

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

(19) World Intellectual Property  
Organization

International Bureau

(43) International Publication Date  
20 June 2019 (20.06.2019)



(10) International Publication Number  
**WO 2019/116346 A1**

(51) International Patent Classification:

C12Q 1/6844 (2018.01) C12N 9/22 (2006.01)

(21) International Application Number:

PCT/IB2018/060121

(22) International Filing Date:

14 December 2018 (14.12.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/599,205 15 December 2017 (15.12.2017) US

(71) Applicant: **NOVARTIS AG** [CH/CH]; Lichtstrasse 35,  
4056 Basel (CH).

(72) Inventors: **BEVERLY, Michael**; Novartis Institutes for  
BioMedical Research, Inc., 250 Massachusetts Avenue,  
Cambridge, Massachusetts 02139 (US). **HAGEN, Caitlin  
Jeanette**; Novartis Institutes for BioMedical Research,  
Inc., 250 Massachusetts Avenue, Cambridge, Massachu-  
setts 02139 (US). **SLACK, Olga**; Novartis Institutes for  
BioMedical Research, Inc., 250 Massachusetts Avenue,  
Cambridge, Massachusetts 02139 (US).

(74) Agent: **NOVARTIS AG**; Lichtstrasse 35, 4056 Basel  
(CH).

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

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

Declarations under Rule 4.17:

— as to applicant's entitlement to apply for and be granted a  
patent (Rule 4.17(ii))

(54) Title: POLYA TAIL LENGTH ANALYSIS OF RNA BY MASS SPECTROMETRY

(57) Abstract: Described herein are methods of analyzing and producing RNA compositions, e.g., mRNA compositions, that include determining the amount of a polyA chain length and/or the relative distribution of polyA chain lengths in a sample from the RNA composition using mass spectrometry, e.g., LC-MS or MALDI-MS.



WO 2019/116346 A1

## POLYA TAIL LENGTH ANALYSIS OF RNA BY MASS SPECTROMETRY

## CROSS-REFERENCE TO RELATED APPLICATION

**[001]** This application claims the priority and the benefit of U.S. Patent Application No. 5 62/599,205, filed December 15, 2017, the contents of which are incorporated herein by their entireties.

## REFERENCE TO A "SEQUENCE LISTING," A TABLE, OR A COMPUTER PROGRAM LISTING APPENDIX SUBMITTED AS AN ASCII FILE

10 **[002]** The Sequence Listing written in file PAT058018-WO-PCT\_SL.TXT, created December 11, 2018, 22,449 bytes in size, machine format IBM-PC, MS-Windows operating system, is hereby incorporated by reference.

## FIELD OF THE INVENTION

15 **[003]** The disclosure relates to methods of analyzing compositions of RNA (e.g., mRNA), e.g., for suitability for use as a therapeutic or for use in making an RNA therapeutic, using mass spectrometry, e.g., liquid chromatography coupled to electrospray mass spectrometry (LC-MS) or matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), to determine the polyA species present in the RNA 20 composition. In particular, the methods can be used to evaluate compositions of RNA and specifically to methods of evaluating compositions of messenger RNA (mRNA) having a polyA tail.

## GOVERNMENT INTERESTS

25 **[004]** This invention was made with government support under contract number HR0011-13-3-0003, awarded by the Defense Advanced Research Projects Agency (DARPA). The government has certain rights in the invention.

## BACKGROUND OF THE INVENTION

30 **[005]** Eukaryotic mRNAs often terminate with a stretch of adenosines on their 3' end. These 3' polyadenosine (polyA) tails can vary in length from 20 to over 200 bases, depending on the species. The polyA tail is necessary for the translation and stability of the mRNA in the cytoplasm, through its interaction with polyA binding protein (PABP), the 5' cap and the eukaryotic initiation factors (eIF).

35 **[006]** Some have postulated that longer tails increase translation efficiency, provide mRNA stability, and can possibly affect immune response, although there has been some recent research suggesting that tail length of mRNA and translation efficiency are linked

only during specific cell development stages. See, Bernstein *et al.*, *Mol. Cell Biol.* **9**: 659-670 (1989); Bernstein & Ross, *Trends Biochem. Sci.* **14**: 373-377 (1989); Koski *et al.*, *J. Immunol.* **172**: 3989-3993 (2004); Beilharz & Preiss, *RNA* **13**: 982-997 (2007); Peng *et al.*, *Methods Mol. Biol.* **419**: 215-230 (2008); Subtelny *et al.*, *Nature* **508**: 66-71 (2014);  
5 and Park *et al.*, *Mol. Cell* **62**: 462-471 (2016). For therapeutic mRNA that produces a specific therapeutic protein or immune stimulating antigen, the longevity and immunogenicity of the mRNA importantly influences the duration or intensity or both of protein expression.

**[007]** Understanding polyA tail length and how it relates to mRNA stability and protein  
10 expression requires methods that accurately characterize the polyA tail length of IVT synthesized mRNA.

#### BRIEF SUMMARY OF THE INVENTION

**[008]** Provided herein is a method of evaluating the quality of an mRNA composition,  
15 including the steps of: (a) providing an evaluation by mass spectrometry of the relative distribution of isolated polyA chain lengths or amount of a polyA chain length that have been cleaved from the mRNA of a sample obtained from the mRNA composition, to provide a test value and (b) providing a determination of whether test value has an relative distribution or amount of a reference value, to thereby evaluate the quality of the  
20 mRNA composition.

**[009]** In some embodiments, the method further comprises performing the mass spectrometry to determine the relative distribution of the isolated polyA chain lengths or the amount of a poly A chain length that have been cleaved from the mRNA of the sample.

25 **[010]** In some embodiments, the mass spectrometry is LC-MS.

**[011]** In some embodiments, the method further comprises providing a sample from the mRNA composition and cleaving the polyA tails from the mRNA in the sample using an enzyme or combination of enzymes that do not cleave adenosine.

**[012]** In some embodiments, the enzyme is ribonuclease T1, RNase CL3 (cusativin),  
30 RNase A or any combination thereof.

**[013]** In some embodiments, the method further comprises isolating the cleaved polyA tails from the mRNA in the sample by hybridizing the cleaved polyA tails to a surface coated substrate conjugated with polynucleotides.

**[014]** In some embodiments, the surface coated substrate is a magnetic bead.

35 **[015]** In some embodiments, the magnetic bead is conjugated with oligo dT.

**[016]** In some embodiments, the mRNA is made by *in vitro* transcription (IVT).

- [017]** Also provided herein is a radiolabel-free method for analyzing the 3'-polyadenosine (polyA) tails of mRNA in an mRNA composition, the method includes the steps of (a) cleaving polyA tails from a sample of the RNA composition using ribonuclease T1, RNase CL3 (cusativin), RNase A or any combination thereof; (b) 5 isolating the cleaved polyA tails by hybridization to surface coated substrate that are conjugated to a polynucleotide; and (c) determining the relative distribution of poly A chain lengths or the amount of a polyA chain in the sample using mass spectrometry, to thereby analyze the polyA tails present in the mRNA composition.
- [018]** In some embodiments, the mass spectrometry is LC-MS.
- 10 **[019]** In some embodiments, the method further includes providing a test value based upon the relative distribution of the polyA chain lengths or amount of the polyA chain in the sample and comparing the test value to a reference value.
- [020]** In some embodiments, the surface coated substrate comprises magnetic beads.
- [021]** In some embodiments, the polynucleotide that is conjugated to the surface coated 15 substrate is oligo dT.
- [022]** In some embodiments, the mRNA is made by *in vitro* transcription (IVT).
- [023]** In some embodiments, the method is completed within five hours.
- [024]** In some embodiments, the polyA tails within the composition range from ~20 A's to ~300 A's in length or longer.
- 20 **[025]** Further provided herein is a method of making an RNA composition, e.g., an mRNA composition, the method includes: (a) providing an RNA sample from the RNA composition; (b) providing a relative distribution of polyA chain lengths or an amount of a polyA chain length from isolated polyA tails from the RNA in the RNA sample, by mass spectrometry, e.g., LC-MS or MALDI-MS, to provide a test value; (c) providing a 25 determination of whether the test value is an amount or has a relative distribution of a reference value; and (d) further processing the RNA composition based upon the determination.
- [026]** In some embodiments, the further processing is one or more of classifying, selecting, accepting or discarding, releasing or withholding, processing into a drug 30 product, shipping, moving to a different location, formulating, labeling, packaging, releasing into commerce, or selling or offering for sale, based upon whether a preselected relationship between the test value and the reference value is met.
- [027]** In some embodiments, the RNA sample has an amount of a polyA chain length or has a profile of polyA chain length distribution of a reference value, and the RNA 35 composition is processed into drug product, formulated, labeled, packaged, or released into commerce based upon the determination.

- [028]** In some embodiments, the RNA sample has an amount of a polyA chain length or has a profile of polyA chain length distribution of a reference value and the production method used to make the RNA composition is used to make an additional batch of the RNA composition.
- 5 **[029]** In some embodiments, the RNA sample does not have an amount of a polyA chain length or does not have a profile of polyA chain length distribution of a reference value, and the RNA composition is discarded or withheld.
- [030]** In some embodiments, the RNA sample does not have an amount of a polyA chain length or does not have a profile of polyA chain length distribution of a reference  
10 value, and the production method used to make the RNA composition is modified.
- [031]** In some embodiments, the production method is modified by one or more of modifying the amount of ATPase used the production process of the RNA composition, e.g., increasing or decreasing, in subsequent batches of the RNA composition and modifying the amount of ATP used in the production process, e.g., increasing or  
15 decreasing, in subsequent batches of the RNA composition.
- [032]** In some embodiments, the method further comprises cleaving the polyA tails from the mRNA in the sample using an enzyme or combination of enzymes that do not cleave adenosine.
- [033]** In some embodiments, the enzyme is ribonuclease T1, RNase CL3 (cusativin),  
20 RNase A or any combination thereof.
- [034]** In some embodiments, the method further comprises isolating the cleaved polyA tails from the mRNA in the sample by hybridizing the cleaved polyA tails to a surface coated substrate conjugated with polynucleotides.
- [035]** In some embodiments, the surface coated substrate is a magnetic bead.
- 25 **[036]** In some embodiments, the magnetic bead is conjugated with oligo dT.
- [037]** In some embodiments, the method further comprises producing the mRNA composition using IVT.
- [038]** In some embodiments, the polyA tails of the mRNA composition are part of the DNA template for IVT.
- 30 **[039]** In some embodiments, the poly A tails are enzymatically added to the IVT produced mRNA in the mRNA composition.
- [040]** In some embodiments, the reference value is a value determined from an RNA sample from a commercially available RNA composition.
- [041]** In some embodiments, the reference value is a value determined from a previous  
35 batch of the RNA composition.
- [042]** In some embodiments, the reference value is or comprises a production standard imposed by a regulatory agency.

**[043]** In some embodiments, the reference value is or comprises a release standard.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**[044]** FIG. 1 is a set of chromatograms showing LC-MS analysis of two methods of making *in vitro* transcribed polyA tailed RNA. The top chromatogram shows that the enzymatically tailed mRNA has a much wider peak, resulting from a larger dispersity of polyA tail lengths, than the bottom chromatogram, which measured mRNA with plasmid-encoded polyA tails.

**[045]** FIG. 2 is an image of a poly-acrylamide gel electrophoretogram, showing the length and integrity of 2100nt RNA with 27 A's (SEQ ID NO: 6), 100 A's (SEQ ID NO: 7) and 117 A's (SEQ ID NO: 8), each plasmid-encoded.

**[046]** FIG. 3 is a chromatogram showing LC-MS analysis of a 10mer and 20mer polyA test sequence (SEQ ID NOS 9 and 10). Equal amounts of each polyA test sequence were measured by the polyA analysis method of T1 cleavage, followed by isolation of the polyA tail on oligo dT 25mer (SEQ ID NO: 14) magnetic beads. The resulting total ion chromatogram from the mass spectrometer shows the T1 cleaved products for the 10 and 20mer polyA sequences (SEQ ID NOS 9 and 10). The presence of single peaks for each polyA sequence shows that digestion was complete after 3hr. FIG. 3 shows that the 10mer is recovered less efficiently than the 20mer.

**[047]** FIG. 4 is a set of three UV250nm chromatograms from the LC-MS analysis of mRNA with polyA tails of 100 (SEQ ID NO: 7), 64 (SEQ ID NO: 11) and 27 A's (SEQ ID NO: 6), respectively. At the 27mer length (the bottom chromatogram), each individual peak represents a single tail species differing by one adenosine. This single nucleotide resolution is lost at the 64mer and the 100mer lengths, where peaks coalesce.

**[048]** FIG. 5 is the resulting deconvoluted electrospray mass spectrum of the single peak in the chromatogram for the 100mer tail length shown in FIG. 4. Mass versus spectral intensity is plotted. The series of peaks separated by the mass of adenosine represents the different polyA tail lengths observed. The peak at mass 33,163.9 (outlined in the box) represents the mass of a plasmid-encoded 100mer polyA tail (SEQ ID NO: 7).

**[049]** FIG. 6 is a graph of average mass spectral signal intensity versus polyA tail length for 3 separate T1 digestions of an mRNA with a polyA tail having a length of 117 A's (SEQ ID NO: 8). FIG. 6 also discloses SEQ ID NOS 15-22, 23-24, 12 and 25-26, respectively, in order of appearance.

**[050]** FIG. 7 is a graph of average mass spectral signal intensity versus polyA tail length for an mRNA with a polyA tail having a length of 64 A's (SEQ ID NO: 11). FIG. 7 also discloses SEQ ID NOS 27-40, respectively, in order of appearance.

**[051]** FIG. 8 is a chromatogram of polyA tails of an mRNA composition made by IVT from plasmids encoding 100mer polyA tails (SEQ ID NO: 7). Minor populations of DNA templates each containing different polyA tail lengths were transcribed into the mRNA and the tail lengths observed by LC-MS closely matched the adenosine lengths found in the sequencing traces. FIG. 8 also discloses SEQ ID NOS 41-53, 15-16 and 19-22, respectively, in order of appearance.

**[052]** FIG. 9 is a chromatogram of polyA tails of an mRNA composition made by IVT from plasmids encoding 27mer tails. The measured tail lengths were longer than expected from the plasmid. FIG. 9 discloses SEQ ID NOS 54, 6 and 55-72, respectively, in order of appearance.

## DETAILED DESCRIPTION OF THE INVENTION

### Introduction

**[053]** *In vivo*, the mRNA 3' polyA tail affects mRNA function in several ways. *In vivo*, polyA tails of mRNA are produced through specific exonuclease trimming of the 3' end followed by poly (A) polymerase.

**[054]** mRNAs can be produced synthetically through *in vitro* transcription (IVT). *In vitro*, mRNAs gain their polyA tails either by encoding the polyA sequence into the template DNA or by having the polyA's added post-synthesis using a polyadenylase. The methods described herein can include a step of producing an mRNA composition, e.g., by either of these methods.

**[055]** Measurement of the tail length in IVT synthesized mRNA can help in understanding how polyA tail length is related to mRNA function. Understanding polyA tail length and how it relates to mRNA stability and protein expression requires methods to accurately characterize the polyA tail length of IVT synthesized mRNA.

**[056]** Assessing the heterogeneity of polyA tails in IVT mRNA can also be important from a clinical standpoint. mRNA is increasingly being produced for therapeutic purposes and having a consistent and well-defined product is important to reproducible activity.

**[057]** There are many techniques to characterize the polyA tail length of mRNA, such as RNase H cleavage, chromatographic methods, PCR based assays like LM-PAT and ePAT, and next generation sequencing methods such as TAIL-seq and PAL-seq. See, e.g., Salles et al., *PCR Methods Appl.* **4**: 317-321 (1995); Meijer et al., *Nucleic Acids Res.* **35**: e132 (2007); Murray & Schoenberg, *Methods Enzymol.* **448**: 483-504 (2008); Janicke et al., *RNA* **18**: 1289-1295 (2012); Chang et al., *Mol. Cell* **53**: 1044-1052 (2014); Subtelny et al., 2014. *Nature* **508**: 66-71. Next generation sequencing approaches have the highest resolution (1 base reported for TAIL-seq), whereas those that use gels or chromatography usually show smears for a population of different tail lengths. However,

PCR-based techniques can be complicated by the indirect nature of their measurement, which must amplify long stretches of adenosines.

**[058]** Unlike other methods, mass spectrometry is a direct measurement technique that does not require labels and has the ability to distinguish between single nucleotides by their differences in mass. The methods described herein use mass spectrometry to study mRNA polyA tail length distributions. Mass spectrometry can provide information about the base composition of known PCR amplified sequence, which allows single nucleotide changes to be identified and multiple PCR products to be distinguished from one another. Furthermore, the methods described herein use mass spectrometry to directly measure many oligonucleotide sequences simultaneously with single nucleotide resolution. The results described in the EXAMPLES used standard LC-MS conditions for oligonucleotides to measure polyA chain lengths within an mRNA composition. The deconvolution of the resulting multiply charged spectra was done by a common software procedure in MS systems.

**[059]** As described in EXAMPLE 1, 2100nt mRNAs with plasmid-encoded tail lengths of 27 (SEQ ID NO: 6), 64 (SEQ ID NO: 11), 100 (SEQ ID NO: 7) or 117 polyA's (SEQ ID NO: 8) were assayed. The results showed that enzymatically tailed mRNA have significant tail length heterogeneity. See, FIG. 1. The methods described herein measured the number of A's in the polyA tails and the numbers closely matched the expected Sanger sequencing results. The methods detected even minor plasmid populations with sequence variations.

**[060]** Moreover, when the plasmid sequence contained a discrete number of polyA's in the tail, the LC-MS analysis revealed a distribution that included tails both longer and shorter than the encoded tail lengths.

25

#### Definitions

**[061]** So that the invention can be more readily understood, some terms are here defined. Additional definitions may be set forth throughout the specification.

**[062]** "3" when used in a nucleotide position refers to a region or position in a polynucleotide or oligonucleotide 3' (*i.e.*, downstream) from another region or position in the same polynucleotide or oligonucleotide. The terms "3' end" and "3' terminus", as used herein in reference to a nucleic acid molecule, refer to the end of the nucleic acid which contains a free hydroxyl group attached to the 3' carbon of the terminal pentose sugar.

**[063]** "5" when used in a nucleotide position refers to a region or position in a polynucleotide or oligonucleotide 5' (*i.e.*, upstream) from another region or position in the same polynucleotide or oligonucleotide. The term "5' end" and "5' terminus", as used herein in reference to a nucleic acid molecule, refers 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. In some embodiments, oligonucleotide primers comprise tracts of poly-adenosine at their 5' termini.

- [001]** "5-methylcytidine" (<sup>5</sup>mC) is a modified nucleoside derived from 5-methylcytosine.
- 5 **5-Methylcytosine** is a methylated form of the DNA base cytosine that may be involved in the regulation of gene transcription. See, WO WO2013/052523 (Moderna Therapeutics).
- [002]** "About" means, approximately the value stated. The term "about" "reflects the inherent uncertainty in any scientific measurement -- i.e., repeated measurements of the same property will not yield exactly the same result due to the limitations of accuracy and
- 10 precision associated with measurement and testing techniques.
- [064]** "Affinity", as used in the art, 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. The skilled artisan will know that several methods have been and can be used to determine affinity. Affinity is a measure of specific binding.
- 15 **[065]** "Cap0" (SEQ ID NO.: 1) is a m7GpppG cap. 5' terminal caps are commercially available, e.g., from TriLink BioTechnologies, Inc., San Diego CA USA.
- [066]** "Chromatography" is a technique for separation of mixtures. The mixture is typically dissolved in a fluid called the "mobile phase," which carries it through a structure holding another material called the "stationary phase." Examples include LC and HPLC.
- 20 **[067]** "IVT" is the *in vitro* transcription of ribonucleic acid (RNA) from a deoxyribonucleic acid (DNA) template. Many IVT techniques are known in the biotechnological arts. For information, see *The Basics: In Vitro Transcription* (2015), available from Thermo Fisher Scientific Inc., Waltham, MA, USA. Many kits for in vitro transcription are commercially available.
- 25 **[003]** "Initiation site" is the initiation site for mRNA transcription. The T7 polymerase promoter best transcribes when the initiating nucleotide is guanosine. It is possible to force transcription to begin with adenosine, but this greatly decreases RNA yield.
- [068]** "LC" is liquid chromatography, a technique used to separate a sample into its individual parts. This separation occurs based on the interactions of the sample with the
- 30 mobile and stationary phases. Many LC techniques are known in the biotechnological arts. For more information, see the *Beginners Guide to UPLC* (2015) and the *HPLC Primer* (2015), both available from Waters Corporation, Milford, MA, USA.
- [004]** "Linearization site" or "linearization sequence". Linearization sequences could include recognition sites for restriction endonucleases (e.g. Dral, BspQI, SapI, BbsI, etc.),
- 35 or ribozyme sequences (e.g. hammerhead, hairpin, hepatitis delta virus, Varkud satellite ribozymes etc.), or T7 polymerase termination sequences. The linearization site consists of a unique restriction enzyme site that, when cut, leaves a precise end for transcription

to run off. Enzymes that cut outside of their recognition sites are most useful for linearization.

**[069]** "Modified" means a changed state or structure of a molecule. A "modified" mRNA contains ribonucleosides that encompass modifications relative to the standard guanine  
5 (G), adenine (A), cytidine (C), and uridine (U) nucleosides. The nonstandard nucleosides can be naturally occurring or non-naturally occurring. RNA can be modified in many ways including chemically, structurally, and functionally, by methods known to those of skill in the biotechnological arts. Such RNA modifications can include, *e.g.*, modifications normally introduced post-transcriptionally to mammalian cell mRNA. Moreover, mRNA  
10 molecules can be modified by the introduction during transcription of natural and non-natural nucleosides or nucleotides, as described in U.S. Pat. No. 8,278,036 (Karikó *et al.*); U.S. Pat. Appl. No. 2013/0102034 (Schrum); U.S. Pat. Appl. No. 2013/0115272 (deFougerolles *et al.*) and U.S. Pat. Appl. No. 2013/0123481 (deFougerolles *et al.*). For examples of incorporation of  $\Psi$  (pseudouridine) or  $m^5C$  (5-methylcytidine) into mRNA,  
15 see, U.S. Pat. No. 8,278,036 (Karikó *et al.*); WO 2015/095351 (Novartis AG); Karikó K *et al. Curr. Opin. Drug Disc. Devel.* 10(5): 523-532 (2007); Karikó K *et al. Mol. Therap.* 16(11): 1833–1840 (2008) and Anderson BR *et al., Nucleic Acids Res.* 38(17): 5884-5892 (2010).

**[070]** "MS" is mass spectrometry, an analytical chemistry technique that helps identify  
20 the amount and type of chemicals present in a sample by measuring the mass-to-charge ratio and abundance of gas-phase ions. A mass spectrum (plural spectra) is a plot of the ion signal as a function of the mass-to-charge ratio. Many MS techniques are known in the biotechnological arts. For more information, see the *MS Primer* (2015) available from Waters Corporation, Milford, MA, USA. See also, Basiri *et al., Bioanalysis* 6(11): 1525-  
25 1542 (2014).

**[071]** "mRNA" is messenger RNA, including eukaryotic messenger RNA. Eukaryotic mRNA can begin at the 5' end with an mRNA cap that is enzymatically synthesized after the mRNA has been transcribed by an RNA polymerase *in vitro*. The mRNA cap facilitates translation initiation while avoiding recognition of the mRNA as foreign and  
30 protects the mRNA from 5' exonuclease mediated degradation.

**[072]** "Nucleoside" or "nucleobase" refer to a base (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 and the nucleobase. When the nucleobase is purine  
35 (*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 usually 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, -NR<sub>2</sub> or halogen groups, where each R is independently H, C<sub>1</sub>-C<sub>6</sub> alkyl or C<sub>5</sub>-C<sub>14</sub> aryl. Ribose examples  
5 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, 2\*-4\*- and 3\*-4\*-linked and other "locked" or "LNA," bicyclic sugar modifications. See, WO 98/22489 (Takeshi Imanishi); WO 98/39352 (Exiqon A/S, Santaris Pharma A/S); and WO 99/14226 (Exiqon A/S).

10 **[073]** "Nucleotide" is a nucleoside in a phosphorylated form (a phosphate ester of a nucleoside), as a monomer unit or within a polynucleotide polymer. A "nucleotide 5'-triphosphate" is 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  
15 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 & Bogdanov, *Advanced Organic Chemistry of Nucleic Acids*  
20 (VCH, New York, 1994).

**[074]** "Nucleic acid", "nucleic acid molecule", "polynucleotide" and "oligonucleotide" refer interchangeably 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  
25 stated, the terms encompass nucleic acid-like structures with synthetic backbones, as well as amplification products. 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 can be linear, branched linear, or circular molecules.

Polynucleotides also have associated counter ions, such as H<sup>+</sup>, NH<sub>4</sub><sup>+</sup>, trialkylammonium,  
30 Mg<sup>+</sup>, Na<sup>+</sup> 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. Throughout the specification, whenever an oligonucleotide is represented by a sequence of letters (chosen, for example, from the four base letters: A (adenosine), C (cytidine), G  
35 (guanosine), and T (thymidine), the nucleotides are presented in the 5' to 3' order from the left to the right. A "polynucleotide sequence" refers to the sequence of nucleotide monomers along the polymer. Unless denoted otherwise, whenever a polynucleotide

sequence is represented, the nucleotides are in 5' to 3' orientation from left to right. 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

5 bases, and which will preferentially hybridize with complementary isoform analog bases. Many such isoform bases are described, *e.g.*, by Benner *et al.*, *Cold Spring Harb. Symp. Quant. Biol.* 52: 53-63 (1987). Analogs of naturally occurring nucleotide monomers include, for example, 7-deazaadenine, 7-deazaguanine, 7-deaza-8-azaguanine, 7-deaza- $\delta$ -azaadenine, 7-methylguanine, inosine, nebularine, nitropyrrole, nitroindole, 2-

10 aminopurine, 2-amino-6-chloropurine, 2,6-diaminopurine, hypoxanthine, pseudouridine ( $\Psi$ ), pseudocytosine, pseudoisocytosine, 5-propynylcytosine, isocytosine, isoguanine), 7-deazaguanine, 2-azapurine, 2-thiopyrimidine, 6-thioguanine, 4-thiothymine, 4-thiouracil, O-6-methylguanine, N6-methyladenine, O-4-methylthymine, 5,6-dihydrothymine, 5,6-dihydrouracil, 4-methylindole, pyrazolo[3,4-D]pyrimidines, "PPG" (U.S. Pat. Nos.

15 6,143,877 (Meyer)) and ethenoadenine (Fasman, in *Practical Handbook of Biochemistry and Molecular Biology*, pp. 385-394 (CRC Press, Boca Raton FL USA, 1989)). See also, the nucleosides described in U.S. Pat. No. 8,278,036 (Karikó *et al.*); WO 2015/095351 (Novartis AG); Karikó K *et al. Curr. Opin. Drug Disc. Devel.* 10(5): 523-532 (2007); Karikó K *et al. Mol. Therap.* 16(11): 1833-1840 (2008) and Anderson BR *et al., Nucleic Acids*

20 *Res.* 38(17): 5884-5892 (2010).

**[075]** "Oligo dT" is a stretch of deoxy-thymidine nucleotides, often used to hybridize to and purify molecules containing polyA, such as mRNA. Oligo dT and oligo Dt magnetic beads are commercially available.

**[005]** "PolyA tail". The polyA tail is important for binding of translational factors and for

25 stability. The polyA tail will be at the 3' end of mRNA can range in length.

**[006]** "Polynucleotide variant" refers to molecules that differ in their nucleotide sequence from a native or reference sequence, which can possess substitutions, deletions, or insertions at certain positions within the amino acid sequence, as shown in WO 2015/006747 A2.

**[076]** "Primers" are short nucleic acid sequences. Polymerase chain reaction (PCR) primers are typically oligonucleotides of short length (*e.g.*, 8-30 nucleotides) that are used in polymerase chain reactions. PCR primers and hybridization probes can readily be developed and produced by those of skill in the art, using sequence information from the target sequence. See, Green & Sambrook, *Molecular Cloning: A Laboratory Manual*,

35 *Fourth Edition* (Cold Spring Harbor Press, Plainview, NY, 2012).

**[077]** A "probe" as used herein is an oligonucleotide probe, a nucleic acid molecule which typically ranges in size from about 50-100 nucleotides to several hundred

nucleotides to several thousand nucleotides in length, in whole number increments. A probe can be any suitable length for use in the method of the invention described herein. Such a molecule is typically used to identify a specific nucleic acid sequence in a sample by hybridizing to the specific nucleic acid sequence under stringent hybridization

5 conditions. Hybridization conditions are known in the biotechnological arts. See, e.g., Green & Sambrook, *Molecular Cloning: A Laboratory Manual, Fourth Edition* (Cold Spring Harbor Press, Plainview, NY, 2012).

**[007]** "Pseudouridine" ( $\Psi$ ) is an isomer of the nucleoside uridine in which the uracil is attached via a carbon-carbon instead of a nitrogen-carbon glycosidic bond. See, WO  
10 WO2013/052523 (Moderna Therapeutics).

**[078]** "RNA" is ribonucleic acid, a ribonucleoside polymer. Each nucleotide in an RNA molecule contains a ribose sugar, with carbons numbered 1' through 5'. A base is attached to the 1' position. In general, the bases are adenine (A), cytosine (C), guanine (G), or uracil (U), although many modifications are known to those of skill in the art. For  
15 example, as described herein, an RNA may contain one or more pseudouracil ( $\Psi$ ) base, such that the pseudouridine nucleotides are substituted for uridine nucleotides. Many other RNA modifications are known to those of skill in the art, as described herein. Procedures for isolating and producing RNA are known to the skilled artisan, such as a laboratory scientist. See, Green & Sambrook, *Molecular Cloning: A Laboratory Manual,*  
20 *Fourth Edition* (Cold Spring Harbor Press, Plainview NY, 2012).

**[079]** "Restriction site for linearization of plasmid DNA template" is a site in the plasmid DNA template that can be cut to generate a linearized DNA, for use in in vitro transcription.

**[080]** A "substitution" is a mutation that exchanges one base for another (*i.e.*, a change  
25 in a single "chemical letter" such as switching an A to a G). Such a substitution could (a) change a codon to one that encodes a different amino acid and cause a small change in the protein produced; (b) change a codon to one that encodes the same amino acid and causes no change in the protein produced ("silent mutations"); or (c) change an amino-acid-coding codon to a single "stop" codon and cause an incomplete protein.

30 **[081]** A "surface coated substrate" is a substrate that is coated with a reagent that binds to a nonradiolabeled tagged probe. In one embodiment, the substrate of the surface coated substrate is a magnetic bead. In one embodiment, the substrate of the surface coated substrate is a polymeric bead. In one embodiment, the substrate of the surface coated substrate is a well-plate.

35 **[082]** "SP6 polymerase" is a DNA-dependent RNA polymerase from the SP6 bacteriophage that catalyzes the formation of RNA in the 5'→ 3'.

- [083] "T1" is ribonuclease (RNase) T1 (EC 3.1.27.3) is a fungal endonuclease that cleaves single-stranded RNA after guanine residues, *i.e.*, on their 3' end. The most commonly studied form of this enzyme is the version found in the mold *Aspergillus oryzae*. RNase T1 is often used to digest denatured RNA prior to sequencing. RNase T1
- 5 is commercially available, *e.g.*, Life Technologies #AM2280 1000units/ $\mu$ L.
- [084] "T3 polymerase" is a DNA-dependent RNA polymerase from the T3 bacteriophage that catalyzes the formation of RNA in the 5'→ 3'.
- [085] "T7 polymerase" is a DNA-dependent RNA polymerase from the T7 bacteriophage that catalyzes the formation of RNA in the 5'→ 3'.
- 10 [086] "T7 polymerase promoter upstream enhancer sequence", SEQ ID NO.: 2, is an enhancer sequence upstream from the T7 polymerase promoter, which helps to increase the yield of RNA in an IVT reaction.
- [087] "T7 polymerase promoter", SEQ ID NO. 3 is a nucleotide sequence for a T7 polymerase to begin transcription. Transcription initiates on the first nucleotide following
- 15 the promoter sequence (typically guanosine).
- [088] The term "target" refers to a molecule of interest. "Target RNA" is an RNA of interest, which can be analyzed by the method of the invention.
- [089] "Transcription initiation nucleotide" is the first nucleotide from which transcription begins. A transcription initiation nucleotide could be A, T, C or G, depending on promoter
- 20 and RNA-polymerase chosen for specific transcript.
- [090] "Transcription" is the first step of gene expression, in which a particular segment of DNA is copied into RNA by the enzyme RNA polymerase.
- [091] "Upstream" refers to the 5' to 3' direction in which RNA transcription takes place, so downstream is toward the 5' end of an RNA molecule.

25

### Methods

- [092] The methods described herein allow for the determination of the polyA tail species present in a composition of RNA, *e.g.*, mRNA, without the need for radiolabeling, by using mass spectrometry, *e.g.*, LC-MS or MALDI-MS, to detect the differences in mass and
- 30 retention time of the polyA tail species within the composition. Applicants discovered that mass spectrometry provides accurate and high-resolution identification of polyA tail lengths in RNA compositions. The methods described herein, are useful for evaluating or processing a RNA composition, *e.g.*, an mRNA composition, to determine whether to accept or reject a batch of RNA, or to guide or control a step in the production of a RNA
- 35 composition, *e.g.*, an mRNA composition.
- [093] Accordingly, in one aspect, the disclosure features a method of analyzing the quality of an RNA composition, *e.g.*, an mRNA composition, *e.g.*, by evaluating or

processing the RNA composition, e.g., mRNA composition, for and/or based upon the polyA chain lengths of the various polyA tails of the RNA, e.g., mRNA, in the composition using mass spectrometry, e.g., LC-MS or MALDI-MS. The analysis of the RNA composition, e.g., mRNA composition, by mass spectrometry can be used to evaluate  
5 processes, intermediates and final products in the production of RNA compositions, e.g., mRNA compositions. The presence, distribution or amount of polyA tail length species can be used in these evaluations.

**[094]** In one embodiment, the method comprises providing an evaluation of a parameter, e.g., the relative distribution of polyA chain lengths or amount of a polyA chain  
10 length in the RNA composition, to provide a test value and, optionally, providing a determination of whether the parameter meets a preselected criteria, e.g., is present in an amount or has a profile of a reference value, thereby evaluating or processing the RNA composition. By way of example, LC-MS analysis of a sample of an RNA composition can resolve mixtures of polyA lengths of the RNA in the composition and provide a profile  
15 of polyA chain lengths within the composition.

**[095]** In one embodiment, the test value, or an indication of whether the preselected relationship is met, can be memorialized, e.g., in a computer readable record.

**[096]** In one embodiment, a decision or step is taken, e.g., the RNA composition is classified, selected, accepted or discarded, released or withheld, processed into a drug  
20 product, shipped, moved to a different location, formulated, labeled, packaged, released into commerce, or sold or offered for sale, depending on whether the preselected relationship is met. E.g., based on the result of the determination, or upon comparison to a reference standard, the RNA composition, e.g., mRNA composition, from which the sample is taken can be processed, e.g., as just described.

**[097]** In one embodiment, the method comprises providing isolated polyA tails cleaved from a sample of the RNA composition and determining a parameter, e.g., the profile and/or the lengths, of the polyA tails from the sample using mass spectrometry, e.g., LC-MS or MALDI-MS, to thereby determine the quality of the RNA composition.

**[098]** In one embodiment, the polyA tail is cleaved from the RNA in the sample using an  
30 enzyme or combination of enzymes that do not cleave adenosine such as, e.g., ribonuclease T1, RNase CL3 (cusativin) and RNase A. In one embodiment, the method further comprises cleaving the polyA tails from the RNA, e.g., mRNA, in the sample, e.g., using an enzyme or combination of enzymes that do not cleave adenosine such as, e.g., ribonuclease T1, RNase CL3 (cusativin) and RNase A. In one embodiment, the method  
35 further comprises isolating the cleaved polyA tails from the RNA in the sample by hybridization to surface coated substrate conjugated with polynucleotides. The method

described herein for polyA tail analysis advantageously provides single nucleotide resolution of the cleaved polyA tails.

**[099]** In one embodiment, the method comprises isolating the cleaved polyA tails from the RNA in the sample and the surface coated substrates comprise magnetic beads. In one embodiment, the polynucleotide conjugated to the surface coated substrates is oligo dT (a stretch of deoxy-thymidine nucleotides). In one embodiment, the surface coated substrates are magnetic beads and the polynucleotide conjugated to the magnetic beads is oligo dT.

**[0100]** In one embodiment, the RNA is made by *in vitro* transcription (IVT), e.g., an *in vitro* transcription method described herein. In one embodiment, the *in vitro* transcribed RNA is an mRNA. In one embodiment, the *in vitro* synthesized RNA comprises modified nucleotides selected from, e.g.:  $\Psi$  (pseudouridine); m<sup>5</sup>C (5-methylcytidine); m<sup>5</sup>U (5-methyluridine); m<sup>6</sup>A (N<sup>6</sup>-methyladenosine); s<sup>2</sup>U (2-thiouridine); Um (2'-O-methyl-U; 2'-O-methyluridine); m<sup>1</sup>A (1-methyladenosine); m<sup>2</sup>A (2-methyladenosine); Am (2'-O-methyladenosine); ms<sup>2</sup>m<sup>6</sup>A (2-methylthio-N<sup>6</sup>-methyladenosine); i<sup>6</sup>A (N<sup>6</sup>-isopentenyladenosine); ms<sup>2</sup>i<sup>6</sup>A (2-methylthio-N<sup>6</sup>-isopentenyladenosine); io<sup>6</sup>A (N<sup>6</sup>-(cis-hydroxyisopentenyl)adenosine); ms<sup>2</sup>i<sup>6</sup>A (2-methylthio-N<sup>6</sup>-(cis-hydroxyisopentenyl)adenosine); g<sup>6</sup>A (N<sup>6</sup>-glycinylylcarbamoyladenosine); t<sup>6</sup>A (N<sup>6</sup>-threonylcarbamoyladenosine); ms<sup>2</sup>t<sup>6</sup>A (2-methylthio-N<sup>6</sup>-threonyl carbamoyladenosine); m<sup>6</sup>t<sup>6</sup>A (N<sup>6</sup>-methyl-N<sup>6</sup>-threonylcarbamoyladenosine); hn<sup>6</sup>A (N<sup>6</sup>-hydroxynorvalylcarbamoyladenosine); ms<sup>2</sup>hn<sup>6</sup>A (2-methylthio-N<sup>6</sup>-hydroxynorvalyl carbamoyladenosine); Ar(p) (2'-O-ribosyladenosine (phosphate)); I (inosine); m<sup>1</sup>I (1-methylinosine); m<sup>1</sup>Im (1,2'-O-dimethylinosine); m<sup>3</sup>C (3-methylcytidine); Cm (2'-O-methylcytidine); s<sup>2</sup>C (2-thiocytidine); ac<sup>4</sup>C (N<sup>4</sup>-acetylcytidine); f<sup>5</sup>C (5-formylcytidine); m<sup>5</sup>Cm (5,2'-O-dimethylcytidine); ac<sup>4</sup>Cm (N<sup>4</sup>-acetyl-2'-O-methylcytidine); k<sup>2</sup>C (lysidine); m<sup>1</sup>G (1-methylguanosine); m<sup>2</sup>G (N<sup>2</sup>-methylguanosine); m<sup>7</sup>G (7-methylguanosine); Gm (2'-O-methylguanosine); m<sup>2</sup><sub>2</sub>G (N<sup>2</sup>,N<sup>2</sup>-dimethylguanosine); m<sup>2</sup>Gm (N<sup>2</sup>,2'-O-dimethylguanosine); m<sup>2</sup><sub>2</sub>Gm (N<sup>2</sup>,N<sup>2</sup>,2'-O-trimethylguanosine); Gr(p) (2'-O-ribosylguanosine (phosphate)); yW (wybutosine); o<sub>2</sub>yW (peroxywybutosine); OHyW (hydroxywybutosine); OHyW\* (undermodified hydroxywybutosine); imG (wyosine); mimG (methylwyosine); Q (queuosine); oQ (epoxyqueuosine); galQ (galactosyl-queuosine); manQ (mannosyl-queuosine); preQ<sub>0</sub> (7-cyano-7-deazaguanosine); preQ<sub>1</sub> (7-aminomethyl-7-deazaguanosine); G<sup>+</sup> (archaeosine); D (dihydrouridine); m<sup>5</sup>Um (5,2'-O-dimethyluridine); s<sup>4</sup>U (4-thiouridine); m<sup>5</sup>s<sup>2</sup>U (5-methyl-2-thiouridine); s<sup>2</sup>Um (2-thio-2'-O-methyluridine); acp<sup>3</sup>U (3-(3-amino-3-carboxypropyl)uridine); ho<sup>5</sup>U (5-hydroxyuridine); mo<sup>5</sup>U (5-methoxyuridine); cmo<sup>5</sup>U (uridine 5-oxyacetic acid); mcmo<sup>5</sup>U (uridine 5-oxyacetic acid methyl ester); chm<sup>5</sup>U (5-(carboxyhydroxymethyl)uridine); mchm<sup>5</sup>U (5-

(carboxyhydroxymethyl)uridine methyl ester); mcm<sup>5</sup>U (5-methoxycarbonylmethyluridine); mcm<sup>5</sup>Um (5-methoxycarbonylmethyl-2'-O-methyluridine); mcm<sup>5</sup>s<sup>2</sup>U (5-methoxycarbonylmethyl-2-thiouridine); nm<sup>5</sup>s<sup>2</sup>U (5-aminomethyl-2-thiouridine); mnm<sup>5</sup>U (5-methylaminomethyluridine); mnm<sup>5</sup>s<sup>2</sup>U (5-methylaminomethyl-2-thiouridine); mnm<sup>5</sup>se<sup>2</sup>U (5-methylaminomethyl-2-selenouridine); ncm<sup>5</sup>U (5-carbamoylmethyluridine); ncm<sup>5</sup>Um (5-carbamoylmethyl-2'-O-methyluridine); cmnm<sup>5</sup>U (5-carboxymethylaminomethyluridine); cmnm<sup>5</sup>Um (5-carboxymethylaminomethyl-2'-O-methyluridine); cmnm<sup>5</sup>s<sup>2</sup>U (5-carboxymethylaminomethyl-2-thiouridine); m<sup>6</sup><sub>2</sub>A (N<sup>6</sup>,N<sup>6</sup>-dimethyladenosine); Im (2'-O-methylinosine); m<sup>4</sup>C(N<sup>4</sup>-methylcytidine); m<sup>4</sup>Cm (N<sup>4</sup>,2'-O-dimethylcytidine); hm<sup>5</sup>C (5-hydroxymethylcytidine); m<sup>3</sup>U (3-methyluridine); cm<sup>5</sup>U (5-carboxymethyluridine); m<sup>6</sup>Am (N<sup>6</sup>,2'-O-dimethyladenosine); m<sup>6</sup><sub>2</sub>Am (N<sup>6</sup>,N<sup>6</sup>,O-2'-trimethyladenosine); m<sup>2,7</sup>G (N<sup>2</sup>,7-dimethylguanosine); m<sup>2,2,7</sup>G (N<sup>2</sup>,N<sup>2</sup>,7-trimethylguanosine); m<sup>3</sup>Um (3,2'-O-dimethyluridine); m<sup>5</sup>D (5-methyldihydrouridine); f<sup>5</sup>Cm (5-formyl-2'-O-methylcytidine); m<sup>1</sup>Gm (1,2'-O-dimethylguanosine); m<sup>1</sup>Am (1,2'-O-dimethyladenosine); tm<sup>5</sup>U (5-taurinomethyluridine); tm<sup>5</sup>s<sup>2</sup>U (5-taurinomethyl-2-thiouridine); imG-14 (4-demethylwyosine); imG2 (isowyosine); andac<sup>6</sup>A (N<sup>6</sup>-acetyladenosine), and combinations thereof.

**[0101]** In one embodiment, the methods described herein are completed within 5 hours, with most of that time spent incubating for enzymatic cleavage.

**[0102]** In one embodiment, the polyA tails present in the sample can be from about 10 A's to about 300 A's (e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, or 300 A's) in length or longer. In one embodiment, the polyA tails present in the sample can be from about 20 A's to about 200 A's in length. In one embodiment, the polyA tails present

in the sample can be from about 20 A's to about 120 A's in length. In one embodiment, the polyA tails present in the sample can be from about 30 A's to about 120 A's in length. In one embodiment, the polyA tails present in the sample can be from about 40 A's to about 120 A's in length. In one embodiment, the polyA tails present in the sample can be from about 50 A's to about 120 A's in length. In one embodiment, the polyA tails present in the sample can be from about 60 A's to about 120 A's in length. In one embodiment, the polyA tails present in the sample can be from about 20 A's to about 190 A's in length. In one embodiment, the polyA tails present in the sample can be from about 20 A's to about 180 A's in length. In one embodiment, the polyA tails present in the sample can be from about 20 A's to about 170 A's in length. In one embodiment, the polyA tails present in the sample can be from about 20 A's to about 160 A's in length. In one embodiment, the polyA tails present in the sample can be from about 20 A's to about 150 A's in length. In one embodiment, the polyA tails present in the sample can be from about 20 A's to about 140 A's in length. In one embodiment, the polyA tails present in the sample can be from about 20 A's to about 130 A's in length.

**[0103]** The methods described herein are useful for analyzing or processing a RNA composition, e.g., to determine whether to accept or reject a batch of a RNA composition or to guide or control a step in the production of a RNA composition.

**[0104]** In one embodiment, the method further comprises a step of further processing the RNA composition. The further processing can be, e.g., one or more of selecting, accepting, processing into a drug product, shipping, formulating, labeling, packaging, or selling the RNA composition. In one embodiment, the further processing comprises processing the RNA composition into a drug product. In another embodiment, the further processing comprises formulating the RNA.

**[0105]** In one embodiment, the RNA composition comprises an RNA drug substance. In one embodiment, the RNA composition is an RNA drug product.

**[0106]** In one embodiment, the evaluation comprises determining if the batch of the RNA composition meets a predetermined reference value, and optionally memorializing the determination, e.g., in a computer readable record.

**[0107]** In one embodiment, the reference value is a value determined from a sample of a commercially available RNA composition. In one embodiment, the reference value is a value determined from a previous batch of the RNA composition. In one embodiment, the reference value is or comprises a production standard imposed by a regulatory agency, e.g., a release standard.

**[0108]** In one embodiment, the method further comprises altering a step in the production of an RNA composition based upon the determination, e.g., modifying, e.g., increasing or decreasing, the amount of ATPase used the production process of the RNA composition

and/or modifying, e.g., increasing or decreasing, the amount of ATP used in the production process.

**[0109]** In another aspect, the disclosure features a method of making an RNA composition, e.g., an mRNA composition, comprising providing an RNA sample from a  
5 RNA composition, providing an evaluation of a parameter, e.g., the relative distribution of polyA chain lengths or amount of a polyA chain length in the RNA sample, by mass spectrometry, e.g., LC-MS or MALDI-MS, to provide a test value; providing a determination of whether the parameter meets a preselected criteria, e.g., is present in an amount or has a profile of a reference value, and further processing the RNA composition  
10 based upon the determination. The further processing can be, e.g., classifying, selecting, accepting or discarding, releasing or withholding, processing into a drug product, shipping, moving to a different location, formulating, labeling, packaging, releasing into commerce, or selling or offering for sale, based upon whether the preselected relationship is met. E.g., based on the result of the determination, or upon comparison to  
15 a reference standard, the RNA composition from which the sample is taken can be processed, e.g., as just described.

**[0110]** In one embodiment, the preselected criteria is met, e.g., the RNA sample has an amount of a polyA chain length or has a profile of polyA chain length distribution of a reference value, and the RNA composition is processed into drug product, formulated,  
20 labeled, packaged, released into commerce based upon the determination. In one embodiment, the preselected criteria is met, e.g., the RNA sample has an amount of a polyA chain length or has a profile of polyA chain length distribution of a reference value and the production method used to make the RNA composition is used to make additional batches of the RNA composition. In one embodiment, when the preselected  
25 criteria is met, it is predictive of or ensures that a batch of RNA composition will meet a release specification.

**[0111]** In one embodiment, the preselected criteria is not met, e.g., the RNA sample does not have an amount of a polyA chain length or does not have a profile of polyA chain length distribution of a reference value, and the RNA composition is discarded or  
30 withheld. In one embodiment, the preselected criteria is not met, e.g., the RNA sample does not have an amount of a polyA chain length or does not have a profile of polyA chain length distribution of a reference value, and the production method used to make the RNA composition is modified. For example, the amount of ATPase used the production process of the RNA composition is modified, e.g., increased or decreased, in  
35 subsequent batches of the RNA composition and/or the amount of ATP used in the production process is modified, e.g., increased or decreased, in subsequent batches of the RNA composition.

**[0112]** In one embodiment, the method comprises evaluating the RNA sample, e.g., by the methods described herein.

**[0113]** In any method described herein, in some embodiments, the method includes a step of cleaving the polyA tails from the mRNA in the sample using an enzyme or  
5 combination of enzymes that do not cleave adenosine. In some embodiments, such cleavage takes about 1 hour, 2 hours, or 3 hours.

#### **[0114] REFERENCE VALUES AND STANDARDS**

**[0115]** In some embodiments, the methods described herein include providing a  
10 comparison of the test value determined with a reference value or values, to thereby evaluate the RNA sample. In preferred embodiments, the comparison includes determining if the test value has a preselected relationship with the reference value, e.g., determining if it meets the reference value. The value need not be a numerical value but, e.g., can be merely an indication of whether the subject entity is present.

**[0116]** In one embodiment, the method includes determining if a test value is equal to or greater than a reference value, if it is less than or equal to a reference value, or if it falls within a range (either inclusive or exclusive of one or both endpoints). By way of example, the amount of the relative distribution of polyA chain lengths in the RNA sample can be determined and, optionally shown to fall within a preselected range, e.g., a range which  
20 corresponds to a range of the reference value.

**[0117]** A reference value, by way of example, can be a value determined from a reference sample (e.g., a commercially available sample or a sample from previous production). The reference value can be numerical or non-numerical. For example, it can be a qualitative value, e.g., yes or no, or present or not present at a preselected level of  
25 detection, or graphic or pictorial. The reference value can also be values for the presence of more than one polyA chain length in an RNA sample. For example, the reference value can be a map of structures present in RNA sample when analyzed by mass spectrometry, e.g., LC-MS, e.g., an LC-MS method described herein. The reference value can also be a release standard (a release standard is a standard which should be  
30 met to allow commercial sale of a product) or production standard, e.g., a standard which is imposed, e.g., by a party, e.g., the FDA, on an RNA composition.

**[0118]** The reference value can be derived from any of a number of sources. The reference value can be one which was set or provided by (either solely or in conjunction with another party, e.g., a regulatory agency, e.g., the FDA), the manufacturer of the drug  
35 or practitioner of a process to make the drug. The reference value can be one which was set or provided by (either solely or in conjunction with another party, e.g., a regulatory agency, e.g., the FDA), a party other than the party manufacturing a drug and practicing a

- method disclosed herein, e.g., another party which manufactures the drug or practices a process to make the drug. The reference value can be one which was set or provided by (either solely or in conjunction with another party) a regulatory agency, e.g., the FDA, to the manufacturer of the drug or practitioner of the process to make the drug, or to another party licensed to market the drug. For example, the reference standard can be a production, release, or product standard required by the FDA. In one embodiment, a reference value is a value required of a pioneer drug (e.g., a drug marketed under an approved NDA) or a generic drug (e.g., a drug marketed or submitted for approval under an ANDA).
- 5
- 10 **[0119]** The reference value can be a statistical function, e.g., an average, of a number of values.
- [0120]** In some embodiments, the reference value refers to a distribution where at least 75% (e.g., 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 95%, 97%, 98%, or 99%) of produced polyAs are within the expected length range.
- 15
- [0121]** In some embodiments, the reference value refers to a distribution where at least 75% of produced polyAs are within the expected length range.
- [0122]** In some embodiments, the reference value refers to a distribution where at least 76% of produced polyAs are within the expected length range.
- 20
- [0123]** In some embodiments, the reference value refers to a distribution where at least 77% of produced polyAs are within the expected length range.
- [0124]** In some embodiments, the reference value refers to a distribution where at least 78% of produced polyAs are within the expected length range.
- [0125]** In some embodiments, the reference value refers to a distribution where at least 79% of produced polyAs are within the expected length range.
- 25
- [0126]** In some embodiments, the reference value refers to a distribution where at least 80% of produced polyAs are within the expected length range.
- [0127]** In some embodiments, the reference value refers to a distribution where at least 81% of produced polyAs are within the expected length range.
- 30
- [0128]** In some embodiments, the reference value refers to a distribution where at least 82% of produced polyAs are within the expected length range.
- [0129]** In some embodiments, the reference value refers to a distribution where at least 83% of produced polyAs are within the expected length range.
- [0130]** In some embodiments, the reference value refers to a distribution where at least 84% of produced polyAs are within the expected length range.
- 35
- [0131]** In some embodiments, the reference value refers to a distribution where at least 85% of produced polyAs are within the expected length range.

- [0132] In some embodiments, the reference value refers to a distribution where at least 86% of produced polyAs are within the expected length range.
- [0133] In some embodiments, the reference value refers to a distribution where at least 87% of produced polyAs are within the expected length range.
- 5 [0134] In some embodiments, the reference value refers to a distribution where at least 88% of produced polyAs are within the expected length range.
- [0135] In some embodiments, the reference value refers to a distribution where at least 89% of produced polyAs are within the expected length range.
- [0136] In some embodiments, the reference value refers to a distribution where at least  
10 90% of produced polyAs are within the expected length range.
- [0137] In some embodiments, the reference value refers to a distribution where at least 91% of produced polyAs are within the expected length range.
- [0138] In some embodiments, the reference value refers to a distribution where at least 92% of produced polyAs are within the expected length range.
- 15 [0139] In some embodiments, the reference value refers to a distribution where at least 93% of produced polyAs are within the expected length range.
- [0140] In some embodiments, the reference value refers to a distribution where at least 94% of produced polyAs are within the expected length range.
- [0141] In some embodiments, the reference value refers to a distribution where at least  
20 95% of produced polyAs are within the expected length range.
- [0142] In some embodiments, the reference value refers to a distribution where at least 96% of produced polyAs are within the expected length range.
- [0143] In some embodiments, the reference value refers to a distribution where at least 97% of produced polyAs are within the expected length range.
- 25 [0144] In some embodiments, the reference value refers to a distribution where at least 98% of produced polyAs are within the expected length range.
- [0145] In some embodiments, the reference value refers to a distribution where at least 99% of produced polyAs are within the expected length range.
- [0146] In some embodiments, the expected length range is  $\pm 20nt$  of the expected  
30 length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 100nt and 140nt.
- [0147] In some embodiments, the expected length range is  $\pm 15nt$  of the expected  
35 length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 105nt and 135nt.

- [0148]** In some embodiments, the expected length range is  $\pm 10nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 110nt and 130nt.
- 5 **[0149]** In some embodiments, the expected length range is  $\pm 5nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 115nt and 125nt.
- [0150]** In some embodiments, the expected length range is  $+5nt$  of expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the
- 10 corresponding DNA template), the expected length range is between 120nt and 125nt.
- [0151]** In some embodiments, the expected length range is  $+6nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 120nt and 126nt.
- [0152]** In some embodiments, the expected length range is  $+7nt$  of the expected length.
- 15 For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 120nt and 127nt.
- [0153]** In some embodiments, the expected length range is  $+8nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 120nt and 128nt.
- 20 **[0154]** In some embodiments, the expected length range is  $+9nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 120nt and 129nt.
- [0155]** In some embodiments, the expected length range is  $+10nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on
- 25 the corresponding DNA template), the expected length range is between 120nt and 130nt.
- [0156]** In some embodiments, the expected length range is  $+11nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 120nt and
- 30 131nt.
- [0157]** In some embodiments, the expected length range is  $+12nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 120nt and 132nt.
- 35 **[0158]** In some embodiments, the expected length range is  $+13nt$  of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on

the corresponding DNA template), the expected length range is between 120nt and 133nt.

**[0159]** In some embodiments, the expected length range is +14nt of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on  
5 the corresponding DNA template), the expected length range is between 120nt and 134nt.

**[0160]** In some embodiments, the expected length range is +15nt of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on  
10 the corresponding DNA template), the expected length range is between 120nt and 135nt.

**[0161]** In some embodiments, the expected length range is +16nt of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on  
the corresponding DNA template), the expected length range is between 120nt and 136nt.

**[0162]** In some embodiments, the expected length range is +17nt of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on  
15 the corresponding DNA template), the expected length range is between 120nt and 137nt.

**[0163]** In some embodiments, the expected length range is +18nt of the expected  
20 length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on the corresponding DNA template), the expected length range is between 120nt and 138nt.

**[0164]** In some embodiments, the expected length range is +19nt of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on  
25 the corresponding DNA template), the expected length range is between 120nt and 139nt.

**[0165]** In some embodiments, the expected length range is +20nt of the expected length. For example, if the expected polyA length is 120nt (SEQ ID NO: 12) (based on  
30 the corresponding DNA template), the expected length range is between 120nt and 140nt.

**[0166]** In some embodiments, the expected length range is within the range of  $-Xnt$  to  $+Ynt$  of the expected length, where X is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30; and Y is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30.

**[0167]** In some embodiments, the expected length or the expected polyA length is the  
35 length of the polyA tail based on the number of nucleotides (e.g., adenosines) in corresponding DNA template.

**[0168] MASS SPECTROMETRY**

**[0169]** The evaluating step can include mass spectral and/or tandem mass spectrometry (MS/MS) techniques. In this technique, parent nucleotide ions are fragmented into  
5 smaller ions which are selected and further fragmented to yield information relating to the nature of the polyA nucleotide mixture. To characterize a type of polyA mixture by mass spectrometry, a type of nucleotide or a particular segment of a type of nucleotide can be given positive and negative charges, or ionized, and volatilized in a mass spectrometer. The ionized, volatilized nucleotide molecules or segment thereof can then analyzed by  
10 the mass spectrometer, which produces a mass spectrum of the nucleotide molecule or segment.

**[0170]** A mass spectrometer determines the weight and/or retention of nucleotide molecules and segments of nucleotide molecules, when a nucleotide molecule or segment is analyzed, the information provided by mass spectrometry can be of used,  
15 e.g., to determine the lengths of various polyA nucleotide segments, present in an RNA sample. Methods such as matrix assisted laser desorption ionization (MALDI), nanospray GC/MS, LC/MS, LC MS/MS are all encompassed within the meaning of mass spectrometry.

**[0171]** Mass spectrometry is a direct measurement technique that does not require  
20 labels. Mass spectrometry can distinguish between single nucleotides by their differences in mass.

**REFERENCES**

**[0172]** The following references may be useful in the practice of the instant invention.

25 **[0173]** PCT patent publication WO 2017/098468 A1 (Novartis AG).

**[0174]** Aitken & Lorsch, A mechanistic overview of translation initiation in eukaryotes. *Nature Struct. Mol. Biol.* **19**: 568-576 (2012).

**[0175]** Beverly, *et al.*, Label-free analysis of mRNA capping efficiency using RNase H probes and LC-MS. *Anal. Bioanal. Chem.* **408**: 5021 ((2016))

30 **[0176]** Gilar, 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).

**[0177]** Harrison *et al.*, PAT-seq: a method to study the integration of 3'-UTR dynamics with gene expression in the eukaryotic transcriptome. *RNA* **21**: 1502-1510 (2015).

35 **[0178]** Moqtaderi *et al.*, Secondary structures involving the poly(A) tail and other 3' sequences are major determinants of mRNA isoform stability in yeast. *Microb. Cell* **1**: 137-139 (2014).

## EXAMPLES

## EXAMPLE 1

- 5 **[0179]** Presented herein are some results from using the LC-MS methods described herein for the characterization of 2100nt IVT synthesized mRNAs, having plasmid-encoded polyA tail lengths between 27 and 117 nucleotides in length (SEQ ID NO: 13).
- [0180]** A sample of 2100nt IVT-synthesized mRNAs was digested with RNase T1. The cleavage fragments containing polyA stretches were then isolated using oligo dT coated  
10 magnetic beads. After washing to remove unbound cleavage products and reaction buffer, the dT bound species were eluted off the beads and then injected into the LC-MS machine for analysis. After analysis, the collected electrospray mass spectra are processed and identified by mass using the known sequence of the mRNA cleavage products.
- 15 **[0181]** The measured distribution of tail lengths closely matched the sequence populations observed in Sanger sequencing.
- [0182]** DNA Template preparation: DNA plasmids were generated using conventional cloning methods. Linearized plasmid DNA templates were analyzed by gel electrophoresis and Sanger nucleotide sequencing prior to mRNA preparation.
- 20 **[0183]** T7, SP3, or T3 RNA polymerase promoter sequences were encoded within the plasmids to generate mRNA using different RNA polymerase forms.
- [0184]** For some plasmids, unique restriction enzyme sites were added to the DNA plasmids to allow for the production of polyA encoded templates or tailless DNA templates. A NotI restriction enzyme site was placed upstream to the polyA tail, so that  
25 linearization with this enzyme generated DNA templates lacking the polyA tail. The BspQI, BbsI, or XhoI restriction sites were placed at the end of the polyA tail and linearization of the DNA with these enzymes generated plasmid DNA templates with polyA tail encoded.
- [0185]** mRNA preparation: mRNA was prepared using standard run-off IVT procedures and then capped (Cap 0; SEQ ID NO: 1) using the *Vaccinia* capping system (Part#  
30 M2080S, New England Biolabs, Ipswich, MA, USA). Following capping, the mRNA was precipitated with LiCl and then brought up in water.
- [0186]** The mRNA DNA template was linearized by either an XhoI or BbsI, or BspQI, or NotI endonucleases to ensure 3' ends of mRNA has either only stretches of adenosines  
35 (BbsI or BspQI), or extra sequence after polyA (XhoI), or no polyA (NotI). The length and integrity of the mRNA was confirmed with gel electrophoresis (BioRad Experion, Hercules, CA, USA). See, FIG. 2.

[0187] Oligonucleotides used to validate the method were purchased from Integrated DNA Technologies (Coralville, IA, USA) and were brought up in DI water and used directly.

[0188] T1 cleavage and polyA tail isolation procedure: 100-150pmol of oligonucleotide or mRNA (previously heated to 95°C and then quickly cooled) in water was added to an Eppendorf tube along with 10µL of 10X RNase H buffer (New England Biolabs, Ipswich, MA, USA) and 2µL of RNase T1 (Life Technologies #AM2280 1000units/µL) followed by water to make a 1X buffer (1X Buffer: 5mM Tris pH 7.5, 0.5mM EDTA, 1M NaCl). The sample was then vortexed briefly and kept at 37°C for 3 hours. After 3 hours 100µL of the mixture was added to 75µL of oligo d) 25 magnetic beads (Dynabeads mRNA Purification Kit, #61006, Invitrogen) and isolated according to the manufacturer's protocol. The final rinse step in the protocol was changed to using 100µL of 100mM ammonium acetate instead of 100µL washing buffer in order to remove any sodium that might interfere with the LC-MS analysis.

[0189] Using the magnet, the ammonium acetate wash was removed from the beads. Next 100 µL of 75% methanol (MeOH) that had been heated to 80°C was added to the beads to release the bound polyA tails. The methanol bead mixture was heated to 80°C on a hot plate for 1 min and placed on a magnet for 1 min. The supernatant containing the cleavage product and released polyA tail was collected and dried down to 10 µL with an evaporative centrifuge at RT for ~45 min. Finally, the sample was resuspended in 50 µL of 100 µM EDTA/1% MeOH for LC-MS analysis.

[0190] LC-MS Analysis: Standard conditions reported in the literature were used for the LC-MS analysis of oligonucleotides. See, Apffel *et al.*, *Anal. Chem.* **69**: 1320-1325 (1997); Gilar, *Anal. Biochem.* **298**: 196-206 (2001); Huber & Oberacher, *Mass Spectrom. Rev.* **20**: 310-343 (2001). Analysis of the cleaved 3' mRNA fragment was conducted with an Acuity UPLC (Waters, Milford, MA, USA) equipped with a TUV detector that was connected to a QExactive orbitrap (Thermo Scientific, Grand Island, NY, USA). Mobile phase A consisted of 200 mM hexafluoro isopropanol + 8.15 mM triethylamine, pH 7.9, and mobile phase B was 100% MeOH See, Apffel *et al.*, *Anal. Chem.* **69**: 1320-1325 (1997). A Waters Acuity C18 BEH, 2.1x100 mm column heated to 75°C with a flow rate of 300 µl/min was used for all analyses. The gradient profile for elution started at 5% B for 1 min followed by a linear ramp to 25% over 12 min. At 12 min, a one min rinse at 90% B began, followed by a return to 5% B at 13 min. UV analysis at 260nm was conducted online prior to the MS.

[0191] All mass spectra were obtained in the negative ion mode, over a scan range of 800-2500  $m/z$  at 35,000 resolution. Source and capillary temperatures were set to 350°C and all spectra were analyzed using Promass Software (Novatia, Newtown, PA, USA). See, Hail *et al.*, *American Biotechnology Laboratory* 12-13 (January 2004). The  
5 deconvolution output mass range was set between 10,000 to 50,000 amu with  $m/z$  input of 900–2000  $m/z$ , and a deconvolution peak width of 1.25 amu.

[0192] Protein expression assay. Luciferase expression of mRNA was assessed by monitoring fluorescence in C2C12 cells 24 h post-transfection. Cells were plated at 5000 cells/well 24 h prior to the transfection, and 25 ng of mRNA was transfected into the cells  
10 using *TransIT* transfection reagent (Mirus, Madison, WI).

[0193] Validation of the method of the invention. Sample mRNA is first digested with RNase T1 which cleaves phosphodiester bonds at the 3' side of guanine. Cleavage fragments containing poly A stretches are then isolated using oligo dT-coated magnetic beads. Magnetic beads functionalized with strands of polythymidine DNA (usually 25 mer in length)  
15 are commonly employed for mRNA isolation and use the poly A tail found in mRNA as a handle for capture via Watson- Crick base pairing. After washing the beads to remove unbound cleavage products and reaction buffer, the bound poly A species are eluted off the beads and then injected into the LC-MS for analysis. After analysis, the collected electrospray mass spectra are processed and using the known sequences of the mRNA  
20 T1 cleavage products are identified by mass.

[0194] The method was validated using two synthetic RNAs, CCUGAAAAAAAAAAAA (SEQ ID NO.: 4) and CCUGAAAAAAAAAAAAAAAAAAAA (SEQ ID NO.: 5). These synthetic RNAs were designed to produce a 10mer or 20mer polyA (SEQ ID NOS 9 and 10) product following T1 digestion. Accordingly, we conducted T1 digestions of a mixture  
25 (100pmol each) of each oligo at each of 1, 3 and 24 hours, which digestions were followed by dT isolation of the polyA tails and LC-MS analysis. Our 1hr and 3hr incubations each produced a single fragment corresponding to the cleaved polyA 10mer and 20mer (SEQ ID NOS 9 and 10) products. Complete digestion of the starting oligo was achieved, as shown by the presence of a single fragment. See, FIG. 3. The 24hr  
30 digestion time however, showed cleavage of the polyA strand itself. Accordingly, we selected the 3hr incubation time for all subsequent analyses. The chromatographic peak of the 10mer was substantially smaller than the 20mer indicating that the smaller strand was captured less efficiently than the 20mer. See, FIG. 3. We expected this less efficient capture result, considering the  $T_m$  and binding strength of the 10mer as compared to the  
35 20mer.

**[0195]** Using the methods described herein, a 2100nt long mRNA with plasmid encoded tail lengths of 27 (SEQ ID NO: 6), 64 (SEQ ID NO: 11), 100 (SEQ ID NO: 7), 117 polyA's (SEQ ID NO: 8), along with two enzymatically tailed samples were examined.

**[0196]** The polyA tails encoded in a DNA plasmid produced much narrower polyA tail distributions than mRNA that was enzymatically tailed post-IVT. In all cases, the enzymatically tailed mRNA produced significantly broader distributions of polyA tails than did plasmid-encoded polyA tails. See, FIG. 1. Enzymatically adenylated mRNAs with tail lengths that ranged from 30 A's to over 120 A's were also generated and analyzed. Our results agreed with those previously reported by Holtkamp *et al.*, 2006. *Blood* **108**: 4009-4017, who also found that enzymatically tailing resulted in a large distribution of tail lengths. Accordingly, and for the purpose of clarity, the inventors did all of our further work in the EXAMPLE using plasmid-encoded polyA tails.

**[0197]** FIG. 4 shows that ion-paired reversed phase chromatographic resolution decreases with increasing tail length and single nucleotide resolution is lost between 27 (SEQ ID NO: 6) and 64 polyA tail length (SEQ ID NO: 11). Chromatographic separation of each distinct tail length is not needed for identification, because the mass spectra from co-eluting polyA species can be distinguished by the deconvolution software.

**[0198]** The processed or deconvoluted mass spectrum shown in FIG. 5 shows the mass for the expected 100mer polyA tail (SEQ ID NO: 7) (mass 33,163 in box), along with a series of masses separated by 329amu (+/- 2 amu), which is the mass of adenosine. Thus, the number of different polyA tail species and their relative abundance can be determined by the number and intensity of peaks separated by the mass of adenosine.

**[0199]** The observed masses are +/- 2 amu from the theoretical 329 difference and this variance arises from the resolution of the mass spectrometer. However, because the next closest nucleotide in mass is guanosine (G), a 16amu difference is easily distinguishable. TABLE 1 shows the masses of the observed and expected T1 cleavage fragments for different tail lengths.

<u>Tail length (CA<sub>n</sub>)</u>	<u>Calculated Molecular weight (avg)</u>	<u>Observed Molecular weight</u>
117	39759.7	38760.6
100	33164.1	33163.9
64	21312.6	21312.2
27	9131.8	9130.8

- [0200]** *PolyA Tail length Distribution*. The reproducibility of the method of the invention was tested on a 2100nt mRNA with a polyA tail length of 117 (SEQ ID NO: 8). Four separate T1 digestions were analyzed. The resulting tail length distributions and the mass spectral intensities showed little variation. See, FIG. 6.
- 5 **[0201]** Our analysis of mRNAs with different polyA tail lengths showed that none of the mRNAs had a single unique tail length. Instead, each mRNA contained a distribution of lengths. See, FIG. 6 and FIG. 7. Closer inspection of our Sanger sequencing results showed populations of plasmids with different polyA tail lengths. For the plasmid encoding 117 A's (SEQ ID NO: 8), the fluorescent traces have a signal for adenosine  
10 ranging from position 109 to 122, despite being called at 117. Specifically, the Sanger sequencing results of the 3' end of the plasmid used to synthesize the mRNA in FIG. 6, which showed fluorescent intensity traces for each base along with the called sequence, was called at having 117 A's (SEQ ID NO: 8). However, close inspection showed that other bases are also present, which decreases the number of A's in the tail to 109 and  
15 that additional A's are present out to 123. These minor populations were transcribed into the mRNA and the observed tail lengths closely matched the adenosine traces in the sequencing. This was also true for plasmids encoding 100mer tails. See, FIG. 8. However, those encoding 27mer tails contained tail lengths that were longer than expected from the plasmid. See, FIG. 9.
- 20 **[0202]** The 3' poly A tail has been shown to affect mRNA function in a variety of ways and measurement of the tail length in IVT- synthesized mRNA can be helpful in understanding how length is related to function. Assessing the heterogeneity of poly A tails in IVT mRNA can also be important from a clinical standpoint as mRNA is increasingly being produced for therapeutic purposes and having a consistent and well- defined product is important to  
25 reproducible activity. In order to characterize IVT mRNA, we developed a simple and rapid LC-MS-based method for directly (no conversion to cDNA) determining poly A tail length with single-nucleotide resolution.
- [0203]** Tail lengths can be analyzed in less than 5 h with most of that time spent incubating for T1 cleavage. We have used magnetic oligo dT beads which are a common  
30 technique for isolating mRNA, and they worked well to capture cleaved poly A tails but did show a bias for capturing longer tails in a mixed population. This bias did not appear to be present for tails around 64 nt and above. Analysis of the isolated poly A tails employed standard LC-MS conditions for oligonucleo- tides and used routine MS data processing to calculate the masses of the tail species. The upper limit of the length of tail that can be  
35 analyzed, the resolution, and the sensitivity of that measurement are dependent on the mass spectrometer. As mentioned earlier, sequences up to 500 nucleotides in length have been analyzed successfully so this is well within the reported range of mRNA poly A tail

lengths. Using the theoretical mass to identify tail length does require knowledge of the 3' sequence in order to calculate the mass of the T1 cleavage product which would prevent it from being used to determine the tail length of unknown mRNAs. Although even without sequence information, the distribution of tail lengths could be obtained from the number of  
5 peaks separated by the mass of a nucleotide.

**[0204]** Our first observations demonstrated that poly A tails encoded in the DNA plasmid produced much narrower poly A tail distributions than mRNA that was enzymatically tailed post-IVT. Using plasmid-encoded poly A tails of 27 (SEQ ID NO: 6), 64 (SEQ ID NO: 11), 100 (SEQ ID NO: 7), and 117 (SEQ ID NO: 8) in length, we found  
10 that the distribution of tail lengths closely matched the sequence populations observed in Sanger sequencing. However, when no variation in tail length was detected in the encoding plasmid, tails greater than the DNA template were observed. The variation in tail length unrelated to the plasmid sequence was attributed to transcriptional slippage by the RNAP. Different RNAPs did not alter tail length distributions and we found that T7, T3, and  
15 SP6 all produced the same amount of slippage. It is possible that mutated or other types of RNAPs could narrow the poly A tail distribution, for example it has been shown that human mitochondrial RNAP has less of tendency to slip than SP6.

**[0205]** The findings here indicate that tying a specific tail length to the amount of protein expression or other mRNA attribute is problematic particularly if enzymatic tailing is used.  
20 This could explain some of the variation between studies looking at optimizing poly A tail length and protein expression. For our specific system, however, we found that protein expression increased up to the maximum 117 poly A tail length (SEQ ID NO: 8) that was studied.

SEQUENCE LISTING

<u>TABLE 2</u> <u>Nucleotide Sequences</u>		
<u>Polynucleotide</u>	<u>Sequence</u>	<u>SEQ ID NO.</u>
Cap0	m7GpppGGGAGACGCGUGUU AAAUAACA	1
T7 polymerase promoter upstream enhancer sequence	GGATCCGGAGGCCGGAGAATTG	2
T7 Polymerase Promoter	TAATACGACTCACTATA	3
synthetic RNA designed to produce a 10mer polyA product following T1 digestion	CCUGAAAAAAAAA	4
synthetic RNA designed to produce a 20mer polyA product following T1 digestion	CCUGAAAAAAAAAAAAAAAAAAAA AA	5

5 **[0207]** The detailed description provided herein is to illustrate the invention, but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the biotechnological arts and are encompassed by the appended claims.

**[0208]** Each of the patents, patent publications, and patent applications, and all documents cited herein or during their prosecution ("application cited documents") and all  
 10 documents cited or referenced in the application cited documents, together with any

instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein and incorporated by reference herein, are hereby incorporated herein by reference, and can be used in the practice of the invention. All documents (e.g., these patents, patent publications and applications and the  
5 application cited documents) are hereby incorporated by reference.

**[0209]** The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element, combination or sub-combination of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

10 **[0210]** While this invention has been disclosed with reference to specific embodiments, other embodiments and variations of this invention can be devised by one of skill in the biotechnological art without departing from the true spirit and scope of the invention. The appended claims are intended to be construed to include all such embodiments and equivalent variations.

15

## CLAIMS

We claim:

- 5 1. A method of evaluating the quality of an mRNA composition, comprising the steps of:
- providing an evaluation by mass spectrometry of the relative distribution of isolated polyA chain lengths or amount of a polyA chain length that have been cleaved from the mRNA of a sample obtained from the mRNA composition, to provide a test value and
- 10 providing a determination of whether test value has an relative distribution or amount of a reference value, to thereby evaluate the quality of the mRNA composition.
2. The method of claim 1, wherein the method further comprises performing the mass spectrometry to determine the relative distribution of the isolated polyA chain
- 15 lengths or the amount of a poly A chain length that have been cleaved from the mRNA of the sample.
3. The method of any of the preceding claims, wherein the mass spectrometry is LC-MS.
- 20 4. The method of any of the preceding claims, wherein the method further comprises providing a sample from the mRNA composition and cleaving the polyA tails from the mRNA in the sample using an enzyme or combination of enzymes that do not cleave adenosine.
- 25 5. The method of claim 4, wherein the enzyme is ribonuclease T1, RNase CL3 (cusativin), RNase A or any combination thereof.
6. The method of any of the preceding claims, wherein the method further comprises
- 30 isolating the cleaved polyA tails from the mRNA in the sample by hybridizing the cleaved polyA tails to a surface coated substrate conjugated with polynucleotides.
7. The method of claim 6, wherein the surface coated substrate is a magnetic bead.
- 35 8. The method of claim 7, wherein the magnetic bead is conjugated with oligo dT.

9. The method of any of the preceding claims, wherein the mRNA is made by *in vitro* transcription (IVT), e.g., by a method described herein.
10. A radiolabel-free method for analyzing the 3'-polyadenosine (polyA) tails of mRNA  
5 in an mRNA composition, comprising the steps of
- (a) cleaving polyA tails from a sample of the RNA composition using ribonuclease T1, RNase CL3 (cusativin), RNase A or any combination thereof;
  - (b) isolating the cleaved polyA tails by hybridization to surface coated substrate that are conjugated to a polynucleotide; and
  - 10 (c) determining the relative distribution of poly A chain lengths or the amount of a polyA chain in the sample using mass spectrometry, to thereby analyze the polyA tails present in the mRNA composition.
11. The method of claim 10, wherein the mass spectrometry is LC-MS.  
15
12. The method of claim 10 or 11, further comprising providing a test value based upon the relative distribution of the polyA chain lengths or amount of the polyA chain in the sample and comparing the test value to a reference value.
- 20 13. The method of claim 10, 11 or 12, wherein the surface coated substrate comprises magnetic beads.
14. The method of any of claims 10, 11, 12 or 13, wherein the polynucleotide that is conjugated to the surface coated substrate is oligo dT.  
25
15. The method of any of claims 10, 11, 12, 13 or 14, wherein the mRNA is made by *in vitro* transcription (IVT).
16. The method of claim any of claims 10, 11, 12, 13,14, or 15, wherein the method is  
30 completed within five hours.
17. The method of any of the preceding claims, wherein the polyA tails within the composition range from ~20 A's to ~200 A's in length.
- 35 18. A method of making an RNA composition, e.g., an mRNA composition, comprising:
- providing an RNA sample from the RNA composition;

providing a relative distribution of polyA chain lengths or an amount of a polyA chain length from isolated polyA tails from the RNA in the RNA sample, by mass spectrometry, e.g., LC-MS or MALDI-MS, to provide a test value;

5 providing a determination of whether the test value is an amount or has a relative distribution of a reference value; and  
further processing the RNA composition based upon the determination.

19. The method of claim 18, wherein the further processing is one or more of classifying, selecting, accepting or discarding, releasing or withholding, processing into a  
10 drug product, shipping, moving to a different location, formulating, labeling, packaging, releasing into commerce, or selling or offering for sale, based upon whether a preselected relationship between the test value and the reference value is met.

20. The method of claim 18, wherein the RNA sample has an amount of a polyA chain  
15 length or has a profile of polyA chain length distribution of a reference value, and the RNA composition is processed into drug product, formulated, labeled, packaged, or released into commerce based upon the determination.

21. The method of claim 18, wherein the RNA sample has an amount of a polyA chain  
20 length or has a profile of polyA chain length distribution of a reference value and the production method used to make the RNA composition is used to make an additional batch of the RNA composition.

22. The method of claim 18, wherein the RNA sample does not have an amount of a  
25 polyA chain length or does not have a profile of polyA chain length distribution of a reference value, and the RNA composition is discarded or withheld.

23. The method of claim 18, wherein the RNA sample does not have an amount of a  
30 polyA chain length or does not have a profile of polyA chain length distribution of a reference value, and the production method used to make the RNA composition is modified.

24. The method of claim 23, wherein the production method is modified by one or  
35 more of modifying the amount of ATPase used the production process of the RNA composition, e.g., increasing or decreasing, in subsequent batches of the RNA composition and modifying the amount of ATP used in the production process, e.g., increasing or decreasing, in subsequent batches of the RNA composition.

25. The method of any of claims 18-24, wherein the method further comprises cleaving the polyA tails from the mRNA in the sample using an enzyme or combination of enzymes that do not cleave adenosine.
- 5
26. The method of claim 25, wherein the enzyme is ribonuclease T1, RNase CL3 (cusativin), RNase A or any combination thereof.
27. The method of any of claims 18-26, wherein the method further comprises  
10 isolating the cleaved polyA tails from the mRNA in the sample by hybridizing the cleaved polyA tails to a surface coated substrate conjugated with polynucleotides.
28. The method of claim 27, wherein the surface coated substrate is a magnetic bead.
- 15 29. The method of claim 28, wherein the magnetic bead is conjugated with oligo dT.
30. The method of any of claims 18-29, wherein the method further comprises producing the mRNA composition using IVT.
- 20 31. The method of claim 30, wherein the polyA tails of the mRNA composition are part of the DNA template for IVT.
32. The method of claim 30, wherein the poly A tails are enzymatically added to the IVT produced mRNA in the mRNA composition.
- 25 33. The method of any of the preceding claims, wherein the reference value is a value determined from an RNA sample from a commercially available RNA composition.
34. The method of any of the preceding claims, wherein the reference value is a value  
30 determined from a previous batch of the RNA composition.
35. The method of any of the preceding claims, wherein the reference value is or comprises a production standard imposed by a regulatory agency.
- 35 36. The method of any of the preceding claims, wherein the reference value is or comprises a release standard.

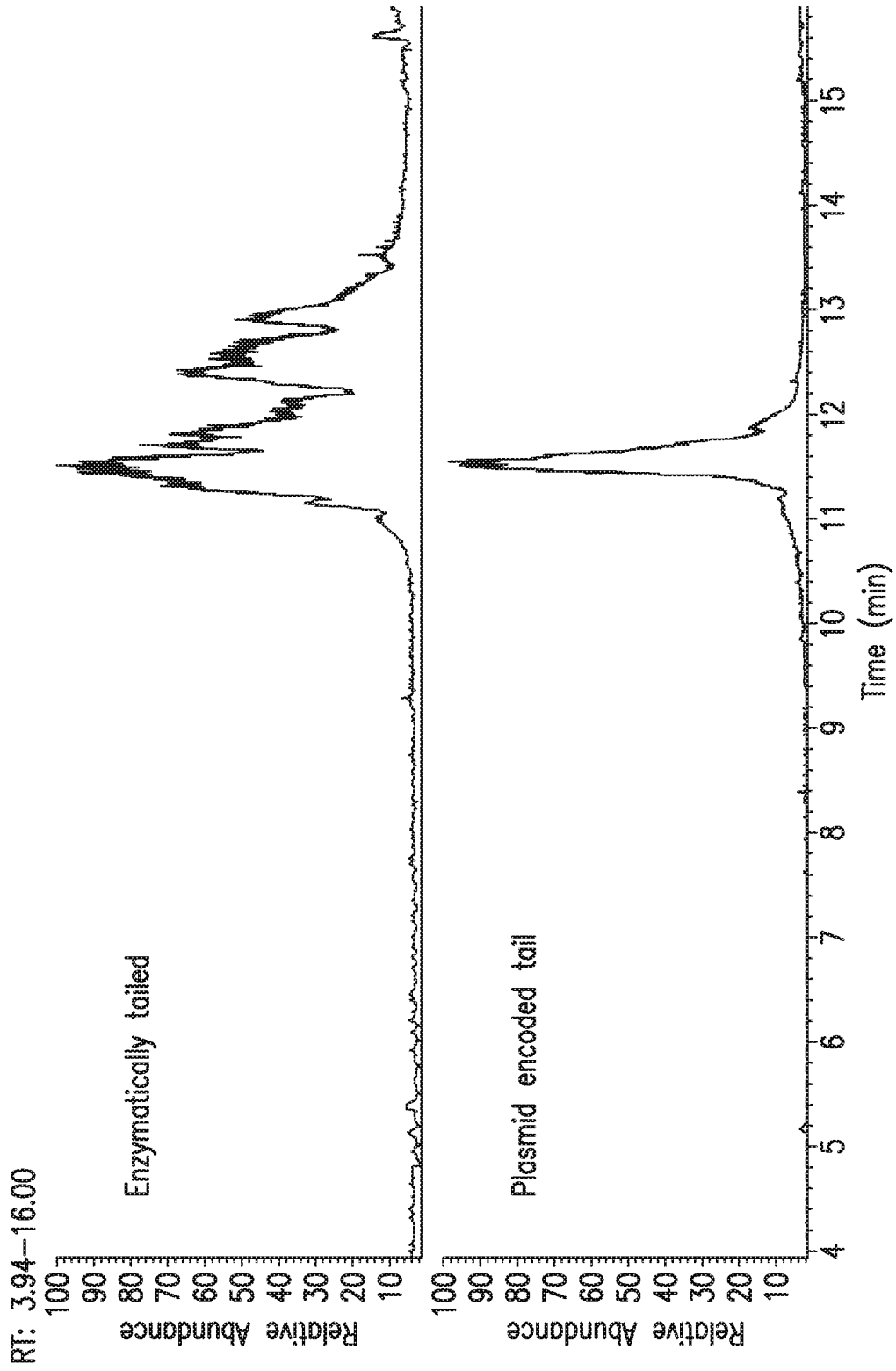


FIG.1

2/9

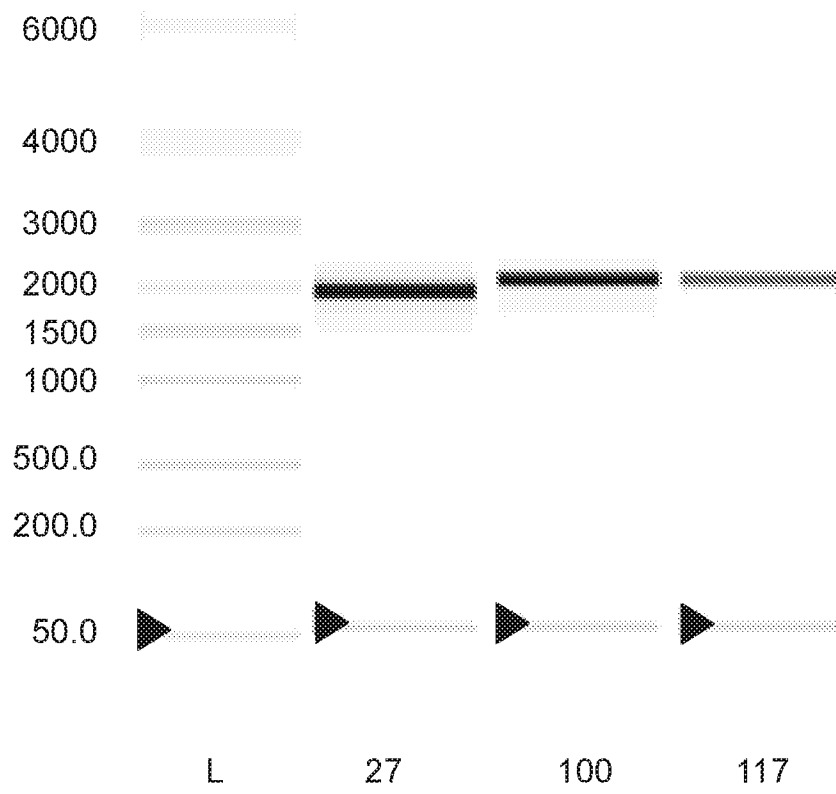


FIG.2

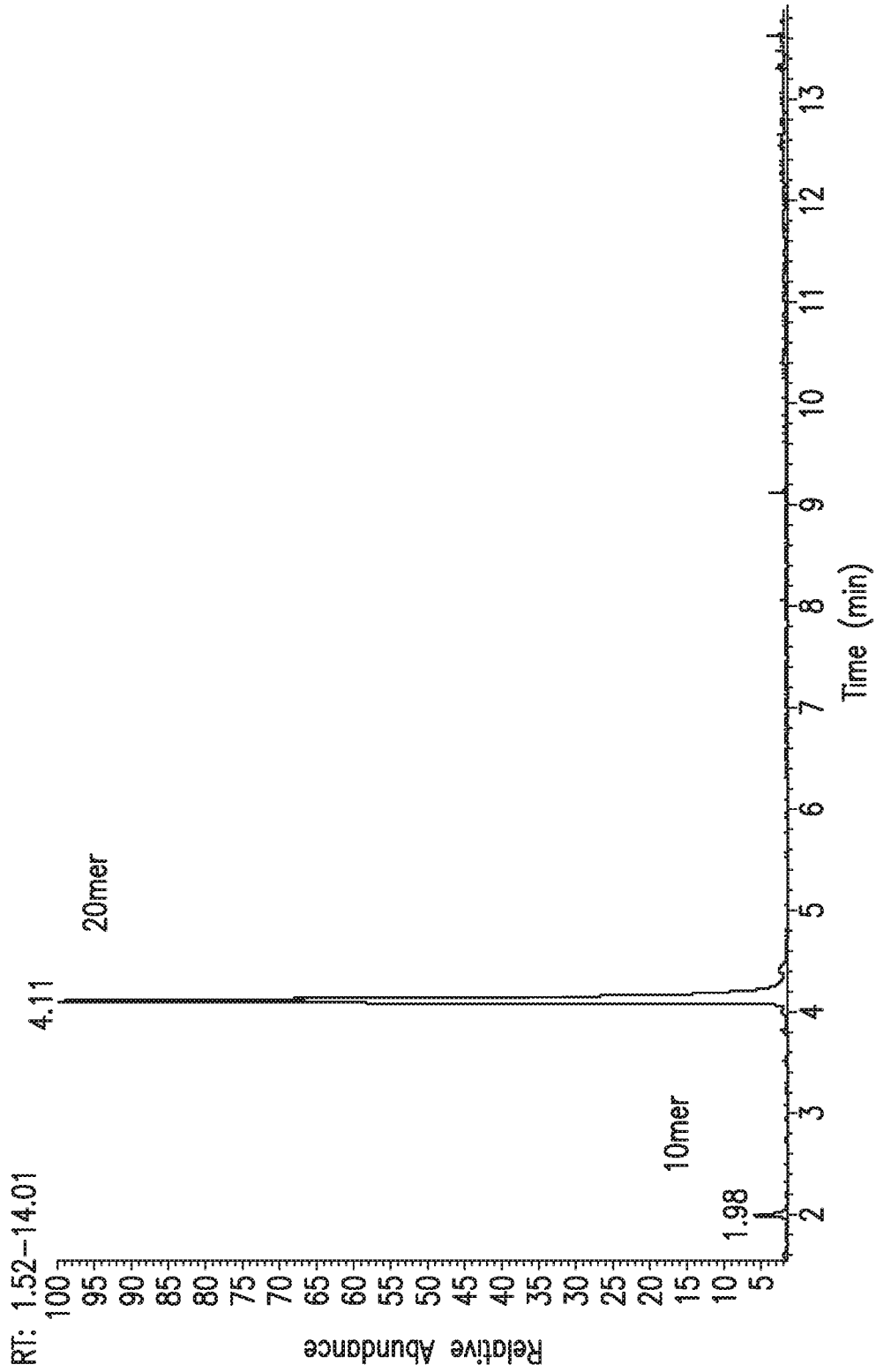


FIG.3

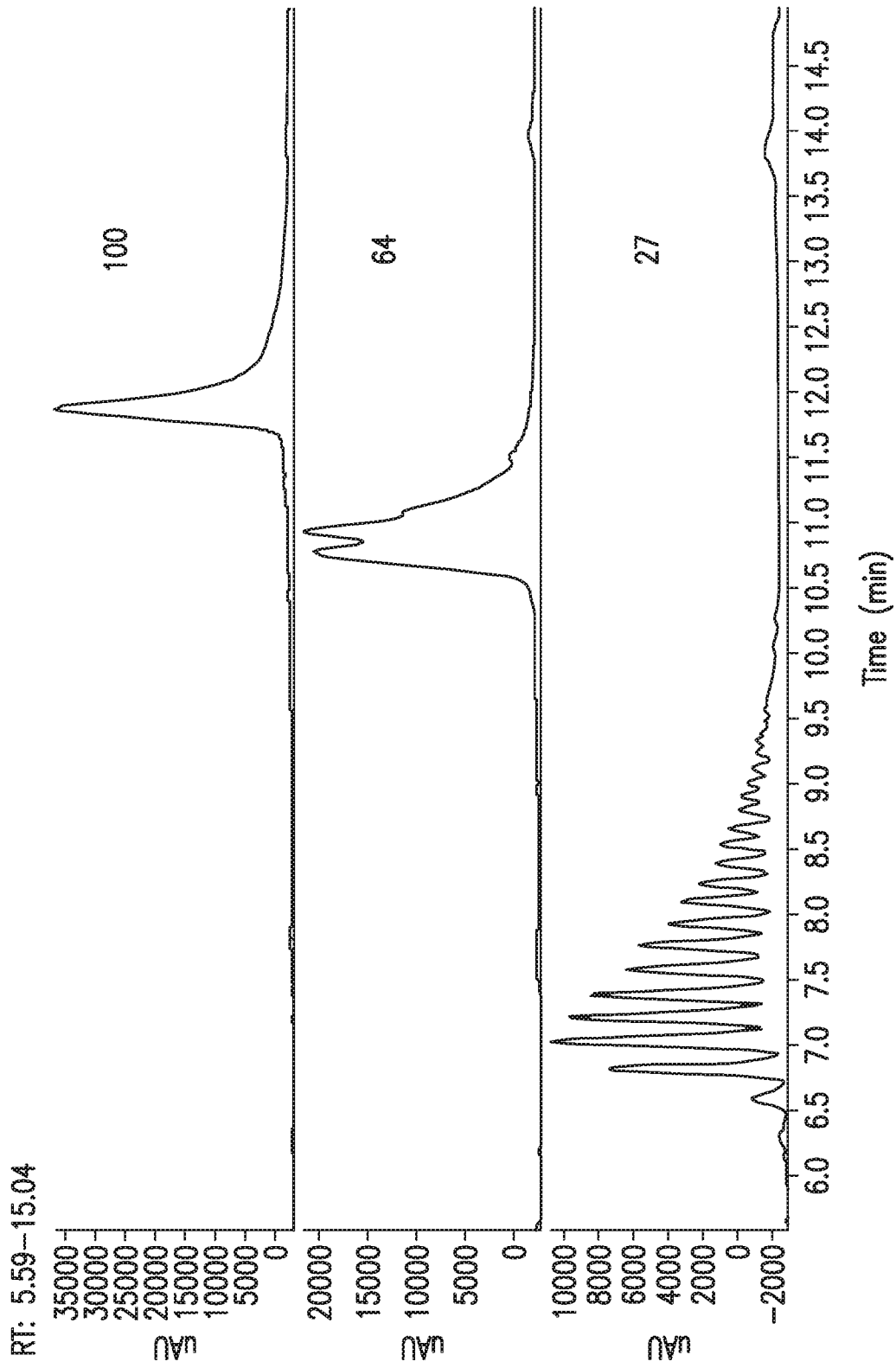


FIG.4

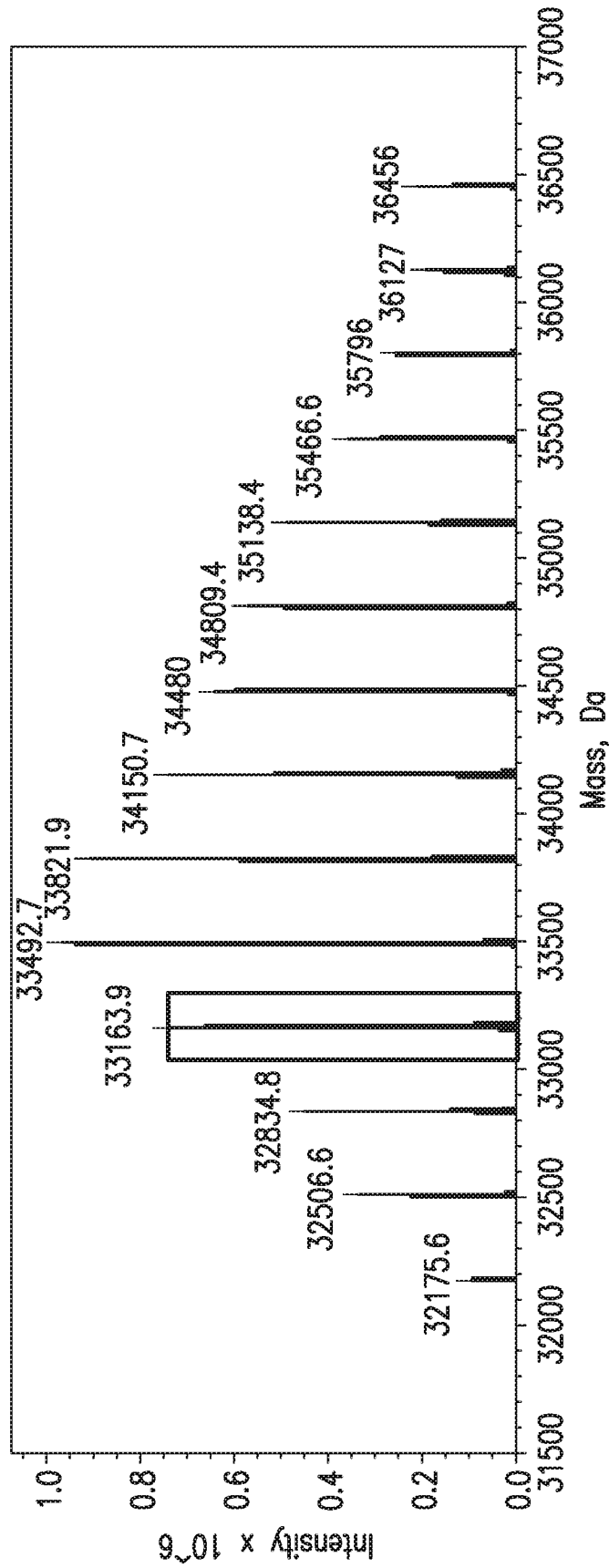


FIG.5

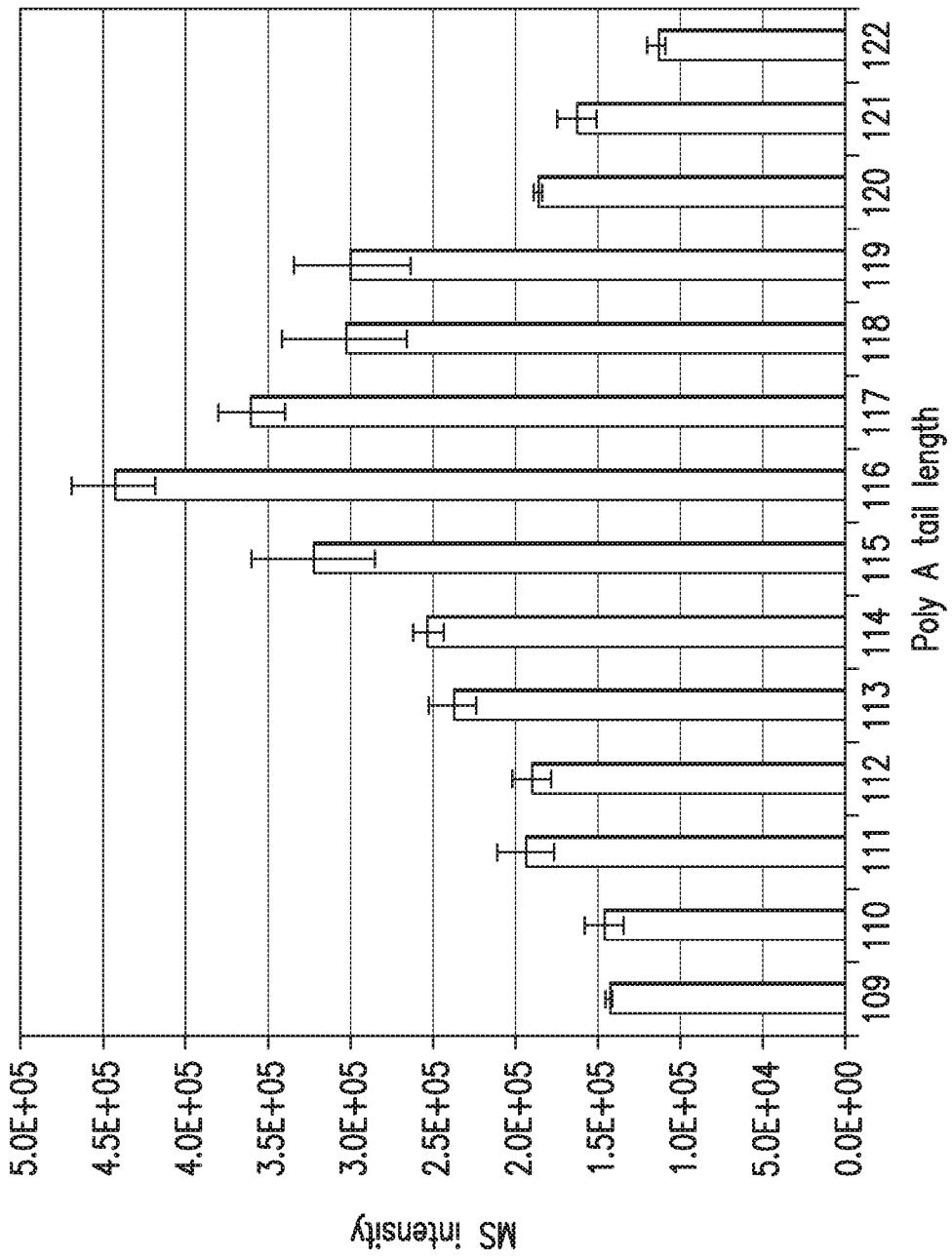


FIG.6

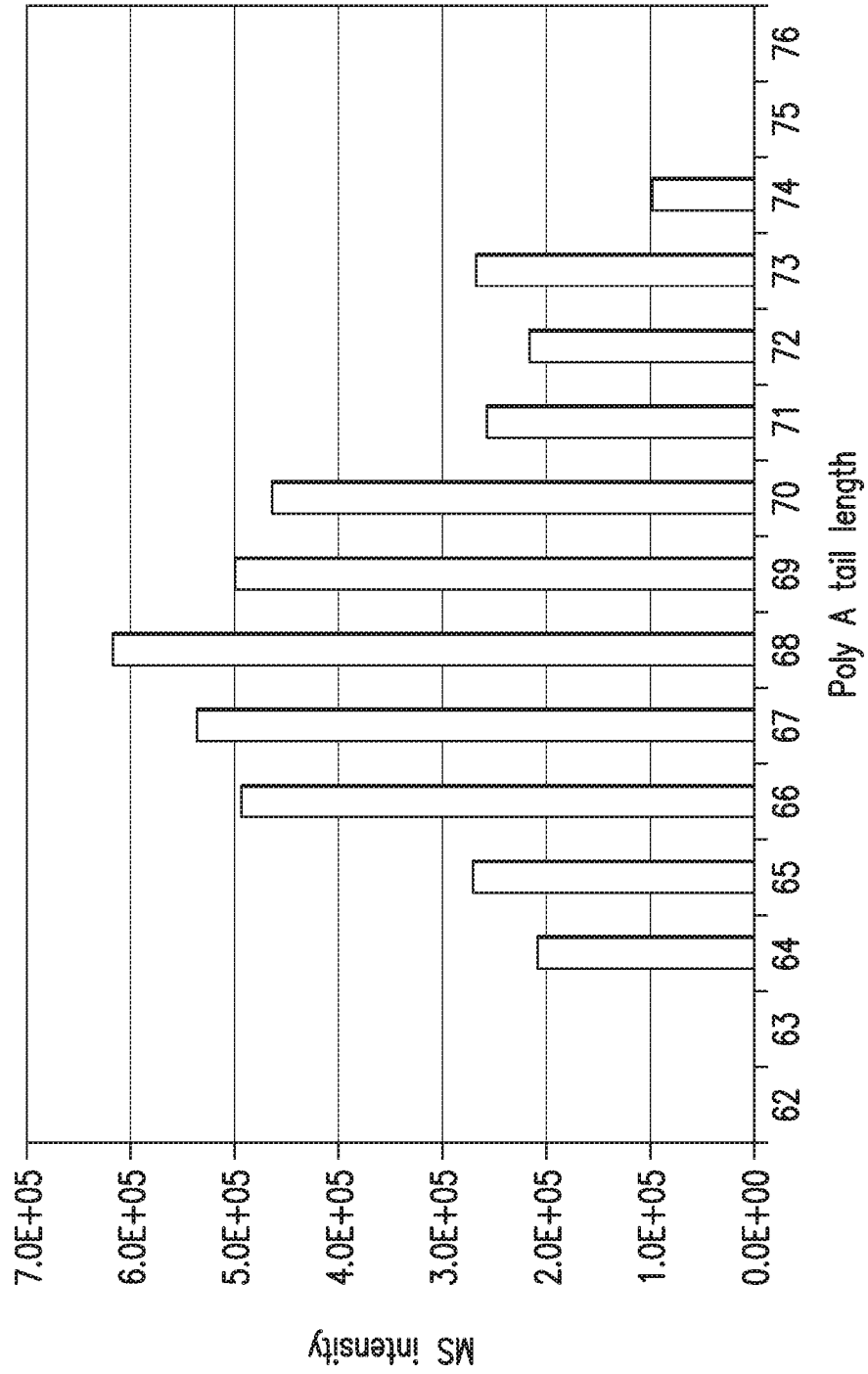


FIG.7

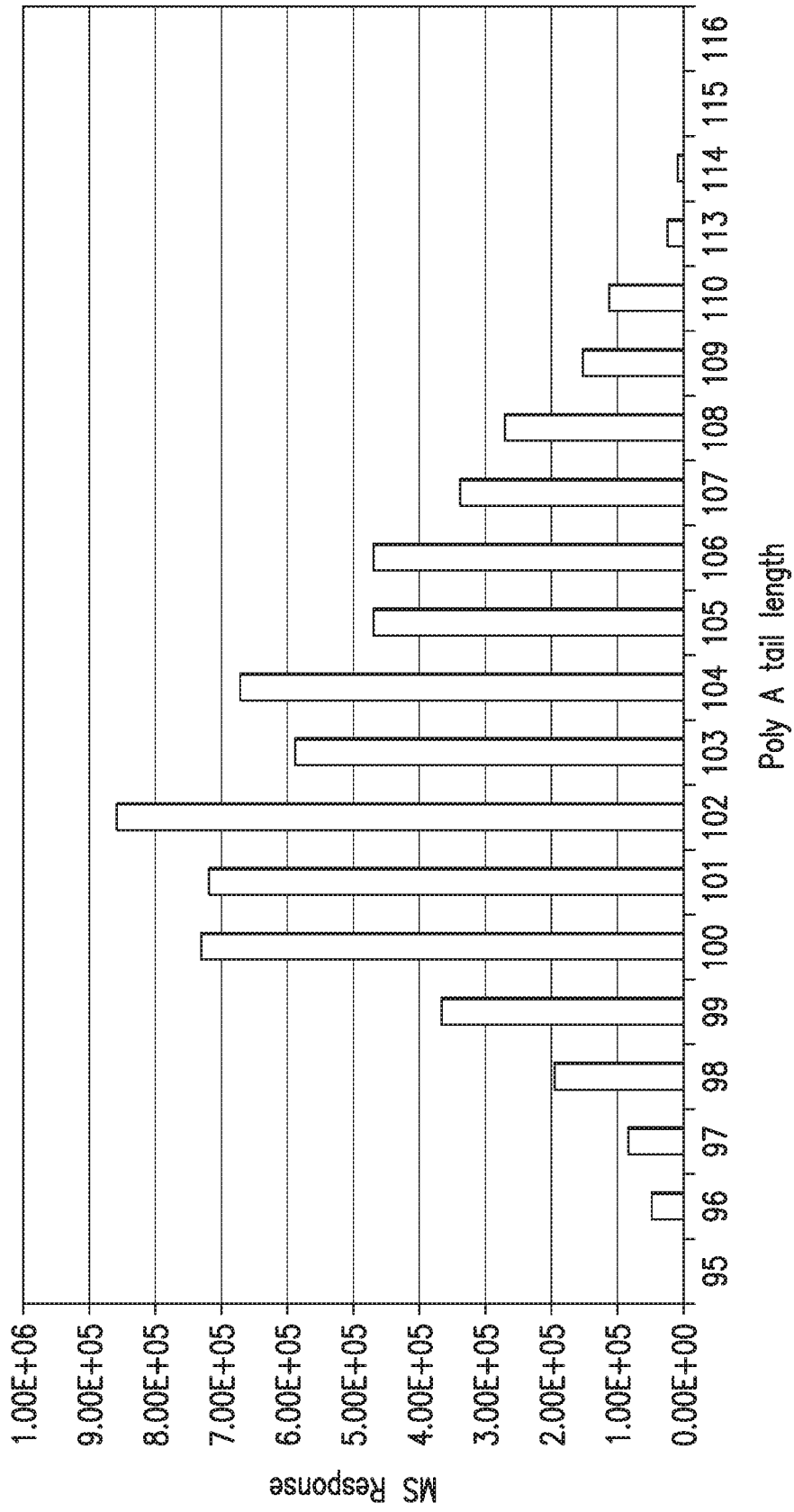


FIG.8

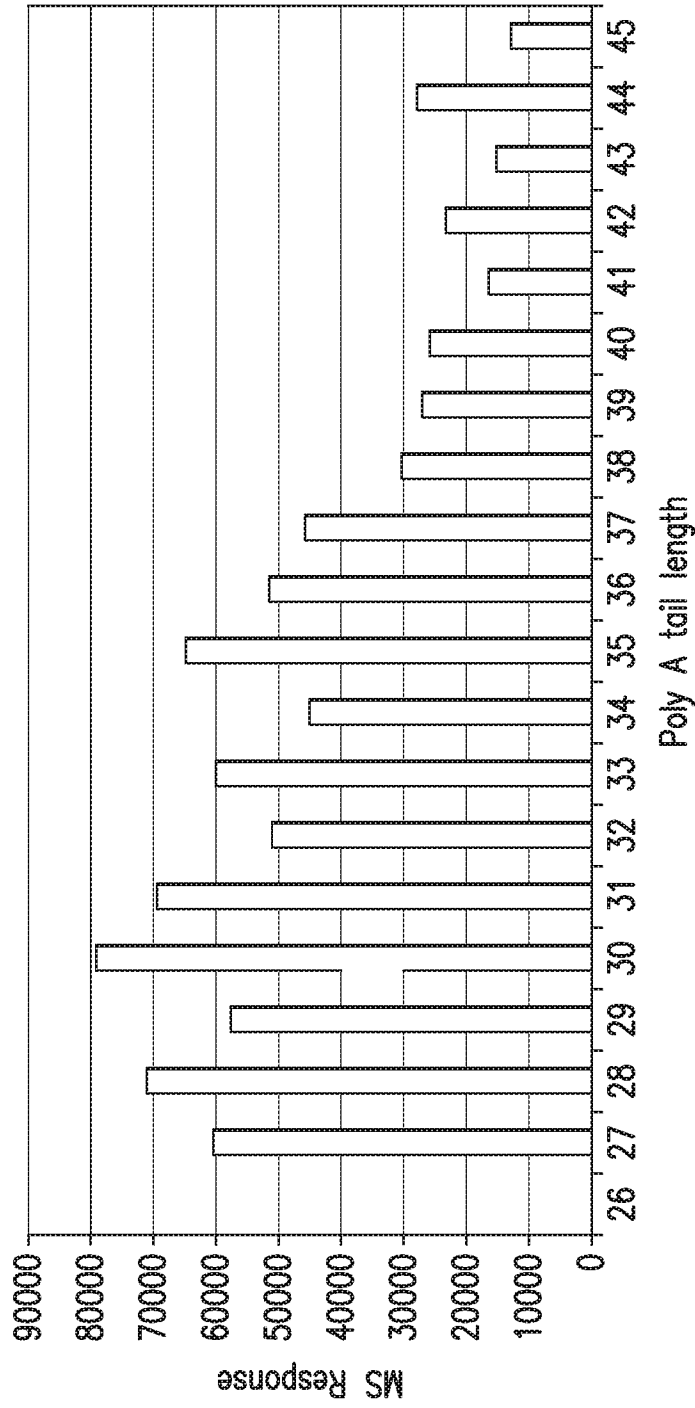


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No  
PCT/IB2018/060121

A. CLASSIFICATION OF SUBJECT MATTER  
INV. C12Q1/6844 C12N9/22  
ADD.  
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED  
Minimum documentation searched (classification system followed by classification symbols)  
C12Q C12N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2016/032273 A1 (SHAHROKH ZAHRA [US] ET AL) 4 February 2016 (2016-02-04) the whole document para. 6, 48, 76, 96-100, 114, 137, 164-166; Fig. 7	1-36
Y	SACHADYN P ET AL: "Naturally occurring mitochondrial DNA heteroplasmy in the MRL mouse", MITOCHONDRION, ELSEVIER, AMSTERDAM, NL, vol. 8, no. 5-6, 1 December 2008 (2008-12-01), pages 358-366, XP025611810, ISSN: 1567-7249, DOI: 10.1016/J.MITO.2008.07.007 [retrieved on 2008-08-12] the whole document p. 363, col. 2, Fig. 5	1-36

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search  5 March 2019	Date of mailing of the international search report  15/03/2019
---	--

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

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/IB2018/060121

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2017/098468 A1 (NOVARTIS AG [CH]) 15 June 2017 (2017-06-15) the whole document para. 7-10, claims 1-15 -----	1-36
Y	ELIZABETH L. MURRAY ET AL: "Assays for determining poly(A) tail length and the polarity of mRNA decay in mammalian cells", METHODS ENZYMOL., vol. 448, 1 January 2008 (2008-01-01), pages 483-504, XP055049447, DOI: 10.1016/S0076-6879(08)02624-4 the whole document p. 5-6 -----	1-36
X,P	WO 2018/081462 A1 (MODERNATX INC [US]) 3 May 2018 (2018-05-03) the whole document p. 5, ll. 21-27; p. 7, ll. 9 - 17; p. 8, ll. 1-4; p. 22, ll. 21-29 -----	1-36
X,P	BEVERLY MICHAEL ET AL: "Poly A tail length analysis of in vitro transcribed mRNA by LC-MS", ANALYTICAL AND BIOANALYTICAL CHEMISTRY, SPRINGER, DE, vol. 410, no. 6, 8 January 2018 (2018-01-08), pages 1667-1677, XP036416023, ISSN: 1618-2642, DOI: 10.1007/S00216-017-0840-6 [retrieved on 2018-01-08] the whole document p. 1668, col. 2 - p. 1669, col. 1 -----	1-36

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2018/060121

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2016032273 A1	04-02-2016	US 2016032273 A1	04-02-2016
		WO 2014144039 A1	18-09-2014
-----			
WO 2017098468 A1	15-06-2017	EP 3387150 A1	17-10-2018
		WO 2017098468 A1	15-06-2017
-----			
WO 2018081462 A1	03-05-2018	US 2018274009 A1	27-09-2018
		WO 2018081462 A1	03-05-2018
-----			