



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

(21) International Application Number:
PCT/US2023/065719

(22) International Filing Date:
13 April 2023 (13.04.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/331,145 14 April 2022 (14.04.2022) US

(71) Applicant: MODERNATX, INC. [US/US]; 200 Technology Square, Cambridge, MA 02139 (US).

(72) Inventors: RABIDEAU, Amy, E.; c/o ModernaTX, Inc., 200 Technology Square, Cambridge, MA 02139 (US). FRANKLIN, Margaret; c/o ModernaTX, Inc., 200 Technology Square, Cambridge, MA 02139 (US). DOUSIS, Athanasios; c/o ModernaTX, Inc., 200 Technology Square, Cambridge, MA 02139 (US).

(74) Agent: LOCKHART, Helen, C. et al.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,

MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: RNA POLYMERASE VARIANTS

(57) Abstract: RNA polymerase variants enable high efficiency transcription of RNA. In some embodiments, the RNA polymerase variants enable RNA transcription with high capping efficiency and/or low levels of double-stranded RNA contamination.

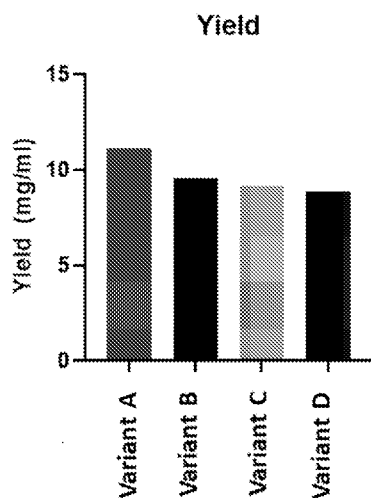


FIG. 1A

RNA POLYMERASE VARIANTS

RELATED APPLICATION

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. provisional application number 63/331,145, filed April 14, 2022, the contents of which are incorporated by reference herein in their entirety.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

The contents of the electronic Sequence Listing (M137870217WO00-SEQ-HJD.xml; Size: 31,370 bytes; and Date of Creation: April 12, 2023) are herein incorporated by reference in their entirety.

BACKGROUND

The emergence of ribonucleic acid (RNA)-based therapeutics requires a polymerase that produces RNA with few byproducts from aberrant activity. Transcripts resulting from *in vitro* transcription using the bacteriophage T7 RNA polymerase exhibit an immune-stimulatory activity that is often undesirable and uncontrollable. This immune-stimulatory activity of T7 transcript is contributed by its aberrant activity to initiate transcription from a promoter-less deoxyribonucleic acid (DNA) end. This activity results in the production of an antisense RNA that is fully complementary to the intended sense RNA product, and consequently a long double-stranded RNA (dsRNA) that can robustly stimulate an unintended immune response. Furthermore, the bacteriophage T7 RNA polymerase produces T7 transcripts having low 5' end capping efficiency in the presence of cap analog(s), in part because the polymerase has low binding affinity for the cap analog(s).

SUMMARY

Some aspects comprise T7 RNA polymerase variants and *in vitro* transcription methods using these variants, which have been shown to reduce dsRNA contaminant and/or improve co-transcriptional 5' end capping efficiency, relative to a control (*e.g.*, wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1).

Some aspects provide an ribonucleic acid (RNA) polymerase variant comprising an amino acid sequence having at least 90% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 2-9, wherein the amino acid sequence comprises an amino acid substitution at position D351 and at least two additional amino acid substitutions, relative to a RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

Some aspects provide an RNA polymerase variant comprising an amino acid sequence that comprises at least one, at least two, at least three, or at least four amino acid substitutions,

relative to a wild-type T7 RNA polymerase (*e.g.*, wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1).

Some aspects provide an RNA polymerase variant comprising an amino acid sequence having at least 90%, at least 95%, at least 98%, or 100% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 2-9.

Some aspects provide an RNA polymerase variant comprising: an amino acid sequence comprising (i) an amino acid substitution at position E350, (ii) an amino acid substitution at D351, and (iii) an amino acid substitution at position K387, position N437, or at position K387 and position N437, relative to a wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the amino acid sequence of the variant comprises an amino acid substitution at position K387.

In some embodiments, the amino acid sequence of the variant comprises an amino acid substitution at position N437.

In some embodiments, the amino acid sequence of the variant comprises an amino acid substitution at position K387 and at position N437.

In some embodiments, the amino acid substitution at position K387 is a polar, neutral amino acid.

In some embodiments, the polar, neutral amino acid is selected from asparagine (N), cysteine (C), glutamine (Q), methionine (M), serine (S), and threonine (T).

In some embodiments, the polar, neutral amino acid is asparagine (K387N).

In some embodiments, the polar, neutral amino acid is cysteine (K387C).

In some embodiments, the polar, neutral amino acid is glutamine (K387Q).

In some embodiments, the polar, neutral amino acid is methionine (K387M).

In some embodiments, the polar, neutral amino acid is serine (K387S).

In some embodiments, the polar, neutral amino acid is threonine (K387T).

In some embodiments, the amino acid substitution at position N437 is an aromatic amino acid.

In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F).

In some embodiments, the aromatic amino acid is tryptophan (N437W).

In some embodiments, the aromatic amino acid is tyrosine (N437Y).

In some embodiments, the aromatic amino acid is phenylalanine (N437F).

Other aspects provide an RNA polymerase variant comprising an amino acid sequence that comprises (i) an amino acid substitution at position E350, (ii) an amino acid substitution at

D351, and (iii) an amino acid substitution at position D653, relative to a wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the amino acid substitution at position D653 is an aromatic amino acid.

5 In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F).

In some embodiments, the aromatic amino acid is tryptophan (D653W).

In some embodiments, the aromatic amino acid is tyrosine (D653Y).

In some embodiments, the aromatic amino acid is phenylalanine (D653F).

10 In some embodiments, the amino acid substitution at position E350 is an aromatic amino acid.

In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F).

In some embodiments, the aromatic amino acid is tryptophan (E350W).

15 In some embodiments, the aromatic amino acid is tyrosine (E350Y).

In some embodiments, the aromatic amino acid is phenylalanine (E350F).

In some embodiments, the amino acid substitution at position D351 is a non-polar, aliphatic amino acid.

20 In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V).

In some embodiments, the non-polar, aliphatic amino acid is alanine (D351A).

In some embodiments, the non-polar, aliphatic amino acid is glycine (D351G).

In some embodiments, the non-polar, aliphatic amino acid is isoleucine (D351I).

In some embodiments, the non-polar, aliphatic amino acid is leucine (D351L).

25 In some embodiments, the non-polar, aliphatic amino acid is proline (D351P).

In some embodiments, the non-polar, aliphatic amino acid is valine (D351V).

Yet other aspects provide an RNA polymerase variant comprising: an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO: 1, wherein the amino acid sequence of the variant comprises (i) an amino acid substitution at position E350, (ii)
30 an amino acid substitution at D351, and (iii) an amino acid substitution at position K387, position N437, or at position K387 and position N437, relative to a wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the amino acid sequence has at least 75%, at least 80%, at least 85%, at least 95%, or at least 98% identity to the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the amino acid sequence of the variant comprises an amino acid substitution at position K387.

In some embodiments, the amino acid sequence of the variant comprises an amino acid substitution at position N437.

5 In some embodiments, the amino acid sequence of the variant comprises an amino acid substitution at position K387 and at position N437.

In some embodiments, the amino acid substitution at position K387 is a polar, neutral amino acid.

10 In some embodiments, the polar, neutral amino acid is selected from asparagine (K387N), cysteine (K387C), glutamine (K387Q), methionine (K387M), serine (K387S), and threonine (K387T).

In some embodiments, the amino acid substitution at position N437 is an aromatic amino acid.

15 In some embodiments, the aromatic amino acid is selected from tryptophan (N437W), tyrosine (N437Y), and phenylalanine (N437F).

Still other aspects provide an RNA polymerase variant comprising: an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO: 1, wherein the amino acid sequence of the variant comprises (i) an amino acid substitution at position E350, (ii) an amino acid substitution at D351, and (iii) an amino acid substitution at position D653, relative to a wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the amino acid sequence has at least 75%, at least 80%, at least 85%, at least 95%, or at least 98% identity to the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the amino acid substitution at position D653 is an aromatic amino acid.

25 In some embodiments, the aromatic amino acid is selected from tryptophan (D653W), tyrosine (D653Y), and phenylalanine (D653F).

In some embodiments, the amino acid substitution at position E350 is an aromatic amino acid.

30 In some embodiments, the aromatic amino acid is selected from tryptophan (E350W), tyrosine (E350Y), and phenylalanine (E350F).

In some embodiments, the amino acid substitution at position D351 is a non-polar, aliphatic amino acid.

35 In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (D351A), glycine (D351G), isoleucine (D351I), leucine (D351L), proline (D351P), and valine (D351V).

Some aspects provide an RNA polymerase variant comprising the amino acid sequence of SEQ ID NO: 2, wherein X^1 is an aromatic amino acid, optionally selected from W, Y, and F; X^2 is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; X^3 is a polar, neutral amino acid, optionally selected from N, C, Q, M, S, and T; and X^4 is an aromatic amino acid, optionally selected from W, Y, and F. In some embodiments, an RNA polymerase variant comprises the amino acid sequence of SEQ ID NO: 6.

Some aspects provide an RNA polymerase variant comprising the amino acid sequence of SEQ ID NO: 3, wherein X^1 is an aromatic amino acid, optionally selected from W, Y, and F; X^2 is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X^4 is an aromatic amino acid, optionally selected from W, Y, and F. In some embodiments, an RNA polymerase variant comprises the amino acid sequence of SEQ ID NO: 7.

Some aspects provide an RNA polymerase variant comprising the amino acid sequence of SEQ ID NO: 4, wherein X^1 is an aromatic amino acid, optionally selected from W, Y, and F; X^2 is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X^3 is a polar, neutral amino acid, optionally selected from N, C, Q, M, S, and T. In some embodiments, an RNA polymerase variant comprises the amino acid sequence of SEQ ID NO: 8.

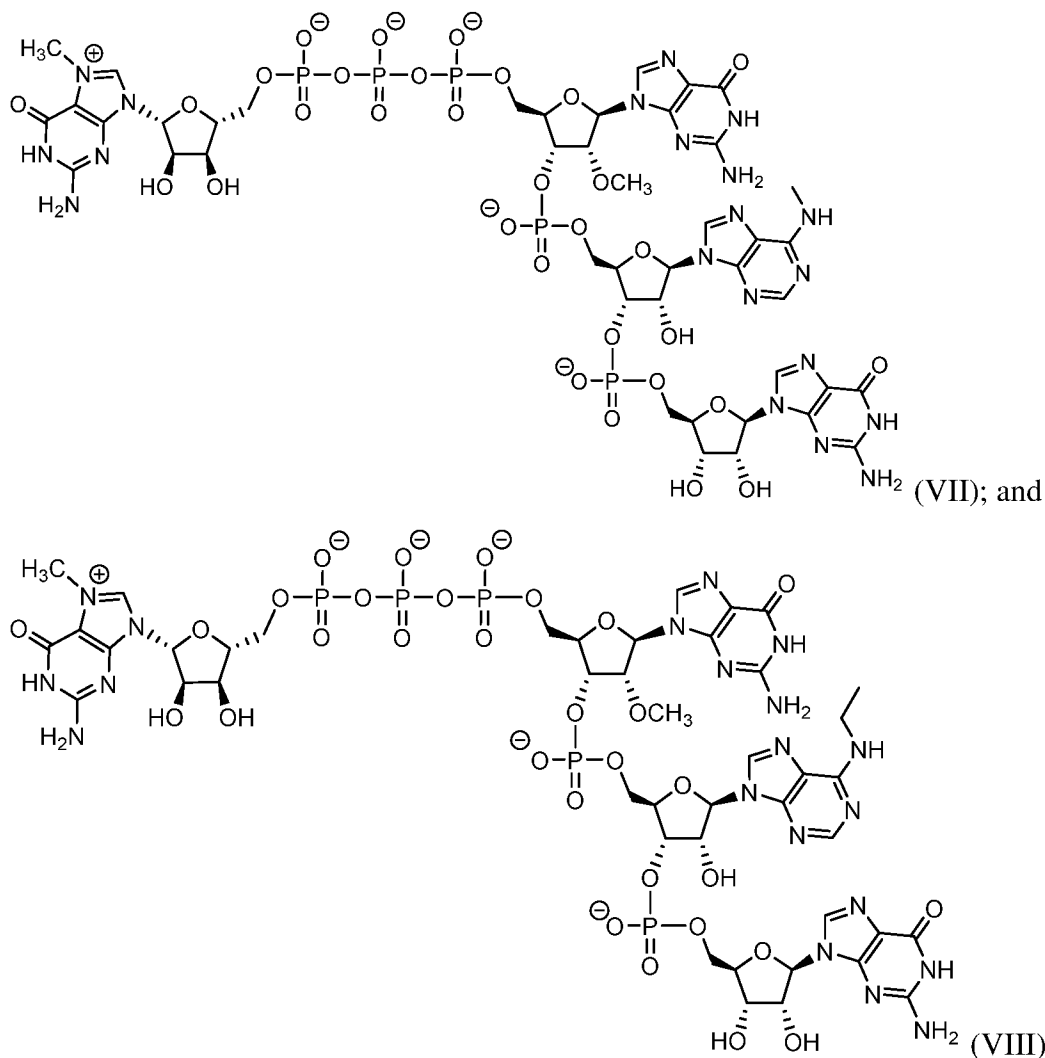
Some aspects provide an RNA polymerase variant comprising the amino acid sequence of SEQ ID NO: 5, wherein X^1 is an aromatic amino acid, optionally selected from W, Y, and F; X^2 is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X^5 is an aromatic amino acid, optionally selected from W, Y, and F. In some embodiments, an RNA polymerase variant comprises the amino acid sequence of SEQ ID NO: 9.

Some aspects provide a method comprising: producing a messenger RNA (mRNA) in an *in vitro* transcription reaction that comprises a DNA, nucleoside triphosphates, the RNA polymerase variant of any one of the preceding paragraphs, and optionally a cap analog.

In some embodiments, the reaction comprises the cap analog.

In some embodiments, the cap analog is a dinucleotide cap analog, a trinucleotide cap analog, or a tetranucleotide cap analog. In some embodiments, the cap analog is a tetranucleotide cap analog.

In some embodiments, the cap analog is a trinucleotide cap analog comprising a GAG sequence. In some embodiments, the GAG cap analog comprises a compound selected from:



In some embodiments, the DNA includes a 2'-deoxythymidine residue or a 2'-deoxycytidine residue at position +1.

Some aspects include a composition or kit comprising the RNA polymerase variant of any one of the preceding paragraphs and an *in vitro* transcription (IVT) reagent selected from the group consisting of a DNA, nucleoside triphosphates, and a cap analog.

Some aspects include a nucleic acid encoding the RNA polymerase variant of any one of the preceding paragraphs.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGs. 1A-1D show graphs depicting the functional characteristics of transcribed RNA products resulting from *in vitro* transcription (IVT) reactions involving exemplary RNA polymerase variants. Following an oligo dT purification, transcribed RNA products were analyzed for yield (FIG. 1A), percent capped RNA (FIG. 1B), percent tailed (*i.e.*, percent of RNA comprising a polyA tail) according to a Tris RP (reverse-phase) method (FIG. 1C), and amount of dsRNA (FIG. 1D).

FIGs. 2A-2C show graphs depicting the functional characteristics of transcribed RNA products resulting from *in vitro* transcription (IVT) reactions involving exemplary RNA polymerase variants in the presence of varying levels of GGAG cap analog. Following an oligo dT purification, transcribed RNA products were analyzed for percent capped RNA (FIG. 2A), yield (FIG. 2B), and percent tailed (*i.e.*, percent of RNA comprising a polyA tail) according to a Tris RP (reverse-phase) method (FIG. 2C).

DETAILED DESCRIPTION

RNA polymerase (*e.g.*, DNA-dependent RNA polymerase) is an enzyme that catalyzes the sequential addition of a ribonucleotide to the 3' end of a growing RNA chain (transcription of RNA in the 5' → 3' direction), with nucleoside triphosphates (NTPs) acting as substrates for the enzyme and with the sequence of nucleotides specified by a DNA template. Transcription relies on the complementary pairing of bases. The two strands of a double helix separate locally, and one of the separated strands serves as a template (DNA template). RNA polymerase then catalyzes the alignment of free nucleotides on the DNA template by their complementary bases in the template. Thus, an RNA polymerase is considered to have RNA polymerase activity if the polymerase catalyzes the sequential addition of a ribonucleotide to the 3' end of a growing RNA chain.

DNA-directed RNA polymerases are capable of initiating synthesis of RNA without primers; the first catalytic stage of initiation is referred to as *de novo* RNA synthesis. *De novo* synthesis is a unique phase in the transcription cycle where the RNA polymerase binds two nucleotides rather than a nascent RNA polymer and a single nucleotide. For bacteriophage T7 RNA polymerase, transcription begins with a marked preference for GTP at the +1 and +2 positions. Initiating nucleotides bind RNA polymerase in locations distinct from those described for elongation complexes (Kennedy WP et al. *J Mol Biol.* 2007; 370(2): 256-68). Selection bias in favor of GTP as an initiating nucleotide is achieved by shape complementarity, extensive protein side-chain, and strong base-stacking interactions for the guanine moiety in the enzyme active site. Thus, an initiating GTP provides the largest stabilization force for the open promoter conformation (Kennedy et al. 2007). The RNA polymerase variants, in some embodiments, comprise one or more amino acid substitution(s) at one or more binding site residue(s) for *de novo* RNA synthesis, which, without being bound by theory, alters RNA polymerase affinity to the cap analog of an *in vitro* transcription reaction, for example, such that there is an improvement in capping efficiency at low cap analog concentrations.

Thus, in some aspects, RNA polymerase variants comprise an RNA polymerase that includes two or more amino acid substitutions at binding site residues for *de novo* RNA

synthesis. An RNA polymerase variant is an enzyme having RNA polymerase activity and at least one substitution and/or modification relative to the counterpart wild-type RNA polymerase. In some embodiments, the amino acid substitution is at a position selected from positions 350, 351, 387, 437, and 653, relative to the wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1.

Structural studies of T7 RNA polymerase have shown that the conformation of the N-terminal domain changes substantially between the initiation phase and elongation phase of transcription. The N-terminal domain comprises a C-helix subdomain and the promoter binding domain, which includes two segments separated by subdomain H. The promoter binding domain and the bound promoter rotate by approximately 45 degrees upon synthesis of an 8-nt RNA transcript, allowing the promoter contacts to be maintained while the active site is expanded to accommodate a growing heteroduplex. The C-helix subdomain moves modestly toward its elongation conformation, whereas subdomain H remains in its initiation- rather than its elongation-phase location, more than 70 angstroms away. Comparison of the structures of the T7 RNA polymerase initiation and elongation complexes reveal extensive conformational changes within the N-terminal 267 residues (N-terminal domain) and little change in the rest of the RNA polymerase. A rigid body rotation of the promoter binding domain as well as the refolding of the N-terminal C-helix (residues 28-71) and H (residues 151-190) subdomains are responsible for abolishing the promoter binding site, enlarging the active site and creating an exit tunnel for the RNA transcript. In particular, residues E42-G47 of T7 RNA polymerase, which exist as a β -loop structure in the initiation complex, adopt an α -helical structure in the elongation complex. The structural changes within the N-terminal domain account for the increased stability and the processivity of the elongation complex (see, *e.g.*, Durniak, K.J. et al., *Science* 322(5901): 553–557, 2008, incorporated herein by reference). T7 RNA polymerase also comprises an ‘N helix’ (residues 374-409) that functions to divert the direction of the 5' end of RNA transcript as it separates from template and influences the stability and processivity of the elongation complex (*e.g.*, through the interactions between residues 385-395 and the ribose backbone). The ‘O helix’ of the RNA polymerase (residues 627-640) functions to stabilize the incoming NTP during insertion and prevent backtracking during synthesis of the RNA transcript. Finally, the ‘Y helix’ (residues 644-661) functions to stabilize the template base at the n+1 position of the growing RNA transcript.

In some aspects are RNA polymerase variants (*e.g.*, T7 RNA polymerase variants) that facilitate the conformational change from the RNA polymerase initiation complex to the RNA polymerase elongation complex. In some embodiments, an RNA polymerase variant comprises at least one, at least two, at least three, or at least four amino acid modifications, relative to wild-

type RNA polymerase, that causes at least one three-dimensional loop structure of the RNA polymerase variant to undergo a conformational change to a helix structure as the RNA polymerase variant transitions from an initiation complex to an elongation complex. Thus, in some embodiments, at least one amino acid modification has a high-helix propensity, relative to wild-type amino acid.

Furthermore, in some aspects are RNA polymerase variants (*e.g.*, T7 RNA polymerase variants) that increase stability and processivity of the elongation complex, prevent backtracking and stabilize the incoming NTPs and template, relative to wild-type T7 RNA polymerase. In some embodiments, an RNA polymerase variant comprises at least one, at least two, at least three, or at least four amino acid modifications, relative to wild-type RNA polymerase, that increase stability and processivity of the elongation complex, prevent backtracking and stabilize the incoming NTPs and template. In some embodiments, an RNA polymerase variant comprises at least one, at least two, at least three, or at least four amino acid modifications, relative to wild-type RNA polymerase, in the 'N helix' (residues 374-409) (*e.g.*, to increase stability and processivity of the elongation complex). In some embodiments, an RNA polymerase variant comprises at least one, at least two, at least three, or at least four amino acid modifications, relative to wild-type RNA polymerase, in the 'O helix' (residues 627-640) (*e.g.*, to stabilize the incoming NTP during insertion and prevent backtracking). In some embodiments, an RNA polymerase variant comprises at least one, at least two, at least three, or at least four amino acid modifications, relative to wild-type RNA polymerase, in the 'Y helix' (residues 644-661) (*e.g.*, to stabilize the growing RNA transcript).

Thus, some aspects provide RNA polymerase variants that comprise multiple amino acid substitutions and/or modifications, relative to wild-type RNA polymerase. In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes (a) an amino acid substitution at a binding site residue for *de novo* RNA synthesis, and (b) an amino acid substitution that facilitates the conformational change from the RNA polymerase initiation complex to the RNA polymerase elongation complex.

Use of the RNA polymerase variants in an *in vitro* transcription reaction, in some embodiments, increases transcription efficiency, relative to a control RNA polymerase. For example, use of an RNA polymerase variant may increase the transcription efficiency (*e.g.*, RNA yield and/or rate of transcription) by at least 20%. In some embodiments, use of an RNA polymerase variant increases the transcription efficiency (*e.g.*, RNA yield and/or rate of transcription) by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 10%. In some embodiments, use of an RNA polymerase variant increases the transcription efficiency by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%,

30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%. In some embodiments, use of an RNA polymerase variant increases the total RNA yield by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 10%. In some embodiments, use of an RNA polymerase variant increases the total RNA yield by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%, 30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%. In some embodiments, use of an RNA polymerase variant increases the rate of transcription by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 10%. In some embodiments, use of an RNA polymerase variant increases the rate of transcription by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%, 30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%. In some embodiments, the control RNA polymerase is a wild-type RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1 ("wild-type T7 RNA polymerase").

Surprisingly, RNA polymerase variants enable the use of a much lower concentration (amount) of cap analog in an *in vitro* transcription reaction to produce an amount of capped RNA equivalent to that produced using the wild-type T7 RNA polymerase. See, for example, FIGs. 1A-2C and Examples 1-2. In some embodiments, use of the RNA polymerase variants in an *in vitro* transcription reaction increases the yield of capped RNA when half the concentration of a cap analog is used in the *in vitro* transcription reaction. In some embodiments, use of the RNA polymerase variants in an *in vitro* transcription reaction increases the yield of capped RNA when only 25%, 50%, or 75% of the concentration of a cap analog is used in the *in vitro* transcription reaction. For example, use of an RNA polymerase variant may increase the yield of capped RNA by at least 20%, when only 25%, 50%, or 75% of the concentration of a cap analog is used in the *in vitro* transcription reaction. In some embodiments, use of an RNA polymerase variant increases the yield of capped RNA by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%, when only 25%, 50%, or 75% of the concentration of a cap analog is used in the *in vitro* transcription reaction. In some embodiments, use of an RNA polymerase variant increases the yield of capped RNA by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%, 30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%, when only 25%, 50%, or 75% of the concentration of a cap analog is used in the *in vitro* transcription reaction. In some embodiments, the control RNA polymerase is a wild-type T7 RNA polymerase.

In some embodiments, use of an RNA polymerase variant increases the total yield of capped RNA by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 10%. In some embodiments, use of an RNA polymerase variant increases the total yield of capped RNA by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%, 30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%.

In some embodiments, use of the RNA polymerase variants in an *in vitro* transcription reaction increases the co-transcriptional capping efficiency. For example, use of an RNA polymerase variant may increase the co-transcriptional capping efficiency (*e.g.*, percentage of transcript comprising cap analog) by at least 20%. In some embodiments, use of an RNA polymerase variant increases the co-transcriptional capping efficiency (*e.g.*, percentage of transcript comprising cap analog) by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In some embodiments, use of an RNA polymerase variant increases the co-transcriptional capping efficiency by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%, 30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%. In some embodiments, the control RNA polymerase is a wild-type T7 RNA polymerase.

In some embodiments, at least 50% of the mRNA comprises a functional cap analog. For example, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 95%, or 100% of the mRNA may comprise a cap analog. In some embodiments, 50%-100%, 50%-90%, 50%-80%, or 50%-70% of the mRNA comprises a cap analog.

In some embodiments, use of the RNA polymerase variants in an *in vitro* transcription reaction improves 3' homogeneity of RNA at half the concentration of a cap analog used in the *in vitro* transcription reaction. For example, use of an RNA polymerase variant may improve 3' homogeneity of RNA by at least 20%, when only 25%, 50%, or 75% of the concentration of a cap analog is used in the *in vitro* transcription reaction. In some embodiments, use of an RNA polymerase variant improves 3' homogeneity by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%, when only 25%, 50%, or 75% of the concentration of a cap analog is used in the *in vitro* transcription reaction. In some embodiments, use of an RNA polymerase variant improves 3' homogeneity by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%, 30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%, when only 25%, 50%, or 75% of the concentration of a cap analog is used in

the *in vitro* transcription reaction. In some embodiments, the control RNA polymerase is a wild-type T7 RNA polymerase.

In some embodiments, at least 50% of the mRNA produced in an *in vitro* transcription reaction that comprises an RNA polymerase variant exhibits 3' homogeneity. For example, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 95%, or 100% of the mRNA exhibits 3' homogeneity. In some embodiments, 50%-100%, 50%-90%, 50%-80%, or 50%-70% of the mRNA exhibits 3' homogeneity.

In some embodiments, the mRNA produced in an *in vitro* transcription reaction that comprises an RNA polymerase variant has greater than a threshold 3' homogeneity. In some embodiments, the threshold is 50% or at least 50%. For example, the threshold may be 55%, 60%, 65%, 70%, 75%, 80%, 85%, or 90%.

In some embodiments, use of the RNA polymerase variants in an *in vitro* transcription reaction improves fidelity (*e.g.*, mutation rate) of transcription. For example, use of an RNA polymerase variant may improve fidelity of transcription by at least 20%. In some embodiments, use of an RNA polymerase variant improves fidelity of transcription by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In some embodiments, use of an RNA polymerase variant improves fidelity of transcription by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%, 30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%. An RNA polymerase variant that improves fidelity of transcription will produce RNA transcript (*e.g.*, mRNA transcript) with a lower rate or total number of mutations than a control RNA polymerase. In some embodiments, the control RNA polymerase is a wild-type T7 RNA polymerase.

In some embodiments, the mRNA produced using an RNA polymerase variant has less than 1 mutation per 100 nucleotides relative to the DNA template. For example, the mRNA produced may have less than 1 mutation per 200, 300, 400, 500, 600, 700, 800, 900 or 1000 nucleotides relative to the DNA template.

In some embodiments, use of the RNA polymerase variants in an *in vitro* transcription reaction lowers the amount of double-stranded RNA (dsRNA) contamination in the *in vitro* transcription reaction. For example, use of an RNA polymerase variant may lower the amount of dsRNA contamination in the *in vitro* transcription reaction by at least 20%. In some embodiments, use of an RNA polymerase variant lowers the amount of dsRNA contamination in the *in vitro* transcription reaction by at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In some embodiments, use of an RNA polymerase variant lowers the amount of dsRNA contamination in the *in vitro* transcription

reaction by 20-100%, 20-90%, 20-80%, 20-70%, 20-60%, 20-50%, 30-100%, 30-90%, 30-80%, 30-70%, 30-60%, 30-50%, 40-100%, 40-90%, 40-80%, 40-70%, 40-60%, 40-50%, 50-100%, 50-90%, 50-80%, 50-70%, or 50-60%. In some embodiments, the control RNA polymerase is a wild-type T7 RNA polymerase.

5 In some embodiments, the concentration of dsRNA contamination is less than 10 ng per 25 µg of mRNA product. In some embodiments, the concentration of dsRNA contamination is less than 5 ng per 25 µg of mRNA product. For example, the concentration of dsRNA contamination may be less than 4 ng per 25 µg of mRNA product, less than 3 ng per 25 µg of mRNA product, less than 2 ng per 25 µg of mRNA product, or less than less than 1 ng per 25 µg of mRNA product. In some embodiments, the concentration of dsRNA contamination is 0.5-1, 10 0.5-2, 0.5-3, 0-4, or 0.5-5 ng per 25 µg of mRNA product.

In some embodiments, the mRNA produced in an *in vitro* transcription reaction that comprises an RNA polymerase variant has lower than a threshold quantity of dsRNA. In some embodiments, the threshold is 10 ng. In some embodiments, the threshold is 5 ng. In some 15 embodiments, the threshold is 4 ng, 3 ng, 2 ng, or 1 ng.

Amino Acid Substitutions and Modifications

RNA polymerase variants include at least one amino acid substitution, preferably at least two amino acid substitutions, relative to the wild type (WT) RNA polymerase. For example, with 20 reference to WT T7 RNA polymerase having an amino acid sequence of SEQ ID NO:1, the glutamic acid (E) at position 350 is considered a “wild-type amino acid,” whereas a substitution of the glutamic acid for tryptophan at position 350 is considered an “amino acid substitution.” In some embodiments, the RNA polymerase variant is a T7 RNA polymerase variant comprising at least one (one or more) amino acid substitution relative to WT RNA polymerase (*e.g.*, WT T7 25 RNA polymerase having an amino acid sequence of SEQ ID NO:1).

In some embodiments, RNA T7 polymerase variants comprise at least two amino acid substitutions. In some embodiments, an RNA T7 polymerase variant comprises at least three amino acid substitutions. In some embodiments, an RNA T7 polymerase variant comprises at least four amino acid substitutions. In some embodiments, an RNA T7 polymerase variant 30 comprises at least five amino acid substitutions.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an (at least one) amino acid modification that causes a loop structure of the RNA polymerase variant to undergo a conformational change to a helix structure as the RNA polymerase variant transitions from an initiation complex to an elongation complex. The amino 35 acid substitution, in some embodiments, is a high propensity amino acid substitution. Examples

of high-helix propensity amino acids include alanine, isoleucine, leucine, arginine, methionine, lysine, glutamine, and/or glutamate.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an (at least one) amino acid substitution to introduce a polar, neutral amino acid. In some embodiments, a polar, neutral amino acid is selected from asparagine (N), cysteine (C), glutamine (Q), methionine (M), serine (S), and threonine (T). In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an (at least one) amino acid substitution to introduce an aromatic amino acid. In some embodiments, an aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an (at least one) amino acid substitution to introduce a non-polar, aliphatic amino acid. In some embodiments, a non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an (at least one) amino acid substitution to introduce a positively charged amino acid. In some embodiments, a positively charged amino acid is selected from lysine (K), arginine (R), and histidine (H). In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an (at least one) amino acid substitution to introduce a negatively charged amino acid. In some embodiments, a negatively charged amino acid is selected from aspartic acid (D) and glutamic acid (E).

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an (at least one) amino acid modification at a position that is not a conserved amino acid residue. Conserved amino acid residues are amino acids or amino acid types (*e.g.*, individual amino acids such as Gly or Ser, or groups of amino acids that share similar properties such as amino acids with acidic functional groups) that are generally shared across multiple homologous sequences of the same protein. Conserved amino acid residues can be identified using sequence alignments of homologous amino acid sequences. A sequence alignment of approximately 1000 RNA polymerase sequences obtained using a Basic Local Alignment search allowed for a determination of the 240 positions of SEQ ID NO: 1 that are most likely to be conserved across RNA polymerase sequences. These 240 positions of SEQ ID NO: 1 that are most likely to be conserved across RNA polymerase sequences are at positions 5-6, 39, 269-277, 279, 281-282, 323-333, 411-448, 454-470, 472-474, 497-516, 532-560, 562-573, 626-646, 691, 693-702, 724-738, 775-794, 805-820, 828-833, 865-867, and 877-879. Accordingly, in some embodiments, an RNA polymerase variant comprises an RNA polymerase that includes an (at least one) amino acid modification at a position that is not one of positions 5-6, 39, 269-277, 279, 281-282, 323-333, 411-448, 454-470, 472-474, 497-516, 532-560, 562-573, 626-646, 691, 693-702, 724-738,

775-794, 805-820, 828-833, 865-867, and 877-879 of SEQ ID NO: 1. In some embodiments, an RNA polymerase variant may further comprise any number of amino acid modifications at any number of positions that are not one of positions 5-6, 39, 269-277, 279, 281-282, 323-333, 411-448, 454-470, 472-474, 497-516, 532-560, 562-573, 626-646, 691, 693-702, 724-738, 775-794, 805-820, 828-833, 865-867, and 877-879 of SEQ ID NO: 1. In some embodiments, an RNA polymerase variant comprising an amino acid sequence of any one of SEQ ID NO: 2-9 may further comprise an (at least one) additional amino acid modification at a position that is not one of positions 5-6, 39, 269-277, 279, 281-282, 323-333, 411-448, 454-470, 472-474, 497-516, 532-560, 562-573, 626-646, 691, 693-702, 724-738, 775-794, 805-820, 828-833, 865-867, and 877-879. Conversely, the amino acid positions that are not conserved are most likely to be modified or mutated. Accordingly, in some embodiments, an RNA polymerase variant comprises an RNA polymerase that includes an (at least one) amino acid modification at positions 1-4, 7-38, 40-268, 278, 280, 283-322, 334-410, 449-453, 471, 475-496, 517-531, 561, 574-625, 647-690, 692, 703-723, 739-774, 795-804, 821-827, 834-864, 868-876, and 880-883. In some embodiments, an RNA polymerase variant comprising an amino acid sequence of any one of SEQ ID NO: 2-9 may further comprise an (at least one) additional amino acid modification at positions 1-4, 7-38, 40-268, 278, 280, 283-322, 334-410, 449-453, 471, 475-496, 517-531, 561, 574-625, 647-690, 692, 703-723, 739-774, 795-804, 821-827, 834-864, 868-876, and 880-883.

In some embodiments, an RNA polymerase variant comprising an amino acid sequence of any one of SEQ ID NO: 2-9 may further comprise an (at least one) amino acid modification at any amino acid position that does not disrupt the secondary or tertiary structure of the RNA polymerase protein. In some embodiments, an RNA polymerase variant comprising amino acid sequence of any one of SEQ ID NO: 2-9 may further comprise an (at least one) amino acid modification at any amino acid position that does not disrupt the ability of the RNA polymerase protein to fold. In some embodiments, an RNA polymerase variant comprising an amino acid sequence of any one of SEQ ID NO: 2-9 may further comprise an (at least one) amino acid modification at any amino acid position that does not disrupt the ability of the RNA polymerase protein to bind to nucleic acids (*e.g.*, DNA).

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position 437 (*e.g.*, N437Y), an amino acid substitution at position 387 (*e.g.*, K387S), an amino acid substitution at position 350 (*e.g.*, E350W), and an amino acid substitution at position 351 (*e.g.*, D351V), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an RNA polymerase variant comprises N437Y, K387S, E350W, and D351V

substitutions, relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position 437 (*e.g.*, N437Y), an amino acid substitution at position (*e.g.*, E350W), and an amino acid substitution at position 351 (*e.g.*, D351V), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an RNA polymerase variant comprises N437Y, E350W, and D351V substitutions, relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position 387 (*e.g.*, K387S), an amino acid substitution at position (*e.g.*, E350W), and an amino acid substitution at position 351 (*e.g.*, D351V) relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an RNA polymerase variant comprises K387S, E350W, and D351V substitutions, relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position 653 (*e.g.*, D653W), an amino acid substitution at position 350 (*e.g.*, E350W), and an amino acid substitution at position 351 (*e.g.*, D351V) relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an RNA polymerase variant comprises D653W, E350W, and D351V substitutions, relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1.

In some embodiments, an amino acid substitution at position K387 is a polar, neutral amino acid. In some embodiments, the polar, neutral amino acid is selected from asparagine (N), cysteine (C), glutamine (Q), methionine (M), serine (S), and threonine (T). Thus, in some embodiments, an amino acid substitution at position K387 is K387N, K387C, K387Q, K387M, K387S, or K387T.

In some embodiments, an amino acid substitution at position N437 is an aromatic amino acid. In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). Thus, in some embodiments, an amino acid substitution at position N437 is N437W, N437Y, or N437F.

In some embodiments, an amino acid substitution at position D653 is an aromatic amino acid. In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine

(Y), and phenylalanine (F). Thus, in some embodiments, an amino acid substitution at position D653 is D653W, D653Y, or D653F.

In some embodiments, an amino acid substitution at position E350 is an aromatic amino acid. In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). Thus, in some embodiments, an amino acid substitution at position E350 is E350W, E350Y, or E350F.

In some embodiments, an amino acid substitution at position D351 is a non-polar, aliphatic amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). Thus, in some embodiments, an amino acid substitution at position D351 is D351A, D351G, D351I, D351L, D351P, or D351V.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position R379 (*e.g.*, R379A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position R379 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, the charged amino acid is a positively charged amino acid (*e.g.*, lysine (K) or histidine (H)) or a negatively charged amino acid (*e.g.*, glutamic acid (E) or aspartic acid (D)). In some embodiments, an amino acid substitution at position R379 is R379A, R379K, R379E, or R379W.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position Y385 (*e.g.*, Y385A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position Y385 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the aromatic amino acid is selected from tryptophan (W) and phenylalanine (F). In some embodiments, the charged amino acid is a positively charged amino acid (*e.g.*, lysine (K), histidine (H), or arginine (R)) or a negatively charged amino acid (*e.g.*, glutamic acid (E) or aspartic acid (D)). In some embodiments, an amino acid substitution at position Y385 is Y385A, Y385K, Y385W, or Y385V.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position R386 (*e.g.*, R386A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position R386 is a non-
5 polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the charged amino acid is a positively charged amino acid (*e.g.*, lysine (K) or histidine (H)) or a negatively charged amino acid (*e.g.*, glutamic acid (E) or aspartic acid (D)).
10 In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, an amino acid substitution at position R386 is R386A, R386K, or R386Y.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position D388 (*e.g.*, D388A), relative to wild-type
15 RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position D388 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some
20 embodiments, the polar, neutral amino acid is selected from asparagine (N), cysteine (C), glutamine (Q), methionine (M), serine (S), and threonine (T). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, an amino acid substitution at position D388 is D388A, D388N, or D388Y.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence
25 that includes an amino acid substitution at position K389 (*e.g.*, K389A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position K389 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from
30 alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the polar, neutral amino acid is selected from asparagine (N), cysteine (C), glutamine (Q), methionine (M), serine (S), and threonine (T). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, an amino acid substitution at position K389 is K389A, K389S, or K389Y.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position R391 (*e.g.*, R391A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position R391 is a non-
5 polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, an amino acid substitution at position R391 is R391A.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence
10 that includes an amino acid substitution at position R394 (*e.g.*, R394A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position R394 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from
15 alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the polar, neutral amino acid is selected from asparagine (N), cysteine (C), glutamine (Q), methionine (M), serine (S), and threonine (T). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, an amino acid substitution at position R394 is R394A, R394Q, or R394Y.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence
20 that includes an amino acid substitution at position R395 (*e.g.*, R395A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position R395 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino
25 acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, an amino acid substitution at position R395 is R395A.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence
30 that includes an amino acid substitution at position D471 (*e.g.*, D471A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position D471 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino
35 acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the charged amino acid is a positively charged amino acid (*e.g.*, lysine (K)),

histidine (H), or arginine (R)) or a negatively charged amino acid (e.g., glutamic acid). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, an amino acid substitution at position D471 is D471A, D471E, or D471Y.

5 In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position R627 (e.g., R627A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position R627 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino
10 acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, an amino acid substitution at position R627 is R627A.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position R631 (e.g., R631A), relative to wild-type
15 RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position R631 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some
20 embodiments, an amino acid substitution at position R631 is R631A.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position R632 (e.g., R632A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position R632 is a non-
25 polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, the polar, neutral amino acid is selected from
30 asparagine (N), cysteine (C), glutamine (Q), methionine (M), serine (S), and threonine (T). In some embodiments, an amino acid substitution at position R632 is R632D, R632A, R632Q, or R632Y.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position G640 (e.g., G640A), relative to wild-type
35 RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of

SEQ ID NO: 1. In some embodiments, an amino acid substitution at position G640 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some
5 embodiments, an amino acid substitution at position G640 is G640A.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position G645 (*e.g.*, G645A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of
10 SEQ ID NO: 1. In some embodiments, an amino acid substitution at position G645 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some
embodiments, an amino acid substitution at position G645 is G645A.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence
15 that includes an amino acid substitution at position Q648 (*e.g.*, Q648A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position Q648 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from
20 alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, the charged amino acid is a positively charged amino acid (*e.g.*, lysine (K), histidine (H), or arginine (R)) or a negatively charged amino acid (*e.g.*, glutamic acid (E) or aspartic acid (D)). In some embodiments, an amino acid substitution at
25 position Q648 is Q648A, Q648R, Q648D, Q648E, or Q648Y.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position Q649 (*e.g.*, Q649A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of
30 SEQ ID NO: 1. In some embodiments, an amino acid substitution at position Q649 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some
embodiments, an amino acid substitution at position Q649 is Q649A.

In some embodiments, an RNA polymerase variant comprises an amino acid sequence
35 that includes an amino acid substitution at position E652 (*e.g.*, E652A), relative to wild-type

RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position E652 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some
5 embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, the charged amino acid is a positively charged amino acid (e.g., lysine (K), histidine (H), or arginine (R)). In some embodiments, an amino acid substitution at position E652 is E652R, E652A, or E652Y.

10 In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position D653 (*e.g.*, D653A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position D653 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino
15 acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, the charged amino acid is a positively charged amino acid (e.g., lysine (K), histidine (H), or arginine (R)). In some embodiments, an amino acid
20 substitution at position D653 is D653K, D653A, or D653Y.

 In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position Q656 (*e.g.*, Q656A), relative to wild-type RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of SEQ ID NO: 1. In some embodiments, an amino acid substitution at position Q656 is a non-
25 polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, the charged amino acid is a positively charged amino
30 acid (e.g., lysine (K), histidine (H), or arginine (R)) or a negatively charged amino acid (e.g., glutamic acid (E) or aspartic acid (D)). In some embodiments, an amino acid substitution at position Q656 is Q656K, Q656A, Q656E, or Q656Y.

 In some embodiments, an RNA polymerase variant comprises an amino acid sequence that includes an amino acid substitution at position P657 (*e.g.*, P657A), relative to wild-type
35 RNA polymerase, wherein the wild-type RNA polymerase comprises the amino acid sequence of

SEQ ID NO: 1. In some embodiments, an amino acid substitution at position P657 is a non-polar, aliphatic amino acid; polar, neutral amino acid; aromatic amino acid; or charged amino acid amino acid. In some embodiments, the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V). In some

5 embodiments, the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F). In some embodiments, the charged amino acid is a positively charged amino acid (e.g., lysine (K), histidine (H), or arginine (R)) or a negatively charged amino acid (e.g., glutamic acid (E) or aspartic acid (D)). In some embodiments, an amino acid substitution at position P657 is P657G, P657A, P657E, or P657Y.

10

Table 1. RNA Polymerase Variants

RNA Polymerase Variants	Amino Acid Sequence	SEQ ID NO
E350X ¹ D351X ² K387X ³ N437X ⁴	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRK MFERQLKAGEVADNAAAKPLITLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIKPEAVAYITIKTTLACLTSAADNTTVQAVASAI GRAIEDEARFGRIRDLEA KHFKKNVEEQLNKRVGHVYKKAFFMQVVEADMLSKGLLGGEAWSSWHKEDSIHV GVRRCIEMLIESTGMVSLHRQNAAGVVGQDSEITELAPEYAEAIATRAGALAGIS PMFQPCVVPPKPKWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPEVYK AINIAQNTAWKINKKVLAVANVITKWKHCPV X¹X² IPAIEREELPMKPEDIDM NPEALTAWKRAAAAVYR X³ DKARKSRRISLEFMLEQANKFANHKAIWFPYNMDW WRGRVYAVSMFNPQ X⁴ DMTKGLLTLAGKPIGKEGYWYWLKIHGANCAGVDKV PFPRIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGL SYNCSLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVN EILQADAINGTDNEVVTVTDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKR SVMTLAYGSKEFGFRQQVLEDITIQPAIDSGKGLMFTQPNQAAGYMAKLIWESV SVTVVAAVEAMNWLKSAKLLAAEVKDKKTGEILRKRCVHVWTPDGFVPWQE YKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHL RKTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLAD FYDQFADQLHESQLDKMPALPAKGNLNLRLDILESDFAF, wherein X ¹ is an aromatic amino acid, optionally selected from W, Y, and F; X ² is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; X ³ is a polar, neutral amino acid, optionally selected from N, C, Q, M, S, and T; and X ⁴ is an aromatic amino acid, optionally selected from W, Y, and F	2
E350X ¹ D351X ² N437X ⁴	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRK MFERQLKAGEVADNAAAKPLITLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIKPEAVAYITIKTTLACLTSAADNTTVQAVASAI GRAIEDEARFGRIRDLEA KHFKKNVEEQLNKRVGHVYKKAFFMQVVEADMLSKGLLGGEAWSSWHKEDSIHV GVRRCIEMLIESTGMVSLHRQNAAGVVGQDSEITELAPEYAEAIATRAGALAGIS PMFQPCVVPPKPKWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPEVYK AINIAQNTAWKINKKVLAVANVITKWKHCPV X¹X² IPAIEREELPMKPEDIDM NPEALTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKAIWFPYNMDW RGRVYAVSMFNPQ X⁴ DMTKGLLTLAGKPIGKEGYWYWLKIHGANCAGVDKVP FPERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLS YNCSLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVN EILQADAINGTDNEVVTVTDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRS VMTLAYGSKEFGFRQQVLEDITIQPAIDSGKGLMFTQPNQAAGYMAKLIWESVS VTVVAAVEAMNWLKSAKLLAAEVKDKKTGEILRKRCVHVWTPDGFVPWQEY KKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLR KTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLADF YDQFADQLHESQLDKMPALPAKGNLNLRLDILESDFAF, wherein X ¹ is an aromatic amino acid, optionally selected from W, Y, and F; X ² is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X ⁴ is an aromatic amino acid, optionally selected from W, Y, and F	3

RNA Polymerase Variants	Amino Acid Sequence	SEQ ID NO
E350X ¹ D351X ² K387X ³	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRK MFERQLKAGEVADNAAKPLITTLPKMIARINDWFEVKAARGKRP TAFQFL QEIKPEAVAYITIKTTLACLT SADNTTVQAVASAI GRAIEDEARFGRIRDLEA KHFKKNVEEQLNKRVRGHVYKKA FMQVVEADMLSKGLLGGEAWSSWHKEDSIHV GVRCIEMLIESTGMVSLHRQ NAGVVGQDSE TIELAPEYAEAIATRAGALAGIS PMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPVEVYK AINIAQNTAWKINKKVLAVANVITKWKHCPV X¹X² IPAIEREELPMKPEDIDM NPEALTAWKRAAAAVYR X³ DKARKSRRISLEFMLEQANKFANHKA IWFPYNMD WRGRVYAVSMFNPQGNDMTKGLLTLAKGKPIGKEGYWYWKI HGANCAGVDKVP FPERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLS YNC SLPLAFD GSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNE ILQADAINGTDNEVTVTDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRS VMTLAYGSKEFGFRQQVLED TIQPAIDSGKGLMFTQPNQAAGYMAKLIWESVS VTVVA AVEAMNWLKSAAKLLAAEVKDKKTGEILRKRC AVHWVTPDGFPVWQ EY KKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLR KTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVD TYESCDVLADF YDQFADQLHESQLDKMPALPAKGNLNL RDILESDF AFA, wherein X ¹ is an aromatic amino acid, optionally selected from W, Y, and F; X ² is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X ³ is a polar, neutral amino acid, optionally selected from N, C, Q, M, S, and T	4
E350X ¹ D351X ² D653X ⁵	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRK MFERQLKAGEVADNAAKPLITTLPKMIARINDWFEVKAARGKRP TAFQFL QEIKPEAVAYITIKTTLACLT SADNTTVQAVASAI GRAIEDEARFGRIRDLEA KHFKKNVEEQLNKRVRGHVYKKA FMQVVEADMLSKGLLGGEAWSSWHKEDSIHV GVRCIEMLIESTGMVSLHRQ NAGVVGQDSE TIELAPEYAEAIATRAGALAGIS PMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPVEVYK AINIAQNTAWKINKKVLAVANVITKWKHCPV X¹X² IPAIEREELPMKPEDIDM NPEALTAWKRAAAAVYRKDKARKSRRISLEFMLEQANKFANHKA IWFPYNMDW RGRVYAVSMFNPQGNDMTKGLLTLAKGKPIGKEGYWYWKI HGANCAGVDKVPF PERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSY NCSLPLAFD GSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEI LQADAINGTDNEVTVTDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRSV MTLAYGSKEFGFRQQVLE X⁵ TIQPAIDSGKGLMFTQPNQAAGYMAKLIWESVS VTVVA AVEAMNWLKSAAKLLAAEVKDKKTGEILRKRC AVHWVTPDGFPVWQ EY KKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLR KTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVD TYESCDVLADF YDQFADQLHESQLDKMPALPAKGNLNL RDILESDF AFA, wherein X ¹ is an aromatic amino acid, optionally selected from W, Y, and F; X ² is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X ⁵ is an aromatic amino acid, optionally selected from W, Y, and F	5
E350W D351V K387S N437Y	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRK MFERQLKAGEVADNAAKPLITTLPKMIARINDWFEVKAARGKRP TAFQFL QEIKPEAVAYITIKTTLACLT SADNTTVQAVASAI GRAIEDEARFGRIRDLEA KHFKKNVEEQLNKRVRGHVYKKA FMQVVEADMLSKGLLGGEAWSSWHKEDSIHV GVRCIEMLIESTGMVSLHRQ NAGVVGQDSE TIELAPEYAEAIATRAGALAGIS PMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPVEVYK AINIAQNTAWKINKKVLAVANVITKWKHCPV WV IPAIEREELPMKPEDIDMNP EALTAWKRAAAAVYR S DKARKSRRISLEFMLEQANKFANHKA IWFPYNMDWRG RVYAVSMFNPQGYDMTKGLLTLAKGKPIGKEGYWYWKI HGANCAGVDKVPFPE RIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNC SLPLAFD GSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEILQ ADAINGTDNEVTVTDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRSVMT LAYGSKEFGFRQQVLED TIQPAIDSGKGLMFTQPNQAAGYMAKLIWESVS VTV VA AVEAMNWLKSAAKLLAAEVKDKKTGEILRKRC AVHWVTPDGFPVWQ EYKKP IQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKTV VWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVD TYESCDVLADF YDQ FADQLHESQLDKMPALPAKGNLNL RDILESDF AFA	6
E350W D351V	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRK MFERQLKAGEVADNAAKPLITTLPKMIARINDWFEVKAARGKRP TAFQFL	7

RNA Polymerase Variants	Amino Acid Sequence	SEQ ID NO
N437Y	QEIKPEAVAYITIKTTLACLTSADNTTVQAVASAI GRAIEDEARFGRIRDLEA KHFKKNVEEQLNKRVGHVYKKA FMQVVEADMLS KGLLGGEAWSSWHKEDSIHV GVRRCIEMLIESTGMVSLHRQNAGVVGQDSE TI ELAPEYAEAIATRAGALAGIS PMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPEVYK AINIAQNTAWKINKKVLAVANVITKWKHCPV WV IPAIEREELPMKPEDIDMNP EALTAWKRAAAAVYRKDKARKSRRI SLEFMLEQANKFANHKA IWFPYNMDWRG RVYAVSMFNPQG Y DMTKGLLTLAKGKPIGKEGYWLKIHGANCAGVDKVPFPE RIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNC SLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEILQ ADAINGTDNEVVTVDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRSVMT LAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAGYMAKLIWESVSVTV VAAVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCVHWVTPDGFPVWQEQYKKP IQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKT VVAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLADFDYDQ FADQLHESQLDKMPALPAKGNLNLRLDILESDFafa	
E350W D351V K387S	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRK MFERQLKAGEVADNAAAKPLITLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIKPEAVAYITIKTTLACLTSADNTTVQAVASAI GRAIEDEARFGRIRDLEA KHFKKNVEEQLNKRVGHVYKKA FMQVVEADMLS KGLLGGEAWSSWHKEDSIHV GVRRCIEMLIESTGMVSLHRQNAGVVGQDSE TI ELAPEYAEAIATRAGALAGIS PMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPEVYK AINIAQNTAWKINKKVLAVANVITKWKHCPV WV IPAIEREELPMKPEDIDMNP EALTAWKRAAAAVYR SD KARKSRRI SLEFMLEQANKFANHKA IWFPYNMDWRG RVYAVSMFNPQGNMTKGLLTLAKGKPIGKEGYWLKIHGANCAGVDKVPFPE RIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNC SLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEILQ ADAINGTDNEVVTVDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRSVMT LAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAGYMAKLIWESVSVTV VAAVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCVHWVTPDGFPVWQEQYKKP IQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKT VVAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLADFDYDQ FADQLHESQLDKMPALPAKGNLNLRLDILESDFafa	8
E350W D351V D653W	MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRK MFERQLKAGEVADNAAAKPLITLLPKMIARINDWFEEVKAKRGKRPTAFQFL QEIKPEAVAYITIKTTLACLTSADNTTVQAVASAI GRAIEDEARFGRIRDLEA KHFKKNVEEQLNKRVGHVYKKA FMQVVEADMLS KGLLGGEAWSSWHKEDSIHV GVRRCIEMLIESTGMVSLHRQNAGVVGQDSE TI ELAPEYAEAIATRAGALAGIS PMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYEDVYMPEVYK AINIAQNTAWKINKKVLAVANVITKWKHCPV WV IPAIEREELPMKPEDIDMNP EALTAWKRAAAAVYRKDKARKSRRI SLEFMLEQANKFANHKA IWFPYNMDWRG RVYAVSMFNPQGNMTKGLLTLAKGKPIGKEGYWLKIHGANCAGVDKVPFPE RIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNC SLPLAFDGSCSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEILQ ADAINGTDNEVVTVDENTGEISEKVKLGTKALAGQWLAYGVTRSVTKRSVMT LAYGSKEFGFRQQVLE WT IQPAIDSGKGLMFTQPNQAAGYMAKLIWESVSVTV VAAVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCVHWVTPDGFPVWQEQYKKP IQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQDGSHLRKT VVAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLADFDYDQ FADQLHESQLDKMPALPAKGNLNLRLDILESDFafa	9

In some embodiments, an RNA polymerase variant further comprises one or more purification tags. For example, an RNA polymerase variant may comprise a histidine purification tag (*e.g.*, an amino acid sequence of -HHHHHH- (SEQ ID NO: 14)) or any other sequence of amino acids useful for purification. A histidine purification tag or similarly charged amino acid sequence is capable of binding to Ni²⁺ resin. In some embodiments, a histidine purification tag comprises the amino acid sequence of -HHHHHHV- (SEQ ID NO: 15). In some embodiments, a

purification tag is an N-terminal purification tag that is covalently attached to the N-terminus of an RNA polymerase variant. In some embodiments, a purification tag is a C-terminal purification tag that is covalently attached to the C-terminus of an RNA polymerase variant. In some embodiments, a protein purification tag is a FLAG tag (*e.g.*, an amino acid sequence of -
 5 DYKDDDK- (SEQ ID NO: 16)) or a hemagglutinin tag. In some embodiments, an RNA polymerase variant comprising an N-terminal His tag comprises any one of SEQ ID NOs: 10-13.

Table 2. RNA Polymerase Variants comprising an N-terminal His tag

RNA Polymerase Variants	Amino Acid Sequence	SEQ ID NO
E350W D351V K387S N437Y	MHHHHHHVNSNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESY EMGEARFRKMFERQLKAGEVADNAAAKPLITLLPKMIARINDWFEEVKAARG KRPTAFQFLQEIKEPAVAYITIKTTLACLTSADNTTVQAVASAI GRAI EDEAR FGRIRDLEAKHFKKNVEEQLNKRVGHVYKKAQFQVVEADMLSKGLLGGEAWSS WHKEDSIHVGVRCEIEMLIESTGMVSLHRQNAVVGQDSEI ELAPEYAEAIAT RAGALAGISPMFQPCVPPKPTGTITGGGYWANGRRPLALVRTHSKKALMRYE DVYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPV WV IPAIEREELPM KPEDIDMNPEALTAWKRAAAAVYR SD KARKSRRISLEFMLEQANKFANHKA IW FPYNMDWRGRVYAVSMFNPQG Y DMTKGLLTLAKGKPIGKEGYWLKIHGANCA GVDKVPFPERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGV QHHGLSYNCSLPLAFDGS CSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIV AKKVNEILQADAINGTDNEVVTVTDENTGEISEKVKLGTKALAGQWLAYGVTR SVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAGYMAKL IWESVSVTVVA AVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCVHVWTPDGF PVWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQ DGSHLRKTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESC DVLADFDYDQFADQLHESQLDKMPALPAKGNNLNRDILESDFAF	10
N437Y E350W D351V	MHHHHHHVNSNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESY EMGEARFRKMFERQLKAGEVADNAAAKPLITLLPKMIARINDWFEEVKAARG KRPTAFQFLQEIKEPAVAYITIKTTLACLTSADNTTVQAVASAI GRAI EDEAR FGRIRDLEAKHFKKNVEEQLNKRVGHVYKKAQFQVVEADMLSKGLLGGEAWSS WHKEDSIHVGVRCEIEMLIESTGMVSLHRQNAVVGQDSEI ELAPEYAEAIAT RAGALAGISPMFQPCVPPKPTGTITGGGYWANGRRPLALVRTHSKKALMRYE DVYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPV WV IPAIEREELPM KPEDIDMNPEALTAWKRAAAAVYR SD KARKSRRISLEFMLEQANKFANHKA IW FPYNMDWRGRVYAVSMFNPQG Y DMTKGLLTLAKGKPIGKEGYWLKIHGANCA GVDKVPFPERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGV QHHGLSYNCSLPLAFDGS CSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIV AKKVNEILQADAINGTDNEVVTVTDENTGEISEKVKLGTKALAGQWLAYGVTR SVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAGYMAKL IWESVSVTVVA AVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCVHVWTPDGF PVWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQ DGSHLRKTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESC DVLADFDYDQFADQLHESQLDKMPALPAKGNNLNRDILESDFAF	11
K387S E350W D351V	MHHHHHHVNSNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESY EMGEARFRKMFERQLKAGEVADNAAAKPLITLLPKMIARINDWFEEVKAARG KRPTAFQFLQEIKEPAVAYITIKTTLACLTSADNTTVQAVASAI GRAI EDEAR FGRIRDLEAKHFKKNVEEQLNKRVGHVYKKAQFQVVEADMLSKGLLGGEAWSS WHKEDSIHVGVRCEIEMLIESTGMVSLHRQNAVVGQDSEI ELAPEYAEAIAT RAGALAGISPMFQPCVPPKPTGTITGGGYWANGRRPLALVRTHSKKALMRYE DVYMPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPV WV IPAIEREELPM KPEDIDMNPEALTAWKRAAAAVYR SD KARKSRRISLEFMLEQANKFANHKA IW FPYNMDWRGRVYAVSMFNPQGNMTKGLLTLAKGKPIGKEGYWLKIHGANCA GVDKVPFPERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGV QHHGLSYNCSLPLAFDGS CSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIV AKKVNEILQADAINGTDNEVVTVTDENTGEISEKVKLGTKALAGQWLAYGVTR SVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAGYMAKL	12

RNA Polymerase Variants	Amino Acid Sequence	SEQ ID NO
	IWESVSVTVVAAVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCVHVWTPDGF PVWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQ DGSHLRKTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESC DVLADFYDQFADQLHESQLDKMPALPAKGNLNLRLDILESDFafa	
D653W E350W D351V	MHHHHHHVNSNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESY EMGEARFRKMFERQLKAGEVADNAAAKPLITTLPKMIARINDWFEEVKAKRG KRPTAFQFLQEIKPEAVAYITIKTTLACLTSAADNTTVQAVASAI GRAIEDEAR FGRIRDLEAKHFKKNVEEQLNKRVGHVYKKAQVVEADMLSKGLLGGEAWSS WHKEDSIHVGVRCEIEMLIESTGMVSLHRQNAAGVVGQDSEIIELAPEYAEAIAT RAGALAGISPMFQPCVVPPKPWTGITGGGYWANGRRPLALVRTHSKKALMRYE DVYMPPEVYKAINIAQNTAWKINKKVLAVANVITKWKHCPVWIPAIEREELPM KPEDIDMNPEALTAWKRAAAVYRKDKARKSRISLEFMLEQANKFANHKAIW FPYNMDWRGRVYAVSMFNPQGNMTKGLLTLAKGKPIGKEGYWLKIHGANCA GVDKVPFPERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGV QHHGLSYNCSLPLAFDGCSCGIQHFSAMLRDEVGGRVNLPLSETVQDIYGIV AKKVNEILQADAINGTDNEVVTVTDENTGEISEKVKLGTKALAGQWLAYGVTR SVTKRSVMTLAYGSKEFGFRQQVLEWTIQPAIDSGKGLMFTQPNQAAGYMAKL IWESVSVTVVAAVEAMNWLKSAAKLLAAEVKDKKTGEILRKRCVHVWTPDGF PVWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEIDAHKQESGIAPNFVHSQ DGSHLRKTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESC DVLADFYDQFADQLHESQLDKMPALPAKGNLNLRLDILESDFafa	13

In some embodiments, RNA polymerase variants have at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity with an RNA polymerase comprising the amino acid sequence of any one of SEQ ID NOs: 2-13. In some embodiments, RNA polymerase variants may share at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95% identity with an RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

The term “identity” refers to a relationship between the sequences of two or more polypeptides (*e.g.* enzymes) or polynucleotides (nucleic acids), as determined by comparing the sequences. Identity also refers to the degree of sequence relatedness between or among sequences as determined by the number of matches between strings of two or more amino acid residues or nucleic acid residues. Identity measures the percent of identical matches between the smaller of two or more sequences with gap alignments (if any) addressed by a particular mathematical model or computer program (*e.g.*, “algorithms”). Identity of related proteins or nucleic acids can be readily calculated by known methods. “Percent (%) identity” as it applies to polypeptide or polynucleotide sequences is defined as the percentage of residues (amino acid residues or nucleic acid residues) in the candidate amino acid or nucleic acid sequence that are identical with the residues in the amino acid sequence or nucleic acid sequence of a second sequence after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent identity. Methods and computer programs for the alignment are well known in the art. It is understood that identity depends on a calculation of percent identity but may differ in value due to gaps and penalties introduced in the calculation. Generally, variants of a particular

polynucleotide or polypeptide (e.g., antigen) have at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% but less than 100% sequence identity to that particular reference polynucleotide or polypeptide as determined by sequence alignment programs and parameters described herein and known to those skilled in the art. Such tools for alignment include those of the BLAST suite (Stephen F. Altschul, *et al* (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402). Another popular local alignment technique is based on the Smith-Waterman algorithm (Smith, T.F. & Waterman, M.S. (1981) "Identification of common molecular subsequences." *J. Mol. Biol.* 147:195-197). A general global alignment technique based on dynamic programming is the Needleman–Wunsch algorithm (Needleman, S.B. & Wunsch, C.D. (1970) "A general method applicable to the search for similarities in the amino acid sequences of two proteins." *J. Mol. Biol.* 48:443-453). More recently a Fast Optimal Global Sequence Alignment Algorithm (FOGSAA) has been developed that purportedly produces global alignment of nucleotide and protein sequences faster than other optimal global alignment methods, including the Needleman–Wunsch algorithm.

RNA Capping

Multivalent RNA compositions may comprise one or more mRNAs having open reading frames that encode proteins or peptides. Each of these mRNAs may have a 5' Cap. The 5' Cap may be added during the co-IVT reaction (e.g., transcriptional co-capping) or after the IVT reaction.

Some aspects also include a polynucleotide that comprises both a 5' Cap and a polynucleotide (e.g., a polynucleotide comprising a nucleotide sequence encoding a polypeptide to be expressed).

The 5' cap structure of a natural mRNA is involved in nuclear export, increasing mRNA stability and binds the mRNA Cap Binding Protein (CBP), for example eIF4E, which is responsible for mRNA stability in the cell and translation competency through the association of CBP with poly(A) binding protein to form the mature cyclic mRNA species. The cap further assists the removal of 5' proximal introns during mRNA splicing.

Endogenous mRNA molecules can be 5'-end capped generating a 5'-ppp-5'-triphosphate linkage between a terminal guanosine cap residue and the 5'-terminal transcribed sense nucleotide of the mRNA molecule. This 5'-guanylate cap can then be methylated to generate an N7-methyl-guanylate residue. The ribose sugars of the terminal and/or antiterminal transcribed nucleotides of the 5' end of the mRNA can optionally also be 2'-O-methylated. 5'-decapping

through hydrolysis and cleavage of the guanylate cap structure can target a nucleic acid molecule, such as an mRNA molecule, for degradation.

In some embodiments, the polynucleotides (e.g., a polynucleotide comprising a nucleotide sequence encoding a polypeptide) incorporate a cap moiety.

5 In some embodiments, polynucleotides comprise a non-hydrolyzable cap structure preventing decapping and thus increasing mRNA half-life. Because cap structure hydrolysis requires cleavage of 5'-ppp-5' phosphodiester linkages, modified nucleotides can be used during the capping reaction. For example, a Vaccinia Capping Enzyme from New England Biolabs (Ipswich, MA) can be used with α -thio-guanosine nucleotides according to the manufacturer's
10 instructions to create a phosphothioate linkage in the 5'-ppp-5' cap. Additional modified guanosine nucleotides can be used such as α -methyl-phosphonate and seleno-phosphate nucleotides.

Additional modifications include, but are not limited to, 2'-O-methylation of the ribose sugars of 5'-terminal and/or 5'-antiterminal nucleotides of the polynucleotide (as mentioned
15 above) on the 2'-hydroxyl group of the sugar ring. Multiple distinct 5'-cap structures can be used to generate the 5'-cap of a nucleic acid molecule, such as a polynucleotide that functions as an mRNA molecule. Cap analogs, which herein are also referred to as synthetic cap analogs, chemical caps, chemical cap analogs, or structural or functional cap analogs, differ from natural (*i.e.*, endogenous, wild-type or physiological) 5'-caps in their chemical structure, while retaining
20 cap function. Cap analogs can be chemically (*i.e.*, non-enzymatically) or enzymatically synthesized and/or linked to the polynucleotides.

For example, the Anti-Reverse Cap Analog (ARCA) cap contains two guanines linked by a 5'-5'-triphosphate group, wherein one guanine contains an N7 methyl group as well as a 3'-O-methyl group (*i.e.*, N7,3'-O-dimethyl-guanosine-5'-triphosphate-5'-guanosine (m^7G -3'mppp-G;
25 which can equivalently be designated 3' O-Me- $m^7G(5')ppp(5')G$). The 3'-O atom of the other, unmodified, guanine becomes linked to the 5'-terminal nucleotide of the capped polynucleotide. The N7- and 3'-O-methylated guanine provides the terminal moiety of the capped polynucleotide.

Another exemplary cap is mCAP, which is similar to ARCA but has a 2'-O-methyl group
30 on guanosine (*i.e.*, N7,2'-O-dimethyl-guanosine-5'-triphosphate-5'-guanosine, m^7Gm -ppp-G).

Another exemplary cap is m^7G -ppp-Gm-AG (*i.e.*, N7,guanosine-5'-triphosphate-2'-O-dimethyl-guanosine-adenosine-guanosine).

In some embodiments, the cap is a dinucleotide cap analog. As a non-limiting example, the dinucleotide cap analog can be modified at different phosphate positions with a
35 boranophosphate group or a phosphoroselenoate group such as the dinucleotide cap analogs

described in U.S. Patent No. US 8,519,110, the contents of which are herein incorporated by reference in its entirety.

In another embodiment, the cap is a cap analog is a N7-(4-chlorophenoxyethyl) substituted dinucleotide form of a cap analog known in the art and/or described herein. Non-limiting examples of a N7-(4-chlorophenoxyethyl) substituted dinucleotide form of a cap analog include a N7-(4-chlorophenoxyethyl)-G(5')ppp(5')G and a N7-(4-chlorophenoxyethyl)-m^{3'}-^oG(5')ppp(5')G cap analog (*See, e.g.*, the various cap analogs and the methods of synthesizing cap analogs described in Kore et al. *Bioorganic & Medicinal Chemistry* 2013 21:4570-4574; the contents of which are herein incorporated by reference in its entirety). In another embodiment, a cap analog is a 4-chloro/bromophenoxyethyl analog.

Polynucleotides can also be capped post-manufacture (whether IVT or chemical synthesis), using enzymes, in order to generate more authentic 5'-cap structures. As used herein, the phrase "more authentic" refers to a feature that closely mirrors or mimics, either structurally or functionally, an endogenous or wild type feature. That is, a "more authentic" feature is better representative of an endogenous, wild-type, natural or physiological cellular function and/or structure as compared to synthetic features or analogs, etc., of the prior art, or which outperforms the corresponding endogenous, wild-type, natural or physiological feature in one or more respects. Non-limiting examples of more authentic 5'cap structures are those that, among other things, have enhanced binding of cap binding proteins, increased half-life, reduced susceptibility to 5' endonucleases and/or reduced 5'decapping, as compared to synthetic 5'cap structures known in the art (or to a wild-type, natural or physiological 5'cap structure). For example, recombinant Vaccinia Virus Capping Enzyme and recombinant 2'-O-methyltransferase enzyme can create a canonical 5'-5'-triphosphate linkage between the 5'-terminal nucleotide of a polynucleotide and a guanine cap nucleotide wherein the cap guanine contains an N7 methylation and the 5'-terminal nucleotide of the mRNA contains a 2'-O-methyl. Such a structure is termed the Cap1 structure. This cap results in a higher translational-competency and cellular stability and a reduced activation of cellular pro-inflammatory cytokines, as compared, *e.g.*, to other 5'cap analog structures known in the art. Cap structures include, but are not limited to, 7mG(5')ppp(5')N,pN2p (cap 0), 7mG(5')ppp(5')NlmpNp (cap 1), and 7mG(5')-ppp(5')NlmpN2mp (cap 2).

As a non-limiting example, capping chimeric polynucleotides post-manufacture can be more efficient as nearly 100% of the chimeric polynucleotides can be capped. This is in contrast to ~80% when a cap analog is linked to a chimeric polynucleotide in the course of an *in vitro* transcription reaction.

In some embodiments, 5' terminal caps can include endogenous caps or cap analogs. In some embodiments, a 5' terminal cap can comprise a guanine analog. Useful guanine analogs

include, but are not limited to, inosine, N1-methyl-guanosine, 2'fluoro-guanosine, 7-deaza-guanosine, 8-oxo-guanosine, 2-amino-guanosine, LNA-guanosine, and 2-azido-guanosine.

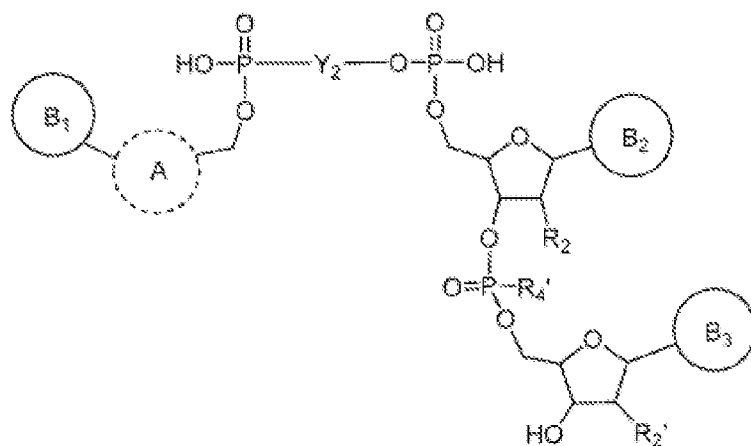
Also described are exemplary caps including those that can be used in co-transcriptional capping methods for ribonucleic acid (RNA) synthesis, using RNA polymerase, e.g., wild type RNA polymerase or variants thereof, e.g., such as those variants described. In one embodiment, caps can be added when RNA is produced in a “one-pot” reaction, without the need for a separate capping reaction. Thus, the methods, in some embodiments, comprise reacting a polynucleotide template with a RNA polymerase variant, nucleoside triphosphates, and a cap analog under *in vitro* transcription reaction conditions to produce RNA transcript.

In some embodiments, the cap analog binds to a polynucleotide template that comprises a promoter region comprising a transcriptional start site having a first nucleotide at nucleotide position +1, a second nucleotide at nucleotide position +2, and a third nucleotide at nucleotide position +3. In some embodiments, the cap analog hybridizes to the polynucleotide template at least at nucleotide position +1, such as at the +1 and +2 positions, or at the +1, +2, and +3 positions.

A cap analog may be, for example, a dinucleotide cap, a trinucleotide cap, or a tetranucleotide cap. In some embodiments, a cap analog is a dinucleotide cap. In some embodiments, a cap analog is a trinucleotide cap. In some embodiments, a cap analog is a tetranucleotide cap. As used here the term “cap” includes the inverted G nucleotide and can comprise additional nucleotides 3' of the inverted G, i.e., 1, 2, or more nucleotides 3' of the inverted G and 5' to the 5' UTR.

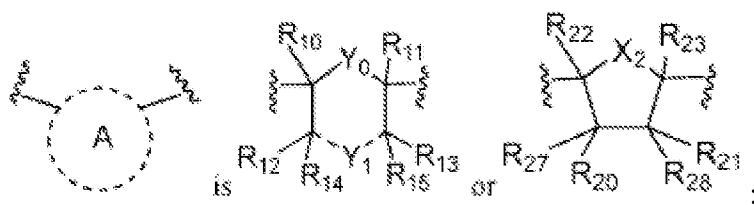
Exemplary caps comprise a sequence GG, GA, or GGA wherein the underlined, italicized G is an inverted G.

A nucleotide cap (e.g., a trinucleotide cap or tetranucleotide cap), in some embodiments, comprises a compound of formula (I)



(I), or a stereoisomer, tautomer or

salt thereof, wherein



ring B₁ is a modified or unmodified Guanine;

ring B₂ and ring B₃ each independently is a nucleobase or a modified nucleobase;

5 X₂ is O, S(O)_p, NR₂₄ or CR₂₅R₂₆ in which p is 0, 1, or 2;

Y₀ is O or CR₆R₇;

Y₁ is O, S(O)_n, CR₆R₇, or NR₈, in which n is 0, 1, or 2;

each --- is a single bond or absent, wherein when each --- is a single bond, Y_i is O, S(O)_n, CR₆R₇, or NR₈; and when each --- is absent, Y₁ is void;

10 Y₂ is (OP(O)R₄)_m in which m is 0, 1, or 2, or -O-(CR₄₀R₄₁)_u-Q₀-(CR₄₂R₄₃)_v-, in which Q₀ is a bond, O, S(O)_r, NR₄₄, or CR₄₅R₄₆, r is 0, 1, or 2, and each of u and v independently is 1, 2, 3 or 4;

each R₂ and R₂' independently is halo, LNA, or OR₃;

each R₃ independently is H, C₁-C₆ alkyl, C₂-C₆ alkenyl, or C₂-C₆ alkynyl and R₃, when
15 being C₁-C₆ alkyl, C₂-C₆ alkenyl, or C₂-C₆ alkynyl, is optionally substituted with one or more of halo, OH and C₁-C₆ alkoxy that is optionally substituted with one or more OH or OC(O)-C₁-C₆ alkyl;

each R₄ and R₄' independently is H, halo, C₁-C₆ alkyl, OH, SH, SeH, or BH₃⁻;

each of R₆, R₇, and R₈, independently, is -Q₁-T₁, in which Q₁ is a bond or C₁-C₃ alkyl
20 linker optionally substituted with one or more of halo, cyano, OH and C₁-C₆ alkoxy, and T₁ is H, halo, OH, COOH, cyano, or R_{s1}, in which R_{s1} is C₁-C₃ alkyl, C₂-C₆ alkenyl, C₂-C₆ alkynyl, C₁-C₆ alkoxy, C(O)O-C₁-C₆ alkyl, C₃-C₈ cycloalkyl, C₆-C₁₀ aryl, NR₃₁R₃₂, (NR₃₁R₃₂R₃₃)⁺, 4 to 12-membered heterocycloalkyl, or 5- or 6-membered heteroaryl, and R_{s1} is optionally substituted with one or more substituents selected from the group consisting of halo, OH, oxo, C₁-C₆ alkyl,
25 COOH, C(O)O-C₁-C₆ alkyl, cyano, C₁-C₆ alkoxy, NR₃₁R₃₂, (NR₃₁R₃₂R₃₃)⁺, C₃-C₈ cycloalkyl, C₆-C₁₀ aryl, 4 to 12-membered heterocycloalkyl, and 5- or 6-membered heteroaryl;

each of R₁₀, R₁₁, R₁₂, R₁₃, R₁₄, and R₁₅, independently, is -Q₂-T₂, in which Q₂ is a bond or C₁-C₃ alkyl linker optionally substituted with one or more of halo, cyano, OH and C₁-C₆ alkoxy, and T₂ is H, halo, OH, NH₂, cyano, NO₂, N₃, R_{s2}, or OR_{s2}, in which R_{s2} is C₁-C₆ alkyl, C₂-C₆
30 alkenyl, C₂-C₆ alkynyl, C₃-C₈ cycloalkyl, C₆-C₁₀ aryl, NHC(O)-C₁-C₆ alkyl, NR₃₁R₃₂, (NR₃₁R₃₂R₃₃)⁺, 4 to 12-membered heterocycloalkyl, or 5- or 6-membered

heteroaryl, and R_{S2} is optionally substituted with one or more substituents selected from the group consisting of halo, OH, oxo, C_1 - C_6 alkyl, $COOH$, $C(O)O$ - C_1 - C_6 alkyl, cyano, C_1 - C_6 alkoxy, $NR_{31}R_{32}$, $(NR_{31}R_{32}R_{33})^+$, C_3 - C_8 cycloalkyl, C_6 - C_{10} aryl, 4 to 12-membered heterocycloalkyl, and 5- or 6- membered heteroaryl; or alternatively R_{12} together with R_{14} is oxo, or R_{13} together with R_{15} is oxo,

each of R_{20} , R_{21} , R_{22} , and R_{23} independently is $-Q_3-T_3$, in which Q_3 is a bond or C_1 - C_3 alkyl linker optionally substituted with one or more of halo, cyano, OH and C_1 - C_6 alkoxy, and T_3 is H, halo, OH, NH_2 , cyano, NO_2 , N_3 , R_{S3} , or OR_{S3} , in which R_{S3} is C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_2 - C_6 alkynyl, C_3 - C_8 cycloalkyl, C_6 - C_{10} aryl, $NHC(O)$ - C_1 - C_6 alkyl, mono- C_1 - C_6 alkylamino, di- C_1 - C_6 alkylamino, 4 to 12-membered heterocycloalkyl, or 5- or 6-membered heteroaryl, and R_{S3} is optionally substituted with one or more substituents selected from the group consisting of halo, OH, oxo, C_1 - C_6 alkyl, $COOH$, $C(O)O$ - C_1 - C_6 alkyl, cyano, C_1 - C_6 alkoxy, amino, mono- C_1 - C_6 alkylamino, di- C_1 - C_6 alkylamino, C_3 - C_8 cycloalkyl, C_6 - C_{10} aryl, 4 to 12-membered heterocycloalkyl, and 5- or 6-membered heteroaryl;

each of R_{24} , R_{25} , and R_{26} independently is H or C_1 - C_6 alkyl;

each of R_{27} and R_{28} independently is H or OR_{29} ; or R_{27} and R_{28} together form $O-R_{30}-O$; each R_{29} independently is H, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, or C_2 - C_6 alkynyl and R_{29} , when being C_1 - C_6 alkyl, C_2 - C_6 alkenyl, or C_2 - C_6 alkynyl, is optionally substituted with one or more of halo, OH and C_1 - C_6 alkoxy that is optionally substituted with one or more OH or $OC(O)$ - C_1 - C_6 alkyl;

R_{30} is C_1 - C_6 alkylene optionally substituted with one or more of halo, OH and C_1 - C_6 alkoxy;

each of R_{31} , R_{32} , and R_{33} , independently is H, C_1 - C_6 alkyl, C_3 - C_8 cycloalkyl, C_6 - C_{10} aryl, 4 to 12-membered heterocycloalkyl, or 5- or 6-membered heteroaryl;

each of R_{40} , R_{41} , R_{42} , and R_{43} independently is H, halo, OH, cyano, N_3 , $OP(O)R_{47}R_{48}$, or C_1 - C_6 alkyl optionally substituted with one or more $OP(O)R_{47}R_{48}$, or one R_{41} and one R_{43} , together with the carbon atoms to which they are attached and Q_0 , form C_4 - C_{10} cycloalkyl, 4- to 14-membered heterocycloalkyl, C_6 - C_{10} aryl, or 5- to 14-membered heteroaryl, and each of the cycloalkyl, heterocycloalkyl, phenyl, or 5- to 6-membered heteroaryl is optionally substituted with one or more of OH, halo, cyano, N_3 , oxo, $OP(O)R_{47}R_{48}$, C_1 - C_6 alkyl, C_1 - C_6 haloalkyl, $COOH$, $C(O)O$ - C_1 - C_6 alkyl, C_1 - C_6 alkoxy, C_1 - C_6 haloalkoxy, amino, mono- C_1 - C_6 alkylamino, and di- C_1 - C_6 alkylamino;

R_{44} is H, C_1 - C_6 alkyl, or an amine protecting group;

each of R_{45} and R_{46} independently is H, $OP(O)R_{47}R_{48}$, or C_1 - C_6 alkyl optionally substituted with one or more $OP(O)R_{47}R_{48}$, and

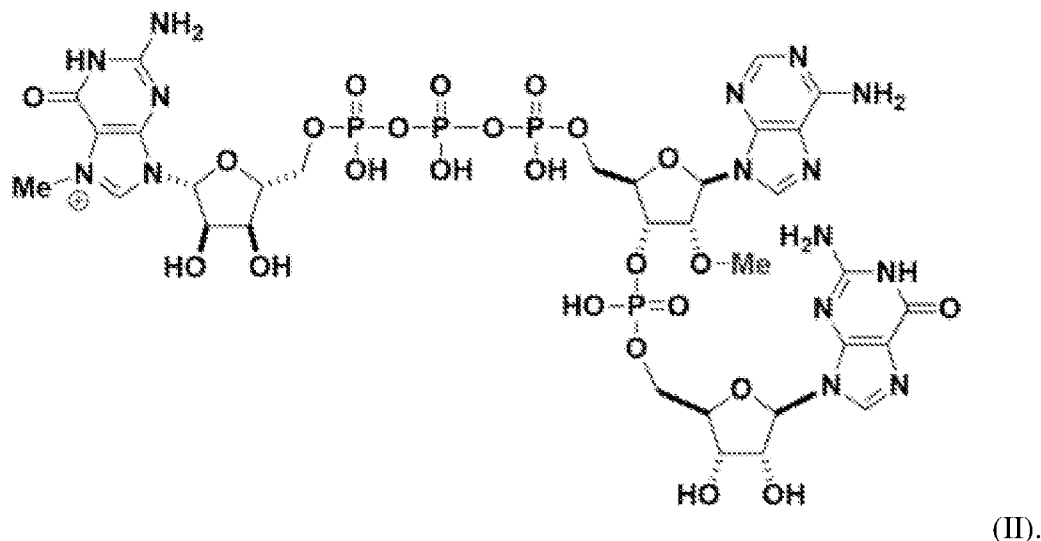
each of R_{47} and R_{48} , independently is H, halo, C_1 - C_6 alkyl, OH, SH, SeH, or BH_3^- .

It should be understood that a cap analog may include any of the cap analogs described in international publication WO 2017/066797, published on 20 April 2017, incorporated by reference herein in its entirety.

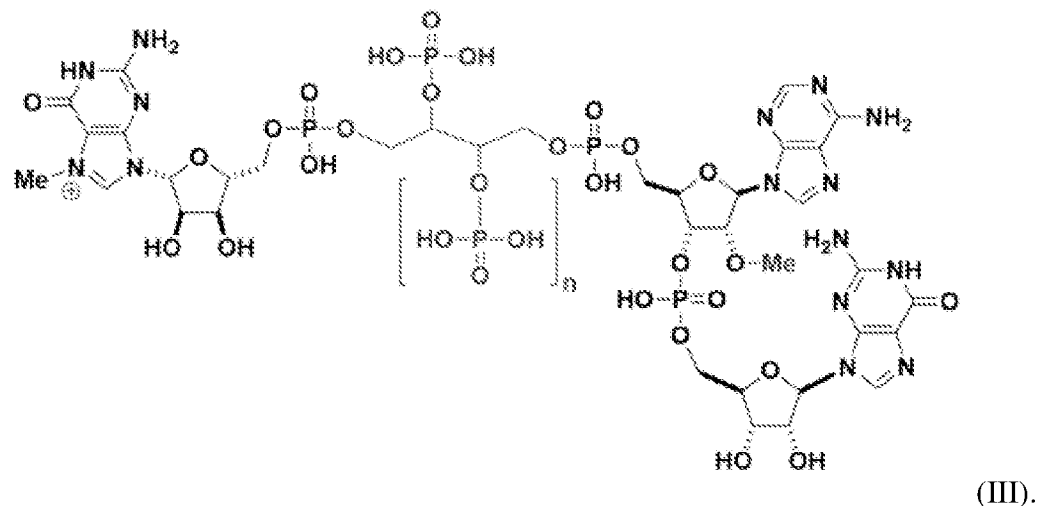
In some embodiments, the B₂ middle position can be a non-ribose molecule, such as
5 arabinose.

In some embodiments R₂ is ethyl-based.

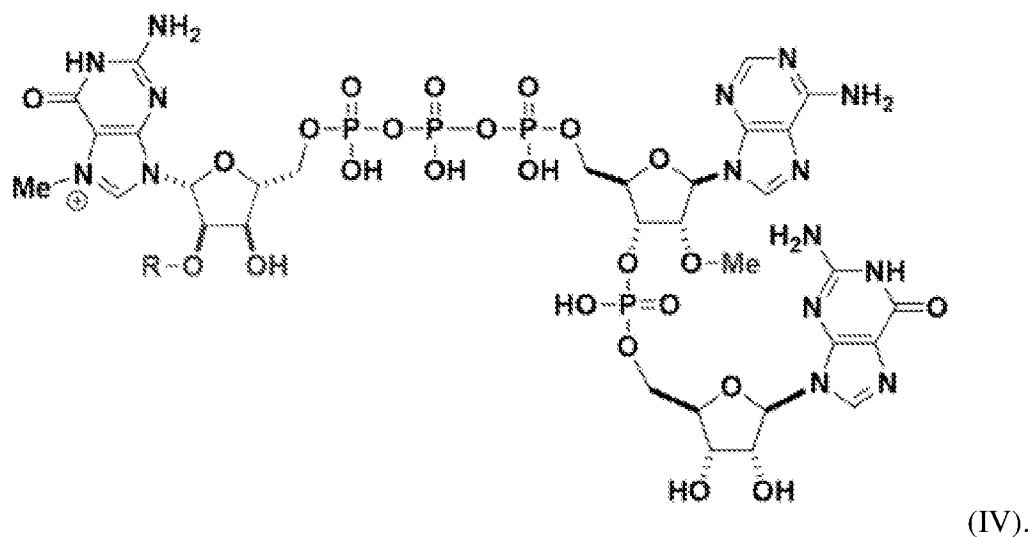
Thus, in some embodiments, a trinucleotide cap comprises the following structure:



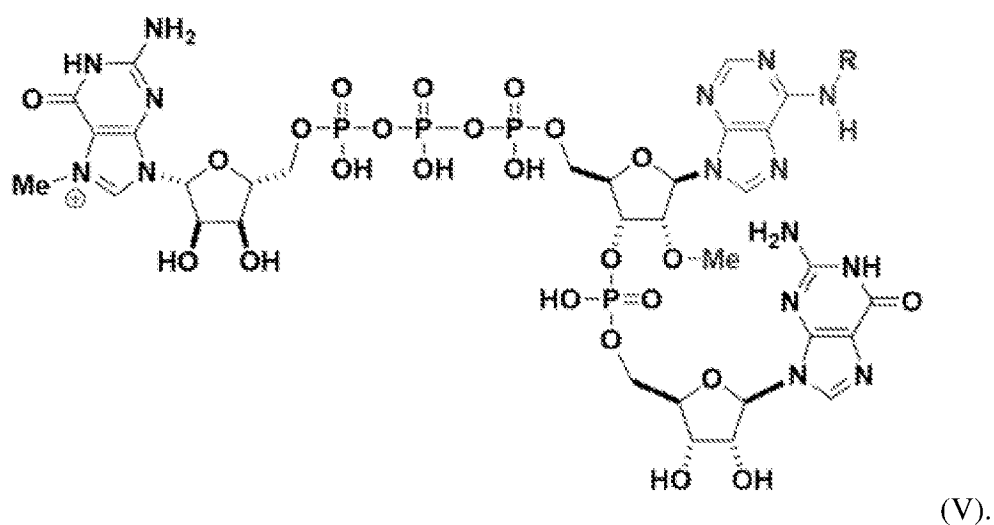
In other embodiments, a trinucleotide cap comprises the following structure:



In yet other embodiments, a trinucleotide cap comprises the following structure:

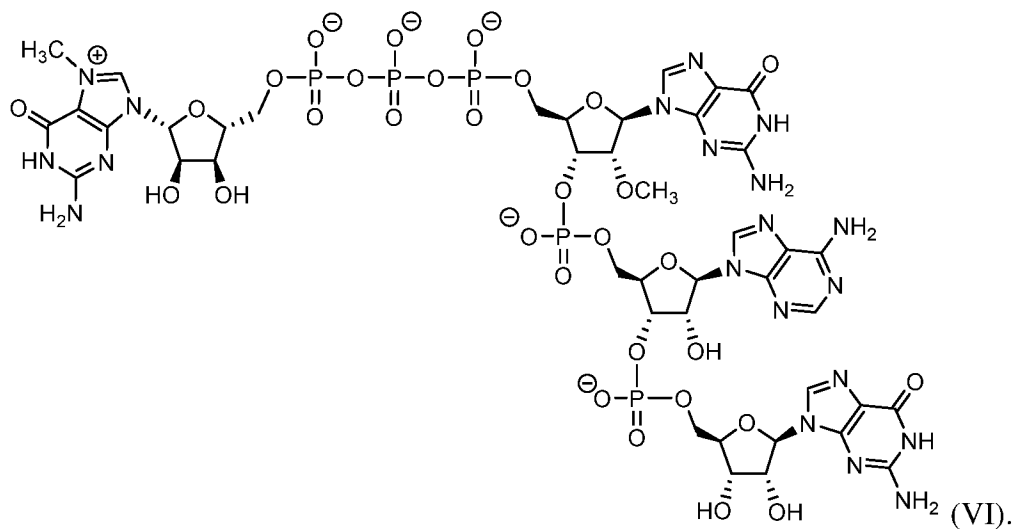


In still other embodiments, a trinucleotide cap comprises the following structure:

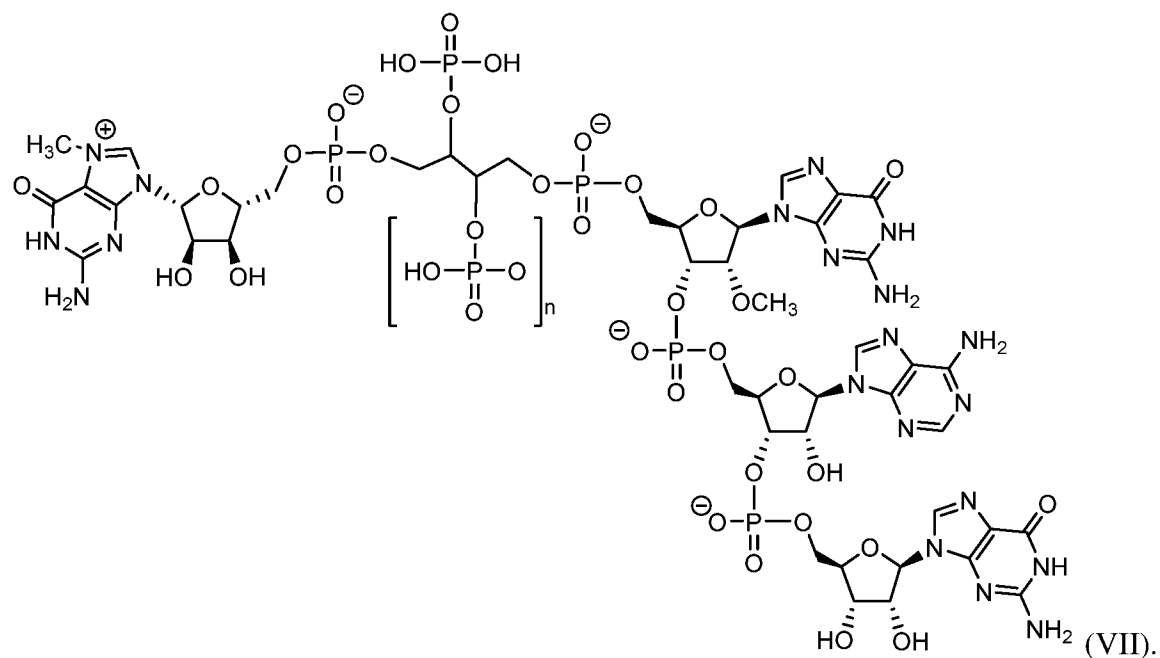


5

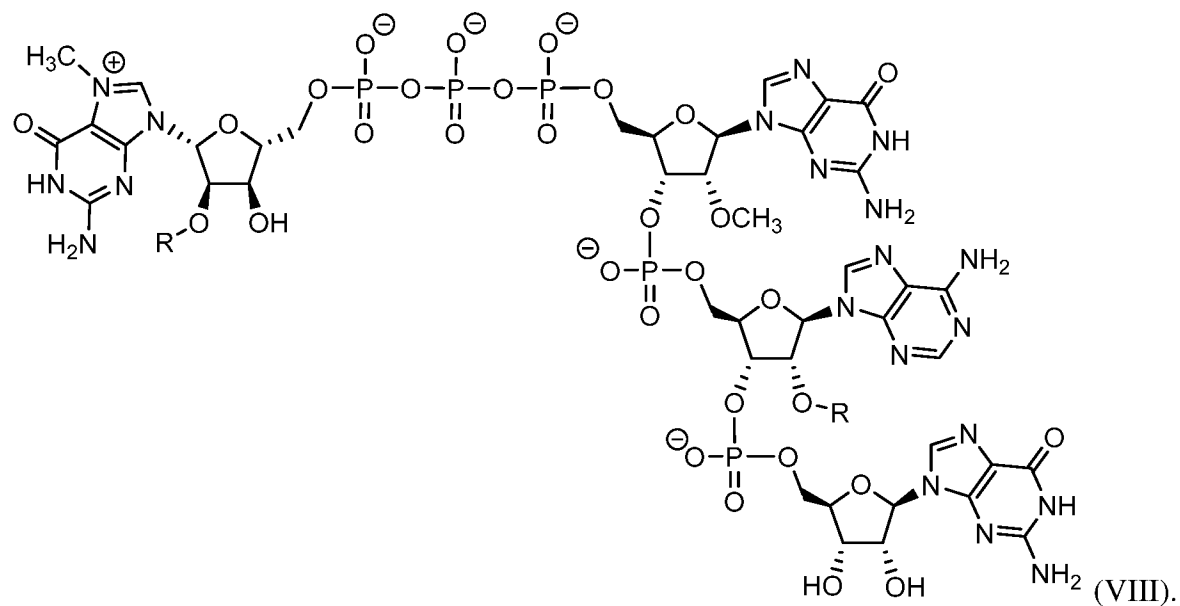
Thus, in some embodiments, a tetranucleotide cap comprises the following structure:



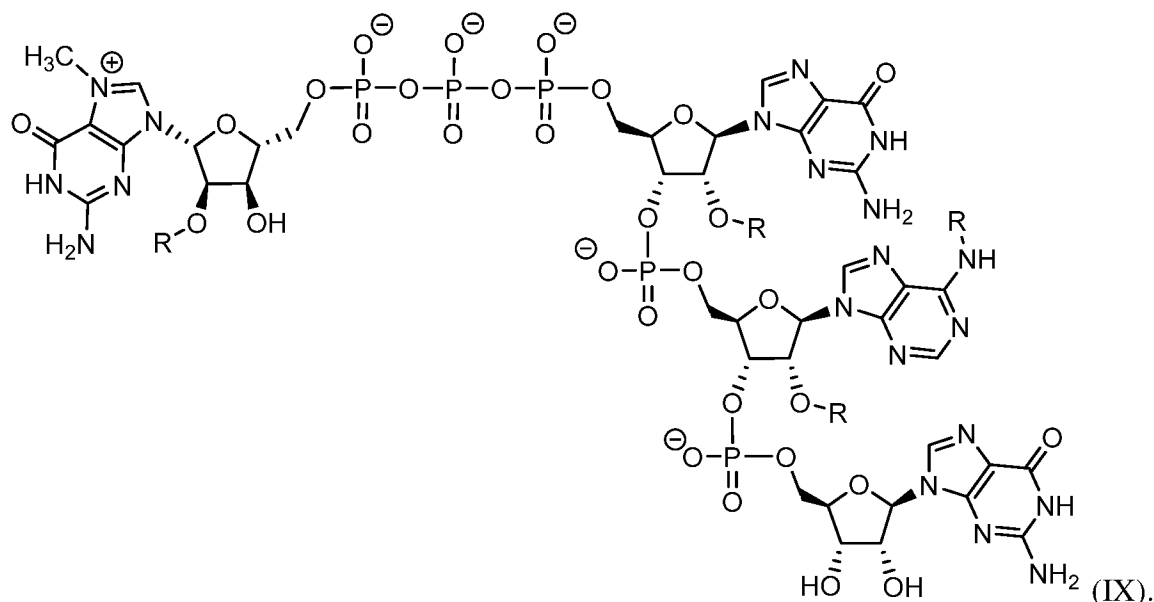
In other embodiments, a tetranucleotide cap comprises the following structure:



In yet other embodiments, a tetranucleotide cap comprises the following structure:



In yet other embodiments, a tetranucleotide cap comprises the following structure:



In some embodiments, R is an alkyl (*e.g.*, C₁-C₆ alkyl). In some embodiments, R is a methyl group (*e.g.*, C₁ alkyl). In some embodiments, R is an ethyl group (*e.g.*, C₂ alkyl). In some embodiments, R is a hydrogen.

5 A trinucleotide cap, in some embodiments, comprises a sequence selected from the following sequences: GAA, GAC, GAG, GAU, GCA, GCC, GCG, GCU, GGA, GGC, GGG, GGU, GUA, GUC, GUG, and GUU. In some embodiments, a trinucleotide cap comprises GAA. In some embodiments, a trinucleotide cap comprises GAC. In some embodiments, a trinucleotide cap comprises GAG. In some embodiments, a trinucleotide cap comprises GAU. In some
10 embodiments, a trinucleotide cap comprises GCA. In some embodiments, a trinucleotide cap comprises GCC. In some embodiments, a trinucleotide cap comprises GCG. In some embodiments, a trinucleotide cap comprises GCU. In some embodiments, a trinucleotide cap comprises GGA. In some embodiments, a trinucleotide cap comprises GGC. In some embodiments, a trinucleotide cap comprises GGG. In some embodiments, a trinucleotide cap comprises GGU. In some
15 embodiments, a trinucleotide cap comprises GUA. In some embodiments, a trinucleotide cap comprises GUC. In some embodiments, a trinucleotide cap comprises GUG. In some embodiments, a trinucleotide cap comprises GUU.

In some embodiments, a trinucleotide cap comprises a sequence selected from the following sequences: m⁷GpppApA, m⁷GpppApC, m⁷GpppApG, m⁷GpppApU, m⁷GpppCpA,
20 m⁷GpppCpC, m⁷GpppCpG, m⁷GpppCpU, m⁷GpppGpA, m⁷GpppGpC, m⁷GpppGpG, m⁷GpppGpU, m⁷GpppUpA, m⁷GpppUpC, m⁷GpppUpG, and m⁷GpppUpU.

In some embodiments, a trinucleotide cap comprises m⁷GpppApA. In some embodiments, a trinucleotide cap comprises m⁷GpppApC. In some embodiments, a trinucleotide cap comprises m⁷GpppApG. In some embodiments, a trinucleotide cap comprises m⁷GpppApU.
25 In some embodiments, a trinucleotide cap comprises m⁷GpppCpA. In some embodiments, a

trinucleotide cap comprises $m^7GpppCpC$. In some embodiments, a trinucleotide cap comprises $m^7GpppCpG$. In some embodiments, a trinucleotide cap comprises $m^7GpppCpU$. In some embodiments, a trinucleotide cap comprises $m^7GpppGpA$. In some embodiments, a trinucleotide cap comprises $m^7GpppGpC$. In some embodiments, a trinucleotide cap comprises $m^7GpppGpG$.

5 In some embodiments, a trinucleotide cap comprises $m^7GpppGpU$. In some embodiments, a trinucleotide cap comprises $m^7GpppUpA$. In some embodiments, a trinucleotide cap comprises $m^7GpppUpC$. In some embodiments, a trinucleotide cap comprises $m^7GpppUpG$. In some embodiments, a trinucleotide cap comprises $m^7GpppUpU$.

A trinucleotide cap, in some embodiments, comprises a sequence selected from the
10 following sequences: $m^7G_3'OMepppApA$, $m^7G_3'OMepppApC$, $m^7G_3'OMepppApG$,
 $m^7G_3'OMepppApU$, $m^7G_3'OMepppCpA$, $m^7G_3'OMepppCpC$, $m^7G_3'OMepppCpG$, $m^7G_3'OMepppCpU$,
 $m^7G_3'OMepppGpA$, $m^7G_3'OMepppGpC$, $m^7G_3'OMepppGpG$, $m^7G_3'OMepppGpU$, $m^7G_3'OMepppUpA$,
 $m^7G_3'OMepppUpC$, $m^7G_3'OMepppUpG$, and $m^7G_3'OMepppUpU$.

In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppApA$. In some
15 embodiments, a trinucleotide cap comprises $m^7G_3'OMepppApC$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppApG$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppApU$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppCpA$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppCpC$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppCpG$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppCpU$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppGpA$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppGpC$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppGpG$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppGpU$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppUpA$. In some embodiments, a trinucleotide cap
20 comprises $m^7G_3'OMepppUpC$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppUpG$. In some embodiments, a trinucleotide cap comprises $m^7G_3'OMepppUpU$.

A trinucleotide cap, in other embodiments, comprises a sequence selected from the following sequences: $m^7G_3'OMepppA_2'OMepA$, $m^7G_3'OMepppA_2'OMepC$, $m^7G_3'OMepppA_2'OMepG$,
 $m^7G_3'OMepppA_2'OMepU$, $m^7G_3'OMepppC_2'OMepA$, $m^7G_3'OMepppC_2'OMepC$, $m^7G_3'OMepppC_2'OMepG$,
30 $m^7G_3'OMepppC_2'OMepU$, $m^7G_3'OMepppG_2'OMepA$, $m^7G_3'OMepppG_2'OMepC$, $m^7G_3'OMepppG_2'OMepG$,
 $m^7G_3'OMepppG_2'OMepU$, $m^7G_3'OMepppU_2'OMepA$, $m^7G_3'OMepppU_2'OMepC$, $m^7G_3'OMepppU_2'OMepG$,
and $m^7G_3'OMepppU_2'OMepU$.

In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppA_{2'OMe}pA$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppA_{2'OMe}pC$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppA_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppA_{2'OMe}pU$. In some embodiments, a trinucleotide cap comprises

5 $m^7G_{3'OMe}pppC_{2'OMe}pA$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppC_{2'OMe}pC$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppC_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppC_{2'OMe}pU$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppG_{2'OMe}pA$. In some embodiments, a trinucleotide cap comprises

10 $m^7G_{3'OMe}pppG_{2'OMe}pC$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppG_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppG_{2'OMe}pU$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppU_{2'OMe}pA$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppU_{2'OMe}pC$. In some embodiments, a trinucleotide cap comprises

15 $m^7G_{3'OMe}pppU_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7G_{3'OMe}pppU_{2'OMe}pU$.

A trinucleotide cap, in still other embodiments, comprises a sequence selected from the following sequences: $m^7GpppA_{2'OMe}pA$, $m^7GpppA_{2'OMe}pC$, $m^7GpppA_{2'OMe}pG$, $m^7GpppA_{2'OMe}pU$, $m^7GpppC_{2'OMe}pA$, $m^7GpppC_{2'OMe}pC$, $m^7GpppC_{2'OMe}pG$, $m^7GpppC_{2'OMe}pU$,

20 $m^7GpppG_{2'OMe}pA$, $m^7GpppG_{2'OMe}pC$, $m^7GpppG_{2'OMe}pG$, $m^7GpppG_{2'OMe}pU$, $m^7GpppU_{2'OMe}pA$, $m^7GpppU_{2'OMe}pC$, $m^7GpppU_{2'OMe}pG$, and $m^7GpppU_{2'OMe}pU$.

In some embodiments, a trinucleotide cap comprises $m^7GpppA_{2'OMe}pA$. In some embodiments, a trinucleotide cap comprises $m^7GpppA_{2'OMe}pC$. In some embodiments, a trinucleotide cap comprises $m^7GpppA_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7GpppA_{2'OMe}pU$. In some embodiments, a trinucleotide cap comprises

25 $m^7GpppC_{2'OMe}pA$. In some embodiments, a trinucleotide cap comprises $m^7GpppC_{2'OMe}pC$. In some embodiments, a trinucleotide cap comprises $m^7GpppC_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7GpppC_{2'OMe}pU$. In some embodiments, a trinucleotide cap comprises $m^7GpppG_{2'OMe}pA$. In some embodiments, a trinucleotide cap comprises

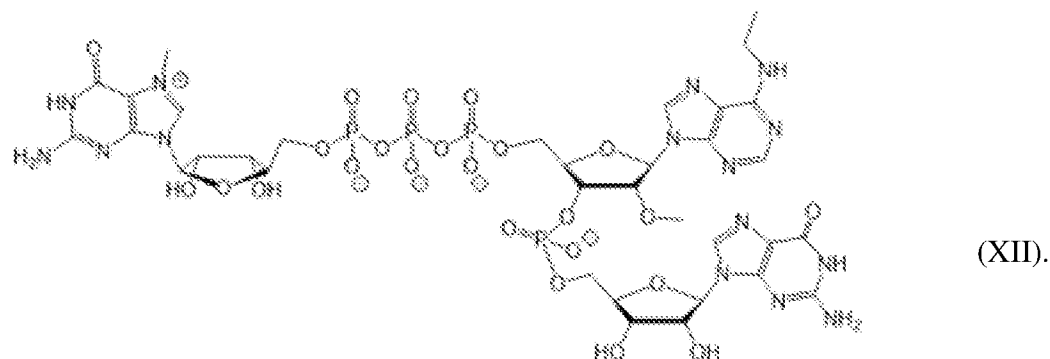
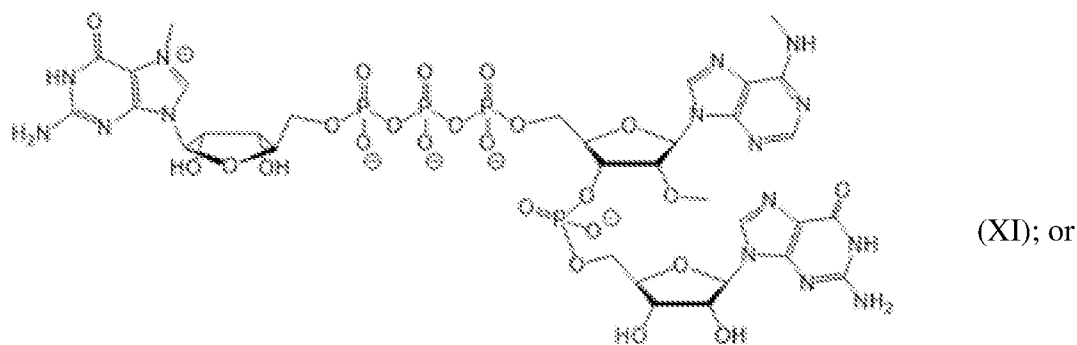
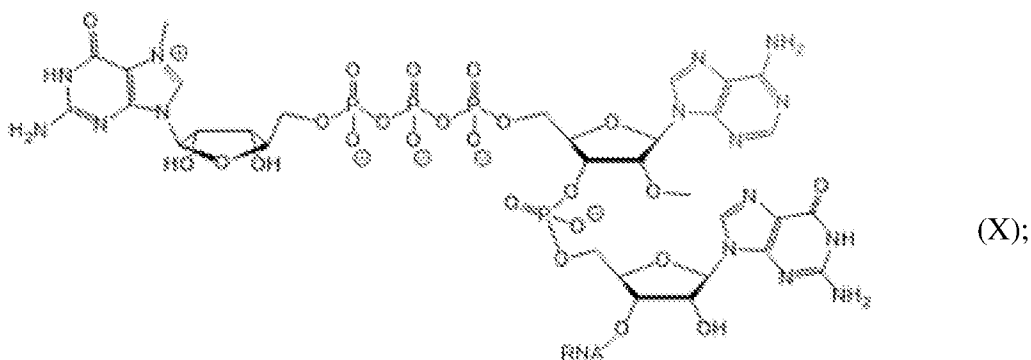
30 $m^7GpppG_{2'OMe}pC$. In some embodiments, a trinucleotide cap comprises $m^7GpppG_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7GpppG_{2'OMe}pU$. In some embodiments, a trinucleotide cap comprises $m^7GpppU_{2'OMe}pA$. In some embodiments, a trinucleotide cap

comprises $m^7GpppU_{2'OMe}pC$. In some embodiments, a trinucleotide cap comprises $m^7GpppU_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7GpppU_{2'OMe}pU$.

In some embodiments, a trinucleotide cap comprises $m^7Gpppm^6A_{2'OMe}pG$. In some embodiments, a trinucleotide cap comprises $m^7Gpppe^6A_{2'OMe}pG$.

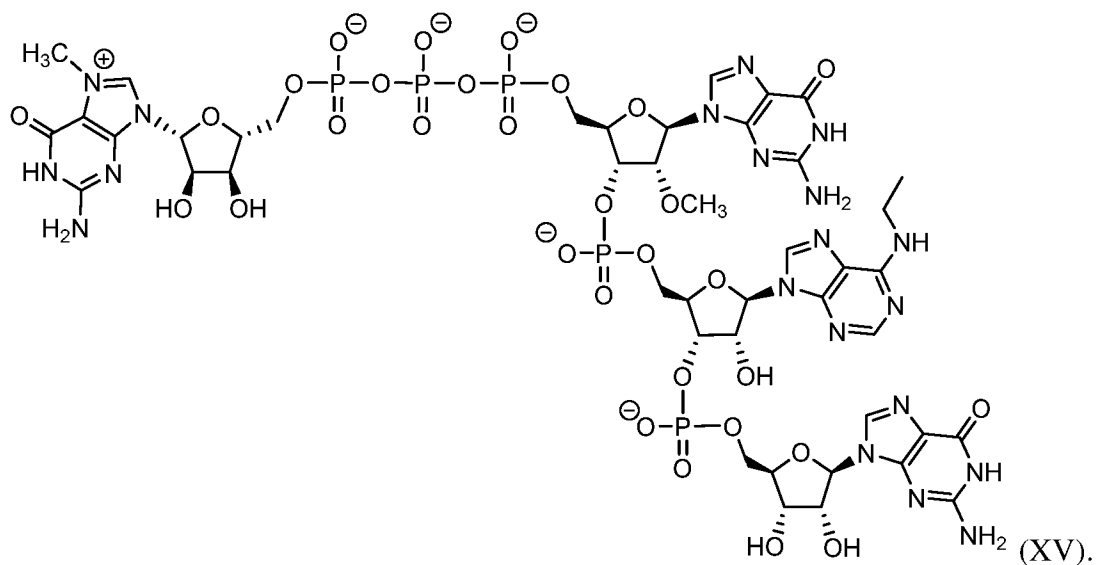
5 In some embodiments, a trinucleotide cap comprises GAG. In some embodiments, a trinucleotide cap comprises GCG. In some embodiments, a trinucleotide cap comprises GUG. In some embodiments, a trinucleotide cap comprises GGG.

In some embodiments, a trinucleotide cap comprises any one of the following structures:



10 In some embodiments, the cap analog comprises a tetranucleotide cap. In some embodiments, the cap analog comprises GGAG.

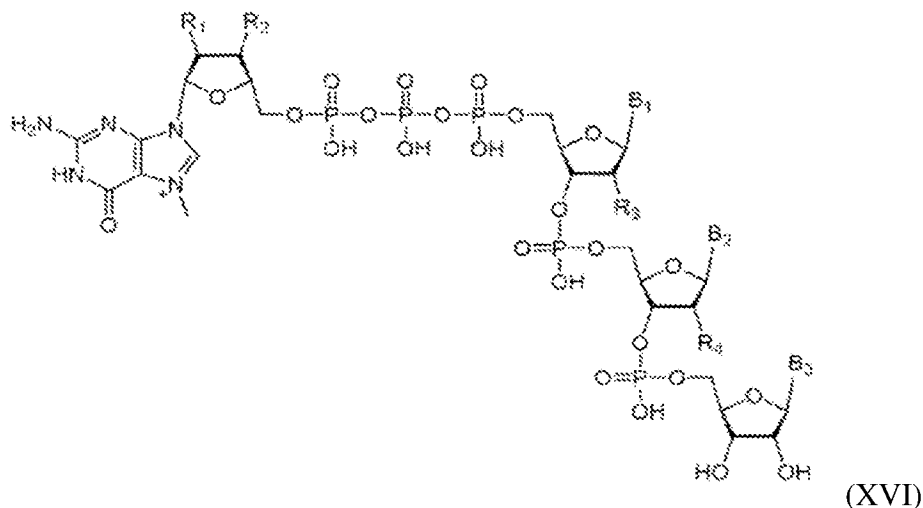
In some embodiments, a tetranucleotide cap comprises any one of the following structures:



In some embodiments, the tetranucleotide cap comprises a trinucleotide as set forth above. In some embodiments, the tetranucleotide cap comprises ${}^{m7}\text{GpppN}_1\text{N}_2\text{N}_3$, where N_1 , N_2 , and N_3 are optional (i.e., can be absent or one or more can be present) and are independently a natural, a modified, or an unnatural nucleoside base. In some embodiments, ${}^{m7}\text{G}$ is further

5 methylated, *e.g.*, at the 3' position. In some embodiments, the ^{m7}G comprises an O-methyl at the 3' position. In some embodiments N₁, N₂, and N₃ if present, optionally, are independently an adenine, a uracil, a guanine, a thymine, or a cytosine. In some embodiments, one or more (or all) of N₁, N₂, and N₃, if present, are methylated, *e.g.*, at the 2' position. In some embodiments, one or more (or all) of N₁, N₂, and N₃, if present have an O-methyl at the 2' position.

In some embodiments, the tetranucleotide cap comprises the following structure:



wherein B₁, B₂, and B₃ are independently a natural, a modified, or an unnatural nucleoside based; and R₁, R₂, R₃, and R₄ are independently OH or O-methyl. In some 10 embodiments, R₃ is O-methyl and R₄ is OH. In some embodiments, R₃ and R₄ are O-methyl. In some embodiments, R₄ is O-methyl. In some embodiments, R₁ is OH, R₂ is OH, R₃ is O-methyl, and R₄ is OH. In some embodiments, R₁ is OH, R₂ is OH, R₃ is O-methyl, and R₄ is O-methyl. In some embodiments, at least one of R₁ and R₂ is O-methyl, R₃ is O-methyl, and R₄ is OH. In some embodiments, at least one of R₁ and R₂ is O-methyl, R₃ is O-methyl, and R₄ is O-methyl.

15 In some embodiments, B₁, B₂, and B₃ are natural nucleoside bases. In some embodiments, at least one of B₁, B₂, and B₃ is a modified or unnatural base. In some embodiments, at least one of B₁, B₂, and B₃ is N6-methyladenine. In some embodiments, B₁ is adenine, cytosine, thymine, or uracil. In some embodiments, B₁ is adenine, B₂ is uracil, and B₃ is adenine. In some embodiments, R₁ and R₂ are OH, R₃ and R₄ are O-methyl, B₁ is adenine, B₂ 20 is uracil, and B₃ is adenine.

In some embodiments the tetranucleotide cap comprises a sequence selected from the following sequences: GAAA, GACA, GAGA, GAUA, GCAA, GCCA, GCGA, GCUA, GGAA, GGCA, GGGA, GGUA, GUCA, and GUUA. In some embodiments the tetranucleotide cap comprises a sequence selected from the following sequences: GAAG, GACG, GAGG, GAUG, 25 GCAG, GCCG, GCGG, GCUG, GGAG, GGCG, GGGG, GGUG, GUUG, GUGG, and GUUG. In some embodiments the tetranucleotide cap comprises a sequence selected from the following sequences: GAAU, GACU, GAGU, GAUU, GCAU, GCCU, GCGU, GCUU, GGAU, GGCU,

GGGU, GGUU, GUAU, GUCU, GUGU, and GUUU. In some embodiments the tetranucleotide cap comprises a sequence selected from the following sequences: GAAC, GACC, GAGC, GAUC, GCAC, GCCC, GCGC, GCUC, GGAC, GGCC, GGGC, GGUC, GUAC, GUCC, GUGC, and GUUC.

5 A tetranucleotide cap, in some embodiments, comprises a sequence selected from the following sequences: $m^7G_3'OMepppApApN$, $m^7G_3'OMepppApCpN$, $m^7G_3'OMepppApGpN$, $m^7G_3'OMepppApUpN$, $m^7G_3'OMepppCpApN$, $m^7G_3'OMepppCpCpN$, $m^7G_3'OMepppCpGpN$, $m^7G_3'OMepppCpUpN$, $m^7G_3'OMepppGpApN$, $m^7G_3'OMepppGpCpN$, $m^7G_3'OMepppGpGpN$, $m^7G_3'OMepppGpUpN$, $m^7G_3'OMepppUpApN$, $m^7G_3'OMepppUpCpN$, $m^7G_3'OMepppUpGpN$, and
10 $m^7G_3'OMepppUpUpN$, where N is a natural, a modified, or an unnatural nucleoside base.

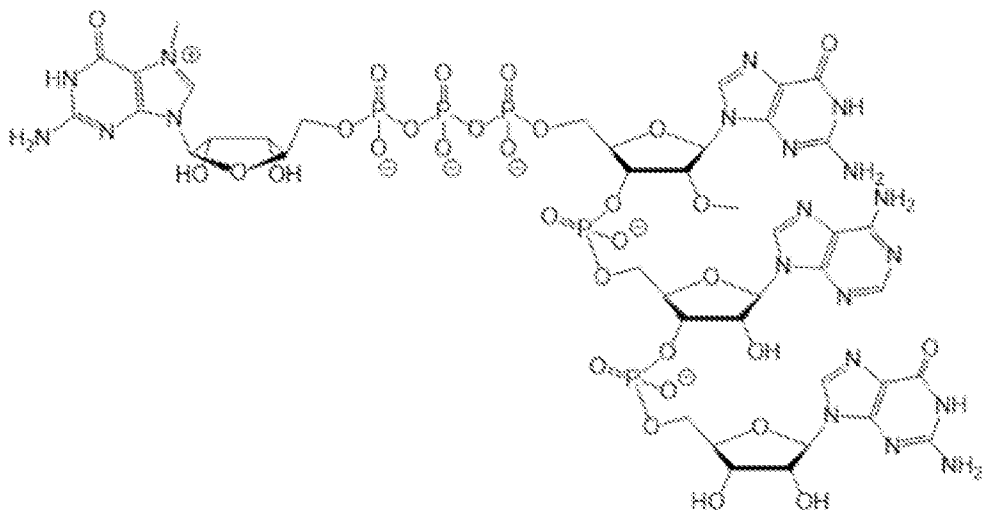
A tetranucleotide cap, in other embodiments, comprises a sequence selected from the following sequences: $m^7G_3'OMepppA_2'OMepApN$, $m^7G_3'OMepppA_2'OMepCpN$, $m^7G_3'OMepppA_2'OMepGpN$, $m^7G_3'OMepppA_2'OMepUpN$, $m^7G_3'OMepppC_2'OMepApN$, $m^7G_3'OMepppC_2'OMepCpN$, $m^7G_3'OMepppC_2'OMepGpN$, $m^7G_3'OMepppC_2'OMepUpN$,
15 $m^7G_3'OMepppG_2'OMepApN$, $m^7G_3'OMepppG_2'OMepCpN$, $m^7G_3'OMepppG_2'OMepGpN$, $m^7G_3'OMepppG_2'OMepUpN$, $m^7G_3'OMepppU_2'OMepApN$, $m^7G_3'OMepppU_2'OMepCpN$, $m^7G_3'OMepppU_2'OMepGpN$, and $m^7G_3'OMepppU_2'OMepUpN$, where N is a natural, a modified, or an unnatural nucleoside base.

A tetranucleotide cap, in still other embodiments, comprises a sequence selected from the following sequences: $m^7GpppA_2'OMepApN$, $m^7GpppA_2'OMepCpN$, $m^7GpppA_2'OMepGpN$,
20 $m^7GpppA_2'OMepUpN$, $m^7GpppC_2'OMepApN$, $m^7GpppC_2'OMepCpN$, $m^7GpppC_2'OMepGpN$, $m^7GpppC_2'OMepUpN$, $m^7GpppG_2'OMepApN$, $m^7GpppG_2'OMepCpN$, $m^7GpppG_2'OMepGpN$, $m^7GpppG_2'OMepUpN$, $m^7GpppU_2'OMepApN$, $m^7GpppU_2'OMepCpN$, $m^7GpppU_2'OMepGpN$, and $m^7GpppU_2'OMepUpN$, where N is a natural, a modified, or an unnatural nucleoside base.

25 A tetranucleotide cap, in other embodiments, comprises a sequence selected from the following sequences: $m^7G_3'OMepppA_2'OMepA_2'OMepN$, $m^7G_3'OMepppA_2'OMepC_2'OMepN$, $m^7G_3'OMepppA_2'OMepG_2'OMepN$, $m^7G_3'OMepppA_2'OMepU_2'OMepN$, $m^7G_3'OMepppC_2'OMepA_2'OMepN$, $m^7G_3'OMepppC_2'OMepC_2'OMepN$, $m^7G_3'OMepppC_2'OMepG_2'OMepN$, $m^7G_3'OMepppC_2'OMepU_2'OMepN$, $m^7G_3'OMepppG_2'OMepA_2'OMepN$, $m^7G_3'OMepppG_2'OMepC_2'OMepN$, $m^7G_3'OMepppG_2'OMepG_2'OMepN$,
30 $m^7G_3'OMepppG_2'OMepU_2'OMepN$, $m^7G_3'OMepppU_2'OMepA_2'OMepN$, $m^7G_3'OMepppU_2'OMepC_2'OMepN$, $m^7G_3'OMepppU_2'OMepG_2'OMepN$, and $m^7G_3'OMepppU_2'OMepU_2'OMepN$, where N is a natural, a modified, or an unnatural nucleoside base.

A tetranucleotide cap, in still other embodiments, comprises a sequence selected from the following sequences: $m^7GpppA_{2'OMe}pA_{2'OMe}pN$, $m^7GpppA_{2'OMe}pC_{2'OMe}pN$, $m^7GpppA_{2'OMe}pG_{2'OMe}pN$, $m^7GpppA_{2'OMe}pU_{2'OMe}pN$, $m^7GpppC_{2'OMe}pA_{2'OMe}pN$, $m^7GpppC_{2'OMe}pC_{2'OMe}pN$, $m^7GpppC_{2'OMe}pG_{2'OMe}pN$, $m^7GpppC_{2'OMe}pU_{2'OMe}pN$, $m^7GpppG_{2'OMe}pA_{2'OMe}pN$, $m^7GpppG_{2'OMe}pC_{2'OMe}pN$, $m^7GpppG_{2'OMe}pG_{2'OMe}pN$, $m^7GpppG_{2'OMe}pU_{2'OMe}pN$, $m^7GpppU_{2'OMe}pA_{2'OMe}pN$, $m^7GpppU_{2'OMe}pC_{2'OMe}pN$, $m^7GpppU_{2'OMe}pG_{2'OMe}pN$, and $m^7GpppU_{2'OMe}pU_{2'OMe}pN$, where N is a natural, a modified, or an unnatural nucleoside base.

In some embodiments, a tetranucleotide cap comprises GGAG. In some embodiments, a tetranucleotide cap comprises the following structure:



(XVII).

The capping efficiency of a post-transcriptional or co-transcriptional capping reaction may vary. The term “capping efficiency” may refer to the amount (e.g., expressed as a percentage) of mRNAs comprising a cap structure relative to the total mRNAs in a mixture (e.g., a post-translational capping reaction or a co-transcriptional calling reaction). In some embodiments, the capping efficiency of a capping reaction is at least 60%, 70%, 80%, 90%, 95%, 99%, or 99.9% (e.g., after the capping reaction at least 60%, 70%, 80%, 90%, 95%, 99%, or 99.9% of the input mRNAs comprise a cap). In some embodiments, multivalent co-IVT reactions do not affect the capping efficiency of the mRNAs resulting from the IVT reaction.

***In vitro* Transcription Methods**

Some aspects relate to methods of producing (e.g., synthesizing) a RNA transcript (e.g., mRNA transcript) comprising contacting a DNA template (e.g., a first input DNA and a second input DNA) with a RNA polymerase (e.g., a T7 RNA polymerase, a T7 RNA polymerase variant, etc.) under conditions that result in the production of the RNA transcript.

Some aspects relate to methods of performing an IVT reaction, comprising contacting a DNA template with the RNA polymerase (*e.g.*, a T7 RNA polymerase, such as a T7 RNA polymerase variant) in the presence of nucleoside triphosphates and buffer under conditions that result in the production of RNA transcripts.

5 Other aspects provide co-transcriptional capping methods that comprise reacting a polynucleotide template with a T7 RNA polymerase variant, nucleoside triphosphates, and a cap analog under *in vitro* transcription reaction conditions to produce RNA transcript.

In some embodiments, a co-transcriptional capping method for RNA synthesis comprises reacting a polynucleotide template with (a) a T7 RNA polymerase variant comprising at least one
10 amino acid substitution, relative to wild-type RNA polymerase (*e.g.*, a T7 polymerase variant comprising amino acid substitutions at positions 437, 387, 350, and 351, relative to SEQ ID NO: 1), (b) nucleoside triphosphates, and (c) a cap analog (*e.g.*, trinucleotide cap comprising sequence GpppA_{2'}OmepG), under *in vitro* transcription reaction conditions to produce RNA transcript, wherein the polynucleotide template includes a 2'-deoxythymidine residue at template
15 position +1.

IVT conditions typically require a purified linear DNA template containing a promoter, nucleoside triphosphates, a buffer system that includes dithiothreitol (DTT) and magnesium ions, and an RNA polymerase. The exact conditions used in the transcription reaction depend on the amount of RNA needed for a specific application. Typical IVT reactions are performed by
20 incubating a DNA template with an RNA polymerase and nucleoside triphosphates, including GTP, ATP, CTP, and UTP (or nucleotide analogs) in a transcription buffer. An RNA transcript having a 5' terminal guanosine triphosphate is produced from this reaction.

The “percent identity,” “sequence identity,” “% identity,” or “% sequence identity” (as they may be interchangeably used herein) of two sequences (*e.g.*, nucleic acid or amino acid)
25 refers to a quantitative measurement of the similarity between two sequences (*e.g.*, nucleic acid or amino acid). Percent identity can be determined using the algorithms of Karlin and Altschul, Proc. Natl. Acad. Sci. USA 87:2264-68, 1990, modified as in Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90:5873-77, 1993. Such algorithms are incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul et al., J. Mol. Biol. 215:403-10, 1990. BLAST
30 protein searches can be performed with the XBLAST program, score=50, word length=3, to obtain amino acid sequences homologous to the protein molecules of interest. Where gaps exist between two sequences, Gapped BLAST can be utilized as described in Altschul et al., Nucleic Acids Res. 25(17):3389-3402, 1997. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. When
35 a percent identity is stated, or a range thereof (*e.g.*, at least, more than, etc.), unless otherwise

specified, the endpoints shall be inclusive and the range (e.g., at least 70% identity) shall include all ranges within the cited range.

The input deoxyribonucleic acid (DNA) serves as a nucleic acid template for RNA polymerase. A DNA template may include a polynucleotide encoding a polypeptide of interest (e.g., an antigenic polypeptide). A DNA template, in some embodiments, includes a RNA polymerase promoter (e.g., a T7 RNA polymerase promoter) located 5' from and operably linked to polynucleotide encoding a polypeptide of interest. A DNA template may also include a nucleotide sequence encoding a polyadenylation (polyA) tail located at the 3' end of the gene of interest. In some embodiments, an input DNA comprises plasmid DNA (pDNA). The terms "plasmid DNA" or "pDNA" may refer to an extrachromosomal DNA molecule that is physically separated from chromosomal DNA in a cell and can replicate independently. In some embodiments, plasmid DNA is isolated from a cell (e.g., as a plasmid DNA preparation). In some embodiments, plasmid DNA comprises an origin of replication, which may contain one or more heterologous nucleic acids, for example nucleic acids encoding therapeutic proteins that may serve as a template for RNA polymerase. Plasmid DNA may be circularized or linear (e.g., plasmid DNA that has been linearized by a restriction enzyme digest).

In some embodiments, each input DNA (e.g., population of input DNA molecules) in a co-IVT reaction is obtained from a different source (e.g., synthesized separately, for example in different cells or populations of cells). In some embodiments, each input DNA (e.g., population of input DNA) is obtained from a different bacterial cell or population of bacterial cells. For example, in a co-IVT reaction having three populations of input DNAs, the first input DNA is produced in bacterial cell population A, the second input DNA is produced in bacterial cell population B, and the third input DNA is produced in bacterial population C, where each of A, B, and C are not the same bacterial culture (e.g., co-cultured in the same container or plate). In another example, two input DNAs obtained from different sources are i) chemically synthesized in separate synthesis reactions, or ii) produced by separate amplification (e.g., polymerase chain reactions (PCR reactions)). Methods of obtaining populations of input DNAs (e.g., plasmid DNAs) are known, for example as described by Sambrook, Joseph. *Molecular Cloning : a Laboratory Manual*. Cold Spring Harbor, N.Y. :Cold Spring Harbor Laboratory Press, 2001.

Some aspects comprise normalizing the amount of DNA used in the multivalent co-IVT reaction. In some embodiments, the normalization is based on the molar mass of the input DNAs. In some embodiments, the normalization is based on the degradation rate of the input DNAs. In some embodiments, the normalization is based on the degradation rate of the resultant mRNAs (e.g., measured based upon polyA variants present in the reaction mixture, or T7 polymerase abortive transcripts or truncated transcripts). In some embodiments, the

normalization is based on the nucleotide content (*e.g.*, amount of A, G, C, U, or any combination thereof) of the input DNAs. In some embodiments, the normalization is based on the purity of the input DNAs. In some embodiments the normalization is based on the polyA-tailing efficiency of the input DNAs. In some embodiments, the normalization is based on the lengths of the input DNAs.

In some embodiments, the normalization is based on the lowest level present in the input DNAs (*e.g.*, lowest molar mass, degradation rate (*e.g.*, of the input DNA and/or output RNA), nucleotide content, purity, and/or polyA-tailing efficiency). In some embodiments, the normalization is based on the highest level present in the input DNAs (*e.g.*, highest molar mass, degradation rate (*e.g.*, of the input DNA and/or output RNA), nucleotide context, purity, and/or polyA-tailing efficiency). In some embodiments, the normalization is based on the rate of RNA production of the input DNAs (*e.g.*, the highest rate of RNA production of an input DNA or the lowest rate of RNA production of an input DNA in a reaction mixture).

Some aspects relate to IVT methods in which the amount of input DNA (*e.g.*, a first DNA or second DNA) is adjusted or normalized in order to improve production of multivalent RNA compositions having a pre-defined mRNA ratio of components. The disclosure is based, in part, on the discovery that certain factors affecting multivalent RNA composition purity, such as large differences in size between input DNAs (*e.g.*, a difference of more than 100, 200, 500, 1000, or more nucleotides in length) and/or polyA-tailing efficiency of a given DNA during IVT, may be addressed prior to the IVT by normalizing the amount of input DNA based upon one or more of those factors. For example, in some embodiments, the amount of two input DNAs is calculated based upon the desired molar ratio of the first RNA to the second RNA that are transcribed from the input DNAs. In some embodiments, the calculating comprises determining a plasmid mass ratio based upon the desired molar ratio of the input DNAs. In some embodiments, the amount of input DNAs is normalized based upon the highest polyA-tailing efficiency of the input DNAs during IVT.

The number of input DNAs (*e.g.*, populations of input DNA molecules) used in an IVT reaction may vary, depending upon the number of different RNA molecules desired to be included in the multivalent RNA composition. In some embodiments, an IVT reaction mixture comprises 2 or more different input DNAs, for example 2, 3, 4, 5, 6, 7, 8, 9, 10, or more different input DNAs. In some embodiments, the IVT reaction comprises more than 10 different input DNAs. The term “different input DNAs” encompasses input DNAs that encode different RNAs, *e.g.*, that have i) different lengths (whether or not the RNAs are identical over the entirety of the shorter of the two lengths), ii) different nucleotide sequences, iii) different chemical modification patterns, or iv) any combination of the foregoing.

The concentration of each of the populations of DNA molecules may also vary. In some embodiments, the concentration of each population of DNA molecules in an IVT reaction ranges from about 0.005 mg/mL to about 0.5 mg/ml. In some embodiments, the concentration of each population of DNA molecules in an IVT reaction ranges from about 0.02 mg/ml to about 0.05 mg/ml, 0.02 to about 0.15 mg/ml, about 0.05 mg/ml to about 0.20 mg/ml, about 0.175 to about 0.3 mg/ml, about 0.2 mg/ml to about 0.5 mg/ml, about 0.3 mg/ml to about 0.6 mg/ml, about 0.5 mg/ml to about 0.75 mg/ml, about 0.5 mg/ml to about 1.0 mg/ml, about 0.75 mg/ml to about 0.9 mg/ml, about 0.75mg/ml to about 1.5 mg/ml, about 0.8 mg/ml to about 1.2 mg/ml, about 1.0 mg/ml to about 1.5 mg/ml, about 1.0 mg/ml to about 2.5 mg/ml, about 1.5 mg/ml to about 3.0 mg/ml, about 2.0 mg/ml to about 4.0 mg/ml, or about 2.5 mg/ml to about 5.0 mg/ml.

In some embodiments, the input DNAs are added to an IVT reaction are a predefined DNA ratio, which may comprise a ratio between 2, 3, 4, 5, 6, 7, 8, 9, 10, or more different input DNAs (e.g., depending on the number of different RNAs in a composition). In some embodiments, a pre-defined input DNA ratio comprises a ratio between more than 10 input DNAs. The term “pre-defined input DNA ratio” may refer to the desired final ratio of DNA molecules in an IVT reaction. The desired final ratio of input DNAs can depend upon the final peptide(s) or polypeptide product(s) encoded by RNAs encoded by the input DNAs. In some embodiments, the input DNAs can have a desired ratio that may comprise between 2 and 8 input DNAs (e.g., a:b, a:b:c, a:b:c:d, a:b:c:d:e, a:b:c:d:e:f, a:b:c:d:e:f:g, a:b:c:d:e:f:g:h, etc., where each of a-h is a number between 1 and 10). In some embodiments, the pre-defined input DNA ratio is different from the pre-defined mRNA ratio.

The size of two or more input DNAs (e.g., DNAs in two or more different populations of input DNAs) may vary. In some embodiments, an input DNA includes from about 15 to about 8,000 base pairs (e.g., from 15 to 50, 15 to 100, 15 to 200, 15 to 300, 15 to 400, 15 to 500, 15 to 600, 15 to 700, 15 to 800, 15 to 900, 15 to 1000, 15 to 1200, 15 to 1400, 15 to 1500, 15 to 1800, 15 to 2000, 15 to 2500, 15 to 3000, 50 to 100, 50 to 200, 50 to 300, 50 to 400, 50 to 500, 50 to 600, 50 to 700, 50 to 800, 50 to 900, 50 to 1000, 50 to 1200, 50 to 1400, 50 to 1500, 50 to 1800, 50 to 2000, 50 to 2500, 50 to 3000, 100 to 200, 100 to 300, 100 to 400, 100 to 500, 100 to 600, 100 to 700, 100 to 800, 100 to 900, 100 to 1000, 100 to 1200, 100 to 1400, 100 to 1500, 100 to 1800, 100 to 2000, 100 to 2500, 100 to 3000, 200 to 300, 200 to 400, 200 to 500, 200 to 600, 200 to 700, 200, to 800, 200 to 900, 200 to 1000, 200 to 1500, 200 to 3000, 500 to 1000, 500 to 1500, 500 to 2000, 500 to 2500, 500 to 3000, 1000 to 1500, 1000 to 2000, 1000 to 2500, 1000 to 3000, 1500 to 3000, 2500 to 3000, 2000 to 3000, 2500 to 4000, 3000 to 5000, 3500 to 6500, 5000 to 7500, or 6500 to 8000 base pairs.

The mass of each population of input DNA molecules in an IVT reaction may vary. In some embodiments, the mass of each population of input DNA ranges based upon the total volume of the IVT reaction mixture. In some embodiments, the mass of each population of each input DNA molecule in an IVT mixture individually varies from about 0.5% to about 99.9% of the total input DNA present in the IVT reaction mixture. In some embodiments, the molar ratio of each population of input DNA molecules in an IVT reaction may vary.

In some embodiments, two or more of the input DNA molecules used in an IVT reaction have a different length (e.g., comprises a different number of nucleotides). In some embodiments, the difference in length between two or more (e.g., 3, 4, 5, 6, 7, 8, 9, 10, or more) of the different input DNA molecules in an IVT reaction mixture is greater than 70 base pairs, 80 base pairs, 90 base pairs, or 100 base pairs (e.g., two input DNAs in a composition are not within 70, 80, 90, or 100 base pairs in length of one another). In some embodiments, the difference in length between two or more (e.g., 3, 4, 5, 6, 7, 8, 9, 10, or more) of the different input DNA molecules is more than 100 base pairs, for example 500 base pairs, 1000 base pairs, 1500 base pairs, 2000 base pairs, 3000 base pairs, 4000 base pairs, 5000 base pairs, 6000 base pairs, 7000 base pairs, 8000 base pairs, or more.

In some embodiments, two or more of the input DNA molecules used in an IVT reaction encode mRNA molecules that have a different length (e.g., comprises a different number of nucleotides). In some embodiments, the difference in length between two or more of the mRNA molecules encoded by different input DNA molecules in an IVT reaction mixture is greater than 70 nucleotides, 80 nucleotides, 90 nucleotides, or 100 nucleotides (e.g., two input DNAs in a composition encode mRNA molecules that are not within 70, 80, 90, or 100 nucleotides in length of one another). In some embodiments, the difference in length between two or more of the mRNA molecules encoded by different input DNA molecules is more than 100 nucleotides, for example 500 nucleotides, 1000 nucleotides, 1500 nucleotides, 2000 nucleotides, 3000 nucleotides, 4000 nucleotides, or more.

In some embodiments, the multivalent IVT comprises co-transcription of at least 2 different input DNAs (e.g., at least 2 of DNA A, B, C, D, E, F, G, H, I, J, etc.) at a ratio of A:B:C:D:E:F:G:H:I:J, wherein if DNA A is normalized to 1, one or more of DNA B, C, D, E, F, G, H, I, J, etc. can each independently be present at an amount (e.g., a concentration) that is from 0.01 to 100 times the amount (e.g., a concentration) of A, such as from 0.05 times to 20 times the amount of A, 0.1 times to 10 times the amount of A, 0.2 times to 5 times the amount of A, 0.3 times to 3 times the amounts of A, 0.5 times to 2 times the amounts of A, 0.75 times to 1.4 times the amount of A, 0.8 times to 1.25 times the amount of A, or 0.9 times to 1.15 times the amount of A. One or more of DNA B, C, D, E, F, G, H, I, or J may also be absent.

In some embodiments, a multivalent RNA composition is produced by combining RNA transcripts (e.g., mRNAs) from separate sources. In some embodiments, a multivalent RNA composition is produced by separately transcribing two or more DNA templates in separate IVT reactions, and combining the transcribed RNAs. In some embodiments, an RNA transcript is produced by IVT, then added to one or more other RNAs. RNAs may be combined in any desired amount to produce a multivalent RNA composition comprising two or more RNAs in a specific ratio.

A RNA transcript, in some embodiments, is the product of an IVT reaction. A RNA transcript, in some embodiments, is a messenger RNA (mRNA) that includes a nucleotide sequence encoding a polypeptide of interest (e.g., a therapeutic protein or therapeutic peptide) linked to a polyA tail. In some embodiments, the mRNA is modified mRNA (mmRNA), which includes at least one modified nucleotide.

The nucleoside triphosphates (NTPs) may comprise unmodified or modified ATP, modified or unmodified UTP, modified or unmodified GTP, and/or modified or unmodified CTP. In some embodiments, NTPs of an IVT reaction comprise unmodified ATP. In some embodiments, NTPs of an IVT reaction comprise modified ATP. In some embodiments, NTPs of an IVT reaction comprise unmodified UTP. In some embodiments, NTPs of an IVT reaction comprise modified UTP. In some embodiments, NTPs of an IVT reaction comprise unmodified GTP. In some embodiments, NTPs of an IVT reaction comprise modified GTP. In some embodiments, NTPs of an IVT reaction comprise unmodified CTP. In some embodiments, NTPs of an IVT reaction comprise modified CTP.

The composition of NTPs in an IVT reaction may also vary. In some embodiments, each NTP in an IVT reaction is present in an equimolar amount. In some embodiments, each NTP in an IVT reaction is present in non-equimolar amounts. For example, ATP may be used in excess of GTP, CTP and UTP. As a non-limiting example, an IVT reaction may include 7.5 millimolar GTP, 7.5 millimolar CTP, 7.5 millimolar UTP, and 3.75 millimolar ATP. In some embodiments, the molar ratio of G:C:U:A is 2:1:0.5:1. In some embodiments, the molar ratio of G:C:U:A is 1:1:0.7:1. In some embodiments, the molar ratio of G:C:A:U is 1:1:1:1. The same IVT reaction may include 3.75 millimolar cap analog (e.g., trinucleotide cap or tetranucleotide cap). In some embodiments, the molar ratio of G:C:U:A:cap is 1:1:1:0.5:0.5. In some embodiments, the molar ratio of G:C:U:A:cap is 1:1:0.5:1:0.5. In some embodiments, the molar ratio of G:C:U:A:cap is 1:0.5:1:1:0.5. In some embodiments, the molar ratio of G:C:U:A:cap is 0.5:1:1:1:0.5. In some embodiments, the amount of NTPs in a co-IVT reaction is calculated empirically. For example, the rate of consumption for each NTP in an IVT reaction may be empirically determined for each

individual input DNA, and then balanced ratios of NTPs based on those individual NTP consumption rates may be added to a co-IVT comprising multiple of the input DNAs.

In some embodiments, an IVT reaction mixture further comprises cap analog. The concentration of nucleoside triphosphates and cap analog present in an IVT reaction may vary. In some embodiments, NTPs and cap analog are present in the reaction at equimolar concentrations. In some embodiments, the molar ratio of cap analog (*e.g.*, trinucleotide cap or tetranucleotide cap) to nucleoside triphosphates in the reaction is greater than 1:1. For example, the molar ratio of cap analog to nucleoside triphosphates in the reaction may be 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 15:1, 20:1, 25:1, 50:1, or 100:1. In some embodiments, the molar ratio of cap analog (*e.g.*, trinucleotide cap or tetranucleotide cap) to nucleoside triphosphates in the reaction is less than 1:1. For example, the molar ratio of cap analog (*e.g.*, trinucleotide cap or tetranucleotide cap) to nucleoside triphosphates in the reaction may be 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:15, 1:20, 1:25, 1:50, or 1:100.

In some embodiments, a RNA transcript (*e.g.*, mRNA transcript) includes a modified nucleobase selected from pseudouridine (ψ), 1-methylpseudouridine ($m^1\psi$), 5-methoxyuridine (mo^5U), 5-methylcytidine (m^5C), α -thio-guanosine and α -thio-adenosine. In some embodiments, a RNA transcript (*e.g.*, mRNA transcript) includes a combination of at least two (*e.g.*, 2, 3, 4 or more) of the foregoing modified nucleobases.

In some embodiments, a RNA transcript (*e.g.*, mRNA transcript) includes pseudouridine (ψ). In some embodiments, a RNA transcript (*e.g.*, mRNA transcript) includes 1-methylpseudouridine ($m^1\psi$). In some embodiments, a RNA transcript (*e.g.*, mRNA transcript) includes 5-methoxyuridine (mo^5U). In some embodiments, a RNA transcript (*e.g.*, mRNA transcript) includes 5-methylcytidine (m^5C). In some embodiments, a RNA transcript (*e.g.*, mRNA transcript) includes α -thio-guanosine. In some embodiments, a RNA transcript (*e.g.*, mRNA transcript) includes α -thio-adenosine.

In some embodiments, the polynucleotide (*e.g.*, RNA polynucleotide, such as mRNA polynucleotide) is uniformly modified (*e.g.*, fully modified, modified throughout the entire sequence) for a particular modification. For example, a polynucleotide can be uniformly modified with 1-methylpseudouridine ($m^1\psi$), meaning that all uridine residues in the mRNA sequence are replaced with 1-methylpseudouridine ($m^1\psi$). Similarly, a polynucleotide can be uniformly modified for any type of nucleoside residue present in the sequence by replacement with a modified residue such as any of those set forth above. Alternatively, the polynucleotide (*e.g.*, RNA polynucleotide, such as mRNA polynucleotide) may not be uniformly modified (*e.g.*, partially modified, part of the sequence is modified). Each possibility represents a separate embodiment.

The buffer system of an IVT reaction mixture may vary. In some embodiments, the buffer system contains tris. The concentration of tris used in an IVT reaction, for example, may be at least 10 mM, at least 20 mM, at least 30 mM, at least 40 mM, at least 50 mM, at least 60 mM, at least 70 mM, at least 80 mM, at least 90 mM, at least 100 mM or at least 110 mM phosphate. In some embodiments, the concentration of phosphate is 20-60 mM or 10-100 mM.

In some embodiments, the buffer system contains dithiothreitol (DTT). The concentration of DTT used in an IVT reaction, for example, may be at least 1 mM, at least 5 mM, or at least 50 mM. In some embodiments, the concentration of DTT used in an IVT reaction is 1-50 mM or 5-50 mM. In some embodiments, the concentration of DTT used in an IVT reaction is 5 mM.

In some embodiments, the buffer system contains magnesium. In some embodiments, the molar ratio of NTP to magnesium ions (Mg^{2+} ; *e.g.*, MgCl_2) present in an IVT reaction is 1:1 to 1:5. For example, the molar ratio of NTP to magnesium ions may be 1:0.25, 1:0.5, 1:1, 1:2, 1:3, 1:4 or 1:5.

In some embodiments, the molar ratio of NTP plus cap analog (*e.g.*, trinucleotide cap, such as GAG) to magnesium ions (Mg^{2+} ; *e.g.*, MgCl_2) present in an IVT reaction is 1:1 to 1:5. For example, the molar ratio of NTP+trinucleotide cap (*e.g.*, GAG) to magnesium ions may be 1:1, 1:2, 1:3, 1:4 or 1:5.

In some embodiments, the buffer system contains Tris-HCl, spermidine (*e.g.*, at a concentration of 1-30 mM), TRITON[®] X-100 (polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether) and/or polyethylene glycol (PEG).

In some embodiments, IVT methods further comprise a step of separating (*e.g.*, purifying) *in vitro* transcription products (*e.g.*, mRNA) from other reaction components. In some embodiments, the separating comprises performing chromatography on the IVT reaction mixture. In some embodiments, the chromatography comprises size-based (*e.g.*, length-based) chromatography. In some embodiments, the chromatography comprises oligo-dT chromatography.

The addition of nucleoside triphosphates (NTPs) to the 3' end of a growing RNA strand is catalyzed by a polymerase, such as T7 RNA polymerase, for example, a T7 RNA polymerase variant (*e.g.*, RNA polymerase comprising D653W/E350W/D351V substitutions). In some embodiments, the RNA polymerase (*e.g.*, T7 RNA polymerase variant) is present in a reaction (*e.g.*, an IVT reaction) at a concentration of 0.01 mg/ml to 1 mg/ml. For example, the RNA polymerase may be present in a reaction at a concentration of 0.01 mg/mL, 0.05 mg/ml, 0.1 mg/ml, 0.5 mg/ml or 1.0 mg/ml.

Surprisingly, use of the combination of a T7 RNA polymerase variant (*e.g.*, RNA polymerase comprising D653W/E350W/D351V substitutions) with a cap analog (*e.g.*,

GpppA_{2'}OmepG), in an *in vitro* transcription reaction, for example, results in the production of RNA transcript, wherein greater than 80% of the RNA transcript produced includes a functional cap. In some embodiments, greater than 85% of the RNA transcript produced includes a functional cap. In some embodiments, greater than 90% of the RNA transcript produced includes a functional cap. In some embodiments, greater than 95% of the RNA transcript produced includes a functional cap. In some embodiments, greater than 96% of the RNA transcript produced includes a functional cap. In some embodiments, greater than 97% of the RNA transcript produced includes a functional cap. In some embodiments, greater than 98% of the RNA transcript produced includes a functional cap. In some embodiments, greater than 99% of the RNA transcript produced includes a functional cap.

Also surprising was the finding that use of a polynucleotide template that includes a 2'-deoxythymidine residue or 2'-deoxycytidine residue at template position +1 results in the production of RNA transcript, wherein greater than 80% (*e.g.*, greater than 85%, greater than 90%, or greater than 95%) of the RNA transcript produced includes a functional cap. Thus, in some embodiments, a polynucleotide (*e.g.*, DNA) template used, for example, in an IVT reaction, includes a 2'-deoxythymidine residue at template position +1. In other embodiments, a polynucleotide (*e.g.*, DNA) template used, for example, in an IVT reaction, includes a 2'-deoxycytidine residue at template position +1.

Applications

The RNA transcripts produced using an RNA polymerase variant may include mRNA (including modified mRNA and/or unmodified RNA), lncRNA, self-replicating RNA, circular RNA, CRISPR guide RNA, and the like. In embodiments, the RNA is RNA (*e.g.*, mRNA or self-replicating RNA) that encodes a polypeptide (*e.g.*, a therapeutic polypeptide). Thus, the RNA transcripts produced using RNA polymerase variants may be used in a myriad of applications.

For example, the RNA transcripts may be used to produce polypeptides of interest, *e.g.*, therapeutic proteins, vaccine antigen, and the like. In some embodiments, the RNA transcripts are therapeutic RNAs. A therapeutic mRNA is an mRNA that encodes a therapeutic protein (the term 'protein' encompasses peptides). Therapeutic proteins mediate a variety of effects in a host cell or in a subject to treat a disease or ameliorate the signs and symptoms of a disease. For example, a therapeutic protein can replace a protein that is deficient or abnormal, augment the function of an endogenous protein, provide a novel function to a cell (*e.g.*, inhibit or activate an endogenous cellular activity, or act as a delivery agent for another therapeutic compound (*e.g.*, an antibody-drug conjugate). Therapeutic mRNA may be useful for the treatment of the following diseases and conditions: bacterial infections, viral infections, parasitic infections, cell

proliferation disorders, genetic disorders, and autoimmune disorders. Other diseases and conditions are encompassed herein.

An RNA transcript produced using an RNA polymerase variant may encode one or more biologics. A biologic is a polypeptide-based molecule that may be used to treat, cure, mitigate, prevent, or diagnose a serious or life-threatening disease or medical condition. Biologics include, but are not limited to, allergenic extracts (*e.g.* for allergy shots and tests), blood components, gene therapy products, human tissue or cellular products used in transplantation, vaccines, monoclonal antibodies, cytokines, growth factors, enzymes, thrombolytics, and immunomodulators, among others.

One or more biologics currently being marketed or in development may be encoded by the RNA produced by an RNA polymerase variant. While not wishing to be bound by theory, it is believed that incorporation of the encoding polynucleotides of a known biologic into the RNA will result in improved therapeutic efficacy due at least in part to the specificity, purity and/or selectivity of the construct designs.

An RNA transcript produced using an RNA polymerase variant may encode one or more antibodies. The term “antibody” includes monoclonal antibodies (including full length antibodies which have an immunoglobulin Fc region), antibody compositions with polyepitopic specificity, multispecific antibodies (*e.g.*, bispecific antibodies, diabodies, and single-chain molecules), as well as antibody fragments. The term “immunoglobulin” (Ig) is used interchangeably with “antibody” herein. A monoclonal antibody is an antibody obtained from a population of substantially homogeneous antibodies, *i.e.*, the individual antibodies comprising the population are identical except for possible naturally occurring mutations and/or post-translation modifications (*e.g.*, isomerizations, amidations) that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site.

Monoclonal antibodies specifically include chimeric antibodies (immunoglobulins) in which a portion of the heavy and/or light chain is identical with or homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is(are) identical with or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired biological activity. Chimeric antibodies include, but are not limited to, “primatized” antibodies comprising variable domain antigen-binding sequences derived from a non-human primate (*e.g.*, Old World Monkey, Ape etc.) and human constant region sequences.

An RNA transcript produced using an RNA polymerase variant may encode one or more vaccine antigens. A vaccine antigen is a biological preparation that improves immunity to a

particular disease or infectious agent. One or more vaccine antigens currently being marketed or in development may be encoded by the RNA. Vaccine antigens encoded in the RNA may be utilized to treat conditions or diseases in many therapeutic areas such as, but not limited to, cancer, allergy and infectious disease. In some embodiments, a cancer vaccine may be a personalized cancer vaccine in the form of a concatemer or individual RNAs encoding peptide epitopes or a combination thereof.

An RNA transcript produced using an RNA polymerase variant may be designed to encode on or more antimicrobial peptides (AMP) or antiviral peptides (AVP). AMPs and AVPs have been isolated and described from a wide range of animals such as, but not limited to, microorganisms, invertebrates, plants, amphibians, birds, fish, and mammals.

In some embodiments, RNA transcripts are used as radiolabeled RNA probes. In some embodiments, RNA transcripts are used for non-isotopic RNA labeling. In some embodiments, RNA transcripts are used as guide RNA (gRNA) for gene targeting. In some embodiments, RNA transcripts (*e.g.*, mRNA) are used for *in vitro* translation and micro injection. In some embodiments, RNA transcripts are used for RNA structure, processing and catalysis studies. In some embodiments, RNA transcripts are used for RNA amplification. In some embodiments, RNA transcripts are used as anti-sense RNA for gene expression experiment.

Wild-type T7 RNA Polymerase

MNTINIAKNDFSDIELAAIPFNTLADHYGERLAREQLALEHESYEMGEARFRKMFERQLKAGEVADNAAAKPLITTL
LPKMIARINDWFEEVKAKRGKRPTAFQFLQEIKPEAVAYITIKTTLACLTSADNTTVQAVASAI GRAIEDEARFGRI
RDLEAKHFKKNVEEQLNKRVGHVYKKAQVVEADMLSKGLLGGEAWSSWHKEDSIHVGVRCEMLIESTGMVSLHR
QNAGVVGQDSEITIELAPEYAEAIATRAGALAGISPMFQPCVVPKPKWTGITGGGYWANGRRPLALVRTHSKKALMRY
EDVYMPEVYKAINIAQNTAWKINKVLAVANVITKWKHCPVEDIPAIEREELPMKPEDIDMNPEALTAWKRAAAVY
RKDKARKSRRISLEFMLEQANKFANHKAIWFPYNMDWRGRVYAVSMFNPQGNDMTKGLLT LAKGKPIGKEGYWLKI
HGANCAGVDKVPFPERIKFIEENHENIMACAKSPLENTWWAEQDSPFCFLAFCFEYAGVQHHGLSYNCSLPLAFDGS
CSGIQHFSAMLRDEVGGRAVNLLPSETVQDIYGIVAKKVNEILQADAINGTDNEVVTVTDENTGEISEKVKLGTKAL
AGQWLAYGVTRSVTKRSVMTLAYGSKEFGFRQQVLEDTIQPAIDSGKGLMFTQPNQAAGYMAKLIWESVSVTVVAAV
EAMNWLKSAAKLLAAEVKDKKTGEILRKRCVHWVTPDGFVPVWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEID
AHKQESGIAPNFVHSQDGSHLRKTVVWAHEKYGIESFALIHDSFGTIPADAANLFKAVRETMVDTYESCDVLADFDYD
QFADQLHESQLDKMPALPAKGNNLNRDILESDFABA (SEQ ID NO: 1)

Additional Embodiments

Additional embodiments are encompassed by the following numbered paragraphs 1-71:

1. A ribonucleic acid (RNA) polymerase variant comprising: an amino acid sequence comprising (i) an amino acid substitution at position E350, (ii) an amino acid substitution at position D351, and (iii) an amino acid substitution at position K387, position N437, or at

position K387 and position N437, relative to a wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

2. The RNA polymerase variant of paragraph 1, wherein the amino acid sequence of the variant comprises an amino acid substitution at position K387.

3. The RNA polymerase variant of paragraph 1, wherein the amino acid sequence of the variant comprises an amino acid substitution at position N437.

4. The RNA polymerase variant of paragraph 1, wherein the amino acid sequence of the variant comprises an amino acid substitution at position K387 and at position N437.

5. The RNA polymerase variant of any one of paragraphs 1-4, wherein the amino acid substitution at position K387 is a polar, neutral amino acid.

6. The RNA polymerase variant of paragraph 5, wherein the polar, neutral amino acid is selected from asparagine (N), cysteine (C), glutamine (Q), methionine (M), serine (S), and threonine (T).

7. The RNA polymerase variant of paragraph 6, wherein the polar, neutral amino acid is asparagine (K387N).

8. The RNA polymerase variant of paragraph 6, wherein the polar, neutral amino acid is cysteine (K387C).

9. The RNA polymerase variant of paragraph 6, wherein the polar, neutral amino acid is glutamine (K387Q).

10. The RNA polymerase variant of paragraph 6, wherein the polar, neutral amino acid is methionine (K387M).

11. The RNA polymerase variant of paragraph 6, wherein the polar, neutral amino acid is serine (K387S).

12. The RNA polymerase variant of paragraph 6, wherein the polar, neutral amino acid is threonine (K387T).

13. The RNA polymerase variant of any one of paragraphs 1-12, wherein the amino acid substitution at position N437 is an aromatic amino acid.

14. The RNA polymerase variant of paragraph 13, wherein the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F).

15. The RNA polymerase variant of paragraph 14, wherein the aromatic amino acid is tryptophan (N437W).

16. The RNA polymerase variant of paragraph 14, wherein the aromatic amino acid is tyrosine (N437Y).

17. The RNA polymerase variant of paragraph 14, wherein the aromatic amino acid is phenylalanine (N437F).

18. A ribonucleic acid (RNA) polymerase variant comprising an amino acid sequence that comprises (i) an amino acid substitution at position E350, (ii) an amino acid substitution at D351, and (iii) an amino acid substitution at position D653, relative to a wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

19. The RNA polymerase variant of paragraph 18, wherein the amino acid substitution at position D653 is an aromatic amino acid.

20. The RNA polymerase variant of paragraph 19, wherein the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F).

21. The RNA polymerase variant of paragraph 20, wherein the aromatic amino acid is tryptophan (D653W).

22. The RNA polymerase variant of paragraph 20, wherein the aromatic amino acid is tyrosine (D653Y).

23. The RNA polymerase variant of paragraph 20, wherein the aromatic amino acid is phenylalanine (D653F).

24. The RNA polymerase variant of any one of paragraphs 1-23, wherein the amino acid substitution at position E350 is an aromatic amino acid.

25. The RNA polymerase variant of paragraph 24, wherein the aromatic amino acid is selected from tryptophan (W), tyrosine (Y), and phenylalanine (F).

26. The RNA polymerase variant of paragraph 25, wherein the aromatic amino acid is tryptophan (E350W).

27. The RNA polymerase variant of paragraph 25, wherein the aromatic amino acid is tyrosine (E350Y).

28. The RNA polymerase variant of paragraph 25, wherein the aromatic amino acid is phenylalanine (E350F).

29. The RNA polymerase variant of paragraphs 1-28, wherein the amino acid substitution at position D351 is a non-polar, aliphatic amino acid.

30. The RNA polymerase variant of paragraph 29, wherein the non-polar, aliphatic amino acid is selected from alanine (A), glycine (G), isoleucine (I), leucine (L), proline (P), and valine (V).

31. The RNA polymerase variant of paragraph 30, wherein the non-polar, aliphatic amino acid is alanine (D351A).

32. The RNA polymerase variant of paragraph 30, wherein the non-polar, aliphatic amino acid is glycine (D351G).

33. The RNA polymerase variant of paragraph 30, wherein the non-polar, aliphatic amino acid is isoleucine (D351I).

34. The RNA polymerase variant of paragraph 30, wherein the non-polar, aliphatic amino acid is leucine (D351L).

35. The RNA polymerase variant of paragraph 30, wherein the non-polar, aliphatic amino acid is proline (D351P).

5 36. The RNA polymerase variant of paragraph 30, wherein the non-polar, aliphatic amino acid is valine (D351V).

37. An RNA polymerase variant comprising: an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO: 1, wherein the amino acid sequence of the variant comprises (i) an amino acid substitution at position E350, (ii) an amino acid substitution at D351, and (iii) an amino acid substitution at position K387, position N437, or at position K387 and position N437, relative to a wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

38. The RNA polymerase variant of paragraph 37, wherein the amino acid sequence has at least 75%, at least 80%, at least 85%, at least 95%, or at least 98% identity to the amino acid sequence of SEQ ID NO: 1.

39. The RNA polymerase variant of paragraph 37 or 38, wherein the amino acid sequence of the variant comprises an amino acid substitution at position K387.

40. The RNA polymerase variant of paragraph 37 or 38, wherein the amino acid sequence of the variant comprises an amino acid substitution at position N437.

20 41. The RNA polymerase variant of paragraph 37 or 38, wherein the amino acid sequence of the variant comprises an amino acid substitution at position K387 and at position N437.

42. The RNA polymerase variant of any one of paragraphs 37-41, wherein the amino acid substitution at position K387 is a polar, neutral amino acid.

43. The RNA polymerase variant of paragraph 42, wherein the polar, neutral amino acid is selected from asparagine (K387N), cysteine (K387C), glutamine (K387Q), methionine (K387M), serine (K387S), and threonine (K387T).

44. The RNA polymerase variant of any one of paragraphs 37-42, wherein the amino acid substitution at position N437 is an aromatic amino acid.

45. The RNA polymerase variant of paragraph 44, wherein the aromatic amino acid is selected from tryptophan (N437W), tyrosine (N437Y), and phenylalanine (N437F).

30 46. An RNA polymerase variant comprising: an amino acid sequence having at least 70% identity to the amino acid sequence of SEQ ID NO: 1, wherein the amino acid sequence of the variant comprises (i) an amino acid substitution at position E350, (ii) an amino acid substitution at D351, and (iii) an amino acid substitution at position D653, relative to a wild-type T7 RNA polymerase comprising the amino acid sequence of SEQ ID NO: 1.

35

47. The RNA polymerase variant of paragraph 46, wherein the amino acid sequence has at least 75%, at least 80%, at least 85%, at least 95%, or at least 98% identity to the amino acid sequence of SEQ ID NO: 1.

48. The RNA polymerase variant of paragraph 46 or 47, wherein the amino acid substitution at position D653 is an aromatic amino acid.

49. The RNA polymerase variant of paragraph 48, wherein the aromatic amino acid is selected from tryptophan (D653W), tyrosine (D653Y), and phenylalanine (D653F).

50. The RNA polymerase variant of any one of paragraphs 37-49, wherein the amino acid substitution at position E350 is an aromatic amino acid.

51. The RNA polymerase variant of paragraph 50, wherein the aromatic amino acid is selected from tryptophan (E350W), tyrosine (E350Y), and phenylalanine (E350F).

52. The RNA polymerase variant of any one of paragraphs 37-51, wherein the amino acid substitution at position D351 is a non-polar, aliphatic amino acid.

53. The RNA polymerase variant of paragraph 52, wherein the non-polar, aliphatic amino acid is selected from alanine (D351A), glycine (D351G), isoleucine (D351I), leucine (D351L), proline (D351P), and valine (D351V).

54. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of SEQ ID NO: 2, wherein X^1 is an aromatic amino acid, optionally selected from W, Y, and F; X^2 is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; X^3 is a polar, neutral amino acid, optionally selected from N, C, Q, M, S, and T; and X^4 is an aromatic amino acid, optionally selected from W, Y, and F.

55. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of SEQ ID NO: 6.

56. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of SEQ ID NO: 3, wherein X^1 is an aromatic amino acid, optionally selected from W, Y, and F; X^2 is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X^4 is an aromatic amino acid, optionally selected from W, Y, and F.

57. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of SEQ ID NO: 7.

58. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of SEQ ID NO: 4, wherein X^1 is an aromatic amino acid, optionally selected from W, Y, and F; X^2 is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X^3 is a polar, neutral amino acid, optionally selected from N, C, Q, M, S, and T.

59. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of SEQ ID NO: 8.

60. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of SEQ ID NO: 5, wherein X^1 is an aromatic amino acid, optionally selected from W, Y, and F; X^2 is selected from a non-polar, aliphatic amino acid, optionally selected from A, G, I, L, P, and V; and X^5 is an aromatic amino acid, optionally selected from W, Y, and F.

61. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of SEQ ID NO: 9.

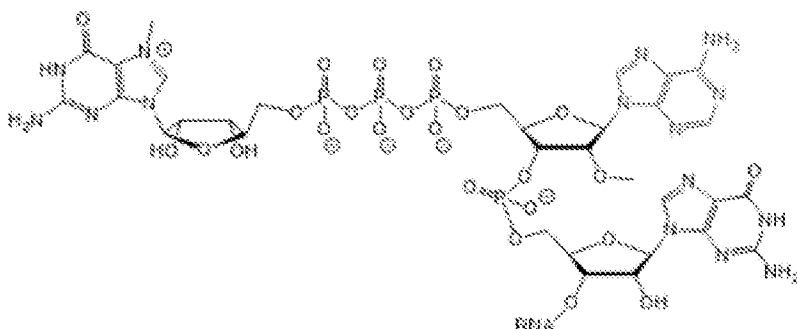
62. A method comprising: producing a messenger RNA (mRNA) in an *in vitro* transcription reaction that comprises a DNA, nucleoside triphosphates, the RNA polymerase variant of any one of paragraphs 1-53, and optionally a cap analog.

63. The method of paragraph 62, wherein the reaction comprises the cap analog.

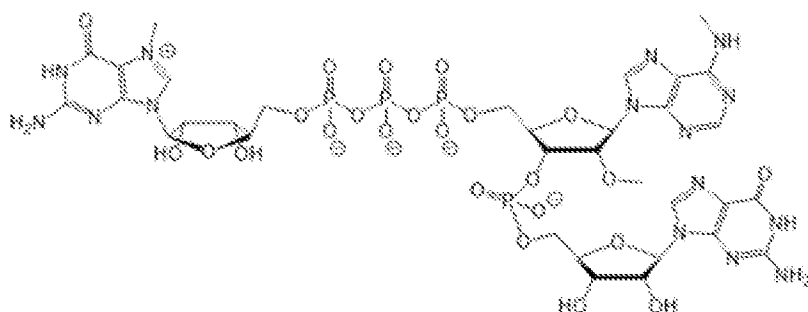
64. The method of paragraph 62 or 63, wherein the cap analog is a dinucleotide cap analog, a trinucleotide cap analog, or a tetranucleotide cap analog.

65. The method of paragraph 64, wherein the cap analog is a trinucleotide cap analog comprising a GAG sequence.

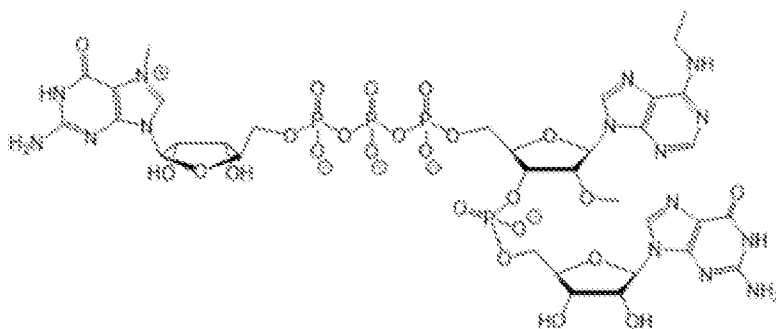
66. The method of paragraph 65, wherein the GAG cap analog comprises a compound selected from:



(X);



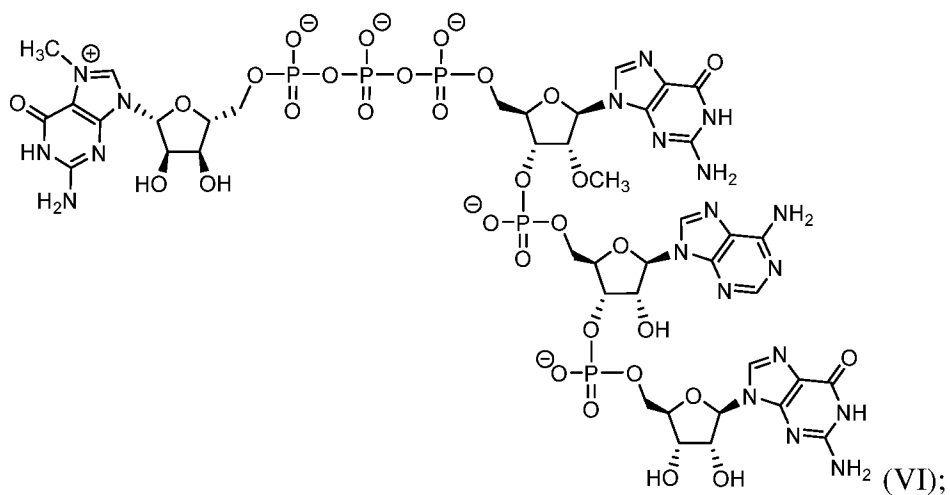
(XI); or



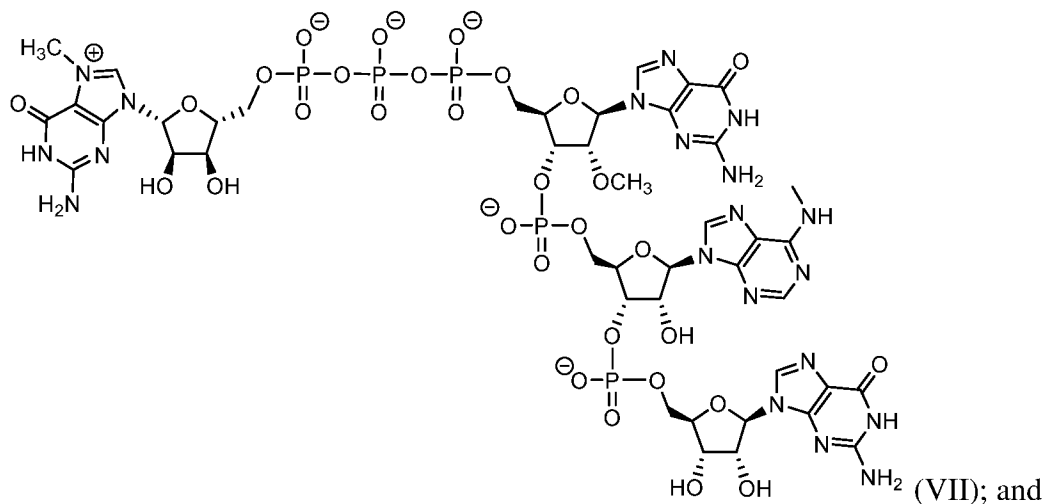
(XII).

67. The method of paragraph 64, wherein the tetranucleotide cap analog comprises a GGAG sequence.

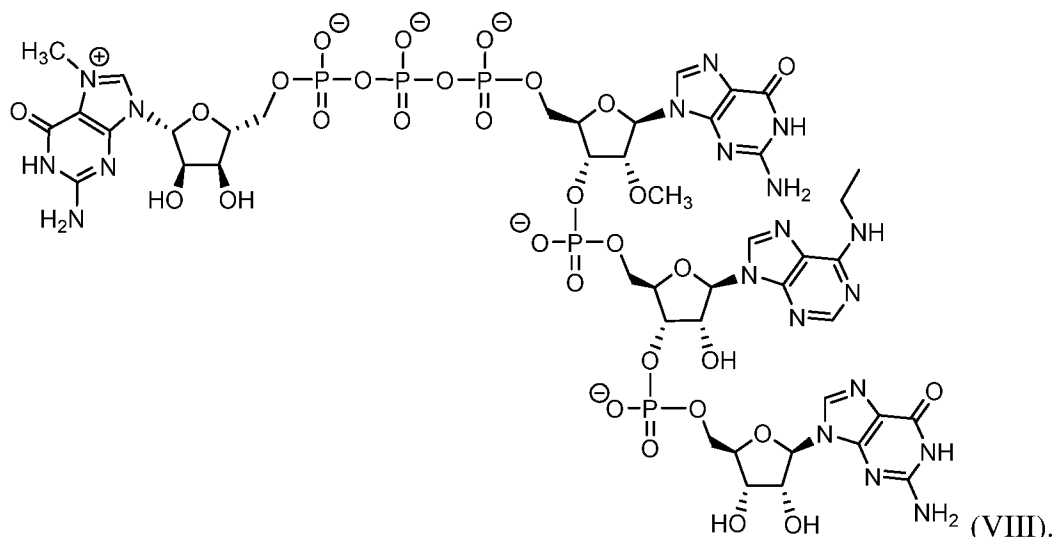
68. The method of paragraph 67, wherein the tetranucleotide cap analog comprises a compound selected from:



(VI);



(VII); and



69. The method of any one of the paragraphs 62-68, wherein the DNA includes a 2'-deoxythymidine residue or a 2'-deoxycytidine residue at position +1.

70. A composition or kit comprising the RNA polymerase variant of any one of paragraphs 1-61 and an *in vitro* transcription (IVT) reagent selected from the group consisting of a DNA, nucleoside triphosphates, and a cap analog.

71. A nucleic acid encoding the RNA polymerase variant of any one of paragraphs 1-61.

EXAMPLES

Example 1. IVT reactions using RNA polymerase variants

In vitro transcription (IVT) reactions were performed using DNA template, GGAG cap analog, and selected individual RNA polymerase variants as provided in Table 1. Specifically, RNA polymerase variants comprising N437Y, K387S, E350W, and D351V substitutions (SEQ ID NO: 6; "Variant A"); N437Y, E350W, and D351V substitutions (SEQ ID NO: 7; "Variant B"); K387S, E350W, and D351V substitutions (SEQ ID NO: 8; "Variant C"); and D653W, E350W, and D351V substitutions (SEQ ID NO: 9; "Variant D") were tested in this Example. Reactions were also performed using control T7 RNA polymerase (SEQ ID NO: 1). Following IVT reactions, transcribed RNA products from each reaction were characterized to address the quality of said RNA products, including total RNA yield, capping efficiency (percentage of total RNA comprising a GGAG cap), dsRNA contamination, and tail purity.

The overall yields of total RNA, following an oligo dT purification, were measured by UV absorption. The total RNA products were analyzed by LC-MS to determine capping efficiency (*i.e.*, percent of transcribed RNA comprising a GGAG cap). A standard ELISA was used to assess dsRNA contaminants (*e.g.*, dsRNA longer than 40 nucleotide base pairs) following IVT reactions in this Example. A Tris RP (reverse-phase) method was used to assess percent tailed RNA (*i.e.*, percent of transcribed RNA comprising a polyA tail).

Each of the tested RNA polymerase variants generated RNA in IVT reactions with at least 80% capped RNA (percentage of total RNA comprising a GGAG cap) and at least ~80% tailed RNA (*i.e.*, percent of transcribed RNA comprising a polyA tail). RNA polymerase variants comprising N437Y, E350W, and D351V substitutions (SEQ ID NO: 7); and K387S, E350W, and D351V substitutions (SEQ ID NO: 8) generated less than 0.007% dsRNA (w:w). Further, the yields of total RNA for each of the tested RNA polymerase variants (greater than 8 mg/mL) was comparable to control T7 RNA polymerase.

Each of the tested RNA polymerase variants performed comparably or better than the control T7 RNA polymerase across each of the tested characteristics. Specifically, N437Y + K387S + E350W + D351V provided RNA with higher capping efficiency (~85% capped RNA), similar yield and similar tailed purity relative to control T7 RNA polymerase. N437Y + E350W + D351V provided RNA with higher capping efficiency (~80% capped RNA), similar yield, similar tailed purity, and similar dsRNA contamination relative to control T7 RNA polymerase. K387S + E350W + D351V provided RNA with higher capping efficiency (~83% capped RNA), similar yield, higher tailed purity (~85% tailed RNA), and less dsRNA contamination (0.00327 dsRNA wt:wt) relative to control T7 RNA polymerase. D653W + E350W + D351V provided RNA with higher capping efficiency (~95% capped RNA) relative to control T7 RNA polymerase.

Data for each of tested RNA polymerases is provided in Table 3 and FIGs. 1A-1D.

Table 3. Characteristics of RNA produced by RNA polymerases used in Example 1

	Yield (mg/ml)	Percent Capping	Percent tailed	Percent dsRNA (w:w)
Control RNAP	9.34	69.51	83.17	0.00624
N437Y + K387S + E350W + D351V	11.14	84.95	79.51	0.01583
N437Y + E350W + D351V	9.56	80.19	80.87	0.006955
K387S + E350W + D351V	9.20	82.88	85.00	0.00327
D653W + E350W + D351V	8.87	94.47	80.30	0.02815

Example 2. RNA polymerase variants produce RNA products with high levels of capping efficiency at low concentrations of GGAG cap analog

In vitro transcription reactions were performed using DNA template, equimolar NTPs, a variable amount of GGAG tetranucleotide cap analog (0.25 mM, 0.5 mM, 0.75 mM, 1 mM, 1.25 mM, 1.5 mM, 3 mM) and T7 RNA polymerase. RNA polymerase variants comprising N437Y, K387S, E350W, and D351V substitutions (SEQ ID NO: 6); N437Y, E350W, and D351V substitutions (SEQ ID NO: 7); K387S, E350W, and D351V substitutions (SEQ ID NO: 8); and

D653W, E350W, and D351V substitutions (SEQ ID NO: 9) were tested in this Example.

Reactions were also performed using control T7 RNA polymerase.

Following the IVT reaction, mRNA products were oligo-dT purified before being analyzed by LC-MS to determine the % capped RNA (*i.e.*, percent of transcribed RNA comprising a cap), by HPLC to determine the RNA yield of the reaction, and by Tris RP (reverse-phase) method to determine percent tailed RNA.

Each of the tested RNA polymerase variants produced RNA with percent capped RNA at higher levels than the control polymerase variant in the presence of GGAG cap analog, regardless of the concentration of the GGAG analog (FIG. 2A). Even at the lowest tested concentrations of GGAG cap analog (0.25 mM), all of tested variants produced at least 50% capped RNA, considerably higher than the ~25% capped RNA produced by the control polymerase variant. At 1.5 mM GGAG cap analog, all of tested variants produced about 80-95% capped RNA.

Each of the tested variants produced RNA with comparable yield (FIG. 2B) and percent tailed RNA (FIG. 2C) relative to the control polymerase variant.

These data demonstrate that each of tested RNA polymerase variants (*i.e.*, N437Y + K387S + E350W + D351V; N437Y + E350W + D351V; K387S + E350W + D351V; and D653W + E350W + D351V) are capable of producing RNA with higher capping efficiency than control T7 RNA polymerase without giving up any yield or tailed content.

EQUIVALENTS AND SCOPE

All references, patents and patent applications disclosed herein are incorporated by reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

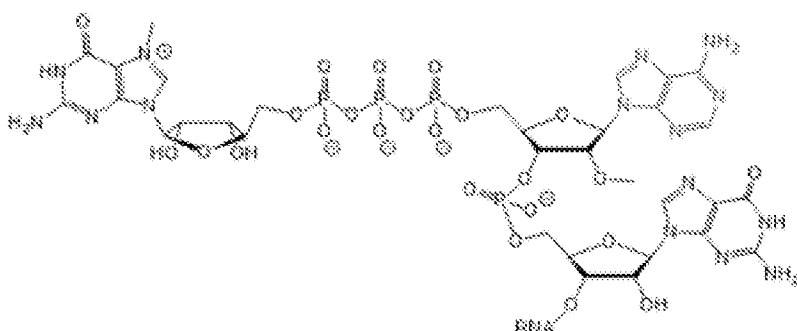
In the claims, as well as in the specification above, all transitional phrases such as “comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,” “composed of,” and the like are to be understood to be open-ended, *i.e.*, to mean including but not limited to. Only the transitional phrases “consisting of” and “consisting essentially of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

CLAIMS

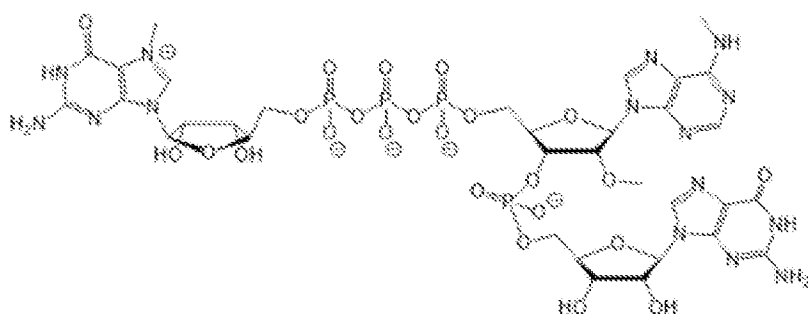
What is claimed is:

1. A ribonucleic acid (RNA) polymerase variant comprising an amino acid sequence having
5 at least 90% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 2-9,
wherein the amino acid sequence comprises an amino acid substitution at position D351 and at
least two additional amino acid substitutions, relative to a RNA polymerase comprising the
amino acid sequence of SEQ ID NO: 1.
- 10 2. The RNA polymerase variant of claim 1, comprising an amino acid sequence having at
least 95% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 2-9.
3. The RNA polymerase variant of claim 1 or 2, comprising an amino acid sequence
comprising the amino acid sequence of any one of SEQ ID NOs: 2-9.
- 15 4. The RNA polymerase variant of any one of claims 1-3, comprising at least one, at least
two, at least three, or at least four amino acid substitutions, relative to a wild-type T7 RNA
polymerase comprising the amino acid sequence of SEQ ID NO: 1.
- 20 5. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of
SEQ ID NO: 6.
6. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of
SEQ ID NO: 7.
- 25 7. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of
SEQ ID NO: 8.
8. A ribonucleic acid (RNA) polymerase variant comprising the amino acid sequence of
30 SEQ ID NO: 9.
9. A method comprising: producing a messenger RNA (mRNA) in an *in vitro* transcription
reaction that comprises a DNA, nucleoside triphosphates, the RNA polymerase variant of any
one of claims 1-8.

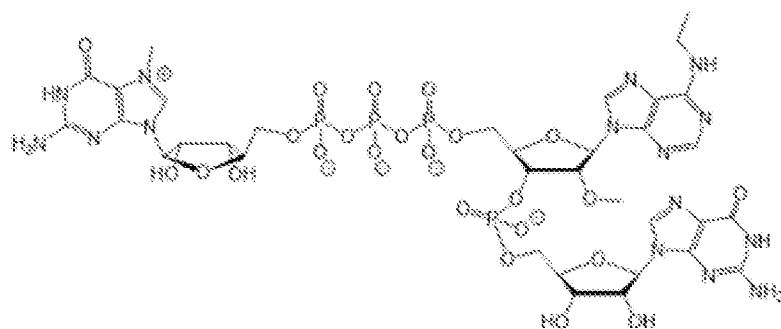
10. The method of claim 9, wherein the reaction further comprises a cap analog.
11. The method of claim 10, wherein the cap analog is a dinucleotide cap analog, a trinucleotide cap analog, or a tetranucleotide cap analog.
12. The method of claim 11, wherein the cap analog is a trinucleotide cap analog comprising a GAG sequence.
13. The method of claim 11, wherein the cap analog comprises a compound selected from:



(X);



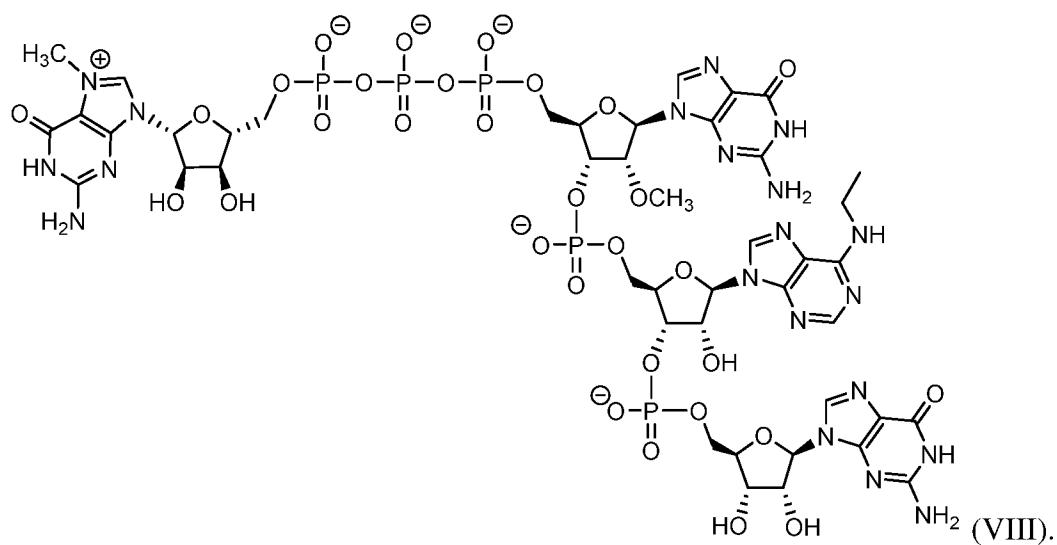
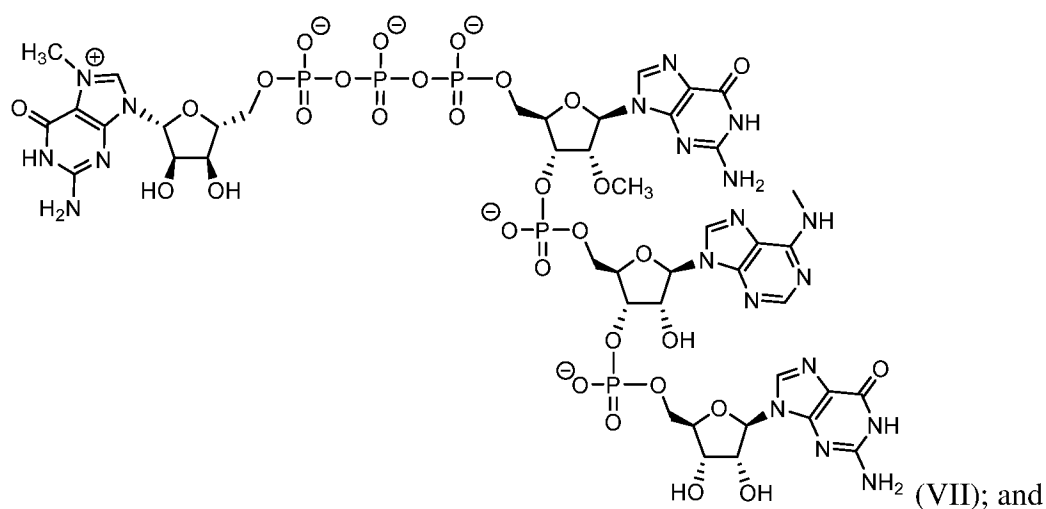
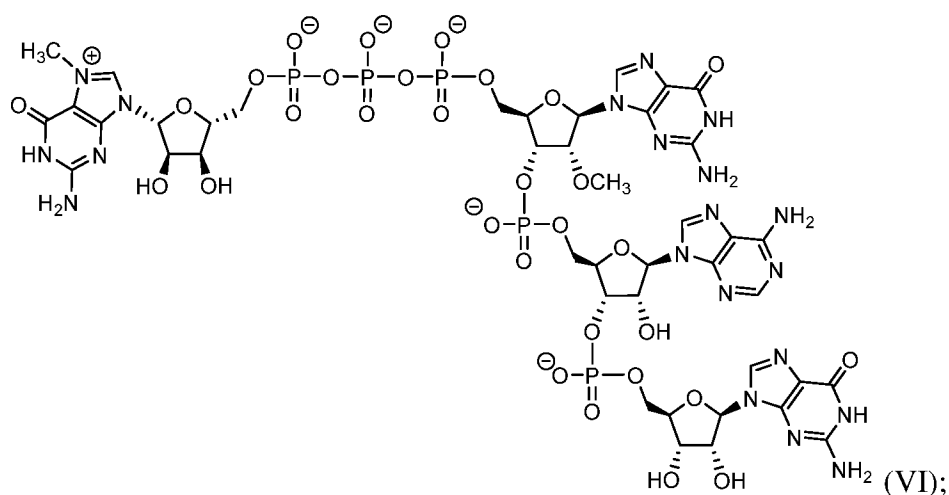
(XI); or



(XII).

14. The method of claim 11, wherein the tetranucleotide cap analog comprises a GGAG sequence.

15. The method of claim 11, wherein the tetranucleotide cap analog comprises a compound selected from:



16. The method of any one of the claims 9-15, wherein the DNA includes a 2'-deoxythymidine residue or a 2'-deoxycytidine residue at position +1.

17. A composition or kit comprising the RNA polymerase variant of any one of claims 1-9 and an *in vitro* transcription (IVT) reagent selected from the group consisting of a DNA, nucleoside triphosphates, and a cap analog.

5 18. A nucleic acid encoding the RNA polymerase variant of any one of claims 1-9.

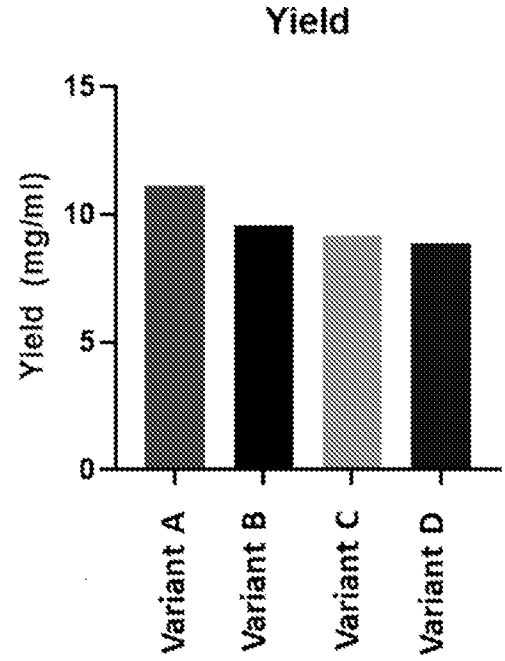


FIG. 1A

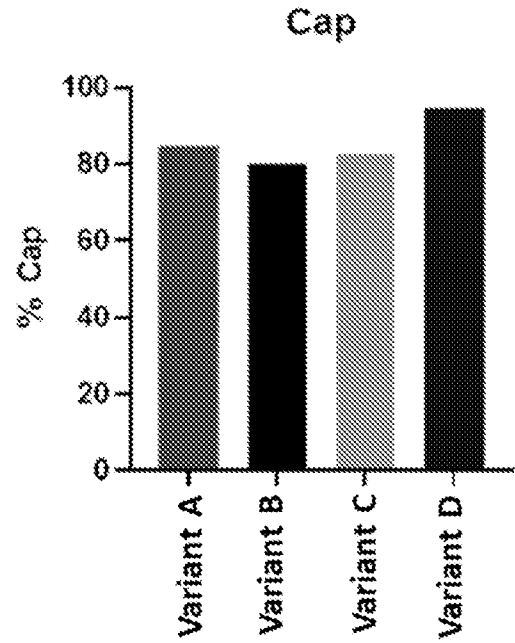


FIG. 1B

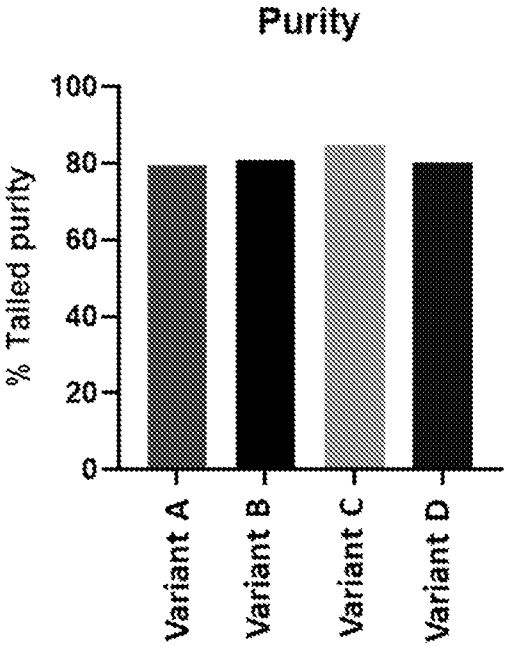


FIG. 1C

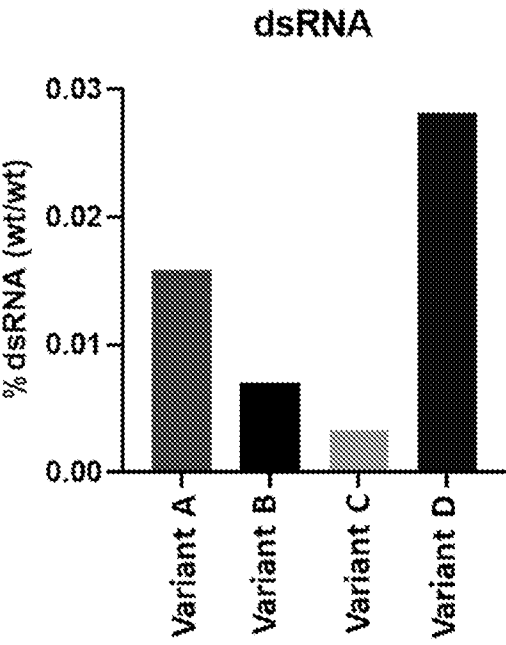


FIG. 1D

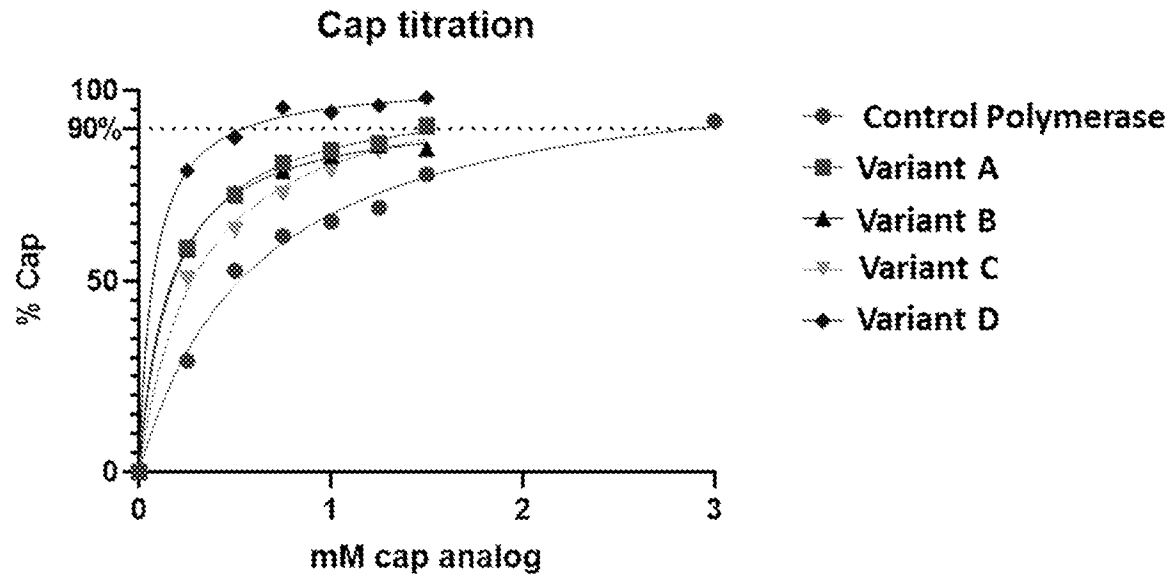


FIG. 2A

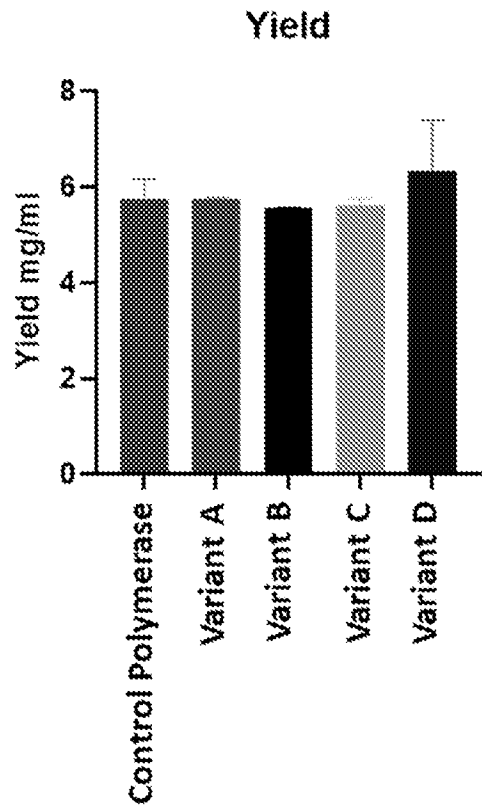


FIG. 2B

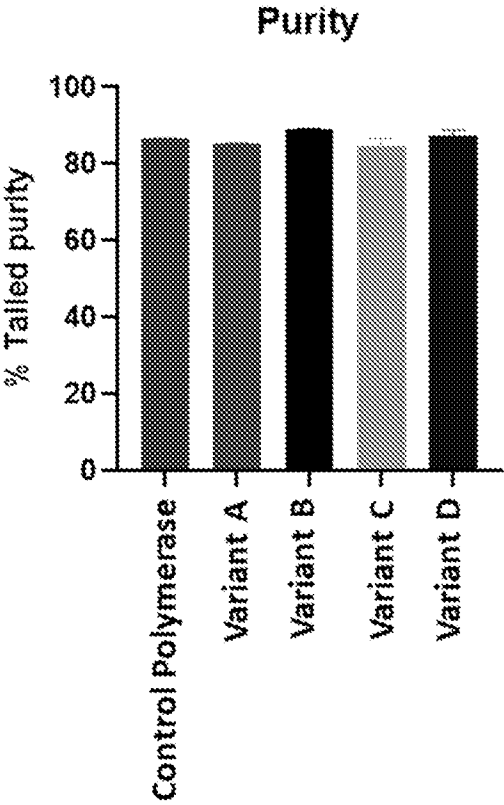


FIG. 2C

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2023/065719

A. CLASSIFICATION OF SUBJECT MATTER**INV. C12N9/12****ADD.**

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N

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, BIOSIS, EMBASE, WPI Data, Sequence Search**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2020/172239 A1 (MODERNATX INC [US]) 27 August 2020 (2020-08-27) example 1 claims 1-52 sequences 127,157,240 & DATABASE Geneseq [Online] 15 October 2020 (2020-10-15), "T7 RNA polymerase mutant G47A/D351V/E350W/K387S/G884, SEQ 127.", retrieved from EBI accession no. GSP:BIF46398 Database accession no. BIF46398 abstract -/--	1-18



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

16 June 2023

Date of mailing of the international search report

30/06/2023

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

Tudor, Mark

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2023/065719

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	& DATABASE Geneseq [Online] 15 October 2020 (2020-10-15), "T7 RNA polymerase mutant G47A/N437Y/G884, SEQ 157.", retrieved from EBI accession no. GSP:BIF46428 Database accession no. BIF46428 abstract & DATABASE Geneseq [Online] 15 October 2020 (2020-10-15), "T7 RNA polymerase mutant D351V/D653T, SEQ 240.", retrieved from EBI accession no. GSP:BIF46511 Database accession no. BIF46511 abstract -----	1-18
	WO 2011/128444 A2 (EUKARYS [FR]; JAIS PHILIPPE [FR]) 20 October 2011 (2011-10-20) claim 1 sequence 40 -----	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2023/065719

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2023/065719

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2020172239 A1	27-08-2020	AU 2020224103 A1	16-09-2021
		CA 3130888 A1	27-08-2020
		CN 113795579 A	14-12-2021
		EP 3927821 A1	29-12-2021
		JP 2022521094 A	05-04-2022
		KR 20210142636 A	25-11-2021
		MA 55037 A	29-12-2021
		US 2021309976 A1	07-10-2021
		US 2023104080 A1	06-04-2023
		WO 2020172239 A1	27-08-2020
WO 2011128444 A2	20-10-2011	CA 2796175 A1	20-10-2011
		DK 2558579 T3	11-11-2013
		EP 2377938 A1	19-10-2011
		EP 2558579 A2	20-02-2013
		ES 2436356 T3	30-12-2013
		IL 222357 A	29-09-2016
		JP 5432415 B2	05-03-2014
		JP 2013531467 A	08-08-2013
		KR 20130069632 A	26-06-2013
		SG 184893 A1	29-11-2012
		US 2013042334 A1	14-02-2013
		US 2015203888 A1	23-07-2015
		WO 2011128444 A2	20-10-2011