

METHODS OF SIMULTANEOUSLY IDENTIFYING OR QUANTIFYING CAPPING AND TAILING MODIFICATIONS OF MESSENGER RNA

CROSS-REFERENCE TO RELATED APPLICATIONS

[1] This application claims the benefit of, and relies on the filing date of, European Patent Application No. 23305946.8 filed 14 June 2023, the entire disclosure of which is incorporated herein by reference.

INCORPORATION-BY-REFERENCE OF SEQUENCE LISTING

[2] The contents of the ST26 Sequence Listing file named "Sequence listing.xml", which was created on June 12, 2024 and is 7,194 bytes in size, is hereby incorporated by reference in its entirety.

BACKGROUND

[3] RNA (e.g., messenger RNA, "mRNA") therapeutics are important for the treatment of a variety of diseases. Rapid development and deployment of RNA vaccines was revolutionized to counter the coronavirus pandemic, and RNA vaccines are increasingly becoming important to treat a variety of diseases, including infectious diseases and cancer.

[4] mRNA vaccines trigger protein production within the patient's body, thereby priming the immune system to trigger a protective immune response when exposed to pathogens. For example, mRNA therapy requires stable and effective delivery of the mRNA and efficient production of protein encoded by the mRNA within the patient's body.

[5] For stable mRNA delivery and protein production *in vivo*, a cap is typically required at the 5' end of the mRNA, and a tail at the 3' end, both of which protect mRNA from degradation and facilitate delivery of the mRNA for *in vivo* protein translation. Development of capping modifications and polyadenosine tail (poly A) have improved mRNA stability and translation in eukaryotic cells.

[6] Current methods of estimating capping and tailing known in the art have various drawbacks. For example, estimation methods may result in loss of product, by rendering it radioactive or destroying it in the process of analysis, making it unavailable for therapeutic applications. Further, although separate quantification reactions could be run alongside a therapeutic synthesis reaction, such methods are inherently variable and error prone. To obtain accurate results, it is desirable to use a representative sample or aliquot taken from the therapeutic synthesis reaction of the product itself.

SUMMARY

[7] The present disclosure provides, among other things, a multi-attribute method to identify and quantify mRNA capping and untailing species simultaneously using liquid chromatography with UV detection (LC-UV), liquid chromatography coupled to mass spectrometry (LC-MS), or liquid chromatography coupled with UV and mass spectrometry (LC-UV-MS). The method provided herein also characterizes in a single sample the poly A tail length and polydispersity by mass spectrometry. The present disclosure is thus useful for characterization of mRNA as an active pharmaceutical ingredient (API) in therapeutic products.

[8] The method of the present disclosure is used to characterize 5' and 3' extremities of RNA, including any type of therapeutic RNA, including mRNA vaccine at different stages, i.e., *in vitro* transcribed mRNA, in-process mRNA during the manufacturing process, drug substance (DS) and drug product (DP) stages after deformation. The method is an important quality control or release assay for manufacturing mRNA products. UV signals are particularly useful in a GMP environment. Quantifying capping and tailing species provides an indication of degradation of RNA at the 3' and 5' extremities.

[9] In some aspects, provided herein is a single sample method of identifying ribonucleic acid (RNA) capping and tailing modifications, the method comprising: (a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail, and wherein said sample further comprises a first oligonucleotide probe complementary to a sequence in the 5' untranslated region (5' UTR) region of the RNA and a second oligonucleotide probe complementary to a sequence in the 3' untranslated region (3' UTR) region of the RNA; (b) annealing the RNA sample with the first oligonucleotide probe complementary to a sequence in the 5' UTR of the RNA, and the second oligonucleotide probe complementary to a sequence in the 3' UTR of the RNA; (c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments; (d) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS using the sample of step (c) for: (i) identifying capping species in the cap fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; (ii) identifying untailing species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; and (iii) identifying tailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; thereby simultaneously identifying RNA capping and tailing modifications in the single sample. In some embodiments, the first base of the first oligonucleotide probe

binds at the penultimate base of the RNA or adjacent to the penultimate base of the RNA, such as at least 2-10 nucleotides from the penultimate base of the RNA.

[10] In some embodiments, capping species in the cap fragments is identified by measuring retention time of peaks in a chromatogram generated by LC-UV. In some embodiments, capping species in the cap fragment is identified by measuring the mass of the cap species using mass spectra generated by MS. In some embodiments, both retention time of peaks and mass of the cap species is measured by LC-UV-MS. In some embodiments, untailed species and tailed species in the tail fragment is identified by measuring retention time of peaks in a chromatogram generated by LC-UV. In some embodiments, untailed species and tailed species in the tail fragment is identified by measuring the mass generated by LC-MS. In some embodiments, untailed species and tailed species in the tail fragment is identified by measuring the mass generated by MS. In some embodiments, both retention time of peaks and mass of tailed species is measured by LC-UV-MS.

[11] In some aspects, provided herein is a single sample method of simultaneously quantifying RNA capping and tailing efficiency, the method comprising: (a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail; (b) annealing the RNA sample with a first oligonucleotide probe complementary to a sequence in a 5' untranslated region (5' UTR) of the RNA, and a second oligonucleotide probe complementary to a sequence in a 3' untranslated region (3' UTR) of the RNA; (c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments; (d) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS; (e) measuring a peak area of each capping species in the cap fragment, and each untailed species in the tail fragment; and (f) quantifying a relative amount of each capping species, quantifying a relative amount of untailed species and characterizing poly A tail in the sample of step (d); thereby simultaneously quantifying RNA capping efficiency and tailing efficiency in the sample. In some embodiments, the first base of the first oligonucleotide probe binds at the penultimate base of the RNA or adjacent to the penultimate base of the RNA, such as at least 2-10 nucleotides from the penultimate base of the RNA.

[12] In some embodiments, provided herein is a method wherein the characterizing poly A tail comprises (i) mass measurements of poly A tail species, deconvoluting a mass spectrometry signal, thereby providing a mass distribution for tailed species, and (ii) determining based on the corresponding mass distribution, a minimum length, a maximum length, an average length and a degree of polydispersity of the poly A tail.

[13] In some embodiments, the RNA is *in vitro* transcribed mRNA.

- [14] In some embodiments, the RNA is obtained from a step of manufacturing.
- [15] In some embodiments, the RNA is obtained from a drug substance (DS) at the final step of manufacturing.
- [16] In some embodiments, the RNA is obtained from a de formulated drug product (DP).
- [17] In some embodiments, the RNA is unmodified. In some embodiments, the RNA comprises no modified nucleotides.
- [18] In some embodiments, the RNA is modified.
- [19] In some embodiments, the nuclease is RNase H.
- [20] In some embodiments, the mass of the cap and the tail fragments is measured by LC-MS.
- [21] In some embodiments, provided herein is a method, wherein the capping species, untailed species and tailed species is identified from a UV signal by comparing the retention time of peaks obtained to the retention time of a reference standard.
- [22] In some embodiments, provided herein is a method, wherein the capping species are Cap1, Cap0, CapG or uncapped. In some embodiments, the capping species is Cap1. In some embodiments, the capping species is Cap0. In some embodiments, the capping species is CapG. In some embodiments, the capping species is uncapped.
- [23] In some embodiments, the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA, GGACA or pppGGACA. In some embodiments, the capping species is m7Gpppm7GGACA. In some embodiments, the capping species is m7GpppGGACA. In some embodiments, the capping species is GpppGGACA. In some embodiments, the capping species is GGACA. In some embodiments, the capping species is pppGGACA.
- [24] In some embodiments, the uncapped species is GGACA or pppGGACA.
- [25] In some embodiments, the untailed species is UGCAUC, wherein U is unmodified uridine.
- [26] In some embodiments, the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine.
- [27] In some embodiments, the relative amount of each of the capping species is a percentage of total amount of capping species in the sample calculated by dividing an area under the peak of the capping species of interest over a sum of total areas under the peak representing Cap1, Cap0, CapG and uncapped species, multiplied by 100.

[28] In some embodiments, the relative amount of untailed species is a percentage calculated by dividing an area under the peak of the untailed species over a sum of areas under the peak of Cap1, Cap0, CapG and uncapped multiplied by 100.

[29] In some embodiments, the oligonucleotide is between about 10 to 40 nucleotides in length. In some embodiments, the oligonucleotide is between about 10-15, 15-20, 20-25, 25-30, or 35-40 nucleotides in length, including all discrete intervening quantities. In some embodiments, the oligonucleotide is 10 nucleotides in length. In some embodiments, the oligonucleotide is 11 nucleotides in length. In some embodiments, the oligonucleotide is 12 nucleotides in length. In some embodiments, the oligonucleotide is 13 nucleotides in length. In some embodiments, the oligonucleotide is 14 nucleotides in length. In some embodiments, the oligonucleotide is 15 nucleotides in length. In some embodiments, the oligonucleotide is 16 nucleotides in length. In some embodiments, the oligonucleotide is 17 nucleotides in length. In some embodiments, the oligonucleotide is 18 nucleotides in length. In some embodiments, the oligonucleotide is 19 nucleotides in length. In some embodiments, the oligonucleotide is 20 nucleotides in length. In some embodiments, the oligonucleotide is 21 nucleotides in length. In some embodiments, the oligonucleotide is 22 nucleotides in length. In some embodiments, the oligonucleotide is 23 nucleotides in length. In some embodiments, the oligonucleotide is 24 nucleotides in length. In some embodiments, the oligonucleotide is 25 nucleotides in length. In some embodiments, the oligonucleotide is 26 nucleotides in length. In some embodiments, the oligonucleotide is 27 nucleotides in length. In some embodiments, the oligonucleotide is 28 nucleotides in length. In some embodiments, the oligonucleotide is 29 nucleotides in length. In some embodiments, the oligonucleotide is 30 nucleotides in length. In some embodiments, the oligonucleotide is 31 nucleotides in length. In some embodiments, the oligonucleotide is 32 nucleotides in length. In some embodiments, the oligonucleotide is 33 nucleotides in length. In some embodiments, the oligonucleotide is 34 nucleotides in length. In some embodiments, the oligonucleotide is 35 nucleotides in length. In some embodiments, the oligonucleotide is 36 nucleotides in length. In some embodiments, the oligonucleotide is 37 nucleotides in length. In some embodiments, the oligonucleotide is 38 nucleotides in length. In some embodiments, the oligonucleotide is 39 nucleotides in length. In some embodiments, the oligonucleotide is 40 nucleotides in length.

[30] In some embodiments, the oligonucleotide is about 10 to 40 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is between about 10-15, 15-20, 20-25, 25-30, or 35-40 nucleotides in length, including all discrete intervening quantities, and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 10 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 11 nucleotides long and comprises 4 DNA bases. In some

embodiments, the oligonucleotide is 12 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 13 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 14 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 15 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 16 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 17 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 18 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 19 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 20 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 21 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 22 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 23 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 24 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 25 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 26 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 27 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 28 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 29 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 30 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 31 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 32 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 33 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 34 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 35 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 36 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 37 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 38 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 39 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 40 nucleotides long and comprises 4 DNA bases.

[31] In some embodiments, the oligonucleotide comprises RNA and DNA bases.

[32] In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1. In some embodiments, the oligonucleotide comprises RNA

and DNA bases in a ratio of about 10:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 9:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 8:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 7:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 6:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 5:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 4:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 3:1.

[33] In some embodiments, the oligonucleotide is 10-40 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is between 10-15, 15-20, 20-25, 25-30, or 35-40 nucleotides long, including all discrete intervening quantities, and comprises 4 DNA bases.

[34] In some embodiments, the oligonucleotide complementary to a sequence in the 5' untranslated region of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein italicized CTGT are DNA bases.

[35] In some embodiments, the oligonucleotide complementary to a sequence in the 3' untranslated region of the RNA is 3'-GGUCGGAACAGGAUUUUUAAATTC-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.

[36] In some embodiments, capping and untailing species are quantified in a single LC-UV or LC-MS or LC-UV-MS analysis. In some embodiments, capping and untailing species are quantified in a single LC-UV analysis. In some embodiments, capping and untailing species are quantified in a single LC-MS analysis. In some embodiments, capping and untailing species are quantified in a single LC-UV-MS analysis. In some embodiments, capping and untailing species are measured by Ultra High Performance Liquid Chromatography-Electrospray Ionization Mass Spectrometry (UHPLC-ESI-MS).

[37] In some embodiments, tailed species are characterized in a single LC-MS or LC-UV-MS analysis. In some embodiments, tailed species are characterized in a single LC-MS analysis. In some embodiments, tailed species are characterized in a single LC-UV-MS analysis.

[38] In some embodiments, one or more steps are automated. In some embodiments, one step is automated. In some embodiments, more than one step are automated. In some embodiments, all steps are automated.

[39] In some aspects, provided herein is an oligonucleotide that is complementary to a sequence in the 5' UTR of the RNA or 3' UTR of the RNA, wherein the oligonucleotide is between about 10 to 40 nucleotides in length and comprises RNA and DNA bases. In some aspects, provided herein is an oligonucleotide that is complementary to a sequence in the 5' UTR of the RNA, wherein the

oligonucleotide is between about 10 to 40 nucleotides in length and comprises RNA and DNA bases. In some aspects, provided herein is an oligonucleotide that is complementary to a sequence in the 3' UTR of the RNA, wherein the oligonucleotide is between about 10 to 40 nucleotides in length and comprises RNA and DNA bases.

[40] In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 9:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 8:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 7:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 6:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 5:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 4:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 3:1.

[41] In some embodiments, the oligonucleotide is about 10 to 40 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is about 10-15, 15-20, 20-25, 25-30, 30-35, or 35-40 nucleotides long and comprises 4 DNA bases.

[42] In some embodiments, the oligonucleotide complementary to a sequence in the 5' UTR of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein the italicized CTGT are DNA bases.

[43] In some embodiments, the oligonucleotide complementary to a sequence in the 3' UTR of the RNA is 3'-GGUCGGAACAGGAUUAUUUAATTCAA-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.

[44] In some aspects, provided herein is a method of manufacturing RNA having a quantified percentage of capped and untail mRNA comprising the steps of: (a) providing *in vitro* synthesized RNA; (b) annealing the *in vitro* synthesized RNA with a first oligonucleotide complementary to a sequence in a 5' UTR of the RNA, and a second oligonucleotide complementary to a sequence in a 3' UTR of the RNA; (c) treating the RNA with a nuclease to cleave the RNA into cap and tail fragments; (d) identifying the capping species and untail species by liquid chromatography with UV detection (LC-UV) and/or identifying and measuring the mass of the capping species, untail species and characterizing the polyA tail by liquid chromatography coupled to mass spectrometry (LC-MS) and/or liquid chromatography with UV detection and coupled to mass spectrometry (LC-UV-MS); and (e) quantifying a relative amount of each capping species, quantifying a relative amount of untail species using LC-UV or LC-MS or LC-UV-MS and characterizing poly A tail using LC-MS or LC-UV-MS in a single sample simultaneously; thereby

manufacturing RNA comprising a quantified percentage of capped and untailed mRNA. In some embodiments, the first base of the first oligonucleotide binds at the penultimate base of the RNA or adjacent to the penultimate base of the RNA, such as at least 2-10 nucleotides from the penultimate base of the RNA.

[45] In some embodiments, provided herein is a method wherein a 5' cap is added enzymatically, post transcription to the RNA from step (a).

[46] In some embodiments, the RNA does not comprise a 5' cap.

[47] In some embodiments, a 3' tail is encoded in a plasmid during *in vitro* synthesis.

[48] In some embodiments, a 3' tail is added enzymatically, post transcription to the RNA from step (a).

[49] In some embodiments, the RNA does not comprise a 3' tail.

[50] In some embodiments, the RNA is unspliced.

[51] In some embodiments, the RNA is spliced.

[52] In some aspects, provided herein is a quality control assay for a manufacturing lot, wherein the assay comprises: (a) providing a manufactured RNA sample comprising a 5' cap and/or a 3' tail; (b) quantifying RNA capping and tailing efficiency by the method provided herein in the sample; (c) comparing the capping and tailing efficiency in step (b) to capping and tailing efficiency in a reference sample; wherein a reduced amount of capping species relative to a reference sample indicates 5' degradation, and wherein an increased amount of untailed species and/or a different poly A tail length or polydispersity indicates 3' degradation, thereby simultaneously monitoring 5' degradation and 3' degradation of an RNA product, and determining product quality of the manufactured lot.

[53] In some embodiments, provided herein is a quality control assay, wherein the reference sample is a stable RNA sample having greater than 75% integrity. In some embodiments, provided herein is a quality control assay, wherein the reference sample is a stable RNA sample having between 75%-100% integrity. In some embodiments, provided herein is a quality control assay, wherein the reference sample is a stable RNA sample having between 75%-99% integrity. In some embodiments, provided herein is a quality control assay, wherein the reference sample is a stable RNA sample having between 75%-95% integrity. In some embodiments, provided herein is a quality control assay, wherein the reference sample is a stable RNA sample having 75%-80%, 80%-85%, 85%-90%, 90%-95% or 95%-100% integrity, including all intervening quantities.

[54] In some embodiments, the RNA product is a therapeutic mRNA.

[55] In some embodiments, the RNA product is an mRNA vaccine.

BRIEF DESCRIPTION OF THE DRAWINGS

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

[57] **FIG. 1** depicts a flowchart showing steps of the method of the present disclosure comprising annealing hybrid RNA oligonucleotides comprising four DNA bases to the 5' and 3' extremities of an mRNA sample in an annealing step, followed by the addition of ribonuclease H (RNase H) enzyme and shrimp alkaline phosphatase (rSAP) to digest mRNA in the RNA:DNA hybrid and subsequent analysis, of the capped and uncapped analytes. Two exemplary hybrid oligonucleotides are: HO2.16 which is a 16 nts custom designed hybrid oligonucleotide 3'-**CCTG**TCUAGCGGACCU-5' (SEQ ID NO: 1), wherein CTGT in bold near the 3' end represents DNA bases; HO2.27 which is a 27 nts custom designed hybrid oligonucleotide HO2.27: 3'-GGUCGGAACAGGAUUUUUAAT**TTCA**A-5' (SEQ ID NO: 2), wherein TTCA in bold near the 5' end represents DNA bases.

[58] **FIG. 2** depicts a flowchart of the annealing step of the method. RNA from an exemplary sample and exemplary custom hybrid oligonucleotides are heated to 75°C which causes denaturation and strand separation. As the temperature is decreased from 75°C to 23°C, hydrogen bonds form between complementary DNA and RNA sequences. Complementary base pairing between hybrid oligonucleotides and mRNA in a specific region results in RNA:DNA hybrid formation at the end of the annealing step.

[59] **FIG. 3** depicts a flowchart of the digestion step of the method. Digestion of the RNA:DNA hybrid is carried out by incubation of RNase H and shrimp alkaline phosphatase (rSAP) at 37°C. RNase H enzymes cleave the phosphodiester bond between double-stranded RNA:DNA hybrids generated in the annealing step. Shrimp alkaline phosphatase (rSAP) catalyzes dephosphorylation of ends to prevent self-ligation.

[60] **FIG. 4A** is an exemplary Extracted Ion Chromatogram (EIC), which shows peaks generated for capped (Cap 0, Cap 1, Cap G), uncapped and untailed species. Methods of quantitation are also provided. Percentage of any specific capped species (Cap 0, Cap 1, Cap G or uncapped) is calculated by measuring the area under the curve of said capped species, relative to the sum of total area under the curve for all capped (Cap0, Cap1, CapG) and uncapped peaks, multiplied by 100. Likewise, percentage of untailed species is calculated by measuring the area under the curve of untailed species, relative to total area under the curve for all capped (Cap0, Cap1, CapG) and uncapped peaks, multiplied by 100. In the exemplary EIC shown in **FIG. 4A**, the results of the quantitation demonstrate the presence of 6.8% untailed, 4.2% uncapped, 0.4% Cap G, 0.0% Cap 0 and 95.4% Cap 1 species. In some embodiments, EIC is

generated, for example, by Liquid chromatography Mass Spectrometry (LC-MS), Liquid chromatography-ultraviolet spectroscopy/mass spectrometry (LC-UV/MS), Ultrahigh performance liquid chromatography-electrospray ionization mass spectrometry (UHPLC-ESI-MS) or other mass spectrometric methods.

[61] **FIG. 4B** is a schematic flowchart showing the synthesis of exemplary mRNA capped structures and an exemplary uncapped structure. Uncapped mRNA (e.g. *in vitro* transcribed mRNA, pppG-mRNA), upon dephosphorylation, loses one phosphate moiety. Subsequent treatment with guanylyltransferase synthesizes CapG (GpppG-mRNA). A Cap G structure in the presence of S-(5'-adenosyl)-L-methionine p-toluenesulfonate salt (SAM-TOS) and N7-methyltransferase, is converted to a Cap 0 structure (m7GpppG-mRNA). A Cap 0 structure, in the presence of 2'-O-ribose-methyltransferase and SAM-TOS is converted to a Cap1 (m7GpppmG-mRNA) structure.

[62] **FIG. 5A** depicts average deconvoluted mass spectra for exemplary mRNA to study poly A tailing, for an exemplary drug substance. In the graph, mass spectrometric signal intensity is plotted relative to m/z (mass to charge ratio).

[63] **FIG. 5B** is a graph of percent mass spectrometric signal intensity relative to poly A tail length. The results show average poly A tail length and poly A polydispersity in an exemplary mRNA sample of a drug substance.

DEFINITIONS

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

[65] *Affinity*: As is known in the art, "affinity" is a measure of the tightness with which a particular ligand binds to (e.g., associates non-covalently with) and/or the rate or frequency with which it dissociates from, its partner. As is known in the art, any of a variety of technologies can be utilized to determine affinity. In many embodiments, affinity represents a measure of specific binding.

[66] *Anneal or hybridization*: As used herein, the terms "anneal," "hybridization," and grammatical equivalent, refer to the formation of complexes (also called duplexes or hybrids) between nucleotide sequences which are sufficiently complementary to form complexes *via* Watson-Crick base pairing or non-canonical base pairing. It will be appreciated that annealing or hybridizing sequences need not have perfect complementary to provide stable hybrids. In many situations, stable hybrids will form where fewer than about 10% of the bases are mismatches. Accordingly, as used herein, the term

“complementary” refers to a nucleic acid molecule that forms a stable duplex with its complement under particular conditions, generally where there is about 90% or greater homology (e.g., about 95% or greater, about 98% or greater, or about 99% or greater homology). Those skilled in the art understand how to estimate and adjust the stringency of hybridization conditions such that sequences that have at least a desired level of complementarity will stably hybridize, while those having lower complementarity will not. For examples of hybridization conditions and parameters, see, for example, Sambrook et al., “*Molecular Cloning: A Laboratory Manual*”, 1989, Second Edition, Cold Spring Harbor Press: Plainview, NY and Ausubel, “*Current Protocols in Molecular Biology*”, 1994, John Wiley & Sons: Secaucus, NJ. Complementarity between two nucleic acid molecules is said to be “complete”, “total” or “perfect” if all the nucleic acid’s bases are matched and is said to be “partial” otherwise.

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

[68] *Chromatography:* As used herein, the term “chromatography” refers to a technique for separation of mixtures. Typically, the mixture is dissolved in a fluid called the “mobile phase,” which carries it through a structure holding another material called the “stationary phase.” Column chromatography is a separation technique in which the stationary bed is within a tube, i.e., a column.

[69] *Compound and Agent:* The terms “compound” and “agent” are used herein interchangeably. They refer to any naturally occurring or non-naturally occurring (i.e., synthetic or recombinant) molecule, such as a biological macromolecule (e.g., nucleic acid, polypeptide or protein), organic or inorganic molecule, or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian, including human) cells or tissues. The compound may be a single molecule or a mixture or complex of at least two molecules.

[70] *Control, Standard or Reference:* As used herein, the term “control”, “standard” or “reference” used interchangeably has its art-understood meaning of being a standard of a known value or quantity against which results are compared. Typically, controls are used to augment integrity in experiments by isolating variables in order to make a conclusion about such variables. In some embodiments, a control is a reaction or assay that is performed simultaneously with a test reaction or

assay to provide a comparator. In one experiment, the “test” (*i.e.*, the variable being tested) is applied. In the second experiment, the “control,” the variable being tested is not applied. In some embodiments, a control is a historical control (*i.e.*, of a test or assay performed previously, or an amount or result that is previously known). In some embodiments, a control is or comprises a printed or otherwise saved record. A control may be a positive control or a negative control.

Extracted Ion Chromatogram: Extracted ion chromatogram (EIC) is created by plotting the intensity of the signal observed at chosen mass-to-charge value or series of values in a series of mass spectra recorded as a function of retention time.

[71] *Kit:* As used herein, the term “kit” refers to any delivery system for delivering materials. Such delivery systems may include systems that allow for the storage, transport, or delivery of various diagnostic or therapeutic reagents (e.g., oligonucleotides, antibodies, enzymes, etc. in the appropriate containers) and/or supporting materials (e.g., buffers, written instructions for performing the assay etc.) from one location to another. For example, kits include one or more enclosures (e.g., boxes) containing the relevant reaction reagents and/or supporting materials. As used herein, the term “fragmented kit” refers to delivery systems comprising two or more separate containers that each contains a subportion of the total kit components. The containers may be delivered to the intended recipient together or separately. For example, a first container may contain an enzyme for use in an assay, while a second container contains oligonucleotides. The term “fragmented kit” is intended to encompass kits containing Analyte Specific Reagents (ASR’s) regulated under section 520(e) of the Federal Food, Drug, and Cosmetic Act, but are not limited thereto. Indeed, any delivery system comprising two or more separate containers that each contains a subportion of the total kit components are included in the term “fragmented kit.” In contrast, a “combined kit” refers to a delivery system containing all of the components in a single container (e.g., in a single box housing each of the desired components). The term “kit” includes both fragmented and combined kits.

[72] *Nucleoside:* The term “nucleoside” or “nucleobase”, as used herein, refers to adenine (“A”), guanine (“G”), cytosine (“C”), uracil (“U”), thymine (“T”) and analogs thereof linked to a carbohydrate, for example D-ribose (in RNA) or 2'-deoxy-D-ribose (in DNA), through an N-glycosidic bond between the anomeric carbon of the carbohydrate (1'-carbon atom of the carbohydrate) and the nucleobase. When the nucleobase is purine, e.g., A or G, the ribose sugar is generally attached to the N9-position of the heterocyclic ring of the purine. When the nucleobase is pyrimidine, e.g., C, T or U, the sugar is generally attached to the N1-position of the heterocyclic ring. The carbohydrate may be substituted or unsubstituted. Substituted ribose sugars include, but are not limited to, those in which one

or more of the carbon atoms, for example the 2'-carbon atom, is substituted with one or more of the same or different Cl, F, --R, --OR, --NR₂ or halogen groups, where each R is independently H, C₁-C₆ alkyl or C₅-C₁₄ aryl. Ribose examples include ribose, 2'-deoxyribose, 2',3'-dideoxyribose, 2'-haloribose, 2'-fluororibose, 2'-chlororibose, and 2'-alkylribose, e.g., 2'-O-methyl, 4'-alpha-anomeric nucleotides, 1'-alpha-anomeric nucleotides (Asseline et al., *Nucleic Acids Research*, 19:4067-74 [1991]), 2'-4'- and 3'-4'-linked and other "locked" or "LNA," bicyclic sugar modifications (WO 98/22489; WO 98/39352; WO 99/14226).

[73] *Nucleotide:* The term "nucleotide" as used herein means a nucleoside in a phosphorylated form (a phosphate ester of a nucleoside), as a monomer unit or within a polynucleotide polymer. "Nucleotide 5'-triphosphate" refers to a nucleotide with a triphosphate ester group at the 5' position, sometimes denoted as "NTP", or "dNTP" and "ddNTP" to particularly point out the structural features of the ribose sugar. The triphosphate ester group may include sulfur substitutions for the various oxygen moieties, e.g., alpha-thio-nucleotide 5'-triphosphates. Nucleotides can exist in the mono-, di-, or tri-phosphorylated forms. The carbon atoms of the ribose present in nucleotides are designated with a prime character (') to distinguish them from the backbone numbering in the bases. For a review of polynucleotide and nucleic acid chemistry see Shabarova, Z. and Bogdanov, A. *Advanced Organic Chemistry of Nucleic Acids*, VCH, New York, 1994.

[74] *Nucleic acid:* The terms "nucleic acid", "nucleic acid molecule", "polynucleotide" or "oligonucleotide" may be used herein interchangeably. They refer to polymers of nucleotide monomers or analogs thereof, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) and combinations thereof. The nucleotides may be genomic, synthetic or semi-synthetic in origin. Unless otherwise stated, the terms encompass nucleic acid-like structures with synthetic backbones, as well as amplification products. As will be appreciated by one skilled in the art, the length of these polymers (i.e., the number of nucleotides it contains) can vary widely, often depending on their intended function or use. Polynucleotides also have associated counter ions, such as H⁺, NH₄⁺, trialkylammonium, Mg⁺, Na⁺ and the like. A polynucleotide may be composed entirely of deoxyribonucleotides, entirely of ribonucleotides, or chimeric mixtures thereof. Polynucleotides may be composed of internucleotide nucleobase and sugar analogs.

[75] *Oligonucleotide:* The term "oligonucleotide" is used herein to denote a polynucleotide that comprises between about 10 and about 40 nucleotides, between about 15 and about 40 nucleotides, or between about 15 and about 30 nucleotides. Throughout the specification, whenever an oligonucleotide is represented by a sequence of letters (chosen, for example, from the four base letters:

A, C, G, and T, which denote adenosine, cytidine, guanosine, and thymidine, respectively), the nucleotides are presented in the 5' to 3' order from the left to the right.

[76] *Polynucleotide*: The term “polynucleotide sequence” refers to the sequence of nucleotide monomers comprising a polymer. Unless denoted otherwise, whenever a polynucleotide sequence is represented, it will be understood that the nucleotides are in 5' to 3' orientation from left to right.

[77] *Modified nucleotides*: Nucleic acids, polynucleotides and oligonucleotides may be comprised of standard nucleotide bases or substituted with nucleotide isoform analogs, including, but not limited to iso-C and iso-G bases, which may hybridize more or less permissibly than standard bases, and which will preferentially hybridize with complementary isoform analog bases. Many such isoform bases are described, for example, by Benner et al., (1987) *Cold Spring Harb. Symp. Quant. Biol.* 52, 53-63. Analogs of naturally occurring nucleotide monomers include, for example, 7-deazaadenine, 7-deazaguanine, 7-deaza-8-azaguanine, 7-deaza-8-azaadenine, 7-methylguanine, inosine, nebularine, nitropyrrole (Bergstrom, J. Amer. Chem. Soc., 117:1201-1209 [1995]), nitroindole, 2-aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, pseudouridine, pseudocytosine, pseudoisocytosine, 5-propynylcytosine, isocytosine, isoguanine (Seela, U.S. Patent No. 6,147,199), 7-deazaguanine (Seela, U.S. Patent No. 5,990,303), 2-azapurine (Seela, WO 01/16149), 2-thiopyrimidine, 6-thioguanine, 4-thiouracil, 0-6-methylguanine, N-6-methyladenine, O-4-methylthymine, 5,6-dihydrothymine, 5,6-dihydrouracil, 4-methylindole, pyrazolo[3,4-D]pyrimidines, “PPG” (Meyer, U.S. Pat. Nos. 6,143,877 and 6,127,121; Gall, WO 01/38584), and ethenoadenine (Fasman (1989) in *Practical Handbook of Biochemistry and Molecular Biology*, pp. 385-394, CRC Press, Boca Raton, Fla.).

[78] *“3' end” and “3' terminus”*: As used herein, the terms refer to the end of the nucleic acid which contains a free hydroxyl group attached to the 3' carbon of the terminal pentose sugar.

[79] *“5' end” and “5' terminus”*: As used herein, the terms refer to the end of the nucleic acid molecule which contains a free hydroxyl or phosphate group attached to the 5' carbon of the terminal pentose sugar.

[80] *Target*: As used herein, the term “target” refers to a molecule of interest.

DETAILED DESCRIPTION

[81] With the growing importance of RNA therapeutics and RNA vaccines to treat a variety of cancers and infectious diseases, there is a need to accurately determine product quality and stability. One of the biggest challenges in RNA therapeutics and RNA vaccines is stability because single stranded RNA

is highly prone to degradation. The 5' cap and 3' tail protect RNA, for example, mRNA from exonucleases and enhance mRNA stability.

[82] There is a need to accurately assess the integrity of the mRNA product and ensure minimal to no degradation, for example, in quality control for releasing manufacturing lots, to ensure high stability standards and batch to batch consistency. The stability of mRNA also needs to be evaluated after varying periods of storage, transport and after thawing frozen drug or mRNA vaccine stored at -70°C, -20°C to warmer temperatures, *e.g.*, 4°C or room temperature. In addition, accurately determining mRNA integrity and stability is important for dosing. Accurate identification, characterization and quantification of mRNA capping and tailing is important for quality control in determining *in vivo* safety and efficacy of mRNA therapeutics. Identification, characterization and quantification of mRNA capping and tailing provides a measure of mRNA quality and stability.

[83] As described in detail below, the present disclosure is, in part, based on simultaneous methods of qualitative identification and quantitation of capping and tailing by chromatography. Thus, the present disclosure provides a simple, reliable and efficient approach for assessing mRNA capping and tailing efficiency at the same time and in the same sample, and among other things, providing a method of manufacturing RNA having quantified capping and tailing efficiency and a quality control method for mRNA therapeutics.

[84] The present disclosure provides, among other things, a multi-attribute method for simultaneously characterizing or quantifying mRNA capping and tailing efficiency from a single sample preparation. The methods of the present disclosure for identification and quantification of mRNA capping and tailing are particularly important for assessing mRNA quality and stability in mRNA therapeutics, including mRNA vaccines.

[85] The present disclosure provides, among other things, a single sample method of identifying ribonucleic acid (RNA) capping and tailing modifications, the method comprising: (a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail, and wherein said sample further comprises a first oligonucleotide probe complementary to a sequence in the 5' untranslated region (5' UTR) region of the RNA and a second oligonucleotide probe complementary to a sequence in the 3' untranslated region (3' UTR) region of the RNA; (b) annealing the RNA sample with the first oligonucleotide probe complementary to a sequence in the 5' UTR of the RNA, and the second oligonucleotide probe complementary to a sequence in the 3' UTR of the RNA; (b) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments; (c) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-

MS using the sample of step (c) for: (i) identifying capping species in the cap fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; (ii) identifying untailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; and (iii) identifying tailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; thereby simultaneously identifying RNA capping and tailing modifications in the single sample. In some embodiments, the first base of the first oligonucleotide probe binds at the penultimate base of the RNA or adjacent to the penultimate base of the RNA, such as at least 2-10 nucleotides from the penultimate base of the RNA.

[86] In some embodiments, capping species in the cap fragment is identified by measuring retention time of peaks in a chromatogram generated by LC-UV. In some embodiments, capping species in the cap fragment is identified by measuring the mass of the cap species using mass spectra generated by MS. In some embodiments, both retention time of peaks and mass of the cap species is measured by LC-UV-MS. In some embodiments, untailed species and tailed species in the tail fragment is identified by measuring retention time of peaks in a chromatogram generated by LC-UV. In some embodiments, untailed species and tailed species in the tail fragment is identified by measuring the mass of the cap species using mass spectra generated by MS. In some embodiments, both retention time of peaks and mass of tailed species is measured by LC-UV-MS.

[87] Provided herein is a single sample method of simultaneously quantifying RNA capping and tailing efficiency, the method comprising: (a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail; (b) annealing the RNA sample with a first oligonucleotide probe complementary to a sequence in a 5' untranslated region (5' UTR) of the RNA, and a second oligonucleotide probe complementary to a sequence in a 3' untranslated region (3' UTR) of the RNA; (c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments; (d) performing liquid chromatography with UV detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS; (e) measuring a peak area of each capping species in the cap fragment, and each untailed and tailed species in the tail fragment; and (f) quantifying a relative amount of each capping species, quantifying a relative amount of untailed species and characterizing poly A tail in the sample of step (d); thereby simultaneously quantifying RNA capping efficiency and tailing efficiency in the sample. In some embodiments, the first base of the first oligonucleotide probe binds at the penultimate base of the RNA or adjacent to the penultimate base of the RNA, such as at least 2-10 nucleotides from the penultimate base of the RNA.

[88] Various embodiments of the present disclosure are useful in identifying and quantitating capping and tailing modifications during *in vitro* mRNA synthesis, and during other stages of mRNA product development and mRNA product release, for example, in-process mRNA during manufacturing, drug substance, and drug product after deformulation, for use in a quality control or release assay. Thus, the present disclosure provides an important quality control approach for manufacturing mRNA and, in particular, for assessing the safety, efficacy and commercial feasibility of mRNAs with therapeutic applications. Use of the method in a release assay provides lot-to-lot or batch-to-batch consistency and ensures product quality of mRNA therapeutics, for example, mRNA vaccines.

[89] Various aspects of the present disclosure are described in detail in the following sections. The use of sections is not meant to limit the present disclosure. Each section can apply to any aspect of the present disclosure. In this application, the use of “or” means “and/or” unless stated otherwise.

[90] As used in this application, the terms “about” and “approximately” are used as equivalents. Any numerals used in this application with or without about/approximately are meant to cover any normal fluctuations appreciated by one of ordinary skill in the relevant art.

[91] Other features, objects, and advantages of the present disclosure are apparent in the detailed description that follows. It should be understood, however, that the detailed description, while indicating embodiments of the present disclosure, is given by way of illustration only, not limitation. Various changes and modifications within the scope of the present disclosure will become apparent to those skilled in the art from the detailed description.

Hybrid Oligonucleotides

[92] The methods of the present disclosure rely, among other things, on design and synthesis of hybrid oligonucleotides specific for a target that are complementary to a sequence in the 5' end of the RNA or 3' end of the RNA, thereby annealing to the complementary sequence at the 5' or 3' end respectively. The hybrid oligonucleotides of the present disclosure comprise RNA and DNA bases, which provide specificity for subsequent digestion of RNA.

[93] Typically, the oligonucleotide is between about 10 to 40 nucleotides in length and comprises RNA and DNA bases. In some embodiments, the oligonucleotide is between about 10 to 40 nucleotides in length. In some embodiments, the oligonucleotide is between about 10-15, 15-20, 20-25, 25-30, or 35-40 nucleotides in length, including all discrete intervening quantities.

[94] In some embodiments, the oligonucleotide is about 10 to 40 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is about 10 to 40 nucleotides long and comprises greater than 4 DNA bases. In some embodiments, the oligonucleotide is about 10 to 40

nucleotides long and comprises 4-10 DNA bases. In some embodiments, the oligonucleotide is about 15 to 30 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is about 15 to 30 nucleotides long and comprises greater than 4 DNA bases. In some embodiments, the oligonucleotide is about 15 to 30 nucleotides long and comprises 4-10 DNA bases. In some embodiments, the oligonucleotide is 15-18 nucleotides, 18-22 nucleotides, 22-25 nucleotides, 25-30 nucleotides long, and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 15-18 nucleotides, 18-22 nucleotides, 22-25 nucleotides, 25-30 nucleotides long, and comprises greater than 4 DNA bases. In some embodiments, the oligonucleotide is 15-18 nucleotides, 18-22 nucleotides, 22-25 nucleotides, 25-30 nucleotides long, and comprises 4-10 DNA bases. In some embodiments, the oligonucleotide is 10-40 nucleotides long and comprises 4 DNA bases. In some embodiments, the oligonucleotide is 10-40 nucleotides long and comprises greater than 4 DNA bases. In some embodiments, the oligonucleotide is 10-40 nucleotides long and comprises 4-10 DNA bases. In some embodiments, the oligonucleotide is between 10-15, 15-20, 20-25, 25-30, or 35-40 nucleotides long, including all discrete intervening quantities, and comprises 4 DNA bases. In some embodiments, the oligonucleotide is between 10-15, 15-20, 20-25, 25-30, or 35-40 nucleotides long, including all discrete intervening quantities, and comprises greater than 4 DNA bases. In some embodiments, the oligonucleotide is between 10-15, 15-20, 20-25, 25-30, or 35-40 nucleotides long, including all discrete intervening quantities, and comprises 4-10 DNA bases.

[95] The oligonucleotide of the present disclosure comprises RNA and DNA bases in a particular ratio, for example, the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 9:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 8:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 7:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 6:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 5:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 4:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 3:1. In some embodiments, the oligonucleotide comprises RNA and DNA bases in any ratio that is suitable for the present methods.

[96] In some embodiments, a hybrid oligonucleotide complementary to a sequence in the 5' untranslated region of the mRNA adjacent to the cap or uncapped mRNA is added to an mRNA sample comprising capped mRNA and uncapped mRNA under conditions that permit the oligonucleotide to

anneal to the specified sequence at the 5' end, including the untranslated region. In some embodiments, the oligonucleotide complementary to a sequence in the 5' untranslated region of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein italicized CTGT are DNA bases.

[97] In some embodiments, hybrid oligonucleotides complementary to a sequence in the 3' untranslated region of the mRNA are added to an mRNA sample comprising a poly A tail and/or untailed mRNA under conditions that permit the oligonucleotide to anneal to the specified sequence at the 3' end, including the untranslated region. In some embodiments, the oligonucleotide complementary to a sequence in the 3' untranslated region of the RNA is 3'-GGUCGGAACAGGAUUUUUAATTCAA-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.

[98] Embodiments of the present disclosure are not limited by the type or size of oligonucleotide. In some embodiments, the oligonucleotide comprises between 10-40 nucleotides; e.g., 10 nucleotides, 15 nucleotides, 20 nucleotides, 25 nucleotides, 30 nucleotides, 35 nucleotides, 40 nucleotide, and all discrete intervening quantities.

[99] The size of the oligonucleotide can be selected to generate a capped fragment (if present) of desired length. The oligonucleotide may also be designed to hybridize to any region of the 5' untranslated region depending on where cleavage is desired; i.e., can be positioned within the 5' untranslated region to produce a capped fragment (if present) of any size. In particular, a suitably designed oligonucleotide comprising a small stretch of DNA bases (e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, or 20 nucleotides) flanked by RNA bases (e.g., 1-15) on each side (i.e., a "gapmer") can be annealed to the mRNA analyte. Designing such an oligonucleotide to bind to the complementary bases of the 5' untranslated region of the mRNA allows for select cleavage via DNA/RNA hybrid recognition by a nuclease, e.g., RNase H. Similarly, a suitably designed oligonucleotide may comprise a small stretch of DNA bases (e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18 or 20 nucleotides) flanked only at one end (e.g., the 5' or 3' end) by RNA bases (e.g., 1-15).

[100] In some embodiments, the oligonucleotide is designed to bind directly adjacent to the cap and/or penultimate base of the mRNA, allowing for the resulting cleaved bases to consist of the initial one, two, three, four (or a few, e.g., up to 15) bases of the mRNA construct. Without wishing to be bound by any particular theory, it is contemplated that small capped and uncapped fragments improve the resolution of chromatographic separation of capped and uncapped fragments and thereby improve accuracy of identification and/or quantification of capped and uncapped species. Generally, the smaller the fragment, the better the separation and quantification. In some embodiments, the first base of the oligonucleotide binds at the penultimate base of the mRNA, adjacent to the cap. In some embodiments,

the first base of the oligonucleotide binds adjacent to the penultimate base of the mRNA. In some embodiments, the first base of the oligonucleotide binds at least 2 nucleotides, at least 3 nucleotides, at least 4 nucleotides, at least 5 nucleotides, at least 6 nucleotides, at least 7 nucleotides, at least 8 nucleotides, at least 9 nucleotides or at least 10 nucleotides from the penultimate base of the mRNA.

Annealing of Hybrid Oligonucleotide to RNA sample

[101] The methods of the present disclosure comprise a step of annealing an RNA sample with a first oligonucleotide probe complementary to a sequence in the 5' UTR of the RNA, and a second oligonucleotide probe complementary to a sequence in the 3' UTR of the RNA.

[102] In some embodiments, RNA from an exemplary sample and exemplary custom hybrid oligonucleotides are heated to 75°C for 10 minutes which causes denaturation and strand separation, following which the temperature is decreased from 75°C to 23°C and maintained at 23°C for 10 minutes, facilitating hydrogen bond formation between complementary DNA and RNA sequences. Complementary base pairing between hybrid oligonucleotides and mRNA in a specific region results in RNA:DNA hybrid formation at the end of the annealing step. (FIG. 2).

[103] In some embodiments, heating is at a temperature of between 60°C-95°C. In some embodiments, heating is at a temperature of between 60°C-80°C. In some embodiments, heating is at a temperature of between 60°C-70°C. In some embodiments, heating is at 65°C-70°C.

[104] In some embodiments, heating is for 5 minutes, 10 minutes, 15 minutes, or 20 minutes. In some embodiments, RNA is heated for 15 minutes at a temperature of between 65°C-70°C. At higher temperatures, RNA is heated for less than 15 minutes. In some embodiments, RNA is heated for 10 minutes at 70°C-75°C.

Digestion of RNA: DNA hybrid

[105] In some embodiments, in the digestion step, the RNA:DNA hybrid is incubated with one or more nucleases that are used to digest or cleave at least one strand of an RNA:DNA hybrid. In some embodiments, multiple nucleases are used in a single reaction to produce, for example, a capped fragment. In some embodiments, multiple nucleases are employed to produce a capped fragment and generate a blunt-ended capped fragment. In some embodiments, multiple nucleases are employed to produce a tailed fragment.

[106] Ribonuclease H is a family of endonuclease enzymes with a shared substrate specificity for the RNA strand of RNA-DNA duplexes. RNase H cleaves phosphodiester bonds in the RNA backbone to generate a 3 hydroxyl and a 5 phosphate group. In some embodiments, a suitable nuclease is RNase H or an enzyme with RNase H-like biochemical activity. RNase H comprises two phylogenetic subtypes, Type

1 and Type 2. RNase H binds a single-stranded (ss) RNA that is hybridized to a complementary DNA single strand, and then degrades the RNA portion of the RNA:DNA hybrid. RNase H plays a role in DNA replication, recombination, and repair. *In vitro*, the enzymes will also bind double-stranded (ds) DNA, ssDNA, ssRNA, and dsRNA, albeit with lower affinities than they bind to RNA: DNA hybrids. There are several sequences for RNase H known in the literature, each of which vary somewhat in their amino acid sequences. The present disclosure utilizes any RNase H enzyme, including but not limited to RNase H disclosed in U.S. Patent No. 5,268,289 and 5,500,370 (thermostable RNase H), U.S. Patent No. 6,376,661 (human RNase H), U.S. Patent No. 6,001,652 (human type 2 RNase H), U.S. Patent No. 6,071,734 (RNase H from HBV polymerase). In the method of the present disclosure, since the DNA bases in the hybrid oligonucleotide are flanked on both sides by one or more RNA nucleotides, nuclease cleavage selectively occurs at a specific site.

[107] In some embodiments, the nuclease selectively degrades an RNA:DNA hybrid and/or unannealed mRNA, resulting in capped and uncapped 5' fragments. In some embodiments, the nuclease selectively degrades an RNA:DNA hybrid resulting in tailed and untailed 3' fragments. In some embodiments, at least a portion of the fragment is double-stranded. In some embodiments, the double-stranded portion is at least partially an RNA:RNA hybrid. In some embodiments, the double-stranded portion is at least partially an RNA:DNA hybrid. Fragments resulting from nuclease treatment may be blunt-ended or staggered. In some embodiments, the fragments are between 2-20 nucleotides (including a cap nucleotide if present); i.e., fragments resulting from nuclease treatment can be 20 nucleotides, 19 nucleotides, 18 nucleotides, 17 nucleotides, 16 nucleotides, 15 nucleotides, 14 nucleotides, 13 nucleotides, 12 nucleotides, 11 nucleotides, 10 nucleotides, 9 nucleotides, 8 nucleotides, 7 nucleotides, 6 nucleotides, 5 nucleotides, 4 nucleotides, 3 nucleotides or 2 nucleotides.

[108] In some embodiments, the capped and uncapped fragments comprise no more than 5 bases of the mRNA. In some embodiments, the capped and uncapped fragments comprise no more than 2 bases of the mRNA. In some embodiments, the tailed and untailed fragments comprise no more than 5 bases of the mRNA. In some embodiments, the tailed and untailed fragments comprise no more than 2 bases of the mRNA.

[109] In some embodiments, a suitable nuclease is S1 nuclease. Additional nucleases used either alone or in combination, include, but are not limited to Benzonase®, Nuclease P1, Phosphodiesterase II, RNase A, and RNase T1. In some embodiments, multiple nucleases are used; e.g., RNase H and S1 nuclease. Some embodiments further comprise addition of a single-stranded DNA nuclease to produce or modify the fragment. In some embodiments, it may be desired to heat the sample

(e.g., to about 60°C to 75°C) or apply the sample to a heated chromatographic column in order to produce the capped, uncapped, tailed or untailing fragments.

[110] In some embodiments, digestion is carried out by incubating with RNase H and shrimp alkaline phosphatase (rSAP) at 37°C for 40 minutes. In some embodiments, digestion is carried out by incubating with RNase H and shrimp alkaline phosphatase (rSAP) at 28°C-45°C. In some embodiments, digestion is carried out by incubating with RNase H and shrimp alkaline phosphatase (rSAP) for between 20-60 minutes. RNase H enzymes cleave the phosphodiester bond between double-stranded RNA:DNA hybrids generated in the annealing step. Shrimp alkaline phosphatase (rSAP) catalyzes dephosphorylation of ends to prevent self-ligation. (FIG. 3).

Chromatographic Analysis

[111] Some embodiments of the present disclosure comprise chromatographic methods of identifying and quantitating mRNA capping and tailing modifications. Chromatographic embodiments of the present disclosure can be used to identify and quantify any of the cap structure and cap analogs described herein, as well as various modifications within the caps and/or presence and absence of a tail.

[112] In some embodiments, capping and untailing species are quantified in a single LC-UV or LC-MS or LC-UV-MS analysis. In some embodiments, capping and untailing species are quantified in a single LC-UV analysis. In some embodiments, capping and untailing species are quantified in a single LC-MS analysis. In some embodiments, capping and untailing species are quantified in a single LC-UV-MS analysis. In some embodiments, capping and untailing species are measured by Ultra High Performance Liquid Chromatography-Electrospray Ionization Mass Spectrometry (UHPLC-ESI-MS).

[113] UV detection is particularly valuable in good manufacturing practice (GMP) environment. In some embodiments, the UV detection is at 260 nm. In some embodiments, the UV detection is at 280 nm. In some embodiments, 0.1 mm to 100 mm UV cell is used. In some embodiments, 10 mm UV cell is used. In some embodiments, 0.01 mm UV cell is used. Besides the sample, in some embodiments, a standard is injected in parallel for Cap1, Cap0, CapG and uncapped species to identify and compare retention time for each species.

[114] In some embodiments, tailed species are characterized in a single LC-MS or LC-UV-MS analysis. In some embodiments, tailed species are characterized in a single LC-MS analysis. In some embodiments, tailed species are characterized in a single LC-UV-MS analysis.

[115] In some embodiments, one or more steps are automated. In some embodiments, one step is automated. In some embodiments, more than one step are automated. In some embodiments, all steps are automated.

[116] The nuclease-treated sample is applied to a chromatographic column to separate, for example, capped and uncapped fragments, tailed and untailed fragments. In addition to separating capped from uncapped fragment, chromatography can resolve and quantitate a methylated cap from an unmethylated guanine cap. In some embodiments, a methylated penultimate base (2'-O-methylated base) can be separated (resolved) and quantitated from a cap unmethylated at that position. The capped and uncapped fragments are resolved (i.e., separated from one another) by chromatography and identified by the peaks. The amount of capped and uncapped fragments can be quantitated by standard quantitative chromatography techniques, for example UHPLC peak integration.

[117] Embodiments of the present disclosure utilize chromatography to provide highly resolved (e.g. single base resolution) capped and uncapped, untailed and tailed fragments. Fragments can be efficiently resolved by ultra-high performance liquid chromatography ("UHPLC"). In the context of the present disclosure, the term "UHPLC" includes various UHPLC methods as well as normal pressure liquid chromatography methods, which may be used to carry out some embodiments of the present disclosure. In some embodiments, fragments may be resolved by reverse phase-ultra high-performance liquid chromatography (RP-UHPLC). In some embodiments of the present disclosure, the major peak in the chromatogram is capped mRNA fragments. Parameters that may be altered or optimized to increase resolution include gradient conditions, organic modifier, counter ion, temperature, column pore size and particle size, solvent composition and flow rate.

[118] In some embodiments, the quantitative methods described herein can include one or more steps of ion exchange chromatography-HPLC (e.g., anion exchange-HPLC and/or cation exchange-HPLC). As will be known by those skilled in the art, ion exchangers (e.g., anion exchangers and/or cation exchangers) may be based on various materials with respect to the matrix as well as to the attached charged groups.

[119] In some embodiments, fragments are resolved by reverse phase-UHPLC. Reversed phase UHPLC consists of a non-polar stationary phase and a moderately polar mobile phase. In some embodiments, the stationary phase is a silica which has been treated with, for example, RMe_2SiCl , where R is a straight chain alkyl group such as $\text{C}_{18}\text{H}_{37}$ or C_8H_{17} . The retention time is therefore longer for molecules which are more non-polar in nature, allowing polar molecules to elute more readily. Retention time is increased by the addition of polar solvent to the mobile phase and decreased by the addition of more hydrophobic solvent. The characteristics of the specific RNA molecule as an analyte may play an important role in its retention characteristics. In general, an analyte having more non-polar functional groups (e.g., methyl groups) results in a longer retention time because it increases the molecule's

hydrophobicity. Protocols for high resolution of RNA species using reverse phase-UHPLC, which may be adapted for use in embodiments of the present disclosure, are known in the art (see, e.g., U.S. publication 2010/0048883; Gilar, M., "Analysis and purification of synthetic oligonucleotides by reversed-phase high-performance liquid chromatography with photodiode array and mass spectrometry detection", *Anal. Biochem.* 298: 196-206 (2001)).

[120] Particular embodiments utilize combinations of the various chromatographic separations disclosed herein. For example, particular embodiments of the present disclosure may utilize reverse-phase ion-pair chromatography, whereby separations are based on both hydrophobicity and on the number of anions associated with the molecule, which may be used to purify fragments in a single HPLC step. Matrices can be silica-based (e.g., Murray et al., *Anal. Biochem.*, 218:177-184 (1994)). Non-porous, inert polymer resins may be used in particular embodiments (see, e.g., Huber, C.G., "High-resolution liquid chromatography of oligonucleotides on nonporous alkylated styrene-divinylbenzene copolymers", *Anal. Biochem.*, 212: 351-358 (1993)). Other combinations may be equally effective and must be evaluated in terms of the size of the fragment and the modifications sought to be resolved.

[121] In some embodiments, a capping efficiency profile and/or methylation profile may be determined by strong anion exchange chromatography using a HPLC system. In general, uncapped mRNA adsorbs onto the fixed positive charge of a strong anion exchange column using a mobile phase at a predetermined flow rate elutes capped species (the cap bearing a lower negative charge than uncapped species) from the column in proportion to the strength of their ionic interaction with the positively charged column. More negatively charged (more acidic) uncapped species elute later than less negatively charged (less acidic) capped species.

[122] In certain embodiments, capped fragments are characterized by the methylation profile associated with the fragment. Typically, methylation profiles reflect and quantitate the efficiency of methylation of the cap guanine base (N-7 position). Additional embodiments can also simultaneously quantitate methylation of the 2'-O position of the ribose ring for the penultimate base (Cap1 structure).

[123] In some embodiments, a methylation profile may be determined by performing reverse phase-HPLC, alone or in combination with ion exchange chromatography. In some embodiments, a "methylation profile" refers to a set of values representing the amount of methylated capped fragment that elutes from a column at a point in time after addition to the column of a mobile phase. As described above, the retention time for methylated caps and penultimate nucleotides, which are more non-polar in nature, is increased relative to polar molecules, which elute more readily. Retention time may be

increased by the addition of polar solvent to the mobile phase and decreased by the addition of more hydrophobic solvent.

[124] In some embodiments, quantitative analysis of capped fragment may also be performed using reverse phase-HPLC columns packed, as a non-limiting example, with 2.5 μm fully porous C18 sorbent, as described in Gilar, M., *Anal. Biochem.* 298: 196-206 (2001). Parameters that may be optimized to enhance oligonucleotide mass transfer in the stationary phase include elevated temperature, small sorbent particle size, and slow mobile phase flow rate. A triethylammonium acetate (TEAA) buffer with UV detection and an optimized TEA-HFIP mobile phase may be used for LC-MS separation and characterization of capped fragments.

[125] In some embodiments, ultra-high performance liquid chromatography ("UHPLC") is used to resolve capped and uncapped fragments, and optionally to provide additional quantitative information on methylation states. UHPLC refers generally to HPLC techniques using resin particle sizes less than 2.5 μm , which provides a significant gain in efficiency even at increased flow rates and linear velocities. By using small particles, speed and peak capacity (number of peaks resolved per unit time) can be extended. Such techniques utilize chromatographic principles to run separations using columns packed with smaller particles and/or higher flow rates for increased speed, with superior resolution and sensitivity. (see, e.g., Swartz, M.E., "Ultra Performance Liquid Chromatography (UPLC): an introduction", *Separation Science Redefined* (2005).)

[126] In some embodiments, hydrophilic interaction chromatography (HILIC) methods are used. Typically, HILIC is a variation of reversed phase chromatography performed using a polar stationary phase with an affinity for polar analytes. A mixture comprising one or more polar analytes to be separated from the mixture is added to the polar stationary phase of the column, and a highly organic mobile phase comprising an alcohol, acetonitrile and/or aprotic solvent comprising a very small percentage of aqueous solvent, buffer or other polar solvent is also added to the column to promote passage of the analytes through the stationary phase. Water present in the mobile phase associates with the polar stationary phase, increasing the affinity of polar analytes such as nucleic acids for the stationary phase. More polar analytes, such as longer nucleic acids, have stronger affinities for the stationary phase, and are thus retained in the column for longer, thereby allowing HILIC methods to separate nucleic acids by length. HILIC methods using a mobile phase that is compatible with downstream applications such as mass spectrometry allow the mass of the purified nucleic acid(s) to be analyzed by mass spectrometry.

[127] HILIC is carried out by various methods known in the art. Some exemplary non-limiting volatile salts used in HILIC include ammonium bicarbonate, ammonium acetate, and ammonium formate.

Organic solvents include, but are not limited to, for example, methanol, acetonitrile, and isopropanol. Ion pairing agents, for example, include, but are not limited to octylamine, nonafluoro-tert-butyl alcohol, diethylammonium acetate, and dibutylammonium acetate. An “ion pairing agent” or an “ion pair” refers to an agent (e.g., a small molecule) that functions as a counter ion to a charged (e.g., ionized or ionizable) functional group on an analyte (e.g., a nucleic acid) and thereby changes the retention time of the analyte as it moves through the stationary phase of a column. Ion pairing agents are classified as cationic ion pairing agents (which interact with negatively charged functional groups) or anionic ion pairing agents (which interact with positively charged functional groups). Typically, a volatile salt and ion pairing agent(s) is dissolved in a solution of the organic solvent in water to prepare a first mobile phase, and in a less concentrated solution of the same organic solvent in water to prepare a second mobile phase. In each mobile phase, a first and second ion pairing agent are combined (for example, in a ratio between 1:10 and 10:1), with each ion pairing agent having a final concentration of 0.1 mM - 100 mM. A buffer containing the volatile salt is added to the mobile phase to a final volatile salt concentration of 1 mM - 100 mM, to promote ionization of the eluted mRNA and reduce the charge state of the mRNA during ionization. mRNA compositions are added to HILIC columns at a temperature between 20°C and 60°C, and a mobile phase was passed through the column. After elution, purified mRNAs are ionized and analyzed by mass spectrometry.

[128] In some aspects, liquid chromatography coupled with UV detection (LC-UV), liquid chromatography coupled with mass spectrometry (LC-MS) or liquid chromatography coupled with UV and mass spectrometry (LC-UV-MS) are used for analysis. Current protocols can differentiate any combination of the RNA capped, uncapped, tailed or untailed species that may arise. Various aspects of chromatographic embodiments are discussed in more detail below.

Liquid Chromatography-UV (LC-UV)

[129] In some aspects, liquid chromatography uses a ultraviolet (UV) detector which an in-line device that measures the UV absorbance of the LC (e.g. High-performance liquid chromatography, HPLC) eluent and provides a continuous signal that can be used to quantify the amount of compounds emerging from the LC column. In some embodiments, the UV detector has a fixed wavelength (e.g., 260 nm, 280 nm, etc.). In some embodiments, the UV detector has a variable wavelength. In some embodiments, the detector is a photodiode array detector. In preferred embodiments, UV detector is used in a good manufacturing practice (GMP) environment.

Liquid Chromatography-Mass Spectrometry (LC-MS)

[130] LC-MS combines the physical separation capabilities of liquid chromatography (or HPLC) with the mass analysis capabilities of MS. LC-MS is a highly sensitive and selective technique. Various ionization sources, mass analyzers, detectors and different statistical methods are employed for data analysis.

[131] LC-MS methods are known in the art for oligonucleotide separation and identification using aqueous triethylammonium–hexafluoroisopropylalcohol (TEA HFIP) buffers compatible with MS detection (Apffel, A., et al., “New procedure for the use of HPLC-ESI MS for the analysis of nucleotides and oligonucleotides”, *J. Chromatogr. A*, 777: 3-21 (1997)). Alternatively, a triethylammonium bicarbonate mobile phase may be used for oligonucleotide separation with postcolumn acetonitrile addition to the eluent. The ion-pairing buffer may be chosen to give the best MS detection sensitivity.

[132] In some embodiments, electrospray ionization (ESI) is used as an ionization system in LC-MS. Briefly, samples are introduced into the LC-MS system through an electrospray probe consisting of a metallic capillary, and a high voltage is applied to the capillary while the sampling orifice is at low voltage. Heat and voltage applied to the probe create a fine spray of sample flow. At low LC flow rates, the potential difference is sufficient to create the spray. In some embodiments, nitrogen gas flow is provided for higher LC flow rates. The electrical field at the capillary tip forms positive or negatively charged droplets of the ionized compounds depending on the polarity of the applied voltage. ESI MS negative mode charges the sample through deprotonation, and positive ion mode, charges the analyte through protonation.

[133] In some embodiments, quantitation of capped fragments and the methylation status thereof, uncapped fragment, tailed and untailed fragment is achieved by automated integration of respective peak area in the HPLC chromatogram. Data may be presented as area percent value, which refers to the percentage of a particular species' integrated peak area relative to the total integrated peak area of the entire chromatograph.

[134] In some embodiments, quantitation of capped, uncapped, tailed and untailed fragments may be achieved through other appropriate methods, for example, liquid chromatography coupled with UV and mass spectroscopy (MS)-based detection (LC-UV-MS).

[135] In some embodiments of the present disclosure, capping and untailed species are quantified in a single LC-UV or LC-MS or LC-UV-MS analysis.

[136] In some embodiments, tailed species are characterized in a single LC-MS or LC-UV-MS analysis.

[137] In some embodiments, one or more steps are automated.

RNA Sample

[138] Embodiments of the present disclosure may be used to quantify the capping and tailing modifications of a wide variety of RNA species, including *in vitro* transcribed mRNA, isolated eukaryotic mRNA, and viral RNA.

[139] In some embodiments, the methods of the present disclosure are used to identify or quantify RNA capping and tailing modifications in linear RNA. In some embodiments, the methods of the present disclosure are used to identify or quantify RNA capping and tailing modifications in messenger RNA. In some embodiments, the methods of the present disclosure are used to identify or quantify RNA capping and tailing modifications in a small RNA.

[140] In some embodiments, RNA is obtained from a manufacturing step. In some embodiments, RNA is obtained from the final step of manufacturing. In some embodiments, RNA is a reformulated drug product.

Synthesis of mRNA

[141] mRNAs may be synthesized according to any of a variety of known methods. For example, mRNAs may be synthesized via *in vitro* transcription (IVT). Briefly, IVT is typically performed with a linear or circular DNA template containing a promoter, a pool of ribonucleotide triphosphates, a buffer system that may include DTT and magnesium ions, and an appropriate RNA polymerase (e.g., T3, T7, or SP6 RNA polymerase), DNase I, pyrophosphatase, and/or RNase inhibitor. The exact conditions will vary according to the specific application.

[142] In some embodiments, for the preparation of mRNA according to the present disclosure, a DNA template is transcribed *in vitro*. A suitable DNA template typically has a promoter, for example a T3, T7 or SP6 promoter, for *in vitro* transcription, followed by desired nucleotide sequence for desired mRNA and a termination signal.

Synthesis of mRNA using T3 RNA Polymerase

In some embodiments, mRNA is produced using T3 RNA Polymerase. T3 RNA Polymerase is a DNA-dependent RNA polymerase from the T3 bacteriophage that catalyzes the formation of RNA from DNA in the 5'→3' direction on either single-stranded DNA or double-stranded DNA, and is able to incorporate modified nucleotide. T3 polymerase is extremely promoter-specific and transcribes only DNA downstream of a T3 promoter. T3 binds to a consensus promoter sequence of 5'-AATTAACCCTCACTAAAGGGAGA-3' (SEQ ID NO: 3).

Synthesis of mRNA using T7 RNA Polymerase

[143] In some embodiments, mRNA is produced using T7 RNA Polymerase. T7 RNA Polymerase is a DNA-dependent RNA polymerase from the T7 bacteriophage that catalyzes the formation of RNA from DNA in the 5'→3' direction. T7 polymerase is extremely promoter-specific and transcribes only DNA downstream of a T7 promoter. T7 binds to a consensus promoter sequence of 5'-TAATACGACTCACTATAGGGAGA-3' (SEQ ID NO: 4). The T7 polymerase also requires a double stranded DNA template and Mg²⁺ ion as cofactor for the synthesis of RNA. It has a very low error rate.

Synthesis of mRNA using SP6 RNA Polymerase

[144] In some embodiments, mRNA is produced using SP6 RNA Polymerase. SP6 RNA Polymerase is a DNA-dependent RNA polymerase with high sequence specificity for SP6 promoter sequences. The SP6 polymerase catalyzes the 5'→3' *in vitro* synthesis of RNA on either single-stranded DNA or double-stranded DNA downstream from its promoter; it incorporates native ribonucleotides and/or modified ribonucleotides and/or labeled ribonucleotides into the polymerized transcript. SP6 binds to a consensus promoter sequence of 5'-ATTTAGGTGACACTATAG-3' (SEQ ID NO: 5). Examples of such labeled ribonucleotides include biotin-, fluorescein-, digoxigenin-, aminoallyl-, and isotope-labeled nucleotides.

DNA Template

[145] Typically, a DNA template is either entirely double-stranded or mostly single-stranded with a suitable promoter sequence (e.g., T3, T7 or SP6 promoter).

[146] Linearized plasmid DNA (linearized via one or more restriction enzymes), linearized genomic DNA fragments (via restriction enzyme and/or physical means), PCR products, and/or synthetic DNA oligonucleotides can be used as templates for *in vitro* transcription, provided that they contain a double-stranded promoter upstream (and in the correct orientation) of the DNA sequence to be transcribed.

[147] In some embodiments, the linearized DNA template has a blunt-end.

[148] In some embodiments, the DNA sequence to be transcribed may be optimized to facilitate more efficient transcription and/or translation. For example, the DNA sequence may be optimized regarding cis-regulatory elements (e.g., TATA box, termination signals, and protein binding sites), artificial recombination sites, chi sites, CpG dinucleotide content, negative CpG islands, GC content, polymerase slippage sites, and/or other elements relevant to transcription; the DNA sequence may be optimized regarding cryptic splice sites, mRNA secondary structure, stable free energy of mRNA, repetitive sequences, RNA instability motif, and/or other elements relevant to mRNA processing and stability; the DNA sequence may be optimized regarding codon usage bias, codon adaptability, internal chi sites,

ribosomal binding sites (e.g., IRES), premature poly A sites, Shine-Dalgarno (SD) sequences, and/or other elements relevant to translation; and/or the DNA sequence may be optimized regarding codon context, codon-anticodon interaction, translational pause sites, and/or other elements relevant to protein folding. Optimization methods known in the art may be used in the present disclosure, e.g., GeneOptimizer by ThermoFisher and OptimumGene™, which are described in US 20110081708, the contents of which are incorporated herein by reference in its entirety.

[149] In some embodiments, the DNA template includes a 5' and/or 3' untranslated region. In some embodiments, a 5' untranslated region includes one or more elements that affect an mRNA's stability or translation, for example, an iron responsive element. In some embodiments, a 5' untranslated region may be between about 50 and 500 nucleotides in length.

[150] In some embodiments, a 3' untranslated region includes one or more of a polyadenylation signal, a binding site for proteins that affect an mRNA's stability of location in a cell, or one or more binding sites for miRNAs. In some embodiments, a 3' untranslated region may be between 50 and 500 nucleotides in length or longer.

[151] Exemplary 3' and/or 5' UTR sequences can be derived from mRNA molecules which are stable (e.g., globin, actin, GAPDH, tubulin, histone, or citric acid cycle enzymes) to increase the stability of the sense mRNA molecule. For example, a 5' UTR sequence may include a partial sequence of a CMV immediate-early 1 (IE1) gene, or a fragment thereof to improve the nuclease resistance and/or improve the half-life of the polynucleotide. Also contemplated is the inclusion of a sequence encoding human growth hormone (hGH), or a fragment thereof to the 3' end or untranslated region of the polynucleotide (e.g., mRNA) to further stabilize the polynucleotide. Generally, these modifications improve the stability and/or pharmacokinetic properties (e.g., half-life) of the polynucleotide relative to their unmodified counterparts, and include, for example modifications made to improve such polynucleotides' resistance to *in vivo* nuclease digestion.

Large-scale mRNA Synthesis

[152] In some embodiments, the mRNA can be synthesized in a large-scale. In some embodiments, mRNA is synthesized in at least 100 mg, 150 mg, 200 mg, 300 mg, 400 mg, 500 mg, 600 mg, 700 mg, 800 mg, 900 mg, 1 g, 5 g, 10 g, 25 g, 50 g, 75 g, 100 g, 250 g, 500 g, 750 g, 1 kg, 5 kg, 10 kg, 50 kg, 100 kg, 1000 kg, or more at a single batch. As used herein, the term "batch" refers to a quantity or amount of mRNA synthesized at one time, e.g., produced according to a single manufacturing setting. A batch may refer to an amount of mRNA synthesized in one reaction that occurs via a single aliquot of enzyme and/or a single aliquot of DNA template for continuous synthesis under one set of conditions.

mRNA synthesized at a single batch would not include mRNA synthesized at different times that are combined to achieve the desired amount.

[153] According to the present disclosure, 1-100 mg of RNA polymerase is typically used per gram (g) of mRNA produced. In some embodiments, about 1-90 mg, 1-80 mg, 1-60 mg, 1-50 mg, 1-40 mg, 10-100 mg, 10-80 mg, 10-60 mg, 10-50 mg of RNA polymerase is used per gram of mRNA produced. In some embodiments, about 5-20 mg of RNA polymerase is used to produce about 1 gram of mRNA. In some embodiments, about 0.5 to 2 grams of RNA polymerase is used to produce about 100 grams of mRNA. In some embodiments, about 5 to 20 grams of RNA polymerase is used to about 1 kilogram of mRNA. In some embodiments, at least 5 mg of RNA polymerase is used to produce at least 1 gram of mRNA. In some embodiments, at least 500 mg of RNA polymerase is used to produce at least 100 grams of mRNA. In some embodiments, at least 5 grams of RNA polymerase is used to produce at least 1 kilogram of mRNA. In some embodiments, about 10 mg, 20 mg, 30 mg, 40 mg, 50 mg, 60 mg, 70 mg, 80 mg, 90 mg, or 100 mg of plasmid DNA is used per gram of mRNA produced. In some embodiments, about 10-30 mg of plasmid DNA is used to produce about 1 gram of mRNA. In some embodiments, about 1 to 3 grams of plasmid DNA is used to produce about 100 grams of mRNA. In some embodiments, about 10 to 30 grams of plasmid DNA is used to produce about 1 kilogram of mRNA. In some embodiments, at least 10 mg of plasmid DNA is used to produce at least 1 gram of mRNA. In some embodiments, at least 1 gram of plasmid DNA is used to produce at least 100 grams of mRNA. In some embodiments, at least 10 grams of plasmid DNA is used to produce at least 1 kilogram of mRNA.

[154] In some embodiments, the concentration of the RNA polymerase in the reaction mixture may be from about 1 to 100 nM, 1 to 90 nM, 1 to 80 nM, 1 to 70 nM, 1 to 60 nM, 1 to 50 nM, 1 to 40 nM, 1 to 30 nM, 1 to 20 nM, or about 1 to 10 nM. In certain embodiments, the concentration of the RNA polymerase is from about 10 to 50 nM, 20 to 50 nM, or 30 to 50 nM. A concentration of 100 to 10000 Units/ml of the RNA polymerase may be used, as examples, concentrations of 100 to 9000 Units/ml, 100 to 8000 Units/ml, 100 to 7000 Units/ml, 100 to 6000 Units/ml, 100 to 5000 Units/ml, 100 to 1000 Units/ml, 200 to 2000 Units/ml, 500 to 1000 Units/ml, 500 to 2000 Units/ml, 500 to 3000 Units/ml, 500 to 4000 Units/ml, 500 to 5000 Units/ml, 500 to 6000 Units/ml, 1000 to 7500 Units/ml, and 2500 to 5000 Units/ml may be used.

[155] The concentration of each ribonucleotide (e.g., ATP, UTP, GTP, and CTP) in a reaction mixture is between about 0.1 mM and about 10 mM, e.g., between about 1 mM and about 10 mM, between about 2 mM and about 10 mM, between about 3 mM and about 10 mM, between about 1 mM and about 8 mM, between about 1 mM and about 6 mM, between about 3 mM and about 10 mM,

between about 3 mM and about 8 mM, between about 3 mM and about 6 mM, between about 4 mM and about 5 mM. In some embodiments, each ribonucleotide is at about 5 mM in a reaction mixture. In some embodiments, the total concentration of rNTPs (for example, ATP, GTP, CTP and UTPs combined) used in the reaction range between 1 mM and 40 mM. In some embodiments, the total concentration of rNTPs (for example, ATP, GTP, CTP and UTPs combined) used in the reaction range between 1 mM and 30 mM, or between 1 mM and 28 mM, or between 1 mM to 25 mM, or between 1 mM and 20 mM. In some embodiments, the total rNTPs concentration is less than 30 mM. In some embodiments, the total rNTPs concentration is less than 25 mM. In some embodiments, the total rNTPs concentration is less than 20 mM. In some embodiments, the total rNTPs concentration is less than 15 mM. In some embodiments, the total rNTPs concentration is less than 10 mM.

[156] The RNA polymerase reaction buffer typically includes a salt/buffering agent, e.g., Tris, HEPES, ammonium sulfate, sodium bicarbonate, sodium citrate, sodium acetate, potassium phosphate sodium phosphate, sodium chloride, and magnesium chloride.

[157] The pH of the reaction mixture may be between about 6 to 8.5, from 6.5 to 8.0, from 7.0 to 7.5, and in some embodiments, the pH is 7.5.

[158] Linear or linearized DNA template (e.g., as described above and in an amount/concentration sufficient to provide a desired amount of RNA), the RNA polymerase reaction buffer, and RNA polymerase are combined to form the reaction mixture. The reaction mixture is incubated at between about 37°C and about 42°C for thirty minutes to six hours, e.g., about sixty to about ninety minutes.

[159] In some embodiments, about 5 mM NTPs, about 0.05 mg/mL RNA polymerase, and about 0.1 mg/ml DNA template in a suitable RNA polymerase reaction buffer (final reaction mixture pH of about 7.5) is incubated at about 37°C to about 42°C for sixty to ninety minutes.

[160] In some embodiments, a reaction mixture contains linearized double stranded DNA template with an RNA polymerase-specific promoter, RNA polymerase, RNase inhibitor, pyrophosphatase, 29 mM NTPs, 10 mM DTT and a reaction buffer (when at 10x is 800 mM HEPES, 20 mM spermidine, 250 mM MgCl₂, pH 7.7) and quantity sufficient (QS) to a desired reaction volume with RNase-free water; this reaction mixture is then incubated at 37°C for 60 minutes. The polymerase reaction is then quenched by addition of DNase I and a DNase I buffer (when at 10x is 100 mM Tris-HCl, 5 mM MgCl₂ and 25 mM CaCl₂, pH 7.6) to facilitate digestion of the double-stranded DNA template in preparation for purification. This embodiment has been shown to be sufficient to produce 100 grams of mRNA.

[161] In some embodiments, a reaction mixture includes NTPs at a concentration ranging from 1-10 mM, DNA template at a concentration ranging from 0.01-0.5 mg/ml, and RNA polymerase at a concentration ranging from 0.01-0.1 mg/ml, e.g., the reaction mixture comprises NTPs at a concentration of 5 mM, the DNA template at a concentration of 0.1 mg/ml, and the RNA polymerase at a concentration of 0.05 mg/ml.

[162] In some embodiments, the RNA comprises one or more modified nucleotides.

[163] In some embodiments, the RNA is unmodified.

Capped RNA

[164] mRNAs and other RNAs bear a 5' terminal "cap" structure, which plays important biological roles in splicing, translation and preventing mRNA degradation, thereby contributing to stability of mRNA. The 5' cap plays a role in processing and maturation of an RNA transcript in the nucleus, e.g., pre-mRNA splicing, mRNA export from the nucleus to the cytoplasm, mRNA stability, and efficient translation of the mRNA to protein.

[165] The 5' cap structure is recognized by eukaryotic translation initiation factor 4E, eIF4E and initiates protein synthesis (Shatkin, A.J., *Cell*, 40: 223-24 (1985); Furuichi, et al., *Nature*, 266: 235 (1977); Sonenberg, N., *Prog. Nuc. Acid Res Mol Biol*, 35: 173-207 (1988)). Specific cap binding proteins exist that are components of the machinery required for initiation of translation of an mRNA (see, e.g., Shatkin, A.J., *Cell*, 40: 223-24 (1985); Sonenberg, N., *Prog. Nuc. Acid Res Mol Biol*, 35: 173-207 (1988)). The cap of mRNA is recognized by the translational initiation factor eIF4E (Gingras, et al., *Ann. Rev. Biochem.* 68: 913-963 (1999); Rhoads, R.E., *J. Biol. Chem.* 274: 30337-3040 (1999)). The 5' cap structure protects the mRNA from 5'-exonuclease activity and resultant degradation (Ross, J., *Mol. Biol. Med.* 5: 1-14 (1988); Green, M.R. et al., *Cell*, 32: 681-694 (1983)). Since the primary transcripts of many eukaryotic cellular genes and eukaryotic viral genes require splicing to remove intervening sequences (introns) within the coding regions of these transcripts, the cap also functions in stabilization of pre-mRNA.

[166] Translation efficiency is increased by capping, and capped RNAs have been reported to be translated more efficiently than uncapped transcripts in a variety of *in vitro* translation systems, such as rabbit reticulocyte lysate or wheat germ translation systems (see, e.g., Shimotohno, K., et al., *Proc. Natl. Acad. Sci. USA*, 74: 2734-2738 (1977); Paterson and Rosenberg, *Nature*, 279: 692 (1979)). Increased mRNA stability and resistance to exonucleases are a contributory factor.

[167] 5' cap also plays a role in pathogen defense, since 2'-O methylation at the 5' penultimate nucleotide of mRNA functions as a molecular signature that discriminates host and invading pathogen mRNA, e.g., viral mRNA, which lacks this feature.

[168] The mRNA cap is typically added enzymatically.

[169] In some embodiments of the method provided herein, the capping species is Cap1, Cap0, CapG or uncapped. In some embodiments, the capping species is Cap1. In some embodiments, the capping species is Cap0. In some embodiments, the capping species is CapG. In some embodiments, the capping species is uncapped.

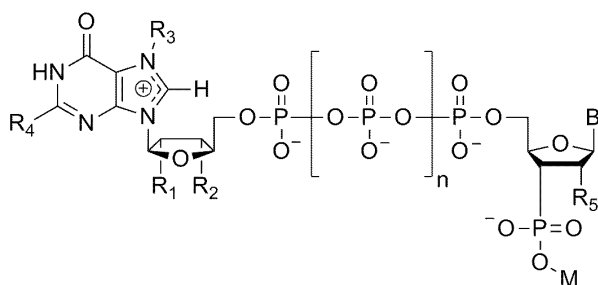
[170] In some embodiments, the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA or GGACA. In some embodiments, the capping species is m7Gpppm7GGACA. In some embodiments, the capping species is m7GpppGGACA. In some embodiments, the capping species is GpppGGACA. In some embodiments, the capping species is GGACA or pppGGACA. In some embodiments, the uncapped species is GGACA or pppGGACA.

[171] In some embodiments, the relative amount of each of the capping species is a percentage of total amount of capping species in the sample calculated by dividing an area under the peak of the capping species of interest over a sum of total areas under the peak representing Cap 1, Cap 0, CapG and uncapped species, multiplied by 100.

[172] Inventive methods described herein are generally amenable to identification and quantification of any type of canonical or non-canonical mRNA cap.

Canonical mRNA cap structures

[173] In some embodiments, the cap has a structure of formula I:



wherein B is a nucleobase, R_1 is selected from a halogen, OH, and OCH_3 , R_2 is selected from H, OH, and OCH_3 , R_3 is CH_3 , CH_2CH_3 , $CH_2CH_2CH_3$ or void, R_4 is NH_2 , R_5 is selected from OH, OCH_3 and a halogen, n is 1, 2, or 3, and M is a nucleotide, i.e., the third base of mRNA. In particular embodiments, B is guanine, but can be any nucleobase. In some embodiments, the cap is m⁷G(5')ppp(5')G in which a 2'-O-methyl residue is present at the 2' OH group of the ribose ring of base 1 (i.e., at the R_5 position of Formula 1).

[174] Naturally occurring cap structures comprise a 7-methyl guanosine that is linked via a triphosphate bridge to the 5'-end of the first transcribed nucleotide, resulting in a dinucleotide cap of $m^7G(5')ppp(5')N$, where N is any nucleoside. In some embodiments, a m^7G cap is $m^7G(5')ppp(5')G$.

[175] The canonical cap on most eukaryotic and viral mRNAs is comprised of N7-methylguanosine (m^7G) linked to the first nucleotide of the RNA by a reverse 5'-5' triphosphate bridge. Cap0 is the predominant form in lower eukaryotes, i.e., m^7GpppN . In higher eukaryotes, the first two 5' nucleotides in the 5' UTR can be 2'-O-methylated to generate $m^7GpppNm$ (Cap1) and $m^7GpppNmpNm$ (Cap 2) structures.

[176] The cap is added in the nucleus and is catalyzed by the enzyme guanylyl transferase. The addition of the cap to the 5' terminal end of RNA occurs immediately after initiation of transcription. The terminal nucleoside is typically a guanosine, and is in the reverse orientation to all the other nucleotides, i.e., $G(5')ppp(5')GpNpNp$.

[177] In some embodiments, the cap is a Cap0 structure. Cap0 structures lack a 2'-O-methyl residue of the ribose attached to bases 1 and 2. In some embodiments, the cap is a Cap1 structure. Cap1 structures have a 2'-O-methyl residue at base 1. In some embodiments, the cap is a Cap2 structure. Cap2 structures have a 2'-O-methyl residue attached to both bases 1 and 2 (FIG. 4B).

Non-canonical mRNA cap structures

[178] Several non-canonical mRNA caps are also recognized, e.g. $GpppX$ variants, nn-methylated guanosine cap ($GpppN$). Some cap variants are additionally methylated at N6 of Am (m^6Am). Multiple methylations also occur on cap 5' G such as di- and trimethylguanosine caps ($m^2,2,7GpppN$) such as small nuclear and nucleolar RNAs, telomerase RNAs and also some viral RNAs. Mammalian U6 and 7SK RNAs comprise phosphate methylation of unprocessed 5' triphosphate ($mpppN$).

[179] A common cap for mRNA produced by *in vitro* transcription is $m^7G(5')ppp(5')G$, which has been used as the dinucleotide cap in transcription with T7 or SP6 RNA polymerase *in vitro* to obtain RNAs having a cap structure in their 5'-termini. The prevailing method for the *in vitro* synthesis of capped mRNA employs a pre-formed dinucleotide of the form $m^7G(5')ppp(5')G$ (" m^7GpppG ") as an initiator of transcription. A disadvantage of using $m^7G(5')ppp(5')G$, a pseudosymmetrical dinucleotide, is the propensity of the 3'-OH of either the G or m^7G moiety to serve as the initiating nucleophile for transcriptional elongation. In other words, the presence of a 3'-OH on both the m^7G and G moieties leads to up to half of the mRNAs incorporating caps in an improper orientation. This leads to the synthesis of two isomeric RNAs of the form $m^7G(5')pppG(pN)_n$ and $G(5')pppm^7G(pN)_n$, in approximately equal

proportions, depending upon the ionic conditions of the transcription reaction. Variations in the isomeric forms can adversely affect *in vitro* translation and are undesirable for a homogenous therapeutic product.

[180] To date, the usual form of a synthetic dinucleotide cap used in *in vitro* translation experiments is the Anti-Reverse Cap Analog ("ARCA"), which is generally a modified cap analog in which the 2' or 3' OH group is replaced with -OCH₃. ARCA and triple-methylated cap analogs are incorporated in the forward orientation. Chemical modification of m⁷G at either the 2' or 3' OH group of the ribose ring results in the cap being incorporated solely in the forward orientation, even though the 2' OH group does not participate in the phosphodiester bond. (Jemielity, J. et al., "Novel 'anti-reverse' cap analogs with superior translational properties", *RNA*, 9: 1108-1122 (2003)). The selective procedure for methylation of guanosine at N7 and 3' O-methylation and 5' diphosphate synthesis has been established (Kore, A. and Parmar, G. *Nucleosides, Nucleotides, and Nucleic Acids*, 25:337-340, (2006) and Kore, A. R., et al. *Nucleosides, Nucleotides, and Nucleic Acids*, 25(3): 307-14, (2006).

Cap analogs

[181] Cap analogs may be or comprise any modified "G" base (e.g., one or more modified guanine nucleotides). Suitable cap analogs include, but are not limited to, a chemical structures selected from the group consisting of m⁷GpppG, m⁷GpppA, m⁷GpppC; unmethylated cap analogs (e.g., GpppG); dimethylated cap analog (e.g., m^{2,7}GpppG), trimethylated cap analog (e.g., m^{2,2,7}GpppG), dimethylated symmetrical cap analogs (e.g., m⁷Gpppm⁷G), or anti reverse cap analogs (e.g., ARCA; m^{7,2'Ome}GpppG, m^{7,2'd}GpppG, m^{7,3'Ome}GpppG, m^{7,3'd}GpppG and their tetraphosphate derivatives).

[182] A variety of m⁷G cap analogs are known in the art, many of which are commercially available. These include the m⁷GpppG described above, as well as the ARCA 3'-OCH₃ and 2'-OCH₃ cap analogs described above (Jemielity, J. et al., *RNA*, 9: 1108-1122 (2003)). Additional cap analogs for use in embodiments of the present disclosure include N7-benzylated dinucleoside tetraphosphate analogs (described in Grudzien, E. et al., *RNA*, 10: 1479-1487 (2004)), phosphorothioate cap analogs (described in Grudzien-Nogalska, E., et al., *RNA*, 13: 1745-1755 (2007)), and cap analogs (including biotinylated cap analogs) described in U.S. Patent Nos. 8,093,367 and 8,304,529, incorporated by reference herein.

[183] In some embodiments, mRNA is uncapped. Uncapped mRNA may be present in a sample (i.e., as a result of incomplete capping in an *in vitro* transcription reaction) and/or may be provided as a standard measured in parallel to quantify the level of uncapped species in a sample.

Production of Capped mRNAs

[184] Capped mRNAs suitable for identification and/or quantification by the methods disclosed herein may be produced by any method known in the art.

[185] In some embodiments, capped mRNA is produced by *in vitro* transcription, originally developed by Krieg and Melton (*Methods Enzymol.*, 1987, 155: 397-415) for the synthesis of RNA using an RNA phage polymerase. Typically, these reactions include at least a phage RNA polymerase (T7, T3 or SP6), a DNA template containing a phage polymerase promoter, nucleotides (ATP, CTP, GTP and UTP), and a buffer containing a magnesium salt. RNA synthesis yields may be optimized by increasing nucleotide concentrations, adjusting magnesium concentrations and by including inorganic pyrophosphatase (U.S. Pat. No. 5,256,555; Gurevich, et al., *Anal. Biochem.*, 195: 207-213 (1991); Sampson, J.R. and Uhlenbeck, O.C., *Proc. Natl. Acad. Sci. USA*, 85, 1033-1037 (1988); Wyatt, J.R., et al., *Biotechniques*, 11: 764-769 (1991)). Some embodiments utilize commercial kits for the large-scale synthesis of *in vitro* transcripts (e.g., MEGAscript®, Ambion). The RNA synthesized in these reactions is usually characterized by a 5' terminal nucleotide that has a triphosphate at the 5' position of the ribose. Typically, depending on the RNA polymerase and promoter combination used, this nucleotide is a guanosine, although it can be an adenosine (see e.g., Coleman, T. M., et al., *Nucleic Acids Res.*, 32: e14 (2004)). In these reactions, all four nucleotides are typically included at equimolar concentrations and none of them is limiting.

[186] In some embodiments, all components are combined and incubated at about 37°C to promote the polymerization of the RNA in a single batch reaction until the reaction terminates. Typically, a batch reaction is used for convenience and to obtain larger quantities of RNA. In some embodiments, a "fed-batch" system (see, e.g., Jeffrey A. Kern, *Batch and Fed-batch strategies for large-scale production of RNA by in vitro transcription* (University of Colorado). (1997)) is used to increase the efficiency of the *in vitro* transcription reaction. All components are combined, but additional amounts of some of the reagents are added over time, such as nucleotides and magnesium, to maintain constant reaction conditions as the reaction is scaled up in a single batch. In some embodiments, additional quantities of reagents needed are not linear increases from a smaller reaction. In addition, in some embodiments, the pH of the reaction may be held at 7.4 by monitoring it over time and adding KOH as needed.

[187] Transcription of RNA usually starts with a nucleoside triphosphate (usually a purine, A or G). *In vitro* transcription typically comprises a phage RNA polymerase such as T7, T3 or SP6, a DNA template containing a phage polymerase promoter, nucleotides (ATP, GTP, CTP and UTP) and a buffer containing magnesium salt. The synthesis of capped RNA includes the incorporation of a cap analog (e.g., m⁷GpppG) in the transcription reaction, which in some embodiments is incorporated by the addition of recombinant guanylyl transferase. Excess m⁷GpppG to GTP (4:1) increases the opportunity that each transcript will have a 5' cap.

[188] Kits for capping of *in vitro* transcribed mRNAs are commercially available, including the mMESSAGE mMACHINE® kit (Ambion, Inc., Austin, Tex.). These kits will typically yield 80% capped RNA to 20% uncapped RNA, although total RNA yields are lower as GTP concentration becomes rate limiting as GTP is needed for the elongation of the transcript.

[189] As the ratio of the cap analog to GTP increases in the reaction, the ratio of capped to uncapped RNA increases proportionally. Increasing the ratio of cap analog to GTP in the transcription reaction produces lower yields of total RNA because the concentration of GTP becomes limiting when holding the total concentration of cap and GTP constant. Thus, the final RNA yield is dependent on GTP concentration, which is necessary for the elongation of the transcript. The other nucleotides (ATP, CTP, UTP) are present in excess.

[190] Thus, the present disclosure provides improved methods of simultaneously identifying and quantitating mRNA capping and tailing modifications in a single sample (e.g., a representative aliquot sample from an *in vitro* synthesis reaction, RNA from a manufacturing step or de formulated drug product), without need for purification of intermediates that result in loss of sample and/or decreased accuracy of measurement.

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

RNA Tailing

[192] The presence of a "tail" at a 3' end serves to protect the mRNA from exonuclease degradation. The 3' tail may be added before, after or at the same time of adding the 5' Cap. Typically, a tail structure includes a poly A and/or poly C tail. (A, adenosine; C, cytosine).

[193] In some embodiments, the poly A tail is added co-transcriptionally. In some embodiments, the poly A tail is added post-transcriptionally. In some embodiments, the poly C tail is added co-transcriptionally. In some embodiments, the poly C tail is added post-transcriptionally.

[194] In some embodiments, the poly A tail is 25-5,000 nucleotides in length. In some embodiments, the poly A tail is 25 nucleotides in length. In some embodiments, the poly A tail is 50

nucleotides in length. In some embodiments, the poly A tail is 75 nucleotides in length. In some embodiments, the poly A tail is 100 nucleotides in length. In some embodiments, the poly A tail is 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000 nucleotides in length, including all discrete intervening quantities. In some embodiments, the poly A tail is between 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, 4000-4500 or 4500-5000 nucleotides in length, including all discrete intervening quantities.

[195] In some embodiments, a poly A or poly C tail on the 3' terminus of mRNA includes at least 25-50, 50-100, 100-200, 200-300, 300-500, or 500-800 adenine or cytosine nucleotides, including all discrete intervening quantities.

[196] In some embodiments, a tail structure includes a combination of poly A and poly C tails with various lengths described herein. In some embodiments, a poly A tail structure includes at least 50%, 55%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 94%, 95%, 96%, 97%, 98%, or 99% adenosine nucleotides. In some embodiments, a poly A tail structure includes at least 50%, 55%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 94%, 95%, 96%, 97%, 98%, or 99% cytosine nucleotides.

[197] In some embodiments, the untailed species is UGCAUC, wherein U is unmodified uridine. In some embodiments, the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine.

[198] In some embodiments, the relative amount of untailed species is a percentage calculated by dividing an area under the peak of the untailed species over a sum of areas under the peak of Cap1, Cap0, CapG and uncapped multiplied by 100.

[199] As described herein, the addition of the 5' cap and/or the 3' tail facilitates the detection of abortive transcripts generated during *in vitro* synthesis because without capping and/or tailing, the size of prematurely aborted mRNA transcripts can be too small to be detected. Thus, in some embodiments, the 5' cap and/or the 3' tail are added to the synthesized mRNA before the mRNA is tested for purity (e.g., the level of abortive transcripts present in the mRNA). In some embodiments, the 5' cap and/or the 3' tail are added to the synthesized mRNA before the mRNA is purified. In other embodiments, the 5' cap and/or the 3' tail are added to the synthesized mRNA after the mRNA is purified.

RNA Modifications

[200] In some embodiments, the linear RNAs of the present disclosure may include one, two, three, or more modifications. In some embodiments, the modified nucleotides are located in coding region(s). In some embodiments, the modified nucleotides are in the untranslated region(s).

[201] In some embodiments, the modifications stabilize the RNA and enhance resistance to degradation as compared to unmodified nucleotides. In some embodiments, modified nucleotides enhance biological functions of nucleic acid molecules, for example, increase binding to an RNA binding protein or increasing translation.

[202] In some embodiments, the modified nucleotide is one or more of N1-methylpseudouridine, 5-methoxyuridine, N6-methyladenosine, pseudouridine or 5-methylcytosine.

[203] In some embodiments, the modified nucleotide is N1-methylpseudouridine. In some embodiments, the modified nucleotide is 5-methoxyuridine. In some embodiments, the modified nucleotide is N6-methyladenosine. In some embodiments, the modified nucleotide is pseudouridine. In some embodiments, the modified nucleotide is 5-methylcytosine.

[204] In some embodiments, the modified nucleotide is 100%. In some embodiments, the modified nucleotide is less than 50%. In some embodiments, the modified nucleotide is less than 20%. In some embodiments, the modified nucleotide is less than 10%.

[205] The linear polynucleotides of the present disclosure may contain from about 0% to about 100% modified nucleotides (either in relation to overall nucleotide content, or in relation to one or more types of nucleotide, i.e. any one or more of A, G, T/U or C) or any intervening percentage (e.g., from 1% to 20%, from 1% to 25%, from 1% to 50%, from 1% to 60%, from 1% to 70%, from 1% to 80%, from 1% to 90%, from 1% to 95%, from 10% to 20%, from 10% to 25%, from 10% to 50%, from 10% to 60%, from 10% to 70%, from 10% to 80%, from 10% to 90%, from 10% to 95%, from 10% to 100%, from 20% to 25%, from 20% to 50%, from 20% to 60%, from 20% to 70%, from 20% to 80%, from 20% to 90%, from 20% to 95%, from 20% to 100%, from 50% to 60%, from 50% to 70%, from 50% to 80%, from 50% to 90%, from 50% to 95%, from 50% to 100%, from 70% to 80%, from 70% to 90%, from 70% to 95%, from 70% to 100%, from 80% to 90%, from 80% to 95%, from 80% to 100%, from 85% to 95%, from 85% to 100%, from 90% to 95%, from 90% to 100%, and from 95% to 100%, including all values and subranges therebetween).

[206] In some embodiments, the polynucleotides are 100% modified. In some embodiments, the polynucleotides are at least 50% modified, e.g., at least 50% of the nucleotides are modified. In some embodiments, the polynucleotides are at least 75% modified, e.g., at least 75% of the nucleotides are modified. In some embodiments, the polynucleotides are at least 20% modified, e.g., at least 20% of the nucleotides are modified. In some embodiments, the polynucleotides are at least 10% modified, e.g., at least 10% of the nucleotides are modified. It is to be understood that since a nucleotide (sugar, base and phosphate moiety, e.g., linkage) may each be modified, any modification to any portion of a nucleotide, or nucleoside, will constitute a modification.

[207] In some embodiments, the modifications are structural modifications and/or chemical modifications. In some embodiments, the chemical modification is a nucleotide and/or nucleoside modification including a nucleobase modification and/or a sugar modification, and a backbone linkage modification (i.e., the internucleoside linkage, e.g., a linking phosphate, a phosphodiester linkage, and a phosphodiester backbone). In some embodiments, the structural modification includes a secondary and/or tertiary structural modification.

[208] In some embodiments, modifications include modifications of ribonucleic acids (RNAs) to deoxyribonucleic acids (DNAs), threose nucleic acids (TNAs), glycol nucleic acids (GNAs), peptide nucleic acids (PNAs), locked nucleic acids (LNAs) or hybrids thereof.

[209] In some embodiments, one, two, or more (optionally different) nucleoside or nucleotide modifications may be incorporated to the polynucleotides of the present disclosure.

[210] In some embodiments, the polynucleotide (e.g., RNA) comprises at least one modification described herein. In other embodiments, the polynucleotides comprise two, three, four, or more (optionally different) chemical modifications described herein. The modifications may be combinations of nucleobase (purine and/or pyrimidine), sugar and backbone (internucleoside) linkage modifications. The modifications may be located on one or more nucleotides of the polynucleotide. In some embodiments, all the nucleotides of the polynucleotide are chemically modified. In some embodiments, all the nucleotides of the nucleic acid sequence with a biological function are chemically modified.

[211] In some embodiments, the polynucleotides are at least 10% modified in only one component of the nucleotide, with such component being the nucleobase, sugar, or linkage between nucleosides. For example, modifications may be made to at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100% of the nucleobases, sugars, or linkages of a polynucleotide described herein.

[212] In some embodiments, the polynucleotides are designed with a patterned array of sugar, nucleobase or linkage modifications.

[213] In some embodiments, the polynucleotides comprise modifications to maximize stability.

[214] In other embodiments, the polynucleotides comprise modifications to decrease stability.

[215] In some embodiments, the modified nucleosides and nucleotides include a modified nucleobase. Examples of nucleobases in RNA include, but are not limited to, adenine(A), guanine(G), cytosine(C), and uracil(U). Examples of nucleobases in DNA include, but are not limited to, adenine(A), guanine(G), cytosine(C), and thymine(T).

[216] In some embodiments, the modified nucleobase is a modified uracil(U). Exemplary nucleobases and nucleosides having a modified uracil include pseudouridine (ψ), pyridin-4-one ribonucleoside, 5-aza-uridine, 6-aza-uridine, 2-thio-5-aza-uridine, 2-thio-uridine (s^2U), 4-thio-uridine (s^4U), 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxy-uridine (ho^5U), 5-aminoallyl-uridine, 5-halo-uridine (e.g., 5-iodo-uridine (I^5U) or 5-bromo-uridine (br^5U)), 3-methyl-uridine (m^3U), 5-methoxy-uridine (mo^5U), uridine 5-oxyacetic acid (cmo^5U), uridine 5-oxyacetic acid methyl ester ($mcmo^5U$), 5-carboxymethyl-uridine (cm^5U), 1-carboxymethyl-pseudouridine, 5-carboxyhydroxymethyl-uridine (chm^5U), 5-carboxyhydroxymethyl-uridine methyl ester ($mchm^5U$), 5-methoxycarbonylmethyl-uridine (mcm^5U), 5-methoxycarbonylmethyl-2-thio-uridine (mcm^5s^2U), 5-aminomethyl-2-thio-uridine (nm^5s^2U), 5-methylaminomethyl-uridine (mnm^5U), 5-methylaminomethyl-2-thio-uridine (mnm^5s^2U), 5-methylaminomethyl-2-seleno-uridine (mnm^5se^2U), 5-carbamoylmethyl-uridine (ncm^5U), 5-carboxymethylaminomethyl-uridine ($cmnm^5U$), 5-carboxymethylaminomethyl-2-thio-uridine ($cmnm^5s^2U$), 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-aurinomethyl-uridine (τm^5U), 1-aurinomethyl-pseudouridine, 5-aurinomethyl-2-thio-uridine (τm^5s^2U), 1-aurinomethyl-4-thio-pseudouridine, 5-methyl-uridine (m^5U , i.e., having the nucleobase deoxythymine), 1-methylpseudouridine ($m^1\psi$), 5-methyl-2-thio-uridine (m^5s^2U), pseudouracil (ψ), 1-methyl-4-thio-pseudouridine ($m^1s^4\psi$), 4-thio-1-methyl-pseudouridine, 3-methyl-pseudouridine ($m^3\psi$), 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine (D), dihydropseudouridine, 5,6-dihydrouridine, 5-methyl-dihydrouridine (m^5D), 2-thio-dihydrouridine, 2-thio-dihydropseudouridine, 2-methoxy-uridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, 4-methoxy-2-thio-pseudouridine, N1-methyl-pseudouridine (also known as 1-methylpseudouridine ($m^1\psi$)), 3-(3-amino-3-carboxypropyl)uridine (acp^3U), 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine ($acp^3\psi$), 5-(isopentenylaminomethyl)uridine (inm^5U), 5-(isopentenylaminomethyl)-2-thio-uridine (inm^5s^2U), α -thio-uridine, 2'-O-methyl-uridine (Um), 5,2'-O-dimethyl-uridine (m^5Um), 2'-O-methyl-pseudouridine (ψm), 2-thio-2'-O-methyl-uridine (s^2Um), 5-methoxycarbonylmethyl-2'-O-methyl-uridine (mcm^5Um), 5-carbamoylmethyl-2'-O-methyl-uridine (ncm^5Um), 5-carboxymethylaminomethyl-2'-O-methyl-uridine ($cmnm^5Um$), 3,2'-O-dimethyl-uridine (m^3Um), 5-(isopentenylaminomethyl)-2'-O-methyl-uridine (inm^5Um), 1-thio-uridine, deoxythymidine, 2'-F-ara-uridine, 2'-F-uridine, 2'-OH-ara-uridine, 5-(2-carbomethoxyvinyl) uridine, and 5-[3-(1-E-propenylamino)uridine.

[217] In some embodiments, the modified nucleobase is a modified cytosine(C). Exemplary nucleobases and nucleosides having a modified cytosine include 5-aza-cytidine, 6-aza-cytidine, pseudoisocytidine, 3-methyl-cytidine (m^3C), N4-acetyl-cytidine (ac^4C), 5-formyl-cytidine (f^5C), N4-methyl-

cytidine (m^4C), 5-methyl-cytidine (m^5C), 5-halo-cytidine (e.g., 5-iodo-cytidine), 5-hydroxymethyl-cytidine (hm^5C), 1-methyl-pseudoisocytidine, pyrrolo-cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine (s^2C), 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxy-cytidine, 2-methoxy-5-methyl-cytidine, 4-methoxy-pseudoisocytidine, 4-methoxy-1-methyl-pseudoisocytidine, lysidine (k_2C), α -thio-cytidine, 2'-O-methyl-cytidine (Cm), 5,2'-O-dimethyl-cytidine (m^5Cm), N4-acetyl-2'-O-methyl-cytidine (ac^4Cm), N4,2'-O-dimethyl-cytidine (m^4Cm), 5-formyl-2'-O-methyl-cytidine (f^5Cm), N4,N4,2'-O-trimethyl-cytidine (m^4_2Cm), 1-thio-cytidine, 2'-F-ara-cytidine, 2'-F-cytidine, and 2'-OH-ara-cytidine.

[218] In some embodiments, the modified nucleobase is a modified adenine(A). Exemplary nucleobases and nucleosides having a modified adenine include 2-amino-purine, 2, 6-diaminopurine, 2-amino-6-halo-purine (e.g., 2-amino-6-chloro-purine), 6-halo-purine (e.g., 6-chloro-purine), 2-amino-6-methyl-purine, 8-azido-adenosine, 7-deaza-adenine, 7-deaza-8-aza-adenine, 7-deaza-2-amino-purine, 7-deaza-8-aza-2-amino-purine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyl-adenosine (m^1A), 2-methyl-adenine (m^2A), N6-methyl-adenosine (m^6A), 2-methylthio-N6-methyl-adenosine (ms^2m^6A), N6-isopentenyl-adenosine (i^6A), 2-methylthio-N6-isopentenyl-adenosine (ms^2i^6A), N6-(cis-hydroxyisopentenyl)adenosine (io^6A), 2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine (ms^2io^6A), N6-glycylcarbamoyl-adenosine (g^6A), N6-threonylcarbamoyl-adenosine (t^6A), N6-methyl-N6-threonylcarbamoyl-adenosine (m^6t^6A), 2-methylthio-N6-threonylcarbamoyl-adenosine (ms^2g^6A), N6,N6-dimethyl-adenosine (m^6_2A), N6-hydroxynorvalylcarbamoyl-adenosine (hn^6A), 2-methylthio-N6-hydroxynorvalylcarbamoyl-adenosine (ms^2hn^6A), N6-acetyl-adenosine (ac^6A), 7-methyl-adenine, 2-methylthio-adenine, 2-methoxy-adenine, α -thio-adenosine, 2'-O-methyl-adenosine (Am), N6,2'-O-dimethyl-adenosine (m^6Am), N6,N6,2'-O-trimethyl-adenosine (m^6_2Am), 1,2'-O-dimethyl-adenosine (m^1Am), 2'-O-ribosyladenosine (phosphate) (Ar(p)), 2-amino-N6-methyl-purine, 1-thio-adenosine, 8-azido-adenosine, 2'-F-ara-adenosine, 2'-F-adenosine, 2'-OH-ara-adenosine, and N6-(19-amino-pentaoxanonadecyl)-adenosine.

[219] In some embodiments, the modified nucleobase is a modified guanine(G). Exemplary nucleobases and nucleosides having a modified guanine include inosine (I), 1-methyl-inosine (m^1I), wyosine (imG), methylwyosine (mimG), 4-demethyl-wyosine (imG-14), isowyosine (imG2), wybutosine (yW), peroxywybutosine (o_2yW), hydroxywybutosine (OHyW), undermodified hydroxywybutosine (OHyW*), 7-deaza-guanosine, queuosine (Q), epoxyqueuosine (oQ), galactosyl-queuosine (galQ), mannosyl-queuosine (manQ), 7-cyano-7-deaza-guanosine ($preQ_0$), 7-aminomethyl-7-deaza-guanosine

(preQ₁), archaeosine (G⁺), 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine (m⁷G), 6-thio-7-methyl-guanosine, 7-methyl-inosine, 6-methoxy-guanosine, 1-methyl-guanosine (m¹G), N2-methyl-guanosine (m²G), N2,N2-dimethyl-guanosine (m^{2,2}G), N2,7-dimethyl-guanosine (m^{2,7}G), N2, N2,7-dimethyl-guanosine (m^{2,2,7}G), 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-methyl-6-thio-guanosine, N2-methyl-6-thio-guanosine, N2,N2-dimethyl-6-thio-guanosine, α-thio-guanosine, 2'-O-methyl-guanosine (Gm), N2-methyl-2'-O-methyl-guanosine (m²Gm), N2,N2-dimethyl-2'-O-methyl-guanosine (m^{2,2}Gm), 1-methyl-2'-O-methyl-guanosine (m¹Gm), N2,7-dimethyl-2'-O-methyl-guanosine (m^{2,7}Gm), 2'-O-methyl-inosine (Im), 1,2'-O-dimethyl-inosine (m¹Im), and 2'-O-ribosylguanosine (phosphate) (Gr(p)).

[220] In some embodiments, the nucleobase of the nucleotide is independently selected from a purine, a pyrimidine, a purine or pyrimidine analog. In some embodiments, the nucleobase and/or analog is each independently selected from adenine, cytosine, guanine, uracil, naturally-occurring and synthetic derivatives of a base, including but not limited to pyrazolo[3,4-d]pyrimidines, 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-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo (e.g., 8-bromo), 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, deazaguanine, 7-deazaguanine, 3-deazaguanine, deazaadenine, 7-deazaadenine, 3-deazaadenine, pyrazolo[3,4-d]pyrimidine, imidazo[1,5-a]1,3,5 triazinones, 9-deazapurines, imidazo[4,5-d]pyrazines, thiazolo[4,5-d]pyrimidines, pyrazin-2-ones, 1,2,4-triazine, pyridazine; and 1,3,5 triazine.

[221] In some embodiments, the polynucleotide comprises a nucleoside modification. In some embodiments, one or more atoms of a pyrimidine nucleobase is replaced or substituted, for example, with optionally substituted amino, optionally substituted thiol, optionally substituted alkyl (e.g., methyl or ethyl), optionally substituted or halo (e.g., chloro or fluoro) atoms or groups.

[222] In some embodiments, uracil nucleosides of the polynucleotide of the present disclosure are all modified. In some embodiments, the guanine nucleosides of the polynucleotide of the present disclosure are all modified. In some embodiments, the cytosine nucleosides of the polynucleotide of the present disclosure are all modified. In some embodiments, the thymine nucleosides of the polynucleotide of the present disclosure are all modified. In some embodiments, the adenine nucleosides of the

polynucleotide of the present disclosure are all modified. In some embodiments, the modification to each nucleobase is the same. In some embodiments, the modification to each nucleobase is different.

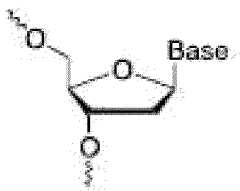
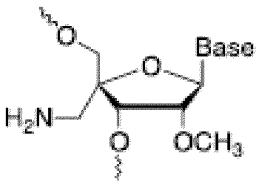
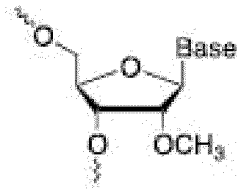
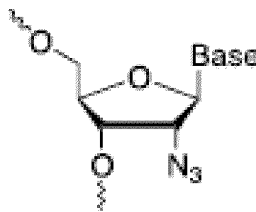
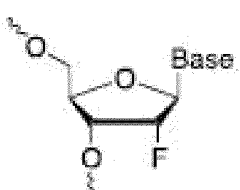
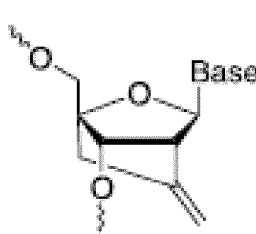
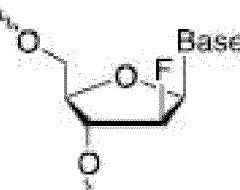
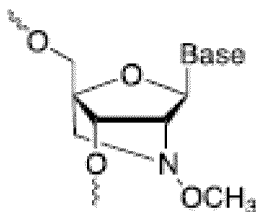
[223] In some embodiments, modifications of the modified nucleosides and nucleotides are present in the sugar subunit. In some embodiments, the polynucleotide described herein comprise at least one sugar modification. Generally, RNA includes the sugar subunit: ribose, which is a 5-membered ring having an oxygen. In some embodiments, the 2' hydroxyl group (OH) can be modified or replaced with a number of different substituents. Exemplary substitutions at the 2'-OH-position include, but are not limited to, H, halo, optionally substituted C₁₋₆ alkyl; optionally substituted C₁₋₆ alkoxy; optionally substituted C₆₋₁₀ aryloxy; optionally substituted C₃₋₈ cycloalkyl; optionally substituted C₃₋₈ cycloalkoxy; optionally substituted C₆₋₁₀ aryloxy; optionally substituted C₆₋₁₀ aryl-C₁₋₆ alkoxy, optionally substituted C₁₋₁₂ (heterocycl)oxy; a sugar (e.g., ribose, pentose, or any described herein); a polyethyleneglycol (PEG)-O(CH₂CH₂O)_nCH₂CH₂OR, where R is H or optionally substituted alkyl, and n is an integer from 0 to 20 (e.g., from 0 to 4, from 0 to 8, from 0 to 10, from 0 to 16, from 1 to 4, from 1 to 8, from 1 to 10, from 1 to 16, from 1 to 20, from 2 to 4, from 2 to 8, from 2 to 10, from 2 to 16, from 2 to 20, from 4 to 8, from 4 to 10, from 4 to 16, and from 4 to 20, including all values and subranges therebetween); and "locked" nucleic acids (LNA) in which the 2'-hydroxyl is connected by a C₁₋₆ alkylene or C₁₋₆ heteroalkylene bridge to the 4'-carbon of the same ribose sugar, where exemplary bridges include methylene, propylene, ether, or amino bridges; aminoalkyl; aminoalkoxy; amino; and amino acid.

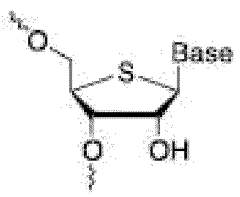
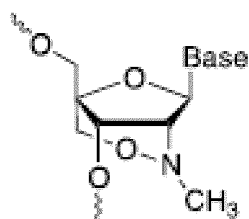
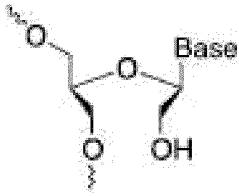
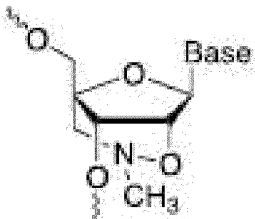
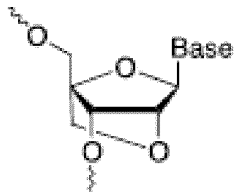
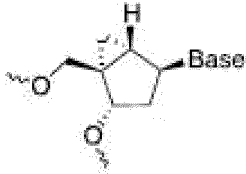
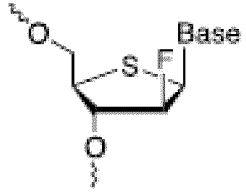
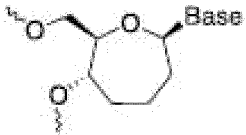
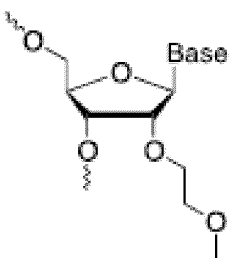
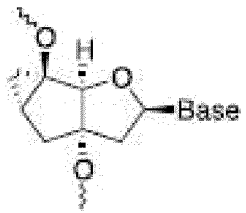
[224] Other exemplary sugar modifications include replacement of the oxygen(O) in ribose (e.g., with S, Se, or alkylene, such as methylene or ethylene); addition of a double bond (e.g., to replace ribose with cyclopentenyl or cyclohexenyl); ring contraction of ribose (e.g., to form a 4-membered ring of cyclobutane or oxetane); ring expansion of ribose (e.g., to form a 6- or 7-membered ring having an additional carbon or heteroatom, such as for anhydrohexitol, altritol, mannitol, cyclohexanyl, cyclohexenyl, and morpholino that also has a phosphoramidate backbone); multicyclic forms (e.g., tricyclo; and "unlocked" forms, such as glycol nucleic acid (GNA) (e.g., R-GNA or S-GNA, where ribose is replaced by glycol units attached to phosphodiester bonds), threose nucleic acid (TNA, where ribose is replaced with α-L-threofuranosyl-(3'→2')) , and peptide nucleic acid (PNA, where 2-amino-ethyl-glycine linkages replace the ribose and phosphodiester backbone).

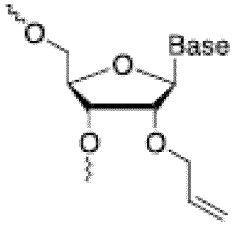
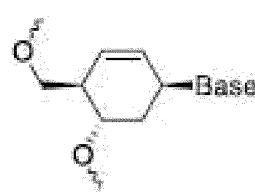
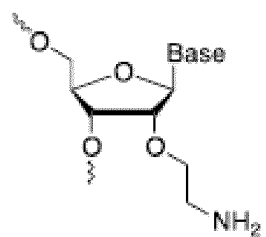
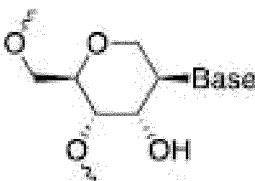
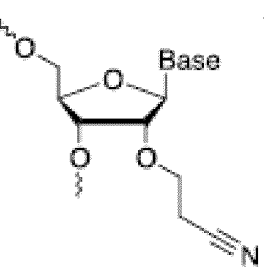
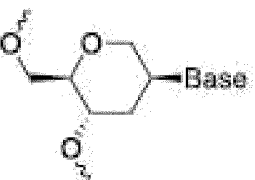
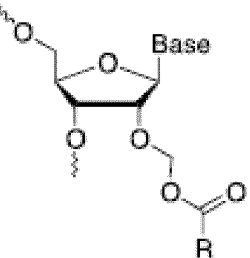
[225] In some embodiments, the sugar subunit contains one or more carbons that possess the opposite stereochemical configuration than that of the corresponding carbon in ribose. In some embodiments, polynucleotides as described herein, include nucleotides containing, e.g., arabinose, as the sugar.

[226] Nonlimiting examples of the sugar modification may include the modifications provided in Table 1. In some embodiments, the polynucleotides of the present disclosure have one or more nucleotides carrying a modification as provided in Table 1. In some embodiments, each of the nucleotides of a polynucleotide described herein carries any one of the modifications as provided in Table 1, or none of the modifications as provided in Table 1.

[227] Table 1. Nucleotide Sugar Modifications

Nucleotide	Structure Depiction	Nucleotide	Structure Depiction
DNA		4'-C-aminomethyl-2'-O-methyl RNA	
2'-O-Methyl (2'-OMe)		2'-azido	
2'-F-RNA		Methylene-cLNA	
2'-F-ANA		N-MeO-amino BNA	

4'S-RNA		N-Me-aminooxy BNA	
UNA		2',4'-BNA ^{NC} [NMe]	
LNA		MC	
4'S-FANA		ONA	
2'-O-Methoxyethyl (2'-MOE)		tc-DNA	

2'-O-Allyl		CeNA	
2'-O-Ethylamine		ANA	
2'-O-Cyanoethyl		HNA	
2'-O-Acetalester			

[228] In some embodiments, at least one of the 2' positions of the sugar (OH in RNA or H in DNA) of a nucleotide of the polynucleotides is substituted with -O- methoxyethyl, referred to as 2'-OMe. In some embodiments, at least one of the 2' positions of the sugar (OH in RNA or H in DNA) of a nucleotide of the polynucleotides is substituted with -F, referred to as 2'-F. In some embodiments, the sugar modification is one or more locked nucleic acids (LNAs). In some embodiments, the polynucleotides are fully 2'-MOE-sugar modified.

[229] In some embodiments, one or more modifications are present in the internucleoside linkage (the linking phosphate or the phosphodiester linkage or the phosphodiester backbone). In the context of the polynucleotide backbone, the phrases “phosphate” and “phosphodiester” are used interchangeably.

[230] In some embodiments, backbone phosphate groups are modified by replacing one or more of the oxygen atoms with a different substituent. In some embodiments, modified nucleosides and nucleotides include replacement of an unmodified phosphate moiety with another internucleoside linkage as described herein. Examples of modified phosphate groups include, but are not limited to, phosphorothioate, methylphosphonates phosphoroselenates, boranophosphates, boranophosphate esters, hydrogen phosphonates, phosphoramidates, phosphorodiamidates, alkyl or aryl phosphonates, and phosphotriesters. Phosphorodithioates have both non-linking oxygens replaced by sulfur. The phosphate linker is also modified by the replacement of a linking oxygen with nitrogen (bridged phosphoramidates), sulfur (bridged phosphorothioates), and carbon (bridged methylene-phosphonates).

[231] The α -thio substituted phosphate moiety is provided to confer stability to RNA and DNA polynucleotides through unnatural phosphorothioate backbone linkages. Phosphorothioate DNA and RNA have increased nuclease resistance and subsequently a longer half-life in a cellular environment. Phosphorothioate linked polynucleotide molecules are expected to also reduce the innate immune response through weaker binding/activation of cellular innate immune molecules.

[232] In some embodiments, the polynucleotides of the present disclosure comprise at least one phosphorothioate linkage, methylphosphonate linkage between nucleotides, 5'-(E)-vinylphosphonate (5'-E-VP), a phosphate mimic, as a modification.

[233] In some embodiments, the internucleoside linkages of the polynucleotides may be partially or fully modified.

[234] In some embodiments, modified nucleotides incorporated in the polynucleotides include, for example, 2'-O-Methyl-modified or 2'-O-Methoxyethyl-modified nucleotides (2'-OMe and 2'-MOE modifications, respectively), an α -thio-nucleoside (e.g., 5'-O-(1-thiophosphate)-adenosine, 5'-O-(1-thiophosphate)-cytidine (α -thio-cytidine), 5'-O-(1-thiophosphate)-guanosine, 5'-O-(1-thiophosphate)-uridine, or 5'-O-(1-thiophosphate)-pseudouridine.

[235] In some embodiments, different sugar modifications, nucleobase modifications, and/or internucleoside linkages (e.g., backbone structures) are introduced at various positions in a polynucleotide described herein. One of ordinary skill in the art will appreciate that the nucleotide analogs

or other modification(s) may be located at any position(s) of a polynucleotide such that the function of the polynucleotide is not substantially decreased.

[236] In some embodiments, the one or more modified nucleotides is a 2'-O-methyl or a phosphorothioate modified nucleotide. Accordingly, in some embodiments, the one or more modified nucleotides comprises a 2'-O-methyl modification. In some embodiments, the one or more modified nucleotides comprises a phosphorothioate modification.

[237] In some embodiments, the one or more modified nucleotides is selected from 2'-O-methyl 3'-phosphorothioate, 2'-O-methyl, 2'-ribo 3'-phosphorothioate, 2'-fluoro, 2'-O-methoxyethyl morpholino (PMO), locked nucleic acid (LNA), deoxy, or 5' phosphate modified nucleotide. Accordingly, in some embodiments, the one or more modified nucleotides is a 2'-O-methyl 3'-phosphorothioate. In some embodiments, the one or more modified nucleotides is a 2'-O-methyl nucleotide. In some embodiments, the one or more modified nucleotides is a 2'-ribo 3'-phosphorothioate. In some embodiments, the one or more modified nucleotides is a 2'-fluoro nucleotide. In some embodiments, the one or more modified nucleotides is a locked nucleic acid (LNA). In some embodiments, the one or more modifications comprises a 2'-O-methoxyethyl morpholino (PMO). In some embodiments, the one or more modifications comprises a deoxy modification. In some embodiments, the one or more modifications comprises a 5' phosphate modification.

[238] Various modified RNA bases are known in the art and include for example, 2'-O-methoxyethyl bases (2'-MOE) such as 2-MethoxyEthoxy A, 2-MethoxyEthoxy MeC, 2-MethoxyEthoxy G, 2-MethoxyEthoxy T. Other modified bases include for example, 2'-O-Methyl RNA bases, and fluoro bases. Various fluoro bases are known, and include for example, Fluoro C, Fluoro U, Fluoro A, Fluoro G bases. Various 2'OMethyl modifications can also be used with the methods described herein. For example, the following RNA comprising one or more of the following 2'-O-Methyl modifications can be used with the methods described: 2'-OMe-5-Methyl-rC, 2'-OMe-rT, 2'-OMe-rl, 2'-OMe-2-Amino-rA, Aminolinker-C6-rC, Aminolinker-C6-rU, 2'-OMe-5-Br-rU, 2'-OMe-5-I-rU, 2-OMe-7-Deaza-rG.

[239] In some embodiments, the RNA comprises one or more of the following modifications: phosphorothioates, 2'-O-methyls, 2' fluoro (2'F), DNA. In some embodiments, the RNA comprises 2'-OMe modifications at the 3'- and 5'-ends. In some embodiments, the RNA comprises one or more of the following modifications: 2'-O-2-Methoxyethyl (MOE), locked nucleic acids, bridged nucleic acids, unlocked nucleic acids, peptide nucleic acids, morpholino nucleic acids. In some embodiments, the RNA comprises one or more of the following base modifications: 2,6-diaminopurine, 2-aminopurine, pseudouracil, N1-methyl-psuedouracil, 5' methyl cytosine, N6-methyladenosine, 2'pyrimidinone (zebularine), thymine.

Other modified bases include for example, 2-Aminopurine, 5-Bromo dU, deoxyUridine, 2,6-Diaminopurine (2-Amino-dA), Dideoxy-C, deoxyInosine, Hydroxymethyl dC, Inverted dT, Iso-dG, Iso-dC, Inverted Dideoxy-T, 5-Methyl dC, 5-Methyl dC, 5-Nitroindole, Super T[®], 2'-F-r(C,U), 2'-NH₂-r(C,U), 2,2'-Anhydro-U, 3'-Desoxy-r(A,C,G,U), 3'-O-Methyl-r(A,C,G,U), rT, rI, 5-Methyl-rC, 2-Amino-rA, rSpacer (Abasic), 7-Deaza-rG, 7-Deaza-rA, 8-Oxo-rG, 5-Halogenated-rU, N-Alkylated-rN.

[240] In some embodiments, other chemically modified RNA is used herein. For example, the RNA can comprise a modified base such as, for example, 5' Int, 3' Azide (NHS Ester); 5' Hexynyl; 5' Int, 3' 5-Octadiynyl dU; 5' Int Biotin (Azide); 5' Int 6-FAM (Azide); and 5' Int 5-TAMRA (Azide). Other examples of RNA nucleotide modifications that can be used with the methods described herein include for example phosphorylation modifications, such as 5'-phosphorylation and 3'-phosphorylation. The RNA can also have one or more of the following modifications: an amino modification, biotinylation, thiol modification, alkyne modifier, adenylation, Azide (NHS Ester), Cholesterol-TEG, and Digoxigenin (NHS Ester).

RNA Manufacturing

[241] In some aspects, the present disclosure provides a method of manufacturing RNA having a quantified percentage of capped and untailed mRNA comprising the steps of: (a) providing *in vitro* synthesized RNA; (b) annealing the manufactured RNA with an oligonucleotide complementary to a sequence in a 5' UTR of the RNA, and an oligonucleotide complementary to a sequence in a 3' UTR of the RNA; (c) treating the RNA with a nuclease to cleave the RNA into cap and tail fragments; (d) identifying the capping species and untailed species by liquid chromatography with UV detection (LC-UV) and/or identifying and measuring the mass of the capping species, untailed species and characterizing the poly A tail by liquid chromatography coupled to mass spectrometry (LC-MS) and/or liquid chromatography with UV detection and coupled to mass spectrometry (LC-UV-MS); and (e) quantifying a relative amount of each capping species, quantifying a relative amount of untailed species using LC-UV or LC-MS or LC-UV-MS and characterizing poly A tail using LC-MS or LC-UV-MS; in a single sample simultaneously; thereby manufacturing RNA comprising a quantified percentage of capped and untailed mRNA.

[242] In some embodiments, RNA is manufactured. In some embodiments, an *in vitro* transcribed RNA is provided. *In vitro* transcription is carried out using a linearized DNA template, nucleotides, and enzymes to synthesize mRNA, for example, in a single use bioreactor. In various embodiments, synthesis includes design of a DNA template, codon optimization, promoter selection (for example, T7, SP6), selection of 5' UTR and 3' UTR sequences, optionally inclusion of modified nucleotides, for example, to enhance protein translation, improve stability, or reduce immunogenicity. Strain optimization and screening is carried out to select a suitable strain to expand suitable templates and

generate a GMP cell bank. In some embodiments, the mRNA product is modified co-transcriptionally or post-synthesis processing is carried out where enzymatic capping and tailing is carried out.

[243] In some embodiments, a 5' cap is added enzymatically, post transcription to the RNA from an *in vitro* synthesis reaction. In some embodiments, the *in vitro* synthesized RNA does not comprise a 5' cap.

[244] In some embodiments, a 3' tail is encoded in a plasmid during *in vitro* synthesis.

[245] In some embodiments, a 3' tail is added enzymatically, post transcription to the RNA from step (a).

[246] In some embodiments, the RNA does not comprise a 3' tail.

[247] In some embodiments, the RNA is unspliced.

[248] In some embodiments, the RNA is spliced.

[249] In some embodiments, a 5' cap is added enzymatically, post transcription to the RNA from step (a) and a 3' tail is encoded in a plasmid during *in vitro* synthesis.

[250] In some embodiments, a 5' cap is added enzymatically, post transcription to the RNA from step (a) and the RNA does not comprise a 3' tail.

[251] In some embodiments, the RNA does not comprise a 5' cap and a 3' tail is encoded in a plasmid during *in vitro* synthesis.

[252] In some embodiments, the RNA does not comprise a 5' cap and the RNA does not comprise a 3' tail.

[253] In various embodiments, nucleotides, reagents, enzymes and process conditions, for example, temperature and incubation times, are selected. The mRNA is then isolated and purified using magnetic bead-based, affinity, or other chromatography and filtration techniques such as ultrafiltration/diafiltration UF/DF.

[254] The mRNA is tested for various quality attributes, for example, the sequence is confirmed as well as the identity of the UTR is verified by PCR. In some embodiments, the RNA content is quantified, for example, by UV measurement. Potency is tested in an activity assay, for example an *in vitro* transcription assay. Purity and integrity are tested by the methods of the present disclosure to evaluate the percent of intact and fragmented mRNA, the 5' capping efficiency and 3' poly A (% or length) and for the presence of undesirable immunostimulatory RNA side products. The manufactured RNA drug substance is further tested for the presence of any residual impurities (e.g., dsRNA, protein or template) and safety.

Quality control for manufactured RNA

[255] In some aspects, the present disclosure provides a quality control assay for a manufacturing lot, wherein the assay comprises, (a) providing a manufactured RNA sample comprising a 5' cap and/or a 3' tail; (b) quantifying RNA capping and tailing efficiency by the methods provided herein; (c) comparing the capping and tailing efficiency in step (b) to capping and tailing efficiency in a reference sample; wherein a reduced amount of capping species relative to a reference sample indicates 5' degradation, and wherein an increased amount of untailing species and/or a different poly A tail length or polydispersity indicates 3' degradation, thereby simultaneously monitoring 5' degradation and 3' degradation of an RNA product, and determining product quality of the manufactured lot.

[256] In some embodiments, the reference sample is a stable RNA sample having greater than 75% integrity. In some embodiments, the reference sample is a stable RNA sample having greater than 75-80%, 80-85%, 85-90%, 90-95% or 95-100%, including all values and subranges therebetween, integrity. In some embodiments, the reference sample is a sample from a manufacturing lot that has previously been tested and found suitable for release. In some embodiments, the reference sample(s) are from one or more previous batches to ensure batch-to-batch consistency.

[257] A purified mRNA therapeutic or vaccine (i.e., mRNA drug product) is formulated by compounding mRNA with a delivery vehicle, such as an LNP or other lipids or carbohydrates. Once formulated, the mRNA containing drug product is processed into the final vaccine or therapeutic, sterilized, aseptically filled, and packaged. Each lot of filled packages are assessed for quality control and batch-to-batch consistency before release of the manufactured lot. A lot that passes the quality control assay is stored in ultra-low temperature freezers for storage, distribution and delivery.

[258] In some embodiments, the RNA product is a therapeutic mRNA.

[259] In some embodiments, the RNA product is an mRNA vaccine.

Kits

[260] The present disclosure further provides kits comprising various reagents and materials useful for carrying out inventive methods according to the present disclosure. The procedures described herein may be performed by diagnostic laboratories, experimental laboratories, or commercial laboratories. The present disclosure provides kits which can be used in these different settings.

[261] For example, materials and reagents for quantifying mRNA capping and tailing efficiency in an mRNA sample by enzymatic manipulation and chromatographic separation coupled with UV and/or mass spectrometry (LC-UV, or LC-MS, or LC-UV-MS) may be assembled together in a kit. In certain embodiments, a kit comprises chromatographic columns, and agents for separating capped, uncapped

and untailed mRNA species on the column, and instructions for using the kit according to a method of the present disclosure.

[262] In some embodiments, each kit comprises a customized reagent which renders the procedure specific to a target. Thus, for detecting/quantifying mRNA capping and tailing efficiency of a particular target, kits comprise hybrid oligonucleotide reagents of designed sequences that specifically anneals adjacent to a 5' end and to a 3' end of a target. In some embodiments, kits comprise nucleases for production of the capped fragments; e.g., RNase H and/or S1 nuclease. In some embodiments, kits further include *in vitro* transcription and capping reagents, enzymes and instructions for using the same.

[263] Kits or other articles of manufacture according to the present disclosure include one or more containers to hold various reagents. Suitable containers include, for example, bottles, vials, syringes (e.g., pre-filled syringes), ampules. The container is formed from a variety of materials such as glass or plastic.

[264] In some embodiments, kits of the present disclosure include suitable standard or control levels or standard or control samples for determining control levels as described herein. In some embodiments, kits of the present disclosure include instructions for using the kit according to one or more methods of the present disclosure. In some embodiments, kits additionally comprise instructions for *in vitro* transcription and capping.

First Set of Representative Embodiments of the Present Disclosure

1. A single sample method of identifying ribonucleic acid (RNA) capping and tailing modifications, the method comprising:

(a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail, and wherein said sample further comprises a first oligonucleotide probe complementary to a sequence in the 5' untranslated region (UTR) region of the RNA and a second oligonucleotide probe complementary to a sequence in the 3' untranslated region (UTR) region of the RNA;

(b) annealing the RNA sample with a first oligonucleotide probe complementary to a sequence in the 5' UTR of the RNA, and a second oligonucleotide probe complementary to a sequence in the 3' UTR of the RNA;

(c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments;

(d) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS using the sample of step (c) for:

- i. identifying capping species in the cap fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS;
- ii. identifying untailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; and
- iii. identifying tailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS;

thereby simultaneously identifying RNA capping and tailing modifications in the single sample.

2. A single sample method of simultaneously quantifying RNA capping and tailing efficiency, the method comprising:

- (a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail;
- (b) annealing the RNA sample with a first oligonucleotide probe complementary to a sequence in a 5' untranslated region (UTR) of the RNA, and a second oligonucleotide probe complementary to a sequence in a 3' untranslated region (UTR) of the RNA;
- (c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments;
- (d) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS;
- (e) measuring a peak area of each capping species in the cap fragment, and each untailed and tailed species in the tail fragment; and
- (f) quantifying a relative amount of each capping species, quantifying a relative amount of untailed species and characterizing poly A tail in the sample of step (d); thereby simultaneously quantifying RNA capping efficiency and tailing efficiency in the sample.

3. The method of numbered embodiment 2, wherein characterizing the poly A tail comprises (i) mass measurements of poly A tail species, deconvoluting a mass spectrometry signal, thereby providing a mass distribution for tailed species, and (ii) determining based on the corresponding mass distribution, a minimum length, a maximum length, an average length and a degree of polydispersity of the poly A tail.
4. The method of numbered embodiments 1 or 2, wherein the RNA is *in vitro* transcribed mRNA.
5. The method of numbered embodiments 1 or 2, wherein the RNA is obtained from a step of manufacturing.
6. The method of numbered embodiments 1 or 2, wherein the RNA is obtained from a drug substance (DS) at the final step of manufacturing.
7. The method of numbered embodiments 1 or 2, wherein the RNA is obtained from a deformulated drug product (DP).
8. The method of any one of the preceding numbered embodiments, wherein the RNA is unmodified.
9. The method of any one of the preceding numbered embodiments, wherein the RNA is modified.
10. The method of numbered embodiments 1 or 2, wherein the nuclease is RNase H.
11. The method of numbered embodiments 1 or 2, wherein the mass of the cap and the tail fragments is measured by LC-MS.

12. The method of numbered embodiments 1 or 2, wherein the capping species, untailed species and tailed species is identified from a UV signal by comparing the retention time of peaks obtained to the retention time of a reference standard.
13. The method of any one of the preceding numbered embodiments, wherein the capping species is Cap1, Cap0, CapG or uncapped.
14. The method of any one of the preceding numbered embodiments, wherein the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA, GGACA or pppGGACA.
15. The method of any one of numbered embodiments 13 or 14, wherein the uncapped species is GGACA or pppGGACA.
16. The method of any one of the preceding numbered embodiments, wherein the untailed species is UGCAUC, wherein U is unmodified uridine.
17. The method of any one of the preceding numbered embodiments, wherein the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine.
18. The method of any one of the preceding numbered embodiments, wherein the relative amount of each of the capping species is a percentage of total amount of capping species in the sample calculated by dividing an area under the peak of the capping species of interest over a sum of total areas under the peak representing Cap 1, Cap 0, CapG and uncapped species, multiplied by 100.

19. The method of any one of the preceding numbered embodiments, wherein the relative amount of untailed species is a percentage calculated by dividing an area under the peak of the untailed species over a sum of areas under the peak of Cap1, Cap0, CapG and uncapped multiplied by 100.
20. The method of any one of the preceding numbered embodiments, wherein the oligonucleotide is between about 10 to 40 nucleotides in length.
21. The method of any one of the preceding numbered embodiments, wherein the oligonucleotide comprises RNA and DNA bases.
22. The method of any one of the preceding numbered embodiments, wherein the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1.
23. The method of any one of the preceding numbered embodiments, wherein the oligonucleotide is 10-40 nucleotides long and comprises 4 DNA bases.
24. The method of numbered embodiment 23, wherein the oligonucleotide complementary to a sequence in the 5' untranslated region of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein italicized CTGT are DNA bases.
25. The method of numbered embodiment 23, wherein the oligonucleotide complementary to a sequence in the 3' untranslated region of the RNA is 3'-GGUCGGAACAGGAUUUUUAATTCAA-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.
26. The method of any one of the preceding numbered embodiments, wherein capping and untailed species are quantified in a single LC-UV or LC-MS or LC-UV-MS analysis.

27. The method of numbered embodiment 1, wherein tailed species are characterized in a single LC-MS or LC-UV-MS analysis.
28. The method of any one of the preceding numbered embodiments, wherein one or more steps is automated.
29. An oligonucleotide that is complementary to a sequence in the 5' UTR of the RNA or 3' UTR of the RNA, wherein the oligonucleotide is between about 10 to 40 nucleotides in length and comprises RNA and DNA bases.
30. The oligonucleotide probe of numbered embodiment 29, wherein the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1.
31. The oligonucleotide of any one of the preceding numbered embodiments, wherein the oligonucleotide is about 10 to 40 nucleotides long and comprises 4 DNA bases.
32. The oligonucleotide of any one of the preceding numbered embodiments, wherein the oligonucleotide complementary to a sequence in the 5' UTR of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein the italicized CTGT are DNA bases.
33. The oligonucleotide of any one of the preceding numbered embodiments, wherein the oligonucleotide complementary to a sequence in the 3' UTR of the RNA is 3'-GGUCGGAACAGGAUUUUUAATTCAA-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.

34. A method of manufacturing RNA having a quantified percentage of capped and untailed mRNA comprising the steps of:

- (a) providing *in vitro* synthesized RNA;
 - (b) annealing the manufactured RNA with an oligonucleotide complementary to a sequence in a 5' UTR of the RNA, and an oligonucleotide complementary to a sequence in a 3' UTR of the RNA;
 - (c) treating the RNA with a nuclease to cleave the RNA into cap and tail fragments;
 - (d) identifying the capping species and untailed species by liquid chromatography with UV detection (LC-UV) and/or identifying and measuring the mass of the capping species, untailed species and characterizing the polyA tail by liquid chromatography coupled to mass spectrometry (LC-MS) and/or liquid chromatography with UV detection and coupled to mass spectrometry (LC-UV-MS); and
 - (e) quantifying a relative amount of each capping species, quantifying a relative amount of untailed species using LC-UV or LC-MS or LC-UV-MS and characterizing poly A tail using LC-MS or LC-UV-MS; in a single sample simultaneously;
- thereby manufacturing RNA comprising a quantified percentage of capped and untailed mRNA.

35. The method of numbered embodiment 34, wherein a 5' cap is added enzymatically, post transcription to the RNA from step (a).

36. The method of numbered embodiment 34, wherein the RNA does not comprise a 5' cap.

37. The method of numbered embodiment 34, wherein a 3' tail is encoded in a plasmid during *in vitro* synthesis.

38. The method of numbered embodiment 34, wherein a 3' tail is added enzymatically, post transcription to the RNA from step (a).

39. The method of numbered embodiment 34, wherein the RNA does not comprise a 3' tail.

40. The method of numbered embodiment 34, wherein the RNA is unspliced.
41. The method of numbered embodiment 34, wherein the RNA is spliced.
42. A quality control assay for a manufacturing lot, wherein the assay comprises:
- (a) providing a manufactured RNA sample comprising a 5' cap and/or a 3' tail;
 - (b) quantifying RNA capping and tailing efficiency by the method of claim 2 in the sample;
 - (c) comparing the capping and tailing efficiency in step (b) to capping and tailing efficiency in a reference sample;
- wherein a reduced amount of capping species relative to a reference sample indicates 5' degradation, and
- wherein an increased amount of untailed species and/or a different poly A tail length or polydispersity indicates 3' degradation,
- thereby simultaneously monitoring 5' degradation and 3' degradation of an RNA product, and determining product quality of the manufactured lot.
43. The quality control assay of numbered embodiment 42, wherein the reference sample is a stable RNA sample having greater than 75% integrity.
44. The quality control assay of numbered embodiment 42, wherein the RNA product is a therapeutic mRNA.
45. The quality control assay of numbered embodiment 43, wherein the RNA product is an mRNA vaccine.

Second Set of Representative Embodiments of the Present Disclosure

1. A single sample method of identifying ribonucleic acid (RNA) capping and tailing modifications, the method comprising:

(a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail, and wherein said sample further comprises a first oligonucleotide probe complementary to a sequence in the 5' untranslated region (UTR) region of the RNA and a second oligonucleotide probe complementary to a sequence in the 3' untranslated region (UTR) region of the RNA;

(b) annealing the RNA sample with a first oligonucleotide probe complementary to a sequence in the 5' UTR of the RNA, and a second oligonucleotide probe complementary to a sequence in the 3' UTR of the RNA;

(c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments;

(d) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS using the sample of step (c) for:

i. identifying capping species in the cap fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS;

ii. identifying untailed species and tailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; and

iii. identifying untailed species and tailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS,

thereby simultaneously identifying RNA capping and tailing modifications in the single sample.

2. A single sample method of simultaneously quantifying RNA capping and tailing efficiency, the method comprising:

(a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail;

(b) annealing the RNA sample with a first oligonucleotide probe complementary to a sequence in a 5' untranslated region (UTR) of the RNA, and a second oligonucleotide probe complementary to a sequence in a 3' untranslated region (UTR) of the RNA;

- (c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments;
- (d) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS;
- (e) measuring a peak area of each capping species in the cap fragment, and each untailing and tailed species in the tail fragment; and
- (f) quantifying a relative amount of each capping species, quantifying a relative amount of untailing species and characterizing poly A tail in the sample of step (d); thereby simultaneously quantifying RNA capping efficiency and tailing efficiency in the sample.

3. The method of numbered embodiment 2, wherein characterizing the poly A tail comprises (i) mass measurements of polyA tail species, deconvoluting a mass spectrometry signal, thereby providing a mass distribution for tailed species, and (ii) determining based on the corresponding mass distribution, a minimum length, a maximum length, an average length and a degree of polydispersity of the poly A tail.

4. The method of any one of the preceding numbered embodiments, wherein one or more of the following conditions are met:

- (a) the RNA is *in vitro* transcribed mRNA
- (b) the RNA is obtained from a step of manufacturing;
- (c) the RNA is obtained from a drug substance (DS) at the final step of manufacturing;
- (d) the RNA is obtained from a deformed drug product (DP).

5. The method of any one of the preceding numbered embodiments, wherein the RNA is unmodified or modified.

6. The method of any one of the preceding numbered embodiments, wherein one or more of the following conditions are met:

- (a) the nuclease is RNase H;

- (b) the mass of the cap and the tail fragments is measured by LC-MS;
- (c) the capping species, untailed species and tailed species are identified from a UV signal by comparing the retention time of peaks obtained to the retention time of a reference standard;
- (d) the capping species is Cap1, Cap0, CapG or uncapped;
- (e) the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA or GGACA;
- (f) the capping species is Cap1, Cap0, CapG or uncapped, and the uncapped species is GGACA;
- (g) the capping species is Cap1, Cap0, CapG or uncapped, and the uncapped species is pppGGACA;
- (h) the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA or GGACA, and the uncapped species is GGACA;
- (i) the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA or GGACA, and the uncapped species is pppGGACA;
- (j) the untailed species is UGCAUC, wherein U is unmodified uridine;
- (k) the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine
- (l) the capping species is Cap1, Cap0, CapG or uncapped, and the untailed species is UGCAUC, wherein U is unmodified uridine;
- (m) the capping species is Cap1, Cap0, CapG or uncapped, and the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine;
- (n) the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA or GGACA, and the untailed species is UGCAUC, wherein U is unmodified uridine;
- (o) the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA or GGACA or pppGACA, and the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine;
- (p) the capping species is Cap1, Cap0, CapG or uncapped, the uncapped species is GGACA or pppGGACA, and the untailed species is UGCAUC, wherein U is unmodified uridine;
- (q) the capping species is Cap1, Cap0, CapG or uncapped, the uncapped species is GGACA or pppGGACA, and the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine;
- (r) the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA or GGACA, the uncapped species is GGACA or pppGGACA, and the untailed species is UGCAUC, wherein U is unmodified uridine;

(s) the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA or GGACA, the uncapped species is GGACA or pppGGACA, and the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine.

7. The method of any one of the preceding numbered embodiments, wherein the relative amount of each of the capping species is a percentage of total amount of capping species in the sample calculated by dividing an area under the peak of the capping species of interest over a sum of total areas under the peak representing Cap 1, Cap 0, CapG and uncapped species, multiplied by 100; and/or the relative amount of untailed species is a percentage calculated by dividing an area under the peak of the untailed species over a sum of areas under the peak of Cap1, Cap0, CapG and uncapped multiplied by 100.

8. The method of any one of the preceding numbered embodiments, wherein one or more of the following conditions are met:

- (a) the oligonucleotide probe is between about 10 to 40 nucleotides in length;
- (b) the oligonucleotide probe comprises RNA and DNA bases;
- (c) the oligonucleotide probe comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1;
- (d) the oligonucleotide probe is 10-40 nucleotides long and comprises 4 DNA bases;
- (e) the oligonucleotide probe comprises RNA and DNA bases, and the oligonucleotide probe complementary to the sequence in the 5' untranslated region of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein italicized CTGT are DNA bases; and
- (f) the oligonucleotide probe comprises RNA and DNA bases, and the oligonucleotide complementary to the sequence in the 3' untranslated region of the RNA is 3'-GGUCGGAACAGGAUUUUUAATTCAA-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.

9. The method of any one of the preceding numbered embodiments, wherein one or more of the following conditions are met:

- (a) capping and untailed species are quantified in a single LC-UV or LC-MS or LC-UV-MS analysis;

- (b) tailed species are characterized in a single LC-MS or LC-UV-MS analysis; and
 - (c) one or more steps is automated.
10. An oligonucleotide that is complementary to a sequence in the 5' UTR of an RNA or 3' UTR of an RNA, wherein the oligonucleotide probe is between about 10 to 40 nucleotides in length and comprises RNA and DNA bases.
11. The oligonucleotide of numbered embodiment 11, wherein one or more of the following conditions are met:
- (a) the oligonucleotide probe comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1;
 - (b) the oligonucleotide probe is about 10 to 40 nucleotides long and comprises 4 DNA bases;
 - (c) the oligonucleotide probe comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1, and the oligonucleotide probe is about 10 to 40 nucleotides long and comprises 4 DNA bases
 - (d) the oligonucleotide probe complementary to the sequence in the 5' UTR of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein the italicized CTGT are DNA bases; and
 - (e) the oligonucleotide probe complementary to the sequence in the 3' UTR of the RNA is 3'-GGUCGGAACAGGAUUUUUAATTCAA-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.
12. A method of manufacturing RNA having a quantified percentage of capped and untailing mRNA comprising the steps of:
- (a) providing *in vitro* synthesized RNA;
 - (b) annealing the *in vitro* synthesized RNA with a first oligonucleotide complementary to a sequence in the 5' UTR of the RNA, and a second oligonucleotide complementary to a sequence in the 3' UTR of the RNA;
 - (c) treating the RNA with a nuclease to cleave the RNA into cap and tail fragments;
 - (d) identifying the capping species and untailing species by liquid chromatography with UV detection (LC-UV) and/or identifying and measuring the mass of the capping species, untailing species and

characterizing the polyA tail by liquid chromatography coupled to mass spectrometry (LC-MS) and/or liquid chromatography with UV detection and coupled to mass spectrometry (LC-UV-MS); and

(e) quantifying a relative amount of each capping species, quantifying a relative amount of untail species using LC-UV or LC-MS or LC-UV-MS and characterizing poly A tail using LC-MS or LC-UV-MS; in a single sample simultaneously;

thereby manufacturing RNA comprising a quantified percentage of capped and untail mRNA.

13. The method of numbered embodiment 12, wherein one or more of the following conditions are met:

- (a) a 5' cap is added enzymatically, post transcription to the RNA from step (a);
- (b) the RNA does not comprise a 5' cap;
- (c) a 3' tail is encoded in a plasmid during *in vitro* synthesis;
- (d) the RNA does not comprise a 3' tail;
- (e) the RNA is unspliced;
- (f) the RNA is spliced;
- (g) a 5' cap is added enzymatically, post transcription to the RNA from step (a) and a 3' tail is encoded in a plasmid during *in vitro* synthesis;
- (h) a 5' cap is added enzymatically, post transcription to the RNA from step (a) and the RNA does not comprise a 3' tail;
- (i) the RNA does not comprise a 5' cap and a 3' tail is encoded in a plasmid during *in vitro* synthesis;
- (j) the RNA does not comprise a 5' cap and the RNA does not comprise a 3' tail.

14. A quality control assay for a manufacturing lot, wherein the assay comprises:

- (a) providing a manufactured RNA sample comprising a 5' cap and/or a 3' tail;
- (b) quantifying RNA capping and tailing efficiency by the method of claim 2 in the sample;
- (c) comparing the capping and tailing efficiency in step (b) to capping and tailing efficiency in a reference sample;

wherein a reduced amount of capping species relative to a reference sample indicates 5' degradation, and

wherein an increased amount of untailed species and/or a different poly A tail length or polydispersity indicates 3' degradation,

thereby simultaneously monitoring 5' degradation and 3' degradation of an RNA product, and determining product quality of the manufactured lot.

15. The quality control assay of numbered embodiment 14, wherein the reference sample is a stable RNA sample having greater than 75% integrity; and/or the RNA product is a therapeutic mRNA, in particular an mRNA vaccine.

EXAMPLES

Example 1. Generation of RNA by *in vitro* synthesis, capping and tailing

[265] Identification or quantification of mRNA capping and tailing as described by the present disclosure is carried out on RNA samples obtained from a variety of sources. For example, in various non-limiting embodiments, the RNA is *in vitro* transcribed mRNA, the RNA is obtained from a step of manufacturing, e.g., the RNA is obtained from the final step of manufacturing or the RNA is obtained from a reformulated drug product. In some embodiments, the RNA is unmodified. In some embodiments, the RNA is modified.

[266] This example illustrates synthesis of RNA by *in vitro* transcription followed by capping and tailing.

[267] Briefly, RNA was synthesized by *in vitro* transcription from a plasmid DNA template comprising the DNA sequence encoding the mRNA. *In vitro* transcription included addition of an exemplary 5' cap structure, e.g., Cap1, which has a 2'-O-methyl residue at the 2' OH group of the ribose ring of base 1, by enzymatic conjugation of GTP via guanylyl transferase. Exemplary mRNA capped structures and an uncapped structure present in various embodiments of the present disclosure and means to derive these enzymatically are shown in FIG. 4B. Uncapped mRNA (for example, *in vitro* transcribed mRNA, pppG-mRNA), Cap G (GpppG-mRNA), Cap 0 (m7GpppG-mRNA) and Cap1 (m7GpppmG-mRNA).

[268] A 3' poly A tail was incorporated through the addition of ATP in conjunction with PolyA polymerase (see detailed reaction conditions below). The *in vitro* transcription product included 5' and 3' untranslated regions.

[269] The synthesis of RNA was conducted under RNase-free conditions. All tubes, vials, pipette tips, pipettes, buffers, etc. were nuclease-free. RNA was synthesized from a linearized DNA template. To produce the desired RNA precursor (IVT) construct, a mixture of about 100 µg of linearized DNA, rNTPs (3.33 mM), DTT (10 mM), T7 RNA polymerase, RNase Inhibitor, Pyrophosphatase and reaction buffer (10x, 800mM Hepes (pH8.0), 20mM Spermidine, 250mM MgCl₂, pH 7.7) was prepared with RNase-free water to a final volume of 2.24 ml. The reaction mixture was incubated at 37°C for between 20-120 minutes. Upon completion, the mixture was treated with DNase I for an additional 15 minutes and quenched accordingly.

[270] The purified RNA product from the aforementioned IVT step was denatured at 65°C for 10 minutes. Separately, portions of GTP (20 mM), S-adenosyl methionine, RNase inhibitor, 2'-O-Methyltransferase and guanylyl transferase were mixed together with reaction buffer (10x, 500mM Tris-HCl (pH 8.0), 60mM KCl, 12.5mM MgCl₂) to a final concentration of 8.3 ml. Upon denaturation, the RNA was cooled on ice and then added to the reaction mixture. The combined solution was incubated at 37°C for 20-90 minutes. Upon completion, aliquots of ATP (20 mM), PolyA Polymerase and tailing reaction buffer (10x, 500mM Tris-HCl (pH 8.0), 2.5M NaCl, 100mM MgCl₂) were added, and the total reaction mixture was further incubated at 37°C for about 20-45 minutes. Upon completion, the final reaction mixture was quenched and purified accordingly.

[271] Overall, this example illustrates synthesis of RNA comprising a 5' cap and 3' tail by *in vitro* synthesis as an exemplary method of generating RNA for identification and/or quantification of capping and tailing efficiency by the present disclosure.

Example 2: Identification and/or Quantification of Capping and Tailing Efficacy of RNA in a Single Sample Method

[272] This example demonstrates a single sample method of simultaneous identification and/or quantification of RNA capping and tailing efficiency.

[273] *Design of custom hybrid oligonucleotides:* First, custom hybrid oligonucleotides were designed that bind the RNA near the 5' and 3' ends. The hybrid oligonucleotides comprised both RNA and DNA bases. In some embodiments, the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1. In some embodiments, the oligonucleotide comprised 4 DNA bases. For example, HO2.16 was a 16 nt custom designed hybrid oligonucleotide 3' -CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), complementary to a sequence in the 3' UTR of the RNA, wherein CTGT in italics near the 3' end represents DNA bases; HO2.27 was an exemplary 27 nt custom designed hybrid oligonucleotide 3'-

GGUCGGAACAGGAUUUUUAATTCAA-5' (SEQ ID NO: 2), complementary to a sequence in the 5' UTR of the RNA, wherein TTCA in italics near the 5' end represents four DNA bases. Specifically, an oligonucleotide that was between 10 to 30 nucleotides in length, for example, between 15 to 30 nucleotides in length was designed and synthesized to be complementary to the 5' UTR of capped RNA, for example, binding within a few (e.g., one or two or up to 15) nucleotides from the 5' cap. Binding of the oligonucleotide to the 5' UTR adjacent to the cap established a well-defined region of DNA:RNA hybrid that is susceptible to RNase H-mediated cleavage. In some embodiments, the oligonucleotide was flanked on either end by 1-15 RNA nucleotides. In some embodiments, the oligonucleotide was flanked on the 3' end by 1-15 RNA nucleotides.

[274] *Annealing of RNA to custom hybrid oligonucleotide forming RNA:DNA hybrid:* In the annealing step, an mRNA sample having a 5' cap and/or a 3' poly A tail, for example, synthesized by *in vitro* transcription as described in Example 1, was annealed with a first oligonucleotide probe complementary to a sequence in a 5' untranslated region (UTR) of the RNA, and a second oligonucleotide probe complementary to a sequence in a 3' untranslated region (UTR) of the RNA (FIG. 2). Briefly, RNA from an exemplary sample and exemplary custom hybrid oligonucleotides were heated to 75°C for 10 minutes which caused denaturation and strand separation, following which the temperature was decreased from 75°C to 23°C and maintained at 23°C for 10 minutes, facilitating hydrogen bond formation between complementary DNA and RNA sequences. Complementary base pairing between hybrid oligonucleotides and mRNA in a specific region resulted in RNA:DNA hybrid formation at the end of the annealing step. (FIG. 2).

[275] *Digestion of RNA:DNA hybrid:* In the digestion step, the RNA:DNA hybrid was incubated with RNase H and shrimp alkaline phosphatase (rSAP) at 37°C for 40 minutes. RNase H enzymes cleaved the phosphodiester bond between double-stranded RNA: DNA hybrids generated in the annealing step. Shrimp alkaline phosphatase (rSAP) catalyzed dephosphorylation of ends to prevent self-ligation. (FIG. 3).

[276] As described above, the hybrid oligonucleotide was designed so that four bases of DNA were flanked on both sides by one or more RNA nucleotides, providing specificity for the site of cleavage by a nuclease. The RNA: DNA hybrid was thereby selectively cleaved by a nuclease like RNase H to produce 5' capped or uncapped fragments that were a few bases long (e.g., 2-10 nucleotides long, including capped nucleotides) without producing appreciable internal fragments of mRNA. The larger mRNA fragments were separated from the shorter 5' capped or uncapped fragments during chromatography because of the drastically different physiochemical properties. Because the larger mRNA fragments distal to the 5' cap were readily removed, the resolution capabilities of chromatography could be focused on the shorter

5' capped and uncapped fragments produced by enzymatic digestion. In certain embodiments, this cleavage was accomplished using an S1 Nuclease or other nuclease to create a blunt-end fragment marked at the 5' end by the cap analytes of choice. Overall, this process provided a smaller molecule with increased resolution of detecting both cap presence and cap modification.

[277] *Identification of Capped, Uncapped and Untailed Species:* In various embodiments, liquid chromatography with UV detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS was employed for identifying capping species in the cap fragment by analyzing peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS. In embodiments using LC-UV, a standard was required for comparison. In other embodiments using LC-MS, the measurement of mass under each peak was used for identification of capping species. Simultaneously, the untailed species and tailed species in the tail fragment in the same sample were identified by analyzing peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS.

[278] *Liquid chromatography:* In some embodiments, the digested RNA:DNA hybrid was analyzed by performing, for example, liquid chromatography. Oligonucleotides generated by digestion were separated, for example, by reverse phase separation which involved separation based on degree of hydrophobicity. For example, an Acquity PREMIER Oligonucleotide BEH C18 130 Å, 1,7 µm, 2.1*100 mm was used to separate oligonucleotides based on ion-pair, reversed-phase chromatography.

[279] *UV Detection:* In some embodiments, UV detection was carried out. UV detection was particularly valuable in good manufacturing practice (GMP) environment. In some embodiments, the UV detection was at 260 nm. In some embodiments, the UV detection was at 280 nm. In some embodiments, 0.1 mm to 100 mm UV cell was used. In some embodiments, 10 mm UV cell was used. In some embodiments, 0.01 mm UV cell was used. Besides the sample, in some embodiments, a standard was injected in parallel for Cap1, Cap0, CapG and uncapped species to identify and compare retention time for each species.

[280] *LC-UV or LC-MS or LC-UV-MS:* In some embodiments, UV detection (LC-UV) and/or LC was coupled to mass spectrometry (LC-MS) or LC-UV-MS for identifying capping species in the cap fragment and untailed species and tailed species in the tail fragment by analyzing peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS.

[281] *Quantification of Capped, Uncapped and Untailed Species:* In some embodiments, the relative amounts of each capping species in the cap fragment, as well as each untailed and tailed species in the tail fragment was quantified by mass spectrometry or UV spectrometry or retention time of chromatographic peaks.

[282] *Mass Spectrometric Analysis:* In some embodiments, analysis of digested RNA: DNA hybrids was carried out quantitatively by mass spectrometry, wherein oligonucleotides were identified and subsequently quantified using the mass to charge ratio (m/z) measurement. Mass spectrometers convert oligonucleotide molecules into a charged ionized state. These ionized molecules and fragments were then detected based on their mass to charge ratios and retention time. Various ionization and ion analysis systems are known in the art including, for example, Electrospray Source Ionization (ESI) Quadrupole Time of Flight (Q-TOF), negative mode, which involves the formation of negative ions for the determination of the mass-to-charge ratio of the sample molecules.

[283] For example, an exemplary Extracted Ion Chromatogram (EIC) showed peaks generated for capped (Cap 0, Cap 1, Cap G), uncapped and untailed species (FIG. 4A). Percentage of any specific capped species (Cap 0, Cap 1, Cap G or uncapped) was calculated by measuring the area under the curve of said capped species, relative to the sum of total area under the curve for capped (Cap 0, Cap 1, Cap G) and uncapped peaks, multiplied by 100. Likewise, percentage of untailed species was calculated by measuring the area under the curve of untailed species, relative to total area under the curve for capped (Cap 0, Cap 1, Cap G) and uncapped peaks, multiplied by 100. In the exemplary EIC shown in FIG. 4A, the results of the quantitation demonstrated the presence of 6.8% untailed, 4.2% uncapped, 0.4% Cap G, 0.0% Cap 0 and 95.4% Cap 1 species. In some embodiments, EIC is generated, for example, by Liquid chromatography Mass Spectrometry (LC-MS), Liquid chromatography-ultraviolet spectroscopy/mass spectrometry (LC-UV/MS), Ultrahigh performance liquid chromatography–electrospray ionization multiple reaction monitoring tandem mass spectrometry (UHPLC-ESI-MS) or other mass spectrometric methods. Mass spectrometric signal intensity was plotted relative to m/z (mass to charge ratio).

[284] To determine poly A tailing, mass measurements of polyA tail species were made, followed by deconvoluting the mass spectrometry signal, thereby providing a mass distribution for tailed species. Average deconvoluted mass spectra for exemplary mRNA to study poly A tailing, in an exemplary drug substance is shown in FIG. 5A. In the graph, mass spectrometric signal intensity is plotted relative to m/z (mass to charge ratio).

[285] Based on the corresponding mass distribution, a minimum length, a maximum length, an average length and a degree of polydispersity of the poly A tail was determined. For example, a plot of percent mass spectrometric signal intensity relative to poly A tail length, reveals poly A polydispersity for exemplary mRNA (FIG. 5B).

[286] In some embodiments, the method was automated.

[287] It is understood that the methods described in this example are applicable to any RNA, including mRNA, therapeutic mRNA from any source, for example, mRNA vaccines targeting any disease, including viral disease. Some exemplary mRNA vaccines include vaccines directed to coronaviruses (e.g., SARS, SARS-CoV-2, MERS), influenza, respiratory syncytial virus, among others, as well as various types of cancer, including pancreatic cancer, colorectal cancer, and melanoma.

[288] Overall, the method of the present disclosure provides a powerful combination of simultaneous qualitative and/or quantitative assessment of capped, uncapped, or untailed species for determining stability of quality of mRNA, applicable for a wide variety of uses, and especially therapeutic uses, for example, mRNA vaccines.

Equivalents and Scope

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

[290] In the claims articles such as “a”, “an” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Thus, for example, reference to “an antibody” includes a plurality of such antibodies, and reference to “the cell” includes reference to one or more cells known to those skilled in the art, and so forth. Claims or descriptions that include “or” between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The present disclosure includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The present disclosure includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process. Furthermore, it is to be understood that the present disclosure encompasses all variations, combinations, and permutations in which one or more limitation, elements, clauses, descriptive terms, etc., from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of using the composition for anyone of the purposes disclosed herein are

included, and methods of making the composition according to any of the methods of making disclosed herein or other methods known in the art are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[291] Where elements are presented as lists, e.g., in Markush group format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should be understood that, in general, where the present disclosure, or aspects thereof, is/are referred to as comprising particular elements, features, etc., certain embodiments of the present disclosure or aspects thereof consist, or consist essentially of, such elements, features, etc. For purposes of simplicity those embodiments have not been specifically set forth in haec verba herein. It is noted that the term “comprising” is intended to be open and permits the inclusion of additional elements or steps.

[292] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub-range within the state ranges in different embodiments of the present disclosure, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

[293] In addition, it is to be understood that any particular embodiment of the present disclosure that falls within the prior art may be explicitly excluded from any one or more of the claims. Since such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular embodiment of the compositions of the present disclosure can be excluded from any one or more claims, for any reason, whether or not related to the existence of prior art.

[294] The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior disclosure.

Other Embodiments

Those of ordinary skill in the art will readily appreciate that the foregoing represents merely certain preferred embodiments of the present disclosure. Various changes and modifications to the procedures and compositions described above can be made without departing from the spirit or scope of the present disclosure, as set forth in the following claims.

CLAIMS

1. A single sample method of identifying ribonucleic acid (RNA) capping and tailing modifications, the method comprising:

(a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail, and wherein said sample further comprises a first oligonucleotide probe complementary to a sequence in the 5' untranslated region (5' UTR) region of the RNA and a second oligonucleotide probe complementary to a sequence in the 3' untranslated region (3' UTR) region of the RNA;

(b) annealing the RNA sample with the first oligonucleotide probe complementary to a sequence in the 5' UTR of the RNA, and the second oligonucleotide probe complementary to a sequence in the 3' UTR of the RNA;

(c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments;

(d) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS using the sample of step (c) for:

i. identifying capping species in the cap fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS;

ii. identifying untailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS; and

iii. identifying tailed species in the tail fragment by measuring retention time of peaks in a chromatogram generated from LC-UV, and/or mass spectra generated by LC-MS or LC-UV-MS;

thereby simultaneously identifying RNA capping and tailing modifications in the single sample.

2. The single sample method of claim 1, wherein the first base of the first oligonucleotide probe binds at the penultimate base of the RNA or adjacent to the penultimate base of the RNA, such as at least 2-10 nucleotides from the penultimate base of the RNA.

3. A single sample method of simultaneously quantifying RNA capping and tailing efficiency, the method comprising:

- (a) providing a sample comprising RNA, optionally having a 5' cap and/or a 3' poly A tail;
- (b) annealing the RNA sample with a first oligonucleotide probe complementary to a sequence in a 5' untranslated region (5' UTR) of the RNA, and a second oligonucleotide probe complementary to a sequence in a 3' untranslated region (3' UTR) of the RNA;
- (c) treating the RNA sample of step (b) with a nuclease to cleave the RNA into cap and tail fragments;
- (d) performing liquid chromatography with ultraviolet detection (LC-UV) or LC coupled to mass spectrometry (LC-MS) or LC-UV-MS;
- (e) measuring a peak area of each capping species in the cap fragment, and each untailed and tailed species in the tail fragment; and
- (f) quantifying a relative amount of each capping species, quantifying a relative amount of untailed species and characterizing poly A tail in the sample of step (d); thereby simultaneously quantifying RNA capping efficiency and tailing efficiency in the sample.

4. The single sample method of claim 3, wherein the first base of the first oligonucleotide probe binds at the penultimate base of the RNA or adjacent to the penultimate base of the RNA, such as at least 2-10 nucleotides from the penultimate base of the RNA.

5. The method of claim 3 or 4, wherein characterizing the poly A tail comprises (i) mass measurements of poly A tail species, deconvoluting a mass spectrometry signal, thereby providing a mass distribution for tailed species, and (ii) determining based on the corresponding mass distribution, a minimum length, a maximum length, an average length and a degree of polydispersity of the poly A tail.

6. The method of any one of claims 1-5, wherein the RNA is *in vitro* transcribed mRNA.

7. The method of any one of claims 1-6, wherein the RNA is obtained from a step of manufacturing.

8. The method of claim 7, wherein the RNA is obtained from a drug substance (DS) at the final step of manufacturing.
9. The method of any one of claims 1-6, wherein the RNA is obtained from a deformulated drug product (DP).
10. The method of any one of claims 1-9, wherein the RNA is unmodified.
11. The method of any one of claims 1-9, wherein the RNA is modified.
12. The method of any one of claims 1-11, wherein the nuclease is RNase H.
13. The method of any one of claims 1-12, wherein the mass of the cap and the tail fragments is measured by LC-MS.
14. The method of any one of claims 1-13, wherein the capping species, untailed species and tailed species is identified from a UV signal by comparing the retention time of peaks obtained to the retention time of a reference standard.
15. The method of any one of claims 1-14, wherein the capping species is Cap1, Cap0, CapG or uncapped.
16. The method of any one of claims 1-15, wherein the capping species is m7Gpppm7GGACA, m7GpppGGACA, GpppGGACA, GGACA or pppGGACA.

17. The method of claim 15 or 16, wherein the uncapped species is GGACA or pppGGACA.
18. The method of any one of claims 1-17, wherein the untailed species is UGCAUC, wherein U is unmodified uridine.
19. The method of any one of claims 1-17, wherein the untailed species is U*GCAU*C, wherein U* is N1-methylpseudouridine.
20. The method of any one of claims 1-19, wherein the relative amount of each of the capping species is a percentage of total amount of capping species in the sample calculated by dividing an area under the peak of the capping species of interest over a sum of total areas under the peak representing Cap1, Cap0, CapG and uncapped species, multiplied by 100.
21. The method of any one of claims 1-20, wherein the relative amount of untailed species is a percentage calculated by dividing an area under the peak of the untailed species over a sum of areas under the peak of Cap1, Cap0, CapG and uncapped multiplied by 100.
22. The method of any one of claims 1-21, wherein the oligonucleotide probe is between about 10 to 40 nucleotides in length.
23. The method of any one of claims 1-22, wherein the oligonucleotide probe comprises RNA and DNA bases.

24. The method of any one of claims 1-23, wherein the oligonucleotide probe comprises RNA and DNA bases in a ratio of about 10:1, about 9:1, about 8:1, about 7:1, about 6:1, about 5:1, about 4:1 or about 3:1.
25. The method of any one of claims 1-24, wherein the oligonucleotide probe is 10-40 nucleotides long and comprises 4 DNA bases.
26. The method of claim 25, wherein the oligonucleotide probe complementary to a sequence in the 5' untranslated region of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein italicized CTGT are DNA bases.
27. The method of claim 25, wherein the oligonucleotide probe complementary to a sequence in the 3' untranslated region of the RNA is 3'-GGUCGGAACAGGAUUUUUAATTCAA-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.
28. The method of any one of claims 1-27, wherein capping and untailing species are quantified in a single LC-UV or LC-MS or LC-UV-MS analysis.
29. The method of claim 1 or 2, wherein tailed species are characterized in a single LC-MS or LC-UV-MS analysis.
30. The method of any one of claims 1-29, wherein one or more steps is automated.
31. An oligonucleotide that is complementary to a sequence in the 5' untranslated region (5' UTR) of the RNA or 3' untranslated region (3' UTR) of the RNA, wherein the oligonucleotide is between about 10 to 40 nucleotides in length and comprises RNA and DNA bases.

32. The oligonucleotide of claim 31, wherein the oligonucleotide comprises RNA and DNA bases in a ratio of about 10:1, 9:1, 8:1, 7:1, 6:1, 5:1, 4:1 or 3:1.
33. The oligonucleotide of claim 31 or 32, wherein the oligonucleotide is about 10 to 40 nucleotides long and comprises 4 DNA bases.
34. The oligonucleotide of any one of claims 31-33, wherein the oligonucleotide complementary to a sequence in the 5' UTR of the RNA is 3'-CCTGTCUAGCGGACCU-5' (SEQ ID NO: 1), wherein the italicized CTGT are DNA bases.
35. The oligonucleotide of any one of claims 31-33, wherein the oligonucleotide complementary to a sequence in the 3' UTR of the RNA is 3'-GGUCGGAACAGGAUUAUUUUAATTCAA-5' (SEQ ID NO: 2), wherein italicized TTCA are DNA bases.
36. A method of manufacturing RNA having a quantified percentage of capped and untailed mRNA comprising the steps of:
- (a) providing *in vitro* synthesized RNA;
 - (b) annealing the *in vitro* synthesized RNA with a first oligonucleotide complementary to a sequence in a 5' untranslated region (5' UTR) of the RNA, and a second oligonucleotide complementary to a sequence in a 3' untranslated region (3' UTR) of the RNA;
 - (c) treating the RNA with a nuclease to cleave the RNA into cap and tail fragments;
 - (d) identifying the capping species and untailed species by liquid chromatography with UV detection (LC-UV) and/or identifying and measuring the mass of the capping species, untailed species and characterizing the poly A tail by liquid chromatography coupled to mass spectrometry (LC-MS) and/or liquid chromatography with UV detection and coupled to mass spectrometry (LC-UV-MS); and

(e) quantifying a relative amount of each capping species, quantifying a relative amount of untail species using LC-UV or LC-MS or LC-UV-MS and characterizing poly A tail using LC-MS or LC-UV-MS in a single sample simultaneously;

thereby manufacturing RNA comprising a quantified percentage of capped and untail mRNA.

37. The method of claim 36, wherein the first base of the first oligonucleotide binds at the penultimate base of the RNA or adjacent to the penultimate base of the RNA, such as at least 2-10 nucleotides from the penultimate base of the RNA.

38. The method of claim 36 or 37, wherein a 5' cap is added enzymatically, post transcription to the RNA from step (a).

39. The method of claim 36 or 37, wherein the RNA does not comprise a 5' cap.

40. The method of any one of claims 36-39, wherein a 3' tail is encoded in a plasmid during *in vitro* synthesis.

41. The method of any one of claims 36-39, wherein a 3' tail is added enzymatically, post transcription to the RNA from step (a).

42. The method of any one of claims 36-39, wherein the RNA does not comprise a 3' tail.

43. The method of any one of claims 36-42, wherein the RNA is unspliced.

44. The method of any one of claims 36-42, wherein the RNA is spliced.

45. A quality control assay for a manufacturing lot, wherein the assay comprises:
- (a) providing a manufactured RNA sample comprising a 5' cap and/or a 3' tail;
 - (b) quantifying RNA capping and tailing efficiency by the method of claim 3 or 4 in the sample;
 - (c) comparing the capping and tailing efficiency in step (b) to capping and tailing efficiency in a reference sample;
- wherein a reduced amount of capping species relative to a reference sample indicates 5' degradation, and
- wherein an increased amount of untailed species and/or a different poly A tail length or polydispersity indicates 3' degradation,
- thereby simultaneously monitoring 5' degradation and 3' degradation of an RNA product, and determining product quality of the manufactured lot.
46. The quality control assay of claim 45, wherein the reference sample is a stable RNA sample having greater than 75% integrity.
47. The quality control assay of claim 45 or 46, wherein the RNA product is a therapeutic mRNA.
48. The quality control assay of claim 45 or 46, wherein the RNA product is an mRNA vaccine.