capped sample.

DETAILED DESCRIPTION

[0043] The presence of a 5' cap on a synthetic RNA molecule may facilitate translation of an encoded target polypeptide. A 5' cap can be incorporated into the synthetic RNA molecule during transcription initiation (i.e., co-transcriptional capping) or added following the completion of transcription (i.e., post-transcriptional capping). However, the process of capping of RNA molecules is not 100% effective. Accordingly, samples of synthetic mRNA vaccines or therapeutics may include a fraction of capped and uncapped mRNA molecules. A sample of synthetic mRNA molecules such as mRNA vaccines or therapeutics is typically assessed for the relative or absolute amount of capped mRNA molecules within the sample.

[0044] The present disclosure relates to methods of measuring the presence or absence of a 5' cap on synthetic RNA molecules within a sample, for example, by analysing the 5' terminal profile and/or length of synthetic RNA molecule, from which the presence or absence of a 5' cap can be deduced. In an embodiment, this provides a means of determining the relative abundance or absolute amounts of capped and uncapped RNA molecules in a sample.

[0045] In an embodiment, the disclosure relates to differentiating a capped RNA molecule from an uncapped RNA molecule in a sample comprising a mixture of capped RNA molecules and uncapped RNA molecules *in vitro* transcribed from a DNA template by an RNA polymerase, wherein the capped RNA molecule has been capped with a capped primer having the general form [cap]-[linker]-N¹[N²]_m[N³]_n; wherein N¹, N² and each instance of N³ are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8; by identifying at least one structural difference the capped RNA molecule and the uncapped RNA molecule from the sample. In an embodiment, N¹ is not the preferred initiation nucleoside for the RNA polymerase.

[0046] The RNA molecules may be transcribed from a DNA template using an RNA polymerase in the presence of the capped primer. The DNA template may have a nucleotide sequence comprising a promoter sequence suitable for use with the RNA polymerase, including a transcription initiation site that facilitates perfectly or substantially complementary base pair binding between the capped primer and the 3' strand (i.e. template strand) of the DNA template during transcription.

[0047] When the polymerase selects the capped primer to initiate transcription of an RNA molecule, the resulting RNA molecule is capped during the transcription reaction. This is referred to as co-transcriptional capping. In this embodiment, the N¹ nucleoside of the capped primer is the 5' terminal nucleoside of the capped RNA molecule (see Figure 1E). Typically, transcription is initiated at the first transcription start site of the modified promoter at the nucleotide corresponding to N¹ (compare Figure 1A and Figure 1C).

[0048] When the polymerase does not select the capped primer to initiate transcription of an RNA molecule an alternative transcription initiating agent (e.g. a nucleotide or analogue) initiates transcription. Typically, the preferred transcription initiation nucleoside triphosphate for that promoter is selected, and an uncapped RNA molecule is produced. In this embodiment, the N¹ nucleoside of the capped primer is not the 5' terminal nucleoside of the uncapped RNA molecule (compare Figure 1A and 1F). Typically, transcription is initiated at the first nucleotide corresponding to the preferred transcription initiation nucleoside of the RNA polymerase.

[0049] In an embodiment, the N¹ nucleoside of the capped primer is not the preferred initiating nucleoside of the RNA polymerase. In this embodiment, when the preferred transcription initiation nucleoside triphosphate is selected to initiate transcription (rather than the capped primer), transcription does not initiate at the nucleotide corresponding to N¹ (i.e., transcription does not start at the first transcription start site, referred to as the +1 position with reference to the DNA template as shown in Figure 1D). Typically, transcription initiates at a second transcription start site (referred to as the +2 position with reference to the DNA template as shown in Figure 1D). Accordingly, in this embodiment, the transcription initiation site for the

capped RNA molecule is distinct from that of the uncapped RNA molecule (compare Figure 1C and 1D).

[0050] The present disclosure relates to the finding that the differential transcription initiation (e.g., incorporation of either the capped primer or a nucleotide to initiate transcription) results in identifiable differences between capped RNA molecules and uncapped RNA molecules.

[0051] In an embodiment, the method comprises identifying a characteristic selected from the group consisting of:

identifying the presence of N^1 as the 5' terminal nucleoside of the RNA molecule, which identifies a capped RNA molecule; and/or identifying the absence of N^1 as the 5' terminal nucleoside which identifies an uncapped RNA molecule;

identifying the 5' terminal nucleoside as corresponding to a first transcription start site of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to a second transcription start site of the DNA template, which identifies an uncapped RNA molecule;

identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +1 position of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +2 position of the DNA template, which identifies an uncapped RNA molecule;

identifying a modified or unnatural nucleoside, which identifies a capped RNA molecule; and/or identifying the absence of a modified or unnatural nucleoside, which identifies an uncapped RNA molecule, where at least one of N^1 , N^2 , and each instance of N^3 are independently any modified or unnatural nucleoside; and

where the uncapped RNA molecule is x nucleotides long, and x is any suitable integer;

identifying an RNA molecule that is at least x+1 nucleotides long, which identifies a capped RNA molecule; and identifying an RNA molecule that is x nucleotides long, which identifies an uncapped RNA molecule.

[0052] In one form, the present disclosure relates to methods of differentiating and/or identifying a capped RNA molecule and an uncapped RNA molecule in a sample. In another form, the presence of the capped RNA molecules and/or uncapped RNA molecules in the sample is quantified. In an embodiment, the absolute amount of the capped RNA molecules and/or uncapped RNA molecules in the sample is quantified. In an embodiment, the abundance of the capped RNA molecules relative to uncapped RNA molecules in the sample is quantified. In an embodiment, the abundance of the capped RNA molecules relative to total RNA molecules in the sample is quantified.

[0053] The relative or absolute quantity of capped and/or uncapped RNA molecules may be determined using a number of different techniques, including capillary electrophoresis, PCR-based or next-generation-sequencing-based techniques.

[0054] The method of the disclosure may be compatible with other methods of distinguishing the capped RNA molecules from the uncapped RNA molecules in a sample.

Definitions

[0055] In the context of this specification, the terms "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

[0056] The term "about", as used herein, when applied to a value of interest, refers to a value that is similar to a stated value. In certain embodiments, it is understood to refer to a range of +/- 10%, preferably +/- 9%, +/- 8%, +/- 7%, +/- 6%, +/- 5%, +/- 4%, +/- 3%, +/- 2%, or +/- 1%; or +/- 0.05% or +/- 0.1% of the stated value unless otherwise stated or otherwise evident from the context.

[0057] As used herein, the term "capped primer" is used to describe a short oligonucleotide molecule comprising, for example, a 5' cap or 5' cap analogue linked via a phosphate group (e.g. a triphosphate bridge) to, e.g., 1 to 10 nucleosides. The capped primer may initiate transcription and be incorporated at the 5' terminus of an RNA molecule co-transcriptionally to provide a 5' cap analogue to the RNA molecule. In an embodiment, the capped primer has a 5' 7-methyl guanosine (m⁷G) linked via a 5'-5' triphosphate bridge to the first nucleotide (N¹), which in turn, may be linked to the 5' end of a further nucleotide (N²), etc. The capped primer typically has a 3' OH group to facilitate linkage with a 3' nucleoside during transcription.

[0058] The term "capped RNA molecule" as used herein refers to an RNA molecule having a 5' cap or 5' cap analogue. In an embodiment, the 5' cap may be bound to the 5' carbon of the 5' terminal nucleotide of the mRNA polynucleotide by a 5' triphosphate bridge. The 5' cap may be a natural 5' cap or a 5' cap analogue.

[0059] The terms "comprise", "comprises", "comprised" or "comprising", "including" or "having" and the like in the present specification and claims are used in an inclusive sense, i.e., to specify the presence of the stated features but not preclude the presence of additional or further features.

[0060] As used herein, the term "complementary" is used to describe the relationship between a first nucleotide sequence and a second nucleotide sequence by the Watson-Crick base-pairing rules, wherein adenine (A) bases pair with uracil (U) bases in an RNA molecule or thymine (T) bases in a DNA molecule; and cytosine (C) bases pair with guanine (G) bases in both RNA and DNA molecules. For example, for a DNA polynucleotide molecule, the sequence "5'-A-G-T-C-3'" is perfectly complementary to the sequence "3'-T-C-A-G-5'"; noting that in an RNA sequence, uracil (U) is typically used in place of thymine (T).

[0061] The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in detection methods that depend upon binding between nucleic acids. The

sequence of a nucleic acid need not be “perfectly” (100%) complementary to that of its target in order to hybridise. Complementarity may be “partial” in which only some of the nucleic acid bases are matched according to the base pairing rules. It will be understood that two sequences are “specifically complementary” when the two molecules can hybridise under appropriate conditions, that is, a first polynucleotide molecule comprising a first nucleotide sequence is specifically complementary to a second polynucleotide molecule comprising a second nucleotide sequence, when the two molecules can hybridise and form a duplex structure under conditions appropriate for the reaction being undertaken (for example, ligation, PCR, sequencing, etc). The term “specifically complementary” can be used interchangeably with “substantially complementary”. It is also to be understood that two nucleotide molecules need not be complementary over their entire length. For example, a portion of a first polynucleotide molecule may be specifically complementary and hybridise with a portion of a second polynucleotide molecule. In this example, the two molecules may not be hybridised over portions that are not specifically complementary. These terms may also be used in reference to individual nucleotides, especially within the context of oligonucleotides. For example, a particular nucleotide within an oligonucleotide may be noted for its complementarity, or lack thereof, to a nucleotide within another nucleic acid strand, in contrast or comparison to the complementarity between the rest of the oligonucleotide and the nucleic acid strand.

[0062] Conditions under which hybridisation occurs may be stringent, such as 400 mM NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA, 50°C or 70°C for 12-16 hours followed by washing. Other conditions, such as physiologically relevant conditions as may be encountered inside an organism, can also apply. Substantial complementarity allows the relevant function of the nucleic acid to proceed, e.g., binding of an oligonucleotide molecule to a polynucleotide molecule during a PCR based assay, sequencing etc. The skilled person will be able to determine the set of conditions most appropriate for a test of complementarity of two sequences in accordance with the ultimate application of the hybridized nucleotides.

[0063] The term “corresponds to” as used herein with reference to the relationship between two nucleotides present at the same or equivalent position within different polynucleotide molecules having the same or complementary nucleobases. Corresponding nucleotides may have different modifications. Further, uracil in an RNA molecule may correspond to thymine in a DNA molecule, as would be understood by those skilled in the art. Accordingly, when the nucleotide sequence of the capped primer corresponds to the nucleotide sequence of the DNA template, then the capped primer would typically bind to the 3' strand (i.e., template strand) of the DNA template at the relevant site of interest using the rules of complementary binding.

[0064] The term “co-transcription” as used herein refers to an *in vitro* transcription reaction where an RNA molecule is transcribed from a DNA template, for example, by an RNA polymerase, where capping of the RNA molecule occurs in the same transcription reaction. In

one form of co-transcription, transcription is initiated with a capped primer such that the capped primer is incorporated at the 5' terminus of the RNA molecule, which both caps the RNA molecule with a 5' cap analogue and initiates transcription.

[0065] As used herein, the term "DNA template" refers to a double stranded DNA molecule comprising at least a promoter sequence and encodes a target RNA molecule of interest, from which an RNA molecule can be transcribed. A DNA template may be a double stranded linear DNA, a partially double stranded linear DNA, circular double stranded DNA, DNA plasmid, PCR amplicon, a modified nucleic acid template which is compatible with RNA polymerase.

[0066] As used herein, the term "effectively" when used with reference to a particular parameter or particular outcome is intended to refer to a sufficient percentage of the parameter or outcome so as to achieve the desired result.

[0067] As used herein, "expression" of a nucleic acid sequence refers to translation of an mRNA into a polypeptide, assembly of multiple polypeptides into an intact protein (e.g., enzyme) and/or post-translational modification of a polypeptide or fully assembled protein (e.g., enzyme). In this application, the terms "expression" and "production," and grammatical equivalents, are used interchangeably.

[0068] The term "isolated" as used herein refers to material that is substantially or essentially free from components that normally accompany it in its native state. For example, an "isolated polynucleotide" as used herein refers to a polynucleotide which has been purified from the sequences which flank it in a naturally occurring state, e.g., a DNA fragment which has been removed from the sequences that are normally adjacent to the fragment. Alternatively, an "isolated peptide" or an "isolated polypeptide" and the like, as used herein, refer to *in vitro* isolation and/or purification of a peptide or polypeptide molecule from its natural cellular environment, and from association with other components of the cell, i.e., it is not associated with *in vivo* substances.

[0069] The term "messenger RNA" or "mRNA" as used herein refers to an RNA polynucleotide molecule that encodes at least one polypeptide. mRNA as used herein encompasses both modified and unmodified RNA. mRNA may contain one or more coding and non-coding regions.

[0070] The term "nucleotide" as used herein, in its broadest sense, refers to a compound and/or substance that is or can be incorporated into a polynucleotide chain. It is understood by those skilled in the art that a nucleotide is typically composed of three distinctive chemical sub-units: a five-carbon sugar molecule (i.e. a pentose-sugar-ring, deoxyribose in DNA or ribose in RNA), a nucleobase (e.g., adenine (A), cytosine (C), guanine (G), thymine (T) or uracil (U)), and a phosphate group (e.g. a monophosphate, di-phosphate or tri-phosphate group). Chemical convention names the carbon atoms in the sugar molecule from 1' to 5', and this convention also dictates that the polynucleotide molecule has a 5' end and a 3' end. In a polynucleotide molecule, the 3' carbon of a first nucleotide is linked to the 5' carbon of the next nucleotide.

Those skilled in the art will understand that polynucleotide sequences are typically read in a 5' to 3' direction, unless specifically stated otherwise.

[0071] In some embodiments, a nucleotide is a compound and/or substance that is or can be incorporated into a polynucleotide chain via a phosphodiester linkage. In some embodiments, “nucleotide” refers to individual nucleic acid residues (e.g., nucleotides and/or nucleosides). The term “nucleotide” can be used interchangeably with “nucleic acid”. “Nucleotide” encompasses RNA (ribonucleotides) as well as single and/or double-stranded DNA and/or cDNA.

[0072] A nucleotide comprises a nucleoside and a phosphate group. However, it will be understood by a person skilled in the art that in common usage, the term “nucleotide” includes nucleosides. One example of a nucleotide is a nucleoside 5' triphosphate (NTP). Other common triphosphate nucleotides are provided in Table 1. Nucleotides include RNA nucleotides (i.e., ribonucleotides) and DNA nucleotides (i.e., deoxyribonucleotides).

[0073] The term “nucleoside” as used herein refers to molecule comprising a pentose sugar (ribose for RNA, deoxyribose for DNA) linked to a nucleobase (nitrogenous base) and includes any suitable natural, modified or unnatural nucleoside. Examples of common natural nucleosides include adenosine, guanosine, cytidine, 5-methyluridine, uridine, and thymidine as shown in Table 1. Naturally occurring nucleobases include purine rings (for example, adenine, guanine, and N⁶-methyladenine) and pyrimidine rings (for example, cytosine, thymine, 5-methylcytosine, pseudouracyl). Naturally occurring nucleosides for example include, but are not limited to, ribo, 2'-O-methyl or 2'-deoxyribo derivatives of adenosine, guanosine, cytidine, thymidine, uridine, inosine, 7-methylguanosine or pseudouridine.

Table 1: Examples of nucleosides and nucleotides

Nucleobase	Purine or Pyrimidine	Nucleosides	Triphosphate Nucleotides
Adenine (A)	Purine	adenosine, deoxyadenosine	adenosine triphosphate, deoxyadenosine triphosphate
Cytosine (C)	pyrimidine	cytidine, deoxycytidine	cytidine triphosphate, deoxycytidine triphosphate
Guanine (G)	Purine	guanosine, deoxyguanosine	guanosine triphosphate, deoxyguanosine triphosphate
Thymine (T)	pyrimidine	5-methyluridine, thymidine	5-methyluridine, thymidine triphosphate
Uracil (U)	pyrimidine	Uridine, deoxyuridine	uridine triphosphate, deoxyuridine triphosphate

[0074] As used herein, the terms “nucleoside analogues”, “modified nucleosides”, or “nucleoside derivatives” include synthetic nucleosides as described herein. Nucleoside derivatives also include nucleosides having modified base or/and sugar moieties, with or without protecting groups and include, for example, 2'-deoxy-2'-fluorouridine, 5-fluorouridine and the like. The compounds and methods provided herein include such base rings and synthetic analogues thereof, as well as unnatural heterocycle-substituted base sugars, and acyclic substituted base sugars. Other nucleoside derivatives that may be utilized with the present disclosure include, for example, LNA nucleosides, halogen-substituted purines (e.g., 6-fluoropurine), halogen-substituted pyrimidines, N⁶-ethyladenine, N⁴-(alkyl)-cytosines, 5-ethylcytosine, and the like.

[0075] As used herein, an “oligonucleotide”, “oligonucleotide molecule”, “oligonucleotide primer” or “primer” is a single-stranded polynucleotide molecule that may be naturally occurring or synthesised to have a user-specified sequence of interest. It may be RNA, DNA or a chimeric RNA/DNA polynucleotide molecule and may contain modified nucleosides. Typically, at least a portion of an oligonucleotide molecule is specifically or perfectly complementary to a polynucleotide sequence of interest and hybridises to a specifically complementary single-stranded polynucleotide molecule. Oligonucleotide molecules are typically considered to be short polynucleotide molecules; however, their length may be varied. Their length may be suitable for use in at least one of a range of applications including polymerase chain reaction (PCR)-based applications, sequencing applications, molecular cloning and molecular probes. Oligonucleotide primers may contain one or more modification groups. Oligonucleotide primers may include RNA, DNA, and/or other modified nucleosides. The person skilled in the art is capable of designing and preparing oligonucleotide primers that are appropriate for transcription of DNA template sequence.

[0076] The term “operably connected” or “operably linked” as used herein refers to the functional relationship between two or more nucleic acid segments such as a gene and a regulatory element including but not limited to a promoter, which then regulates the expression of the gene.

[0077] The term “pharmaceutically acceptable” as used herein refers to substances that do not cause substantial adverse allergic or immunological reactions when administered to a subject. A “pharmaceutically acceptable carrier” includes, but is not limited to, solvents, coatings, dispersion agents, wetting agents, isotonic and absorption delaying agents and disintegrants.

[0078] As used herein, the term “polynucleotide molecule” refers to a DNA or RNA nucleic acid molecule comprising a chain of nucleotides and may include an oligonucleotide molecule or a target nucleic acid of interest. The terms “RNA nucleic acid molecule”, “RNA molecule” and “RNA polynucleotide molecule” can be used interchangeably. Similarly, the terms “DNA nucleic

acid molecule", "DNA molecule" and "DNA polynucleotide molecule" can be used interchangeably.

[0079] The term "polynucleotide variant" refers to polynucleotides displaying substantial sequence identity with a reference polynucleotide sequence or polynucleotides that hybridize with a reference sequence under stringent conditions. The term also encompasses polynucleotides that are distinguished from a reference polynucleotide by the addition, deletion or substitution of at least one nucleotide. Accordingly, the term "polynucleotide variant" includes polynucleotides in which one or more nucleotides have been added or deleted, or replaced with different nucleotides. In this regard, it is well understood in the art that certain alterations inclusive of mutations, additions, deletions and substitutions can be made to a reference polynucleotide whereby the altered polynucleotide retains the biological function or activity of the reference polynucleotide. The term "polynucleotide variant" also includes naturally occurring allelic variants. The terms "peptide variant" and "polypeptide variant" and the like refer to peptides and polypeptides that are distinguished from a reference peptide or polypeptide by the addition, deletion or substitution of at least one amino acid residue. In certain examples, a peptide or polypeptide variant is distinguished from a reference peptide or polypeptide by one or more substitutions, which may be conservative or non-conservative. In certain examples, the peptide or polypeptide variant comprises conservative substitutions and, in this regard, it is well understood in the art that some amino acids may be changed to others with broadly similar properties without changing the nature of the activity of the peptide or polypeptide. Peptide and polypeptide variants also encompass peptides and polypeptides in which one or more amino acids have been added or deleted, or replaced with different amino acid residues.

[0080] As used herein, the term "preferred initiation nucleoside" or "preferred initiation nucleotide" refers to the nucleoside or nucleotide, for example, a nucleoside triphosphate, with which a particular RNA polymerase typically initiates transcription.

[0081] The term "promoter" as used herein refers to a region of DNA template that directs and controls the initiation of transcription of a particular DNA sequence to produce an RNA molecule. Promoters are located on the same strand and upstream on the DNA (towards the 5' region of the sense strand). Promoters are typically immediately adjacent to (or partially overlap with) the DNA sequence to be transcribed. Nucleotide positions in the promoter are designated relative to the transcriptional start site, where transcription of DNA into RNA begins (position +1).

[0082] As used herein, the term "specific" when used in reference to an initiating capped oligonucleotide primer sequence and its ability to hybridize to a DNA template is a sequence that has at least 50% sequence identity with a portion of the DNA template when the initiating capped oligonucleotide primer and DNA strand are aligned. Higher levels of sequence identity

that may be preferred include at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, and most preferable 100% sequence identity.

[0083] The term "synthetic" as applied to polynucleotide molecules is intended to mean that the polynucleotide molecules are produced by an *in vitro* transcription reaction, which does not occur naturally within a cell.

[0084] As used herein, the term "synthetic RNA" or "synthesised RNA" refers to an RNA polynucleotide molecule that has been produced using non-natural processes, such as *in vitro* transcription, wherein polymerase enzymes, DNA template and ribonucleotides are used in a chemical reaction to synthesise the synthetic RNA. Synthetic mRNA as used herein can include non-natural nucleosides, non-natural phosphate backbones, non-natural 5' caps.

[0085] For example, synthetic RNA can comprise non-natural nucleosides nucleoside analogues such as analogues having chemically modified bases or sugars, backbone modifications, etc. In some embodiments, an mRNA is or comprises natural nucleosides (e.g., adenosine, guanosine, cytidine, uridine); nucleoside analogues (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, a synthetic mRNA molecule incorporates modified, artificial or unnatural nucleotides, nucleoside analogues, nucleotide derivatives, modified uridine, N1-methyl-pseudouridine, etc.

[0086] A "therapeutically effective amount" is at least the minimum concentration or amount required to effect a measurable improvement of a particular disease or condition. A therapeutically effective amount herein may vary according to factors such as the disease state, age, sex and weight of the patient. A therapeutically effective amount is also one in which any toxic or detrimental effects are outweighed by the therapeutically beneficial effects.

[0087] As used herein, the terms "transcription" or "transcription reaction" refers to methods known in the art for enzymatically producing RNA molecules that are complementary to a DNA template, thereby producing RNA copies of a DNA sequence. The RNA molecule synthesized in transcription reaction is called an "RNA transcript" or "transcript".

[0088] As used herein, the term "transcription start site" refers to a particular nucleotide within or following the promoter region of a DNA template from which transcription initiates.

[0089] The term “uncapped RNA molecule” as used herein refers to an RNA molecule without a 5' cap. In an embodiment, the uncapped mRNA molecule comprises a 5' triphosphate group, that is a triphosphate group bound to the 5' carbon of the 5' terminal nucleotide.

[0090] As used herein, the term “unsubstituted” or “unmodified” in the context of the capped primer and NTPs refers to a capped primer and NTPs that have not been modified.

[0091] As used herein, the term “upstream” when used in connection with a polynucleotide molecule, refers to 5' direction. The term “downstream” refers to the 3' direction.

[0092] As used herein, the term “5' terminus” of a polynucleotide refers to the 5' carbon of the first nucleotide in the chain (i.e. the 5' terminal nucleotide) and any chemical group attached to the 5' carbon. The “5' end” of a polynucleotide molecule as used herein refers to the terminal portion of the molecule having a 5' carbon at its terminus. In an embodiment, the 5' end is the 5' terminus. In an embodiment, the 5' terminal portion includes the 5' terminus, a 5' cap if present, the 5' terminal nucleotide and a number of nucleotides immediately adjacent the 5' terminal portion, for example, less than 50 nucleotides, less than 40 nucleotides, less than 30 nucleotides, less than 20 nucleotides, or less than 10 nucleotides.

[0093] The group at the 5' carbon may be a phosphate group (e.g. a monophosphate, diphosphate, triphosphate, etc); however, it is to be understood that other groups may be present at the 5' end. Capped mRNA molecules of the present disclosure have a cap linked to a 5' triphosphate bridge at the 5' carbon.

[0094] The “3' terminus” of a polynucleotide molecule as used herein refers to 3' carbon of the last nucleotide in the polynucleotide chain (i.e. the 3' terminal nucleotide) and any chemical group attached to the 3' carbon. The 3' carbon typically has a hydroxyl group; however, it is to be understood that other groups may be attached to the 3' carbon.

[0095] As used herein, the term “+1 position” refers to the first nucleoside in the capped RNA molecule, and the term “+2 position” refers to the second nucleoside in the capped RNA molecule, etc. The +1 position of the capped RNA molecule corresponds to the first transcription start site in the DNA template, and the +2 position of the capped RNA molecule corresponds to the second transcription start site as shown in Figure 1E. For an uncapped RNA molecule of the present disclosure, transcription may start from the second transcription start site, which corresponds to the +2 position relative to the capped RNA molecule as shown in parentheses Figure 1F. The +1 position with respect to the DNA template is the nucleotide from which transcription of the capped primer is initiated, i.e., the first transcriptional start site (TSS). The nucleotide immediately 3' of the first TSS on the DNA template is at the +2 position.

Synthetic RNA molecules

[0096] The success of mRNA vaccines and therapies has been realised, in part, by advances in manufacturing that enabled billions of doses to be safely produced during the COVID-19 pandemic. The synthesis of mRNA vaccines and therapies typically employs *in vitro* transcription.

Synthetically produced RNAs including messenger RNA, transfer RNA, small nucleolar RNA (snoRNA), guide RNAs, etc, may be generated by *in vitro* transcription of a DNA template by an RNA polymerase in a reaction containing nucleoside-5'-triphosphate (NTPs) under suitable conditions. A 5' cap, such as a 5' m7G cap or analogue, can be incorporated during *in vitro* transcription (i.e., co-transcriptionally) or can be added post-transcriptionally using enzymes such as Faustovirus capping enzyme, vaccinia virus capping enzyme, etc. The target RNA molecules vaccine may be purified to eliminate residual contaminants and formulated for delivery.

[0097] In an embodiment, the RNA molecule may be an mRNA molecule, a transfer RNA molecule, a small nucleolar RNA (snoRNA) molecule, a guide RNA molecule, etc. In an embodiment, the RNA molecule may be an mRNA molecule.

[0098] A synthetic messenger (mRNA) molecule of the present disclosure may include a number of features typically found in naturally produced mRNA molecules, for example:

- a. a 5' cap, for example, a 5' 7-methyl-guanosine cap or a cap analogue, which facilitates ribosome recognition and improves translation and mRNA stability;
- b. a 5' untranslated region (UTR), which aids or regulates translation by a ribosome, the sequence of which may be natural or artificial;
- c. a coding region that includes an open reading frame that encodes a protein of interest;
- d. a 3' untranslated region, which facilitates improved translation and stability, the sequence of which may be natural or artificial; and
- e. a polyA tail, the length of which is associated with mRNA stability and expression.

[0099] Figure 2 illustrates features that may be present in a synthetic mRNA molecule.

[00100] To enhance translation efficiency, the sequence of the open reading frame may undergo codon optimization to avoid rare codons and contain fewer uracil residues. Synthetic mRNA vaccines and therapies also often incorporate unnatural modified nucleotides, such as N1-methyl-pseudouridine, methoxyl uridine, etc, for example to minimize recognition by the innate immune response and/or reduce susceptibility to nuclease digestion.

[00101] The mRNA sequence may be modified to reduce RNA secondary structures. The synthetic mRNA molecule may have an internal ribosome entry site (IRES) that can facilitate translation initiation in a 5' cap-independent process. In an embodiment, the mRNA molecule may have a polyA tail, which may be of length greater than 10 nucleotides, and which can be encoded within the DNA template and transcribed. Alternatively, a polyA tail may be added following transcription using enzymes (such as polyadenylase).

[00102] The synthetic mRNA can, for example, encode a vaccine or antigen sequence, or a therapeutic protein, such as an enzyme, antibody (or fragment thereof) or peptide. In an embodiment, the synthetic mRNA molecule is a vaccine and/or therapeutic reagent. In an

embodiment, the present disclosure provides a pharmaceutical composition comprising the synthetic mRNA molecule that has been tested using a method of the present disclosure.

[00103] The presence of the 5' cap and other RNA quality features on mRNA molecules for a vaccine or therapy is typically rigorously analysed to ensure its quality, safety and effectiveness (Analytical Procedures for mRNA Vaccine Quality- Draft Guidelines, USP, 2023). The USP currently recommends using Reverse Phase Ion-Pair Chromatography (IP-RP-HPLC) to verify the presence of the 5' cap. Alternatively, Liquid Chromatography-Mass Spectrometry (LC-MS) can also be used for this purpose (Galloway *et al.*, 2020; Beverly *et al.*, 2016). This technique first uses Antarctic phosphatase to convert uncapped monophosphate diphosphates and triphosphates to a 5'OH to facilitate analysis. The technique uses a complementary DNA oligonucleotide that hybridises with the mRNA to direct RNase H cleavage, with the products then analysed using LC-MS. The analysis of the LC-MS trace shows peaks corresponding to uncapped (5' OH after phosphatase treatment) and capped, indicated by mass.

[00104] The performance of different capping technologies, including their impact on mRNA translation and stability, is commonly assayed functionally using reporter mRNAs such as luciferase. These functional assays measure the impact of the 5' cap on mRNA expression, thereby providing an indirect inference of capping status rather than directly detecting the 5' cap. Further, both HPLC and MS-based methods are time-consuming and expensive, and provide a poor quantitative measurement of the fraction of capped and uncapped mRNA molecules in a mixed sample.

5' caps

[00105] While not wanting to be bound by theory, the 5' cap is typically considered to facilitate translation of an open reading form encoded by a mRNA molecule. It is understood that uncapped mRNA molecules, even if otherwise intact, cannot typically be translated. Moreover, uncapped mRNAs may activate RIG-I, a cytoplasmic antiviral innate immune sensor, leading to an interferon response (Hornung *et al.* 2006). The presence of an internal ribosome entry site (IRES) may allow initiation of translation in a cap-independent pathway. The 5' cap may also have additional functions, such as regulating nuclear export of mRNA and/or preventing degradation of mRNA by exonucleases.

[00106] Naturally occurring cap structures typically comprise a 7-methyl guanosine (m7G) cap that is linked via a triphosphate bridge to the 5'-end of the first transcribed nucleotide, resulting in a dinucleotide cap of m7G(5')ppp(5')N, where N is any nucleoside. This is also represented interchangeably as m7GpppN. The 5' terminal nucleoside (i.e., the cap-adjacent nucleotide) *in vivo* is often a guanosine. *In vivo*, the cap is added in the nucleus enzymatically immediately post-transcription, in a reaction typically catalysed by the enzyme guanylyl transferase.

[00107] The 5' cap facilitates the effectiveness of synthetic mRNA vaccines and therapies. For synthetic mRNAs that are generated using *in vitro* transcription, a 5' cap is typically added to an

mRNA molecule either during the initiation of *in vitro* transcription (co-transcriptionally), wherein a synthetic 5' cap analogue is incorporated at transcription initiation; or after mRNA synthesis (i.e. post-transcriptionally), wherein an enzyme performs the transferase and methylation reaction required to convert the 5' mRNA termini into a Cap0 or Cap1 structure.

[00108] While not wanting to be bound by theory, the 5' cap is typically considered to facilitate translation of the open reading frame encoded by a mRNA molecule. It is understood that uncapped mRNA molecules, even if otherwise intact, cannot typically be translated. Moreover, uncapped mRNAs may activate RIG-I, a cytoplasmic antiviral innate immune sensor, leading to an interferon response (Hornung et al. 2006). Additionally, the 5' cap is involved in nuclear export, stability and degradation of the mRNA molecule.

[00109] Within eukaryotic cells, the cap is added post-transcriptionally to the nascent RNA shortly after transcription initiation by guanylyl transferase. During the capping reaction, a first intermediate known as Cap0 (m⁷GpppN) is produced. Cap0 structures lack a 2'-O-methyl residue at the first and second 5' terminal (i.e. cap adjacent) nucleotides. The Cap0 formation may subsequently undergo methylation at the 2'-O position of the first 5' terminal nucleoside, creating Cap1 (m⁷GpppNmN). The Cap1 structure can be further methylated at the 2'-O position of the second 5' terminal nucleoside to form a Cap2 (m⁷GpppNmNm) structure. The presence of either Cap0, Cap1 and Cap2 is important for self/non-self-recognition of the mRNA by the cellular innate immune system. When the first 5' nucleoside is adenosine, the adenosine can additionally be N⁶ methylated to form an m⁶Am cap.

[00110] In an embodiment, the capped primer of the present disclosure comprises a Cap0 structure. In an embodiment, the capped primer of the present disclosure comprises a Cap1 structure. Cap1 structures have a 2'-O-methyl residue at the first cap adjacent (5' terminal) nucleotide. In an embodiment, the capped primer of the present disclosure comprises a Cap2 structure. Cap2 structures have a 2'-O-methyl residue attached to both the first and second cap adjacent (5' terminal) nucleotides. In an embodiment, the capped primer of the present disclosure comprises a capM6 structure. CapM6 structures have adenosine as the first cap adjacent (5' terminal) nucleotide, which is methylated at the ribose 2'O and N6 positions.

Table 2: Methylation of Cap0, Cap1, Cap2, Cap M6 and TMG Cap structures

Cap type	Structure/Chemical name	Comment
Cap0	m ⁷ GpppN ¹ m ⁷ G(5')pppN N7-methyl guanosine 5'- Triphosphate-5'-[nucleoside]	The initiating/cap adjacent nucleoside lacks methylation.
Cap1	m ⁷ GpppN ¹ m	Methylation on the 2'O position of the initiating/cap adjacent nucleotide (N ¹)

Cap2	m7GpppN ¹ mN ² m	Methylation on the 2'O position of the initiating/cap adjacent nucleotide (N ¹) and methylation on the 2'O position the next downstream nucleotide (N ²)
CapM6 (m ⁶ A _m cap)	m7Gpppm ⁶ Am	The initiating nucleotide is adenosine (A ¹) having methylation on the 2'O position and N ⁶ methylated (N 6-methyladenosine).
TMG cap	m ₃ ^{2,2,7} GpppN ¹	Contains 2,2,7-trimethylated guanosine (TMG).

Wherein:

m indicates methylation

G indicates guanosine

p indicates a phosphate group

N¹ and N² indicates any nucleoside

A indicates adenosine.

[00111] The 5' cap of a synthetic mRNA molecule can comprise the same 5' cap as *in vivo* produced mRNA molecules (e.g. m7G(5')ppp(5')N, where N is any nucleoside). Alternatively, the 5' cap of a synthetic mRNA molecule may comprise a synthetic 5' cap analogue. A variety of m7G cap analogues are known in the art. Cap structures may include ARCA 3'-OCH₃, ARCA 2'-OCH₃ cap analogues (Jemielity, J. *et al.*, 2003), N7-benzylated dinucleoside tetraphosphate analogues (described in Grudzien, E. *et al.*, 2004), phosphorothioate cap analogues (described in Grudzien-Nogalska, E., *et al.*, 2007), and cap analogues (including biotinylated cap analogues) described in U.S. Patent Nos. 8093367, 8304529 and 11,377,642.

Capped primers

[00112] The capped primers of the present disclosure may comprise the general form [cap]-[linker]-N¹[N²]_m[N³]_n; wherein N¹, N² and each instance of N³ are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8. In an embodiment, N¹ is not the preferred initiation nucleoside for the RNA polymerase. The capped primers may comprise a 5' cap or cap analogue linked to at least one nucleoside that may be incorporated with an RNA molecule co-transcriptionally to provide a 5' cap analogue to the RNA molecule.

[00113] The capped primers of the present disclosure may comprise a 7-methyl guanosine (m⁷G) cap or an analogue thereof, or an alternative cap structure as detailed herein.

[00114] The cap or analogue thereof may be linked to the at least one nucleoside by any suitable linker. In one form, the linker is a phosphate group. In an embodiment, the linker is a triphosphate. In one form, the linker is a reverse 5' to 5' triphosphate linkage.

[00115] The at least one nucleoside may comprise any nucleoside suitable for incorporating into the 5' end of an RNA molecule. It may include any suitable natural, modified or unnatural nucleoside, e.g., naturally occurring RNA and DNA nucleosides, synthetic nucleosides, or any suitable modified nucleosides, nucleoside analogues, nucleoside derivatives, or modified nucleobases. Examples include but are not limited to, base and sugar modified nucleosides, nucleotides, and nucleic acids, such as inosine, 7-deazaguanosine, 2'-O-methylguanosine, 2'-fluoro-2'-deoxycytidine, pseudouridine, Locked Nucleic Acids (LNA), and Peptide Nucleic Acids (PNA).

[00116] The modified nucleoside may comprise a modified, synthetic or natural nucleobase such as deoxy-thymine (dT), 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo, particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-daazaadenine and 3-deasaguanine and 3-deazaadenine.

[00117] Modified nucleosides may have one or more substituted sugar moieties, for example, it may include one of the following at the 2'-position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl can be substituted or unsubstituted C₁ to C₁₀ alkyl or C₂ to C₁₀ alkenyl and alkynyl.

[00118] The capped primer may contain natural internucleotide phosphodiester linkages or modifications thereof, or combination thereof. Examples of oligonucleotide internucleotide linkage modifications including phosphorothioate, phosphotriester and methylphosphonate derivatives.

[00119] In an aspect, a capped primer having the general form [cap]-[linker]-N¹[N²]_m[N³]_n; wherein the cap comprises any suitable cap as disclosed herein; the linker comprises any suitable linker as disclosed herein; N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8.

[00120] In an aspect, the capped primer has the general form m⁷GpppN¹[N²]_m[N³]_n; wherein m⁷G is N⁷-methylated guanosine or a guanosine analogue; ppp is a triphosphate; N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; and m is 0 or 1; and n is any integer from 0 to 10.

[00121] In an embodiment, the capped primer may be a dinucleotide, for example, comprising a 7-methyl guanosine (m⁷G) linked via a triphosphate bridge to the 5' end of a first nucleoside

(N¹), resulting in a dinucleotide cap of m⁷G(5')ppp(5')N¹, where N¹ is any nucleoside.

Commercially available examples of dinucleotide capped primers are shown in Table 3.

Table 3: Examples of synthetic dinucleotide capped primers

Commercial Name	Chemical names	Comment	Initiator sequence
mCap (TriLink Biotechnologies)	m ⁷ GpppG; m ⁷ G(5')ppp(5')G; N ⁷ -Methyl-Guanosine-5'-Triphosphate-5'-Guanosine	Dimer N ¹ = G (monomethylated)	G
Anti-Reverse Cap Analogue (ARCA) (TriLink Biotechnologies; ThermoFisher; Invitrogen)	m ⁷ G3'OmpppG m ⁷ (3')-O-Me-G(5')ppp(5')G; N ⁷ -Methyl-3'-O-Methyl-Guanosine-5'-Triphosphate-5'-Guanosine; where 2' or 3' OH group is replaced with —OCH ₃	Dimer; N ¹ = G N ¹ = G analogue of ARCA, where 2' or 3' OH group is replaced with —OCH ₃	G

[00122] In an embodiment, m = 1 and n = 0.

[00123] In an embodiment, the capped primer may be a trinucleotide, for example, comprising a 7-methyl guanosine (m⁷G) linked via a triphosphate bridge to the 5' end of a first nucleoside (N¹), which is linked to a second nucleotide (N²) resulting in a trinucleotide cap of m⁷G(5')ppp(5')N¹N², where N¹ and N² are independently any nucleoside. Commercially available examples of trinucleotide capped primers are shown in Table 4.

Table 4: Examples of synthetic trinucleotide capped primers

Commercial Name	Chemical names	Comment	Initiator sequence
CleanCap Reagent AG	m ⁷ GpppAmG m ⁷ G(5')ppp(5')(2'OMeA)pG	Trimer; Cap1 N ¹ = 2'OMeA; N ² = pG	AG
CleanCap reagent AG (3'Ome)	m ⁷ G3'OmpppAmG m ⁷ (3'OMeG)(5')ppp(5')(2'OMeA)pG includes 5' N7-Methyl-3'-O-methylguanosine	Trimer Cap1 N ¹ = 2'OMeA N ² = pG	AG

CleanCap Reagent AU	m7GpppAmU m7G(5')ppp(5')(2'OMeA)pU	Trimer; Cap1 N ¹ = 2'OMeA; N ² = pU	AU
CleanCap Reagent M6; CleanCap Reagent m ⁶ AG (3'OMe)	m7G3'Ompppm6AmG m7(3'OMeG)(5')ppp(5')m6(2'OMeA)pG includes 5' N7-Methyl-3'-O- methylguanosine	Dimer; Cap1 N ¹ = m6(2'OMeA) N ² = pG	AG

Wherein:

m or Me indicates methylation,
G indicates guanosine,
p indicates a phosphate group,
N indicates any nucleoside, and
A indicates adenosine.

[00124] The CleanCap reagents are commercially available from TriLink Biotechnologies (San Diego, USA).

[00125] In an embodiment, the capped primer is selected from m7GpppAmG, m7G3'OmpppAmG, m7GpppAmU and m7G3'Ompppm6AmG.

[00126] Other synthetic trinucleotide capped primers that may be suitable for use in a method of the disclosure are selected from the group consisting of m7GpppApA, m7GpppApC, m7GpppApG, m7GpppApU, m7GpppCpA, m7GpppCpC, m7GpppCpG, m7GpppCpU, m7GpppGpA, m7GpppGpC, m7GpppGpG, m7GpppGpU, m7GpppUpA, m7GpppUpC, m7GpppUpG, m7GpppUpU, m7G3'OmepppApA, m7G3'OmepppApC, m7G3'OmepppApG, m7G3'OmepppApU, m7G3'OmepppCpA, m7G3'OmepppCpC, m7G3'OmepppCpG, m7G3'OmepppCpU, m7G3'OmepppGpA, m7G3'OmepppGpC, m7G3'OmepppGpG, m7G3'OmepppGpU, m7G3'OmepppUpA, m7G3'OmepppUpC, m7G3'OmepppUpG, m7G3'OmepppUpU, m7G3'OmepppA2'OmepA, m7G3'OmepppA2'OmepC, m7G3'OmepppA2'OmepG, m7G3'OmepppA2'OmepU, m7G3'OmepppC2'OmepA, m7G3'OmepppC2'OmepC, m7G3'OmepppC2'OmepG, m7G3'OmepppC2'OmepU, m7G3'OmepppG2'OmepA, m7G3'OmepppG2'OmepC, m7G3'OmepppG2'OmepG, m7G3'OmepppG2'OmepU, m7G3'OmepppU2'OmepA, m7G3'OmepppU2'OmepC, m7G3'OmepppU2'OmepG, m7G3'OmepppU2'OmepU, m7GpppA2'OmepA, m7GpppA2'OmepC, m7GpppA2'OmepG, m7GpppA2'OmepU, m7GpppC2'OmepA, m7GpppC2'OmepC, m7GpppC2'OmepG, m7GpppC2'OmepU, m7GpppG2'OmepA, m7GpppG2'OmepC, m7GpppG2'OmepG, m7GpppG2'OmepU, m7GpppU2'OmepA, m7GpppU2'OmepC, m7GpppU2'OmepG and m7GpppU2'OmepU.

[00127] In an embodiment, the capped primer of the present disclosure comprises a structure selected from a Cap0 structure, a Cap1 structure, a Cap2 structure, CapM6 structure or TMG Cap structure.

[00128] In an embodiment, the capped primers comprise a structure selected from m7GpppG, m7GpppN, m7G3'OmpppG, m7G3'OmpppA, modified ARCA, and β -S-ARCA. In an embodiment, the capped primers comprise a structure selected from m7GpppNN, m7G3'OpppNN, m7GpppNmN, m7G3'OpppNmN, m7GpppAN, m7G3'OpppAN, m7G3'OmpppAmN, m7Gpppm6AmN, m7G3'Opppm6AmN, m7GpppAmN, m7G3'OmpppAmN, m7GpppAG, m7G3'OpppAG, m7G3'OmpppAmG, m7Gpppm6AmG, m7G3'Opppm6AmG, m7GpppAmG, m7G3'OmpppAmG, m7GpppAU, m7G3'OpppAU, m7G3'OmpppAmU, m7Gpppm6AmU, m7G3'Opppm6AmU, m7GpppAmU, m7G3'OmpppAmU. In an embodiment, the capped primers comprise a structure selected from the preceding claims, wherein the capped primer is selected from the group consisting of 5'm7GpppAN, m7Gpppm2AN, m7Gpppm2Am2N, m7Gpppm6AN, and m7Gpppm6Am2N, m7G(5')ppp(5')G, m7(3')-O-Me-G(5')ppp(5')G, m7G(5')ppp(5')(2'OMeA)pG and m7G(5')ppp(5')(2'OMeA)pU, m7(3'OMeG)(5')ppp(5')m6(2'OMeA)pG, m7Gpppm6A mN, m7GpppN mN, and 7mG3OMe5'ppp5'N.

[00129] In an embodiment, N is any nucleoside. In an embodiment, the capped primers may further comprise a further 1 to 8 nucleosides.

[00130] In an embodiment, the capped primer may comprise natural RNA or DNA nucleosides, or modified nucleoside analogues, natural or solidified phosphodiester linkages, or one or more modified sugars.

[00131] In some embodiments, the capped primer comprises at least one modified nucleoside. In an embodiment, the modified nucleoside comprises a 2'-O-methyl modification (2'Ome, 2'Om). In an embodiment, the modified nucleoside comprises a methyl group at position 6 (m6A).

[00132] In an embodiment, the N¹ nucleoside comprises a 2'-O-methyl modification, which may provide a Cap1 structure when incorporated during transcription initiation. In an embodiment, the N¹ nucleoside comprises a methyl group at position 6 (m6A). In an embodiment, the N¹ nucleoside comprises adenosine having a 2'-O-methyl modification and a methyl group at position 6 (m6(2'OMeA)). In an embodiment, the N¹ nucleoside comprises adenosine having a methyl group on position 6 (m6A). The m6A modification may further increase protein expression relative to CleanCap AG or CleanCap AG (3'OMe). It has been hypothesized that the m6A modification adjacent to the 7-methylguanosine cap may positively influence mRNA stability by preventing enzyme-mediated decapping.

[00133] In an embodiment, the N² nucleoside comprises a 2'-O-methyl modification. In an embodiment, the N² nucleoside comprises adenosine having a methyl group at position 6

(m6A). In an embodiment, the N¹ nucleoside comprises a 2'-O-methyl modification and the N² nucleoside includes a 2'-O-methyl modification.

[00134] In an embodiment, the N¹ nucleoside comprises a methyl group at position 6 (m6A) and the N² nucleoside comprises a methyl group at position 6 (m6A).

[00135] In an embodiment, the 7-methyl guanosine is m7-methyl-3'-O-methyl-guanosine (7mG3'Ome).

[00136] In an embodiment, the capped primer comprises an ARCA analogue that carries a modified 7MeG residue where the 3' and/or 2' position is blocked on the ribose to facilitate directional initiation of *in vitro* transcription.

[00137] Use of dinucleotide (m7GpppN), trinucleotide (m7GpppNN), tetranucleotide (m7GpppNNN), etc, capped primers may improve the incorporation frequency of capped primers into RNA molecules. In an embodiment, the dinucleotide, trinucleotide, or tetranucleotide capped primers of the disclosure may comprise a further 1 to 8 nucleosides. The capped primers maybe, for example, 2 to 10 nucleotides long. In an embodiment, the capped primers have a form selected from m7GpppN, m7GpppNN, m7GpppNNN, m7GpppNNNN, m7GpppNNNNN, m7GpppNNNNNN, m7GpppNNNNNNN, m7GpppNNNNNNNN, m7GpppNNNNNNNNN, m7GpppNNNNNNNNNN, etc, where each instance of N is independently any natural, modified or unnatural nucleoside. The m7G cap may also be modified as described herein.

[00138] In an embodiment, the capped primer comprises nucleotides that correspond to the DNA template at the transcription start site, which may overlap with promoter sequence. In an embodiment, N¹ corresponds to a first transcription start site in the DNA template. In an embodiment, at least one instance of N² or N³ correspond to a second transcription start site in the DNA template. In an embodiment, N¹ is any natural, modified or unnatural nucleoside wherein N¹ is not the preferred initiation nucleoside for the RNA polymerase. In an embodiment, at least one instance of N² or N³ correspond to the preferred initiation nucleoside for the RNA polymerase. In an embodiment, N¹ is not the same nucleoside as at least one instance of N² or N³.

[00139] In some embodiments, the capped primer comprises nucleotides that correspond to the DNA template sequence at the site that is transcribed the 5' end of the transcribed RNA (i.e. the target RNA). In an embodiment, the capped primers have a 3' terminal OH group that is a valid substrate for RNA Polymerase (RNAP) and can enable elongation of the RNA molecule being transcribed.

DNA template

[00140] A DNA template may be a double stranded linear DNA, a partially double stranded linear DNA, circular double stranded DNA, a DNA plasmid, PCR amplicon, or a modified nucleic acid template which is compatible with RNA polymerase. The DNA template of the disclosure is

typically a double stranded DNA molecule comprising at least a promoter sequence and a sequence from which an RNA molecule of interest can be transcribed (e.g. a mRNA molecule; see Figure 1).

[00141] A person skilled in the art will understand that during transcription, RNA polymerase binds to a specific promoter region of a DNA molecule and transcribes an RNA molecule from a transcription start site using the 3' strand (i.e., template strand) of the double stranded DNA molecule as a template for transcription. Nucleosides are added to the RNA strand, typically using the rules of complementary base pairing. The promoter sequence of the DNA template comprises, or is closely followed by, one or more transcription start sites (TSS), from which transcription may be initiated (see Figures 1 and Figures 3-5). The transcription start site of the DNA template may also be referred to as the initiating nucleotide and is the nucleotide from which transcription initiates. As shown in Figures 1 and 3-5, the corresponding nucleotide in the RNA molecule is the 5' terminal nucleotide of the RNA molecule, which may also be referred to as the +1 transcript nucleotide of the RNA molecule. The next transcribed nucleotide of the RNA molecule is referred to as the +2 transcript nucleotide, and so on. The +1 transcript nucleotide corresponds to a +1 nucleotide in the DNA template, and the +2 transcript nucleotide corresponds to a +2 nucleotide in the DNA template (see Figures 1 and 3-5).

[00142] In an embodiment, the nucleoside at the +1 position of the DNA template comprises adenosine and N¹ comprises adenosine. In an embodiment, the nucleoside at the +1 position of the DNA template comprises cytidine and N¹ comprises cytidine. In an embodiment, the nucleoside at the +1 position of the DNA template comprises thymidine and N¹ is uridine. In an embodiment, the nucleoside at the +1 position of the DNA template is adenosine and N¹ is selected from the group consisting of adenosine, N2-methyladenosine, or N6-methyladenosine. In an embodiment, nucleoside at the +2 position of the DNA template is guanosine and N² is guanosine. In an embodiment, nucleoside at the +2 position of the DNA template comprises thymidine and N² comprises guanosine. In an embodiment, the nucleoside at the +2 position of the DNA template comprises thymidine and N² is uridine. In an embodiment, the nucleoside at the +2 position of the DNA template comprises cytidine and N² comprises cytidine.

[00143] The promoter sequence of the DNA template may be modified depending on the RNA polymerase to be used and/or the capped primer to be used. The sequence of the DNA template at the transcription start site may be modified to improve the effectiveness of initiating transcription with a capped primer, which produces a capped RNA molecule, rather than with a nucleoside triphosphate such as guanosine triphosphate, which produces an uncapped RNA molecule. In particular, the DNA template sequence may be varied at the transcription start site to optimise the hybridisation between the capped primer and the nucleotides of the 3' strand (i.e. template strand) of the DNA template at the transcription start site to initiate transcription with the capped primer.

[00144] Accordingly, in an embodiment, the sequence of the DNA template may be modified at the transcription start site to correspond to the nucleotides in the capped primer (see Figure Figures 1 and 3-5). For example, when the capped primer has the form m7GpppAG, the 5' to 3' sequence of the 5' strand of the transcription start site of the DNA template may be modified to AG to optimise hybridisation of the capped primer to the DNA template during transcription in order to facilitate incorporation of the capped primer into the produced RNA molecules.

[00145] In another example, when the capped primer has the form m7GpppAU, the 5' to 3' sequence of the 5' strand of the transcription start site of the DNA template may be modified to AU to optimise hybridisation of the capped primer to the DNA template during transcription in order to facilitate incorporation of the capped primer into the produced RNA molecules. Persons skilled in the art will appreciate that the transcription start site may be modified to correspond to the nucleosides present in other capped primers.

[00146] In an embodiment, the DNA template may comprise a native or modified promoter sequence. Examples of native and modified promoter sequences for T7, T3 and SP6 RNA polymerases, and for self-amplifying RNA are shown in Table 6. The native T7 promoter sequence is TAATACGACTCACTATAGG (SEQ ID NO: 13). In an embodiment, a modified T7 (HG) promoter sequence is TAATACGACTCACTATAHG (SEQ ID NO: 14), where H = A, C or T (A is most commonly used), and the preferred initiation nucleoside is G. The native T3 promoter sequence is AATTAACCCTCACTAAAG (SEQ ID NO: 16). In an embodiment, a modified T3 promoter sequence is TAATACGACTCACTAAAH (SEQ ID NO: 17), where H = A, C or T, and the preferred initiation nucleoside is G. The native SP6 promoter sequence is ATTTAGGTGACACTATAGA (SEQ ID NO: 18). In an embodiment, a modified SP6 promoter sequence is ATTTAGGTGACACTATAH (SEQ ID NO: 20), where H = A, C or T, and the preferred initiation nucleoside is G. The native self-amplifying RNA promoter is TAATACGACTCACTATAGG (SEQ ID NO: 21). In an embodiment, a modified self-amplifying RNA promoter sequence is TAATACGACTCACTATAHT (SEQ ID NO: 22), where H = A, C or T, and the preferred initiation nucleoside is G. In an embodiment, the (+) strand genomes of self-amplifying viruses start with a 5'-AU. Underlining represents alternative transcription start sites. Bold indicates the transcription start site for an RNA molecule when co-transcribed with a capped primer of the form m7GpppN¹[N²]_m[N³] where N¹ is X.

[00147] In an embodiment, the DNA template comprises a sequence selected from the group consisting of TAATACGACTCACTATA (SEQ ID NO: 23), TAATACGACTCACTATAAG (SEQ ID NO: 15), and TAATACGACTCACTATAAT (SEQ ID NO: 24) wherein the RNA molecules have been transcribed by T7 RNA polymerase; a sequence selected from AATTAACCCTCACTAAA (SEQ ID NO: 25), AATTAACCCTCACTAAAG (SEQ ID NO: 26) and AATTAACCCTCACTATA (SEQ ID NO: 27) wherein the RNA molecules have been transcribed by T3 RNA polymerase; or a sequence ATTTAGGTGACACTATA (SEQ ID NO: 28) wherein the

RNA molecules have been transcribed by Sp6 RNA polymerase. In an embodiment, the preferred initiation nucleoside for T7 RNA polymerase is guanosine.

[00148] It will be understood by a person skilled in the art that the methods of this disclosure may be used with any suitable RNA polymerase by incorporating the appropriate promoter sequence at a suitable location within the DNA template.

[00149] In an embodiment, the nucleotides at the transcription start sites of the promoter sequences for at least the T7, T3 and SP6 RNA polymerases can be modified to correspond to the sequence of the capped primer to optimise incorporation of the capped primer into the RNA molecule. Similarly, the nucleotides at the transcription start sites of the promoter sequences for other polymerases or for self-amplifying RNA can be modified to correspond to the nucleosides present in a particular capped primer, which may enhance incorporation of the capped primer into the RNA molecule.

[00150] In an embodiment, the DNA template comprises a promoter having the sequence [N]_yTATA, wherein N is any nucleoside; and y is any suitable integer.

[00151] Persons skilled in the art will understand that the methods of the disclosure can be applied to a range of capped primers, polymerases and target RNA molecules by optimising the DNA template sequence and transcription reactions.

In vitro transcription

[00152] Synthetically produced RNAs including messenger RNA, transfer RNA, small nucleolar RNA (snoRNA), guide RNAs, etc, may be generated by *in vitro* transcription of a DNA template by an RNA polymerase in a suitable reaction containing nucleoside-5'-triphosphate (NTPs). Phage RNA Polymerases, such as T3 polymerase, T7 polymerase, SP6 polymerase and other polymerases are commonly used to drive transcription from a sequence-specific promoter, upstream of the template mRNA sequence of interest in the DNA template. For example, a plasmid DNA template may be synthesised, which includes an RNA polymerase promoter sequence (e.g., a T7 promoter sequence), followed by a sequence encoding the target RNA sequence of interest, and optionally, a restriction enzyme site. In an embodiment, this DNA template may then be amplified using either *in vitro* or bacterial amplification methods and linearized through restriction enzyme digestion before being purified. The DNA template may undergo *in vitro* transcription with an RNA polymerase to synthesize the target RNA molecule. In an embodiment, suitable polymerases include polymerases derived from T7, T3, SP6, K1-5, K1E, K1F or K11 bacteriophages. In an embodiment, the RNA polymerase is selected from T7 RNA polymerase, T3 RNA polymerase and SP6 RNA polymerase. In an embodiment, the RNA molecules have been transcribed by T7 RNA polymerase. In an embodiment, the RNA molecule is an mRNA molecule.

[00153] In an embodiment, RNA molecules are capped post-transcriptionally following *in vitro* transcription using a capping enzyme, such as vaccinia virus capping enzyme, which results in Cap0 or Cap1 RNA molecules.

[00154] Synthetic RNA molecules found in mRNA vaccines and therapies may undergo *in vitro* transcription using a capped primer as described herein to initiate transcription and co-transcriptionally incorporate the 5' cap analogue into the RNA molecule. In an embodiment, RNA molecules are capped co-transcriptionally, for example, using an excess of a capped primers (e.g. having the general form $m7GpppN^1[N^2]_m[N^3]_n$ as described herein) in the transcription reaction. In this embodiment, the capped primer may initiate transcription. While not wanting to be bound by theory, in one embodiment, it is understood that in the absence of a capped primer, T7 RNA polymerase has a strong preference for initiating transcription using guanosine triphosphate (GTP) resulting in guanosine being incorporated as the 5' terminal nucleoside producing an uncapped RNA molecule. However, in the presence of capped primer having the general form $m7GpppN^1[N^2]_m[N^3]_n$, T7 polymerase may preferentially initiate transcription with a capped primer, allowing initiation of transcription with a variety of 5' sequences and co-transcriptional capping of the RNA molecule.

[00155] In an embodiment, transcription initiates at a different TSS when initiation occurs with the capped primer as compared to in the absence of the capped primer. For example, when incorporating the capped primer comprising the form $m7GpppAG$, and the TSS has been modified to AG, the T7 RNA polymerase initiates transcription at the A nucleotide of the AG of the TSS region (see Figures 1 and 4). However, when initiating transcription with a nucleoside triphosphate, T7 may preferentially initiate transcription with a guanosine triphosphate. Accordingly, transcription may initiate from the next guanosine nucleoside downstream from the promoter region. In this example, transcription may initiate from the G nucleoside of the AG of the TSS region (see Figures 1 and 5).

[00156] In one example, when the capped primer is $m7GmpppAmG$ (e.g., CleanCap Reagent AG) and the polymerase is T7, the DNA template may be designed to include the promoter sequence of T7 polymerase with an AG at the transcription start site, in order to facilitate specific binding of the capped primer to the transcription start site, i.e., TAATACGACTCACTATAAG (SEQ ID NO: 15).

[00157] In this example, when the capped primer initiates transcription (i.e., co-transcription), the resulting RNA molecule is a capped RNA molecule (i.e., capped with the capped primer). The 5' terminal nucleoside of the capped RNA molecule is adenosine, which is methylated. The 5' terminal nucleotide of the RNA molecule corresponds to the adenosine at a first transcriptional start site of the DNA template (i.e., the +1 A in the AG TSS region). The RNA molecule is longer than an RNA molecule that started at a second (i.e., downstream) transcriptional start site.

[00158] In the same example, when the capped primer does not initiate transcription, T7 preferentially initiates transcription with a guanosine triphosphate. The resulting RNA molecule is uncapped. The 5' terminal nucleoside of the uncapped RNA molecule is guanosine, which is expected to be unmethylated. The 5' terminal nucleotide of the RNA molecule corresponds to the guanosine at a second transcriptional start site of the DNA template (i.e., the +2 G in the AG TSS region in this example). The RNA molecule is shorter than an RNA molecule that started at a first (i.e., upstream) transcriptional start site.

RNA samples

[00159] In an embodiment, the present disclosure provides a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template by an RNA polymerase in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules. In an embodiment, the sample comprises a mixture of capped RNA molecules and uncapped RNA molecules of a single population of RNA molecule having an RNA sequence of interest, transcribed from a corresponding DNA template.

[00160] In an embodiment, the sample comprises more than one population of RNA molecule, wherein each population is transcribed from a different (distinct) DNA template, such that each population of RNA molecule has a distinct RNA sequence of interest. Persons skilled in the art will understand that a sample of RNA molecules comprising a population of RNA molecule, following *in vitro* transcription from a DNA template, will comprise a mixture of capped and uncapped RNA molecules having the distinct RNA sequence of the relevant population.

[00161] In an embodiment, the sample comprises a mixture of capped RNA molecules and uncapped RNA molecules of two populations of RNA molecules, each having an RNA sequence of interest wherein each population is *in vitro* transcribed from its corresponding DNA template. In an embodiment, the sample comprises a mixture of capped RNA molecules and uncapped RNA molecules of three populations of RNA molecules, each having an RNA sequence of interest wherein each population is *in vitro* transcribed from its corresponding DNA template. In an embodiment, the sample comprises a mixture of capped RNA molecules and uncapped RNA molecules of four populations of RNA molecules, each having an RNA sequence of interest wherein each population is *in vitro* transcribed from its corresponding DNA template. In an embodiment, the sample comprises a mixture of capped RNA molecules and uncapped RNA molecules of multiple (e.g., two, three, four, five, six, seven, eight, nine, ten, etc) populations of RNA molecules, each having an RNA sequence of interest wherein each population is *in vitro* transcribed from its corresponding DNA template.

Identifying capped and uncapped RNA molecules

[00162] In an aspect, the present disclosure relates to a method of differentiating a capped RNA molecule from an uncapped RNA molecule following *in vitro* transcription from a DNA by identifying at least one difference between the capped RNA molecule and the uncapped RNA

molecule associated with the differential transcription initiating agent utilised to start transcription (i.e., the capped primer or a nucleotide or analogue thereof) and the associated differential transcription start site.

[00163] In one aspect, the present disclosure provides method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template by an RNA polymerase in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules; wherein the capped RNA molecule has been capped with a capped primer having the general form [cap]-[linker]-N¹[N²]_m[N³]_n; wherein N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8; the method comprising identifying the capped RNA molecule and/or the uncapped RNA molecule from the sample, wherein the identifying comprises a step selected from a group consisting of:

identifying the presence of N¹ as the 5' terminal nucleoside of the RNA molecule, which identifies a capped RNA molecule; and/or identifying the absence of N¹ as the 5' terminal nucleoside which identifies an uncapped RNA molecule;

identifying the 5' terminal nucleoside as corresponding to a first transcription start site of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to a second transcription start site of the DNA template, which identifies an uncapped RNA molecule;

identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +1 position of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +2 position of the DNA template, which identifies an uncapped RNA molecule;

where at least one of N¹, N², and each instance of N³ are independently any modified or unnatural nucleoside, identifying a modified or unnatural nucleoside, which identifies a capped RNA molecule; and/or identifying the absence of a modified or unnatural nucleoside, which identifies an uncapped RNA molecule; and

where the uncapped RNA molecule is x nucleotides long, and x is any suitable integer;

identifying an RNA molecule that is at least x+1 nucleotides long, which identifies a capped RNA molecule; and identifying an RNA molecule that is x nucleotides long, which identifies an uncapped RNA molecule;

wherein identifying step identifies a capped RNA molecule and/or an uncapped RNA molecule.

[00164] In an embodiment, the capped primer may have the general form m7GpppN¹[N²]_m[N³]_n; wherein m7G is N7-methylated guanosine or a guanosine analogue; ppp is a triphosphate; N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8.

[00165] The structural difference may be identified using any suitable method capable of distinguishing or identifying the structural difference known to the persons skilled in the art. In

an embodiment, the method is selected from the group selected from polymerase chain reaction (PCR) based approaches (including Reverse transcription PCR (RT-PCR), Quantitative PCR (qPCR), digital PCR (dPCR), Nested PCR, Multiplex PCR, Touchdown PCR, Hot start PCR, High-fidelity PCR); next-generation sequencing (NGS) approaches (Illumina sequencing, Ion Torrent sequencing, PacBio sequencing, Nanopore sequencing, Sanger sequencing and RNA-seq sequencing); direct RNA sequencing; sequence-by-synthesis; ligation or probe hybridisation methods; RNA-sequencing, RNA ligation, cDNA sequencing, isothermal amplification methods (such as LAMP), and cDNA synthesis with a specific template switching primer or hybridization with a specific oligonucleoside. The person skilled in the art will understand how to design appropriate primers and/or probes and/or optimise reaction conditions in connection with these techniques.

[00166] In an aspect, the method further comprises quantifying the identified capped RNA molecules and/or uncapped RNA molecules. In an embodiment, the quantifying comprises determining the absolute amount of the capped RNA molecules in the sample; determining the relative abundance of the capped RNA molecules and/or uncapped RNA molecules in the sample; and/or determining the relative abundance of the capped RNA molecules relative to total RNA molecules in the sample.

Identifying capped RNA molecules by identifying the 5' terminal nucleoside

Identifying capped RNA molecules by identifying presence or absence of N¹ as the 5' terminal nucleoside

[00167] In one aspect, the present disclosure provides a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template by an RNA polymerase in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules; wherein the capped RNA molecule has been capped with a capped primer having the general form m7GpppN¹[N²]_m[N³]_n; wherein m7G is N7-methylated guanosine or a guanosine analogue; ppp is a triphosphate; N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8; the method comprising identifying the presence of N¹ as the 5' terminal nucleoside of the RNA molecule, which identifies a capped RNA molecule; and/or identifying the absence of N¹ as the 5' terminal nucleoside which identifies an uncapped RNA molecule.

[00168] For example, where the capped primer has the general form m7GpppAN², N¹ may be selected from the group consisting of adenosine, N2-methyladenosine, or N6-methyladenosine, or an adenosine having an alternative modification, and identifying an adenosine as the 5' terminal nucleoside of the RNA molecule would identify a capped RNA molecule. Accordingly, in an embodiment, where N¹ comprises a modified nucleoside, identifying the presence of the base nucleoside (e.g., by sequencing techniques) may identify the presence of N¹ as the 5' terminal nucleoside.

[00169] The presence or absence of N¹ as the 5' terminal nucleoside can be determined using methods known to those skilled in the art, for example, using methods described elsewhere herein.

[00170] In an embodiment, the identifying comprises identifying at least one to three 5' nucleosides of the RNA molecule. In an embodiment, the identifying comprises identifying at least one to ten 5' nucleosides of the RNA molecule. In an embodiment, the identifying comprises identifying at least one to 20 5' nucleosides of the RNA molecule. In an embodiment, the identifying comprises identifying at least one to 50 5' nucleosides of the RNA molecule. In an embodiment, the identifying comprises identifying at least one to fifty 5' nucleosides of the RNA molecule. In an embodiment, the identifying comprises identifying at least one to 100 5' nucleosides of the RNA molecule. In an embodiment, the identifying in step comprises identifying essentially all of the nucleosides in the RNA molecule of the RNA molecule.

Identifying capped RNA molecules by identifying +1 or +2 nucleoside corresponding to the DNA template as the 5' terminal nucleoside

[00171] In an aspect, the present disclosure provides a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template by an RNA polymerase in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules; wherein the capped RNA molecule has been capped with a capped primer having the general form m7GpppN¹[N²]_m[N³]_n; wherein m7G is N7-methylated guanosine or a guanosine analogue; ppp is a triphosphate; N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8; the method comprising identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +1 position of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +2 position of the DNA template, which identifies an uncapped RNA molecule.

[00172] For example, where the DNA template comprises a T7 promoter with the sequence TAATACGACTCACTATAAG (SEQ ID NO: 15), and the capped primer has the general form m7GmpppAG, T7 polymerase is understood to preferentially initiate transcription using the capped primer. The 5' terminal nucleoside of the resulting capped RNA molecule is an adenosine, and the nucleoside at the +1 position of the DNA template is adenosine, and the nucleoside at the +2 position of the DNA template is guanosine. When transcription is initiated with a nucleotide rather than the capped primer, T7 polymerase is understood to preferentially initiate transcription with a guanosine at the G(+2) position (underlined) of the DNA template. In this embodiment, identifying the 5' terminal nucleoside as adenosine corresponds to the nucleoside at the +1 position of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as guanosine corresponds to the nucleoside at the +2 position of the DNA template, which identifies an uncapped RNA molecule.

[00173] In an embodiment, the nucleoside at the +1 position of the DNA template comprises adenosine and N¹ comprises adenosine. In an embodiment, the nucleoside at the +1 position of the DNA template comprises cytidine and N¹ comprises cytidine. In an embodiment, the nucleoside at the +1 position of the DNA template comprises thymidine and N¹ is uridine. In an embodiment, the nucleoside at the +1 position of the DNA template is adenosine and N¹ is selected from the group consisting of adenosine, N2-methyladenosine, or N6-methyladenosine.

Identifying capped RNA molecules by identifying a nucleoside corresponding to the first or second transcription start site of the DNA template as the 5' terminal nucleoside

[00174] In an aspect, the present disclosure provides a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template by an RNA polymerase in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules; wherein the capped RNA molecule has been capped with a capped primer having the general form m7GpppN¹[N²]_m[N³]_n; wherein m7G is N7-methylated guanosine or a guanosine analogue; ppp is a triphosphate; N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8; the method comprising identifying the 5' terminal nucleoside as corresponding to a first transcription start site of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to a second transcription start site of the DNA template, which identifies an uncapped RNA molecule.

[00175] For example, where the DNA template comprises a T7 promoter with the sequence TAATACGACTCACTATA**AG** (SEQ ID NO: 15), and the capped primer has the general form m7GmpppAG, T7 polymerase is understood to preferentially initiate transcription using the capped primer at the first transcription start site (underlined, bold). The 5' terminal nucleoside of the resulting capped RNA molecule is an adenosine, corresponding to a first transcription start site of the DNA template. When transcription is initiated with a nucleotide rather than the capped primer, T7 polymerase is understood to preferentially initiates transcription with a guanosine at the second transcription start site (underlined) of the DNA template. In this embodiment, identifying the 5' terminal nucleoside as adenosine corresponds to first transcription start site of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as guanosine corresponds to the nucleoside at the second transcription start site of the DNA template, which identifies an uncapped RNA molecule.

Identifying capped RNA molecules by length

[00176] In an aspect, the present disclosure provides a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template by an RNA polymerase in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules; wherein the capped RNA molecule has been capped with a capped primer having the general form m7GpppN¹[N²]_m[N³]_n; wherein m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate; N^1 , N^2 , and each instance of N^3 are independently any natural, modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8; the method comprising: where the uncapped RNA molecule is x nucleotides long, and x is any suitable integer; identifying an RNA molecule that is at least $x+1$ nucleotides long, which identifies a capped RNA molecule; and identifying an RNA molecule that is x nucleotides long, which identifies an uncapped RNA molecule.

[00177] For example, where the DNA template comprises a T7 promoter with the sequence TAATACGACTCACTATA**AG** (SEQ ID NO:15), and the capped primer has the general form m7GmpppAG, T7 polymerase is understood to preferentially initiate transcription using the capped primer at the first transcription start site (underlined, bold). When transcription is initiated with a nucleotide rather than the capped primer, T7 polymerase is understood to preferentially initiate transcription with a guanosine at the second transcription start site (underlined) of the DNA template. In this embodiment, if the uncapped RNA molecule is 1000 nucleotides in length, the capped RNA molecule will be 1001 nucleotides in length, as transcription initiated one nucleotide upstream as compared to the uncapped molecule. In this embodiment, identifying an RNA molecule that is 1001 nucleotides long identifies a capped RNA molecule; and identifying an RNA molecule that is 1000 nucleotides long identifies an uncapped RNA molecule. Persons skilled in the art will understand that this embodiment can be adapted to RNA molecules of different lengths. Additionally, in an embodiment, it can be adapted to transcription start sites being non-contiguous. For example, where the second transcription start sites is two nucleotides downstream from the first transcription start site, the capped primer will be two nucleotides longer than the uncapped RNA molecule, and so on.

Identifying capped RNA molecules by nucleotide modification

[00178] In one form, the present disclosure relates to a method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules; wherein the capped RNA molecule has been capped with a capped primer having the general form m7Gppp N^1 [N^2] m [N^3] n ; wherein m7G is N7-methylated guanosine or a guanosine analogue; ppp is a triphosphate; N^1 , N^2 , and each instance of N^3 are independently any natural, modified or unnatural nucleoside wherein at least one of N^1 , N^2 , and each instance of N^3 present in the capped primer comprises any modified or unnatural nucleoside; m is 0 or 1; and n is any integer from 0 to 8; the method comprising: identifying the presence of a modified or unnatural nucleoside identifies a capped RNA molecule; and identifying the absence of a modified or unnatural nucleoside identifies an uncapped RNA molecule. In an embodiment, the modified or unnatural nucleoside is detectable by the methods of the disclosure.

[00179] For example, where the capped primer comprises one or more modified nucleoside(s), for example, a methylated adenosine such as found in 7GmpppAmG, the capped RNA

molecules produced in accordance with the disclosure will comprise the modified nucleoside(s), whereas the uncapped RNA molecules comprises unmodified nucleosides from the transcription reaction. In this embodiment, identifying the presence of a modified or unnatural nucleoside as the 5' terminal nucleoside identifies a capped RNA molecule; whereas identifying the absence of a modified or unnatural nucleoside as the 5' terminal nucleoside identifies an uncapped RNA molecule. Persons skilled in the art will understand that this embodiment of the disclosure may identify modified nucleosides at different positions within the capped primer.

[00180] In an embodiment, identifying the presence of a modified or unnatural nucleoside at the 5' end of the RNA molecule identifies a capped RNA molecule; and identifying the absence of a modified or unnatural nucleoside at the 5' end of the RNA molecule an uncapped RNA molecule. In an embodiment, identifying the presence of a modified or unnatural nucleoside as the 5' terminal nucleoside of the RNA molecule identifies a capped RNA molecule; and identifying the absence of a modified or unnatural nucleoside as the 5' terminal nucleoside of the RNA molecule identified an uncapped RNA molecule. In an embodiment, N¹ is a modified or unnatural nucleoside, and identifying the presence of the modified or unnatural nucleoside as the 5' terminal nucleoside identifies a capped RNA molecule; and identifying the absence of the modified or unnatural nucleoside as the 5' terminal nucleoside identifies an uncapped RNA molecule.

[00181] In an embodiment, the modified or unnatural nucleoside is as described elsewhere herein. In an embodiment, the modified or unnatural nucleoside forms part of the 5' cap structure as described elsewhere herein.

[00182] In an embodiment, the modified or unnatural nucleoside comprises a methyl group at position 6 (m6A). In an embodiment, the modified or unnatural nucleoside comprises adenosine having a methyl group on position 6 (N6-methyladenosine, m6A).

[00183] In an embodiment, the modified or unnatural nucleoside comprises N2-methyladenosine.

[00184] In an embodiment, the modified or unnatural nucleoside comprises a 2'-O-methyl modification (2'Ome, 2'Om).

[00185] In an embodiment, the modified or unnatural nucleoside comprises m7-methyl-3'-O-methyl-guanosine (7mG3'Ome). In an embodiment, the modified or unnatural nucleoside comprises modified 7MeG residue where the 3' and/or 2' position is blocked on the ribose to facilitate directional initiation of *in vitro* transcription.

[00186] In an embodiment, the modified or unnatural nucleoside is selected from N2-methyladenosine or N6-methyladenosine.

[00187] In an embodiment, the method may detect multiple modified or unnatural nucleosides to identify the presence of a capped RNA molecule. In an embodiment, multiple 2'-O-methyl modifications may be detected. In an embodiment, multiple m6A modifications may be detected.

Methods for detecting capped and/or uncapped RNA molecules

[00188] In an aspect of the present disclosure, the identifying of an RNA molecule as a capped or uncapped RNA molecule can be performed using methods known to those skilled in the art. In an embodiment, the identifying comprises a method selected from the group selected from polymerase chain reaction (PCR) based approaches (including Reverse transcription PCR (RT-PCR), Quantitative PCR (qPCR), digital PCR (dPCR), Nested PCR, Multiplex PCR, Touchdown PCR, Hot start PCR, High-fidelity PCR); next-generation sequencing approaches (Illumina sequencing, Ion Torrent sequencing, PacBio sequencing, Nanopore sequencing, Sanger sequencing and RNA-seq sequencing); direct RNA sequencing; sequence-by-synthesis; ligation or probe hybridisation methods; RNA-sequencing, RNA ligation, cDNA sequencing, isothermal amplification methods (such as LAMP), and cDNA synthesis with a specific template switching primer or hybridization with a specific oligonucleoside.

Conversion into complementary DNA molecules

[00189] In certain embodiments, the RNA molecules may be converted into complementary DNA (cDNA) molecules using reverse transcription enzymes (such as MMLV Reverse Transcriptase, Omniscript Reverse Transcriptase (Qiagen), EnzScript Reverse Transcriptase (Qiagen), StableScript Reverse Transcriptase (Qiagen), SuperScript III Reverse Transcriptase (Invitrogen), SuperScript IV Reverse Transcriptase (Invitrogen), ProtoScript II Reverse Transcriptase (NEB), Induro Reverse Transcriptase (NEB), WarmStart Reverse Transcriptase (NEB), AMV Reverse Transcriptase (NEB)). Accordingly, in an embodiment of a method of the disclosure, the method further comprises converting the RNA molecules into cDNA molecules using a reverse transcription enzyme. In an embodiment of the present disclosure, the identifying step further comprises reverse transcribing the RNA molecule from the sample into a complementary DNA (cDNA) and analysing the cDNA molecule to identify a capped or uncapped RNA molecule.

[00190] The conversion of the cDNA molecule may be utilised for further downstream analysis of the RNA molecule capillary-electrophoresis-based, PCR-based or NGS-based approaches.

[00191] In an embodiment, the identifying step further comprises reverse transcribing the RNA molecule from the sample into a complementary DNA (cDNA) and analysing the cDNA molecule to determine the 5' terminal nucleoside of the RNA molecule, and thereby identify a capped RNA molecule or uncapped RNA molecule in accordance with the methods of the disclosure. A person skilled in the art will appreciate that the 5' terminal nucleoside of the RNA molecule can be determined by analysing the sequence of the cDNA molecule. In an embodiment, the identifying step further comprises reverse transcribing an RNA molecule from the sample into a complementary DNA (cDNA) molecule and identifying at least the nucleoside of the cDNA molecule corresponding to the 5' terminal nucleoside of the RNA molecule, wherein identifying at least the nucleoside of the cDNA molecule corresponding to the 5' terminal nucleoside of the

RNA molecule identifies at least the 5' terminal nucleoside of the RNA molecule. In an embodiment, the identifying comprises identifying at least a terminal nucleoside of the cDNA molecule corresponding to the 5' terminal nucleoside of the RNA molecule, wherein identifying at least the terminal nucleoside of the cDNA molecule identifies at least the 5' terminal nucleoside of the RNA molecule.

[00192] In an embodiment, the identifying step further comprises reverse transcribing the RNA molecule from the sample into a complementary DNA (cDNA) and analysing the cDNA molecule to identify the length of the RNA molecule, and thereby identifying a capped RNA molecule or uncapped RNA molecule in accordance with the methods of the disclosure. A person skilled in the art will appreciate that the length of the RNA molecule can be determined by analysing the sequence of the cDNA molecule.

Next Generation Sequencing

[00193] In certain embodiments, the capped or uncapped RNA molecules can be identified using next generation sequencing (NGS) methods. In an embodiment, the output RNA molecules are reversed transcribed into cDNA, which may then undergo library preparation and sequencing conducted following the manufacturer's instructions for the specific NGS platform being used. These methods include but are not limited to, Illumina sequencing, Ion Torrent sequencing, PacBio sequencing, Nanopore sequencing, Sanger sequencing, direct RNA sequencing, RNA sequencing and cDNA sequencing.

[00194] In certain embodiments, the presence of the cap may be determined using direct RNA nanopore sequencing such as those provided by Oxford Nanopore Technologies. This method allows for the direct analysis of the RNA molecules (with or without a cap) without the requirement for reverse transcription into cDNA.

[00195] The use of next-generation sequencing (NGS) techniques, such as Illumina, Pacific Biosciences (PacBio), and Oxford Nanopore Technologies (ONT) may be used to determine the sequence of RNA molecules. This facilitates the identification of nucleotides that are at the 5' termini of the RNA molecules. Upon sequencing the heterogeneous sample of capped and uncapped RNA molecules, the read data sets can be bioinformatically aligned with the reference DNA template sequence. By analysing the nucleotides at the 5' end of the RNA molecules (e.g. the +1, +2, +3, +4, +5, etc nucleotides) of the sequenced read alignments, the TSS can be identified. This may identify whether the capped primer was incorporated into the RNA molecule and, accordingly, whether the RNA molecule is capped or uncapped in accordance with the methods described herein.

[00196] In an embodiment, NGS can detect the presence of modifications of nucleotides at the 5' end of the RNA molecules (e.g., the cap-proximal terminal positions, including the +1, +2 and +3 nucleotides, etc). The detection of nucleotide modifications can be performed using numerous methods, including direct RNA sequencing, cDNA sequencing, and mass

spectrophotometry methods. For example, direct RNA sequencing can be used to detect the m⁷GTP cap on the 5' terminal of the mRNA molecule. Direct RNA sequencing can also be used to detect an N⁶-methylation of adenosine at the first cap-proximal nucleotide (i.e., 5' terminal nucleoside), thereby indicating the incorporation and presence of a capped primer, such as m⁷Gpppm⁶A. In an embodiment, direct RNA sequencing can identify the presence of N⁶-methyladenosine at transcript position 1, a 2'O-methyl nucleoside residue at transcript position 1 (Cap1), presence of 2'O-methyl nucleoside residue at transcript position 2 (Cap2), presence of 6'O-methyl nucleoside residue at transcript position 1 (Cap1), and/or the presence of 6'O-methyl nucleoside residue at transcript position 2 (Cap2).

[00197] In an embodiment, NGS analysis can be used to detect the length of the RNA molecule, which can identify whether the RNA molecule is capped or uncapped, as disclosed herein.

Quantitative Polymerase Chain Reaction

[00198] In an embodiment, the 5' terminal nucleotide and capping status of an RNA molecule can be analysed using polymerase chain reaction (PCR) based methods. The PCR based methods may be selected from the group consisting of Reverse Transcription Polymerase Chain Reaction (RT-PCR), Real-Time Quantitative RT-PCR (qRT-PCR), TaqMan qPCR, Multiplex RT-PCR, digital RT-PCR (dRT-PCR), Nested RT-PCR, RT-Loop-Mediated Isothermal Amplification (RT-LAMP), RT-Droplet Digital PCR (RT-ddPCR), RT-PCR followed by High-Resolution Melt (HRM) Analysis etc.

[00199] These methods may identify the 5' nucleotide of RNA molecules (e.g. in a synthetic mRNA vaccine) and thereby identify the presence or absence of N¹ as the 5' terminal nucleotide, identify whether the 5' terminal nucleotide corresponds to the first transcription start site or the second transcription start site, and/or identify whether the 5' terminal nucleoside as corresponds to the nucleoside at the +1 position of the DNA template or the +2 position of the DNA template, and/or another position on the DNA template. In an embodiment, this in turn identifies whether the RNA molecule is capped or uncapped. This approach is generally applicable to any PCR or probe hybridisation technique that can sufficiently measure differences in the terminal nucleotide, or length of the mRNA molecules.

[00200] In certain embodiments, the relative abundance of RNA molecules containing a presence or absence of an adaptor oligonucleotide can be measured using polymerase chain reaction (PCR) based approaches (including Reverse transcription PCR (RT-PCR), Quantitative PCR (qPCR), digital PCR (dPCR), Nested PCR, Multiplex PCR, Touchdown PCR, Hot start PCR, High-fidelity PCR). PCR requires short oligonucleotide primers that anneal and amplify the sequence or sequences of interest.

[00201] In an embodiment, the 5' terminal nucleotide and capping status can be determined by Taqman probe assays. In an embodiment, the Taqman assay comprises the use of probes and

or primers selected from the group consisting of a Taqman forward primer according to SEQ ID NO: 1, a Taqman probe for identifying capped RNA molecules according to SEQ ID NO: 2, a Taqman reverse primer according to SEQ ID NO: 3 and a Taqman probe for identifying uncapped RNA molecules according to SEQ ID NO: 4. Persons skilled in the art will appreciate that other primers and probes may be suitable for use in a Taqman probe assay.

[00202] In an embodiment, the 5' terminal nucleotide and capping status can be determined by a SYBR green assay. In an embodiment, the SYBR green assay comprises the use of primers selected from the group consisting of SYBR forward primer for detecting capped molecules according to SEQ ID NO: 5, a forward primer for detecting uncapped molecules according to SEQ ID NO: 6, a first reverse primer according to SEQ ID NO: 7, a second reverse primer according to SEQ ID NO: 8, a third reverse primer according to SEQ ID NO: 9, and a fourth reverse primer according to SEQ ID NO: 10. Persons skilled in the art will appreciate that other primers and probes may be suitable for use in a SYBR green assay.

Quantifying capped and/or uncapped RNA molecules

[00203] In an aspect, the present disclosure provides methods for quantifying the capped or uncapped RNA molecules identified as described herein.

[00204] In an embodiment, the quantifying comprises determining the absolute amount of the capped RNA molecules in the sample. In an embodiment, the quantifying comprises determining the relative abundance of the capped RNA molecules and/or uncapped RNA molecules in the sample. In an embodiment, the quantifying comprises determining the relative abundance of the capped RNA molecules relative to total RNA molecules in the sample.

[00205] In an embodiment, the quantifying comprises a comparison of the quantity of RNA molecules identified as capped RNA molecules with the quantity of RNA molecules identified as uncapped RNA molecules to determine the relative abundance of the capped RNA molecules in the sample.

[00206] In an embodiment, the method of quantifying of the identified capped and/or uncapped RNA molecules utilises suitable methods for quantifying identified RNA molecules as known to those skilled in the art. In an embodiment, the method of quantifying the identified RNA molecules comprises the detection methods as described herein. In an embodiment, the quantifying comprises a method selected from the group consisting of polymerase chain reaction (PCR) based approaches (including Reverse transcription PCR (RT-PCR), Quantitative PCR (qPCR), digital PCR (dPCR), Nested PCR, Multiplex PCR, Touchdown PCR, Hot start PCR, High-fidelity PCR); next-generation sequencing approaches (Illumina sequencing, Ion Torrent sequencing, PacBio sequencing, Nanopore sequencing, Sanger sequencing and RNA-seq sequencing); direct RNA sequencing; sequence-by-synthesis; oligonucleotide probe hybridisation methods; RNA-sequencing, RNA ligation, cDNA sequencing, isothermal amplification methods (such as LAMP), Digestion with Restriction enzymes and cDNA synthesis

with a specific template switching primer or hybridization with a specific oligonucleoside. In an embodiment, the method of quantifying the identified RNA molecules comprises utilises next generation sequencing methods, Taqman probe assays, SYBR green assays and/or qRT-PCR assays.

Bioinformatics analysis

[00207] Analysis of NGS data sets may determine the amount of capped and uncapped mRNA molecules in a sample. By measuring the count of the identified capped or uncapped nucleotides, the proportion of capped and uncapped RNA molecules in the RNA sample can be quantified. The first step may involve the analysis of the output data sets, e.g. in FASTQ format, that indicate the sequenced reads. These reads may be aligned using software (such as minimap2, bwa Bowtie2, HISAT2, or STAR) to a reference sequence, that may include the plasmid DNA, the DNA template sequence, the RNA molecule sequence, and/or the adaptor oligonucleotide sequences. Comparison between the aligned reads and the reference sequence may indicate differences caused by mutations and errors. Alternatively, k-mer counting approaches (such as Kallisto or Salmon) can be used to count the occurrence of k-mer sequences (short DNA or RNA sequences of length k) within the sequencing data that correspond to either capped or uncapped mRNA. For example, measuring the count of k-mer sequences that start from either the +1 or +2 nucleotide at the 5' termini of the RNA molecules can indicate the count of capped or uncapped mRNA molecules, respectively. The counting of k-mer sequences matching or complementary to identified RNA molecules can indicate determine the abundance of capped and/or uncapped RNA molecules.

Measuring other mRNA quality features

[00208] In an embodiment, the analysis of mRNA capping status using the methods described above can be performed simultaneously for measuring other mRNA quality features using a single assay. For example, the analysis of mRNA capping status can be performed simultaneously in addition to one or more other analyses including, mRNA intactness, sequence identity, polyA tail, and contamination. This method enables the analysis of many different mRNA quality features using a single, streamlined method. When used with single-molecule sequence approaches, this method enables the capping status to be compared to other quality features for individual molecules. In an embodiment, this method enables the analysis of capping status for different mRNA sequences that together comprise a multivalent pharmaceutical composition.

Identifying capped RNA molecules in multivalent RNA samples

[00209] The methods of the present disclosure can be adapted to detect capped RNA molecules from multiple populations of RNA with an RNA sample. For example, multiple distinct primers may be required to detect multiple, different RNA molecules within a single sample by PCR-based

methods (e.g., a multiplex PCR amplification) that may be used during library preparation for next-generation sequencing, or for PCR-based detection methods, such as qRT-PCR, ddPCR etc.

Pharmaceutical compositions

[00210] In an embodiment of the disclosure, the RNA molecule identified using the methods disclosed herein is intended for use as a vaccine molecule, a scientific or research reagent, or a therapeutic reagent in a pharmaceutical composition.

[00211] In an embodiment, the present disclosure provides a pharmaceutical composition comprising the RNA molecule identified by the methods of the present disclosure, and optionally a pharmaceutically acceptable carrier. In an embodiment, the pharmaceutical composition comprises a capped RNA molecule identified by the methods disclosed herein. In an embodiment, the pharmaceutical composition comprises an uncapped RNA molecule identified by the methods disclosed herein.

[00212] In some examples, the present disclosure provides a multivalent pharmaceutical composition comprising a plurality of populations of capped RNA molecules identified by the methods described herein. It will be understood that the plurality of RNA molecules in a multivalent pharmaceutical composition have different sequences and can therefore serve as templates for the translation of different polypeptides. In an embodiment, the pharmaceutical composition of the present disclosure comprises different populations of RNA molecules identified by the method of the present disclosure, wherein each population is encoded by a distinct RNA sequence of interest. In an embodiment, the pharmaceutical composition comprises two or more different populations of RNA molecules. In an embodiment, the pharmaceutical composition is a divalent formulation comprising two populations of RNA molecules. In an embodiment, the pharmaceutical composition is a trivalent composition comprising three populations of RNA molecules. In an embodiment, the pharmaceutical composition is a quadrivalent composition comprising four populations of RNA molecules. In an embodiment, the pharmaceutical composition is a multivalent composition comprising multiple populations of RNA molecules. For example, the composition may comprise one, two, three, four, five, six, seven, eight, nine or ten, etc, populations of RNA molecule.

[00213] In one embodiment, the method of the disclosure can determine the abundance of capped and uncapped mRNA molecules for each population that together comprise a multivalent composition. In another embodiment, the method of the present disclosure can be utilised to quantify the capped RNA molecules for each population of RNA of interest individually, and then, optionally, different populations of RNA molecules may be combined into a single multivalent pharmaceutical composition.

[00214] The compositions may comprise a pharmaceutically acceptable carrier, excipient, diluent and/or adjuvant. Pharmaceutically acceptable carriers, excipients, diluents and/or adjuvants as contemplated herein are substances which do not produce adverse reaction(s)

when administered to a particular recipient such as a human or non-human animal.

Pharmaceutically acceptable carriers, excipients, diluents and adjuvants are generally also compatible with other components of the composition.

[00215] In an embodiment, the pharmaceutical composition further comprises a lipid nanoparticle.

[00216] In an embodiment, the present disclosure provides a method of prophylactically or therapeutically treating a disease comprising administering a therapeutically effective amount of the RNA molecule identified by the methods of the present disclosure to a subject in need of treatment.

[00217] In an embodiment, the present disclosure provides a use of the RNA molecules identified by the methods of the present disclosure for use in the manufacture of a medicament for the prophylactic or therapeutic treatment of a disease.

[00218] In an embodiment, the present disclosure provides a RNA molecules identified by the methods of the present disclosure for use the prophylactic or therapeutic treatment of a disease.

EXAMPLES

[00219] Nucleic acid sequences relevant to the present disclosure are listed in Table 5.

Table 5: Polynucleotide sequences – primers and probes

SEQ ID No.	Description	Sequence
1	TaqMan_Fwd1_primer	CTAATCATTGCAAGCAGTGGTATC
2	TaqMan_cap1_probe	ACGCAGAGTACATGGGAGGACTCT
3	TaqMan_uncap1_probe	ACGCAGAGTACATGGGGGACTCT
4	TaqMan_Rev1_primer	CATGGTGGCGGGTTCTC
5	Fwd1_SYBR_cap	GTGGTATCAACGCAGAGTACATGGGA
6	Fwd2_SYBR_uncap	GTGGTATCAACGCAGAGTACATGGGG
7	Rev1_SYBR	GCGGGTTCTCTCTGAGTCTGT
8	Rev2_SYBR	CTGAGTCTGTGGGGACCAGAAG
9	Rev3_SYBR	CGCTGAACTTGTGGCCGTTT
10	Rev4_SYBR	TGAACAGCTCCTCGCCCTTG
11	Fwd1_eGFP_poscon (Leveque-Sergerie 2007)	CTGCTGCCCCGACAACCAC
12	Rev1_eGFP_poscon (Leveque-Sergerie 2007)	TGTGATCGCGCTTCTCGTT
29	TSO primer	GCTAATCATTGCAAGCAGTGGTATCA ACGCAGAGTACATrGrGrG
30	Oligo dT(20) primer	TTTTTTTTTTTTTTTTTTTT

Table 6: Polynucleotide sequences – promoter sequences

SEQ ID No.	Description	Sequence
13	Native T7 promoter	TAATACGACTCACTATAGG
14	Modified T7 promoter (HG) where H = A, C or T	TAATACGACTCACTATAHG
15	Modified T7 promoter (AG)	TAATACGACTCACTATAAG
16	Native T3 promoter	AATTAACCCTCACTAAAG
17	Modified promoter where H = A, C or T	TAATACGACTCACTAAAH
18	Native SP6 promoter	ATTTAGGTGACACTATAGA
19	Modified SP6 promoter (AH) where H = A, C or T	ATTTAGGTGACACTATAHA
20	Modified SP6 promoter where H = A, C or T	ATTTAGGTGACACTATAH
21	Native self-amplifying RNA promoter	TAATACGACTCACTATAGG
22	Modified self-amplifying RNA promoter where H = A, C or T	TAATACGACTCACTATAHT
23	Promoter sequence	TAATACGACTCACTATA
24	Promoter sequence	TAATACGACTCACTATAAT
25	Promoter sequence	AATTAACCCTCACTAAA
26	Promoter sequence	AATTAACCCTCACTAAAG
27	Promoter sequence	AATTAACCCTCACTATA
28	Promoter sequence	ATTTAGGTGACACTATA

Example 1: Identifying capped molecules using next generation sequencing

[00220] The DNA template with artificial T7 promoter used in Example 1 is provided by SEQ ID NO: 31. DNA template with native T7 promoter used in Example 1 is provided by SEQ ID NO: 32. Expected RNA transcript in Example 1 for capped RNA transcribed from artificial T7 promoter is provided by SEQ ID NO: 33. Expected RNA transcript in Example 1 for uncapped RNA transcribed from artificial T7 promoter is provided in SEQ ID NO: 34. Expected RNA transcript in Example 1 for uncapped RNA transcribed from native T7 promoter is provided in SEQ ID NO: 35.

In vitro transcription of capped and uncapped mRNA

[00221] mRNA with modified nucleotides was produced by *in vitro* transcription (IVT) using T7 RNA polymerase following protocols described in Henderson *et al.* (2021) and according to the manufacturer's instructions (NEB, E2080S). Firstly, plasmid DNA that encoded the following components; a modified T7 promoter (i.e. A+1;G+2), a modified alpha-globin 5' UTR, an open reading frame encoding green fluorescent protein (GFP), a modified alpha-globin 3' UTR, and a segmented poly(A) tail (Trepotec *et al.* 2019) followed by a restriction digestion site (BsaI) was

prepared. As a control, we also generated a matching plasmid DNA that encodes the native T7 promoter (i.e., G+1; G+2), with a single nucleotide difference at the transcription start site (from A+1 to a G+1).

[00222] Plasmid DNA templates were linearised and purified and used as template for an IVT reaction at 32°C for 3 hr with 16 µg/mL T7 RNA polymerase (NEB M0251), ribonucleotides (6 mM ATP, 5 mM CTP, 5 mM GTP; NEB, N0450), 5 mM N1-methylpseudouridine-5'- Triphosphate (TriLink BioTechnologies, TRN108110), or 5 mM UTP for matched unmodified controls, transcription buffer (40 mM Tris-HCl pH 8.0, 16.5 mM magnesium acetate, 10 mM dithiothreitol (DTT), 20 mM spermidine, 0.002% (v/v) Triton X-100), 2 U/mL yeast Inorganic pyrophosphatase (NEB, M2403) and 1000 U/mL murine Rnase inhibitor (NEB, M0314).

[00223] For AG capped mRNA, Cap1 analogue was co-transcriptionally incorporated to the mRNA 5' end by addition of 4 mM CleanCap AG reagents (TriLink, TRN711310) to the reaction. The uncapped mRNA was *in vitro* transcribed in the absence of any synthetic cap analogue. The mRNA IVT reaction was stopped by addition of 200 units DnaseI (NEB, M0303) per mL of IVT reaction and incubation at 37°C for 15 min.

[00224] For enzymatical capping, the mRNA was prepared from purified, uncapped eGFP mRNA. Briefly, 10 µg of uncapped mRNA was combined with 1x FCE capping buffer, 0.2 mM SAM, 0.5 mM GTP, 25 U Faustovirus Capping Enzyme (FCE, NEB M2081) and 100 U Cap2'-O-me (2'-O-me, NEB M0366). The reaction was incubated at 37°C for 60 min. The capped mRNA was then cleaned up using the Zymo clean and concentrator– 5kit (Zymo R1013) and eluted in 15 µL of nuclease-free water. The prepared capped mRNA samples (derived from either co-transcriptional or enzymatic capping) were then prepared for library preparation and sequencing as described above.

[00225] Enzymatic decapping of either purified co-transcriptionally capped mRNA, or enzymatically capped eGFP mRNA was performed as follows. Briefly, we used mRNA Decapping Enzyme (NEB M0608S) according to manufacturer's instructions. Briefly the mRNA Decapping Enzyme was combined with Reaction Buffer (10x), mRNA vaccine sample, and nuclease-free water, and incubated for 37°C for 30 minutes. The prepared de-capped samples (derived from either co-transcriptional or enzymatic capped mRNA) were then prepared for library preparation and sequencing as described above.

[00226] The resulting mRNAs were purified using Monarch RNA cleanup kits (NEB, T2050) according to the manufacturer's instructions with the final elution in distilled ultrapure water (ThermoFisher Scientific, 10977015). The yield, length and purity of the IVT mRNAs was evaluated using a range of different analytical methods. mRNA was quantified by UV spectrophotometry analysis using a NanoPhotometer N120 (Implen) and the size distribution was evaluated using an TapeStation electrophoresis with RNA ScreenTapes (Agilent Technologies, USA, 5067-5576).

Nanopore library preparation and detecting using next generation sequencing

[00227] cDNA-PCR sequencing was used to determine the accuracy and purity of the *in vitro* transcribed mRNAs. First, mRNA concentration was calculated using the Qubit RNA BR kit (ThermoFisher Scientific). mRNAs were diluted in nuclease free water to an appropriate concentration for library preparation (~1ng/μL), and concentrations were confirmed using a Qubit RNA HS kit (ThermoFisher). Barcoded ONT cDNA-PCR libraries (SQK-PCS111.24) were prepared (see Figure 6A and 6B) according to manufacturer's instructions (Oxford Nanopore Technologies), with the following exceptions. Evaporation during the cDNA synthesis step was identified by measuring reaction volume, and tubes were topped up with nuclease-free water where appropriate. The cDNA was amplified for 14-16 cycles (recommendation is 14-18 cycles). Finally, libraries were eluted in 8 μL of Elution Buffer (rather than the recommended 12 μL volume) to boost the final concentration of the libraries. This was beneficial, as libraries prepared with templates containing modified bases appear to produce lower output libraries than those prepared with unmodified bases.

[00228] The resulting libraries were quantified via a Qubit instrument (Invitrogen) with the dsDNA HS kit and qualitative analysis of fragment length distribution was conducted using D5000 ScreenTapes (Agilent Technologies, USA). The results of the quantitative and qualitative analysis were used to adjust library concentrations for pooling and loading. Barcoded libraries were sequenced on R9.4.1 (FLO-MIN106D) Flow Cells, with High Accuracy live base-calling enabled (Guppy v5.1.13 and MinKNOW Core 4.5.4). All nanopore reads with a quality score greater than 9 were allocated as passed and proceeded to further analysis.

Bioinformatic Analysis

[00229] Oxford Nanopore pDNA sequencing data were mapped and analysed using a custom pipeline. The quality-filtered, concatenated FASTQ reads were aligned to the plasmid reference via Minimap2 (Release 2.20-r1064) with -ax map-ont for Nanopore³⁴. The resulting SAM alignment files were processed via SAMtools v1.15 to generate sorted and indexed BAM files, as well as various other mapping analysis files³⁵. The generated BAM files were viewed and analysed utilising Integrative Genomics Viewer (IGV v2.12.3)³⁶. Further run and sample quality statistics were acquired via NanoPlot v1.38.1 and pycoQC v2.5.237,38.

Results and Discussion

[00230] Next-generation sequencing was used to measure the 5' terminal nucleotide of the mRNA and thereby determine 5' capping status. An mRNA vaccine molecule was transcribed from the artificial T7 promoter in the presence of the CleanCap reagent AG cap-analogue wherein transcription initiates at the first TSS nucleotide (i.e. A (+1)) nucleotide (Figure 6A, upper panel) or in the absence of any synthetic Cap Analogue wherein transcription initiates at the second TSS nucleotide (i.e. G (+2) nucleotide) (Figure 6A, lower panel, indicated). For

comparison, the transcription of the mRNA vaccine molecule initiates from the first TSS nucleotide (i.e. G (+1) nucleotide from the native T7 promoter).

[00231] As shown in Figure 6C, analysis of next generation sequencing indicated that the 5' terminal nucleotide of capped mRNA molecules is adenosine, corresponding to transcription initiation at the first TSS nucleotide (i.e. +1A), and that the 5' terminal nucleotide of uncapped mRNA molecules is guanosine, corresponding to transcription initiation at the second TSS nucleotide (i.e. +2G).

[00232] This confirmed that the detection of 5' terminal nucleotides of capped RNA molecules and uncapped RNA molecules can be detected and quantified. It also confirmed that transcription initiated for capped RNA molecules at the first TSS (A +1) and at the second TSS (G+2) for uncapped RNA molecule as shown in Figure 9.

[00233] As shown in Figure 7, a genome browser view displays alignments derived from sequenced mRNA molecules that have been aligned to the reference plasmid sequence. It shows that mRNA incorporating the co-transcriptional 5' cap structure initiates from the +1 A nucleotide (upper panel), whereas mRNA that does not incorporate the co-transcriptional 5' cap structure initiates from the +2 G nucleotide (middle panel). Meanwhile, mRNA molecules that incorporated the co-transcriptional 5' cap structure which had then been subsequently enzymatically decapped also start at the +1 A nucleotide (lower panel). While not wanting to be bound by theory, this may indicate that the +1 A nucleotide represents a different transcription initiation start site, rather than direct detection of the 5' cap itself by the next-generation sequencing method.

[00234] As shown in Figure 8, a genome browser view displays alignments derived from sequenced mRNA molecules that have been aligned to the reference plasmid sequence. It shows that mRNA lacking the co-transcriptional 5' cap structure initiate transcription from the +2 G nucleotide (top panel). However, if a 5' cap is enzymatically added post-transcriptionally to the mRNA molecule, an impact on sequencing alignments is not observed. While not wanting to be bound by theory, this may indicate that the sequencing method does not directly detect the 5' cap. Additionally, if this enzymatically added 5' cap is subsequently removed with enzymatic decapping, no impact on the alignments is observed. This may indicate that the sequencing method does not directly detect the 5' cap. Together, this supports the interpretation that the change in the 5' alignment start observed during the sequencing method is not due to direct detection of the 5' cap, but instead reflects a difference in transcription initiation preference by the RNA polymerase in the presence of co-transcriptional 5' cap structure.

[00235] As shown in Figure 10, a substantial proportion of capped RNA molecules were 1187 nucleotides long, whereas a substantial proportion of uncapped RNA molecules were 1186 nucleotides long. This confirms that the difference in transcriptional start sites can be detected and quantified by analysing the length of the RNA molecule.

[00236] Figure 11 shows that there is a direct correlation between the proportion of RNA molecules having an adenosine as the 5' terminal nucleoside and the proportion of capped molecules in a sample, providing further evidence that identifying the nucleotide corresponding to the first transcription start site identifies a capped RNA molecule.

[00237] The incorporation of the capped primer during transcription initiation impacts the TSS selected by the RNA polymerase. For example, when incorporating the m7GpppAG capped primer, the T7 RNA Polymerase initiates at the A (+1) nucleotide (see Figure 1 and 4). This preference is due to the promoter sequence and RNA polymerase favouring the incorporation of the dinucleotide m7GpppAG analogue with greater efficiency than other adenine, thymine, uracil, guanine nucleotides otherwise present in the reaction mixture.

[00238] Conversely, during *in vitro* transcription, wherein no synthetic capped primer is present, the RNA polymerase prefers to initiate transcription from a different nucleotide. For example, in the absence of any synthetic capped primer, the T7 RNA Polymerase prefers to initiate transcription at the G (+2) nucleotide (see Figure 1 and 5). Notably, this preferred transcription start site is more similar to the native promoter sequence of the T7 RNA Polymerase (GG) (see Figures 1 and 3).

[00239] Therefore, the 5' terminal nucleoside of a synthetic mRNA molecule (e.g. an RNA vaccine or therapy) indicates the transcription start site. The choice of transcription start site by the RNA polymerase indicates whether or not a synthetic capped primer was incorporated during transcription initiation. Therefore, by measuring the 5' terminal nucleotide of an mRNA molecule, it is possible to determine whether a synthetic capped primer has been incorporated during transcription initiation. For example, when using a capped primer having the form m7GpppAmG, by determining whether an mRNA has a 5' terminal A(+1) ribonucleotide or G(+2) ribonucleotide (corresponding to the transcript numbering of the capped RNA molecule as shown in Figure 1), it is possible to determine whether transcription is initiated from the A (i.e. first TSS) or G (i.e. second TSS), and thereby determine whether or not a synthetic m7GpppAG cap has been incorporated during *in vitro* transcription by T7 RNA Polymerase. Given that G is often the TSS nucleotide preferred by RNA Polymerases, this approach is optimised wherein guanosine is not the first cap-proximal nucleoside within the capped primer, and wherein the first TSS in the modified T7 promoter is not guanosine. This approach can be adapted to detect incorporation of other suitable cap analogues wherein the +1 nucleotide is a non-preferred initiator nucleotide.

[00240] In addition, the synthetic capped primer may include nucleoside modifications that are incorporated into the 5' termini of the mRNA. For example, the first and second cap-proximal nucleotides can be methylated to improve performance and reduce innate immune recognition. These include, m7Gpppm2AN, m7Gpppm2Am2N, m7Gpppm6AN, and m7Gpppm6Am2N. The detection of these nucleotide modifications on the 5' termini of the mRNA molecules indicate the

correct incorporation of the synthetic capped primer during co-transcriptional initiation.

Conversely, the absence of nucleotide modifications at the 5' termini of the mRNA molecule indicates the absence of incorporation of synthetic capped primer.

[00241] Determining the presence of the 5' cap is a key feature in mRNA manufacture and quality control. By measuring the 5' termini nucleotide of the synthetic mRNA vaccine or therapy, it is possible to determine whether the mRNA molecule has successfully incorporated a capped primer. By determining the quantity of mRNA initiating from different TSS, it is possible to determine the fraction of synthetic mRNA vaccines that incorporate a 5' cap, and the fraction of uncapped mRNA, within a mixed sample. This provides an accurate, quantitative method to indicate the fraction of capped mRNA molecules within a sample and measure the efficiency of 5' capping during mRNA manufacture.

[00242] There are numerous methods to measure the terminal nucleotides of the mRNA vaccine or therapy, including but not limited to next-generation sequencing, PCR amplification, ligation or probe hybridisation methods. The identity of the first nucleotide can be detected using various methods, including RNA-sequencing, RNA-ligation, cDNA-sequencing, PCR-based techniques, isothermal amplification methods (such as LAMP), cDNA synthesis with a specific template switching primer or hybridization with a specific oligonucleotide.

[00243] The use of next-generation sequencing (NGS) techniques such as Illumina, Pacific Biosciences (PacBio), and Oxford Nanopore Technologies (ONT) can be used to determine the sequence of mRNA vaccine samples. This permits the identification of ribonucleotides that are at the 5' termini of the mRNA molecules (Figure 6). Upon sequencing the heterogenous sample, the read data sets can be bioinformatically aligned with the reference DNA template sequence. By analysing the +1, +2 and +3 nucleotides of the sequenced read alignments, the TSS is identified which can determine whether a capped primer (synthetic 5' cap Analog) was correctly incorporated during transcription initiation.

[00244] The mRNA vaccine sample was analysed using full-length complementary DNA (cDNA) sequencing. In this case, the mRNA molecules were first converted to cDNA using a reverse transcription reaction before the ligation of library adaptors and sequencing using the Oxford Nanopore Technology (GridION; Figure 6). Sequenced read data sets were then processed (including trimming of library adaptors) and aligned to the reference DNA template. The analysis of the sequenced read alignments allows us to determine the 5' terminal (+1) nucleotide for the mRNA molecules within a sample, indicating the presence or absence of a 5' cap Analog on the RNA molecule (Figure 3). Kits including NGS reagents are required to detect initiating nucleotide.

[00245] NGS can provide a quantitative measure of the fraction of capped and uncapped mRNA molecules in a mixed sample. For example, by comparing the relative fraction of mRNA molecules initiating with A (+1), compared to mRNA molecules starting with the G (+2), the

fraction of capped and uncapped mRNAs in a mixture, respectively (Figure 11). The fraction of mRNA molecules can indicate the efficiency of capping steps during mRNA manufacture.

Example 2: Identifying capped molecules using Taqman Probe and SYBR green based detection methods

Preparing cDNA from mRNA molecules

[00246] The sequences for the primers and probes used in this Example can be found in Table 5. Double stranded cDNA was synthesised from capped and uncapped mRNAs encoding the GFP coding region, that were *in vitro* transcribed as described in Example 1. Briefly, 3.5 µg RNA was combined with 1 µM Oligo(dT)20 (Invitrogen, 18418020) and 1 mM dNTPs, and incubated at 70°C for 5 min, then held at 4°C until the following step. Next, 1x Template Switching RT buffer, 3.75 µM Template Switching Oligo (TSO) and 1x Template Switching RT enzyme mix (NEB M0466) were added for first strand cDNA synthesis. The Template Switching Oligo added a unique, 39 nucleotide (nt) adaptor to the 5' end of each cDNA molecule. The reaction was incubated at 42°C for 90 min, 5 min at 85°C, then held at 4°C until the following step. The single stranded cDNA was then combined with 1x Q5 Hot Start High Fidelity Master Mix (NEB M0494) and 25 U *E. coli* RNase H (NEB #M0523). The RNA was then hydrolysed, through incubating at 37°C for 15 min, and second strand synthesis proceeded through incubation at 95°C for 1 min and 65°C for 10 min. Double stranded cDNAs were stored at -20°C.

Design of qPCR primers

[00247] Primers and probes were designed to quantify capped and uncapped mRNAs, using both (1) TaqMan probes, and (2) SYBR green based detection methods.

1) TaqMan probe assay

[00248] A TaqMan probe assay was designed to quantify the relative proportions of capped and uncapped mRNAs in two single-plexed qRT-PCR reactions. A single primer pair was designed to amplify all eGFP mRNAs (regardless of capping status), while two TaqMan probes were designed with different fluorophores to differentially label capped and uncapped mRNAs. The forward (Fwd) primer was designed to bind the 5' adaptor sequence introduced by the Template Switching Oligo, which is upstream of the mRNA 5' terminus (Figure 12). The Reverse primer is template specific, flanking the human alpha globin 5' UTR, KOZAK region and the eGFP coding region. When used together in a PCR, they were expected to generate an 87 nucleotide PCR product. The two Taqman probes were designed to flank the TSS of the capped (i.e. A (+1)) and uncapped (i.e., G (+2)) mRNAs, the 5' adaptor introduced by the Template Switching Oligo and the human alpha globin 5' UTR. The capped mRNA detection probe was labelled with a HEX fluorophore and was designed to overlap the TSS (i.e. A (+1)) introduced during the 5' capping reaction. The alternative TSS (i.e., G (+2)) of uncapped mRNAs means that the cap

probe had a mismatch for uncapped mRNAs and was not expected to anneal stably during qRT-PCR. The uncapped mRNA detection probe, labelled with a FAM fluorophore, lacked the capped mRNA TSS (i.e., A (+1)). As the uncapped probe does not incorporate the capped mRNA TSS (A (+1)), it was not expected to bind stably to capped mRNAs during qRT-PCR.

(2) SYBR green assay

[00249] A SYBR green assay was also designed to quantify the relative proportions of capped and uncapped mRNAs, in two single-plexed qRT-PCR reactions. Two Fwd primers were designed to differentially bind capped and uncapped mRNAs (Figure 13). The two primers differed only in the 3' terminal nucleotides, with the cap-detecting Fwd primer (Fwd1_SYBR_cap) terminating in GGGA, and the uncap detecting Fwd primer (Fwd2_SYBR_uncap) lacking the A, but rather terminating in GGGG. These primers are predominantly complementary to the TSO adaptor sequence, with the exception of complementarity to the TSS of the capped mRNA (i.e., A (+1)), or the TSS of the uncapped mRNA (i.e., G (+2)). The two Fwd primers were designed to be used with a single Rev primer, designed either the human alpha-globin 5' UTR, KOZAK or eGFP coding region. After testing four Reverse primers, the Rev3_SYBR was identified as producing the most reliable results with the Fwd1_SYBR_cap and Fwd2_SYBR_uncap primers, according to melt curve and Efficiency curve analyses. When used in a PCR, Fwd1_SYBR_cap and Rev3_SYBR produce a 156 nucleotide PCR product, and PCR, Fwd2_SYBR_uncap and Rev3_SYBR produce a 155 nucleotide PCR product. All presented SYBR capping assay results include these two primer combinations. The expression of the cap dependent SYBR green primers was compared to a cap-independent positive control eGFP primer pair (Fwd1_eGFP_poscon and Rev1_eGFP_poscon (Leveque-Serve *et al.*, 2007)). The control eGFP primers acted as an internal control, similar to a housekeeping control gene.

qRT-PCR assay of samples with known capping status

[00250] The SYBR green qRT-PCR assay was tested on mixtures of capped and uncapped eGFP cDNAs, to test whether the calculated proportions of capped and uncapped cDNAs matched the cDNA inputs. The cDNA mixtures included: 1) 100% capped eGFP cDNA, 2) 95% capped, 3) 75% capped, 4) 50% capped, 5) 0% capped. The capped and uncapped PCR primers were amplified in separate reactions, alongside a positive control primer pair, which was independent of capping status. All three primer pairs were amplified using the following reaction conditions: 95°C for 3 min, then 40 cycles of 95°C for 15 sec and 68.5°C for 30 sec. A melt curve was then performed to confirm that each well amplified a single product. The Ct of each reaction was compared to that of the positive control primer, using the delta Ct method. The delta Ct method involves normalisation against a control sample. The 100% capped sample was used, so the delta Ct values are expressed as a ratio to the 100% capped sample.

[00251] Next, the TaqMan qRT-PCR assay was tested on mixtures of capped and uncapped cDNAs, using the cDNA inputs described above. The capped and uncapped PCR probes were assayed with a single primer pair. Capped and uncapped probes were used in two separate reactions, and amplified using the following reaction conditions: 95°C for 3 min, then 40 cycles of 95°C for 15 sec and 60°C for 30 sec. The Ct of each reaction was compared to that of the positive control primer (described above), using the delta Ct method. The delta Ct method involves normalisation against a control sample. The 100% capped sample was used, so the delta Ct values are expressed as a ratio to the 100% capped sample.

[00252] The determination of the 5' terminal nucleotide and capping status of an mRNA can be analysed using polymerase chain reaction (PCR) based methods, for example, Reverse Transcription Polymerase Chain Reaction (RT-PCR), Real-Time Quantitative RT-PCR (qRT-PCR), TaqMan qPCR, Multiplex RT-PCR, Digital RT-PCR (dRT-PCR), Nested RT-PCR, RT-Loop-Mediated Isothermal Amplification (RT-LAMP), RT-Droplet Digital PCR (RT-ddPCR), RT-PCR followed by High-Resolution Melt (HRM) Analysis etc. This approach measures the 5' nucleotide of synthetic mRNA vaccines and thereby indicates the TSS and whether a 5' Cap Analog has been incorporated during transcription initiation. This approach is generally applicable to any PCR- or probe hybridisation technique that can sufficiently measure differences in the terminal nucleotide, or length of the mRNA molecules. Kits including qPCR reagents are required to detect initiating nucleotide.

[00253] A TaqMan probe assay was designed to quantify the relative proportions of capped and uncapped mRNAs in a multiplexed qRT-PCR reaction. In this case, whether mRNA vaccines synthesised in the presence of m7Gpppm2A started with either the A (+1) or G (+2) was investigated to determine whether m7Gpppm2A was incorporated or not, respectively. By designing probes to bind mRNA vaccines starting with either the A (+1) or G (+2) nucleotides, either capped or uncapped mRNAs molecules was detected, respectively (Figure 4).

[00254] When expected and calculated capping proportions were plotted for the SYBR green qRT-PCR assay, an R^2 value of 0.8 was observed, demonstrating that the calculated proportions of capped and uncapped cDNA reflect the capping proportion in the input sample (Figure 14A).

[00255] When expected and calculated capping proportions were plotted for the TaqMan qRT-PCR assay, an R^2 value of 0.7711 was observed, confirming that the calculated proportions of capped and uncapped cDNA reflect the capping proportion in the input sample (Figure 14B).

[00256] These results confirm that the presence of the additional adenosine present due to transcription initiation from the first TSS found within the capped RNA molecules can be used to distinguish and quantify capped and uncapped RNA molecules using qRT-PCR and TaqManqRT-PCR assays.

[00257] By determining mRNA 5' terminal nucleotide using PCR-based methods, the fraction of capped or uncapped mRNA molecules in a mixed sample was determined. By measuring the fraction of mRNA molecules starting with an A (+1) or G (+2) nucleotide, the fraction of capped and uncapped mRNA molecules in a mixed sample, respectively, can be determined. By comparing the fraction of mRNA molecules starting with an A (+1) or G (+2) nucleotide compared to a reference standard mixture with known proportions of capped and uncapped mRNA, the absolute fraction of capped or uncapped mRNA molecules in a test sample can be determined.

FORMS OF THE DISCLOSURE

[00258] Form 1: A method of differentiating a capped RNA molecule from an uncapped RNA molecule in a sample comprising a mixture of capped RNA molecules and uncapped RNA molecules *in vitro* transcribed from a DNA template by an RNA polymerase in a sample of RNA molecules;

wherein the capped RNA molecule has been capped with a capped primer having the general form [cap]-[linker]-N¹[N²]_m[N³]_n; wherein

N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

- a) identifying the capped RNA molecule and/or the uncapped RNA molecule from the sample, wherein the identifying comprises a step selected from a group consisting of:
 - identifying the presence of N¹ as the 5' terminal nucleoside of the RNA molecule, which identifies a capped RNA molecule; and/or identifying the absence of N¹ as the 5' terminal nucleoside which identifies an uncapped RNA molecule;
 - identifying the 5' terminal nucleoside as corresponding to a first transcription start site of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to a second transcription start site of the DNA template, which identifies an uncapped RNA molecule;
 - identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +1 position of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +2 position of the DNA template, which identifies an uncapped RNA molecule;
 - where at least one of N¹, N², and each instance of N³ are independently any modified or unnatural nucleoside, identifying the presence of a modified or unnatural nucleoside at the 5' end of the RNA molecule identifies a capped RNA molecule; and identifying the

absence of a modified or unnatural nucleoside at the 5' end of the RNA molecule an uncapped RNA molecule; and

where the uncapped RNA molecule is x nucleotides long, and x is any suitable integer; identifying an RNA molecule that is at least x+1 nucleotides long, which identifies a capped RNA molecule; and identifying an RNA molecule that is x nucleotides long, which identifies an uncapped RNA molecule;

wherein the identifying step differentiates a capped RNA molecule from an uncapped RNA molecule.

[00259] Form 2: A method of identifying a capped RNA molecule and/or an uncapped RNA molecule *in vitro* transcribed from a DNA template by an RNA polymerase in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules; wherein the capped RNA molecule has been capped with a capped primer having the general form [cap]-[linker]-N¹[N²]_m[N³]_n; wherein

N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

a) identifying the capped RNA molecule and/or the uncapped RNA molecule from the sample, wherein the identifying comprises a step selected from the group consisting of:

identifying the presence of N¹ as the 5' terminal nucleoside of the RNA molecule, which identifies a capped RNA molecule; and/or identifying the absence of N¹ as the 5' terminal nucleoside which identifies an uncapped RNA molecule;

identifying the 5' terminal nucleoside as corresponding to a first transcription start site of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to a second transcription start site of the DNA template, which identifies an uncapped RNA molecule;

identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +1 position of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +2 position of the DNA template, which identifies an uncapped RNA molecule;

identifying a modified or unnatural nucleoside, which identifies a capped RNA molecule; and/or identifying the absence of a modified or unnatural nucleoside, which identifies an uncapped RNA molecule; where at least one of N¹, N², and each instance of N³ are independently any modified or unnatural nucleoside; and

where the uncapped RNA molecule is x nucleotides long, and x is any suitable integer; identifying an RNA molecule that is at least $x+1$ nucleotides long, which identifies a capped RNA molecule; and identifying an RNA molecule that is x nucleotides long, which identifies an uncapped RNA molecule;

wherein:

the identifying step identifies a capped RNA molecule and/or an uncapped RNA.

[00260] Form 3: A method of identifying capped RNA molecule and/or an uncapped RNA molecule in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules from multiple populations of RNA molecules, wherein each of the multiple populations of RNA molecules is *in vitro* transcribed from a DNA template by an RNA polymerase;

wherein the capped RNA molecules have been capped with a capped primer having the general form $[\text{cap}]-[\text{linker}]-\text{N}^1[\text{N}^2]_m[\text{N}^3]_n$; wherein

N^1 , N^2 , and each instance of N^3 are independently any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

a) identifying the capped RNA molecules and/or the uncapped RNA molecules from the sample, wherein the identifying comprises a step selected from the group consisting of:

identifying the presence of N^1 as the 5' terminal nucleoside of the RNA molecule, which identifies a capped RNA molecule; and/or identifying the absence of N^1 as the 5' terminal nucleoside which identifies an uncapped RNA molecule;

identifying the 5' terminal nucleoside as corresponding to a first transcription start site of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to a second transcription start site of the DNA template, which identifies an uncapped RNA molecule;

identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +1 position of the DNA template, which identifies a capped RNA molecule; and/or identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +2 position of the DNA template, which identifies an uncapped RNA molecule;

identifying a modified or unnatural nucleoside, which identifies a capped RNA molecule; and/or identifying the absence of a modified or unnatural nucleoside, which identifies an uncapped RNA molecule; where at least one of N^1 , N^2 , and each instance of N^3 are independently any modified or unnatural nucleoside; and

where the uncapped RNA molecule is x nucleotides long, and x is any suitable integer; identifying an RNA molecule that is at least $x+1$ nucleotides long, which identifies a capped RNA molecule; and identifying an RNA molecule that is x nucleotides long, which identifies an uncapped RNA molecule;

wherein the identifying step identifies a capped RNA molecule and/or an uncapped RNA molecule.

[00261] Form 4: The method of any one of the preceding forms further comprising

b) quantifying the identified capped RNA molecules and/or uncapped RNA molecules.

[00262] Form 5: The method of any one of the preceding forms wherein the quantifying in step b) comprises:

determining the absolute amount of the capped RNA molecules in the sample;

determining the relative abundance of the capped RNA molecules and/or uncapped RNA molecules in the sample;

determining the relative abundance of the capped RNA molecules relative to total RNA molecules in the sample.

[00263] Form 6: The method of form 3 further comprising

b) quantifying the identified capped RNA molecules and/or uncapped RNA molecules;

wherein the quantifying in step b) comprises:

determining the absolute amount of the capped RNA molecules in the sample;

determining the relative abundance of the capped RNA molecules and/or uncapped RNA molecules in the sample;

determining the relative abundance of the capped RNA molecules relative to total RNA molecules in the sample;

determining the absolute amount of the capped RNA molecules and/or uncapped RNA molecules of each of the multiple populations of RNA molecules in the sample;

determining the relative abundance of the capped RNA molecules and/or uncapped RNA molecules of each of the multiple populations of RNA molecules in the sample;

determining the relative abundance of the capped RNA molecules relative to total RNA molecules of each of the multiple populations of RNA molecules in the sample; and/or

determining the relative abundance of capped RNA molecules and/or uncapped RNA molecules of each of the multiple populations of RNA molecules.

REFERENCES

- [00264] Beverly, Michael, Amy Dell, Parul Parmar, and Leslie Houghton. 2016. "Label-Free Analysis of mRNA Capping Efficiency Using RNase H Probes and LC-MS." *Analytical and Bioanalytical Chemistry* 408 (18): 5021–30.
- [00265] Galloway, Alison, Abdelmadjid Atrih, Renata Grzela, Edward Darzynkiewicz, Michael A. J. Ferguson, and Victoria H. Cowling. 2020. "CAP-MAP: Cap Analysis Protocol with Minimal Analyte Processing, a Rapid and Sensitive Approach to Analysing mRNA Cap Structures." *Open Biology* 10 (2): 190306.
- [00266] Grudzien, E. *et al.*, *RNA*, 10: 1479-1487 (2004).
- [00267] Grudzien-Nogalska, E., *et al.*, *RNA*, 13: 1745-1755 (2007).
- [00268] Henderson, Jordana M., Andrew Ujita, Elizabeth Hill, Sally Yousif-Rosales, Cory Smith, Nicholas Ko, Taylor McReynolds, Charles R. Cabral, Julianne R. Escamilla-Powers, and Michael E. Houston. 2021. "Cap1 Messenger RNA Synthesis with Co-Transcriptional CleanCap® Analogue by in Vitro Transcription." *Current Protocols* 1 (2): e39.
- [00269] Hornung, Veit, Jana Ellegast, Sarah Kim, Krzysztof Brzózka, Andreas Jung, Hiroki Kato, Hendrik Poeck, *et al.* 2006. "5'-Triphosphate RNA Is the Ligand for RIG-I." *Science (New York, N. Y.)* 314 (5801): 994–97.
- [00270] Jemielity, J. *et al.*, "Novel 'anti-reverse' cap analogues with superior translational properties", *RNA*, 9: 1108-1122 (2003)
- [00271] Levesque-Sergerie, JP., Duquette, M., Thibault, C. *et al.* Detection limits of several commercial reverse transcriptase enzymes: impact on the low- and high-abundance transcript levels assessed by quantitative RT-PCR. *BMC Molecular Biol* 8, 93 (2007).
- [00272] Trepotec, Z. *et al.*, "Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life." *RNA* 25(4):507-518 (2019)
- [00273] Analytical Procedures for mRNA Vaccine Quality- Draft Guidelines, USP, 2023

CLAIMS

1. A method of identifying capped RNA molecules *in vitro* transcribed from a DNA template by an RNA polymerase in a sample comprising a mixture of capped RNA molecules and uncapped RNA molecules;
wherein the capped RNA molecules have been capped with a capped primer having the general form $m7GpppN^1[N^2]_m[N^3]_n$; wherein

m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate;

N^1 , N^2 , and each instance of N^3 are independently any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

a) identifying at least the 5' terminal nucleoside of a plurality of RNA molecules from the sample by:

identifying the presence of N^1 as the 5' terminal nucleoside and then identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +1 position of the DNA template, which identifies capped RNA molecules; and/or

identifying the absence of N^1 as the 5' terminal nucleoside and then identifying the 5' terminal nucleoside as corresponding to the nucleoside at the +2 position of the DNA template, which identifies uncapped RNA molecules,

wherein:

the nucleoside at the +1 position of the DNA template comprises adenosine and N^1 comprises adenosine;

the nucleoside at the +1 position of the DNA template comprises cytidine and N^1 comprises cytidine; or

the nucleoside at the +1 position of the DNA template comprises thymidine and N^1 is uridine; and

b) quantifying the presence of N^1 and thereby quantifying the capped RNA molecules in the sample.

2. The method of claim 1, wherein the nucleoside at the +1 position of the DNA template is adenosine and N^1 is selected from the group consisting of adenosine, N2-methyladenosine, or N6-methyladenosine, and the nucleoside at the +2 position of the DNA template comprises guanosine and N^2 comprises guanosine.
3. The method of claim 1 or claim 2, wherein the quantifying in step b) comprises:
determining the absolute amount of the capped RNA molecules in the sample;

determining the abundance of the capped RNA molecules relative to uncapped RNA molecules in the sample; and/or

determining the abundance of the capped RNA molecules relative to total RNA molecules in the sample.

4. The method of any one of the preceding claims wherein N¹ is not a preferred initiation nucleoside for the RNA polymerase.
5. The method of any one of the preceding claims, wherein step a) further comprises identifying the length of the RNA molecule, wherein:
 - the uncapped RNA molecule is x nucleotides long, and
 - x is any suitable integer; and
 wherein:
 - identifying an RNA molecule that is at least x+1 nucleotides long identifies a capped RNA molecule; and
 - identifying an RNA molecule that is x nucleotides long identifies an uncapped RNA molecule.
6. The method of any one of the preceding claims, wherein N¹ is a modified or unnatural nucleoside, and step a) further comprises identifying the modified or unnatural nucleoside, wherein:
 - identifying the presence of the modified or unnatural nucleoside as the 5' terminal nucleoside identifies a capped RNA molecule; and
 - identifying the absence of the modified or unnatural nucleoside as the 5' terminal nucleoside identifies an uncapped RNA molecule.
7. The method of any one of the preceding claims, wherein m = 1 and n = 0.
8. The method of any one of the preceding claims, wherein the capped primer is selected from the group consisting of 5'm7GpppAN, m7Gpppm2AN, m7Gpppm2Am2N, m7Gpppm6AN, and m7Gpppm6Am2N, m7G(5')ppp(5')G, m7(3')-O-Me-G(5')ppp(5')G), m7G(5')ppp(5')(2'OMeA)pG and m7G(5')ppp(5')(2'OMeA)pU), m7(3'OMeG)(5')ppp(5')m6(2'OMeA)pG), m7Gpppm6A mN, m7GpppN mN, and 7mG3OMe5'ppp5'N.
9. The method of any one of the preceding claims, wherein the RNA molecules have been transcribed by T7 RNA polymerase.
10. The method of claim 9, wherein the DNA template comprises a sequence selected from the group consisting of TAATACGACTCACTATA (SEQ ID NO: 23), TAATACGACTCACTATAAG (SEQ ID NO: 15), and TAATACGACTCACTATAAT (SEQ ID NO: 24).
11. The method of any one of the preceding claims, wherein the identifying of step a) further comprises

reverse transcribing an RNA molecule from the sample to form a complementary DNA (cDNA) molecule, and

identifying at least a terminal nucleoside of the cDNA molecule wherein identifying the terminal nucleoside of the cDNA molecule identifies the at least 5' terminal nucleoside of the RNA molecule.

12. The method of any one of the preceding claims, wherein the identifying in step a) comprises identifying at least one to three 5' nucleosides of the RNA molecule, identifying at least one to ten 5' nucleosides of the RNA molecule, or identifying essentially all of the nucleosides of the RNA molecule.

13. A method of quantifying a capped RNA molecule *in vitro* transcribed from a DNA template in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules;

wherein the capped RNA molecule has been capped with a capped primer having the general form $m7GpppN^1[N^2]_m[N^3]_n$; wherein

m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate;

N1 is a modified or unnatural nucleoside, N2 is a modified or unnatural nucleoside, and N3 is any natural, modified or unnatural nucleoside;

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

identifying whether at least the 5' terminal nucleoside of a plurality of RNA molecules from the sample is a natural, modified or unnatural nucleoside;

wherein:

identifying the presence of a modified or unnatural nucleoside as the 5' terminal nucleoside identifies a capped RNA molecule; and

identifying the absence of a modified or unnatural nucleoside as the 5' terminal nucleoside identifies an uncapped RNA molecule; and

quantifying the presence of a modified or unnatural nucleoside as the 5' terminal nucleoside and thereby quantifying the capped RNA molecules in the sample.

14. A method of identifying a capped RNA molecule *in vitro* transcribed from a DNA template in a sample of RNA molecules comprising a mixture of capped RNA molecules and uncapped RNA molecules;

wherein the capped RNA molecule has been capped with a capped primer having the general form $m7GpppN^1[N^2]_m[N^3]_n$; wherein

m7G is N7-methylated guanosine or a guanosine analogue;

ppp is a triphosphate;

N¹ is any natural, modified or unnatural nucleoside wherein N¹ is not a preferred initiation nucleoside for the RNA polymerase;

N¹, N², and each instance of N³ are independently any natural, modified or unnatural nucleoside; and

m is 0 or 1; and

n is any integer from 0 to 8;

the method comprising:

identifying the length of the RNA molecule, wherein:

the uncapped RNA molecule is x nucleotides long, and

x is any suitable integer; and

wherein:

identifying an RNA molecule that is at least x+1 nucleotides long identifies a capped RNA molecule; and

identifying an RNA molecule that is x nucleotides long identifies an uncapped RNA molecule.

15. A pharmaceutical composition comprising a capped RNA molecule identified by the method of any one of the preceding claims.

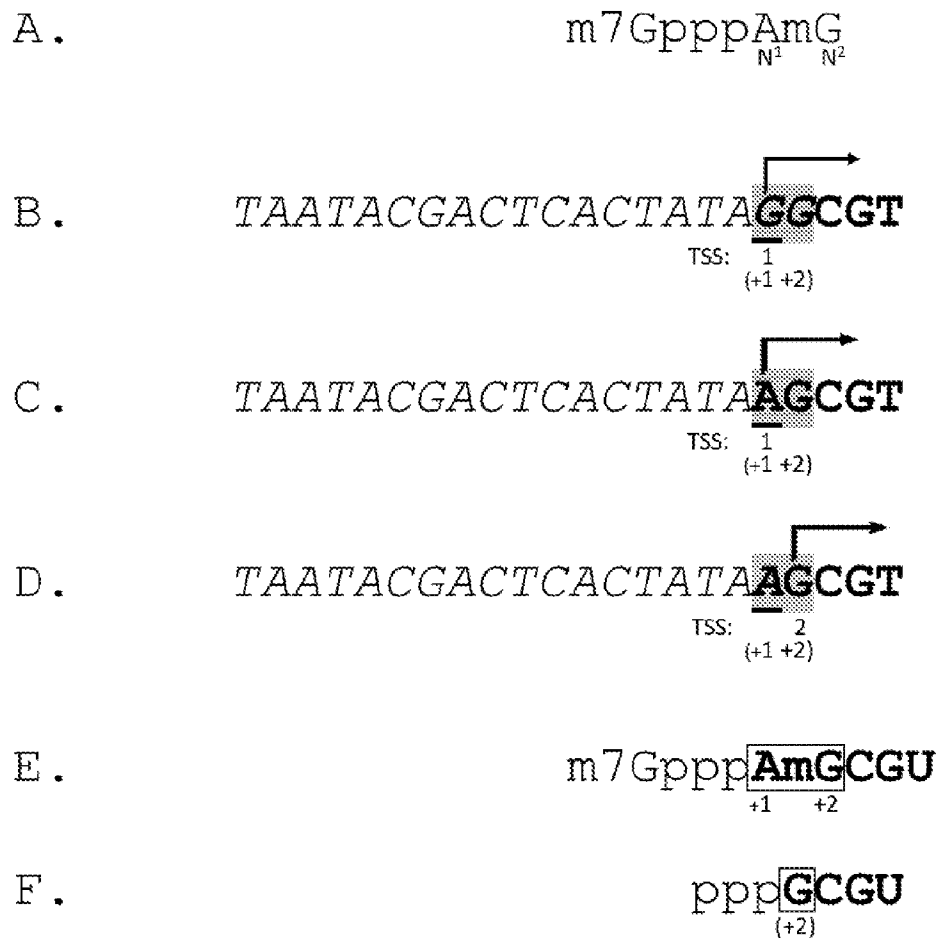


FIGURE 1

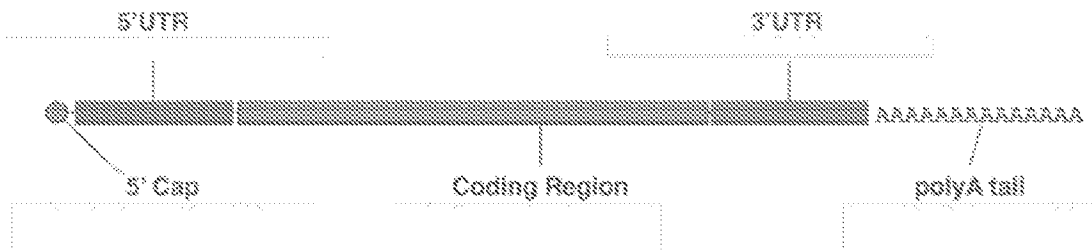


FIGURE 2

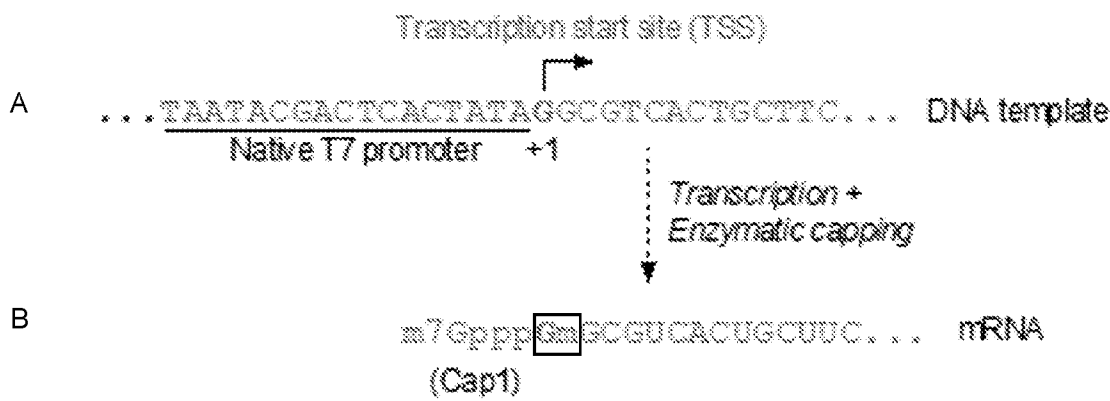


FIGURE 3

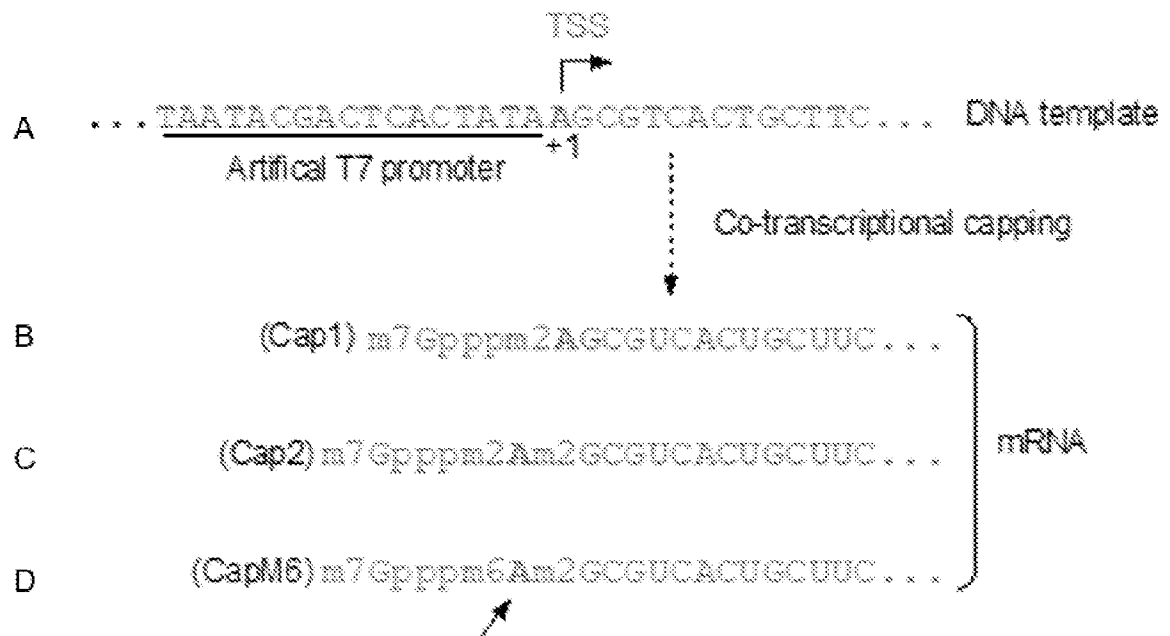


FIGURE 4

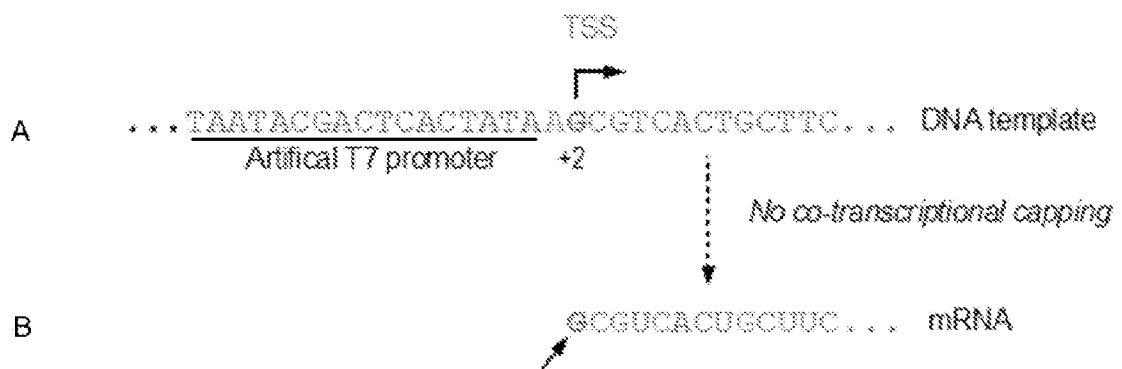


FIGURE 5

Capped mRNAs m7GpppAGCGUCACUGCUUC . . .

Uncapped mRNAs GCGUCACUGCUUC . . .

FIGURE 6A

. . . UACGGACGAUACGCGUCACUGCUUC . . . mRNA
Library Adaptor . . . UACGGACGAUACGCGUCACUGCUUC . . .

FIGURE 6B

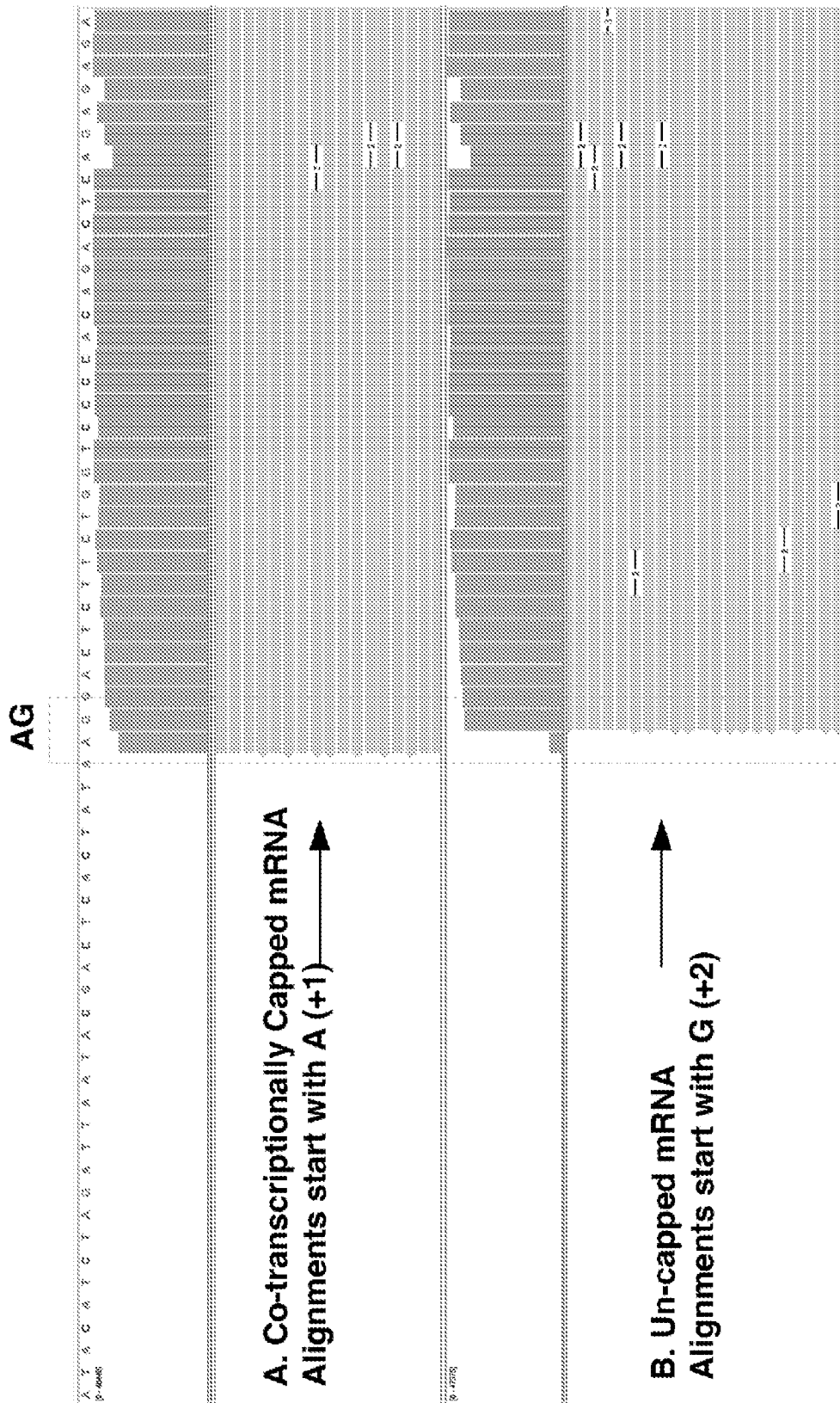


FIGURE 6C

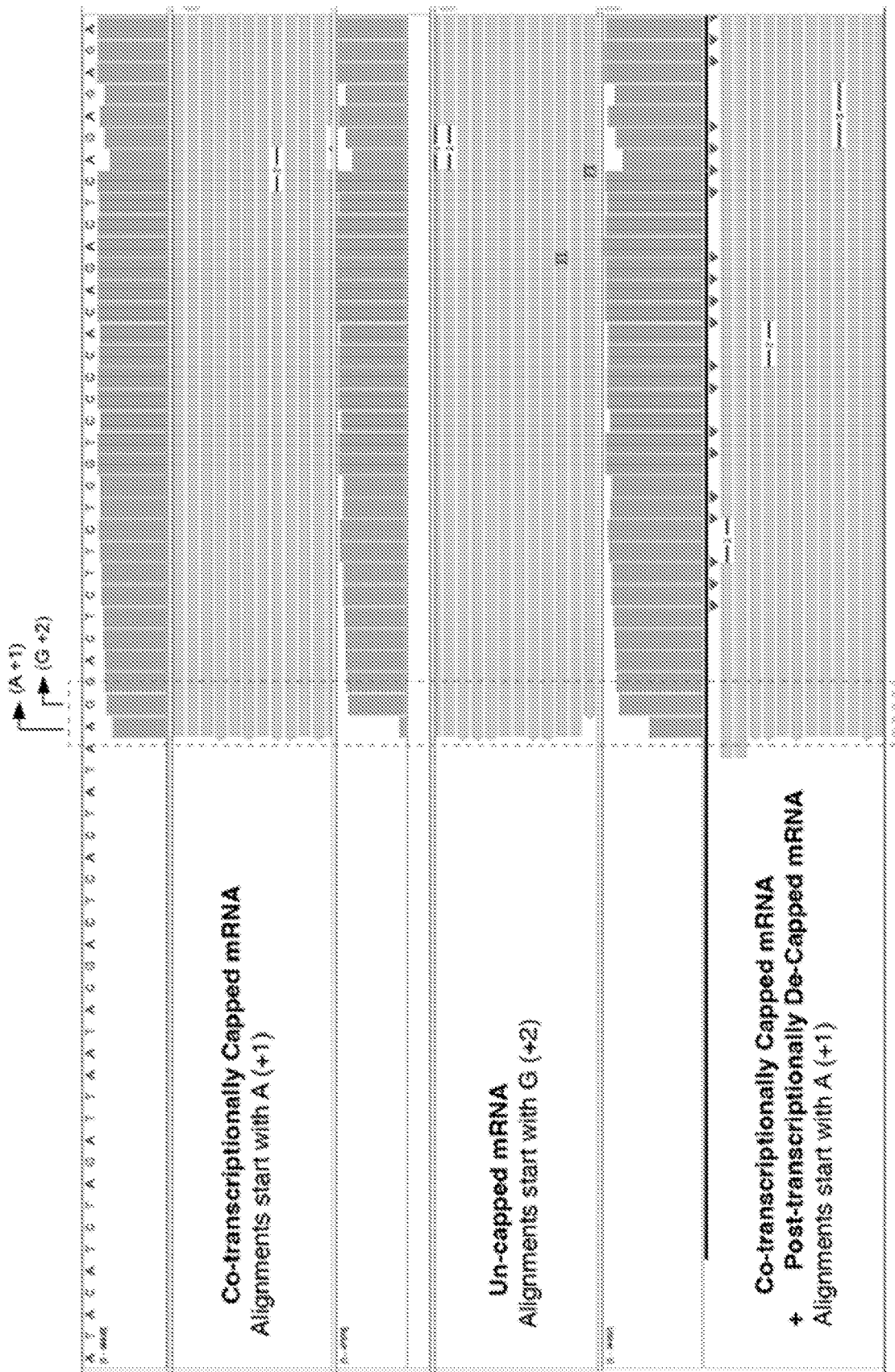


FIGURE 7

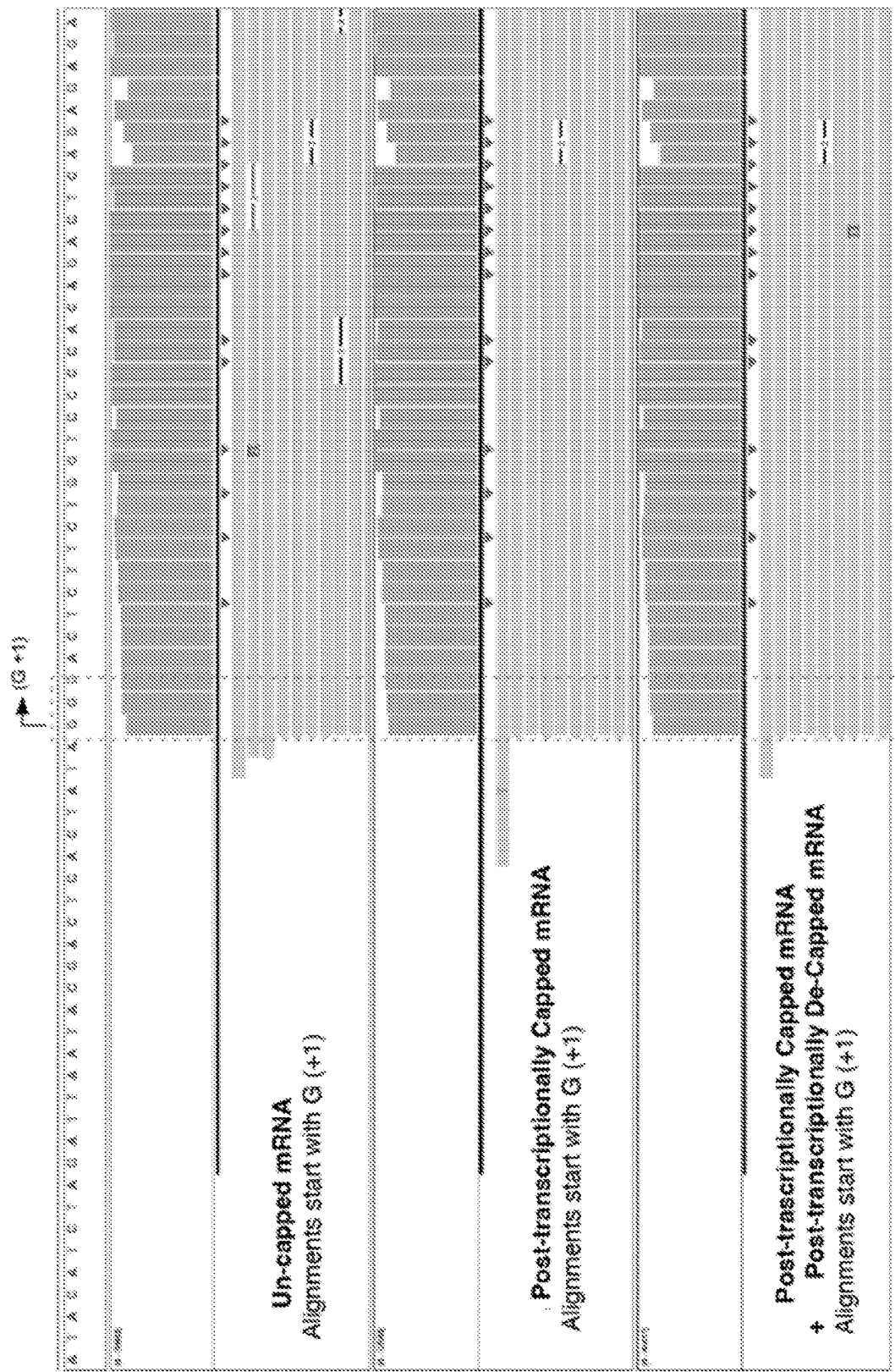


FIGURE 8

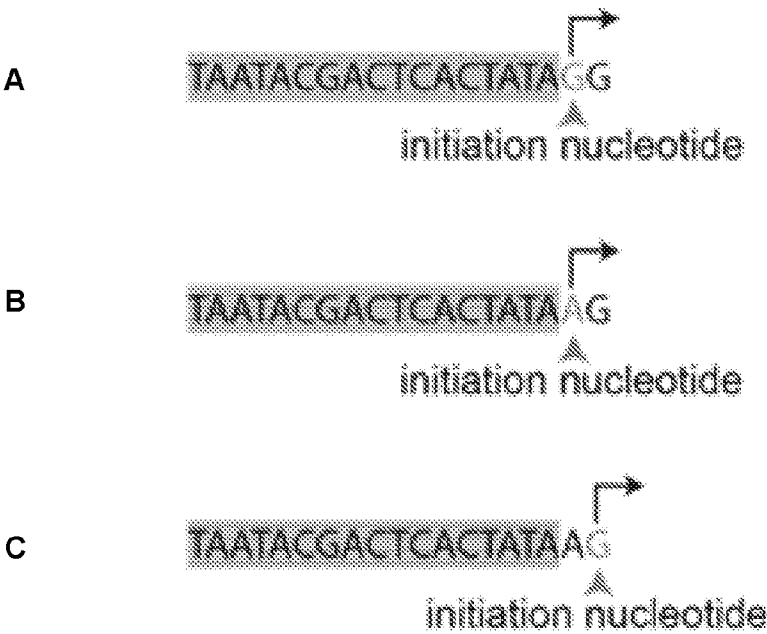


FIGURE 9

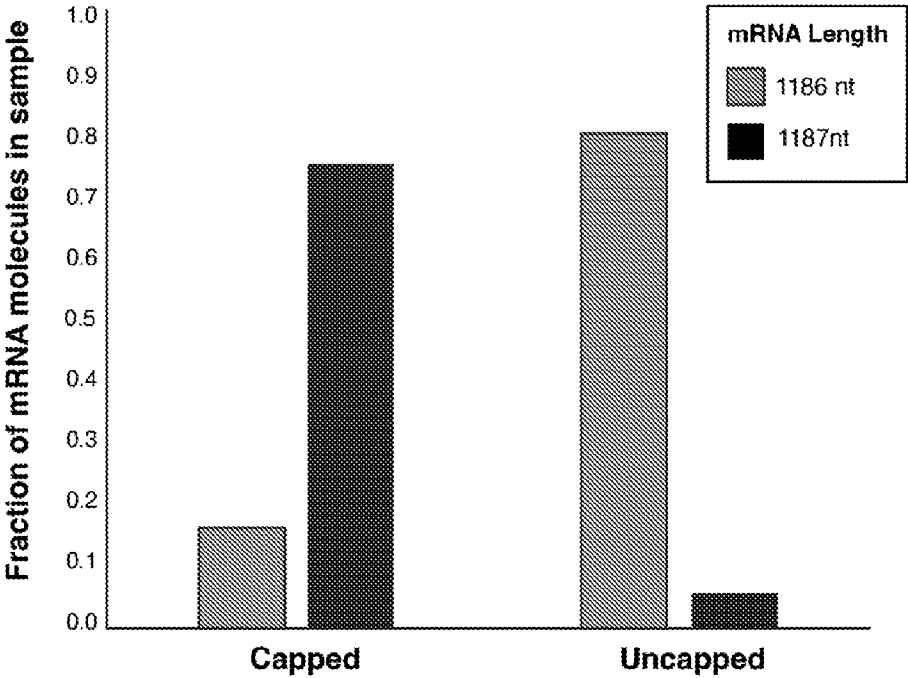


FIGURE 10

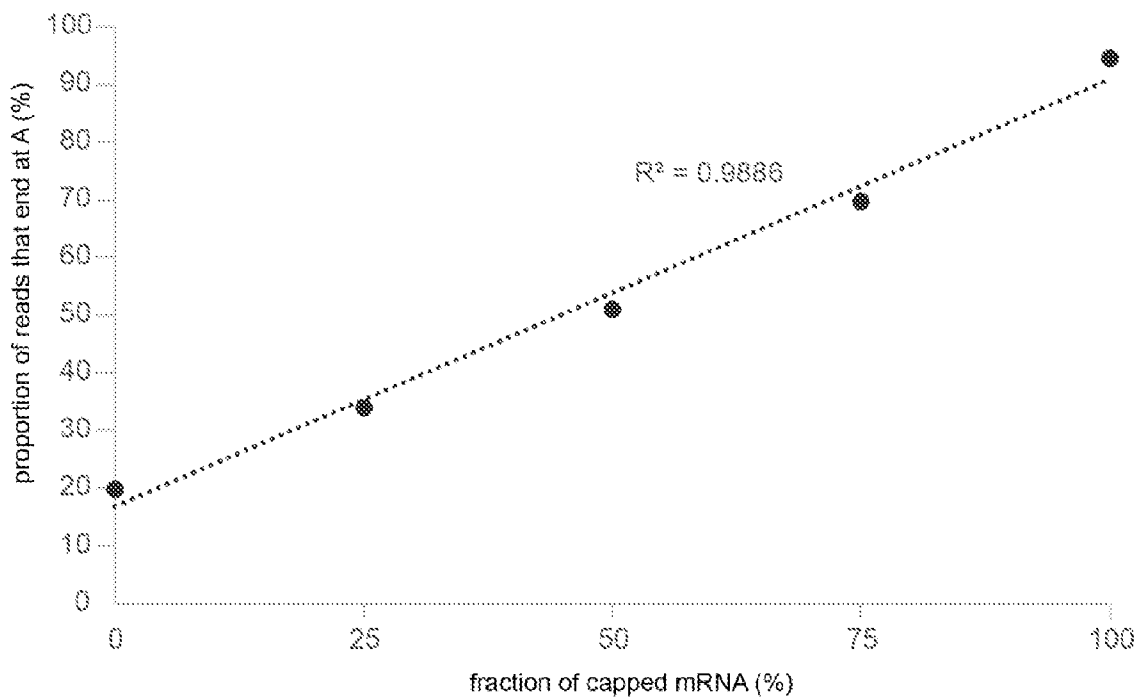


FIGURE 11

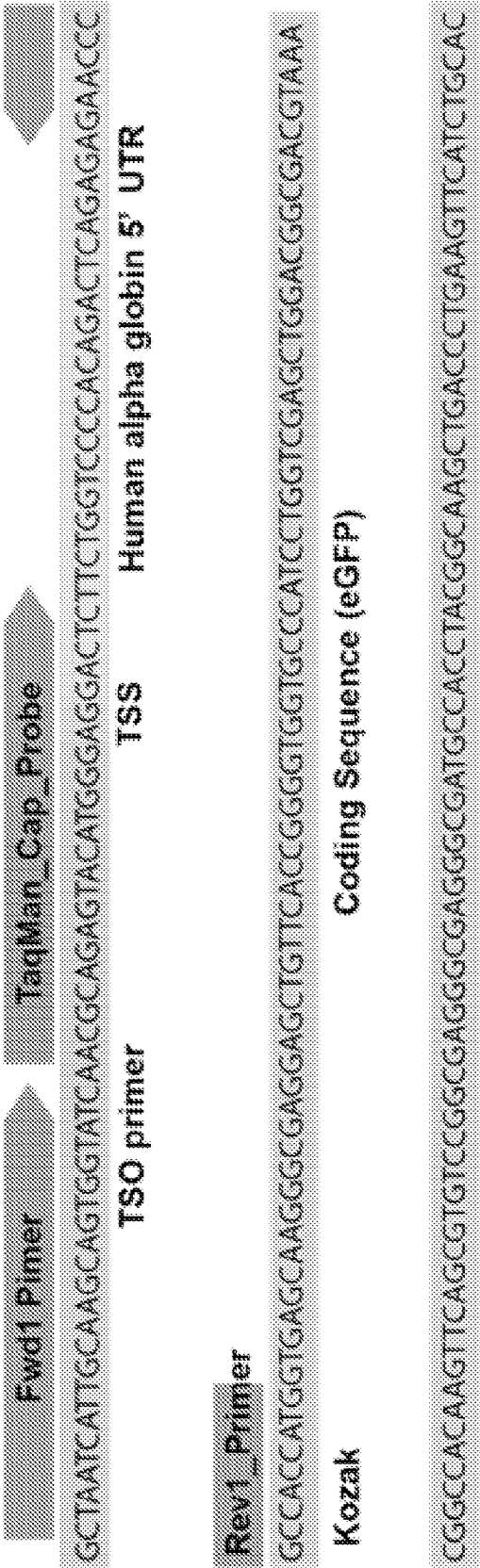


FIGURE 12A

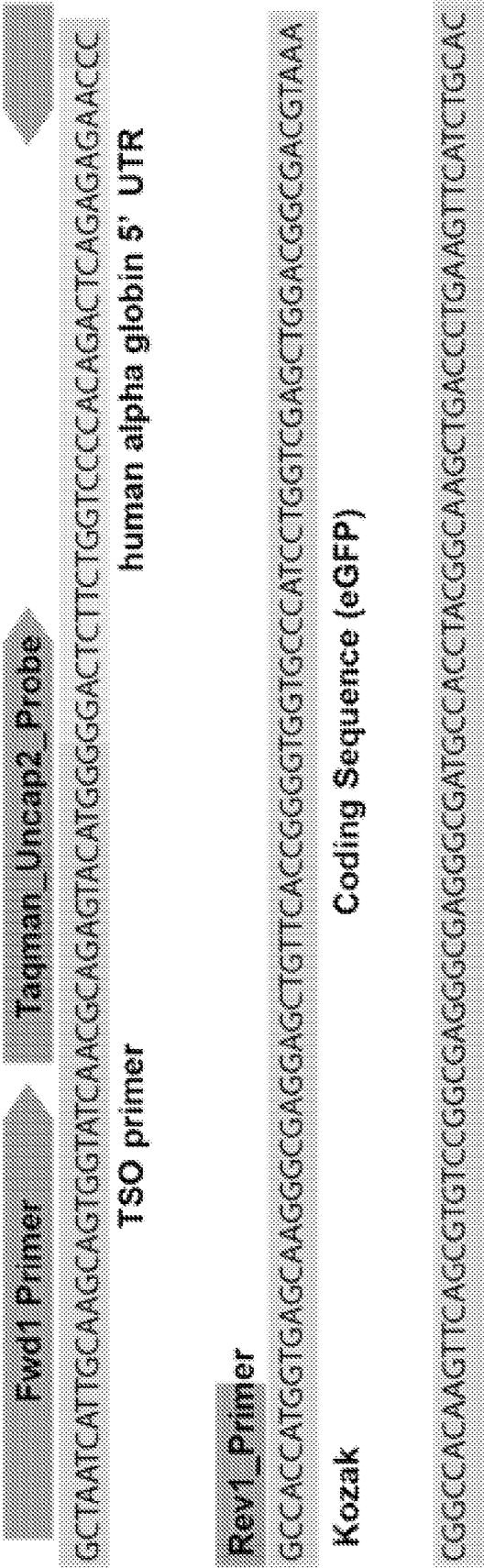


FIGURE 12B

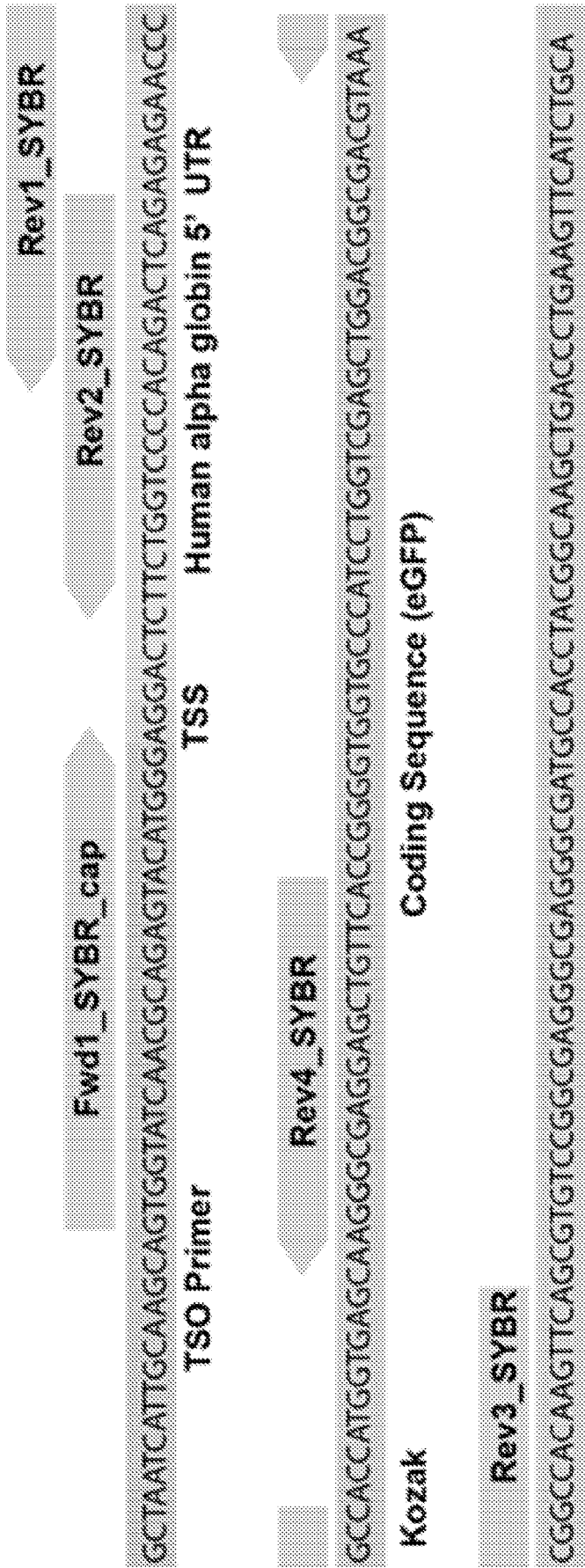


FIGURE 13A

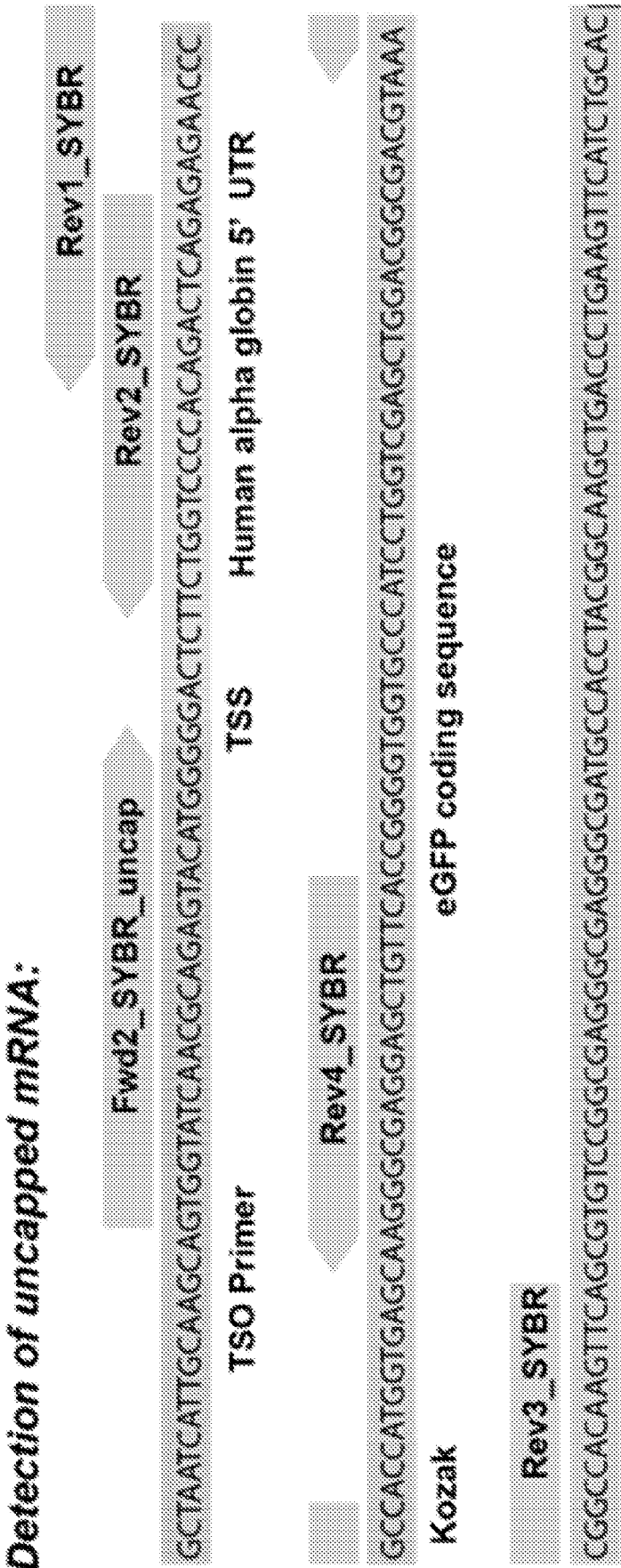


FIGURE 13B

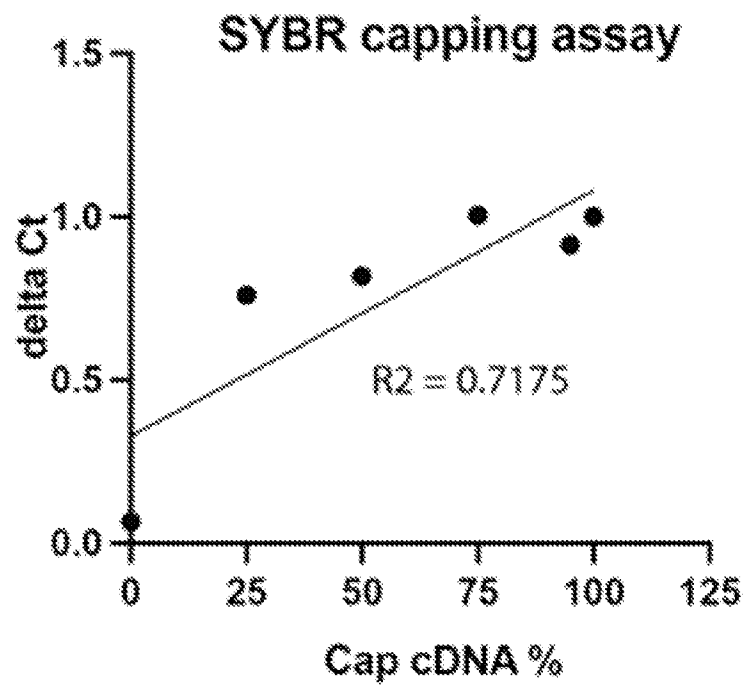


FIGURE 14A

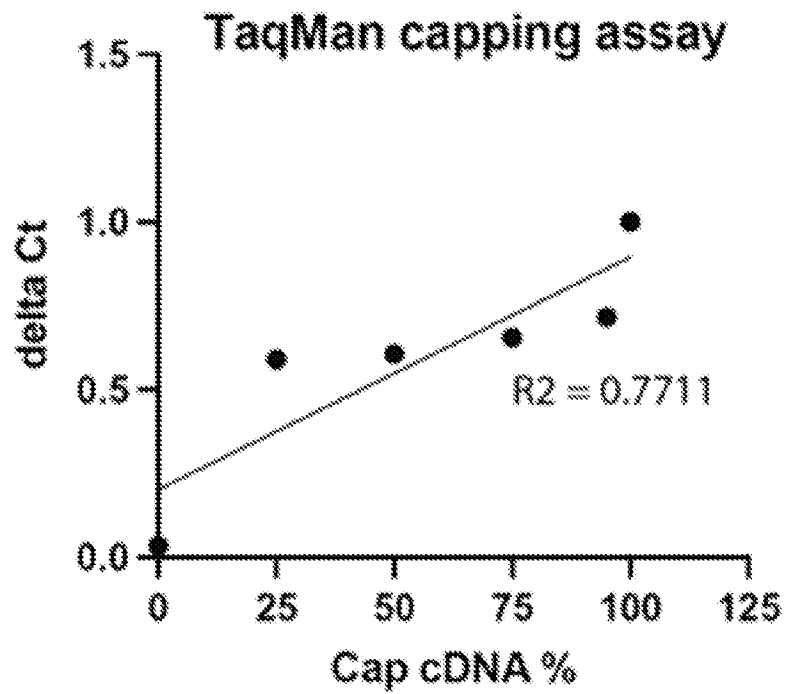


FIGURE 14B

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/6876 (2018.01) A61K 31/7115 (2006.01) C12Q 1/6869 (2018.01)

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)

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)

PATENW, MEDLINE, CAPLUS, EMBASE & keywords: RNA, cap, capping efficiency and like terms; CPC/IPC: C12Q1/68

PUNMED, ESPACENET, Internal Databases: Inventor and Applicant names

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

☒ Further documents are listed in the continuation of Box C☒ 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	"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
"D" document cited by the applicant in the international application	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 19 December 2024	Date of mailing of the international search report 19 December 2024
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA Email address: pct@ipaustalia.gov.au	Authorised officer Richard Filmer AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No. +61 2 6283 2735

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☐ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13^{ter}.1(a)).
☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

A sequence listing was filed with the application but not used for the purposes of the search.

INTERNATIONAL SEARCH REPORT Information on patent family members		International application No. PCT/AU2024/051208	
This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.			
Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	Publication Date	Publication Number	Publication Date
CN 116356002 A	30 June 2023	CN 116356002 A	30 Jun 2023
CN 115896239 A	04 April 2023	CN 115896239 A	04 Apr 2023
End of Annex			
Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001. Form PCT/ISA/210 (Family Annex)(July 2019)			