



US012390524B2

(12) **United States Patent**
Sullivan et al.

(10) **Patent No.:** US 12,390,524 B2
(45) **Date of Patent:** *Aug. 19, 2025

(54) **COMPOSITIONS AND METHODS FOR INDUCING IMMUNE RESPONSES**(71) Applicant: **Arcturus Therapeutics, Inc.**, San Diego, CA (US)(72) Inventors: **Sean Michael Sullivan**, Escondido, CA (US); **Daiki Matsuda**, San Diego, CA (US); **Kiyoshi Tachikawa**, San Diego, CA (US); **Padmanabh Chivukula**, San Diego, CA (US); **Priya Prakash Karmali**, San Diego, CA (US); **Jared Henry Davis**, Poway, CA (US); **Yanjie Bao**, San Diego, CA (US)(73) Assignee: **Arcturus Therapeutics, Inc.**, San Diego, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **18/351,392**(22) Filed: **Jul. 12, 2023**(65) **Prior Publication Data**

US 2024/0115692 A1 Apr. 11, 2024

Related U.S. Application Data

(63) Continuation of application No. 17/196,890, filed on Mar. 9, 2021, now Pat. No. 11,759,515.

(60) Provisional application No. 63/073,900, filed on Sep. 2, 2020, provisional application No. 62/987,191, filed on Mar. 9, 2020.

(51) **Int. Cl.**

A61K 39/215	(2006.01)
A61K 9/51	(2006.01)
A61K 38/00	(2006.01)
A61K 39/00	(2006.01)
A61K 39/12	(2006.01)
A61K 47/10	(2017.01)
A61K 47/20	(2006.01)
A61K 47/26	(2006.01)
C07K 14/005	(2006.01)
C07K 14/18	(2006.01)
C12N 7/00	(2006.01)
C12N 15/86	(2006.01)

(52) **U.S. Cl.**

CPC	A61K 39/215 (2013.01); A61K 9/5123 (2013.01); A61K 39/12 (2013.01); A61K 47/10 (2013.01); A61K 47/20 (2013.01); A61K 47/26 (2013.01); C07K 14/005 (2013.01); C07K 14/1808 (2013.01); C12N 7/00 (2013.01); C12N 15/86 (2013.01); A61K 38/00 (2013.01); A61K 2039/53 (2013.01); C12N 2770/20022 (2013.01); C12N 2770/20034 (2013.01); C12N 2770/36122 (2013.01); C12N 2770/36134 (2013.01); C12N 2830/42 (2013.01); C12N 2830/50 (2013.01)
-----------	--

2770/36122 (2013.01); C12N 2770/36134 (2013.01); C12N 2830/42 (2013.01); C12N 2830/50 (2013.01)

(58) **Field of Classification Search**

CPC	A61K 39/215; A61K 9/5123; A61K 39/12; A61K 47/10; A61K 47/20; A61K 47/26; A61K 38/00; A61K 2039/53; A61K 39/395; A61K 2039/507; A61K 2039/884; A61K 2039/55555; A61K 2039/572; A61K 2039/575; C07K 14/005; C07K 14/1808; C07K 2317/76; C07K 16/2818; C07K 16/2827; C12N 7/00; C12N 15/86; C12N 2770/20022; C12N 2770/20034; C12N 2770/36122; C12N 2770/36134; C12N 2830/42; C12N 2830/50; C12N 2740/13071; C12N 2760/16134; C12N 2760/16171; C12N 2740/13034; A61P 35/00; A61P 31/14; A61P 31/16; Y02A 50/30
----------	--

See application file for complete search history.

(56)

References Cited**U.S. PATENT DOCUMENTS**

7,332,322 B2	2/2008	Frolov et al.
7,425,337 B2	9/2008	Smith et al.
7,442,381 B2	10/2008	Smith et al.
8,093,367 B2	1/2012	Kore et al.
8,158,601 B2	4/2012	Chen et al.
8,304,529 B2	11/2012	Kore et al.
8,961,995 B2	2/2015	Frolov et al.
9,254,265 B2	2/2016	Geall et al.
9,295,646 B2	3/2016	Brito et al.
9,730,997 B2	8/2017	Perri et al.
9,770,463 B2	9/2017	Geall et al.

(Continued)

FOREIGN PATENT DOCUMENTS

EP	2591114 B1	6/2016
EP	3471778 A2	4/2019

(Continued)

OTHER PUBLICATIONS

Shustov AV, et. al. VEEV replicon vector YFV-C1, complete sequence. GenBank: DQ322637.1. Dep. Jan. 16, 2006. (Year: 2006).*

(Continued)

Primary Examiner — Rachel B Gill

(74) Attorney, Agent, or Firm — Greenberg Traurig, LLP

(57)

ABSTRACT

Provided herein are nucleic acid molecules encoding viral replication proteins and antigenic proteins or fragments thereof. Also provided herein are compositions that include nucleic acid molecules encoding viral replication and antigenic proteins, and lipids. Nucleic acid molecules provided herein are useful for inducing immune responses.

22 Claims, 11 Drawing Sheets**Specification includes a Sequence Listing.**

(56)	References Cited						
U.S. PATENT DOCUMENTS							
10,238,733 B2	3/2019	Brito et al.	WO	2018222926 A1	12/2018		
10,487,105 B2	11/2019	Chivukula et al.	WO	2019023566 A1	1/2019		
11,135,283 B2	10/2021	Berglund et al.	WO	2020014654 A1	1/2020		
11,744,887 B2	9/2023	Sullivan et al.	WO	2020035609 A2	2/2020		
11,759,515 B2	9/2023	Sullivan et al.	WO	2020254535 A1	12/2020		
2009/0075384 A1 *	3/2009	Kamrud	A61P 37/04	WO	2020254804 A1	12/2020	
			435/235.1	WO	2020255055 A1	12/2020	
2009/0155301 A1	6/2009	Mason et al.	WO	2021067181 A1	4/2021		
2011/0171255 A1	7/2011	Kliver et al.	WO	2021183563 A1	9/2021		
2011/0207223 A1	8/2011	Tang et al.	WO	2021183564 A1	9/2021		
2011/0256175 A1	10/2011	Hope et al.	WO	2023010128 A2	2/2023		
2012/0027803 A1	2/2012	Manoharan et al.	OTHER PUBLICATIONS				
2012/0128760 A1	5/2012	Manoharan et al.	Keyer VV, et. al. Non-structural polyprotein [Cloning vector pCMV-VEE-GFP]. GenBank: QCD25069.1, Dep. Apr. 27, 2019. (Year: 2019).*				
2012/0156251 A1	6/2012	Brito et al.	(Apr. 27, 2019) "Cloning Vector pCMV-VEE-GFP", Complete Sequence, GenBank ID: MH891622.1, 7 pages.				
2013/0171241 A1	7/2013	Geall	International Search Report and Written Opinion for Application No. PCT/US21/21572, mailed on Jul. 20, 2021, 11 pages.				
2013/0195968 A1	8/2013	Geall et al.	International Search Report and Written Opinion for Application No. PCT/US21/21573, mailed on Jul. 1, 2021, 12 pages.				
2014/0227346 A1	8/2014	Geall et al.	International Search Report and Written Opinion for Application No. PCT/US22/74337, mailed on Dec. 30, 2022, 16 pages.				
2014/0242152 A1	8/2014	Geall et al.	(Jul. 18, 2020) "Surface Glycoprotein [Severe Acute Respiratory Syndrome Coronavirus 2]", GenBank ID: YP_009724390, 3 pages.				
2015/0024002 A1	1/2015	Perri et al.	Altschul et al. (Oct. 5, 1990) "Basic Local Alignment Search Tool", Journal of Molecular Biology, 215(3):403-410.				
2016/0074500 A1	3/2016	Pushko et al.	Altschul et al. (Sep. 1, 1997) "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs", Nucleic Acids Research, 25(17):3389-3402.				
2016/0348132 A1	12/2016	Rayner et al.	Baden et al. (Feb. 4, 2021) "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine", The New England Journal of Medicine, 384(5):403-416.				
2018/0036398 A1	2/2018	Hagen et al.	Bochicchio et al. (2014) "Liposomes as siRNA Delivery Vectors", Current Drug Metabolism, 15(9):882-892.				
2018/0104359 A1	4/2018	Kamrud	Boles et al. (2017) "Synthetic Construct H7N9 HA Gene, Complete CDS", GenBank KY199425.1, National Library of Medicine, 4 pages.				
2018/0169268 A1	6/2018	Payne et al.	Both et al. (Mar. 1, 1975) "Methylation-Dependent Translation of Viral Messenger RNAs In Vitro", Proceedings of the National Academy of Sciences, 72(3): 1189-1193.				
2018/0171340 A1	6/2018	Kamrud et al.	Bouloy et al. (Jul. 1, 1980) "Both The 7-Methyl and The 2'-O-Methyl Groups in the Cap of mRNA Strongly Influence Its Ability to Act as Primer for Influenza Virus RNA Transcription", Proceedings of the National Academy of Sciences, 77(7):3952-3956.				
2018/0273576 A1	9/2018	Hogrefe et al.	Chan et al. (Sep. 19, 2016) "Cross-reactive antibodies enhance live attenuated virus infection for increased immunogenicity", Nature Microbiology, 1(12):16164 (10 pages).				
2018/0327471 A1	11/2018	Limphong et al.	Chan et al. (Oct. 5, 2017) "Early Molecular Correlates of Adverse Events Following Yellow Fever Vaccination", JCI Insight, 2(19):e96031 (12 pages).				
2019/0091329 A1	3/2019	Brito et al.	Chan et al. (Aug. 2019) "Metabolic Perturbations and Cellular Stress Underpin Susceptibility to Symptomatic Live-attenuated Yellow Fever Infection", Nature Medicine, 25(8):1218-1224 (21 pages).				
2019/0224299 A1	7/2019	Kamrud et al.	Chu et al. (Aug. 1978) "Paradoxical Observations on the 5' Terminus of Ovalbumin Messenger Ribonucleic Acid", Journal of Biological Chemistry, 253(15):5228-5231.				
2019/0321458 A1	10/2019	Sahin et al.	Cirelli et al. (May 16, 2019) "Slow Delivery Immunization Enhances Hiv Neutralizing Antibody and Germinal Center Responses via Modulation of Immunodominance", Cell, 177(5):1153-1171.e28 (57 pages).				
2019/0374650 A1	12/2019	Moon et al.	Conticello et al. (Aug. 22, 2008) "Interaction Between Antibody-diversification Enzyme Aid and Spliceosome-associated Factor CTNNBL1", Molecular Cell, 31(4):474-484.				
2020/0010849 A1	1/2020	Blair et al.	Corbett et al. (Oct. 22, 2020) "SARS-CoV-2 mRNA Vaccine Design Enabled by Prototype Pathogen Preparedness", Nature, 586(7830):567-571.				
2020/0113830 A1	4/2020	Geall et al.	Corbett et al. (Jun. 11, 2020) "SARS-CoV-2 mRNA Vaccine Development Enabled by Prototype Pathogen Preparedness", bioRxiv., 39 pages.				
2020/0113831 A1	4/2020	Geall et al.					
2020/0222332 A1	7/2020	Irvine et al.					
2020/0230058 A1	7/2020	Geall et al.					
2020/0230225 A1 *	7/2020	Vogels	C12N 15/86				
2020/0297634 A1	9/2020	Karmali et al.					
2020/0330585 A1	10/2020	Mogler et al.					
2021/0030859 A1	2/2021	Bucala et al.					
2021/0284974 A1	9/2021	Chivukula et al.					
2021/0290752 A1	9/2021	Sullivan et al.					
2021/0290756 A1	9/2021	Sullivan et al.					
2022/0347298 A1	11/2022	Sullivan et al.					
2022/0395570 A1 *	12/2022	Rauch	A61K 47/543				
2022/0401550 A1	12/2022	Simon-loriere et al.					
2023/0219996 A1	7/2023	Matsuda et al.					
2024/0115691 A1	4/2024	Sullivan et al.					
FOREIGN PATENT DOCUMENTS							
EP	3433369 B1	3/2020					
EP	2729126 B1	12/2020					
WO	2008119827 A1	10/2008					
WO	2009079185 A2	6/2009					
WO	2009086558 A1	7/2009					
WO	2009127060 A1	10/2009					
WO	2010048536 A2	4/2010					
WO	2010054406 A1	5/2010					
WO	2010088537 A2	8/2010					
WO	2010129709 A1	11/2010					
WO	2011153493 A2	12/2011					
WO	2012006369 A2	1/2012					
WO	2012006380 A2	1/2012					
WO	2012170431 A2	12/2012					
WO	2014170493 A2	10/2014					
WO	2015051169 A2	4/2015					
WO	2015061491 A1	4/2015					
WO	2016184822 A1	11/2016					
WO	2017083356 A1	5/2017					
WO	2017223085 A2	12/2017					
WO	2018078053 A1	5/2018					
WO	2018208856 A1	11/2018					
WO	2018222890 A1	12/2018					

(56)

References Cited**OTHER PUBLICATIONS**

- Dabkowska et al. (Mar. 7, 2012) "The Effect of Neutral Helper Lipids on the Structure of Cationic Lipid Monolayers", *Journal of the Royal Society Interface*, 9(68):548-561.
- Dua et al. (Apr.-Jun. 2012) "Liposome: Methods of Preparation and Applications", *International Journal of Pharmaceutical Studies and Research*, 3(3):14-20.
- Dupuis et al. (Sep. 1, 2000) "Distribution of Dna Vaccines Determines Their Immunogenicity After Intramuscular Injection in Mice", *The Journal of Immunology*, 165(5):2850-2858.
- Ehrchen et al. (Sep. 2009) "The Endogenous Toll-like Receptor 4 Agonist S1OOA8/S1OOA9 (Calprotectin) as Innate Amplifier of Infection, Autoimmunity, and Cancer", *Journal of Leukocyte Biology*, 86(3):557-566.
- Enright et al. (Dec. 12, 2003) "MicroRNA targets in Drosophil", *Genome Biology*, 5:R1 (14 pages).
- Geall et al. (Sep. 4, 2012) "Nonviral Delivery of Self-amplifying Rna Vaccines", *Proceedings of the National Academy of Sciences*, 109(36):14604-14609.
- Groom et al. (Mar. 10, 2011) "CXCR3 in T Cell Function", *Experimental Cell Research*, 317(5):620-631.
- Gustafsson et al. (Jul. 2004) "Codon Bias and Heterologous Protein Expression", *Trends in Biotechnoloov*, 22(7):346-353.
- Hashem et al. (Oct. 8, 2019) "A Highly Lmmunogenic, Protective, and Safe Adenovirus-based Vaccine Expressing Middle East Respiratory Syndrome Coronavirus S1-cd40I Fusion Protein in a Transgenic Human Diptidyl Peptidase 4 Mouse Model", *The Journal of Infectious Diseases*, 220(10):1558-1567.
- Hassett et al. (Apr. 15, 2019) "Optimization of Lipid Nanoparticles for Intramuscular Administration of mRNA Vaccines", *Molecular Therapy—Nucleic Acids*, 15:1-11.
- Higgins et al. (Apr. 2019) "Programming Isotype-specific Plasma Cell Function", *Trends Immunology*, 40(4):345-357.
- Honda-Okubo et al. (Mar. 2015) "Severe Acute Respiratory Syndrome-associated Coronavirus Vaccines Formulated with Delta Inulin Adjuvants Provide Enhanced Protection While Ameliorating Lung Eosinophilic Immunopathology", *Journal of Virological Methods*, 89(6):2995-3007.
- Hsieh et al. (Sep. 18, 2020) "Structure-based design of prefusion-stabilized SARS-CoV-2 spikes", *Science*, 369(6510):1501-1505 (10 pages).
- Huang et al. (Aug. 15, 2011) "In Vivo Delivery of RNAi with Lipid-Based Nanoparticles", *Annual Review of Biomedical Engineering*, 13:507-530.
- Hyde et al. (Aug. 3, 2015) "The 5' and 3' Ends of Alphavirus RNAs-non-coding Is Not Non-functional", *Virus Research*, 206:99-107 (8 pages).
- Ishikawa et al. (Sep. 27, 2009) "Preparation of Eukaryotic mRNA having Differently Methylated Adenosine at the 5'-Terminus and the Effect of the Methyl Group in Translation", *Nucleic Acids Symposium*, 53(1):129-130.
- Jackson et al. (Feb. 4, 2020) "The Promise of mRNA Vaccines: a Biotech and Industrial Perspective", *NPJ Vaccines*, 5:11 (6 pages).
- Jin et al. (Jul. 5, 2010) "Immunomodulatory Effects of dsRNA and Its Potential as Vaccine Adjuvant", *Journal of Biomedicine and Biotechnology*, 2010:690438.
- Jokerst et al. (Jun. 2011) "Nanoparticle PEGylation for Imaging and Therapy", *Nanomedicine (Lond)*, 6(4):715-728.
- Kalnins et al. (2021) "Immunogenicity and efficacy of mRNA COVID-19 vaccine MRT5500 in preclinical animal models", *NPJ Vaccines*, 6(61):12 pages.
- Karlin et al. (Jun. 1993) "Applications and Statistics for Multiple High-Scoring Segments in Molecular Sequences", *Proceedings of the National Academy of Sciences*, 90(12):5873-5877.
- Karlin et al. (Mar. 1990) "Methods for Assessing the Statistical Significance of Molecular Sequence Features by Using General Scoring Schemes", *Proceedings of the National Academy of Sciences*, 87(6):2264-2268.
- Kasturi et al. (Feb. 24, 2011) "Programming the Magnitude and Persistence of Antibody Responses with Innate Immunity", *Nature*, 470(7335):543-547 (20 pages).
- Kawabata et al. (1995) "The Fate of Plasmid DNA After Intravenous Injection in Mice: Involvement of Scavenger Receptors in Its Hepatic Uptake", *Pharmaceutical Research*, 12:825-830.
- Keech et al. (Dec. 10, 2020) "Phase 1-2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine", *The New England Journal of Medicine*, 383:2320-2332.
- Kirchdoerfer et al. (Oct. 24, 2018) "Stabilized Coronavirus Spikes Are Resistant to Conformational Changes Induced by Receptor Recognition or Proteolysis", *Science Reports*. 8(1):15701 (11 pages).
- Kowalski et al. (Apr. 2019) "Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery", *Molecular Therapy*, 27(4):710-728.
- Kozak Marilyn. (Nov. 1990) "Downstream Secondary Structure Facilitates Recognition of Initiator Codons by Eukaryotic Ribosomes", *Proceedings of the National Academy of Sciences*, 87(21):8301-8305.
- Kozak Marilyn. (Jul. 1988) "Leader Length and Secondary Structure Modulate mRNA Function Under Conditions of Stress.", *Molecular and Cellular Biology*, 8(7):2737-2744.
- Kozak Marilyn. (Oct. 25, 1991) "Structural Features In Eukaryotic mRNAs That Modulate The Initiation of Translation", *Journal of Biological Chemistry*, 266(30):19867-19870.
- Kozak Marilyn. (Feb. 1989) "The Scanning Model for Translation: An Update", *Journal of Cell Biology*, 108(2):229-241.
- Kreiter et al. (Jan. 1, 2008) "Increased Antigen Presentation Efficiency by Coupling Antigens to MHC Class I Trafficking Signals", *Journal of Immunology*, 180(1):309-318.
- Kulasegaran-Shylini et al. (Apr. 25, 2009) "The 5'UTR-specific Mutation in VEEV TC-83 Genome Has a Strong Effect on RNA Replication and Subgenomic RNA Synthesis, but Not on Translation of the Encoded Proteins", *Virology*, 387(1):211-221 (24 pages).
- Kulkarni et al. (Jun. 2018) "Lipid Nanoparticles Enabling Gene Therapies: From Concepts to Clinical Utility", *Nucleic Acid Therapeutics*, 28(3):146-157.
- Lasic Dand. (Jul. 1, 1998) "Novel Applications of Liposomes", *Trends in Biotechnology*, 16(7):307-321.
- Li et al. (Jan. 2011) "Biosynthesis of Nanoparticles by Microorganisms and Their Applications", *Journal of Nanomaterials*, Article ID 270974, 2011:17 Pages.
- Li et al. (Aug. 3, 2010) "Stealth Nanoparticles: High Density but Sheddable PEG is a Key for Tumor Targeting", *Journal of Controlled Release*, 145(3):178-181.
- Lin et al. (Jan. 2014) "Lipid-based Nanoparticles in the Systemic Delivery of siRNA", *Nanomedicine*, 9(1):105-120.
- Love et al. (Feb. 2, 2010) "Lipid-like materials for low-dose, In Vivo Gene Silencing", *Proceedings of the National Academy of Sciences*, 107(5):1864-1869.
- Magini et al. (Aug. 15, 2016) "Self-amplifying mRNA Vaccines Expressing Multiple Conserved Influenza Antigens Confer Protection Against Homologous and Heterosubtypic Viral Challenge", *PLoS one*, 11(8):e0161193 (25 pages).
- Maruggi et al. (Dec. 2013) "Engineered Alphavirus Replicon Vaccines Based on Known Attenuated Viral Mutants Show Limited Effects on Immunogenicity", *Virology*, 447(1-2):254-264.
- Maruggi et al. (Apr. 10, 2019) "mRNA as a Transformative Technology for Vaccine Development to control Infectious Diseases", *Molecular Therapy*, 27(4):757-772.
- Muthukrishnan et al. (May 1, 1975) "5'-Terminal 7-Methylguanosine in Eukaryotic mRNA is Required for Translation", *Nature*, 255:33-37.
- Olmedillas et al. (May 6, 2021) "Structure-based design of a highly stable, covalently-linked SARS-CoV-2 spike trimer with improved structural properties and immunogenicity", *bioRxiv*, 51 pages.
- Patil et al. (Jan. 2014) "Novel Methods for Liposome Preparation", *Chemistry and Physics of Lipids*, 177:8-18.
- Pearson et al. (2013) "An Introduction to Sequence Similarity ("Homology") Searching", *Current Protocols in Bioinformatics Book*, 42:3.1.1-3.1.8.

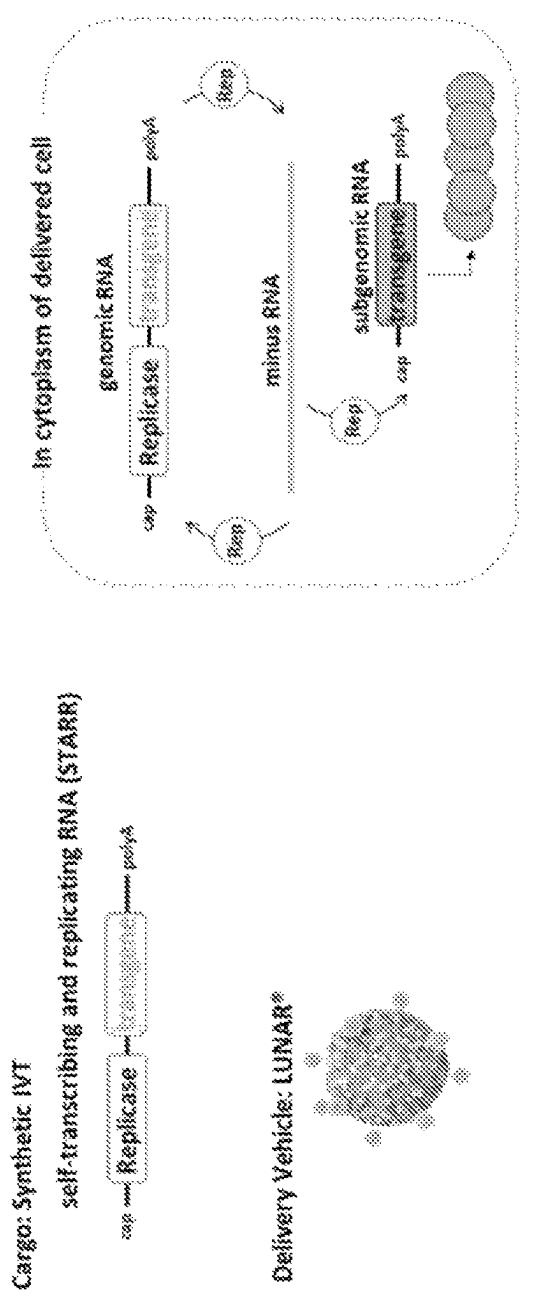
(56)

References Cited**OTHER PUBLICATIONS**

- Pearson et al. (Apr. 1988) "Improved Tools for Biological Sequence Comparison", *Proceedings of the National Academy of Sciences*, 85:2444-2448.
- Pepini et al. (May 15, 2017) "Induction of an IFN-mediated Antiviral Response by a Self-amplifying RNA Vaccine: Implications for Vaccine Design", *The Journal of Immunology*, 198(10):4012-4024 (13 pages).
- Petkov et al. (Jun. 4, 2018) "DNA Immunization Site Determines the Level of Gene Expression and the Magnitude, but Not the Type of the Induced Immune Response", *PLoS One*, 13(6): e0197902 (22 pages).
- Polack et al. (Dec. 31, 2020) "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine", *The New England Journal of Medicine*, 383:2603-2615.
- Querec et al. (Jan. 2009) "Systems Biology Approach Predicts Immunogenicity of the Yellow Fever Vaccine in Humans", *National Immunology*, 10(1):116-125 (26 pages).
- Querec et al. (2007) "Understanding the Role of Innate Immunity in the Mechanism of Action of the Live Attenuated Yellow Fever Vaccine 17D", *Advances in Experimental Medicine and Biology*, 590:43-53.
- Ramanathan et al. (Sep. 19, 2016) "mRNA Capping: Biological Functions and Applications", *Nucleic Acids Research*, 44(16):7511-7526.
- Rodriguez-Gascon et al., (Apr. 10, 2014) "Development of Nucleic Acid Vaccines: Use of Self-Amplifying RNA in Lipid Nanoparticles", *International Journal of Nanomedicine*, 9:1833-1843.
- Sahin et al. (Dec. 11, 2020) "BNT162b2 Induces SARS-CoV-2-Neutralising Antibodies and T cells in Humans", *medRxiv*, 49 pages.
- Sahin et al. (May 27, 2021) "BNT162b2 vaccine induces neutralizing antibodies and poly-specific T cells in humans", *Nature*, 595:572-577.
- Saltl et al. (Dec. 15, 2011) "Granzyme B Regulates Antiviral CD8+ T Cell Responses", *Journal of Immunology*, 187(12):6301-6309 (19 pages).
- Sercombe et al. (Dec. 1, 2015) "Advances and Challenges of Liposome Assisted Drug Delivery", *Frontiers in Pharmacology*, 6(286):13 Pages.
- Slansky et al. (Oct. 2000) "Enhanced Antigen-Specific Antitumor Immunity with Altered Peptide Ligands that Stabilize the MHC-Peptide-TCR Complex", *Immunity*, 13(4):529-538.
- Tam et al. (Oct. 4, 2016) "Sustained Antigen Availability During Germinal Center Initiation Enhances Antibody Responses to Vaccination", *Proceedings of the National Academy of Sciences*, 113(43):e6639-e6648.
- Taverniti et al. (Jan. 9, 2015) "Elimination of Cap Structures Generated by mRNA Decay Involves the New Scavenger mRNA Decapping Enzyme Aph1/FHIT Together with DcpS", *Nucleic Acids Research*, 43(1):482-492.
- Thompson et al. (Mar. 7, 2006) "Mucosal and Systemic Adjuvant Activity of Alphavirus Replicon Particles", *Proceedings of the National Academy of Sciences*, 103(10):3722-3727.
- U.S. Appl. No. 16/823,212, "Method of Making Lipid-Encapsulated RNA Nanoparticles", 95 pages.
- Villalobos et al. (Jun. 6, 2006) "Gene Designer: A Synthetic Biology Tool for Constructing Artificial DNA Segments", *BMC Bioinformatics*, 7:285 (8 pages).
- Von Herrath et al. (Jun. 2003) "Immune Responsiveness, Tolerance and dsRNA: Implications for Traditional Paradigms", *Trends in Immunology*, 24(6):289-293 (4 pages).
- Wootton et al. (Jun. 1993) "Statistics of Local Complexity in Amino Acid Sequences and Sequence Databases", *Computers & Chemistry*, 17(2):149-163.
- Wrapp et al. (Mar. 13, 2020) "Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation", *Science*, 367(6483):1260-1263.
- Wu et al. (Mar. 12, 2020) "A New Coronavirus Associated with Human Respiratory Disease in China", *Nature*, 579(7798):265-269.
- Yu et al. (Sep. 2000) "April and TALL-I and Receptors BCMA and TACI: System for Regulating Humoral Immunity", *Nature Immunology*, 1(3):252-256.
- Extended European Search Report for Application No. EP 21767978. 6, mailed on Apr. 30, 2024, 11 pages.
- Extended European Search Report for Application No. EP 21768525. 4, mailed on Mar. 22, 2024, 10 pages.
- Kim et al. (2014) "Enhancement of Protein 1-15 Expression by Alphavirus Replicons by Designing Self-replicating Subgenomic RNAs", *Proceedings of the National Academy of Sciences of the United States of America*, 111 (29):10708-10713.
- Lundstrom, Kenneth (2016) "Replicon RNA Viral 1-15 Vectors as Vaccines", *Vaccines*, 4(4):39.
- Lundstrom, Kenneth (2018) "Self-Replicating RNA Viruses for RNA Therapeutics", *Molecules*, 23(12):3310.

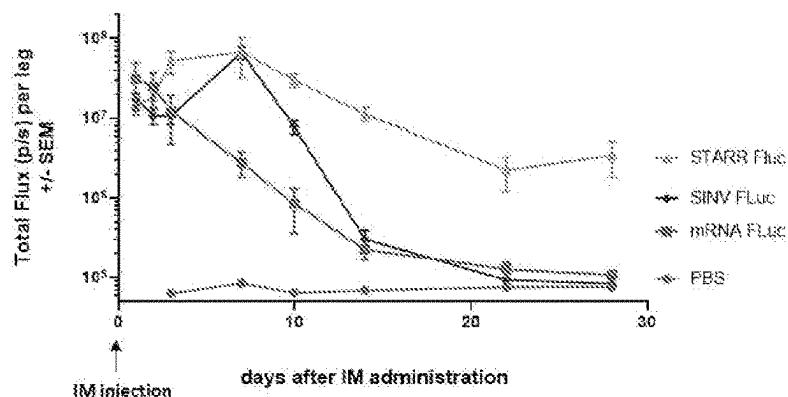
* cited by examiner

[STARRTM] self-replicating RNA

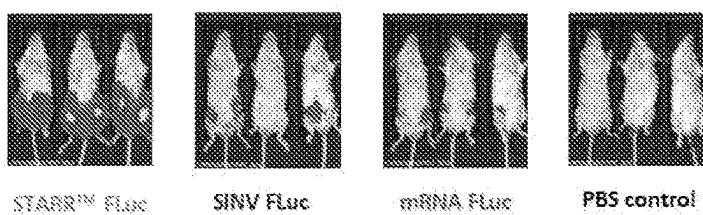


STARR technology™ can be used to generate a positive feedback loop that drives protein expression

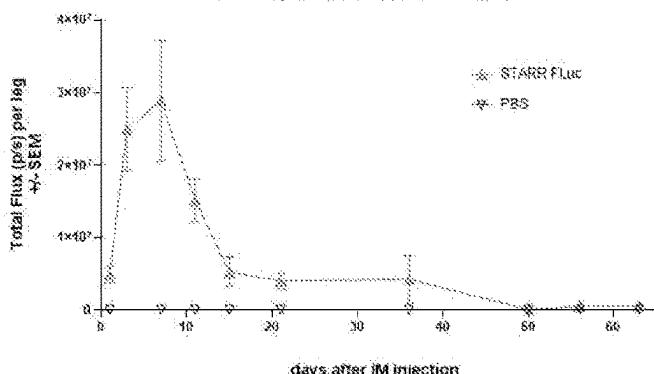
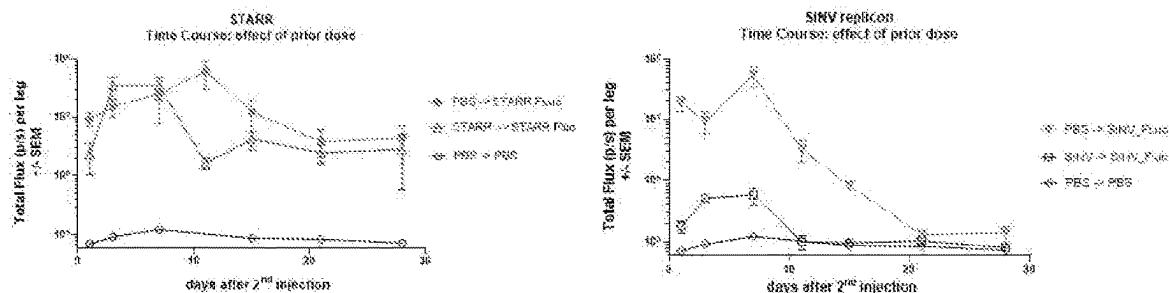
FIG. 1

FIG. 2A Time Course: FLuc expression in BALB/c mice**FIG. 2B**

IVIS: Day 14 post dosing

**FIG. 2C**

FLuc Expression Duration in BALB/c mice

**FIG. 2D**

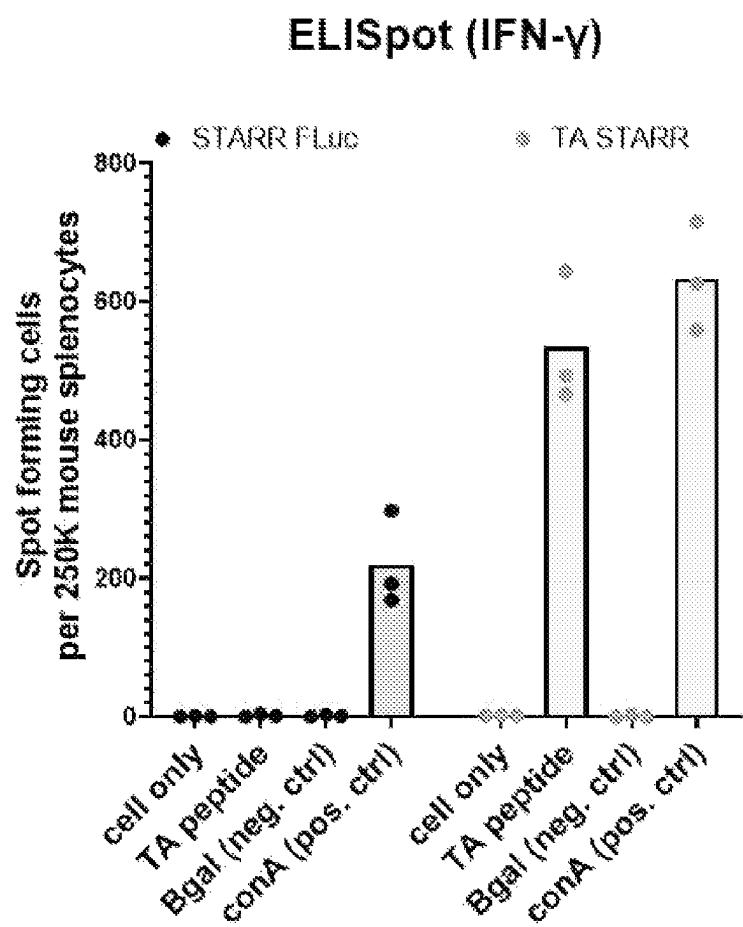
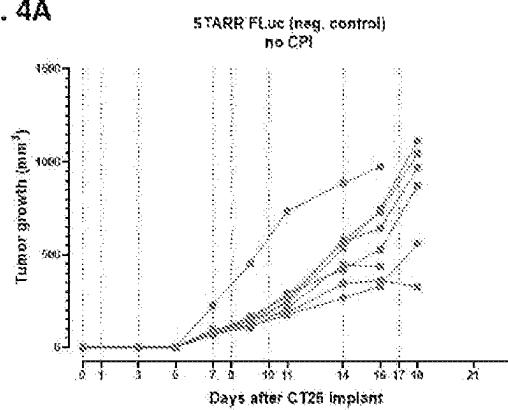
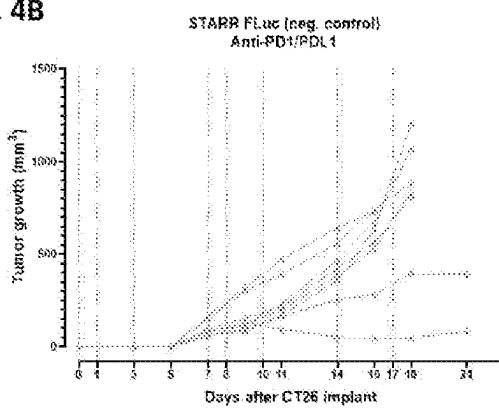
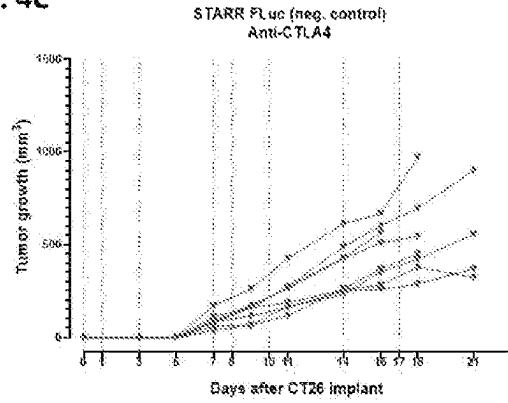
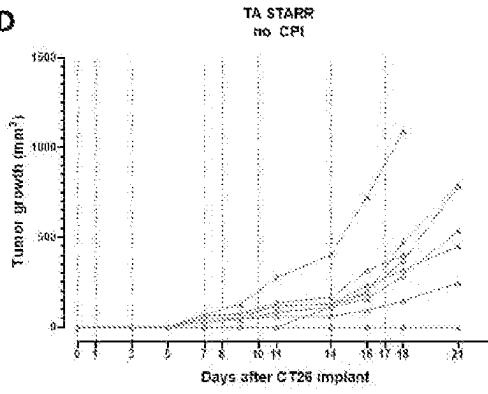
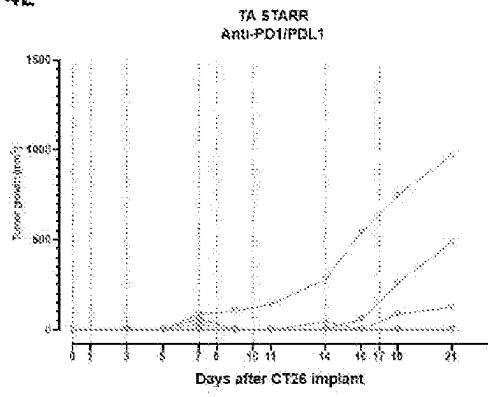
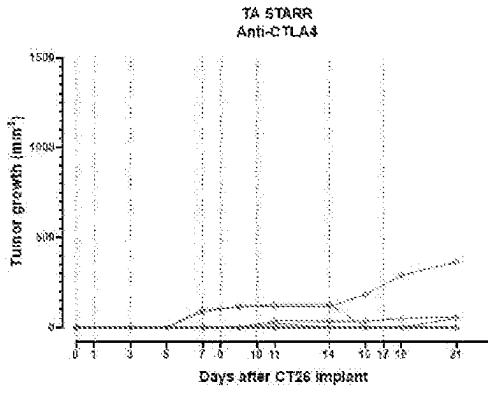


FIG. 3

FIG. 4A**FIG. 4B****FIG. 4C****FIG. 4D****FIG. 4E****FIG. 4F**

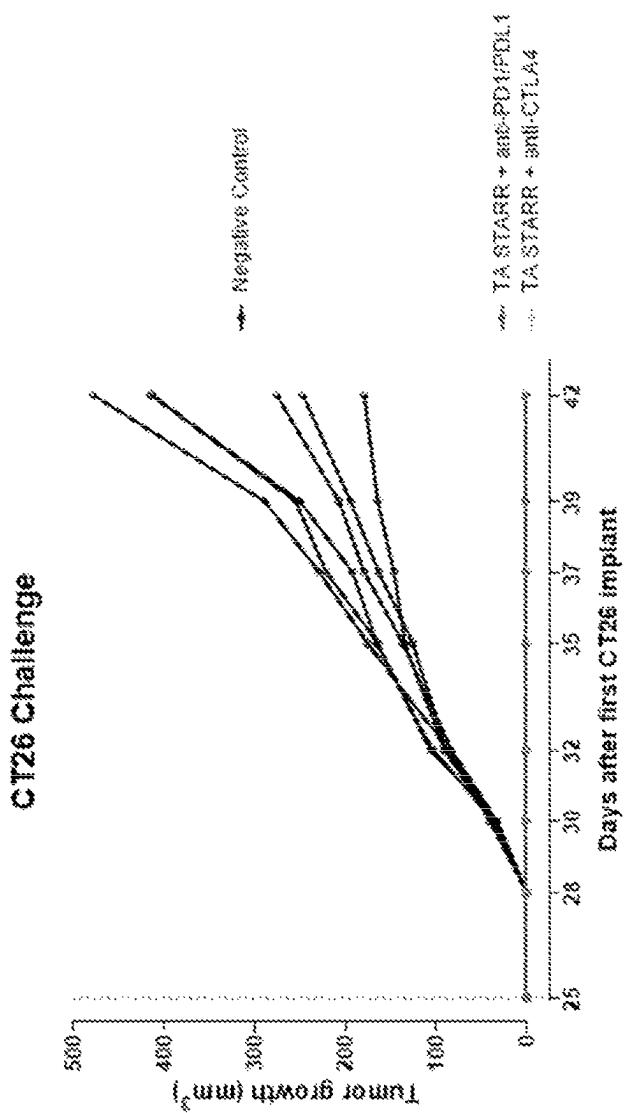
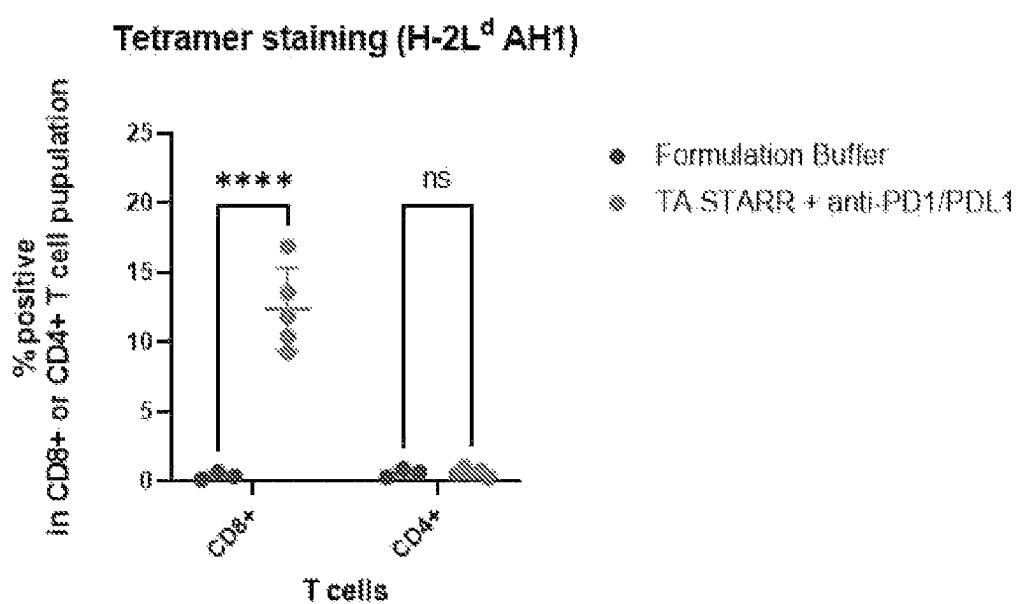


FIG. 5

**FIG. 6A**

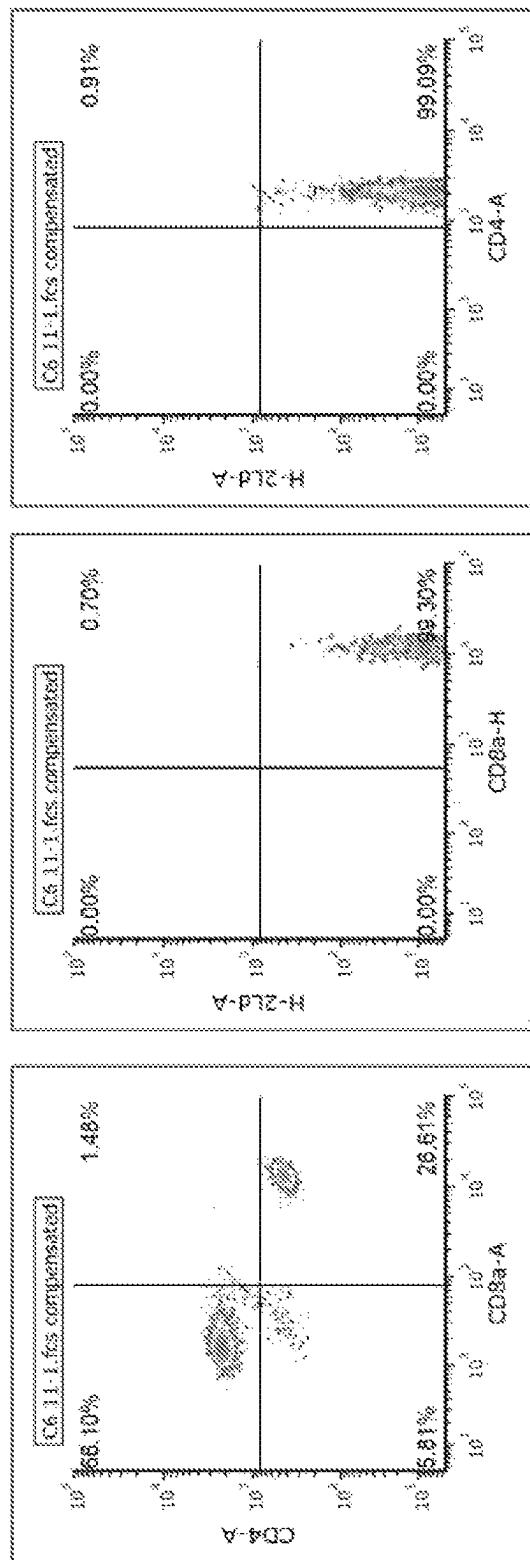


FIG. 6B

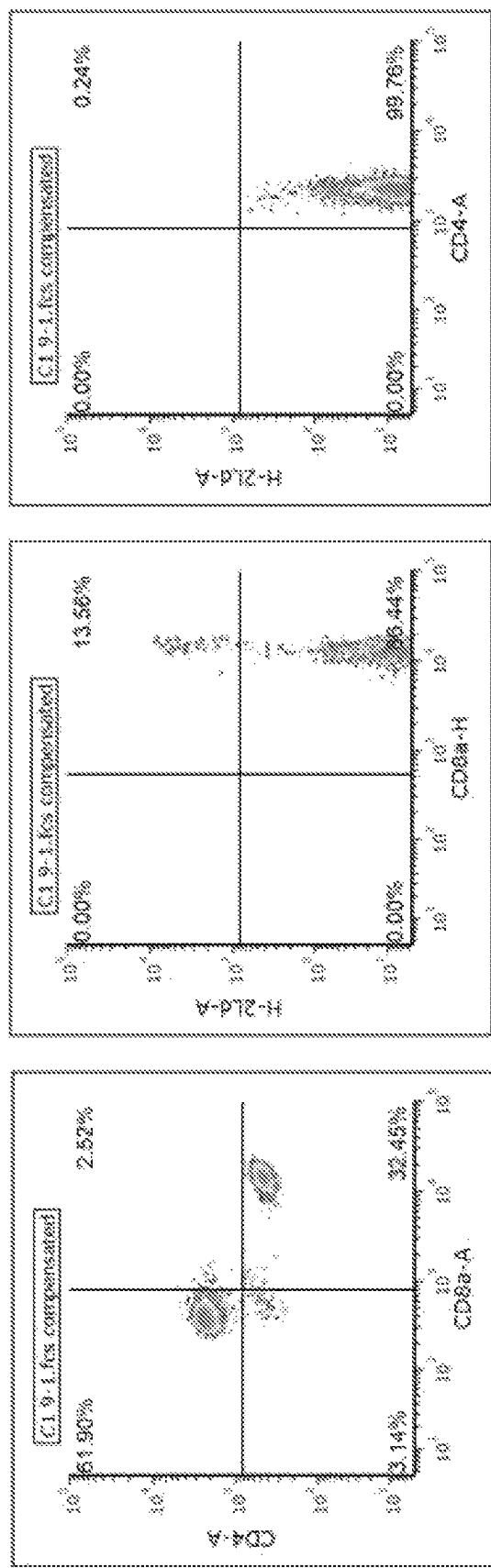


FIG. 6C

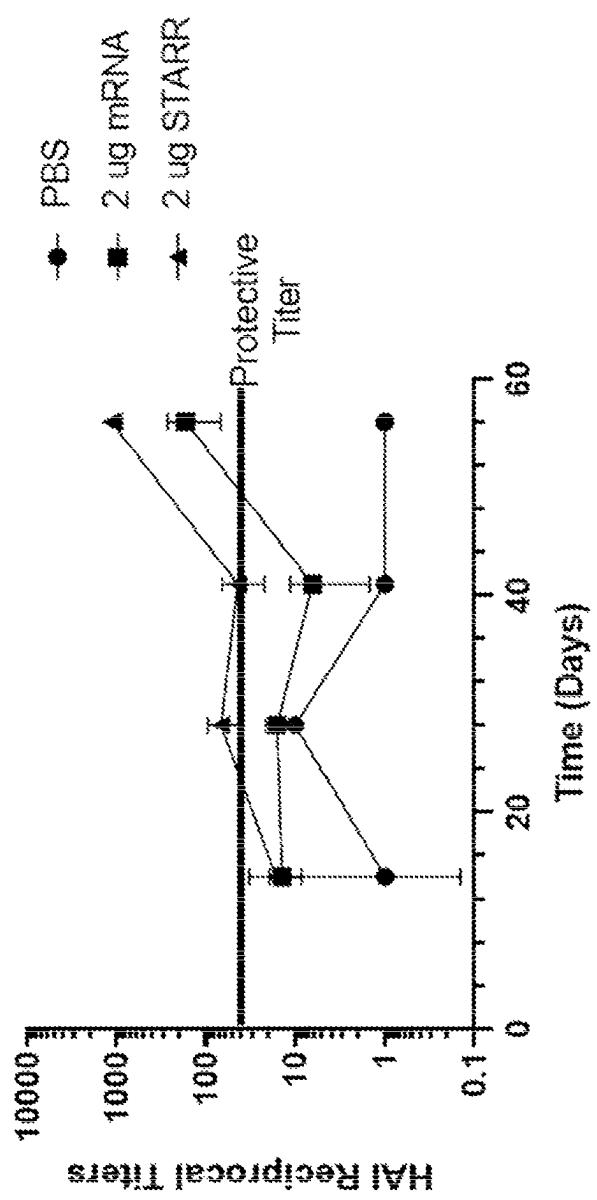


FIG. 7

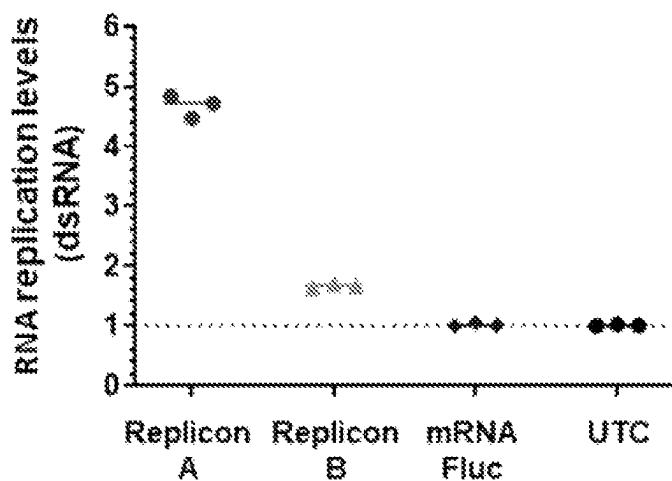


FIG. 8A

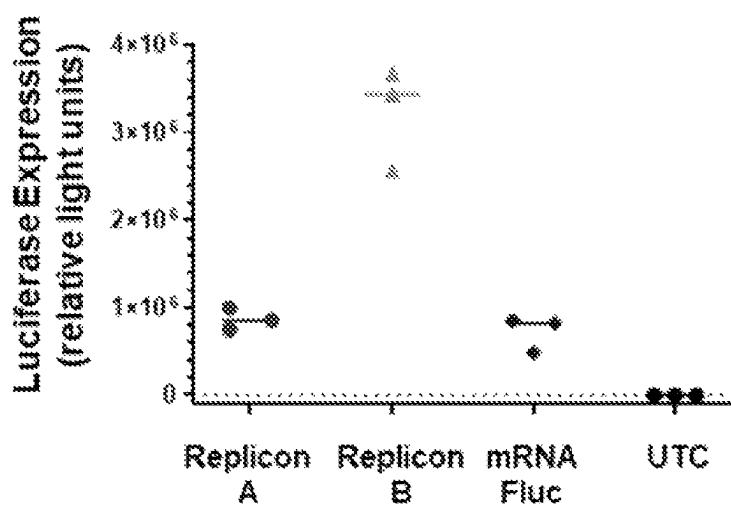
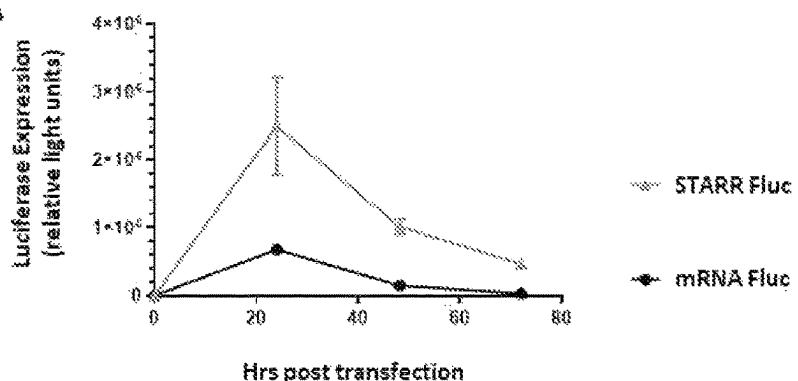
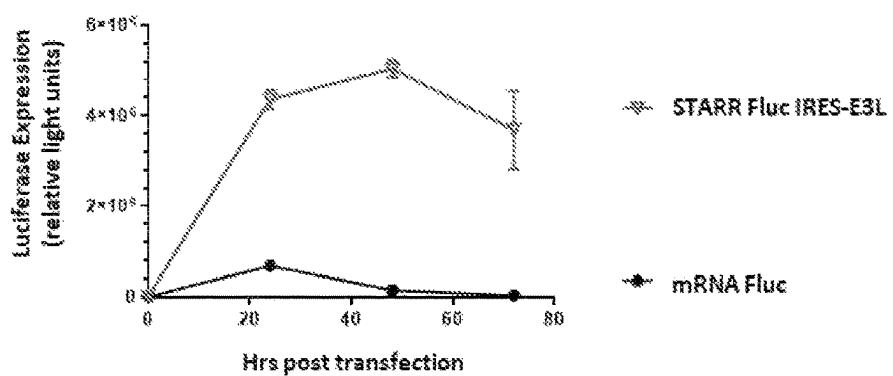
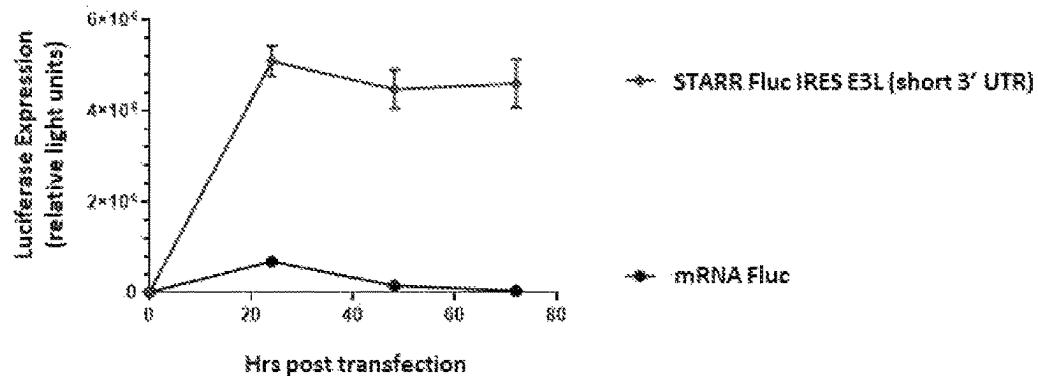


FIG. 8B

FIG. 9A**FIG. 9B****FIG. 9C**

1**COMPOSITIONS AND METHODS FOR INDUCING IMMUNE RESPONSES****CROSS-REFERENCES TO RELATED APPLICATIONS**

This application claims the benefit of U.S. Provisional Application No. 62/987,191, filed Mar. 9, 2020 and U.S. Provisional Application No. 63/073,900, filed Sep. 2, 2020.

REFERENCE TO A SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on Jul. 12, 2023 is named 2023 Jul. 11 Sequence_Listing_ST26 049386-530C01US.xml and is 322,134 bytes in size.

Reference is also made to the Sequence Listing filed with U.S. application Ser. No. 17/196,890, which was submitted electronically in ASCII format and is also hereby incorporated by reference in its entirety. Said ASCII copy, created on Mar. 8, 2021 is named 049386-530001US_SequenceListing_ST25.txt and is 390,698 bytes in size.

TECHNICAL FIELD

The present disclosure relates generally to inducing immune responses against infectious agents and tumor antigens and more specifically to self-transcribing and replicating RNA for antigen expression.

BACKGROUND

Infectious diseases and cancer represent significant burdens on health worldwide. According to the World Health Organization (WHO), lower respiratory tract infection was the deadliest infectious disease worldwide in 2016, causing approximately 3 million deaths. Current control measures to curb the rapid worldwide spread of infection diseases, such as national lockdowns, closure of work places and schools, and reduction of international travel are threatening to result in a global economic recession to an extent not seen since the Great Depression.

Cancer is the second leading cause of death globally, accounting for approximately 9.6 million deaths worldwide in 2018. Cancer is a large group of diseases that can affect almost any organ or tissue in the body. Cancer burden continues to grow globally, exerting physical, emotional, and financial strains on patients and health care providers. Self-replicating ribonucleic acids (RNAs), e.g., derived from viral replicons, are useful for expression of proteins, such as heterologous proteins, for a variety of purposes, such as expression of therapeutic proteins and expression of antigens for vaccines. A desirable property of such replicons is the ability for sustained expression of the protein.

Few treatments for infections caused by viruses and eukaryotic organisms are available, and resistance to antibiotics for the treatment of bacterial infections is increasing. In addition, rapid responses, including rapid vaccine development, are required to effectively control emerging infectious diseases and pandemics. Moreover, many cancer treatments include costly and painful surgeries and chemotherapies that are often unsuccessful or only modestly

2

prolong life despite serious side effects. Thus, there exists a need for the prevention and/or treatment of infectious diseases and cancer.

SUMMARY

In one aspect, the present disclosure provides a nucleic acid molecule comprising a first polynucleotide encoding one or more viral replication proteins, wherein the first 10 polynucleotide is codon-optimized as compared to a wild-type polynucleotide encoding the one or more viral replication proteins; and a second polynucleotide comprising a first transgene encoding a first antigenic protein or a fragment thereof.

15 In some embodiments, the one or more viral replication proteins may be alphavirus proteins or rubivirus proteins.

In some embodiments, the alphavirus proteins are from Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus 20 (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MIDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiama Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (SINV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus 30 (SAV), Buggy Creek Virus (BCRV), or any combination thereof.

35 In some embodiments, the first polynucleotide encodes a polyprotein comprising an alphavirus nsP1 protein, an alphavirus nsP2 protein, an alphavirus nsP3 protein, an alphavirus nsP4 protein, or any combination thereof.

In some embodiments, the first polynucleotide encodes a 40 polyprotein comprising an alphavirus nsP1 protein, an alphavirus nsP2 protein, an alphavirus nsP3 protein, or any combination thereof, and an alphavirus nsP4 protein.

In some embodiments, the nucleic acid molecule further 45 comprises a first intergenic region between a sequence encoding the polyprotein comprising an alphavirus nsP1 protein, an alphavirus nsP2 protein, an alphavirus nsP3 protein, or any combination thereof, and a sequence encoding an alphavirus nsP4 protein.

In some embodiments, the first intergenic region comprises an alphavirus sequence.

In some embodiments, the first polynucleotide comprises 50 a sequence having at least 80% identity to a sequence of SEQ ID NO:72.

In some embodiments, the nucleic acid molecule further 55 comprises a 5' untranslated region (UTR), such as a viral 5' UTR, a non-viral 5' UTR, or a combination of viral and non-viral 5' UTR sequences. In some embodiments, the 5' UTR comprises an alphavirus 5' UTR.

In some embodiments, the alphavirus 5' UTR comprises a Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MIDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiama Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (SINV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-

agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), or Buggy Creek Virus (BCRV) 5' UTR sequence.

In some embodiments, the 5' UTR comprises a sequence of SEQ ID NO:73, SEQ ID NO:74, or SEQ ID NO:75.

In some embodiments, the nucleic acid molecule further comprises a 3' untranslated region (UTR). In some embodiments, the 3' UTR comprises a viral 3' UTR, a non-viral 3' UTR, or a combination of viral and non-viral 3' UTR sequences. In some embodiments, the 3' UTR comprises an alphavirus 3' UTR.

In some embodiments, the alphavirus 3' UTR comprises a Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiyma Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (SINV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), or Buggy Creek Virus (BCRV) 3' UTR sequence.

In some embodiments, the 3' UTR comprises a poly-A sequence. In some embodiments, the 3' UTR comprises a sequence of SEQ ID NO:76.

In some embodiments, the antigenic protein is a viral protein, a bacterial protein, a fungal protein, a protozoan protein, a parasite protein, or a tumor protein.

In some embodiments, the viral protein is an orthomyxovirus protein, a paramyxovirus protein, a picornavirus protein, a flavivirus protein, a filovirus protein, a rhabdovirus protein, a togavirus protein, an arterivirus protein, a bunyavirus protein, an arenavirus protein, a reovirus protein, a bornavirus protein, a retrovirus protein, an adenovirus protein, a herpesvirus protein, a polyomavirus protein, a papillomavirus protein, a poxvirus protein, or a hepadnavirus protein.

In some embodiments, the antigenic protein is an influenza virus protein, a respiratory syncytial virus (RSV) protein, a human immunodeficiency virus (HIV) protein, a hepatitis C virus (HCV) protein, a cytomegalovirus (CMV) protein, a Lassa Fever Virus (LFV) protein, an Ebola Virus (EBOV) protein, a *Mycobacterium* protein, a *Bacillus* protein, a *Yersinia* protein, a *Streptococcus* protein, a *Pseudomonas* protein, a *Shigella* protein, a *Campylobacter* protein, a *Salmonella* protein, a *Plasmodium* protein, or a *Toxoplasma* protein.

In some embodiments, the tumor protein is a kidney cancer, renal cancer, urinary bladder cancer, prostate cancer, uterine cancer, breast cancer, cervical cancer, ovarian cancer, lung cancer, liver cancer, stomach cancer, colon cancer, rectal cancer, oral cavity cancer, pharynx cancer, pancreatic cancer, thyroid cancer, melanoma, skin cancer, head and neck cancer, brain cancer, hematopoietic cancer, leukemia, lymphoma, bone cancer, or sarcoma protein.

In some embodiments, the second polynucleotide comprises at least two transgenes.

In some embodiments, a second transgene encodes a second antigenic protein or a fragment thereof or an immunomodulatory protein.

In some embodiments, the second polynucleotide further comprises a sequence encoding a 2A peptide, an internal ribosomal entry site (IRES), or a combination thereof, located between transgenes.

5 In some embodiments, the immunomodulatory protein is a cytokine, a chemokine, or an interleukin.

In some embodiments, the first and second transgenes encode viral proteins, bacterial proteins, fungal proteins, protozoan proteins, parasite proteins, tumor proteins, immunomodulatory proteins, or any combination thereof.

10 In some embodiments, the first polynucleotide is located 5' of the second polynucleotide.

In some embodiments, the nucleic acid molecule further comprises a second intergenic region located between the 15 first polynucleotide and the second polynucleotide.

In some embodiments, the second intergenic region comprises a sequence having at least 85% identity to a sequence of SEQ ID NO:77.

20 In some embodiments, the nucleic acid molecule is a DNA molecule; or an RNA molecule, wherein T is substituted with U.

In some embodiments, the DNA molecule further comprises a promoter. In some embodiments, the promoter is located 5' of the 5'UTR.

25 In some embodiments, the promoter is a T7 promoter, a T3 promoter, or an SP6 promoter.

In some embodiments, the RNA molecule is a self-replicating RNA molecule.

In some embodiments, the RNA molecule further comprises a 5' cap. In some embodiments, the 5' cap has a Cap 1 structure, a Cap 1 (m6A) structure, a Cap 2 structure, a Cap 0 structure, or any combination thereof.

30 In another aspect, provided herein is a nucleic acid molecule comprising a sequence of SEQ ID NO:78; or a sequence of SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:76, and SEQ ID NO:77, wherein T is substituted with U.

35 In some embodiments, the nucleic acid molecule is an RNA molecule.

In some embodiments, the nucleic acid molecule further 40 comprises a 5' cap having a Cap 1 structure.

In yet another aspect, provided herein is a nucleic acid molecule comprising a first polynucleotide comprising a sequence having at least 80% identity to a sequence of SEQ ID NO:72; and a second polynucleotide comprising a first 45 transgene encoding a first antigenic protein or a fragment thereof.

In some embodiments, the nucleic acid molecule further comprises a 5' untranslated region (UTR).

50 In some embodiments, the 5' UTR comprises a viral 5' UTR, a non-viral 5' UTR, or a combination of viral and non-viral 5' UTR sequences. In some embodiments, the 5' UTR comprises an alphavirus 5' UTR.

In some embodiments, the alphavirus 5' UTR comprises a Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiyma Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (SINV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), or Buggy Creek Virus (BCRV) 5' UTR sequence.

55 In some embodiments, the alphavirus 5' UTR comprises a Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiyma Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (SINV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), or Buggy Creek Virus (BCRV) 5' UTR sequence.

5

In some embodiments, the 5' UTR comprises a sequence of SEQ ID NO:73, SEQ ID NO:74, or SEQ ID NO:75.

In some embodiments, the nucleic acid molecule further comprises a 3' untranslated region (UTR).

In some embodiments, the 3' UTR comprises a viral 3' UTR, a non-viral 3' UTR, or a combination of viral and non-viral 3' UTR sequences. In some embodiments, the 3' UTR comprises an alphavirus 3' UTR.

In some embodiments, the alphavirus 3' UTR comprises a Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MIDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiymaya Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (SINV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), or Buggy Creek Virus (BCRV) 3' UTR sequence.

In some embodiments, the 3' UTR comprises a poly-A sequence.

In some embodiments, the 3' UTR comprises a sequence of SEQ ID NO:76.

In some embodiments, the antigenic protein is a viral protein, a bacterial protein, a fungal protein, a protozoan protein, a parasite protein, or a tumor protein.

In some embodiments, the viral protein is an orthomyxovirus protein, a paramyxovirus protein, a picornavirus protein, a flavivirus protein, a filovirus protein, a rhabdovirus protein, a togavirus protein, an arterivirus protein, a bunyavirus protein, an arenavirus protein, a reovirus protein, a bornavirus protein, a retrovirus protein, an adenovirus protein, a herpesvirus protein, a polyomavirus protein, a papillomavirus protein, a poxvirus protein, or a hepadnavirus protein.

In some embodiments, the antigenic protein is an influenza virus protein, a respiratory syncytial virus (RSV) protein, a human immunodeficiency virus (HIV) protein, a hepatitis C virus (HCV) protein, a cytomegalovirus (CMV) protein, a Lassa Fever Virus (LFV) protein, an Ebola Virus (EBOV) protein, a *Mycobacterium* protein, a *Bacillus* protein, a *Yersinia* protein, a *Streptococcus* protein, a *Pseudomonas* protein, a *Shigella* protein, a *Campylobacter* protein, a *Salmonella* protein, a *Plasmodium* protein, or a *Toxoplasma* protein.

In some embodiments, the tumor protein is a kidney cancer, renal cancer, urinary bladder cancer, prostate cancer, uterine cancer, breast cancer, cervical cancer, ovarian cancer, lung cancer, liver cancer, stomach cancer, colon cancer, rectal cancer, oral cavity cancer, pharynx cancer, pancreatic cancer, thyroid cancer, melanoma, skin cancer, head and neck cancer, brain cancer, hematopoietic cancer, leukemia, lymphoma, bone cancer, or sarcoma protein.

In some embodiments, the second polynucleotide comprises at least two transgenes. In some embodiments, a second transgene encodes a second antigenic protein or a fragment thereof or an immunomodulatory protein.

In some embodiments, the second polynucleotide further comprises a sequence encoding a 2A peptide, an internal ribosomal entry site (IRES), or a combination thereof, located between transgenes.

6

In some embodiments, the immunomodulatory protein is a cytokine, a chemokine, or an interleukin.

In some embodiments, the first and second transgenes encode viral proteins, bacterial proteins, fungal proteins, protozoan proteins, parasite proteins, tumor proteins, immunomodulatory proteins, or any combination thereof.

In some embodiments, the first polynucleotide is located 5' of the second polynucleotide.

In some embodiments, the nucleic acid molecule further comprises a second intergenic region located between the first polynucleotide and the second polynucleotide. In some embodiments, the second intergenic region comprises a sequence having at least 85% identity to a sequence of SEQ ID NO:77.

In some embodiments, the nucleic acid molecule is a DNA molecule; or an RNA molecule, wherein T is substituted with U.

In some embodiments, the DNA molecule further comprises a promoter. In some embodiments, the promoter is located 5' of the 5'UTR.

In some embodiments, the promoter is a T7 promoter, a T3 promoter, or an SP6 promoter.

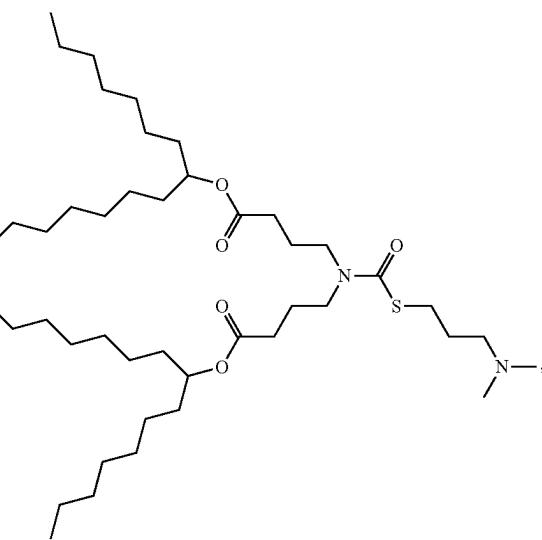
In some embodiments, the RNA molecule is a self-replicating RNA molecule.

In some embodiments, the RNA molecule further comprises a 5' cap. In some embodiments, the 5' cap has a Cap 1 structure, a Cap 1 (m6A) structure, a Cap 2 structure, a Cap 0 structure, or any combination thereof.

In yet another aspect, provided herein is a composition comprising any one of the nucleic acid molecules described herein.

In some embodiments, the lipid comprises an ionizable cationic lipid. In some embodiments, the ionizable cationic lipid has a structure of

40



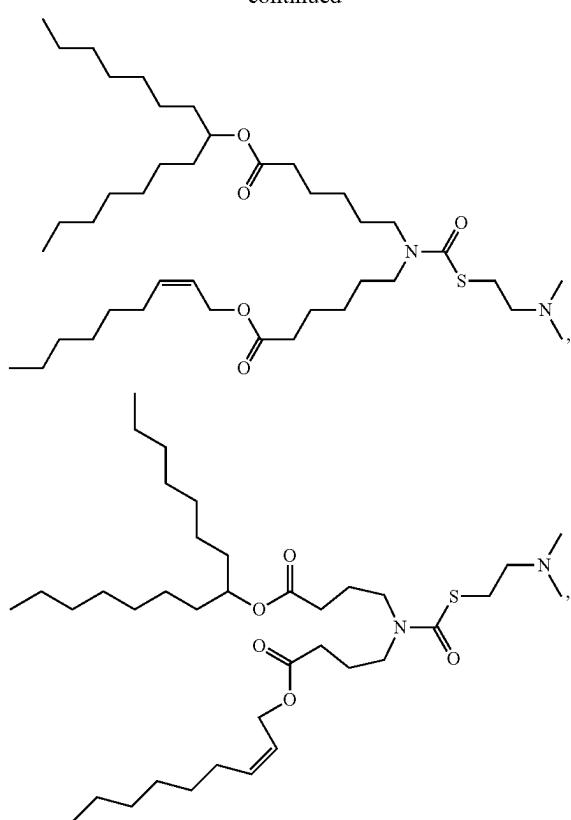
45

50

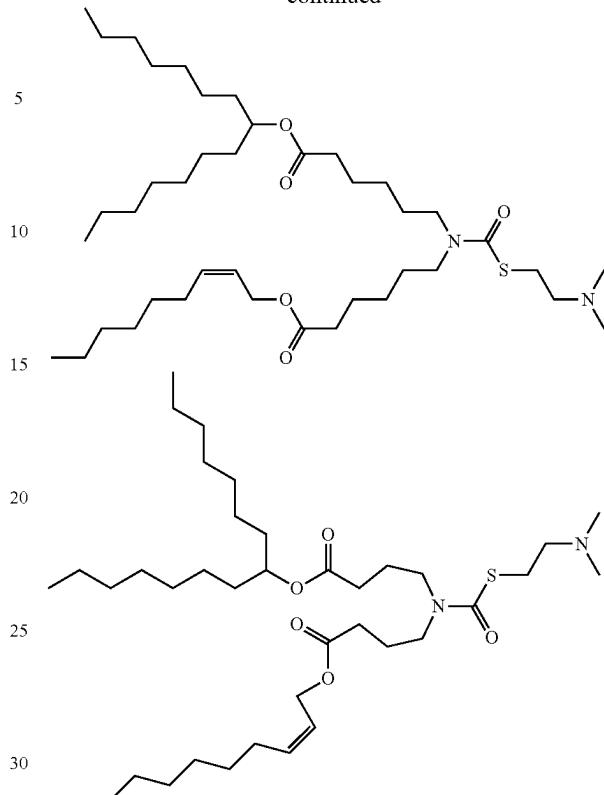
60

65

7
-continued



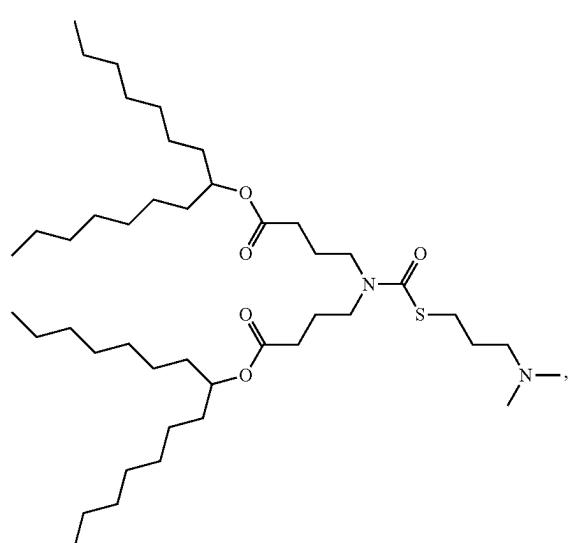
8
-continued



or a pharmaceutically acceptable salt thereof.

In yet another aspect, provided herein is a pharmaceutical composition comprising any one of the nucleic acid molecules described herein, and a lipid formulation.

In some embodiments, the lipid formulation comprises an ionizable cationic lipid. In some embodiments, the ionizable cationic lipid has a structure of



or a pharmaceutically acceptable salt thereof.

35 In yet another aspect, provided herein is a method of inducing an immune response in a subject comprising administering to the subject an effective amount of any one of the nucleic acid molecules described herein.

In some embodiments, the method comprises administering the nucleic acid molecule intramuscularly, subcutaneously, intradermally, transdermally, intranasally, orally, sublingually, intravenously, intraperitoneally, topically, by aerosol, or by a pulmonary route.

40 In yet another aspect, provided herein is a method of inducing an immune response in a subject comprising administering to the subject an effective amount of any one of the compositions described herein.

45 In some embodiments, the method comprises administering the composition intramuscularly, subcutaneously, intra-dermally, transdermally, intranasally, orally, sublingually, intravenously, intraperitoneally, topically, by aerosol, or by a pulmonary route.

50 In yet another aspect, provided herein is a method of inducing an immune response in a subject comprising administering to the subject an effective amount of any one of the pharmaceutical compositions described herein.

55 In some embodiments, the method may comprise administering the pharmaceutical composition intramuscularly, subcutaneously, intradermally, transdermally, intranasally, orally, sublingually, intravenously, intraperitoneally, topically, by aerosol, or by a pulmonary route.

60 In yet another aspect, the present disclosure provides any of the nucleic acid molecules described herein for use in inducing an immune response to the first antigenic protein or fragment thereof.

65 In yet another aspect, the present disclosure provides use of any one of the nucleic acid molecules described herein in

the manufacture of a medicament for inducing an immune response to the first antigenic protein or fragment thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a schematic illustrating one aspect of STARR™ technology.

FIGS. 2A-2D show characterization of STARR™ technology with firefly luciferase transgene expression. (2A) Firefly luciferase (FLuc) expression from STARR™ Fluc, SINV FLuc, and mRNA FLuc was monitored up to day 28 by In Vivo Imaging System (IVIS). The average of total flux (p/s) from 6 injection sites in a mouse group was plotted at each time point with a standard error of mean, SEM. (2B) IVIS picture of three mice (6 injection sites) per group on day 14 is shown for each group that was administered with the test article labeled below the picture. (2C) Luciferase expression from mice that were intramuscularly injected with STARR™ FLuc was monitored by IVIS up to 63 days post administration. (2D) Effect of prior administration of replicon backbone was examined for STARR™ (upper panel) and SINV (lower panel). Replicon encoding FLuc was IM injected at 7 days post dose of replicon with homologous backbone with an irrelevant gene/sequence (labeled STARR™ irr or SINV irr) at day 0. As a reference, a mouse group with PBS administration at day 0 was included in each of STARR™ and SINV group.

FIG. 3 shows that STARR™ elicits antigen-specific IFN-gamma response. Enzyme-linked immune absorbent spot ELISpot was used to count the number of splenocytes that were specifically stimulated by an antigen peptide of the same amino acid sequence encoded in TA STARR™. Neither no peptide (cell only) nor irrelevant peptide (Bgal) elicited significant IFN-gamma from splenocytes from mice vaccinated with STARR™ FLuc or TA STARR™. Stimulation with AH1-A5 peptide resulted in the detection of IFN-gamma-producing cells specifically from the mice that were vaccinated with TA STARR™. Concanavalin A (ConA) was used as a positive control of IFN-gamma production.

FIGS. 4A-4F illustrate reduced tumor growth rate by TA STARR™ vaccination in a CT26 syngeneic mouse model. CT26 murine colorectal carcinoma cells (5×10^5) were subcutaneously implanted in 10-week old female BALB/c mice (n=8 per group). On days 1 and 8, the mice were vaccinated with STARR™ FLuc, a negative control, or TA STARR™, which encodes AH1A5 epitope. Tumor growth was monitored in mice vaccinated with (4A) STARR™ FLuc without checkpoint inhibitor treatment; (4B) STARR™ FLuc with a combination anti-PD1/PDL1 treatment; (4C) STARR™ FLuc with a combination anti-CTLA4 treatment; (4D) STARR™ vaccine without checkpoint inhibitor treatment; (4E) STARR™ vaccine with a combination treatment of anti-PD1 and anti-PDL1; and (4F) STARR™ vaccine with a combination treatment of anti-CTLA4. The individual tumor growth curves from a mouse group that were administered with STARR™ FLuc and TA STARR™ are shown in upper and lower panels, respectively.

FIG. 5 illustrates prolonged protection by combination treatment of TA STARR™ Vaccine with checkpoint inhibitors. Mice that were treated with TA STARR™ combined with anti-PD1/PDL1 or anti-CTLA4 were found to be resistant to tumor development following the CT26 challenge at day 25 to 42. Naïve mice were used as a control for the CT26 tumor growth.

FIGS. 6A-6C show results from AH1-tetramer staining of CD8+ T-cells in the form of (6A) a graph and (6B and 6C)

plots. Splenocytes from the mice group with combination treatment of TA STARR™ and anti-PD1/PDL1 at day 42 were stained with AH1 (H-2Ld)-tetramer. The staining was specific to CD8+ T cells from the mouse group with TA STARR™ treatment, and the population represented 9-17% of total CD8+ T cells from the splenocytes.

FIG. 7 shows HAI titers obtained for self-replicating RNA (STARR™) and mRNA constructs encoding the hemagglutinin of influenza virus A/California/07/2009 (H1N1).

FIGS. 8A-8B show (8A) RNA replication levels and (8B) luciferase reporter gene expression levels for the indicated self-replicating (replicon) RNAs as compared to mRNA.

FIGS. 9A-9C shows duration of luciferase reporter gene expression for self-replicating (replicon) RNA (STARR™), such as (9A) STARR™ FLuc, (9B) STARR™ FLuc IRES-E3L, and (9C) STARR™ FLuc IRES E3L (short 3' UTR) as compared to mRNA.

DETAILED DESCRIPTION

The present disclosure relates to self-replicating RNAs and nucleic acids encoding the same for expression of transgenes such as antigenic proteins and tumor antigens, for example. Also provided herein are methods of administration (e.g., to a host, such as a mammalian subject) of self-replicating RNAs, whereby the self-replicating RNA is translated in vivo and the heterologous protein-coding sequence is expressed and, e.g., can elicit an immune response to the heterologous protein-coding sequence in the recipient or provide a therapeutic effect, where the heterologous protein-coding sequence is a therapeutic protein. Self-replicating RNAs provided herein are useful as vaccines that can be rapidly generated and that can be effective at low and/or single doses. The present disclosure further relates to methods of inducing an immune response using self-replicating RNAs provided herein.

Definitions

As used herein, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise. Thus, for example, references to “the method” includes one or more methods, and/or steps of the type described herein which will become apparent to those persons skilled in the art upon reading this disclosure and so forth.

“About” as used herein when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of +20%, or $\pm 10\%$, or $\pm 5\%$, or even $\pm 1\%$ from the specified value, as such variations are appropriate for the disclosed methods or to perform the disclosed methods.

As used herein, the term “fragment,” when referring to a protein or nucleic acid, for example, means any shorter sequence than the full-length protein or nucleic acid. Accordingly, any sequence of a nucleic acid or protein other than the full-length nucleic acid or protein sequence can be a fragment. In some aspects, a protein fragment includes an epitope. In other aspects, a protein fragment is an epitope.

As used herein, the term “nucleic acid” refers to any deoxyribonucleic acid (DNA) molecule, ribonucleic acid (RNA) molecule, or nucleic acid analogues. A DNA or RNA molecule can be double-stranded or single-stranded and can be of any size. Exemplary nucleic acids include, but are not limited to, chromosomal DNA, plasmid DNA, cDNA, cell-free DNA (cfDNA), mitochondrial DNA, chloroplast DNA,

11

viral DNA, mRNA, tRNA, rRNA, long non-coding RNA, siRNA, micro RNA (miRNA or miR), hnRNA, and viral RNA. Exemplary nucleic analogues include peptide nucleic acid, morpholino- and locked nucleic acid, glycol nucleic acid, and threose nucleic acid. As used herein, the term “nucleic acid molecule” is meant to include fragments of nucleic acid molecules as well as any full-length or non-fragmented nucleic acid molecule, for example. As used herein, the terms “nucleic acid” and “nucleic acid molecule” can be used interchangeably, unless context clearly indicates otherwise.

As used herein, the term “protein” refers to any polymeric chain of amino acids. The terms “peptide” and “polypeptide” can be used interchangeably with the term protein, unless context clearly indicates otherwise, and can also refer to a polymeric chain of amino acids. The term “protein” encompasses native or artificial proteins, protein fragments and polypeptide analogs of a protein sequence. A protein may be monomeric or polymeric. The term “protein” encompasses fragments and variants (including fragments of variants) thereof, unless otherwise contradicted by context.

In general, “sequence identity” or “sequence homology,” which can be used interchangeably, refer to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Typically, techniques for determining sequence identity include determining the nucleotide sequence of a polynucleotide and/or determining the amino acid sequence encoded thereby or the amino acid sequence of a polypeptide, and comparing these sequences to a second nucleotide or amino acid sequence.

As used herein, the term “percent (%) sequence identity” or “percent (%) identity,” also including “homology,” refers to the percentage of amino acid residues or nucleotides in a sequence that are identical with the amino acid residues or nucleotides in a reference sequence after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Thus, two or more sequences (polynucleotide or amino acid) can be compared by determining their “percent identity,” also referred to as “percent homology.” The percent identity to a reference sequence (e.g., nucleic acid or amino acid sequences), which may be a sequence within a longer molecule (e.g., polynucleotide or polypeptide), may be calculated as the number of exact matches between two optimally aligned sequences divided by the length of the reference sequence and multiplied by 100. Percent identity may also be determined, for example, by comparing sequence information using the advanced BLAST computer program, including version 2.2.9, available from the National Institutes of Health. The BLAST program is based on the alignment method of Karlin and Altschul, Proc. Natl. Acad. Sci. USA 87:2264-2268 (1990) and as discussed in Altschul et al., J. Mol. Biol. 215:403-410 (1990); Karlin and Altschul, Proc. Natl. Acad. sci. USA 90:5873-5877 (1993); and Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997). Briefly, the BLAST program defines identity as the number of identical aligned symbols (i.e., nucleotides or amino acids), divided by the total number of symbols in the shorter of the two sequences. The program may be used to determine percent identity over the entire length of the sequences being compared. Default parameters are provided to optimize searches with short query sequences, for example, with the blastp program. The program also allows use of an SEG filter to mask-off segments of the query sequences as determined by the SEG program of Wootton and Federhen,

12

Computers and Chemistry 17: 149-163 (1993). Ranges of desired degrees of sequence identity are approximately 80% to 100% and integer values in between. Percent identities between a reference sequence and a claimed sequence can be at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, or at least 99.9%. In general, an exact match indicates 100% identity over the length of the reference sequence. Additional programs and methods for comparing sequences and/or assessing sequence identity include the Needleman-Wunsch algorithm (see, e.g., the EMBOSS Needle aligner available at ebi.ac.uk/Tools/psa/emboss_needle/, optionally with default settings), the Smith-Waterman algorithm (see, e.g., the EMBOSS Water aligner available at ebi.ac.uk/Tools/psa/emboss_water/, optionally with default settings), the similarity search method of Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85, 2444, or computer programs which use these algorithms (GAP, BESTFIT, FASTA, BLAST P, BLAST N and TFASTA in Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.). In some aspects, reference to percent sequence identity refers to sequence identity as measured using BLAST (Basic Local Alignment Search Tool). In other aspects, ClustalW is used for multiple sequence alignment. Optimal alignment may be assessed using any suitable parameters of a chosen algorithm, including default parameters.

The term “expression” refers to the process by which a nucleic acid sequence or a polynucleotide is transcribed from a DNA template (such as into mRNA or other RNA transcript) and/or the process by which a transcribed mRNA or other RNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as “gene product.”

As used herein, “operably linked,” “operable linkage,” “operatively linked,” or grammatical equivalents thereof refer to juxtaposition of genetic elements, e.g., a promoter, an enhancer, a polyadenylation sequence, etc., wherein the elements are in a relationship permitting them to operate in the expected manner. For instance, a regulatory element, which can comprise promoter and/or enhancer sequences, is operatively linked to a coding region if the regulatory element helps initiate transcription of the coding sequence. There may be intervening residues between the regulatory element and coding region so long as this functional relationship is maintained.

As used herein, the term “drug” or “medicament,” means a pharmaceutical formulation or composition as described herein.

The phrases “administered in combination” or “combined administration” means that two or more agents are administered to a subject at the same time or within an interval such that there may be an overlap of an effect of each agent on the patient. In some embodiments, they are administered within about 60, 30, 15, 10, 5, or 1 minute of one another. In some embodiments, the administrations of the agents are spaced sufficiently closely together such that a combinatorial (e.g., a synergistic) effect is achieved.

As used herein, the terms “self-replicating RNA,” “self-transcribing and self-replicating RNA,” “self-amplifying RNA (saRNA),” and “replicon” may be used interchangeably, unless context clearly indicates otherwise. Generally, the term “replicon” or “viral replicon” refers to a self-replicating subgenomic RNA derived from a viral genome that includes viral genes encoding non-structural proteins important for viral replication and that lacks viral genes

13

encoding structural proteins. A self-replicating RNA can encode further subgenomic RNAs that are not able to self-replicate.

Nucleic Acid Molecules

In some embodiments, provided herein are nucleic acid molecules comprising: (i) a first polynucleotide encoding one or more viral replication proteins, wherein the first polynucleotide is codon-optimized as compared to a wild-type polynucleotide encoding the one or more viral replication proteins; and (ii) a second polynucleotide comprising a first transgene encoding a first antigenic protein or a fragment thereof.

An RNA molecule can encode a single polypeptide immunogen or multiple polypeptides. Multiple immunogens can be presented as a single polypeptide immunogen (fusion polypeptide) or as separate polypeptides. If immunogens are expressed as separate polypeptides from a replicon then one or more of these may be provided with an upstream IRES or an additional viral promoter element. Alternatively, multiple immunogens may be expressed from a polyprotein that encodes individual immunogens fused to a short autocatalytic protease (e.g., foot-and-mouth disease virus 2A protein), or as inteins.

Also provided herein, in some embodiments, are nucleic acid molecules comprising: (i) a first polynucleotide comprising a sequence having at least 80% identity to a sequence of SEQ ID NO:72; and (ii) a second polynucleotide comprising a first transgene encoding a first antigenic protein or a fragment thereof.

Codon Optimization

In some embodiments, first polynucleotides of nucleic acid molecules provided herein encoding one or more viral replication proteins include codon-optimized sequences. As used herein, the term "codon-optimized" means a polynucleotide, nucleic acid sequence, or coding sequence has been redesigned as compared to a wild-type or reference polynucleotide, nucleic acid sequence, or coding sequence by choosing different codons without altering the amino acid sequence of the encoded protein. Accordingly, codon-optimization generally refers to replacement of codons with synonymous codons to optimize expression of a protein while keeping the amino acid sequence of the translated protein the same. Codon optimization of a sequence can increase protein expression levels (Gustafsson et al., Codon bias and heterologous protein expression. 2004, Trends Biotechnol 22: 346-53) of the encoded proteins, for example, and provide other advantages. Variables such as codon usage preference as measured by codon adaptation index (CAI), for example, the presence or frequency of U and other nucleotides, mRNA secondary structures, cis-regulatory sequences, GC content, and other variables may correlate with protein expression levels (Villalobos et al., Gene Designer: a synthetic biology tool for constructing artificial DNA segments. 2006, BMC Bioinformatics 7:285).

Any method of codon optimization can be used to codon optimize polynucleotides and nucleic acid molecules provided herein, and any variable can be altered by codon optimization. Accordingly, any combination of codon optimization methods can be used. Exemplary methods include the high codon adaptation index (CAI) method, the Low U method, and others. The CAI method chooses a most frequently used synonymous codon for an entire protein coding sequence. As an example, the most frequently used codon for each amino acid can be deduced from 74,218 protein-coding genes from a human genome. The Low U method targets U-containing codons that can be replaced with a synonymous codon with fewer U moieties, generally with-

14

out changing other codons. If there is more than one choice for replacement, the more frequently used codon can be selected. Any polynucleotide, nucleic acid sequence, or codon sequence provided herein can be codon-optimized.

5 This method may be used in conjunction with the disclosed RNAs to design coding sequences that are to be synthesized with, for example, 5-methoxyuridine or N1-methyl pseudouridine. Methods of codon optimization in combination with the use of a modified nucleotide monomer are 10 described in U.S. 2018/0327471, the contents of which are herein incorporated by reference.

In some embodiments, the nucleotide sequence of any region of the RNA or DNA templates described herein may be codon optimized. Preferably, the primary cDNA template 15 may include reducing the occurrence or frequency of appearance of certain nucleotides in the template strand. For example, the occurrence of a nucleotide in a template may be reduced to a level below 25% of said nucleotides in the template. In further examples, the occurrence of a nucleotide 20 in a template may be reduced to a level below 20% of said nucleotides in the template. In some examples, the occurrence of a nucleotide in a template may be reduced to a level below 16% of said nucleotides in the template. Preferably, the occurrence of a nucleotide in a template may be reduced to a level below 15%, and preferably may be reduced to a level below 12% of said nucleotides in the template.

In some embodiments, the nucleotide reduced is uridine. For example, the present disclosure provides nucleic acids 25 with altered uracil content wherein at least one codon in the wild-type sequence has been replaced with an alternative codon to generate a uracil-altered sequence. Altered uracil sequences can have at least one of the following properties:

- 30 (i) an increase or decrease in global uracil content (i.e., the percentage of uracil of the total nucleotide content in the nucleic acid of a section of the nucleic acid, e.g., the open reading frame);
- (ii) an increase or decrease in local uracil content (i.e., changes in uracil content are limited to specific subsequences);
- (iii) a change in uracil distribution without a change in the global uracil content;
- (iv) a change in uracil clustering (e.g., number of clusters, location of clusters, or distance between clusters); or
- (v) combinations thereof.

In some embodiments, the percentage of uracil nucleobases in the nucleic acid sequence is reduced with respect to the percentage of uracil nucleobases in the wild-type nucleic acid sequence. For example, 30% of nucleobases may be uracil in the wild-type sequence but the nucleobases that are uracil are preferably lower than 15%, preferably lower than 12% and preferably lower than 10% of the nucleobases in the nucleic acid sequences of the disclosure. The percentage uracil content can be determined by dividing the number of uracil in a sequence by the total number of nucleotides and multiplying by 100.

In some embodiments, the percentage of uracil nucleobases in a subsequence of the nucleic acid sequence is reduced with respect to the percentage of uracil nucleobases in the corresponding subsequence of the wild-type sequence. 60 For example, the wild-type sequence may have a 5'-end region (e.g., 30 codons) with a local uracil content of 30%, and the uracil content in that same region could be reduced to preferably 15% or lower, preferably 12% or lower and preferably 10% or lower in the nucleic acid sequences of the disclosure. These subsequences can also be part of the wild-type sequences of the heterologous 5' and 3' UTR sequences of the present disclosure.

15

In some embodiments, codons in the nucleic acid sequence of the disclosure reduce or modify, for example, the number, size, location, or distribution of uracil clusters that could have deleterious effects on protein translation. Although lower uracil content is desirable in certain aspects, the uracil content, and in particular the local uracil content, of some subsequences of the wild-type sequence can be greater than the wild-type sequence and still maintain beneficial features (e.g., increased expression).

In some embodiments, the uracil-modified sequence induces a lower Toll-Like Receptor (TLR) response when compared to the wild-type sequence. Several TLRs recognize and respond to nucleic acids. Double-stranded (ds) RNA, a frequent viral constituent, has been shown to activate TLR3. Single-stranded (ss)RNA activates TLR7. RNA oligonucleotides, for example RNA with phosphorothioate internucleotide linkages, are ligands of human TLR8. DNA containing unmethylated CpG motifs, characteristic of bacterial and viral DNA, activate TLR9.

As used herein, the term “TLR response” is defined as the recognition of single-stranded RNA by a TLR7 receptor, and preferably encompasses the degradation of the RNA and/or physiological responses caused by the recognition of the single-stranded RNA by the receptor. Methods to determine and quantify the binding of an RNA to a TLR7 are known in the art. Similarly, methods to determine whether an RNA has triggered a TLR7-mediated physiological response (e.g., cytokine secretion) are well known in the art. In some embodiments, a TLR response can be mediated by TLR3, TLR8, or TLR9 instead of TLR7. Suppression of TLR7-mediated response can be accomplished via nucleoside modification. RNA undergoes over a hundred different nucleoside modifications in nature. Human rRNA, for example, has ten times more pseudouracil ('P) and 25 times more 2'-O-methylated nucleosides than bacterial rRNA. Bacterial RNA contains no nucleoside modifications, whereas mammalian RNAs have modified nucleosides such as 5-methylcytidine (m5C), N6-methyladenosine (m6A), inosine and many 2'-O-methylated nucleosides in addition to N7-methylguanosine (m7G).

In some embodiments, the uracil content of polynucleotides disclosed herein is less than about 50%, 49%, 48%, 47%, 46%, 45%, 44%, 43%, 42%, 41%, 40%, 39%, 38%, 37%, 36%, 35%, 34%, 33%, 32%, 31%, 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% or 1% of the total nucleobases in the sequence in the reference sequence. In some embodiments, the uracil content of polynucleotides disclosed herein is between about 5% and about 25%. In some embodiments, the uracil content of polynucleotides disclosed herein is between about 15% and about 25%.

In some embodiments, first polynucleotides of nucleic acid molecules provided herein comprise a sequence having at least 80%, at least 85%, 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 97.5%, at least 98%, at least 98.5%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, and any number or range in between, identity to a sequence of SEQ ID NO:72. In some embodiments, first polynucleotides of nucleic acid molecules provided herein comprise a sequence of SEQ ID NO:72.

In some aspects, first polynucleotides and second polynucleotides of nucleic acid molecules provided herein are included in the same (i.e., a single) or in separate nucleic acid molecules. Generally, first polynucleotides and second

16

polynucleotides of nucleic acid molecules provided herein are included in a single nucleic acid molecule. In one aspect, the first polynucleotide is located 5' of the second polynucleotide. In one aspect, first polynucleotides and second polynucleotides of nucleic acid molecules provided herein are included in separate nucleic acid molecules. In yet another aspect, first polynucleotides and second polynucleotides are included in two separate nucleic acid molecules.

In some aspects, first polynucleotides and second polynucleotides are included in the same (i.e., a single) nucleic acid molecule. First polynucleotides and second polynucleotides of nucleic acid molecules provided herein can be contiguous, i.e., adjacent to each other without nucleotides in between. In one aspect, an intergenic region is located between the first polynucleotide and the second polynucleotide. In another aspect, the intergenic region located between the first polynucleotide and the second polynucleotide is a second intergenic region, with a first intergenic region included in the first polynucleotide as described below. As used herein, the terms “intergenic region” and “intergenic sequence” can be used interchangeably, unless context clearly indicates otherwise.

An intergenic region located between the first polynucleotide and the second polynucleotide can be of any length and can have any nucleotide sequence. As an example, the intergenic region between the first polynucleotide and the second polynucleotide can include about one nucleotide, about two nucleotides, about three nucleotides, about four nucleotides, about five nucleotides, about six nucleotides, about seven nucleotides, about eight nucleotides, about nine nucleotides, about ten nucleotides, about 11 nucleotides, about 12 nucleotides, about 13 nucleotides, about 14 nucleotides, about 15 nucleotides, about 16 nucleotides, about 17 nucleotides, about 18 nucleotides, about 19 nucleotides, about 20 nucleotides, about 21 nucleotides, about 22 nucleotides, about 23 nucleotides, about 24 nucleotides, about 25 nucleotides, about 26 nucleotides, about 27 nucleotides, about 28 nucleotides, about 29 nucleotides, about 30 nucleotides, about 31 nucleotides, about 32 nucleotides, about 33 nucleotides, about 34 nucleotides, about 35 nucleotides, about 36 nucleotides, about 37 nucleotides, about 38 nucleotides, about 39 nucleotides, about 40 nucleotides, about 41 nucleotides, about 42 nucleotides, about 43 nucleotides, about 44 nucleotides, about 45 nucleotides, about 46 nucleotides, about 47 nucleotides, about 48 nucleotides, about 49 nucleotides, about 50 nucleotides, about 60 nucleotides, about 70 nucleotides, about 80 nucleotides, about 90 nucleotides, about 100 nucleotides, about 125 nucleotides, about 150 nucleotides, about 175 nucleotides, about 200 nucleotides, about 250 nucleotides, about 300 nucleotides, about 350 nucleotides, about 400 nucleotides, about 450 nucleotides, about 500 nucleotides, about 600 nucleotides, about 700 nucleotides, about 800 nucleotides, about 1,000 nucleotides, about 1,500 nucleotides, about 2,000 nucleotides, about 2,500 nucleotides, about 3,000 nucleotides, about 3,500 nucleotides, about 4,000 nucleotides, about 4,500 nucleotides, about 5,000 nucleotides, about 6,000 nucleotides, about 7,000 nucleotides, about 8,000 nucleotides, about 9,000 nucleotides, about 10,000 nucleotides, and any number or range in between. In one aspect, the intergenic region between first and second polynucleotides includes about 10-100 nucleotides, about 10-200 nucleotides, about 10-300 nucleotides, about 10-400 nucleotides, or about 10-500 nucleotides. In another aspect, the intergenic region between first and second polynucleotides includes about 1-10 nucleotides, about 1-20 nucleotides, about 1-30 nucleotides, about 1-40 nucleotides, or about 1-50 nucleotides. In

yet another aspect, the region includes about 44 nucleotides. In one aspect, the intergenic region between first and second polynucleotides of nucleic acid molecules provided herein is a second intergenic region.

In one aspect, the intergenic region between first and second polynucleotides includes a viral sequence. The intergenic region between first and second polynucleotides can include a sequence from any virus, such as alphaviruses and rubiviruses, for example. In one aspect, the intergenic region between the first polynucleotide and the second polynucleotide comprises an alphavirus sequence, such as a sequence from Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MIDV), Chikungunya Virus (CHIKV), ONYong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiyama Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (STNV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), Buggy Creek Virus (BCRV), or any combination thereof. In another aspect, the intergenic region between first and second polynucleotides comprises a sequence from Venezuelan Equine Encephalitis Virus (VEEV). In yet another aspect, the intergenic region between first and second polynucleotides comprises a sequence having at least 85%, 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 97.5%, at least 98%, at least 98.5%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, and any number or range in between, identity to SEQ ID NO:77. In a further aspect, the intergenic region between first and second polynucleotides comprises a sequence of SEQ ID NO:77. In yet a further aspect, the intergenic region between first and second polynucleotides is a second intergenic region comprising a sequence having at least 85% identity to SEQ ID NO:77.

Natural and Modified Nucleotides

A self-replicating RNA of the disclosure can comprise one or more chemically modified nucleotides. Examples of nucleic acid monomers include non-natural, modified, and chemically-modified nucleotides, including any such nucleotides known in the art. Nucleotides can be artificially modified at either the base portion or the sugar portion. In nature, most polynucleotides comprise nucleotides that are "unmodified" or "natural" nucleotides, which include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). These bases are typically fixed to a ribose or deoxy ribose at the 1' position. The use of RNA polynucleotides comprising chemically modified nucleotides have been shown to improve RNA expression, expression rates, half-life and/or expressed protein concentrations. RNA polynucleotides comprising chemically modified nucleotides have also been useful in optimizing protein localization thereby avoiding deleterious bio-responses such as immune responses and/or degradation pathways.

Examples of modified or chemically-modified nucleotides include 5-hydroxycytidines, 5-alkylcytidines, 5-hydroxymethylcytidines, 5-carboxycytidines, 5-formylcytidines, 5-alkoxycytidines, 5-alkynylcytidines, 5-halocytidines, 2-thiocytidines, N4-alkylcytidines, N4-aminocytidines, N4-acetylcytidines, and N4,N4-dialkylcytidines.

Examples of modified or chemically-modified nucleotides include 5-hydroxycytidine, 5-methylcytidine, 5-hydroxymethylcytidine, 5-carboxycytidine, 5-formylcytidine, 5-methoxycytidine, 5-propynylcytidine, 5-bromocytidine, 5-iodocytidine, 2-thiocytidine; N4-methylcytidine, N4-aminocytidine, N4-acetylcytidine, and N4,N4-dimethylcytidine.

Examples of modified or chemically-modified nucleotides include 5-hydroxyuridines, 5-alkyluridines, 5-hydroxyalkyluridines, 5-carboxyuridines, 5-carboxyalkylesteruridines, 5-formyluridines, 5-alkoxyuridines, 5-alkynyluridines, 5-halouridines, 2-thiouridines, and 6-alkyluridines.

Examples of modified or chemically-modified nucleotides include 5-hydroxyuridine, 5-methyluridine, 5-hydroxymethyluridine, 5-carboxyuridine, 5-carboxymethylesteruridine, 5-formyluridine, 5-methoxyuridine (also referred to herein as "5MeOU"), 5-propynyluridine, 5-bromouridine, 5-fluorouridine, 5-iodouridine, 2-thiouridine, and 6-methyluridine.

Examples of modified or chemically-modified nucleotides include 5-methoxycarbonylmethyl-2-thiouridine, 5-methylaminomethyl-2-thiouridine, 5-carbamoylmethyluridine, 5-carbamoylmethyl-2'-O-methyluridine, 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine, 5-methylaminomethyl-2-selenouridine, 5-carboxymethyluridine, 5-methyldihydouridine, 5-taurinomethyluridine, 5-taurinomethyl-2-thiouridine, 5-(isopentenylaminomethyl)uridine, 2'-O-methylpseudouridine, 2-thio-2'-O-methyluridine, and 3,2'-O-dimethyluridine.

Examples of modified or chemically-modified nucleotides include N6-methyladenosine, 2-aminoadenosine, 3-methyladenosine, 8-azaadenosine, 7-deazaadenosine, 8-oxoadenosine, 8-bromo-adenosine, 2-methylthio-N6-methyladenosine, N6-isopentenyladenosine, 2-methylthio-N6-isopentenyladenosine, N6-(cis-hydroxyisopentenyl)

adenosine, 2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine, N6-glycylcarbamoyladenine, N6-threonylcarbamoyl-adenosine, N6-methyl-N6-threonylcarbamoyl-adenosine, 2-methylthio-N6-threonylcarbamoyl-adenosine, N6,N6-dimethyladenosine, N6-hydroxynorvalylcarbamoyladenine, 2-methylthio-N6-hydroxynorvalylcarbamoyl-adenosine, N6-acetyl-adenosine, 7-methyl-adenine, 2-methylthio-adenine, 2-methoxy-adenine, alpha-thio-adenosine, 2'-O-methyl-adenosine, N6,2'-O-dimethyl-adenosine, N6,N6,2'-O-trimethyl-adenosine, 1,2'-O-dimethyl-adenosine, 2'-O-ribosyladenosine, 2-amino-N6-methyl-purine, 1-thio-adenosine, 2'-F-ara-adenosine, 2'-F-adenosine, 2'-OH-ara-adenosine, and N6-(19-amino-pentaoxanonadecyl)-adenosine.

Examples of modified or chemically-modified nucleotides include N1-alkylguanosines, N2-alkylguanosines, thienoguanosines, 7-deazaguanosines, 8-oxoguanosines, 8-bromoguanosines, O6-alkylguanosines, xanthosines, inosines, and N1-alkylinosines.

Examples of modified or chemically-modified nucleotides include N1-methylguanosine, N2-methylguanosine, thienoguanosine, 7-deazaguanosine, 8-oxoguanosine, 8-bromoguanosine, O6-methylguanosine, xanthosine, inosine, and N1-methylinosine.

Examples of modified or chemically-modified nucleotides include pseudouridines. Examples of pseudouridines include Nhalkylpseudouridines, N1-cycloalkylpseudouridines, N1-hydroxypseudouridines, N1-phenylalkylpseudouridines, N1-aminoalkylpseudouridines, N3-alkylpseudouridines, N6-alkylpseudouridines, N6-alkoxypseudouridines, N6-hydroxypseudouridines, N6-hydroxyalkylpseudouridines, N6-morpholinopseudouri-

dines, N6-phenylpseudouridines, and N6-halopseudouridines. Examples of pseudouridines include N1-alkyl-N6-alkylpseudouridines, N1-alkyl-N6-alkoxypseudouridines, N1-alkyl-N6-hydroxypseudouridines, N1-alkyl-N6-hydroxyalkylpseudouridines, N1-alkyl-N6-morpholinopseudouridines, N1-alkyl-N6-phenylpseudouridines, and N1-alkyl-N6-halopseudouridines. In these examples, the alkyl, cycloalkyl, and phenyl substituents may be unsubstituted, or further substituted with alkyl, halo, haloalkyl, amino, or nitro substituents.

Examples of pseudouridines include N1-methylpseudouridine (also referred to herein as "N1MPU"), N1-ethylpseudouridine, N1-propylpseudouridine, N1-cyclopropylpseudouridine, N1-phenylpseudouridine, N1-aminomethylpseudouridine, N3-methylpseudouridine, N1-hydroxypseudouridine, and N1-hydroxymethylpseudouridine.

Examples of nucleic acid monomers include modified and chemically-modified nucleotides, including any such nucleotides known in the art.

Examples of modified and chemically-modified nucleotide monomers include any such nucleotides known in the art, for example, 2'-O-methyl ribonucleotides, 2'-O-methyl purine nucleotides, 2'-deoxy-2'-fluoro ribonucleotides, 2'-deoxy-2'-fluoro pyrimidine nucleotides, 2'-deoxy ribonucleotides, 2'-deoxy purine nucleotides, universal base nucleotides, 5-C-methyl-nucleotides, and inverted deoxyabasic monomer residues.

Examples of modified and chemically-modified nucleotide monomers include 3'-end stabilized nucleotides, 3'-glyceryl nucleotides, 3'-inverted abasic nucleotides, and 3'-inverted thymidine.

Examples of modified and chemically-modified nucleotide monomers include locked nucleic acid nucleotides (LNA), 2'-O,4'-C-methylene-(D-ribofuranosyl) nucleotides, 2'-methoxyethoxy (MOE) nucleotides, 2'-methyl-thio-ethyl, 2'-deoxy-2'-fluoro nucleotides, and 2'-O-methyl nucleotides. In an exemplary embodiment, the modified monomer is a locked nucleic acid nucleotide (LNA).

Examples of modified and chemically-modified nucleotide monomers include 2',4'-constrained 2'-O-methoxyethyl (cMOE) and 2'-O-Ethyl (cEt) modified DNAs.

Examples of modified and chemically-modified nucleotide monomers include 2'-amino nucleotides, 2'-O-amino nucleotides, 2'-C-allyl nucleotides, and 2'-O-allyl nucleotides.

Examples of modified and chemically-modified nucleotide monomers include N6-methyladenosine nucleotides.

Examples of modified and chemically-modified nucleotide monomers include nucleotide monomers with modified bases 5-(3-amino)propyluridine, 5-(2-mercapto)ethyluridine, 5-bromouridine; 8-bromoguanosine, or 7-deazaadenosine.

Examples of modified and chemically-modified nucleotide monomers include 2'-O-aminopropyl substituted nucleotides.

Examples of modified and chemically-modified nucleotide monomers include replacing the 2'-OH group of a nucleotide with a 2'-R, a 2'-OR, a 2'-halogen, a 2'-SR, or a 2'-amino, where R can be H, alkyl, alkenyl, or alkynyl.

Example of base modifications described above can be combined with additional modifications of nucleoside or nucleotide structure, including sugar modifications and linkage modifications. Certain modified or chemically-modified nucleotide monomers may be found in nature.

Preferred nucleotide modifications include N1-methylpseudouridine and 5-methoxyuridine.

Viral Replication Proteins and Polynucleotides Encoding Them

Provided herein, in some embodiments, are nucleic acid molecules comprising a first polynucleotide encoding one or more viral replication proteins. As used herein, the term "replication protein" or "viral replication protein" refers to any protein or any protein subunit of a protein complex that functions in replication of a viral genome. Generally, viral replication proteins are non-structural proteins. Viral replication proteins encoded by nucleic acid molecules provided herein can function in the replication of any viral genome. The viral genome can be a single-stranded positive-sense RNA genome, a single-stranded negative-sense RNA genome, a double-stranded RNA genome, a single-stranded positive-sense DNA genome, a single-stranded negative-sense DNA genome, or a double-stranded DNA genome. Viral genomes can include a single nucleic acid molecule or more than one nucleic acid molecule. Nucleic acid molecules provided herein can encode one or more viral replication proteins from any virus or virus family, including animal viruses and plant viruses, for example. Viral replication proteins encoded by first polynucleotides included in nucleic acid molecules provided herein can be expressed from self-replicating RNA.

First polynucleotide sequences of nucleic acid molecules provided herein can encode one or more togavirus replication proteins. In some aspects, the one or more viral replication proteins encoded by first polynucleotides of nucleic acid molecules provided herein are alphavirus proteins. In some embodiments, the one or more viral replication proteins encoded by first polynucleotides of nucleic acid molecules provided herein are rubivirus proteins. First polynucleotide sequences of nucleic acid molecules provided herein can encode any alphavirus replication protein and any rubivirus replication protein. Exemplary replication proteins from alphaviruses include proteins from Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagi-yama Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (SINV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzylagach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), Buggy Creek Virus (BCRV), and any combination thereof. Exemplary rubivirus replication proteins include proteins from rubella virus.

Viral replication proteins encoded by first polynucleotides of nucleic acid molecules provided herein can be expressed as one or more polyproteins or as separate or single proteins. Generally, polyproteins are precursor proteins that are cleaved to generate individual or separate proteins. Accordingly, proteins derived from a precursor polyprotein can be expressed from a single open reading frame (ORF). As used herein, the term "ORF" refers to a nucleotide sequence that begins with a start codon, generally ATG, and that ends with a stop codon, such as TAA, TAG, or TGA, for example. It will be appreciated that T is present in DNA, while U is present in RNA. Accordingly, a start codon of ATG in DNA corresponds to AUG in RNA, and the stop codons TAA, TAG, and TGA in DNA correspond to UAA, UAG, and UGA in RNA. It will further be appreciated that for any sequence provided in the present disclosure, T is present in

DNA, while U is present in RNA. Accordingly, for any sequence provided herein, T present in DNA is substituted with U for an RNA molecule, and U present in RNA is substituted with T for a DNA molecule.

The protease cleaving a polyprotein can be a viral protease or a cellular protease. In some aspects, the first polynucleotide of nucleic acid molecules provided herein encodes a polyprotein comprising an alphavirus nsP1 protein, an alphavirus nsP2 protein, an alphavirus nsP3 protein, an alphavirus nsP4 protein, or any combination thereof. In other aspects, the first polynucleotide of nucleic acid molecules provided herein encodes a polyprotein comprising an alphavirus nsP1 protein, an alphavirus nsP2 protein, an alphavirus nsP3 protein, or any combination thereof, and an alphavirus nsP4 protein. In some aspects, the polyprotein is a VEEV polyprotein. In other aspects, the alphavirus nsP1, nsP2, nsP3, and nsP4 proteins are VEEV proteins.

In one aspect, first polynucleotides of nucleic acid molecules provided herein lack a stop codon between sequences encoding an nsP3 protein and an nsP4 protein. Accordingly, in some aspects, first polynucleotides of nucleic acid molecules provided herein encode a P1234 polyprotein comprising nsP1, nsP2, nsP3, and nsP4. First polynucleotides of nucleic acid molecules provided herein can also include a stop codon between sequences encoding an nsP3 and an nsP4 protein. Accordingly, in some aspects, first polynucleotides of nucleic acid molecules provided herein encode a P123 polyprotein comprising nsP1, nsP2, and nsP3 and a P1234 polyprotein comprising nsP1, nsP2, nsP3, and nsP4 as a result of stop codon readthrough, for example. In other aspects, first polynucleotides of nucleic acid molecules provided herein encode a polyprotein having at least 80%, at least 85%, 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 97.5%, at least 98%, at least 98.5%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, and any number or range in between, identity to a sequence of SEQ ID NO:79. In some embodiments, first polynucleotides of nucleic acid molecules provided herein encode a polyprotein having a sequence of SEQ ID NO:79. Further exemplary polyproteins comprise a sequence of SEQ ID NO:80 or SEQ ID NO:81. In one aspect, nsP2 and nsP3 proteins include mutations. Exemplary mutations include G1309R and S1583G mutations of VEEV proteins. In another aspect, the nsP1, nsP2, and nsP4 proteins are VEEV proteins, and the nsP3 protein is a chikungunya virus (CHIKV) nsP3 protein.

In some aspects, first polynucleotides of nucleic acid molecules provided herein can include a first intergenic region. In some aspects, the first intergenic region is located between a sequence encoding a polyprotein comprising an alphavirus nsP1 protein, an alphavirus nsP2 protein, an alphavirus nsP3 protein, or any combination thereof, and a sequence encoding an alphavirus nsP4 protein. A first intergenic region can comprise any sequence, such as any viral or non-viral sequence. In one aspect, the first intergenic region comprises a viral sequence. In another aspect, the first intergenic region comprises an alphavirus sequence. In yet another aspect, the alphavirus is VEEV. In one aspect, nsP2 and nsP3 proteins include mutations. Exemplary mutations include G1309R and S1583G mutations of VEEV proteins. In another aspect, the nsP1, nsP2, and nsP4 proteins are VEEV proteins, and the nsP3 protein is a chikungunya virus (CHIKV) nsP3 protein.

5' Untranslated Region (5' UTR)

Nucleic acid molecules provided herein can further comprise untranslated regions (UTRs). Untranslated regions,

including 5' UTRs and 3' UTRs, for example, can affect RNA stability and/or efficiency of RNA translation, such as translation of cellular and viral mRNAs, for example. 5' UTRs and 3' UTRs can also affect stability and translation of 5' viral genomic RNAs and self-replicating RNAs, including virally derived self-replicating RNAs or replicons. Exemplary viral genomic RNAs whose stability and/or efficiency of translation can be affected by 5' UTRs and 3' UTRs include the genome nucleic acid of positive-sense RNA viruses. Both genome nucleic acid of positive-sense RNA viruses and self-replicating RNAs, including virally derived self-replicating RNAs or replicons, can be translated upon infection or introduction into a cell.

In some aspects, nucleic acid molecules provided herein further include a 5' untranslated region (5' UTR). Any 5' UTR sequence can be included in nucleic acid molecules provided herein. In some embodiments, nucleic acid molecules provided herein include a viral 5' UTR. In one aspect, nucleic acid molecules provided herein include a non-viral 5' UTR. Any non-viral 5' UTR can be included in nucleic acid molecules provided herein, such as 5' UTRs of transcripts expressed in any cell or organ, including muscle, skin, subcutaneous tissue, liver, spleen, lymph nodes, antigen-presenting cells, and others. In another aspect, nucleic acid molecules provided herein include a 5' UTR comprising viral and non-viral sequences. Accordingly, a 5' UTR included in nucleic acid molecules provided herein can comprise a combination of viral and non-viral 5' UTR sequences. In some aspects, the 5' UTR included in nucleic acid molecules provided herein is located upstream of or 5' of the first polynucleotide that encodes one or more viral replication proteins. In other aspects, the 5' UTR is located 5' of or upstream of the first polynucleotide of nucleic acid molecules provided herein that encodes one or more viral replication proteins, and the first polynucleotide is located 5' of or upstream of the second polynucleotide of nucleic acid molecules provided herein.

In one aspect, the 5' UTR of nucleic acid molecules provided herein comprises an alphavirus 5' UTR. A 5' UTR from any alphavirus can be included in nucleic acid molecules provided herein, including 5' UTR sequences from Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus (SFV), Pixuna Virus (PIXV), Middleburg Virus (MIDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiyama Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (STNV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), or Buggy Creek Virus (BCRV). In another aspect, the 5' UTR comprises a sequence having at least 80%, at least 85%, 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 97.5%, at least 98%, at least 98.5%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, and any number or range in between, identity to a sequence of SEQ ID NO:73, SEQ ID NO:74, or SEQ ID NO:75. In yet another aspect, the 5' UTR comprises a sequence of SEQ ID NO:73, SEQ ID NO:74, or SEQ ID NO:75.

In some embodiments, the 5' UTR comprises a sequence selected from the 5' UTRs of human IL-6, alanine amino-

transferase 1, human apolipoprotein E, human fibrinogen alpha chain, human transthyretin, human haptoglobin, human alpha-1-antichymotrypsin, human antithrombin, human alpha-1-antitrypsin, human albumin, human beta globin, human complement C3, human complement C5, SynK (thylakoid potassium channel protein derived from the cyanobacteria, *Synechocystis* sp.), mouse beta globin, mouse albumin, and a tobacco etch virus, or fragments of any of the foregoing. Preferably, the 5' UTR is derived from a tobacco

etch virus (TEV). Preferably, an mRNA described herein comprises a 5' UTR sequence that is derived from a gene expressed by *Arabidopsis thaliana*. Preferably, the 5' UTR sequence of a gene expressed by *Arabidopsis thaliana* is AT1G58420. Examples of 5' UTRs and 3' UTRs are described in PCT/US2018/035419, the contents of which are herein incorporated by reference. Preferred 5' UTR sequences comprise SEQ ID NOs: 5, 25-27 and 28-45: as shown in Table 1.

TABLE 1

5' UTR Sequences		
Name	Sequence	Seq ID No.:
EV	UCAACACAACAUAAUACAAAAACAAACGAAUCUCAAGCAAUC AAGCAUUCUACUUUCUAAUUGCAGCAUUUAAAUAUUCU UUUAAAGCAAAGCAUUUCUGAAUAAAUCACCAAUU ACGAACGGAUAG	SEQ ID NO: 5
AT1G58420	AUUUAUUAUCAUAAAACAAAAGCCGCCA	SEQ ID NO: 6
ARC5-2	CUUUAGGGGGCGCUGCCUACGGAGGGUGGCAGGCCAUCCU UCUCGGCAUCAAGCUUACCAUGGUCCCCAGGGCCUGUC UUGGUCCCCUGCUGGUUGUCCCCUUCUGCUUCGGCAAGU UCCCCAUCAUCAACCAUCCCCGACAAGCUGGGGGCGUGGAG CCCCAUCAUCACAUCCCCACCUGUCUCCUACCAACCUUG UGGUCAUGGACGAGGGCUGCACCAACCUGAGCGGGGUUCU CUAC	SEQ ID NO: 7
HCV	UGAGUGUCGU ACAGGCCUCCA GCCCCCCCC UCCCGGGAGA GCCAUAGUGG UCUGCGGAACCGGGAGAGUAC ACCGGAAUUG CCGGGAAGAC UGGGUCCUUU CUUGGAUAAA CCCAUCUUAUGCCCCGCCAU UGGGGCGUGC CCCCGCAAGA CUGCUAGCCG AGUAGUGUUG GGUUGCG	SEQ ID NO: 8
HUMAN ALBUMIN	AAUUUAUUGGUAAAAGAAGUUAUUAUGUGCUAAUUCU CCGUUUGUCCUAGCUUUUCUUCUUCUGCUAACCCACACGC CUUUGGCACA	SEQ ID NO: 9
EMCV	CCCCCUCCCC CCCCUUAAC GUUACUGGCC GAAGCCGUU GGAAUAAGGC CGGUGUGCGU UUGUCUUAU GUUAUUUUCC ACCAUAUUGC CGUCUUUUGG CAAUGUGAGG CCCCGGAAAC CUGGCCUUGU CUUCUUGACG AGCAUCCUA GGGGCUUUUC CCCUCUCGCC AAAGGAUUGC AAGGUUCGUU GAAUGUCUG AAGGAAGCAG UUCCUCUGGA AGCUUCUUGA AGACAACCAA CGUGUGUAGC GACCUUUUGC AGGCAGCGGA ACCCCCCACC UGGCGACAGG UGCCUCUGCG GCCAAAGGCC ACGUGUUAAGA GAUACACCUG CAAAGGGCGC ACAACCCCG AGCCACGUUG UGAGUUGGAU AGUUGUGGAA AGAGUAAAU GGCUCUCCUC AAGCGUAUUC AACAAAGGGC UGAAGGAUGC CCAGAAGGU CCCCAUUGUA UGGGUACUGA UCUGGGGCCU CGGUGCACAU GCUUUACUGG UGUUUAUGUCG AGGUUAAA ACGUCUAGGC CCCCGAACC ACGGGGACGU GGUUUUCUUU UGAAAACAC GAUGAUAAU	SEQ ID NO: 10
AT1G67090	CACAAAGAGUAAAAGAAGAACAA	SEQ ID NO: 25
AT1G35720	AACACUAAAAGUAGAAGAAAA	SEQ ID NO: 26
AT5G45900	CUCAGAAAGUAAGAACAGCC	SEQ ID NO: 27
AT5G61250	AACCAAUCGAAAGAACACAAA	SEQ ID NO: 28
AT5G46430	CUCUAAUCACCAGGAGUAAA	SEQ ID NO: 29
AT5G47110	GAGAGAGAUCUUAACAAAAAA	SEQ ID NO: 30
AT1G03110	UGUGUAACAAACAACAAACA	SEQ ID NO: 31
AT3G12380	CCGCAGUAGGAAGAGAACGCC	SEQ ID NO: 32
AT5G45910	AAAAAAAAAGAAAUCAUAAA	SEQ ID NO: 33

TABLE 1-continued

5' UTR Sequences		
Name	Sequence	Seq ID No.:
AT1G07260	GAGAGAAGAAGAAGAAGACG	SEQ ID NO: 34
AT3G55500	CAAUUUAAAUAUCUUAACAAA	SEQ ID NO: 35
AT3G46230	GCAAAACAGAGUAAGCGAAACG	SEQ ID NO: 36
AT2G36170	GCGAAGAAGACGAACGCAAAG	SEQ ID NO: 37
AT1G10660	UUAGGACUGUAUUGACUGGCC	SEQ ID NO: 38
AT4G14340	AUCAUCGGAAUCGGAAAAAG	SEQ ID NO: 39
AT1G49310	AAAACAAAAGUUAAGCAGAC	SEQ ID NO: 40
AT4G14360	UUUAUCUCAAAUAAGAAGGCA	SEQ ID NO: 41
AT1G28520	GGUGGGGAGGUGAGAUUUCUU	SEQ ID NO: 42
AT1G20160	UGAUUAGGAAACUACAAAGCC	SEQ ID NO: 43
AT5G37370	CAUUUUUCAUUUCAUAAAAC	SEQ ID NO: 44
AT4G11320	UUACUUUUUAGCCCAACAAAA	SEQ ID NO: 45
AT5G40850	GGCGUGUGUGUGUGUUGUUGA	SEQ ID NO: 46
AT1G06150	GUGGUGAAGGGGAAGGUUAG	SEQ ID NO: 47
AT2G26080	UUGUUUUUUUUUGGUUUGGUU	SEQ ID NO: 48

3' Untranslated Region (3' UTR)

In some aspects, nucleic acid molecules provided herein further include a 3' untranslated region (3' UTR). Any 3' UTR sequence can be included in nucleic acid molecules provided herein. In one aspect, nucleic acid molecules provided herein include a viral 3' UTR. In another aspect, nucleic acid molecules provided herein include a non-viral 3' UTR. Any non-viral 3' UTR can be included in nucleic acid molecules provided herein, such as 3' UTRs of transcripts expressed in any cell or organ, including muscle, skin, subcutaneous tissue, liver, spleen, lymph nodes, antigen-presenting cells, and others. In some aspects, nucleic acid molecules provided herein include a 3' UTR comprising viral and non-viral sequences. Accordingly, a 3' UTR included in nucleic acid molecules provided herein can comprise a combination of viral and non-viral 3' UTR sequences. In one aspect, the 3' UTR is located 3' of or downstream of the second polynucleotide of nucleic acid molecules provided herein that comprises a first transgene encoding a first antigenic protein or a fragment thereof. In another aspect, the 3' UTR is located 3' of or downstream of the second polynucleotide of nucleic acid molecules provided herein that comprises a first transgene encoding a first antigenic protein or a fragment thereof, and the second polynucleotide is located 3' of or downstream of the first polynucleotide of nucleic acid molecules provided herein.

In one aspect, the 3' UTR of nucleic acid molecules provided herein comprises an alphavirus 3' UTR. A 3' UTR from any alphavirus can be included in nucleic acid molecules provided herein, including 3' UTR sequences from Venezuelan Equine Encephalitis Virus (VEEV), Eastern Equine Encephalitis Virus (EEEV), Everglades Virus (EVEV), Mucambo Virus (MUCV), Semliki Forest Virus

(SFV), Pixuna Virus (PIXY), Middleburg Virus (MDV), Chikungunya Virus (CHIKV), O'Nyong-Nyong Virus (ONNV), Ross River Virus (RRV), Barmah Forest Virus (BFV), Getah Virus (GETV), Sagiyma Virus (SAGV), Bebaru Virus (BEBV), Mayaro Virus (MAYV), Una Virus (UNAV), Sindbis Virus (SINV), Aura Virus (AURAV), Whataroa Virus (WHAV), Babanki Virus (BABV), Kyzyl-agach Virus (KYZV), Western Equine Encephalitis Virus (WEEV), Highland J Virus (HJV), Fort Morgan Virus (FMV), Ndumu Virus (NDUV), Salmonid Alphavirus (SAV), or Buggy Creek Virus (BCRV). In another aspect, the 3' UTR comprises a sequence having at least 80%, at least 85%, 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 97.5%, at least 98%, at least 98.5%, at least 99%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, at least 99.9%, and any number or range in between, identity to a sequence of SEQ ID NO:5. In yet another aspect, the 3' UTR comprises a poly-A sequence. In a further aspect, the 3' UTR comprises a sequence of SEQ ID NO:5.

In some embodiments, the 3' UTR comprises a sequence selected from the 3' UTRs of alanine aminotransferase 1, human apolipoprotein E, human fibrinogen alpha chain, human haptoglobin, human antithrombin, human alpha globin, human beta globin, human complement C3, human growth factor, human hepcidin, MALAT-1, mouse beta globin, mouse albumin, and *Xenopus* beta globin, or fragments of any of the foregoing. In some embodiments, the 3' UTR is derived from *Xenopus* beta globin. Exemplary 3' UTR sequences include SEQ ID NOS: 16-22 as shown in Table 2.

TABLE 2

3' UTR sequences.		
Name	Sequence	Seq ID No.:
XBG	CUAGUGACUGACUAGGAUCUGGUUACCACUAACAG CCUCAAGAACACCGAAUGGGAGUCUUAAGCUACAU AUACCAACUUACACUUACAAAAGUUGUCCCCAAAA UGUAGGCCAUUCGUUAUCUGCUCCUAAUAAAAGAAAGU UUCUUCACAU	SEQ ID NO: 16
HUMAN HAPTOGLOBIN	UGCAAGGCUGGCCGGAAGCCCUGCCUGAAAGCAAGA UUUCAGCUGGAAGAGGGCAAAGUGGACGGGAGUGG ACAGGGAGUGGAUGCGAUAAGAUGGGUUGAAGCUG AUGGGUGCCACCCUGCAUUGCUGACUCAUCAAUA AGAGCUUUCUUUUGACCCAU	SEQ ID NO: 17
HUMAN APOLIPOPROTEINE	ACGCCGAAGGCCUGCAGCCAUGCGACCCCCACGCCACCCC GUGCUCUCUGCCUCCGCGCAGCCUGCAGCGGGAGACC CUGUCCCCGCCCCAGCCGUCCUCCUGGGUGGACCCU AGUUUUAAAAGAUUCACCAAGUUUCACGCA	SEQ ID NO: 18
HCV	UAGAGCGGCAACCCUAGCUACACUCCAUAGCUAGUU UCUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU UCCUCUUUUUUUUGGUUGGUCCAUUCUAGGCCUAGUC ACGGCUAGCUGUGAAAGGUCCUGAGCCGAUGACUG CAGAGAGUGCCGUAACUGGUUCUCUGCAGAUCAUGU	SEQ ID NO: 19
MOUSE ALBUMIN	ACACAUCACAACCAACCUUCUCAGGUACCCUGAG AAAAAAAGACAUAGAAGACUCAGGACUCAUUUUCUG UUGGUAAAUAACACACCUAAGGAACACAAUUC UUAAAACAUUUGACUUUCUUGUCUGCUGCAUUA AUAAAAAAUUGGAAAGAAUCUAC	SEQ ID NO: 20
HUMAN ALPHA GLOBIN	GCUGGAGCCUCGGUAGCCGUUCCCCUCCUGGCCUGGG CCUCCCAACGGGCCCUCCCUCCUUGCACCGGCC UCCUGGUUUUGAAUAAAGUCUGAGUGGGCAGCA	SEQ ID NO: 21
EMCV	UAGUGCAGUCAC UGGCACAACG CGUUGCCGG UAAGCCAACU GGGUAUACAC GGUCGUCAUACUGCAGACAG GGUUCUUCUA CUUUGCAAGA UAGCUAGAG UAGUAAAAUA AAUAGUUAAG	SEQ ID NO: 22

Triple Stop Codon

In some embodiments, the self-replicating RNA may comprise a sequence immediately downstream of a coding region (i.e., ORF) that creates a triple stop codon. A triple stop codon is a sequence of three consecutive stop codons. The triple stop codon can ensure total insulation of an expression cassette and may be incorporated to enhance the efficiency of translation. In some embodiments, a self-replicating RNA of the disclosure may comprise a triple combination of any of the sequences UAG, UGA, or UAA immediately downstream of a ORF described herein. The triple combination can be three of the same codons, three different codons, or any other permutation of the three stop codons.

Translation Enhancers and Kozak Sequences

For translation initiation, proper interactions between ribosomes and mRNAs must be established to determine the exact position of the translation initiation region. However, ribosomes also must dissociate from the translation initiation

region to slide toward the downstream sequence during mRNA translation. Translation enhancers upstream from initiation sequences of mRNAs enhance the yields of protein biosynthesis. Several studies have investigated the effects of translation enhancers. In some embodiments, an mRNA described herein comprises a translation enhancer sequence. These translation enhancer sequences enhance the translation efficiency of a self-replicating RNA of the disclosure and thereby provide increased production of the protein encoded by the mRNA. The translation enhancer region may be located in the 5' or 3' UTR of an mRNA sequence. Examples of translation enhancer regions include naturally-occurring enhancer regions from the TEV 5' UTR and the *Xenopus* beta-globin 3' UTR. Exemplary 5' UTR enhancer sequences include but are not limited to those derived from mRNAs encoding human heat shock proteins (HSP) including HSP70-P2, HSP70-M1 HSP72-M2, HSP17.9 and HSP70-P1. Preferred translation enhancer sequences used in accordance with the embodiments of the present disclosure are represented by SEQ ID Nos: 11-15 as shown in Table 3.

TABLE 3

5' UTR Enhancers		
Name	Sequence	Seq ID No.:
HSP70-P2	GUCAGCUUCAACUCUUUGUUUCUUGUUUGAUUGAGAA	SEQ ID NO: 11 UA

TABLE 3-continued

5' UTR Enhancers		
Name	Sequence	Seq ID No.:
HSP70-M1	CUCUCGCCUGAGAAAAAAAUCACGAACCAUUUCAGCA	SEQ ID NO: 12
	ACCAGCAGCACG	
HSP72-M2	ACCUGUGAGGGUUCGAAAGGAUAGCAGUGUUUUUUCGUCCU	SEQ ID NO: 13
	AGAGGAAGAG	
HSP17.9	ACACAGAAACAUUCGCAAAACAAAUCCCAGUAUCAAAAUU	SEQ ID NO: 14
	CUUCUCUUUUUUCAUAUUCGCAAAGAC	
HSP70-P1	CAGAAAAAUUUGCUACAUUGUUUCACAAACUCAAAUUAU	SEQ ID NO: 15
	UCAUUUUUUU	

In some embodiments, a self-replicating RNA of the disclosure comprises a Kozak sequence. As is understood in the art, a Kozak sequence is a short consensus sequence centered around the translational initiation site of eukaryotic mRNAs that allows for efficient initiation of translation of the mRNA. See, for example, Kozak, Marilyn (1988) Mol. and Cell Biol., 8:2737-2744; Kozak, Marilyn (1991) J. Biol. Chem., 266: 19867-19870; Kozak, Marilyn (1990) Proc Natl. Acad. Sci. USA, 87:8301-8305; and Kozak, Marilyn (1989) J. Cell Biol., 108:229-241. It ensures that a protein is correctly translated from the genetic message, mediating ribosome assembly and translation initiation. The ribosomal translation machinery recognizes the AUG initiation codon in the context of the Kozak sequence. A Kozak sequence may be inserted upstream of the coding sequence for the protein of interest, downstream of a 5' UTR or inserted upstream of the coding sequence for the protein of interest and downstream of a 5' UTR. In some embodiments, a self-replicating RNA described herein comprises a Kozak sequence having the amino acid sequence GCCACC (SEQ ID NO: 23). Preferably a self-replicating RNA described herein comprises a partial Kozak sequence "p" having the amino acid sequence GCCA (SEQ ID NO: 24).

Transgenes

Transgenes included in nucleic acid molecules provided herein can encode an antigenic protein or a fragment thereof. In some embodiments, second polynucleotides of nucleic acid molecules provided herein comprise a first transgene. A first transgene included in second polynucleotides of nucleic acid molecules provided herein can encode a first antigenic protein or a fragment thereof. A transgene included in second polynucleotides of nucleic acid molecules provided herein can comprise a sequence encoding the full amino acid sequence of an antigenic protein or a sequence encoding any suitable portion or fragment of the full amino acid sequence of an antigenic protein. Any antigenic protein can be encoded by transgenes included in nucleic acid molecules provided herein. In one aspect, the antigenic protein is a viral protein, a bacterial protein, a fungal protein, a protozoan protein, a parasite protein, or a tumor protein or tumor antigen. Transgenes included in nucleic acid molecules provided herein can be expressed from a subgenomic RNA.

In another embodiment, the antigenic protein, when administered to a mammalian subject, raises an immune response to a pathogen, optionally wherein the pathogen is bacterial, viral, fungal, protozoan, or cancerous. In some more particular embodiments, the antigenic protein is expressed on the outer surface of the pathogen; while in other more particular embodiments, the antigen may be a non-surface antigen, e.g., useful as a T-cell epitope. The

immunogen may elicit an immune response against a pathogen (e.g. a bacterium, a virus, a fungus or a parasite) but, in some other embodiments, it elicits an immune response against an allergen or a tumor antigen. The immune response may comprise an antibody response (usually including IgG) and/or a cell mediated immune response. The polypeptide immunogen will typically elicit an immune response that recognizes the corresponding pathogen (or allergen or tumor) polypeptide, but in some embodiments, the polypeptide may act as a mimotope to elicit an immune response that recognizes a saccharide. The immunogen will typically be a surface polypeptide e.g. an adhesin, a hemagglutinin, an envelope glycoprotein, a spike glycoprotein, etc.

Any viral, bacterial, fungal, protozoan, parasite, or tumor protein can be encoded by transgenes included in nucleic acid molecules provided herein. A protein from any infectious agent can be encoded by transgenes included in nucleic acid molecules provided herein. As used herein, the term "infectious agent" refers to any agent capable of infecting an organism, including humans and animals, and causing disease or deterioration in health. The terms "infectious agent" and "infectious pathogen" may be used interchangeably, unless context clearly indicates otherwise.

In some aspects, the viral protein encoded by transgenes included in nucleic acid molecules provided herein is an orthomyxovirus protein, a paramyxovirus protein, a picornavirus protein, a flavivirus protein, a filovirus protein, a rhabdovirus protein, a togavirus protein, an arterivirus protein, a bunyavirus protein, an arenavirus protein, a reovirus protein, a bornavirus protein, a retrovirus protein, an adenovirus protein, a herpesvirus protein, a polyomavirus protein, a papillomavirus protein, a poxvirus protein, or a hepadnavirus protein. In other aspects, the antigenic protein is an influenza virus protein, a respiratory syncytial virus (RSV) protein, a human immunodeficiency virus (HIV) protein, a hepatitis C virus (HCV) protein, a cytomegalovirus (CMV) protein, a Lassa Fever Virus (LFV) protein, an Ebola Virus (EBOV) protein, a *Mycobacterium* protein, a *Bacillus* protein, a *Yersinia* protein, a *Streptococcus* protein, a *Pseudomonas* protein, a *Shigella* protein, a *Campylobacter* protein, a *Salmonella* protein, a *Plasmodium* protein, or a *Toxoplasma* protein.

In one aspect, the antigenic protein is from a prokaryotic organism, including gram positive bacteria, gram negative bacteria, or other bacteria, such as *Bacillus* (e.g., *Bacillus anthracis*), *Mycobacterium* (e.g., *Mycobacterium tuberculosis*, *Mycobacterium leprae*), *Shigella* (e.g., *Shigella sonnei*, *Shigella dysenteriae*, *Shigella flexneri*), *Helicobacter* (e.g., *Helicobacter pylori*), *Salmonella* (e.g., *Salmonella enterica*, *Salmonella typhi*, *Salmonella typhimurium*), *Neis-*

seria (e.g., *Neisseria gonorrhoeae*, *Neisseria meningitidis*), *Moraxella* (e.g., *Moraxella catarrhalis*), *Haemophilus* (e.g., *Haemophilus influenzae*), *Klebsiella* (e.g., *Klebsiella pneumoniae*), *Legionella* (e.g., *Legionella pneumophila*), *Pseudomonas* (e.g., *Pseudomonas aeruginosa*), *Acinetobacter* (e.g., *Acinetobacter baumannii*), *Listeria* (e.g., *Listeria monocytogenes*), *Staphylococcus* (e.g., *Staphylococcus aureus*), *Streptococcus* (e.g., *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae*), *Corynebacterium* (e.g., *Corynebacterium diphtheriae*), *Clostridium* (e.g., *Clostridium botulinum*, *Clostridium tetani*, *Clostridium difficile*), *Chlamydia* (e.g., *Chlamydia pneumonia*, *Chlamydia trachomatis*), *Caplylobacter* (e.g., *Caplylobacter jejuni*), *Bordetella* (e.g., *Bordetella pertussis*), *Enterococcus* (e.g., *Enterococcus faecalis*, *Enterococcus faecum*), *Vibrio* (e.g., *Vibrio cholerae*), *Yersinia* (e.g., *Yersinia pestis*), *Burkholderia* (e.g., *Burkholderia cepacia* complex), *Coxiella* (e.g., *Coxiella burnetti*), *Francisella* (e.g., *Francisella tularensis*), and *Escherichia* (e.g., enterotoxigenic, enterohemorrhagic or Shiga toxin producing *E. coli*, such as ETEC, EHEC, EPEC, EIEC, and EAEC). In another aspect, the antigenic protein is from a eukaryotic organism, including protists and fungi, such as *Plasmodium* (e.g., *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, *Plasmodium* diarrhea), *Candida* (e.g., *Candida albicans*), *Aspergillus* (e.g., *Aspergillus fumigatus*), *Cryptococcus* (e.g., *Cryptococcus neoformans*), *Histoplasma* (e.g., *Histoplasma capsulatum*), *Pneumocystis* (e.g., *Pneumocystis jirovecii*), and *Coccidioides* (e.g., *Coccidioides immitis*).

In one aspect, the antigenic protein encoded by first transgenes of second polynucleotides included in nucleic acid molecules provided herein is an influenza virus protein or a fragment thereof. In another aspect, the second polynucleotide includes one or more transgenes encoding one or more influenza virus proteins or fragments thereof. Exemplary influenza virus proteins that can be encoded by transgenes of second polynucleotides included in nucleic acid molecules provided herein include proteins from any human or animal virus, including influenza A virus, influenza B virus, influenza C virus, influenza D virus, or any combination thereof. Exemplary influenza proteins include hemagglutinin (HA), neuraminidase (NA), M2, M1, NP, NS1, NS2, PA, PB1, PB2, and PB1-F2. Hemagglutinin proteins from any influenza virus subtype, such as H1-H18 and any emerging hemagglutinin, and neuraminidase proteins from any influenza virus subtype, such as N1-N11 and any emerging neuraminidase, can be antigenic proteins encoded by transgenes included in second polynucleotides of nucleic acid molecules provided herein. Any suitable fragment of influenza virus proteins can be encoded by transgenes included in second polynucleotides of nucleic acid molecules provided herein, including, for example, one or more helper T lymphocyte (HTL) epitope, one or more cytotoxic T lymphocyte (CTL) epitope, or any combination thereof.

Transgenes included in second polynucleotides of nucleic acid molecules provided herein can express tumor proteins or tumor antigens. Tumor proteins or tumor antigens can be from any tumor, including solid and liquid tumors, for example. As used herein, the terms “tumor protein” or “tumor antigen,” which may be used interchangeably unless context clearly indicates otherwise, refer to any protein antigen that is present on the surface of a tumor cell or expressed in a tumor cell. Tumor proteins or tumor antigens include tumor-specific antigens that are present on tumor cells but generally not on other cells and tumor-associated antigens that are generally present on tumor cells and on

normal cells. Tumor proteins or tumor antigens also include neoantigens. As used herein, the term “neoantigen” refers to tumor-specific mutations that can be unique to a patient’s cancer or tumor.

As used herein, the term “tumor” refers to a mass or lump of tissue that is formed by an accumulation of abnormal cells. A tumor can be benign (i.e., not cancer), malignant (i.e., cancer), or premalignant (i.e., precancerous). The terms “tumor” and “neoplasm” can be used interchangeably. Generally, a cancerous tumor is malignant. As used herein, the term “solid tumor” refers to an abnormal mass of tissue that usually does not contain cysts or liquid areas. Exemplary solid tumors include sarcomas and carcinomas. As used herein, the term “liquid tumors” refers to tumors or cancers present in body fluids such as blood and bone marrow. Exemplary liquid tumors include hematopoietic tumors, such as leukemias and lymphomas, notwithstanding the ability of lymphomas to grow as solid tumors by growing in a lymph node, for example. The term “liquid tumor” can be used interchangeably with the term “blood cancer,” unless context clearly indicates otherwise.

Exemplary tumor proteins or tumor antigens include products of mutated oncogenes, products of mutated tumor suppressor genes, products of mutated genes other than oncogenes or tumor suppressors, tumor antigens produced by oncogenic viruses, altered cell surface glycoproteins, oncofetal antigens, and others. Tumor proteins or tumor antigens also include immune regulatory molecules, such as immune checkpoint inhibitors and immune stimulatory molecules. Tumor antigens further include altered cell surface glycolipids. In some aspects, the tumor protein or tumor antigen encoded by transgenes included in second polynucleotides of nucleic acid molecules provided herein is a kidney cancer, renal cancer, urinary bladder cancer, prostate cancer, uterine cancer, breast cancer, cervical cancer, ovarian cancer, lung cancer, liver cancer, stomach cancer, colon cancer, rectal cancer, oral cavity cancer, pharynx cancer, pancreatic cancer, thyroid cancer, melanoma, skin cancer, head and neck cancer, brain cancer, hematopoietic cancer, leukemia, lymphoma, bone cancer, or sarcoma protein. Exemplary tumor proteins or tumor antigens include KRAS, NRAS, HRAS, HER2, BRCA1, BRCA2, carcinoembryonic antigen (CEA), MUC1, guanylyl-cyclase C, NY-ESO-1, melanoma-associated antigen (e.g., MAGE-1, MAGE-3), p53, survivin, alphafetoprotein (AFP), CA-125, epithelial tumor antigen (ETA), tyrosinase, prostate-specific antigen (PSA), prostate-specific membrane antigen (PSMA), prostate stem cell antigen (PSCA), human aspartyl (asparaginyl) β -hydroxylase (HAAH), EphA2, and others.

In some aspects, the tumor protein or tumor antigen encoded by transgenes included in nucleic acid molecules provided herein is a wild-type protein or a fragment or epitope thereof. In other aspects, the tumor protein or tumor antigen encoded by transgenes included in second polynucleotides of nucleic acid molecules provided herein is a mutant protein or a fragment or epitope thereof. In one aspect, the tumor protein or tumor antigen is KRAS, NRAS, HRAS, HER2, BRCA1, BRCA2, carcinoembryonic antigen (CEA), MUC1, guanylyl-cyclase C, NY-ESO-1, melanoma-associated antigen (e.g., MAGE-1, MAGE-3), p53, survivin, alphafetoprotein (AFP), CA-125, epithelial tumor antigen (ETA), tyrosinase, prostate-specific antigen (PSA), prostate-specific membrane antigen (PSMA), prostate stem cell antigen (PSCA), human aspartyl (asparaginyl) β -hydroxylase (HAAH), EphA2, or any mutant thereof. In another aspect, the tumor protein or tumor antigen is KRAS or a fragment or epitope thereof. In another aspect, the tumor protein or

33

tumor antigen is KRASG12D, KRASG12C, KRASG12V, or KRASG13D. Any KRAS that includes any mutation can be encoded by transgenes included in second polynucleotides of nucleic acid molecules provided herein.

In some aspects, transgenes included in second polynucleotides of nucleic acid molecules provided herein encode a reporter or a marker, including selectable markers. Reporters and markers can include fluorescent proteins, such as green fluorescent protein (GFP), red fluorescent protein (RFP), yellow fluorescent protein (YFP), luciferase enzymes, such as firefly and *Renilla* luciferases, and antibiotic selection markers, for example.

In some aspects, the second polynucleotide of nucleic acid molecules provided herein comprises at least two transgenes. Any number of transgenes can be included in second polynucleotides of nucleic acid molecules provided herein, such as one, two, three, four, five, six, seven, eight, nine, ten, or more transgenes. In one aspect, the second polynucleotide of nucleic acid molecules provided herein includes a second transgene encoding a second antigenic protein or a fragment thereof or an immunomodulatory protein. In one aspect, the second polynucleotide further comprises an internal ribosomal entry site (IRES), a sequence encoding a 2A peptide, or a combination thereof, located between transgenes. As used herein, the term "2A peptide" refers to a small (generally 18-22 amino acids) sequence that allows for efficient, stoichiometric production of discrete protein products within a single reading frame through a ribosomal skipping event within the 2A peptide sequence. As used herein, the term "internal ribosomal entry site" or "IRES" refers to a nucleotide sequence that allows for the initiation of protein translation of a messenger RNA (mRNA) sequence in the absence of an AUG start codon or without using an AUG start codon. An IRES can be found anywhere in an mRNA sequence, such as at or near the beginning, at or near the middle, or at or near the end of the mRNA sequence, for example.

Any number of transgenes included in second polynucleotides of nucleic acid molecules provided herein can be expressed via any combination of 2A peptide and IRES sequences. For example, a second transgene located 3' of a first transgene can be expressed via a 2A peptide sequence or via an IRES sequence. As another example, a second transgene located 3' of a first transgene and a third transgene located 3' of the second transgene can be expressed via 2A peptide sequences located between the first and second transgenes and the second and third transgenes, via an IRES sequence located between the first and second transgenes and the second and third transgenes, via a 2A peptide sequence located between the first and second transgenes and an IRES located between the second and third transgenes, or via an IRES sequence located between the first and second transgenes and a 2A peptide sequence located between the second and third transgenes. Similar configurations and combinations of 2A peptide and IRES sequences located between transgenes are contemplated for any number of transgenes included in second polynucleotides of nucleic acid molecules provided herein. In addition to expression via 2A peptide and IRES sequences, two or more transgenes included in nucleic acid molecules provided herein can also be expressed from separate subgenomic RNAs.

A second, third, fourth, fifth, sixth, seventh, eighth, ninth, tenth, etc., transgene included in second polynucleotides of nucleic acid molecules provided herein can encode an immunomodulatory protein or a functional fragment or functional variant thereof. Any immunomodulatory protein

34

or a functional fragment or functional variant thereof can be encoded by a transgene included in second polynucleotides.

As used herein, the terms "functional variant" or "functional fragment" refer to a molecule, including a nucleic acid or protein, for example, that comprises a nucleotide and/or amino acid sequence that is altered by one or more nucleotides and/or amino acids compared to the nucleotide and/or amino acid sequences of the parent or reference molecule. For a protein, a functional variant is still able to function in a manner that is similar to the parent molecule. In other words, the modifications in the amino acid and/or nucleotide sequence of the parent molecule do not significantly affect or alter the functional characteristics of the molecule encoded by the nucleotide sequence or containing the amino acid sequence. The functional variant may have conservative sequence modifications including nucleotide and amino acid substitutions, additions and deletions. These modifications can be introduced by standard techniques known in the art, such as site-directed mutagenesis and random PCR-mediated mutagenesis. Functional variants can also include, but are not limited to, derivatives that are substantially similar in primary structural sequence, but which contain, e.g., in vitro or in vivo modifications, chemical and/or biochemical, that are not found in the parent molecule. Such modifications include, inter alia, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI-anchor formation, hydroxylation, iodination, methylation, myristylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA-mediated addition of amino acids to proteins such as arginylation, ubiquitination, and the like.

In one aspect, a second transgene included in second polynucleotides of nucleic acid molecules provided herein encodes a cytokine, a chemokine, or an interleukin. Exemplary cytokines include interferons, TNF- α , TGF- β , G-CSF, and GM-CSF. Exemplary chemokines include CCL3, CCL26, and CXCL7. Exemplary interleukins include IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12, IL-15, IL-18, IL-21, and IL-23. Any transgene or combination of transgenes encoding any cytokine, chemokine, interleukin, or combinations thereof, can be included in second polynucleotides of nucleic acid molecules provided herein.

In one aspect, first and second transgenes included in second polynucleotides of nucleic acid molecules provided herein encode viral proteins, bacterial proteins, fungal proteins, protozoan proteins, parasite proteins, tumor proteins, immunomodulatory proteins, or any combination thereof. In yet another aspect, first, second, third, fourth, fifth, sixth, seventh, eighth, ninth, tenth, or more transgenes included in second polynucleotides of nucleic acid molecules provided herein encode viral proteins, bacterial proteins, fungal proteins, protozoan proteins, parasite proteins, tumor proteins, immunomodulatory proteins, or any combination thereof. DNA and RNA Molecules

Nucleic acid molecules provided herein can be DNA molecules or RNA molecules. It will be appreciated that T present in DNA is substituted with U in RNA, and vice versa. In one aspect, nucleic acid molecules provided herein are DNA molecules. In another aspect, DNA molecules provided herein further comprise a promoter. As used herein,

the term “promoter” refers to a regulatory sequence that initiates transcription. A promoter can be operably linked to first and second polynucleotides of nucleic acid molecules provided herein. Generally, promoters included in DNA molecules provided herein include promoters for in vitro transcription (IVT). Any suitable promoter for in vitro transcription can be included in DNA molecules provided herein, such as a T7 promoter, a T3 promoter, an SP6 promoter, and others. In one aspect, DNA molecules provided herein comprise a T7 promoter. In another aspect, the promoter is located 5' of the 5' UTR included in DNA molecules provided herein. In yet another aspect, the promoter is a T7 promoter located 5' of the 5' UTR included in DNA molecules provided herein. In yet another aspect, the promoter overlaps with the 5' UTR. A promoter and a 5' UTR can overlap by about one nucleotide, about two nucleotides, about three nucleotides, about four nucleotides, about five nucleotides, about six nucleotides, about seven nucleotides, about eight nucleotides, about nine nucleotides, about ten nucleotides, about 11 nucleotides, about 12 nucleotides, about 13 nucleotides, about 14 nucleotides, about 15 nucleotides, about 16 nucleotides, about 17 nucleotides, about 18 nucleotides, about 19 nucleotides, about 20 nucleotides, about 21 nucleotides, about 22 nucleotides, about 23 nucleotides, about 24 nucleotides, about 25 nucleotides, about 26 nucleotides, about 27 nucleotides, about 28 nucleotides, about 29 nucleotides, about 30 nucleotides, about 31 nucleotides, about 32 nucleotides, about 33 nucleotides, about 34 nucleotides, about 35 nucleotides, about 36 nucleotides, about 37 nucleotides, about 38 nucleotides, about 39 nucleotides, about 40 nucleotides, about 41 nucleotides, about 42 nucleotides, about 43 nucleotides, about 44 nucleotides, about 45 nucleotides, about 46 nucleotides, about 47 nucleotides, about 48 nucleotides, about 49 nucleotides, about 50 nucleotides, or more nucleotides.

In some aspects, DNA molecules provided herein include a promoter for in vivo transcription. Generally, the promoter for in vivo transcription is an RNA polymerase II (RNA pol II) promoter. Any RNA pol II promoter can be included in DNA molecules provided herein, including constitutive promoters, inducible promoters, and tissue-specific promoters. Exemplary constitutive promoters include a cytomegalovirus (CMV) promoter, an EF1 α promoter, an SV40 promoter, a PGK1 promoter, a Ubc promoter, a human beta actin promoter, a CAG promoter, and others. Any tissue-specific promoter can be included in DNA molecules provided herein. In one aspect, the RNA pol II promoter is a muscle-specific promoter, skin-specific promoter, subcutaneous tissue-specific promoter, liver-specific promoter, spleen-specific promoter, lymph node-specific promoter, or a promoter with any other tissue specificity. DNA molecules provided herein can also include an enhancer. Any enhancer that increases transcription can be included in DNA molecules provided herein.

In some aspects, nucleic acid molecules provided herein are RNA molecules. An RNA molecule provided herein can be generated by in vitro transcription (IVT) of DNA molecules provided herein. In one aspect, RNA molecules provided herein are self-replicating RNA molecules. In another aspect, RNA molecules provided herein further comprise a 5' cap. Any 5' cap can be included in RNA molecules provided herein, including 5' caps having a Cap 1 structure, a Cap 1 (m6A) structure, a Cap 2 structure, a Cap 0 structure, or any combination thereof. In one aspect, RNA molecules provided herein include a 5' cap having Cap 1 structure. In yet another aspect, RNA molecules provided herein are self-replicating RNA molecules comprising a 5'

cap having a Cap 1 structure. In a further aspect, RNA molecules provided herein comprise a cap having a Cap 1 structure, wherein a m7G is linked via a 5'-5' triphosphate to the 5' end of the 5' UTR. In yet a further aspect, RNA molecules provided herein comprise a cap having a Cap 1 structure, wherein a m7G is linked via a 5'-5' triphosphate to the 5' end of the 5' UTR comprising a sequence of SEQ ID NO:73. Any method of capping can be used, including, but not limited to using a Vaccinia Capping enzyme (New England Biolabs, Ipswich, Mass.) and co-transcriptional capping or capping at or shortly after initiation of in vitro transcription (IVT), by for example, including a capping agent as part of an in vitro transcription (IVT) reaction. (Nuc. Acids Symp. (2009) 53:129).

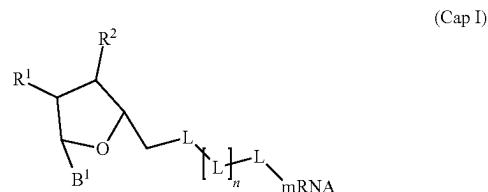
Provided herein, in some embodiments, are nucleic acid molecules comprising (a) a sequence of SEQ ID NO:78; or (b) a sequence of SEQ ID NO:72, SEQ ID NO:73, SEQ ID NO:76, and SEQ ID NO:77, wherein T is substituted with U. In one aspect, nucleic acid molecules provided herein are RNA molecules. In another aspect, RNA molecules provided herein further comprise a 5' cap having a Cap 1 structure. Any RNA molecules provided herein can be self-replicating RNA molecules.

Only those mRNAs that carry the Cap structure are active in Cap dependent translation; “decapitation” of mRNA results in an almost complete loss of their template activity for protein synthesis (Nature, 255:33-37, (1975); J. Biol. Chem., vol. 253:5228-5231, (1978); and Proc. Natl. Acad. Sci. USA, 72:1189-1193, (1975)).

Another element of eukaryotic mRNA is the presence of 2'-O-methyl nucleoside residues at transcript position 1 (Cap 1), and in some cases, at transcript positions 1 and 2 (Cap 2). The 2'-O-methylation of mRNA provides higher efficacy of mRNA translation in vivo (Proc. Natl. Acad. Sci. USA, 77:3952-3956 (1980)) and further improves nuclease stability of the 5'-capped mRNA. The mRNA with Cap 1 (and Cap 2) is a distinctive mark that allows cells to recognize the bona fide mRNA 5' end, and in some instances, to discriminate against transcripts emanating from infectious genetic elements (Nucleic Acid Research 43: 482-492 (2015)).

Some examples of 5' cap structures and methods for preparing mRNAs comprising the same are given in WO2015/051169A2, WO/2015/061491, US 2018/0273576, and U.S. Pat. Nos. 8,093,367, 8,304,529, and 10,487,105. In some embodiments, the 5' cap is m7GpppAmpG, which is known in the art. In some embodiments, the 5' cap is m7GpppG or m7GpppGm, which are known in the art. Structural formulas for embodiments of 5' cap structures are provided below.

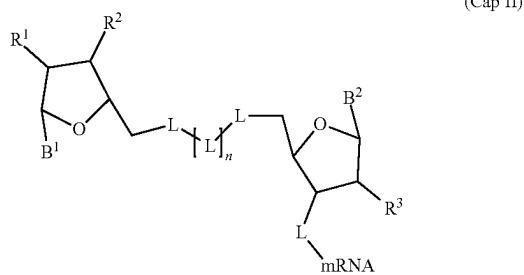
In some embodiments, a self-replicating RNA of the disclosure comprises a 5' cap having the structure of Formula (Cap I).



wherein B¹ is a natural or modified nucleobase; R¹ and R² are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophos-

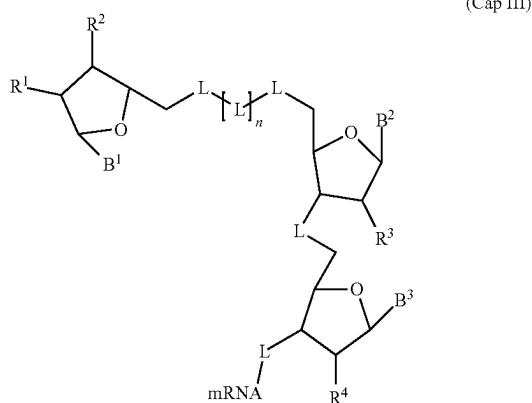
phate wherein each L is linked by diester bonds; n is 0 or 1. and mRNA represents an mRNA of the present disclosure linked at its 5' end. In some embodiments B¹ is G, m⁷G, or A. In some embodiments, n is 0. In some embodiments n is 1. In some embodiments, B¹ is A or m⁶A and R¹ is OCH₃; wherein G is guanine, m⁷G is 7-methylguanine, A is adenine, and m⁶A is N⁶-methyladenine.

In some embodiments, a self-replicating RNA of the disclosure comprises a 5' cap having the structure of Formula (Cap II).



wherein B¹ and B² are each independently a natural or modified nucleobase; R¹, R², and R³ are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments B¹ is G, m⁷G, or A. In some embodiments, n is 0. In some embodiments, n is 1. In some embodiments, B¹ is A or m⁶A and R¹ is OCH₃; wherein G is guanine, m⁷G is 7-methylguanine, A is adenine, and m⁶A is N⁶-methyladenine.

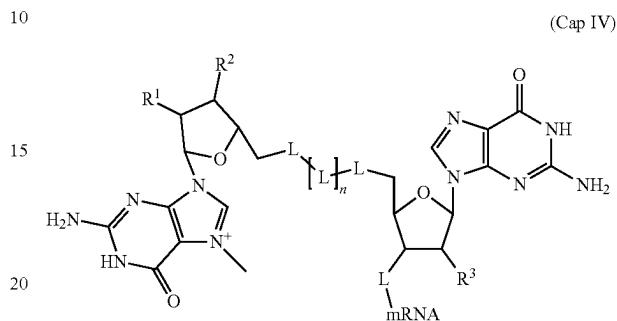
In some embodiments, a self-replicating RNA of the disclosure comprises a 5' cap having the structure of Formula (Cap III).



wherein B¹, B², and B³ are each independently a natural or modified nucleobase; R¹, R², R³, and R⁴ are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments, at least one of R¹, R², R³, and R⁴ is OH.

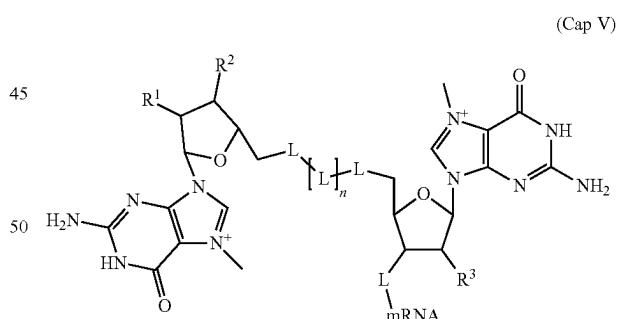
In some embodiments B¹ is G, m⁷G, or A. In some embodiments, B¹ is A or m⁶A and R¹ is OCH₃; wherein G is guanine, m⁷G is 7-methylguanine, A is adenine, and m⁶A is N⁶-methyladenine. In some embodiments, n is 1.

In some embodiments, a self-replicating RNA of the disclosure comprises a m7GpppG 5' cap analog having the structure of Formula (Cap IV).



wherein, R¹, R², and R³ are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; n is 0 or 1. In some embodiments, at least one of R¹, R², and R³ is OH. In some embodiments, the 5' cap is m⁷GpppG wherein R¹, R², and R³ are each OH, n is 1, and each L is a phosphate. In some embodiments, n is 1. In some embodiments, the 5' cap is m⁷GpppGm, wherein R¹ and R² are each OH, R³ is OCH₃, each L is a phosphate, mRNA is the mRNA encoding an enzyme having OTC activity linked at its 5' end, and n is 1.

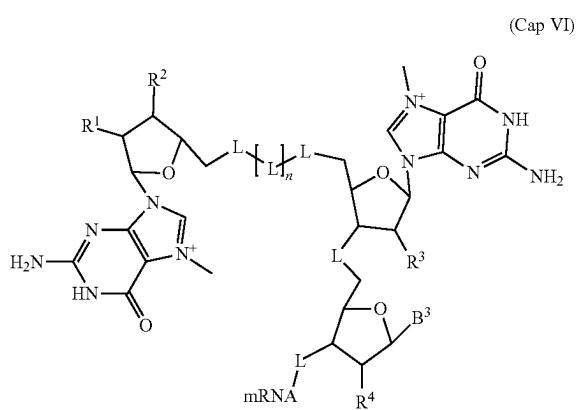
In some embodiments, a self-replicating RNA of the disclosure comprises a m7Gpppm7G 5' cap analog having the structure of Formula (Cap V).



wherein, R¹, R², and R³ are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments, at least one of R¹, R², and R³ is OH. In some embodiments, n is 1.

In some embodiments, a self-replicating RNA of the disclosure comprises a m7Gpppm7GpN, 5' cap analog, wherein N is a natural or modified nucleotide, the 5' cap analog having the structure of Formula (Cap VI).

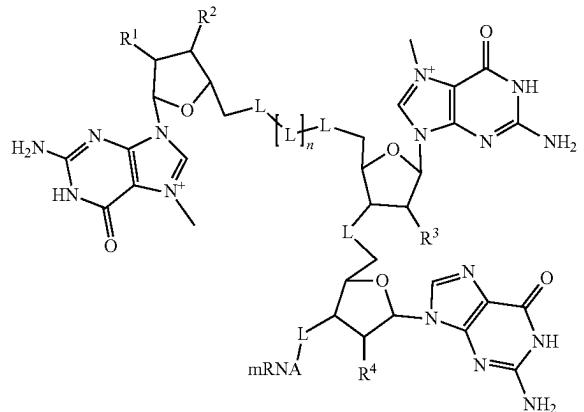
39



wherein B³ is a natural or modified nucleobase; R¹, R², R³, and R⁴ are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 3. In some embodiments, at least one of R¹, R², R³, and R⁴ is OH. In some embodiments B¹ is G, m⁷G, or A. In some embodiments, B¹ is A or m⁶A and R¹ is OCH₃; wherein G is guanine, m⁷G is 7-methylguanine, A is adenine, and m⁶A is N⁶-methyladenine. In some embodiments, n is 1.

In some embodiments, a self-replicating RNA of the disclosure comprises a m7Gpppm7GpG 5' cap analog having the structure of Formula (Cap VII).

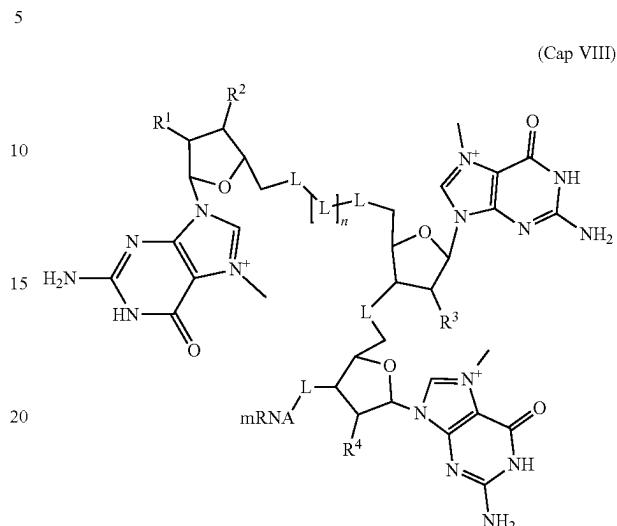
(Cap VII)



wherein, R¹, R², R³, and R⁴ are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments, at least one of R¹, R², R³, and R⁴ is OH. In some embodiments, n is 1.

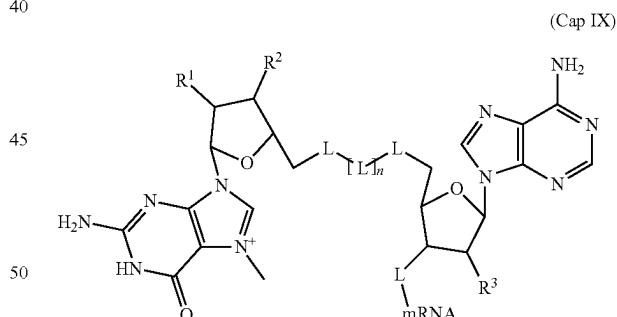
40

In some embodiments, a self-replicating RNA of the disclosure comprises a m7Gpppm7Gpm7G 5' cap analog having the structure of Formula (Cap VIII).



wherein, R¹, R², R³, and R⁴ are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; n is 0 or 1. In some embodiments, at least one of R¹, R², R³, and R⁴ is OH. In some embodiments, n is 1.

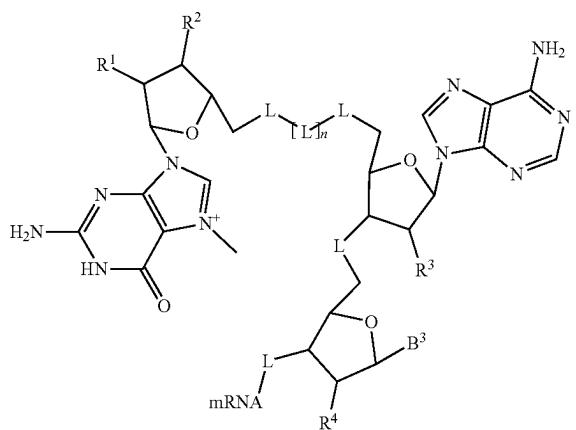
In some embodiments, a self-replicating RNA of the disclosure comprises a m7GpppA 5' cap analog having the structure of Formula (Cap IX).



wherein, R¹, R², and R³ are each independently selected from a halogen, OH, and OCH₃; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments, at least one of R¹, R², and R³ is OH. In some embodiments, n is 1.

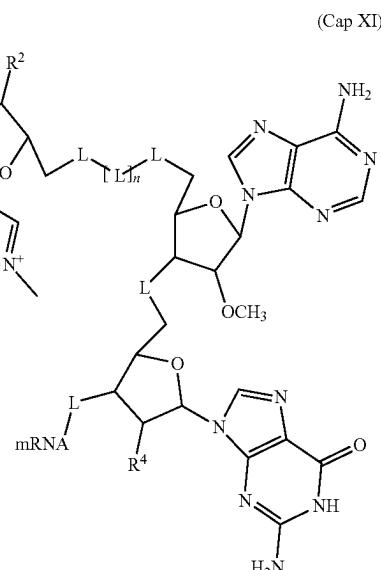
In some embodiments, a self-replicating RNA of the disclosure comprises a m7GpppApN 5' cap analog, wherein N is a natural or modified nucleotide, and the 5' cap has the structure of Formula (Cap X).

41



wherein B^3 is a natural or modified nucleobase; R^1 , R^2 , R^3 , and R^4 are each independently selected from a halogen, OH, and OCH_3 ; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments, at least one of R^1 , R^2 , R^3 , and R^4 is OH. In some embodiments B^3 is G, m^7G , A or m^6A ; wherein G is guanine, m^7G is 7-methylguanine, A is adenine, and m^6A is N^6 -methyladenine. In some embodiments, n is 1.

In some embodiments, a self-replicating RNA of the disclosure comprises a $m^7GpppAmpG$ 5' cap analog having the structure of Formula (Cap XI).

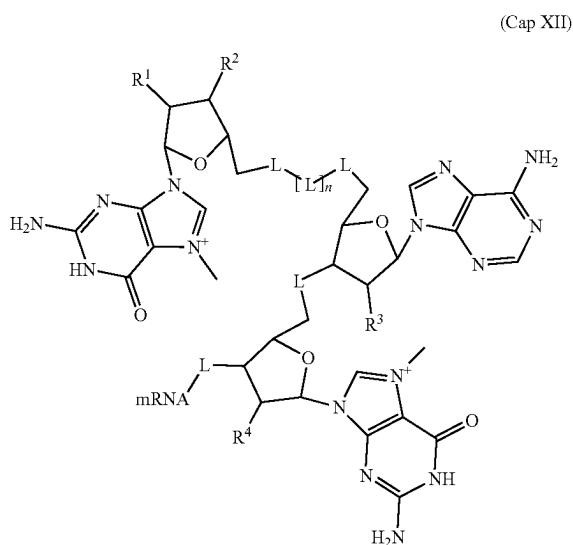


wherein, R^1 , R^2 , and R^4 are each independently selected from a halogen, OH, and OCH_3 ; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments, at least one of R^1 , R^2 , and R^4 is OH. In some embodiments, the compound of Formula Cap XI is

42

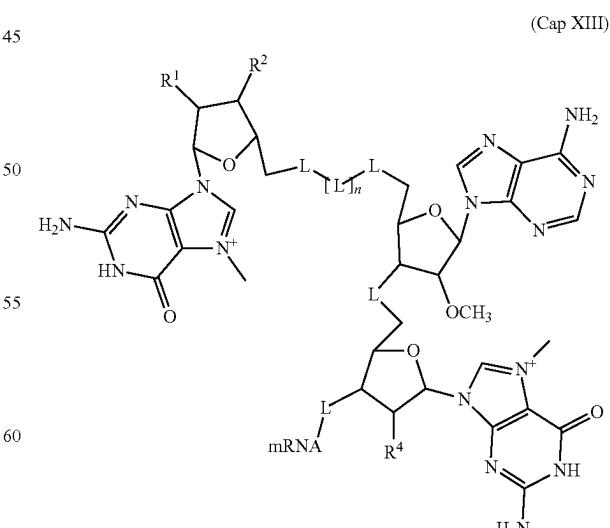
$m^7GpppAmpG$, wherein R^1 , R^2 , and R^4 are each OH, n is 1, and each L is a phosphate linkage. In some embodiments, n is 1.

In some embodiments, a self-replicating RNA of the disclosure comprises a $m^7GpppApm7G$ 5' cap analog having the structure of Formula (Cap XII).



wherein, R^1 , R^2 , R^3 , and R^4 are each independently selected from a halogen, OH, and OCH_3 ; each L is independently selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments, at least one of R^1 , R^2 , R^3 , and R^4 is OH. In some embodiments, n is 1.

In some embodiments, a self-replicating RNA of the disclosure comprises a $m^7GpppApm7G$ 5' cap analog having the structure of Formula (Cap XIII).



wherein, R^1 , R^2 , and R^4 are each independently selected from a halogen, OH, and OCH_3 ; each L is independently

selected from the group consisting of phosphate, phosphorothioate, and boranophosphate wherein each L is linked by diester bonds; mRNA represents an mRNA of the present disclosure linked at its 5' end; and n is 0 or 1. In some embodiments, at least one of R¹, R², and R⁴ is OH. In some embodiments, n is 1.

Poly-Adenine (Poly-A) Tail

Polyadenylation is the addition of a poly(A) tail, a chain of adenine nucleotides usually about 100-120 monomers in length, to a mRNA. In eukaryotes, polyadenylation is part of the process that produces mature mRNA for translation and begins as the transcription of a gene terminates. The 3'-most segment of a newly made pre-mRNA is first cleaved off by a set of proteins; these proteins then synthesize the poly(A) tail at the 3' end. The poly(A) tail is important for the nuclear export, translation, and stability of mRNA. The tail is shortened over time, and, when it is short enough, the mRNA is enzymatically degraded. However, in a few cell types, mRNAs with short poly(A) tails are stored for later activation by re-polyadenylation in the cytosol.

Preferably, a self-replicating RNA of the disclosure comprises a 3' tail region, which can serve to protect the RNA from exonuclease degradation. The tail region may be a 3'poly(A) and/or 3'poly(C) region. Preferably, the tail region

is a 3' poly(A) tail. As used herein a "3' poly(A) tail" is a polymer of sequential adenine nucleotides that can range in size from, for example: 10 to 250 sequential adenine nucleotides; 60-125 sequential adenine nucleotides, 90-125 sequential adenine nucleotides, 95-125 sequential adenine nucleotides, 95-121 sequential adenine nucleotides, 100 to 121 sequential adenine nucleotides, 110-121 sequential adenine nucleotides; 112-121 sequential adenine nucleotides; 114-121 adenine sequential nucleotides; or 115 to 121 sequential adenine nucleotides. Preferably, a 3' poly(A) tail as described herein comprise 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, or 125 sequential adenine nucleotides. 3' Poly(A) tails can be added using a variety of methods known in the art, e.g., using poly(A) polymerase to add tails to synthetic or in vitro transcribed RNA. Other methods include the use of a transcription vector to encode poly(A) tails or the use of a ligase (e.g., via splint ligation using a T4 RNA ligase and/or T4 DNA ligase), wherein poly(A) may be ligated to the 3' end of a sense RNA. In some embodiments, a combination of any of the above methods is utilized.

Design and Synthesis of Self-Replicating RNA

The constructs for exemplary self-replicating RNA sequences of the present disclosure are provided in Table 4.

TABLE 4

Comparison of STARR TM self-replicating RNA of the disclosure with comparative self-replicating RNA as described			
Construct	Position	Sequence Type	Sequence
STARR TM (SEQ ID NO: 49)	5' UTR	nucleotide	ATGGGCGGCATGAGAGAAGCCCAGACCAATTACCT ACCCAAA
STARR TM (SEQ ID NO: 50)	non-structural gene ORF	nucleotide	ATGGAGAAAAGTTCACTGACATCGAGGAAGCACGCC CATTCCTCAGAGCTTGCAGCGAGCTCCCGCAGTT GAGGTAGAGCCAAGCAGGTCACTGATAATGACCATG CTAATGCCAGAGCGTTTCGCATCTGGCTTCAAAGCT ATCGAAACGGAGGTGACCCATCCGACACGATCCTTG ACATTGGAAGTGCGCCGCCGCGCAGAATGTATCTAA GACAAGTATCATTGATCTGCGATGAGATGTGGG AAGATCCGGACAGATTGATAAGTATGCAACTAACGCT GAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTG GACAAGAAAATGAAAGGAGCTGGCCGCGTCATGAGCG ACCCCTGACCTGAAACTGAGACTATGCGCTCAGGA CGACGAGTCGTGCTACGAAGGGCAAGTCGCTGTT TACCAAGGATGATACGCGCTGACGGCCCCACAGCC TGTAACCACAGGCCAACAAAGGGCGTGAAGGGTGGCCTA CTGGATCGGCTTCGACACCAACCCCTCATGTTCAAGA ACCTGGCGGCCCTACCCAGCTACAGCACCAACTG GGCGCAGAGACCGTGTGACCGCCAGGAACATCGGC CTGTGAGCAGCGACCTGATGAGAGAGGGCGGAGAG GCATGAGCATCTGAGGAAGAAATACTGAGCCAG CAACAACGTGTTGAGCTGGCAGCACCATCTAC CACGAGAAAGAGGGACCTGCTCAGGAGCTGGCACCTGC CCAGCGTGTCCACCTGAGGGGCAAGCAGAACATCAC CTGCAGGTGCGAGACCATCGTGAAGCTGCGACGGCTAC GTGGTGAAGAGGATGCCCATCAGCCCCGGCTGTACG GCAAGGCCAGCGCTACGCCGTACAATGACAGGGA GGGCTTCTGCTGCAAGGTGACCCGACACCCCTGAAC GGCGAGAGGGTGAAGCTTCCCGCTGTGACCTACGTGC CGGCCACCCCTGTCGACAGATGACGGCATCTGCG CACCGACGTGAGCGCCGACGACGCCAGAAGCTGCTC GTGGGCTGAACTGAGGATCGTGGTCAACCGCAGGA CCCAGAGGAACCCAACACATGAAAGAAACTACCTGCT GCCCGTGGTGGCCAGGCTTCGCCAGGTGGGCTGTGCT GAGTACAAGGAGGACAGGAAAGCAGAGAGGCCCTG GGCCTGAGGGACAGGAGCTGGTGAAGGGCTGTGCT GGCCCTTCAGGGGGCACAAAGATCACAGCATCTACAA GAGGCCGACACCCAGACCATCTCAAGGTGACAGC GACTTCCACAGCTTCGTGCTGCCAGGATCGGCAGCA ACACCCCTGGAGATCGGCCCTGAGGACCCGGATCAGGAA

TABLE 4 -continued

Comparison of STARRTM self-replicating RNA of the disclosure with comparative self-replicating RNA as described

Construct	Position	Sequence Type	Sequence
			GATGCTGGAGGAACACAAGGAGCCCAGCCCCACTGATC ACCGCCGAGGAGCTGCAGGAGGCCAAGTGCCTGCCG ACGAGGCCAAGGAGCTGAGGGAGGCCAGGAACCTGA GGGCCGCCCTGCCACCCCTGGCTGCCGACGTGGAG ACCCACCCCTGAAAGGCCACGTGGACCTGATGCTGCAG GAGGCCGCCGGCGGAAGCTGGAGAGACACCCAGGGC CTGATCAAGGTGACCAGCTACGACGGCAGGACAAGA TCGGCACGCTACCCCGTCTGAGCCCCACAGGCCGTGCT GAAGTCCGAGAACGCTGAGCTGCATCCACCCACTGCC GAGCAGGTGATCGTGTACCCACAGGGCAGGAAGG GCAGGTACGCCGTGGAGCCCTACCCAGGCCAAGGTGGT CGTGCCGAGGCCACGCCATCCCGTGCAGGACTTC CAGGCCCTGAGCGAGAGGCCACCATGTTACAACG AGAGGGAGTTCTGTAACAGGTACCTGCACCATATGCC CACCCACGGCGGAGCCCTGAAACACCGACGGAAATAC TACAAGACCGTGAAGGCCAGGGAGCACGACGGCAGT ACCTGTACGACATCGACAGGAAGCAGTGCCTGAAGAA AGAGCTGTTGACGGCTGGACTGACGGCAGGCTG GTGGACCCACCTTCCACAGGTTCTGCTACAGAGACCT GAGGACAGAACCGCCGCTCCCTACAGGTGCCACC ATCGGCGTGTACGGCGTGGCCCGAGCGGAAAGAGCG GCATCATCAAGAGCGCCGTGACCAAGAAAGACCTGGT GGTCAGCGCCAAGAAAGAGAAACTGCGCCGAGATCATC AGGGACGTGAAGAGATGAAAGGCCCTGAGCTGAAAC GCCGCGACCGTGGACAGCTGCTGAACCGCTGCA AGCACCCCGTGGAGACCCCTGTACATCGACGAGGCCCT CGCTTGCCACGCCGACCCCTGAGGGCCCTGATCGCC ATCATCAGGCCAAGAAAGCCGTGCTGTGCCGCCACC CCAAGCAGTGGGCTTCTTCAACATGATGTCCTGAAG GTGCACTTCAACCACGAGATCTGCACCCAGGTGTTCCA CAAGAGCATCAGCAGGGGTGACCAAAGAGCGTGA AGCGTCGTGAGCACCCTGTTCTACGACAAGAAAATGA GGACCAACCAACCCCAAGGAGCACAAATCGTATCGA CACACAGGCAGCACCAAGCCCAGCAGGAGCACCTG ATCCTGACCTGCTTCAGGGCTGGTGAAGCAGCTGC AGATCGACTACAAGGGCAACAGAGATCATGACCGCCGC TGCCAGCCAGGGCTGACCGAGGAAGGGCTGTACGCC GTGAGGTACAAGGTGAACGAGAACCCACTGTACGCC CCACCGAGCGCACGTGACACTGCTGACCGAG CGAGGACAGGATCTGTGAGAACCCCTGCCGCCGAC CCCTGGATCAACCCCTGACCCCAAGTACCCCGCA ACTTCACGCCACCATCGAAGAGTGGCAGGCCGAGCA CGACGCCATCATGAGGACACATCTGGAGAGGCCGAC CCCACCGACGTGTTCCAGAAACAAGGCCAACGTGTGCT GGGCAAGGCCCTGGTCCCGCTGCTGAAGACGCCGG CATCGACATGACCACAGAGCAGTGGAACACCCGTGGAC TACTTCAGAGACCGAACAGGCCACAGGCCAGATCG TGCTGAACCGAGCTGTGCGTGAGGTTCTCGGCTGGAC CTGGACAGCGGCCCTGTTCAAGGCCCTGCCACT GAGCATCAGGAACAACCAACTGGGACAACAGCCCCAGC CCAAACATGTACGGCTGAAACAGGGAGGTGTCAGGC AGCTGAGCAGGGTACCCACAGCTGCCAGGGCGT GGCCACCCGGCAGGGTGTACGACATGAACACCCGGCACC CTGAGGAACATCGACCCAGGATCAACCTGGTGGCCCG TGAACAGGGCTGCCACGCCCTGGTGTGACCA CAACGAGCACCCACAGAGCGACTTCAGCTCTTGTG AGCAAGCTGAAAGGCAGGACCGTGTGGCTGG AGAAGCTGAGCGTGCCTGGCAAGATGGTGGACTGGCT GAGCGACAGGCCAGGGCACCTCCGGCAGGGT GACCTCGGCATCCCCGGCAGCGTGCCAACGTACGACA TCATCTTGTGAACTGTCAGGACCCATACAGTACAC CATTACCACTGAGTCAGGAGGACACGCCATCAAGCTGA GCATGCTGACCAAGAACGGCTGCCCTGCCACCTGAACCC CGGAGGGACCTGCGTGTGAGCATCGGCTACGGCTACGCC GACAGGGCAGCGAGAGCATATTGGGCCATCGCCA GGCTGTCAAGTTCAAGGAGGTGTGAAACCCAAGAG CAGCCCTGGAGGAACCGAGGTGTGTTCTGTTAC GGCTACGACCCGGAAAGGCCAGGACCCACAACCCCTACA AGCTGAGCAGCACCCCTGACAAACATCTACACCCGGCAG CAGGCTGACCGAGGCCGGCTGCCGCCAGTAC GTGCTGAGGGCGATACTGCCACGCCACCGAGGGC TGATCATCAACGCTGCCAACAGCAAGGGCCAGGCC AGGCGGAGTGTGCGGCCCTGTAAGAAGTCCCC GAGAGCTTCGACCTGCAAGCCCATCGAGGTGGCAGG

TABLE 4 -continued

Comparison of STARRTM self-replicating RNA of the disclosure with
comparative self-replicating RNA as described

Construct	Position	Sequence Type	Sequence
			CCAGGCTGGTGAAGGGCGCCGTAAGCACACATCATCCA CGCCGTGGGCCCAACTTCACAAGGTGAGGGAGGTG GAAGGGACAACGAGCTGGCGAAGCTACAGAGGC ATGCCAAAGATCGTGAACGACAATAACTACAAGAGCG TGGCCATCCCCTGCTCAGCACCGGCATCTTCAGCGGC AACAAAGGACAGGCTGACCCAGAGCCTAACCCACCTGC TCACCGCCCTGGACACCACCGATGCCGACGTGGCCAT CTACTGAGGGACAAGAAGTGGGAGATGACCTGAAG GAGGCCGTGGCCAGGGGGAGGGCGTGGAAAGAGATCT GCATCAGGGACGACTCCAGCGTGACCCAGGCCAGCC CGAGCTGGTGAGGGTGCACCCCCAAGAGCTCCCTGGCC GGCAGGAAGGGTACAGCACAGCAGCGAACAGACCT TCAGCTACCTGGAGGGCACCAAGTTCCACCAAGGCCGC TAAGGACATCGCCGAGATCAACGCTATGTGCCCGTGC GCCACCGAGGCCAACAGCAGCAGGTGTCATGTACATCC TGGGCAGAGGAGATGTCCAGCATCAGGAGCAAGTGC CGTGGAGGAAAGCAGGCCAGCACACCACCCAGCACC CTGCCCCCTGCCCTGTCATCCACGGCTATGACACCCGAGAG GGTGCAGGGCTGAAGGGCCACAGGCCGAGCAGATC ACCGTGTGAGCTCTCCCACTGCCAAGTACAGGAT CACCGGGCTGCAAGATTCAGTGCAGCCAGGCCATC CTGTTCAGCCAAAGGTGCCGCCATCATCCACCCAG GAAGTACCTGGTGGAGACCCCAACCCGTGGACGAGACA CCCGAGGCCAGGCCAGAACAGCAGCACCCAGGGC ACACCCGAGCAGCACCCCTGATCACCGAGGAGCAGA CAAGGACCCGGACCCAGAGGCCATATTATCGAGGA AGAGGAAGAGGACAGCATCACCTGCTGAGGGACGCC CCCCACCCACAGGTGTCAGGTGGAGGCCACATCC ACGGCCACCCAGCGTGTCCAGCTCCAGCTGGAGCAT CCCCACGCCAGCGACTTCGACGTGGACAGCCTGAGC ATCCTGGACACCCCTGGAGGGGCCACCGTGAACCTCCG GCGCCACCGGCCGGAGACCAACAGCTACTTCGCAA GAGCATGGAGTTCTGTCAGGCCAGGCTGCGAGCTCCC AGGACCGTGTTCAGGAACCCACCCACCCAGCTCCCA GGACCAAGGACCCCAAGCCTGCTCCAGCAGGCCCTG CAGCAGGACAGCTGGTGAGCACCCCAACCCGGCGTGA AACAGGGTGTACCCAGGGAGGAACCTGGAGGCCCTGA CACCCAGCAGGACCCCAAGCAGGTCCGTGAGCAGGAC TAGCTGGTGTCAAACCCACCCGGCTGAACAGGGTGA ATCACCAAGGAGGAATTGAGGCTTCTGCCCCAGC AACAGAGACGGTTGAGCCGGCGCTACATCTTCAG CAGCGACACCCGGCAGGGACACCTGCAAGCAAAGAGC GTGAGGGAGACAGCTGAGGCTGAGGGAGGTGCTGGAGA GGACCGAGCTGGAAATCAGCTACGCCACCCAGGCTGGA CCAGGAGAAGGAGGAACCTGCTCAGGAAGAAACTGCA GCTGAACCCCACCCAGCCAAACAGGAGCAGGTACCG AGCAGGAAGGTGGAGAACATGAAGGGCCTACCGCCA GGCGGATCTGCAAGGCTGGGACACTACCTGAAAGGC CGAGGGCAAGGTGGAGTGTCAAGGACCCCTGCCACCC GTGCCACTGTACAGCTCAGCGTGAACAGGGCTTCTC CAGCCCCAAGGTGGCCCTGGAGGCCCTGCAACGCTATG CTGAAGGGAGAACTTCCCCACCGTGGCAGCTACTGCA TCATCCCCAGGTACGACGCCATCTGGACATGGTGG CGGCGCAGCTGCTGCCCTGGACACGCCAGCTTCTGCC CCGGCAAGCTGAGGAGCTTCCCCAAGAAACACAGCTA CTTGGAGCCCCACCATCAGGAGGCCCTGCCCCAGCGCC ATCCAGAACACCCCTGCAAGACGTGCTGGCCCTGCCA CCAAGAGGAACCTGCAACGTGACCCAGATGAGGGAGCT GCCCGTGTGACAGGCTGCCCTAACCTGGAGGTGCT TCAAGAAAATACGCCATGCAACAAACAGGACTTGGGAGAC CTTCAAGGAGAACCCCATCAGGCTGACCGAGAGAAC GTGGTGAACATACATCAGGCTGACGGGCCAAGG CCGCTGCCCTGTTGCTGAAGACCCACAACCTGAAACATG CTGCAGGACATCCCAATGGACAGGTGCTGATGGACC TGAAGAGGGACCTGAGGCTGACACCCGGCACCAAGCA CACCGAGGAGGGCCAAGGTGAGGTGATCCAGGCC GCTGACCCACTGCCACCGCTACCTGTCAGGCCATCCA CAGGGAGCTGGTGAAGGGCGGCTGAACGCCGTGCTG CCCAACATCCACACCCCTGTTGACATGAGGCCGAGG ACTTCGACGCCATCATGCCAGCACTTCCAGGCC GACTGCCGTGAGGAGACGCCGACATGCCAGCTTCGACA AGAGCGAGGATGACGCTATGCCCTGACCGCTCTGAT GATCCCTGGAGGACCTGGCGTGGACGCCGAGCTGCTC ACCCCTGATCGAGGCTGCCCTCGCGAGATCAGCTCCAT

TABLE 4 -continued

Comparison of STARRTM self-replicating RNA of the disclosure with comparative self-replicating RNA as described

Construct	Position	Sequence Type	Sequence
STARR TM (SEQ ID NO: 51)	non- structural gene ORF	amino acid	CCACCTGCCACCAAGACCAAGTTCAAGTCGGCGCT ATGATGAAAAGCGGAATGTTCTGACCCCTGTCGTGA ACACCGTGATCACATCTGTGATCGCCAGCAGGGTGC GCGGGAGGGCTGACCGCGACGCCCTGCGCTGCCCTC ATCGGCAGCAGAACATCGTAAGGGCGTAAAAGCG ACAAGCTGATGGCCGACAGGTGCGCACCTGGCTGAA CATGGAGGTGAAGATCATCGACGCCGTGGGGCGAG AAGGCCCTACTTCTCGGGCGATTATCCTGTGCGA CAGCGTACCGGCACCGCTGCAGGGTGGCCGACCCC CTGAAGAGGGCTTCAAGCTGGCAAGCCACTGGCCG CTGACGATGAGCACGAGATGACAGGGGGAGGGCCCT GACAGGAAAGCACCAAGGTGAACAGGGTGGGCAT CCTGAGCGAGCTGTGCAAGGGCGTGGAGAGCAGGTAC GAGACCGTGGCACCGCATCATCGTATGCTATGA CCACACTGGCAGCTCGTCAAGAGCTTCTCTACACTG AGGGGGCCCTATAACTCTACGGCTAA
			MEKVHVDIEEDSPFLRLAQRSFPQFEVEAKQVTDNDHAN ARAFSHLASLKIETEVDPDSITLDIGSAPARRMYSKHKYH C1CPMRCAEDPDRLYKYATKLKKNCKEITDKELDKMK ELAAVMSDPDLETEMCLHDDESCRYEGQVAVYQDVY AVDGPTSLYHQANKGVRVAYWIGFDTPFMFKNLAGAY PSYSTNWADETFLTARNIGLCSSDVMERSRRGMSILRK YLKPSNNVLFSVGSTIYHEKRDLLSWHLPSPFHRLRGKQ NYTCRCETIVSCDGYVVKRIAISPGLYGPSPGYAATMHR EGFLCCKVTDLNGERVSFPVCTYVPATLCDQMTGILAT DVSADDAQKLVLGLNQRIVVNNGRTQRTNTMKNYLLPV VAQAFARWAKEYKEDQEDERPLGLRDRQLVMGCCWAF RRHKITSIYKRPTDTQTIIKVNSDFHSFVLPRIGSNTEIGLR TRIRKMLEEHKEPSPLITAEDVQEAKCAADEAKEVREAE ELRAALPPLAADVEEPTEADVDLMLQEAGAGSVETPRG LIKVTSYDGEDKIGSYAVLSPQAVLKSEKLSCIHPPLAEQVI VITHSGRKGRYAVEPYHGKVVPPEGHAIPVQDFQALSES ATIVYNEREVFNRYLHHIATHGGALNTDEEYYKTVKPSE HDGEYLYDIDRKQCVKKELVTGLGLTGELVDPFFHFAY ESLRTRPAAPYQVPTIGVYGVPGSGKGSIIKSAVTKKDLV VSAKKENCAETIRDVKKMKGLDVNARTVDSVLLNGKH PVETLYIDEAFACHAGTLRALIAIIRPKKAVLCGDPQCG FFNMMCLVKVHNHEICTOQVFHKSIISRCTKSUTSVSTLF YDKKMRRTNPKETKVIDTTGSTKPKQDDLIITCPFRGW KQLQIDYKGNEIMTAAASQGLTRKGVYAVRYKVNENPL YAPTSEHVNVLRLTRTEDRIVWKTLAGDPWIKLTAKYPG NFTATIEWQAEHDAIMRHILERPDPTDVFQNKAANVCWA KALVPVLKTAGIDMTTEQWNTVDYFETDKAHSAEIVLN QLCVRFPGLDLDSGLFSAPTVPLSIRNNHWNSPNPMY GLNKEVVQLSRRYPQLPRAVATGRVYDMNTGTLRNVD PRINLVNVNRRLPHALVLHNNHEHPQSDFSSFSVSKLKGRTV LUVGEKLSVPGMVWDWLSDRPEATFRRALDLGIPGDVP KYDIIFVNVRTPYKYHHYQQCEDHAIKLSMLTKKACLHL NPGGTCVSIGGYADRASESIIIGAIARLFKSRVCKPKSSL EETEVLFVFIGYDRKARTHNPYKLSSTLTNIYTGSRLHEA GCAPSYHVVRGDIATATEGVINIANSKGQPGGGVCGAL YKKFPESFDLQPIEVGKARLVKGAAKHIIVAGPNPNKVS EVEGDKQLAEAYESIAKIVNDNNYKSVAIPLLSTGIFSGN KDRLTQSLNHLLTALDTDAVAIYCRDKKWEMLTKEA VARREAVEEECISDDSSVTEPDAELVRVHPKSSLAGRKY STSDGKTFSYLEGTKFHQAARDIAEINAMWPVATEANEQ VCMYILGESMSSIRSCKPVEESEASTPPSTLPCLCIHAMTP ERVQRLKASRPBQITVCSSFLPLPKYRITGVQKIQCSQPILFS PKVPAIYHPRKYLVETPPVDETPEPSAENQSTEGTPEQPPL ITEDETRTRPPIIIEEEEEDSISLLSDGPTHQVLQVEADIH GPPSVSSSSWSIPHASDFDVDSLISLDTLEGASVTSGATSA ETNSYFAKSMEFLARPVPAPRTVFRNPHPAPRTRTPSLA PSRACSRTSLVSTPPGVNRVITREELEALTPSRTPSRSVSR TSLVSNPVGVRNVRTBEEFEAFAVQQQRRFDAGAYIFSSD TGQGHLQQKSVRQTVLSEVVILERTEISYAPRLDQEKE ELLRKKLQLNPTPANRSRYQSRKVENMKAITARRILQGL GHYLKAEKGKVECYRTLHPVPLYSSSVNRAFPSSPKVAEEA CNAMLKENFPVTASYCIIPEYDAYLMDVGDASCCLDTAS FCPAKLRSPFPKKHSYLEPTIRSAVPSAIQNTLQNVLAAT KRNCRNTQMRELPLVLDAAFNVECFKKYACNNEYWTF KENPIRLTEENVNYITKLKGPKAAALFAKTHNLNMLQD IPMDRFVMDLKRDVKVTPGTKHTEERPKVQVIQAADPL ATAYLCGIHRELVRRNAVLLPNIHTLFDMAEDFDAIIA

TABLE 4-continued

Comparison of STARR TM self-replicating RNA of the disclosure with comparative self-replicating RNA as described			
Construct	Position	Sequence Type	Sequence
			EHFQPGDCVLETDIASFDKSEDDAMALTALMILEDLGVD AELLTLIEAAFGEISSIHLPTKTKFGAMMKSGMFLTLF VNVTVINIVIASRVLRERLTGSPCAAFIGDDNIVKGVKSDK LMADRCATWLNEVVKIIDAVVGEKAPYFCGGFILCDSVT GTACRVADEPLKRLFKLGKPLAADDEHDDDRRALHEES TRWNNRVGILSELCKAVERSYETVGTTSIIVMAMTTLASSV KSFSYLRGAPITLYG*
STARR TM (SEQ ID NO: 52)	intergenic region	nucleotide	CCTGAATGGACTACGACATAGTCTAGTCGCCAAGGC CGCCACC
STARR TM	transgene ORF	nucleotide	n/a (depends on gene of our interest)
STARR TM (SEQ ID NO: 53)	3' UTR	nucleotide	ACTCGAGTATGTTACGTGCAAAGGTGATTGTCACCCCC CGAAAGACCATATTGTGACACACCCTCAGTATCACGC CCAAACATTACAGCCGGTGTCAAAAACCGCGTGG ACGTGGTTAACATCCCTGCTGGGAGGATCAGCGTAA TTATTATAATTGGCTTGTTGCTGGCTACTATTGTGGCC ATGTACGTGCTGACCAACCAGAAACATAATTGAATAC AGCAGAAATTGCAAGCTGCTTACATAGAACCTCGCG CGATTGGCATGGCCTTAAATTTTATTTTT CTTTCTTTCCGAAATGGATTTTGTTTTAATATTCA AAAAAAAAAAAAAAAATCTAGAAAAAA AAAAAAAAAAAAAAA AAAAAAAAAAAAAAA AAAAAAAAAAAAAAA
Comparitive	5' UTR	nucleotide	unknown
Original (SEQ ID NO: 54)	non-structural gene ORF	nucleotide	ATGCCCGAGAAGGTGCACGTGGACATCGAGGAGGACA GCCCTTCCTGAGGGCCTGAGAGGAGCTCCACA GTTCGAAGTGGAGGCCAAGCAGGTGACCGACAACGAC CACGCCAACGCCAGGGCCTTCAAGCACCTGCCCAGCA AGCTGATCGAGACCGGGCTTGAGCCCCAGCGACACCAT CTTGGACATCGGAGGCCAGGGCAGCAGGAGATGTAC AGCAAGCACAAGTACCACTGCATCTGCCCATGAGGT GCCCGAGGCCAGGGACAGGCTGTACAAGTACGCCAC CAAATGAAGAGAACTGCAAGGAGATCACCACAA GGAGCTGGACAAAGAAATGAAGGAGCTGGCCCGCTG ATGAGCGACCCGACCTGGAGACCGAGACAATGTGCC TGCAAGCAGCAGAGAGCTGCAAGTACAGGGCCAGGT GGCCGTCTACCAAGACGTGTACGCCCTGACGGCCCC ACCAGCTGTACCCACAGGCCAACAAAGGCCTGAGGG TGGCCTACTGGATCGGCTCGACACCACACCTCATG TTCAAGAACCTGGCCGGCCCTACCCAGCTACAGCA CCAACTGGGCCACGAGACCGCTGCTGACCCAGGAA CATCGGCTGTGCAAGCAGCTGATGGAGAGGAGC CGGAGAGGCATGAGCATCTGAGGAAGAAATACCTGA AGCCCAAGCAACACAGTCTGCTTCAAGCTGGCAGCAC CATCTACACAGAGAAGAGGGGACCTGCTCAGGAGCTGG CACCTGCCAGCGTGTCCACCTGAGGGCAAGCAGA ACTACACCTGCAAGTGTGAGACCATCTGAGCTGCGA CGGCTACGTGGTGAAGAGGGATGCCATCAGCCCCGGC CTGTACGGCAAGCCCCAGCGGTACGCCCTACATGCA ACAGGGAGGGCTTCTGTGCTCAAGGTGACCGACAC CCTGAACGGCGAGAGGGTGAAGCTCCCGTGTGCA TAGGTGCCACCGACGTGAGGCCAGGATGACCGCA TCCGGCCACCGACGTGAGGCCAGCAGGCCAGAA GCTGCTGTGGGCTGAACCAAGGGATCGTGGTCAAC GGCAGGACCCAGAGGAACACAAACACAATGAAGAAC TACCTGCTGCCCGTGGTGGCCAGGCTTCCAGGTG GGCCAAGGAGTACAAGGGGACAGGCAGCTGGTGTG GCCCTGGGCCACCGACGGGAGGAGACAGCAGAG TGTGCTGGGCTTCAGGGCAGCAGCTGGTGTG TCTACAAGAGGCCAGACCCAGACCATCATCAAGGT GAACAGCAGACTTCCACAGCTTGTGCTGCCAGGATC GGCAGCAACACCCCTGGAGATCGGCCCTGAGGCCAGG TCAGGAAGATGTGGAGGAACACAAGGGCCAGGCC ACTGATACCGCCGAGGACAGTGTGAGGAGGCCAG GCTGCCAGGCCAGGGCAAGGAGGTGAGGGAGGCCAG GAACGTAGGGGCCCTGCCACCCCTGGCTGCCAG TGGAGGAACCCACCCCTGGAGGCCAGGTGGACCTGAT

TABLE 4 -continued

Comparison of STARRTM self-replicating RNA of the disclosure with comparative self-replicating RNA as described

Construct	Position	Sequence Type	Sequence
			GCTGCAGGAGGCCGGCGCCGGAAGCGTGGAGACACCC AGGGGCTGATCAAGGTGACAGCTACGACGGCGAGG ACAAGATCGCAGCTACGCCGTGCTGAGCCCACAGGC CTGCTGAAGTCCGAGAAGCTGAGCTGCATCCACCCA CTGGCCGAGCAGGTGATCGTATCACCCACAGCGGCA GGAAGGGCAGGTACGCCGTGAGGCCCTACCCACGGCA GGTGGTCGTGCCCAGGGCCACGCCATCCCCGTGAG GACTTCCAGGCCCTGAGCGAGAGCGCACCATCGTGT ACAACGAGAGGGAGTTCTGTAACAGGTACCTGCACCA TATCGCACCACCGGCCGAGGCCCTGAACACCGACGAG GAATACTACAAGACCGTGAAGGCCACGCGAGCACGAC GCGAGTACCTGTACGACATCGACAGAAAGCATGCGT GAAGAAAGAGCTGGTACCGGCTGGGACTGACCGGC GAGCTGGTGGACCCACCTTCCACGAGTTCCCTACGA GAGCCTGAGGACCGAGACCCGCCGCTCCCTACCGGTG CCCACCATCGGCGTGTACGGCGTGCCTGGCAGCGGAA AGAGCGGCATCATCAAGAGCCCGTGACCAAGAAAGA CTTGGTGTACGGCCAAGAAAGAGAACTGCGCGAG ATCATCAGGGACGTAAAGAAGATGAAAGGCCCTGGAGC TGAACGCCGACCCCGTGGACAGCGTGTGCTGAAACGG CTGCAAGCACCCCGTGGAGACCCCTGTACATCGACGAG GCCTTCGCTTGCCACGCCGACCCCTGAGGCCCTGAT CGCCATCATCAGGCCAACAGAACGGCTGTGCGGCC GACCCCAAGCAGTGGGCTTCTTCACATGTATGTGCTT GAAGGTGCACTTCAACCACGAGATCTGACCCAGGTG TTCCACAAAGAGCATCAGCAGGGGTGACCAAGAGCG TGACCAAGCGTGTGAGCACCCTGTTCTACGACAAGAA AAATGAGGACCAACACCCAAAGGAGACCAAATCGTG ATCGACACACAGGCAGCACCAAGGCCAACAGCAGGACG ACCTGATCTGACCTGCTTCAGGGCTGGGTGAAGCA GCTGCAGATCGACTACAAGGCCAACAGAGATCATGACC GCCGCTGCCAGCCAGGGCTGACCGAGGAAGGGCGTGT ACCCCGTGAAGGTACAAGGTGAACGAGAACCAACTGTA CGCTCCACCAGCAGCACGTGAACGTGCTGCTGACC AGGACCGAGGAAGGATCGTGTGGAAGACCCCTGGCG GCGACCCCTGATCAAGACCTGACCGCCAAGTACCC CGGCAACTTCACCGCCACCATGAGGCACATCTGGAGAGGC GAGCACGACGCCATCATGAGGCACATCTGGAGAGGC CCGACCCCAACCGACGCTTCCAGAACAAAGGCCAACGT GTGCTGGCCCAAGCCCCCTGGTCCCCCTGCTGAAGACC GCCGGCATCGACATGACCCAGAGCAAGGCCCCACAGCGCG TGGACTACTTCGAGACCGACAAGGCCAACAGCGCGA GATCGTGTGAACCGAGCTGTGCGTGAAGGTTCTCGGCC TGGACCTGGACAGCGGCCCTGTTCAGGCCCCCCACCGT GCCACTGAGCATCAGGAACAAACACTGGACAAACAGC CCCAGCCCAAACATGTACGGCTGAACAAGGAGGTGG TCAGGCAGCTGAGCAGGCCGTACCCACAGCTGCCAG GGCGCTGGCCACCGGCCAGGGTGTACGACATGAACACC GGCACCTGAGGAACATCGACCCCCAGGATCAACCTGG TGCCCCTGAAACAGGCCGTGGCCACGCCCTGGTGCT GCACCCAAACAGCAGCACCCACAGAGCGACTTCAGCTCC TTCGTGAGCAAGCTGAAAGGCAGGACCGTGTGGTGT TGGCGAGAAGCTGAGCGTGCCTGGCAAGATGGTGG CTGGCTGAGCGACAGGCCGAGGCCACCTTCCGGGCC AGGCTGGACCTCGGCATCCCCGGCAGCTGGCCAAGT ACGACATCATCTCGTGAACGCTGAGGACCCATACAA GTACCAACATTACCGAGCTGAGGACCGCCAC AAGCTGAGCATGCTGACCAAGAAGGCCCTGCTGACCC TGAACCCCGGAGGGCACCTGCGTGAAGCATCGCTACGG CTACGCCGACAGGGCAGCGAGAGCATCATTGGCGCC ATGCGCAGGCTGTTCAAGTTAGCAGGGTGTGCAAC CCAAGAGCAGGCTGGAGGAACCGAGGTGCTGGTGT GTTCATCGCTGACCGGAAGGCCAGGACCCACAC CCCTACAAGCTGAGCGACACCTGACAAACATCTACA CCGGCAGCAGGCTGACAGGCCGGCTGCGCCCCAG CTACCGTGGTCAGGGCGATATGCCACCGCACC GAGGGCTGATCATCAAGCTGCCAACAGCAAGGGCC AGCCCGGAGGGCGAGTGTGCGGCCCTGTACAAGAA GTTCCCCGAGACTTCGACCTGCGCCATCGAGGTG GGCAAGGCCAGGCTGGTGAAGGGCGCGCTAACGACA TCATCCACGCCGTGGGCCCAACTTAAACAAGGTGAG CGAGGTGGAAGGCAGACAAGCAGCTGGCCGAAGCCTAC GAGAGCATGCCAAGATCGTGAACGACAATAACTACA AGAGCGTGGCCATCCACTGCTCAGCACCGCATTTC

TABLE 4 -continued

Comparison of STARRTM self-replicating RNA of the disclosure with comparative self-replicating RNA as described

Construct	Position	Sequence Type	Sequence
			AGCGGCAACAAGGACAGGCTGACCCAGAGCTGAACC ACCTGCTCACGCCCTGGACACCACCGATGCCGACGT GGCCATCTACTGCAGGGACAAGAAAGTGGGAATGACC CTGAAGGGAGGGCGTGGCAGGGGGAGGCCCTGGAA GAGATCTGCATCAGCAGCAGTCCAGCGTGAACCGAGC CCGACGCCAGCTGGTGGAGGTCACCCCAAGAGCTC CCTGGCCGGCAGGAAGGGCTACAGCACCAGCGACGGC AAAGACCTTCAGCTACCTGGAGGGCACCAAGTTCACC AGGCCGCTAAGGACATCGCCGAGATCAACGCTATGTG GCCCGTGCCCACCGAGGCCAACAGCAGGAGCTGCACTG TACATCTGGCGAGAGCATGTCCAGCATCAGGAGCA AGTGGCCCGTGGAGGAAGCGAGGCCAGCACCCACC CAGCACCCCTGCCCTGCCTGTGCATCCACGCTATGACAC CCGAGAGGGTGAGCCGCTGAAGGCCAGCAGGCCGA GCAGATCACCGTGTGCAGCTCTCCACTGCCCAAGT ACAGGATCACCGCGTGCAGAAAGATCAGTCAGCCA GCCCATCTGTTCAAGCCAAAGGTGCCCTACATCC ACCCCAAGGAAGTACCTGTGGAGAACCCACCGTGG CGAGACACCCAGGCCAACGCCAGAGAACAGAACCC GAGGGCACACCCGAGCAGGCCACCCCTGATACCGAGG ACGAGACAAGGACCCGGAGCCAGCATCACCTGCTGAG CGAGGAAGAGGAAGAGGACAGCATCACCTGCTGAG CGACGGCCCCACCCACCCAGGTGCTGCAGGTGGAGGCC GACATCCACGGCCACCCAGGTGCTGCAGCTCCAGCT GGAGCATCCCCACAGCAGCGACTTCGACGTGGACAG CTTGAGCATCTGGACACCCCTGGAGGGCGCAGCGTG ACCTCCGGCACCAGCAGGCCAGAGACCAACAGCTACT TCGCCAAGAGCATGGAGTTCTGGCCAGGCCGTGCC AGCTCCCAAGGACCGTGTCAAGAACCCACCCACCA GCTCCCAGGACAGGACCCCAAGCCTGGCTCCAGCA GGGCCCTGAGCAGGAGCAGCAGCTGGTGGACACCCACC CGCGTGAACAGGTGATCACCCAGGGAAACTGGAG GCCCTGACACCCAGCAGGCCAGGACCCAGGTGCGTGA GCAGGACTAGTGTGGTCAACCCACCCGGCGTGA CAGGGTATCACCCAGGGAGGAATTGAGGCTTCGTG GCCAGCAACAGAGACGGTTCGACGCCGGCGCTACA TCTTCAGCAGGACACCCGGCAGGGACACTTCAGCA AAAGAGCGTGGAGGAGCAGCCGTGCTGAGCGAGGTGG CTGGAGAGGAGCCGAGCTGGAAATCAGCTACCCACCA GGCTGGACCCAGGAAGGAGGAACCTCAGGAAGA AACTGCACTGAACCCACCCAGCAGGAGCAG GTACCAAGAGCAGGAAGGTGGAGAACATGAAGGCCATC ACCGCCAGGGGATCTGCAGGGCTGGGACACTACC TGAAGGCCAGGGCAAGGTGGAGTGTCTACAGGACCC GCACCCCGTGCACACTGACAGCTCACCGTGAACAGG GCCCTTCTCCAGCCCCAAGGTGGCGTGGAGGCCCTGCA ACGCTATGCTGAGGAGAACCTCCACCGTGGCCAG CTACTGCATCATCCCCAGTACGACGCCAACCTGGACA TGGTGGACGGGCCAGCTGCTGCCAGGCC CTTCTGCCCGCCAAGCTGAGGAGCTCCCCAAGAAA CACAGCTACCTGGAGGCCACCATCAGGAGGCCGTG CCAGCGCCATCCAGAACACCCCTGCAGAACGCTGG CGCTGCCACCAAGAGGAACCTGCAACCTGACCCAGATG AGGGAGCTGCCGTGCTGGACAGCGCTGCCTCAACG TGGAGTCTCAAGAAATACCGCTGCAACACAGAGTA CTGGGAGAACCTCAAGGAGAACCCCATCAGGCTGACC GAAGAGAACGCTGTAACATCACCAAGCTGAAGG GCCCCAAGGGCGCTGCCCTGTTGCTAAGACCCACAA CTTGAACATGCTGCAGGACATCCCAATGGACAGGTT GTGATGGACCTGAAGAGGGACGCTGAAGGTGACACCC GCACCAAGCACCCAGGAGGAGGCCAAGGTGCAAGGT GATCCAGGCCGTGACCCACTGGCCACCGCCTACCTGT GCCGCATCACAGGGAGCTGGTGGAGGGGGCTGAACGC CGTGCCTGCTGCCAACATCCACACCCCTGGACATGA GCCCGAGGAACCTGCAACGCCATCATGCCGAGCACTT CCAGCCCGCGACTGCGTGTGGAGACCGACATGCC AGCTTCGACAAGAGCAGGATGACGTATGCCCTGA CCGCTCTGATGATCCTGGAGGGACCTGGGCGTGGACGC CGAGCTGCTCACCCCTGATCGAGGCTGCTTCGGCGAG ATCAGCTCCATCCACCTGCCAACAGACCAAGTTCAA GTTCGGGCTATGATGAAAGGGGAATGTTCTGACC CTGTTCTGAACACCGTGTACACATTGTGATGCCAG CAGGGTGTGGGGAGGGCTGACCGGCAGCCCTGC GCTGCCCTCATCGGCCAGCACACATCGTGAAGGGCG

TABLE 4 -continued

Comparison of STARRTM self-replicating RNA of the disclosure with comparative self-replicating RNA as described

Construct	Position	Sequence Type	Sequence
			TGAAAAGCACAAGCTGATGGCCGACAGGTGCGCAC CTGGCTGAACATGGAGGTGAAGATCATCGACGCCGTG GTGGGCAGAAGGCCCTACTTCTGCGCCGATTCA TCCTGTGCGACAGCGTGACCGCCACCGCCTCAGGGT GGCCGACCCCTGAAGAGGCTGTTCAAGCTGGGCAAG CCACTGGCCGCTGACGATGACGACGAGATGACAGGO GGAGGGCCCTGCGACGAGGAAGCACCAGGTGGAACA GGGTGGGCATCTTGAGCGAGCTGTGCAAGGGCGTGG GAGCAGGTACGAGACCGTGGCACCAGCATACCGT ATGGCTATGACCAACTGGCAGCTCGTCAAGAGCTT CTCCTACCTGAGGGGGGCCCTATAACTCTACGGCT AA
Comparative (SEQ ID NO: 55)	non- structural gene ORF	amino acid	MPEKVHVDI EEDSPFLRALQRSFPQFEVAKQVTNDH NARAFSHLASKLIETEVDPSTDILDIGSAPARRMYSKHKY HC1CPMRCAEDPDRLYKATKLKKNCKEITDKELDKMM KELAAVMSPDPLETETMCLHHDESCRYEGQAVYQDV YAVDGPTSLYHQANKGVRVAYWIGFTTPMPFKNLAGA YPSYSTNWADETVL TARNIGLCSSDVMERSRGRMSILRK KYLKPSNNVLFVGSTIYHEKRDLRSWHLPSVFHLRGK QNYTCRCETIVSCDGYVVKRIAISPGLYKGPSGYATMH REGFLCCVKTDTLNGERVSFPVCTYPATLCDQMTGILA TDVSADDAQKLLVGLNQRIVVNGRQTQRNTNTMKNYLLP VVAQAFARWAEKYKEQDEDERPLGLRDRQLVMGCCWA FRRRHKITSIYKRPTDQTIIKVNNSDFHSFVLPRIGSNTLEIGL RTRIRKMLEENKEPSPSLITAEDVQEAKCAADEAKEVREA EELRAALPPLAADVEEPTLEADVDMLQEAAGGSVETPR GLIKVTSYDGEDKIGSYAVLSPQAVLKSEKLSCHIPLAEQ VIVITHSGRKRYAVEPYHGKVVPPEGHAI PVQDFQALS ESATIVYNEREFVNRYLHHIAHGALNTDEEYYKTVKP SEHDGEYLIDDRKCVKKELVLTGLGLTGEVLVDPPHEF AYESLRTRPAAPYQVPTIGVYVGPGSGKSGI IKSATKKD LVSAKKENCAEIIRDVKKMGLDVNARTVDVSLLN KHPVETLYIDEAFACHAGTLRALIAIIRPKKAVLCGDPKQ CGFFNMCMCLVKHFHNHEICTQVFHKSI SRRCTKS TLFYDKKMRRTNP KETKIVIDTTGSTKPQDDLI LTCFRG WKQLOQIDYKGNEIMTAASQGLTRKGVYAVRYK NPLYAPTSEHVNLLTRTEDRIWKTLAGDPWI KLTAK YPGNFTATIEWQAEHDAIMRHLIERPDPTDV FQNKANV CWAKALVPLKTAGIDMTTEQWNTDVYFETDKAHS VNLQLCVRFGLDLSGFSAPTVPLSIRNNHWDNSP MYGLNKEVVRQLSRRYQPQLPRAVATGRVYDMNTG NYPDPRINLVPVNRRLPHALV LHHNEHPQSDESSFVSKL GRTVLVVGEKLSVPGKMDWLSDRPEATF RARLDLGIP GDVPKYD II FVNVRTPYKYHYQQCEDHAI KLSMLTKKA CLHLPGGTCVSI GYGYADRASESII GAIARLFKFSRVCKP KSSLEETEVLFVFIGYDRKARTHNPYKLSLT NITYTGSRL HEAGCAPSYHVRGDIATATEGV INAANSKGQPGGGVC GALYKKPESFDLQPIEVGKARLVKA KHI IHAVGPNF NKVSEVEGDQKLA EAYESIAKI VNDNNYK SVAI PLLST GFSGNKDRLTQSLNHL LTA DADVAIYCRD KKWEMT LKEA VARRE AEE ICIS DSD SSV TEP DAEL V R V H P K S L A G V E A F V A Q Q R R D A G A Y I F S S D T G Q G H L Q Q K S V R Q T V L S E V V L E R T E L I S Y A P R L D M V D G A S C C L D T A S P C A K L R S F P K K H S L E P T I R S A V P S A I Q N T L Q N V L A A T K R C N C V T Q M R E L P V L D S A F N V E C F K K Y A C N N E Y W E T F K E N P I R L T E E N V N Y I T K L K G P K A A L F A K T H N L N M L Q D I P M D R F V M D L K R D V K V P G T K H T E E R P K V Q V I Q A A D P L A T Y L C G I H R E L V R R L A V L P N I H T L F L E V N T V I N I V I A S R V L R E L T G S P C A A F I G D D N I V K G V K S D K L M A D R C A T W L N M E V K I I D A V V G E K A P Y F C G G F I L

TABLE 4 -continued

Comparison of STARR™ self-replicating RNA of the disclosure with comparative self-replicating RNA as described			
Construct	Position	Sequence Type	Sequence
			CDSVTGTACRVADPLKRLFKLGKPLAADDEHDDDRRA LHEESTRWNRVGILSELCKAVESRYETVGTTSIIVMAMTTL ASSVKSFSYLRGAPITLYG*
Comparative	intergenic region	nucleotide	unknown
Comparative	3' UTR	nucleotide	unknown

15

RNA sequences can include any combination of the RNA sequences listed in Table 4. In some embodiments, RNA sequences of the present disclosure include any combination of the RNA sequences listed in Table 4 in which 0% to 100%, 1% to 100%, 25% to 100%, 50% to 100% and 75% to 100% of the uracil nucleotides of the mRNA sequences are modified. In some embodiments, 1% to 100% of the uracil nucleotides are N1-methylpseudouridine or 5-methoxyuridine. In some embodiments, 100% of the uracil nucleotides are N1-methylpseudouridine. In some embodiments, 100% of the uracil nucleotides are 5-methoxyuridine.

A self-replicating RNA of the disclosure may be obtained by any suitable means. Methods for the manufacture of self-replicating RNA are known in the art and would be readily apparent to a person of ordinary skill. A self-replicating RNA of the disclosure may be prepared according to any available technique including, but not limited to chemical synthesis, in vitro transcription (IVT) or enzymatic or chemical cleavage of a longer precursor, etc.

In some embodiments, a self-replicating RNA of the disclosure is produced from a primary complementary DNA (cDNA) construct. The cDNA constructs can be produced on an RNA template by the action of a reverse transcriptase (e.g., RNA-dependent DNA-polymerase). The process of design and synthesis of the primary cDNA constructs described herein generally includes the steps of gene construction, RNA production (either with or without modifications) and purification. In the IVT method, a target polynucleotide sequence encoding a self-replicating RNA of the disclosure is first selected for incorporation into a vector which will be amplified to produce a cDNA template. Optionally, the target polynucleotide sequence and/or any flanking sequences may be codon optimized. The cDNA template is then used to produce a self-replicating RNA of the disclosure through in vitro transcription (IVT). After production, the self-replicating RNA of the disclosure may undergo purification and clean-up processes. The steps of which are provided in more detail below.

The step of gene construction may include, but is not limited to gene synthesis, vector amplification, plasmid purification, plasmid linearization and clean-up, and cDNA template synthesis and clean-up. Once a protein of interest is selected for production, a primary construct is designed. Within the primary construct, a first region of linked nucleosides encoding the polypeptide of interest may be constructed using an open reading frame (ORF) of a selected nucleic acid (DNA or RNA) transcript. The ORF may comprise the wild type ORF, an isoform, variant or a fragment thereof. As used herein, an "open reading frame" or "ORF" is meant to refer to a nucleic acid sequence (DNA

or RNA) which is capable of encoding a polypeptide of interest. ORFs often begin with the start codon, ATG and end with a nonsense or termination codon or signal.

The cDNA templates may be transcribed to produce a self-replicating RNA of the disclosure using an in vitro transcription (IVT) system. The system typically comprises a transcription buffer, nucleotide triphosphates (NTPs), an RNase inhibitor and a polymerase. The NTPs may be selected from, but are not limited to, those described herein including natural and unnatural (modified) NTPs. The polymerase may be selected from, but is not limited to, T7 RNA polymerase, T3 RNA polymerase and mutant polymerases such as, but not limited to, polymerases able to incorporate modified nucleic acids.

The primary cDNA template or transcribed RNA sequence may also undergo capping and/or tailing reactions. A capping reaction may be performed by methods known in the art to add a 5' cap to the 5' end of the primary construct. Methods for capping include, but are not limited to, using a Vaccinia Capping enzyme (New England Biolabs, Ipswich, Mass.) or capping at initiation of in vitro transcription, by for example, including a capping agent as part of the IVT reaction. (Nuc. Acids Symp. (2009) 53:129). A poly(A) tailing reaction may be performed by methods known in the art, such as, but not limited to, 2' O-methyltransferase and by methods as described herein. If the primary construct generated from cDNA does not include a poly-T, it may be beneficial to perform the poly(A)-tailing reaction before the primary construct is cleaned.

The present disclosure also provides expression vectors comprising a nucleotide sequence encoding a self-replicating RNA that is preferably operably linked to at least one regulatory sequence. Regulatory sequences are art-recognized and are selected to direct expression of the encoded polypeptide.

Accordingly, the term regulatory sequence includes promoters, enhancers, and other expression control elements. The design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed.

The present disclosure also provides polynucleotides (e.g. DNA, RNA, cDNA, mRNA, etc.) directed to a self-replicating RNA of the disclosure that may be operably linked to one or more regulatory nucleotide sequences in an expression construct, such as a vector or plasmid. In certain embodiments, such constructs are DNA constructs. Regulatory nucleotide sequences will generally be appropriate for a host cell used for expression. Numerous types of appropriate expression vectors and suitable regulatory sequences are known in the art for a variety of host cells.

61

Typically, said one or more regulatory nucleotide sequences may include, but are not limited to, promoter sequences, leader or signal sequences, ribosomal binding sites, transcriptional start and termination sequences, translational start and termination sequences, and enhancer or activator sequences. Constitutive or inducible promoters as known in the art are contemplated by the embodiments of the present disclosure. The promoters may be either naturally occurring promoters, or hybrid promoters that combine elements of more than one promoter.

An expression construct may be present in a cell on an episome, such as a plasmid, or the expression construct may be inserted in a chromosome. In some embodiments, the expression vector contains a selectable marker gene to allow the selection of transformed host cells. Selectable marker genes are well known in the art and will vary with the host cell used.

The present disclosure also provides a host cell transfected with a self-replicating RNA or DNA described herein. The host cell may be any prokaryotic or eukaryotic cell. For example, a polypeptide encoded by a self-replicating RNA may be expressed in bacterial cells such as *E. coli*, insect cells (e.g., using a baculovirus expression system), yeast, or mammalian cells. Other suitable host cells are known to those skilled in the art.

The present disclosure also provides a host cell comprising a vector comprising a polynucleotide which encodes a self-replicating RNA sequence provided herein.

A host cell transfected with an expression vector comprising a self-replicating RNA of the disclosure can be cultured under appropriate conditions to allow expression of

62

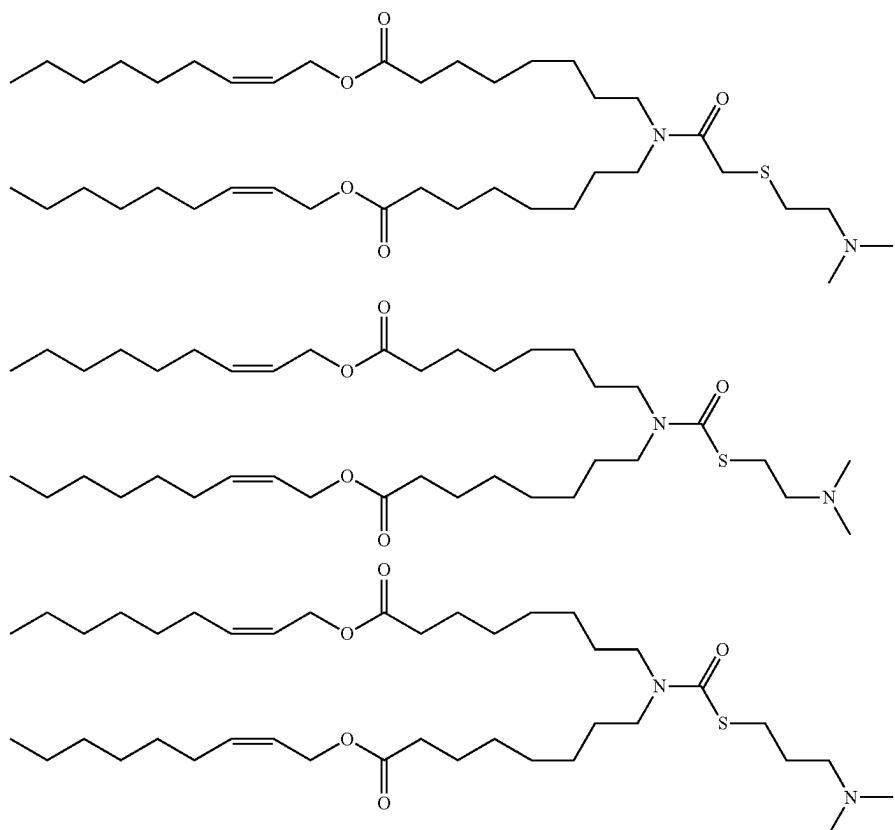
the amplification of the self-replicating RNA and translation of the polypeptide to occur. The polypeptide may be secreted and isolated from a mixture of cells and medium containing the polypeptides. Alternatively, the polypeptides may be retained in the cytoplasm or in a membrane fraction and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art.

The expressed proteins described herein can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins, including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for particular epitopes of the polypeptide.

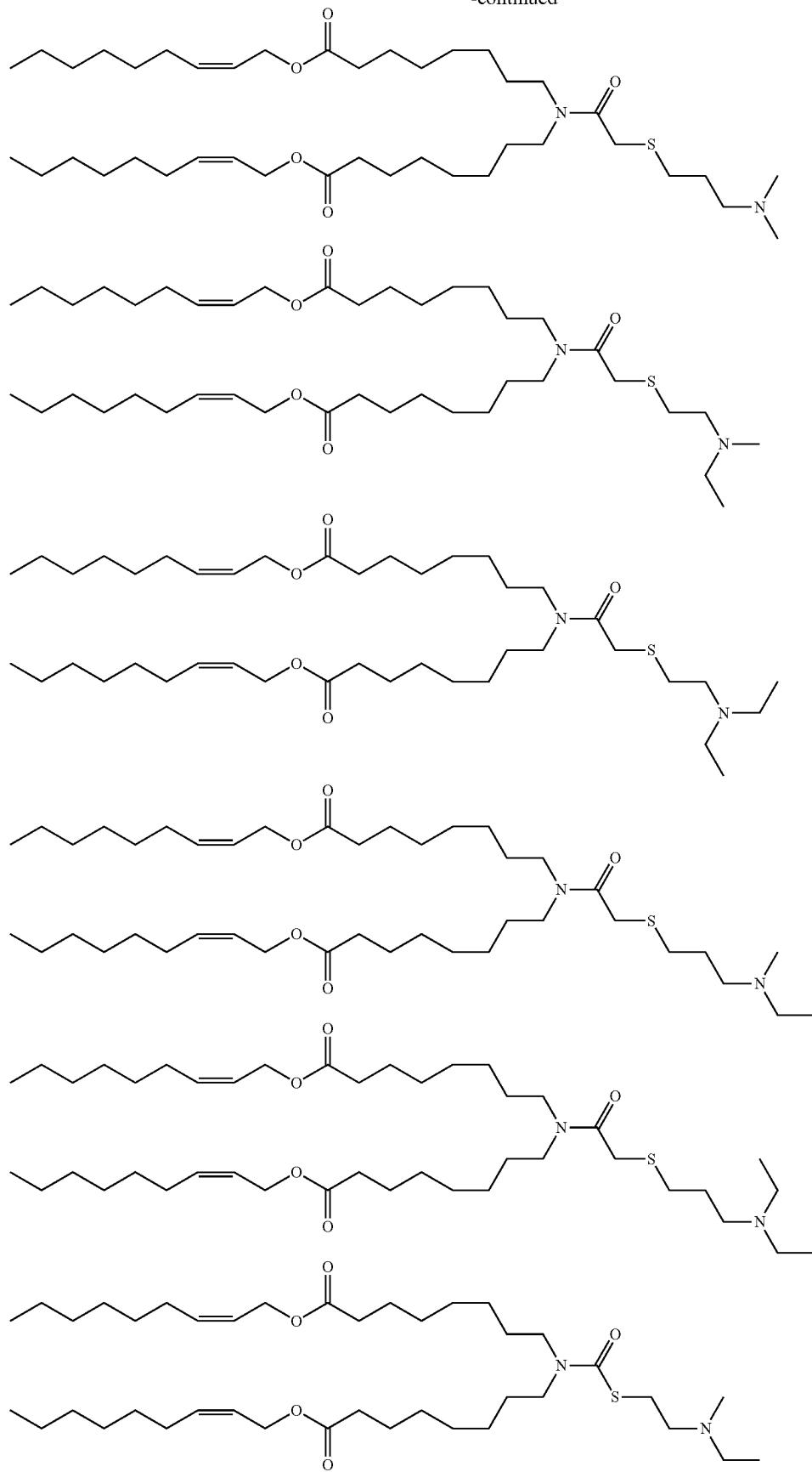
15 Compositions and Pharmaceutical Compositions

Provided herein, in some embodiments, are compositions comprising any of the nucleic acid molecules provided herein. Compositions provided herein can include a lipid. Any lipid can be included in compositions provided herein. In one aspect, the lipid is an ionizable cationic lipid. Any ionizable cationic lipid can be included in compositions comprising nucleic acid molecules provided herein.

Also provided herein, in some embodiments, are pharmaceutical compositions comprising any of the nucleic acid molecules provided herein and a lipid formulation. Any lipid can be included in lipid formulations of pharmaceutical compositions provided herein. In one aspect, lipid formulations of pharmaceutical compositions provided herein include an ionizable cationic lipid. Exemplary ionizable cationic lipids of compositions and pharmaceutical compositions provided herein include the following:

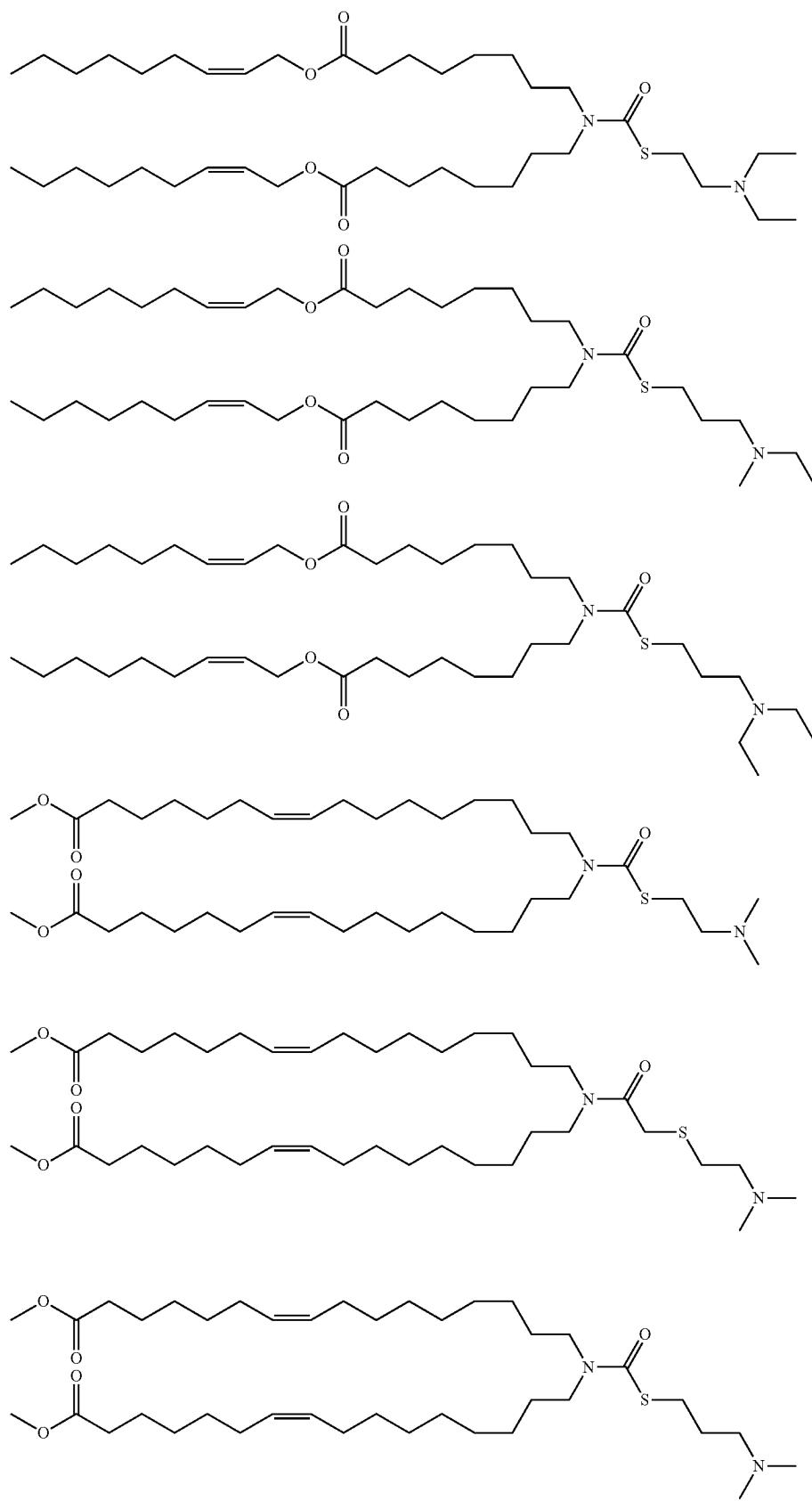


-continued

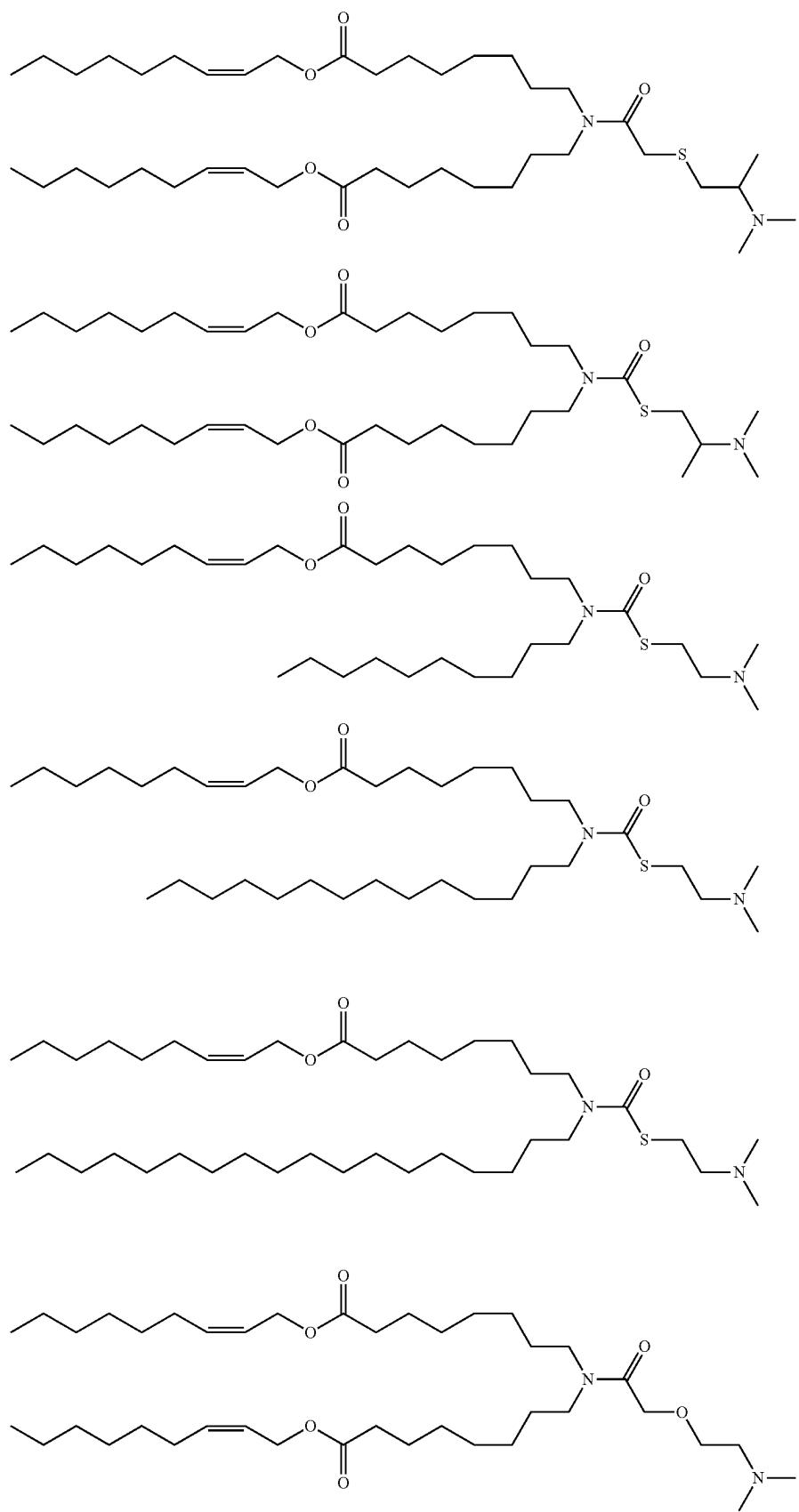


65**66**

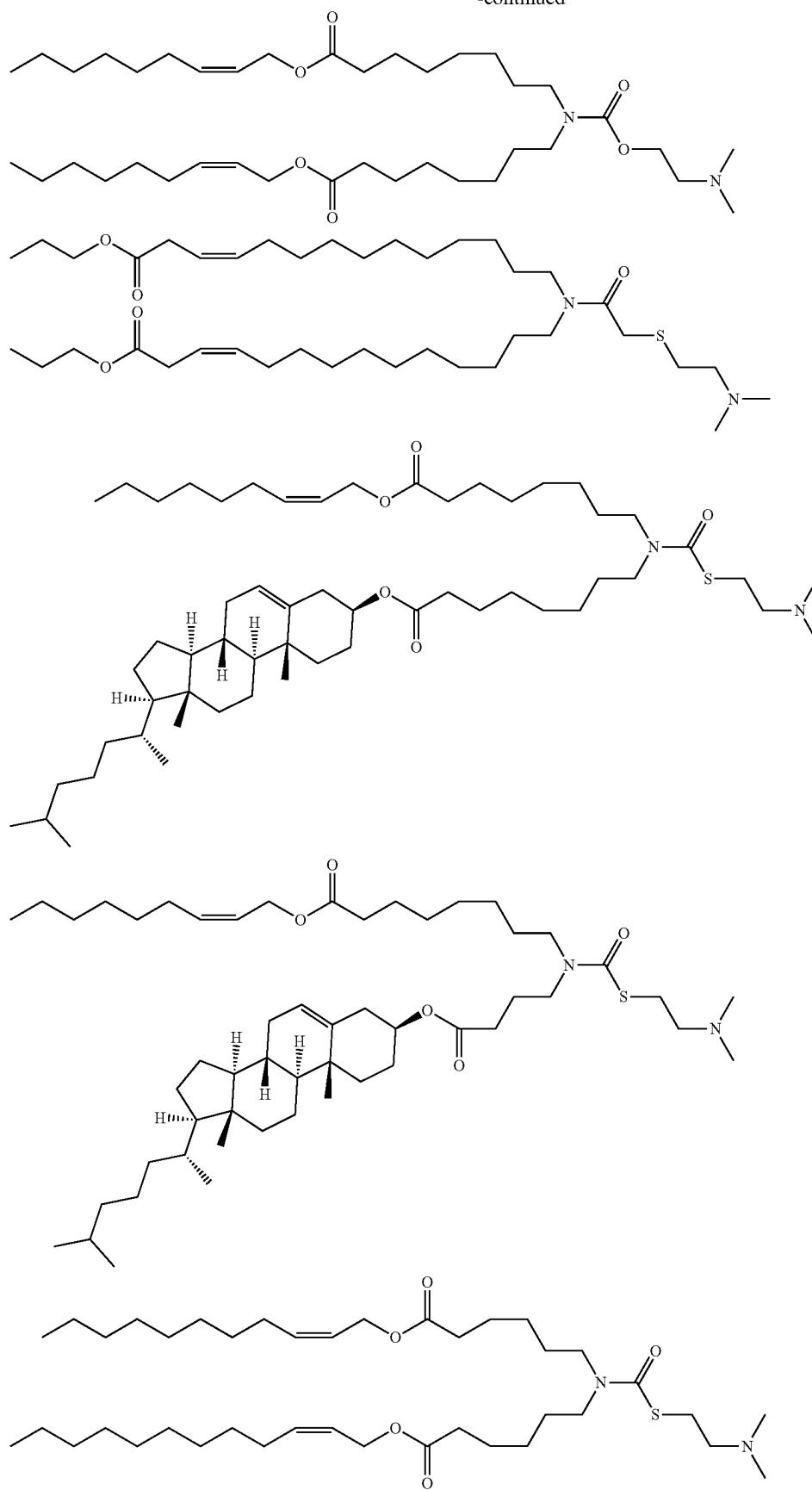
-continued



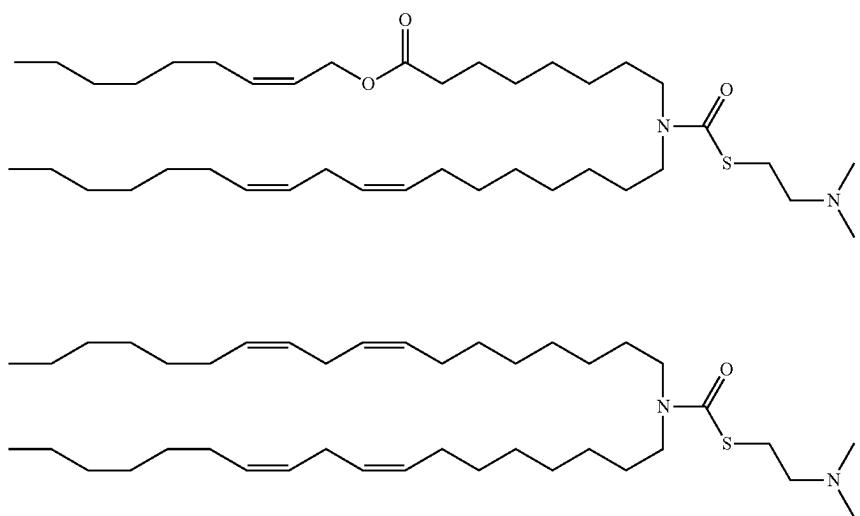
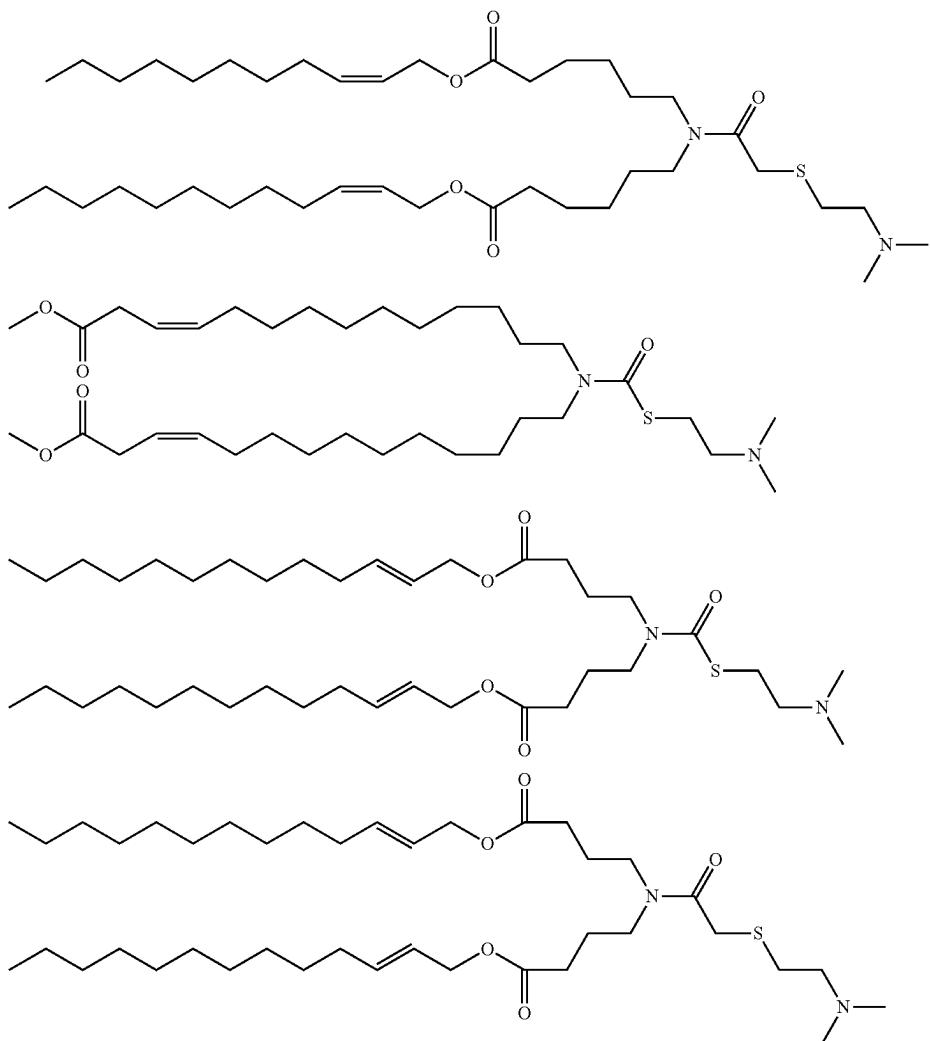
-continued



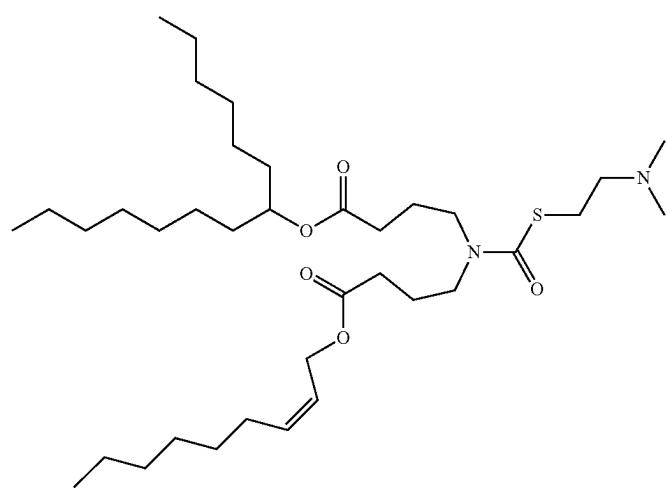
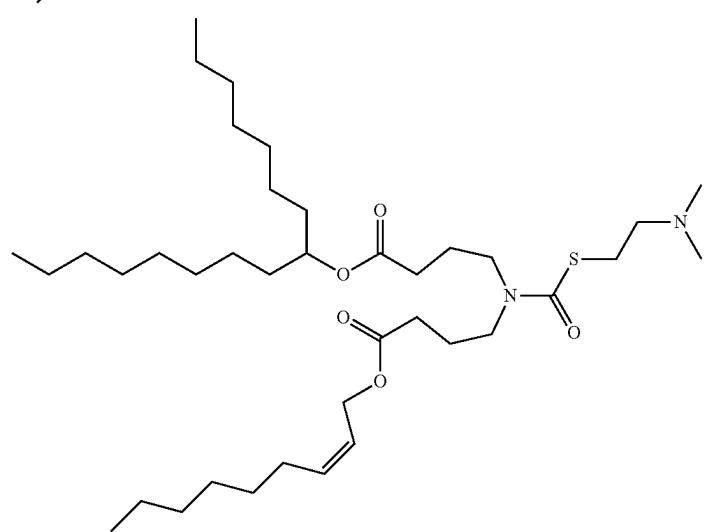
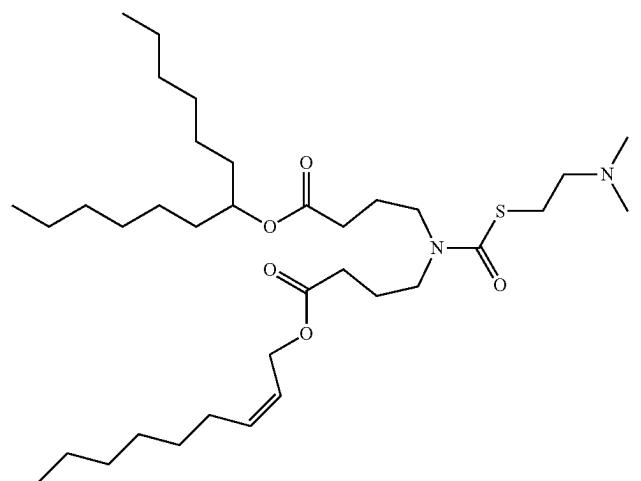
-continued



-continued

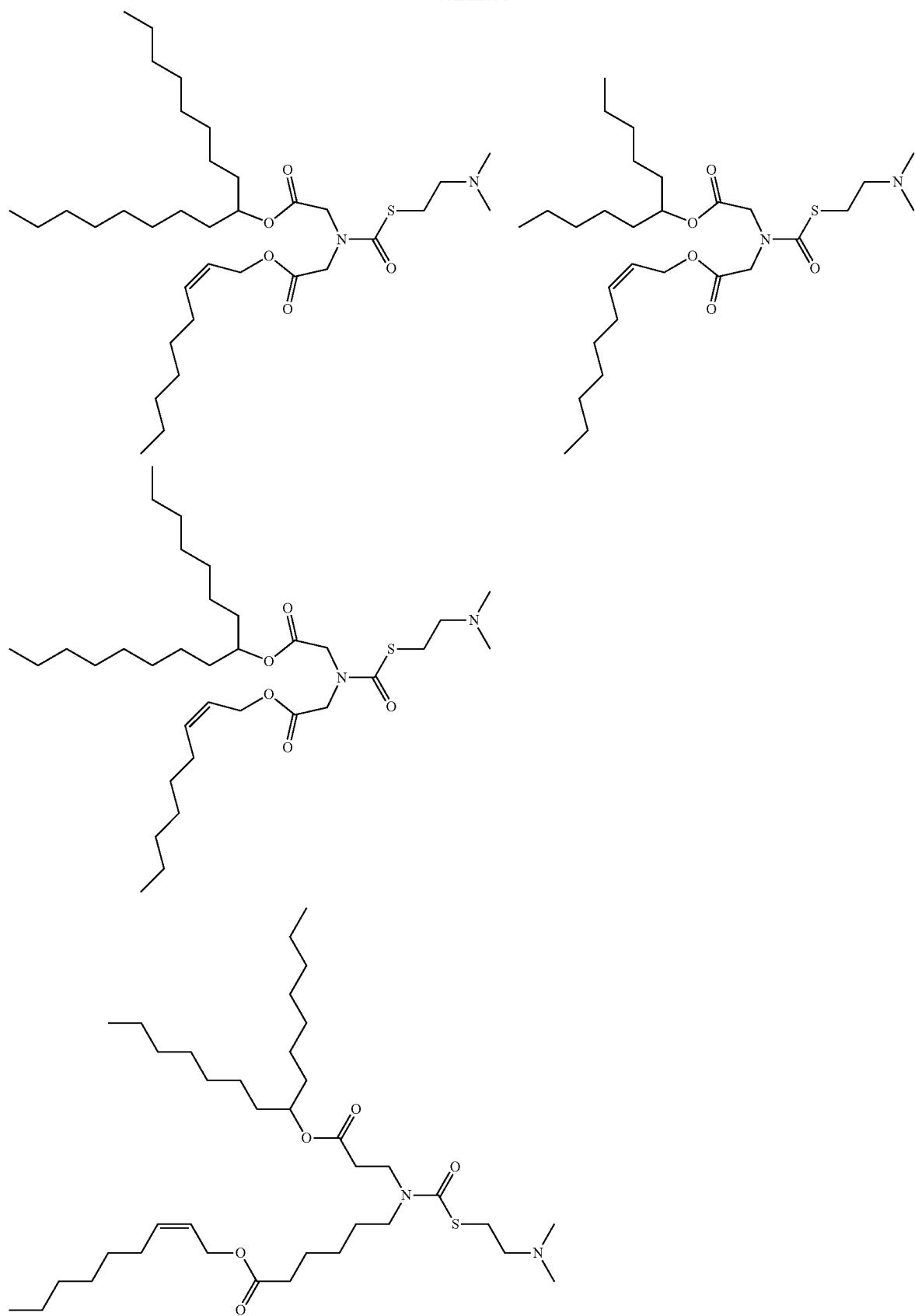


-continued

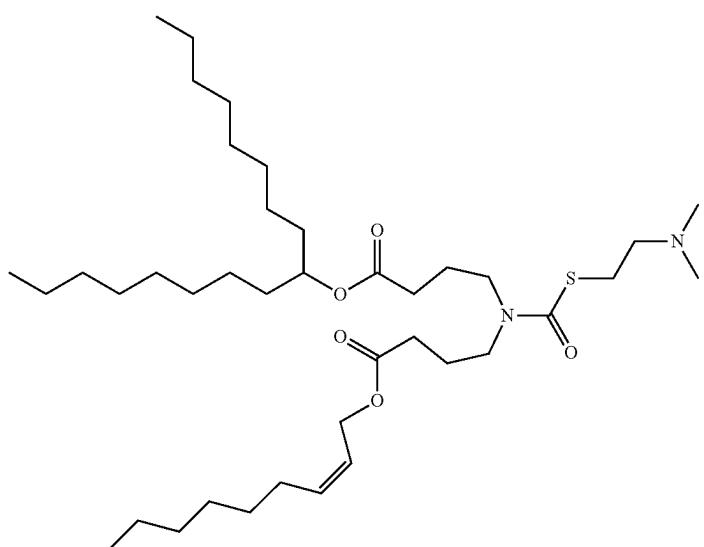
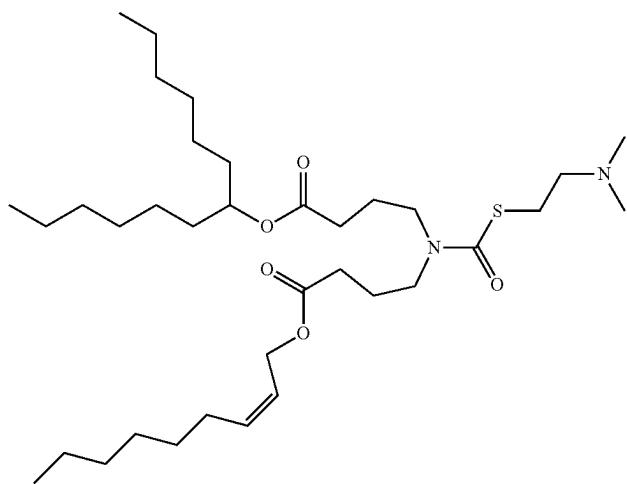
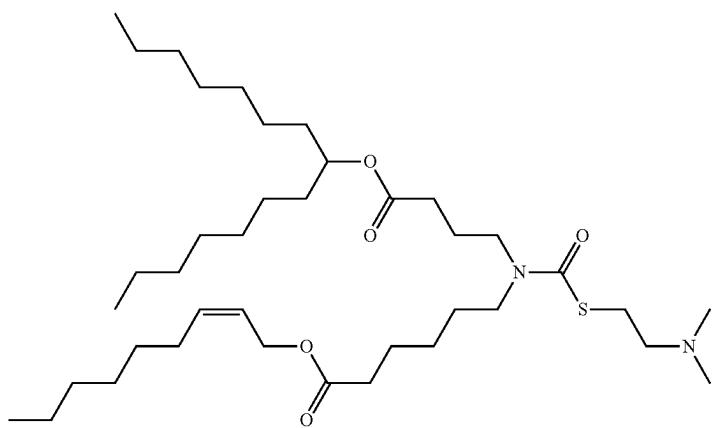


75**76**

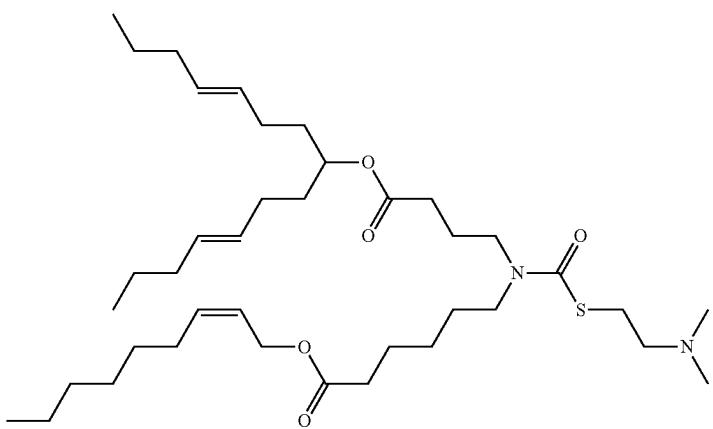
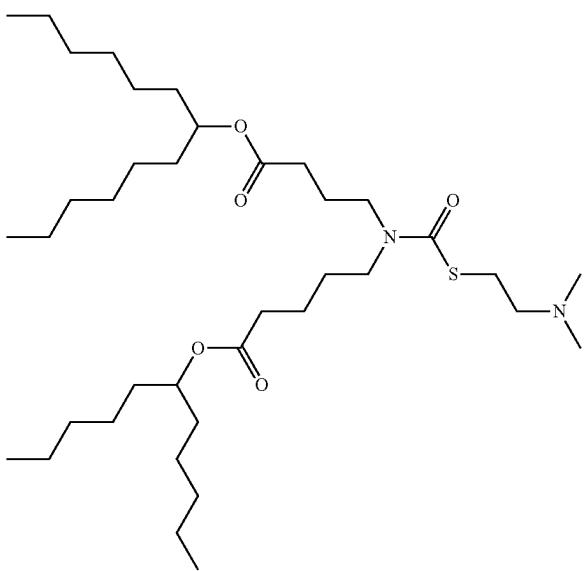
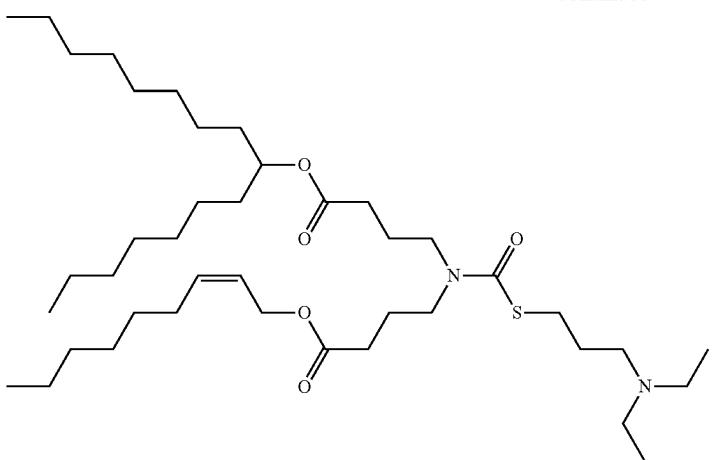
-continued



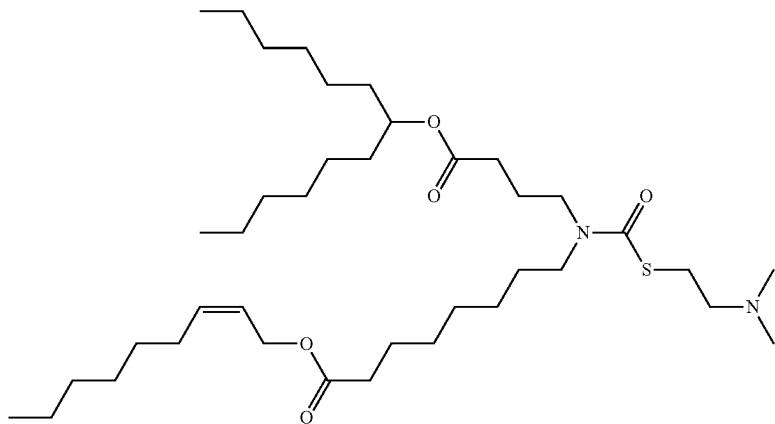
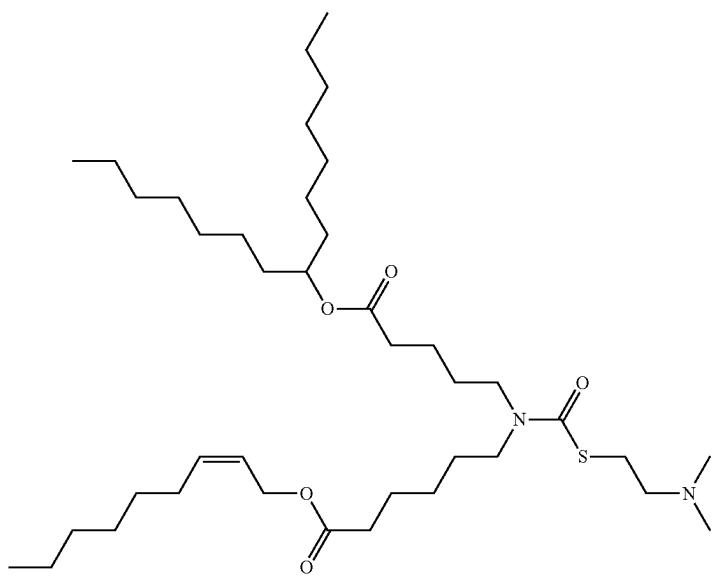
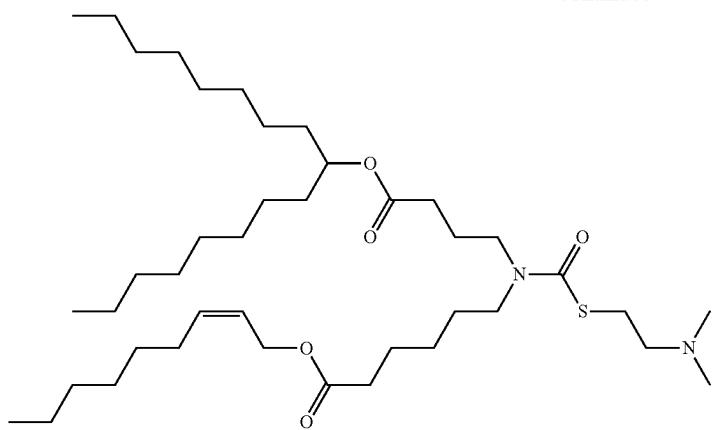
-continued



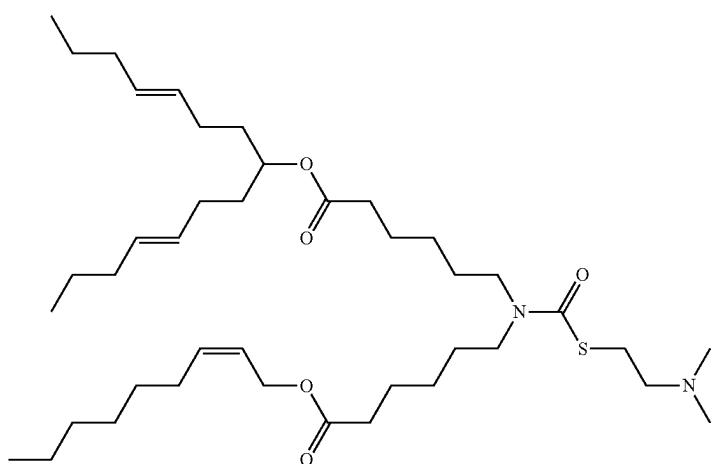
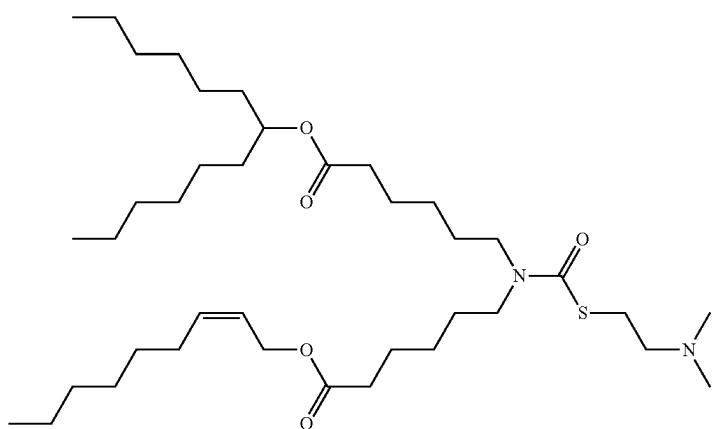
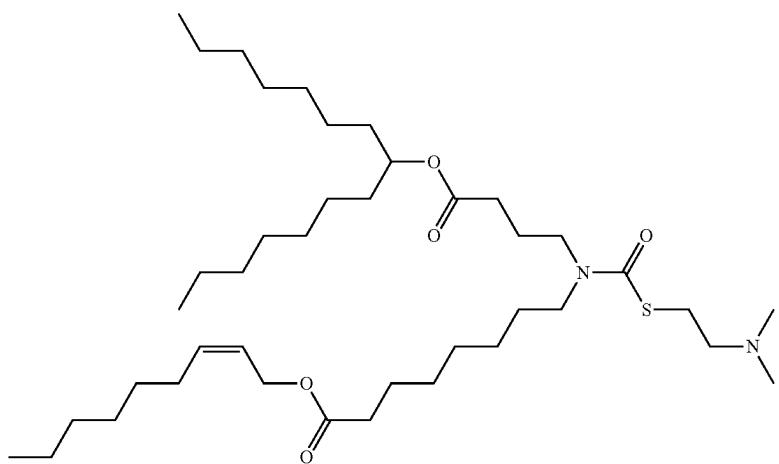
-continued



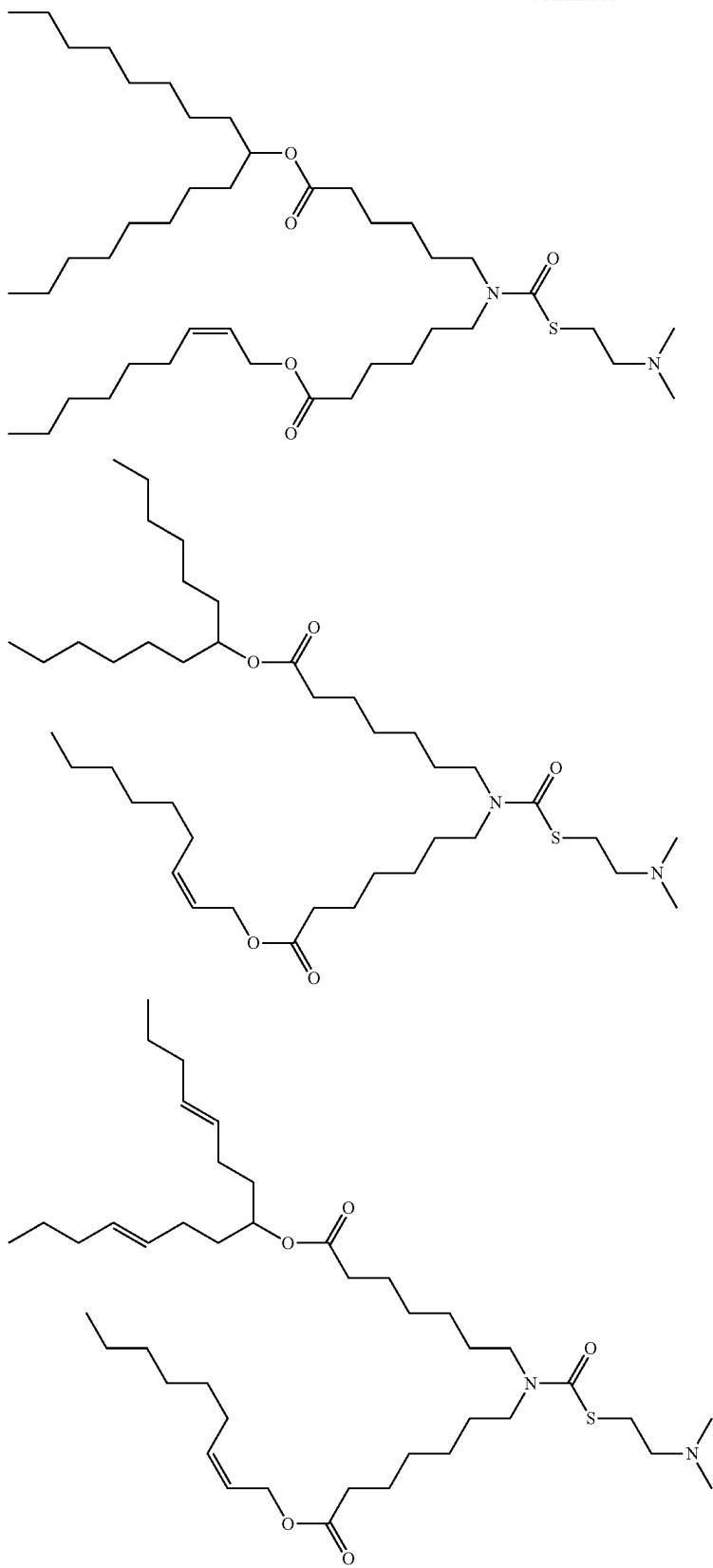
-continued



-continued

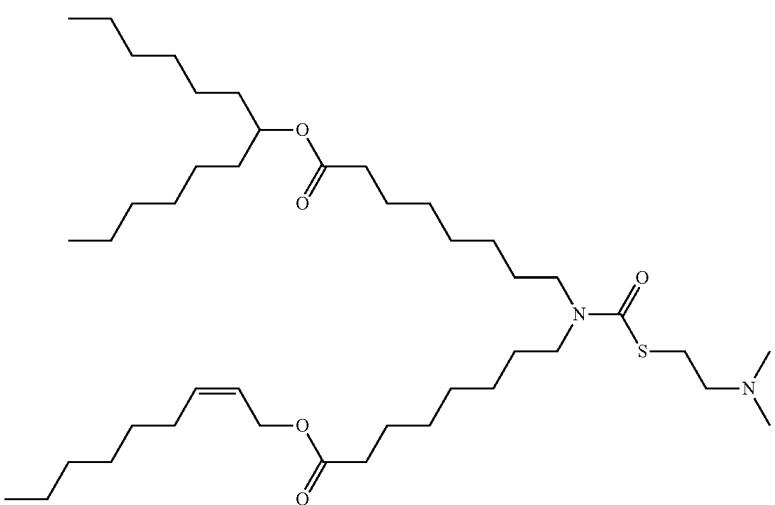
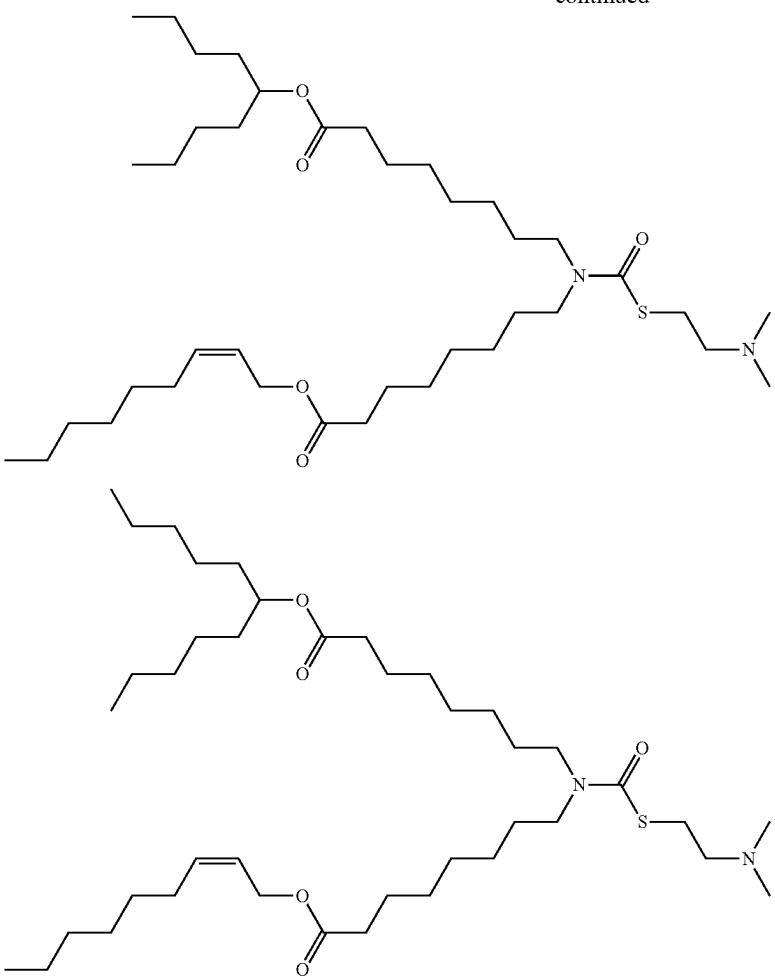


-continued

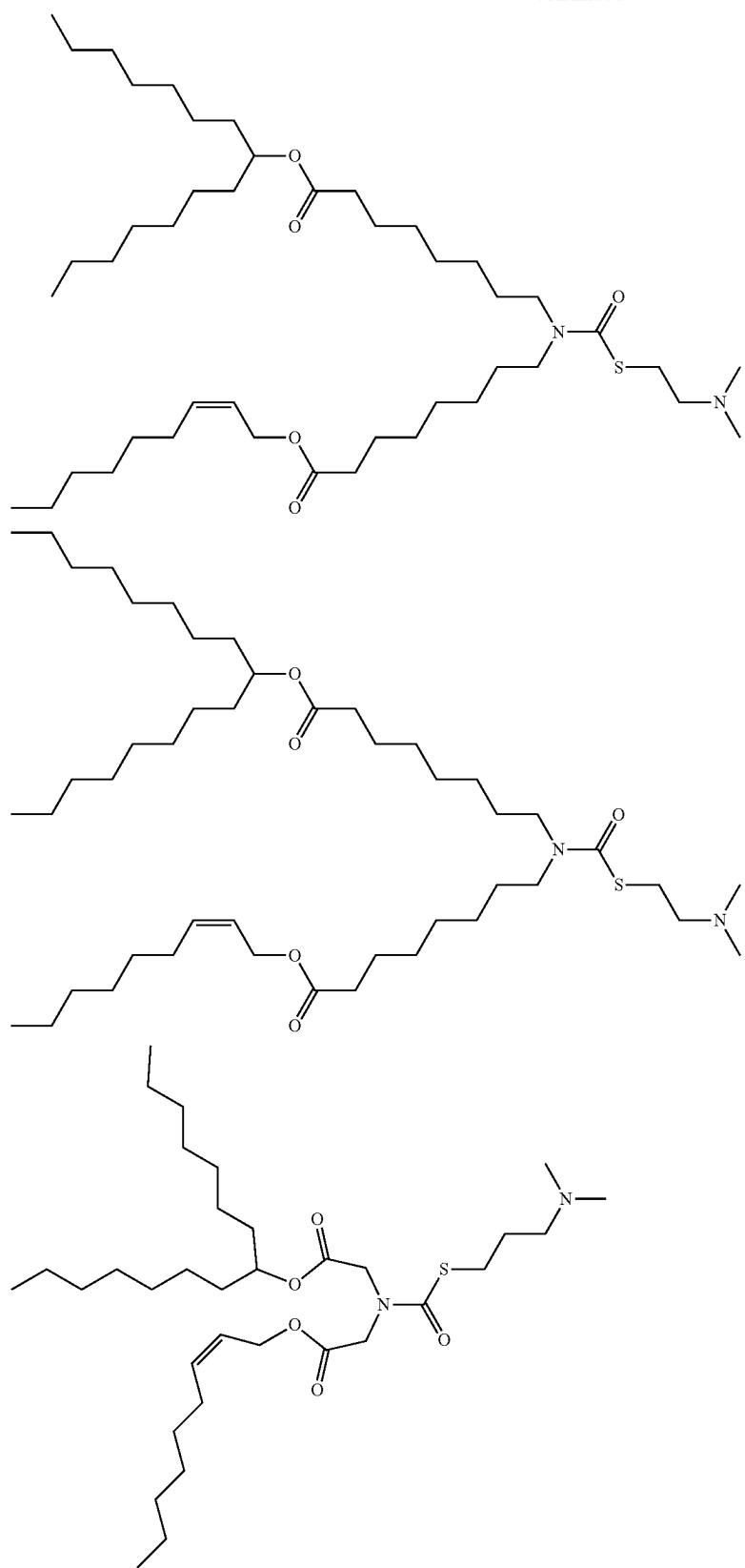


87**88**

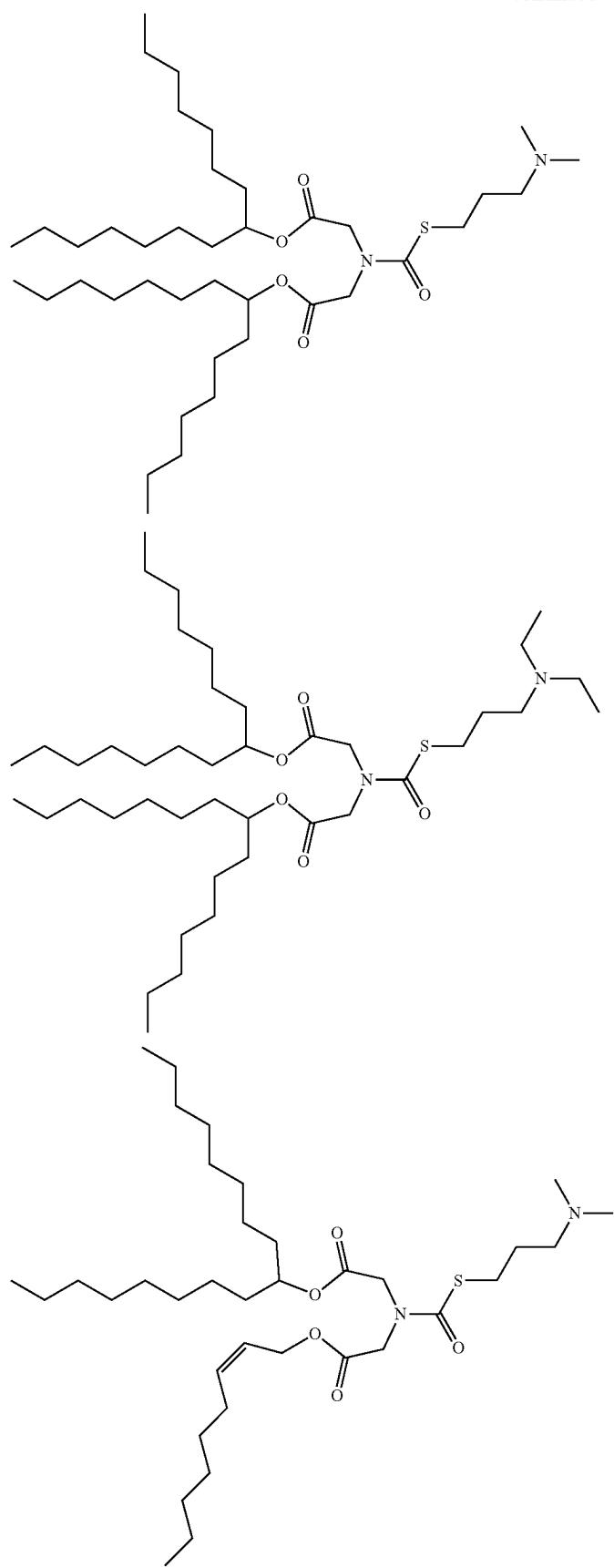
-continued



-continued



-continued

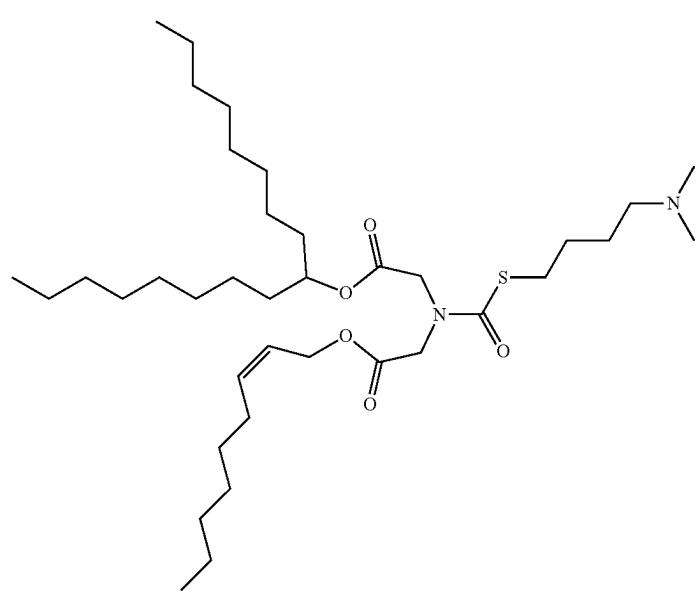
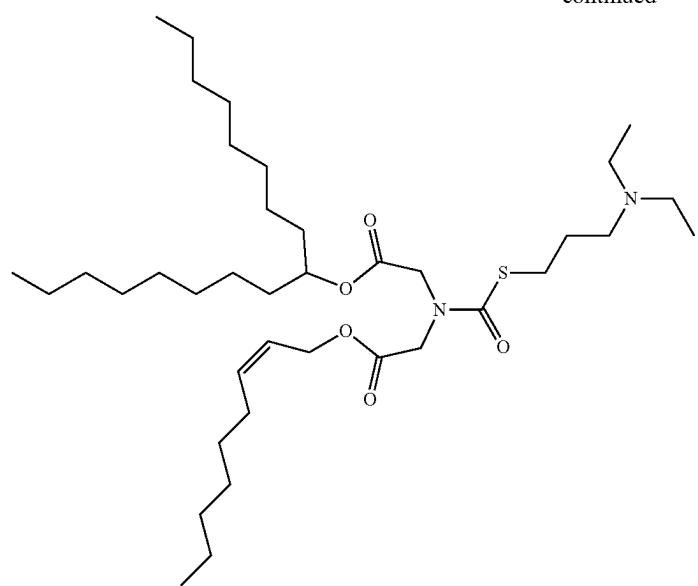


US 12,390,524 B2

93

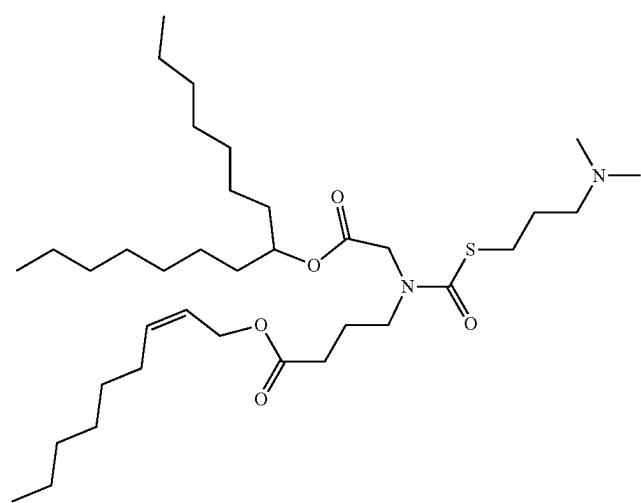
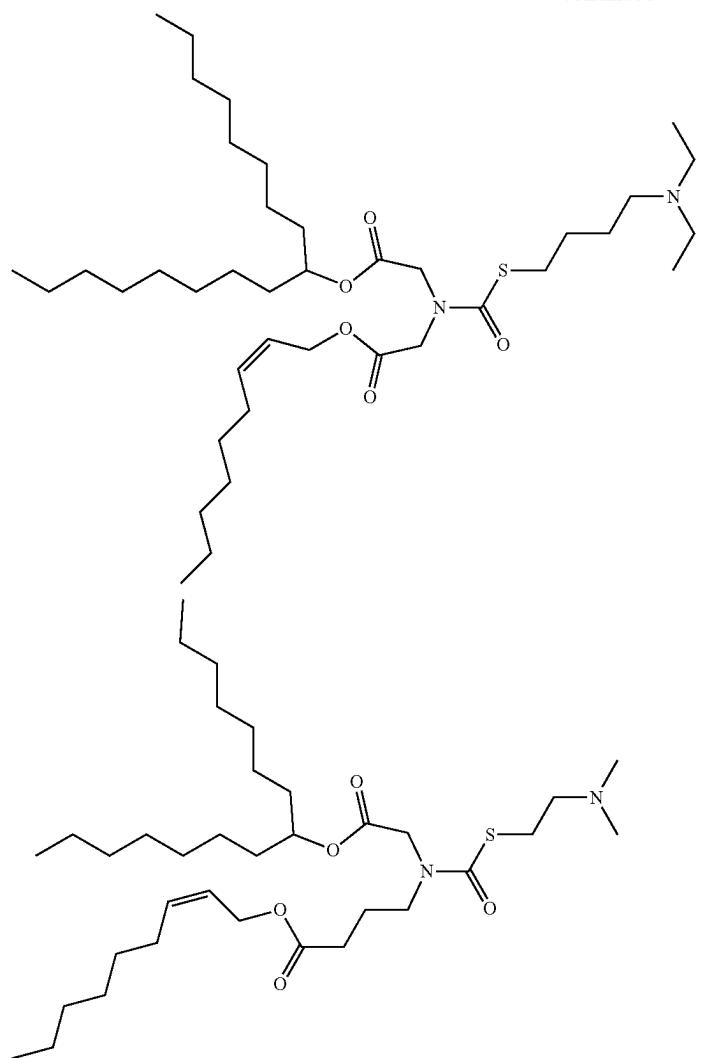
94

-continued

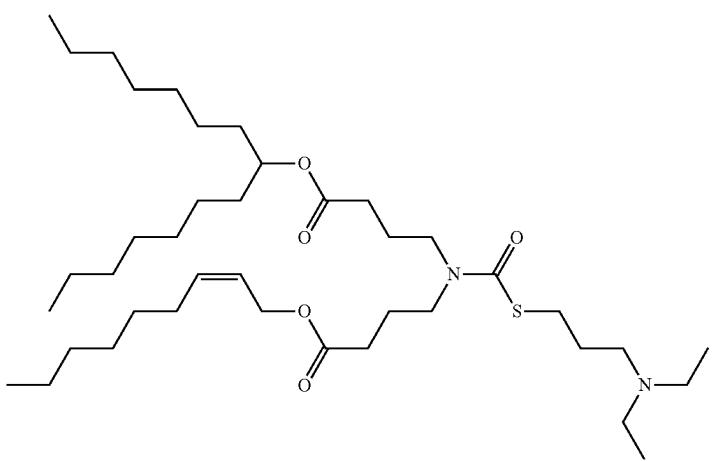
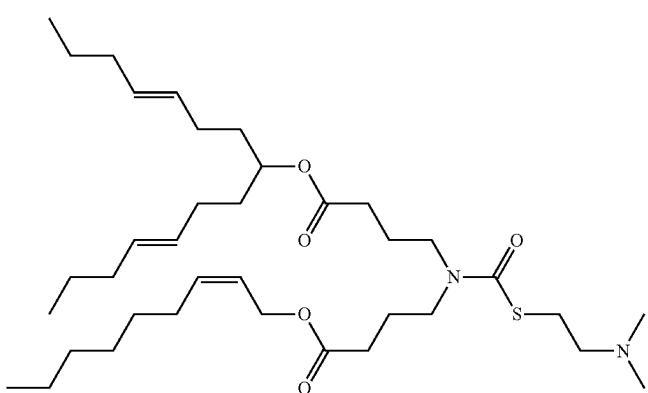
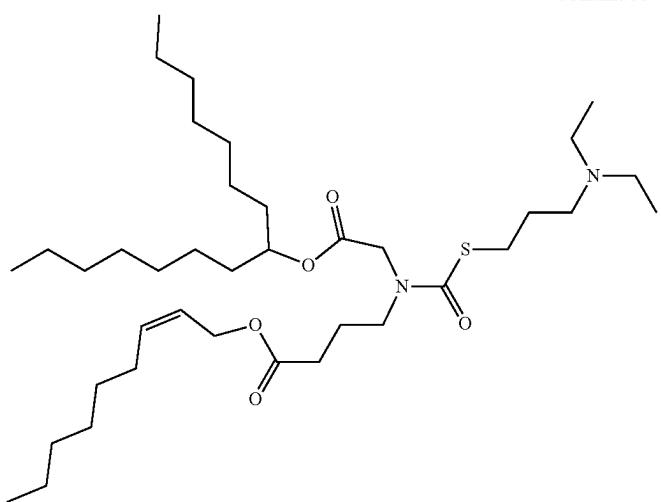


95**96**

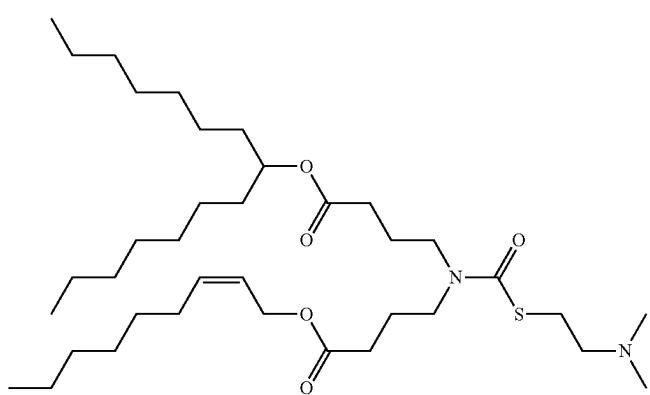
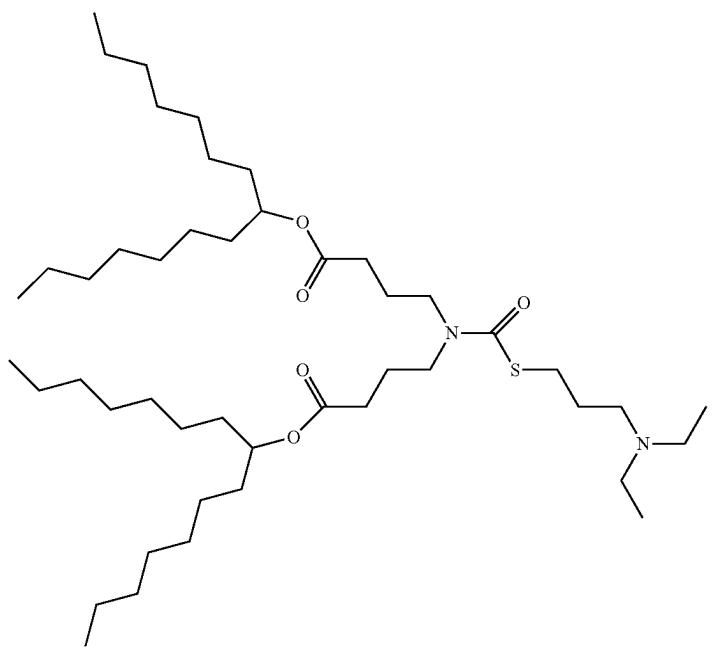
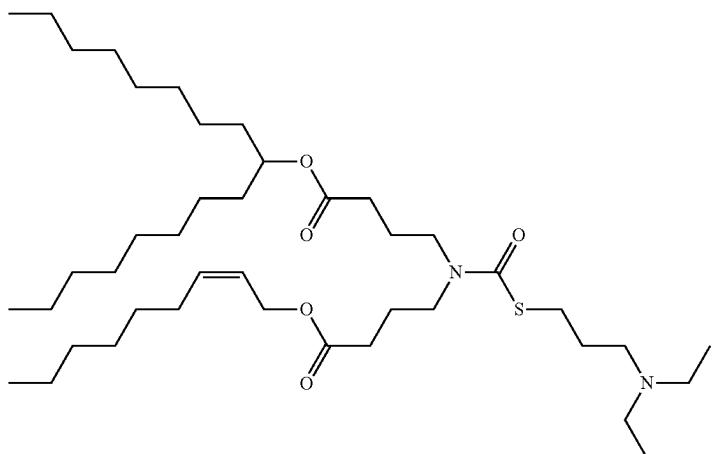
-continued



-continued



-continued

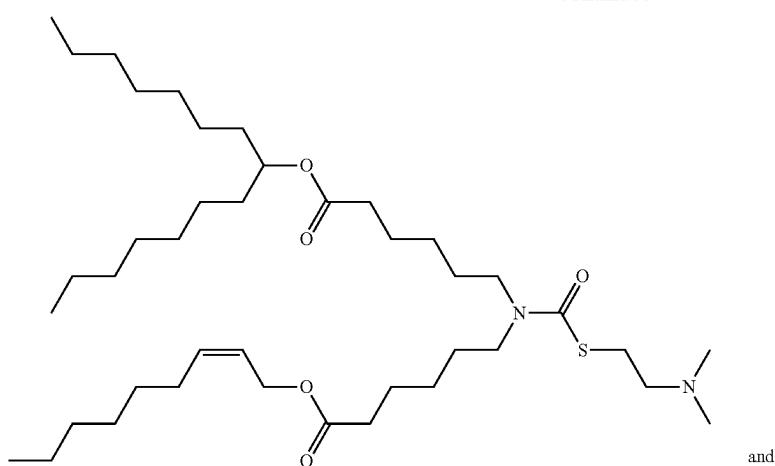


US 12,390,524 B2

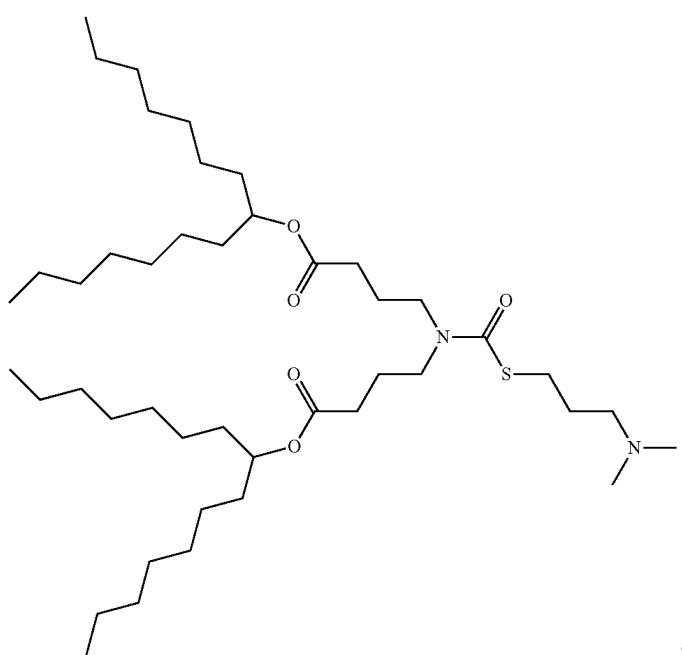
101

102

-continued

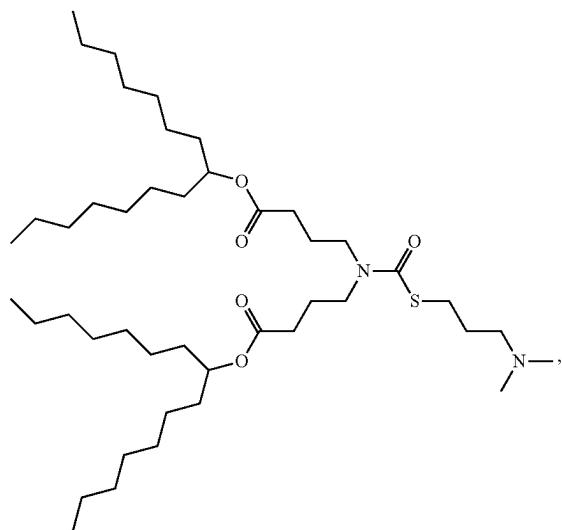
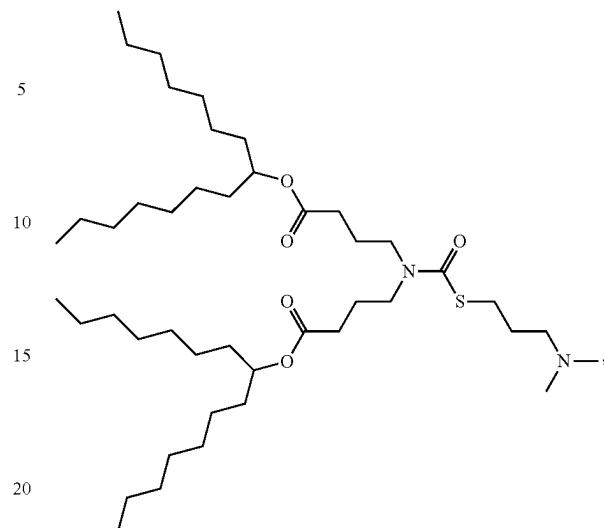


and



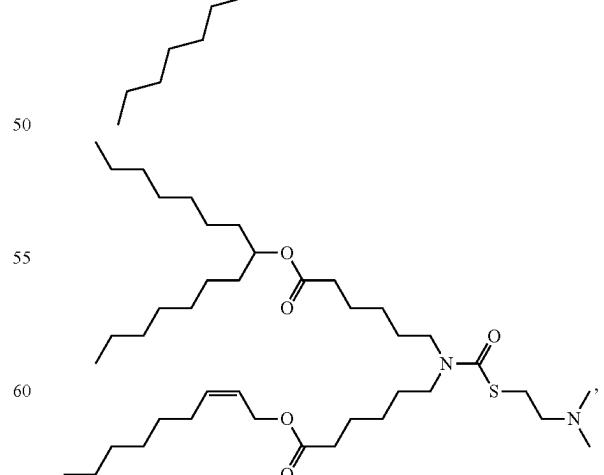
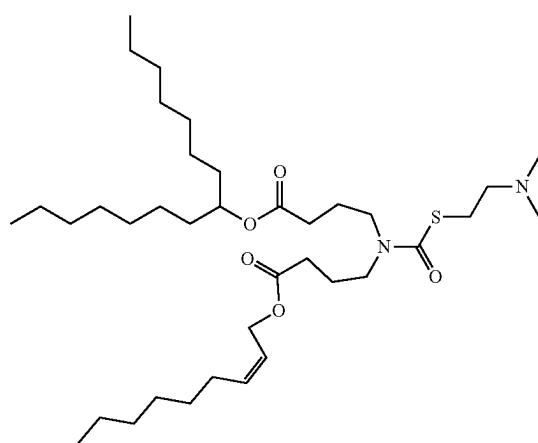
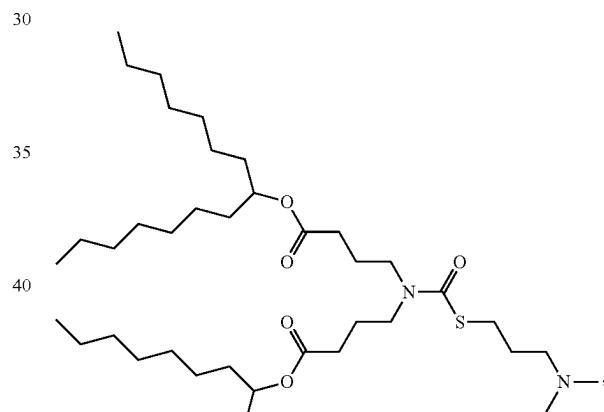
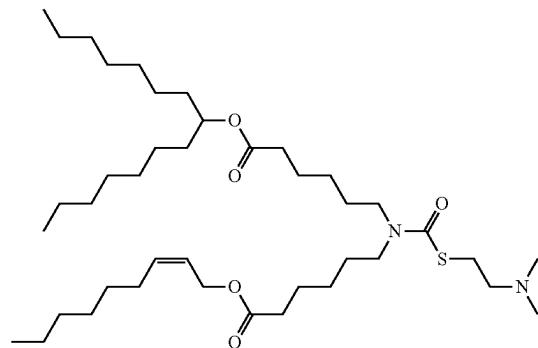
103

In one aspect, the ionizable cationic lipid of compositions provided herein has a structure of

**104**

or a pharmaceutically acceptable salt thereof.

²⁵ In one aspect, the ionizable cationic lipid included in lipid formulations of pharmaceutical compositions provided herein has a structure of



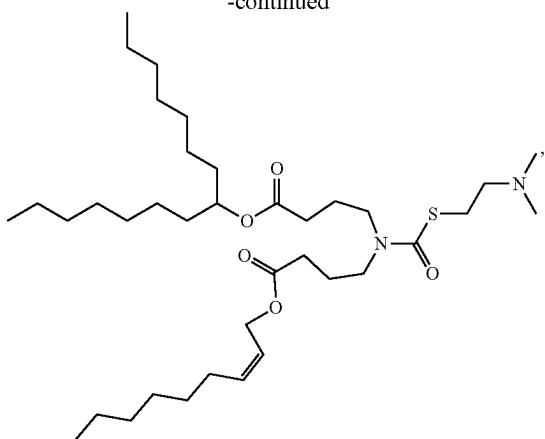
or a pharmaceutically acceptable salt thereof.

65

In another aspect, the ionizable cationic lipid of compositions provided herein has a structure of

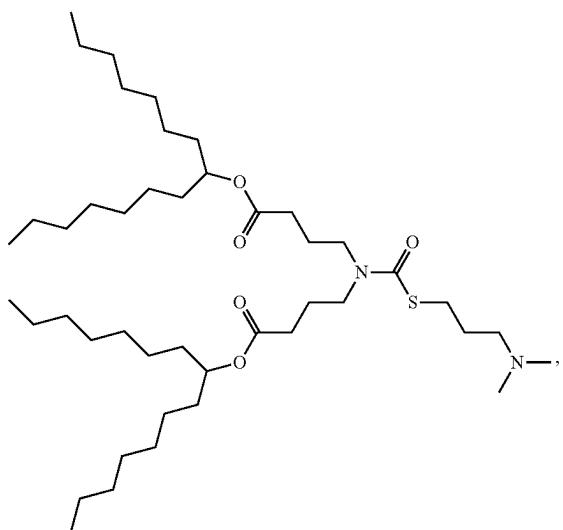
105

-continued



or a pharmaceutically acceptable salt thereof.

In another aspect, the ionizable cationic lipid included in lipid formulations of pharmaceutical compositions provided herein has a structure of



or a pharmaceutically acceptable salt thereof.

Lipid Formulations/LNPs

Therapies based on the intracellular delivery of nucleic acids to target cells face both extracellular and intracellular barriers. Indeed, naked nucleic acid materials cannot be easily systemically administered due to their toxicity, low stability in serum, rapid renal clearance, reduced uptake by target cells, phagocyte uptake and their ability in activating the immune response, all features that preclude their clinical development. When exogenous nucleic acid material (e.g., mRNA) enters the human biological system, it is recognized by the reticuloendothelial system (RES) as foreign pathogens and cleared from blood circulation before having the chance to encounter target cells within or outside the vascular system. It has been reported that the half-life of naked nucleic acid in the blood stream is around several minutes (Kawabata K, Takakura Y, Hashida M Pharm Res. 1995 June; 12(6):825-30). Chemical modification and a proper delivery method can reduce uptake by the RES and protect nucleic acids from degradation by ubiquitous nucleases,

106

which increase stability and efficacy of nucleic acid-based therapies. In addition, RNAs or DNAs are anionic hydrophilic polymers that are not favorable for uptake by cells, which are also anionic at the surface. The success of nucleic acid-based therapies thus depends largely on the development of vehicles or vectors that can efficiently and effectively deliver genetic material to target cells and obtain sufficient levels of expression in vivo with minimal toxicity.

Moreover, upon internalization into a target cell, nucleic acid delivery vectors are challenged by intracellular barriers, including endosome entrapment, lysosomal degradation, nucleic acid unpacking from vectors, translocation across the nuclear membrane (for DNA), release at the cytoplasm (for RNA), and so on. Successful nucleic acid-based therapy thus depends upon the ability of the vector to deliver the nucleic acids to the target sites inside of the cells in order to obtain sufficient levels of a desired activity such as expression of a gene.

While several gene therapies have been able to successfully utilize a viral delivery vector (e.g., AAV), lipid-based formulations have been increasingly recognized as one of the most promising delivery systems for RNA and other nucleic acid compounds due to their biocompatibility and their ease of large-scale production. One of the most significant advances in lipid-based nucleic acid therapies happened in August 2018 when Patisiran (ALN-TTR02) was the first siRNA therapeutic approved by the Food and Drug Administration (FDA) and by the European Commission (EC). ALN-TTR02 is an siRNA formulation based upon the so-called Stable Nucleic Acid Lipid Particle (SNALP) transfecting technology. Despite the success of Patisiran, the delivery of nucleic acid therapeutics, including mRNA, via lipid formulations is still under ongoing development.

Some art-recognized lipid-formulated delivery vehicles for nucleic acid therapeutics include, according to various embodiments, polymer based carriers, such as polyethyleneimine (PEI), lipid nanoparticles and liposomes, nanoliposomes, ceramide-containing nanoliposomes, multivesicular liposomes, proteoliposomes, both natural and synthetically-derived exosomes, natural, synthetic and semi-synthetic lamellar bodies, nanoparticulates, micelles, and emulsions. These lipid formulations can vary in their structure and composition, and as can be expected in a rapidly evolving field, several different terms have been used in the art to describe a single type of delivery vehicle. At the same time, the terms for lipid formulations have varied as to their intended meaning throughout the scientific literature, and this inconsistent use has caused confusion as to the exact meaning of several terms for lipid formulations. Among the several potential lipid formulations, liposomes, cationic liposomes, and lipid nanoparticles are specifically described in detail and defined herein for the purposes of the present disclosure.

Liposomes

Conventional liposomes are vesicles that consist of at least one bilayer and an internal aqueous compartment. Bilayer membranes of liposomes are typically formed by amphiphilic molecules, such as lipids of synthetic or natural origin that comprise spatially separated hydrophilic and hydrophobic domains (Lasic, Trends Biotechnol., 16: 307-321, 1998). Bilayer membranes of the liposomes can also be formed by amphiphilic polymers and surfactants (e.g., polymerosomes, niosomes, etc.). They generally present as spherical vesicles and can range in size from 20 nm to a few microns. Liposomal formulations can be prepared as a colloidal dispersion or they can be lyophilized to reduce stability risks and to improve the shelf-life for liposome-

107

based drugs. Methods of preparing liposomal compositions are known in the art and would be within the skill of an ordinary artisan.

Liposomes that have only one bilayer are referred to as being unilamellar, and those having more than one bilayer are referred to as multilamellar. The most common types of liposomes are small unilamellar vesicles (SUV), large unilamellar vesicle (LUV), and multilamellar vesicles (MLV). In contrast to liposomes, lysosomes, micelles, and reversed micelles are composed of monolayers of lipids. Generally, a liposome is thought of as having a single interior compartment, however some formulations can be multivesicular liposomes (MVL), which consist of numerous discontinuous internal aqueous compartments separated by several non-concentric lipid bilayers.

Liposomes have long been perceived as drug delivery vehicles because of their superior biocompatibility, given that liposomes are basically analogs of biological membranes, and can be prepared from both natural and synthetic phospholipids (Int J Nanomedicine. 2014; 9:1833-1843). In their use as drug delivery vehicles, because a liposome has an aqueous solution core surrounded by a hydrophobic membrane, hydrophilic solutes dissolved in the core cannot readily pass through the bilayer, and hydrophobic compounds will associate with the bilayer. Thus, a liposome can be loaded with hydrophobic and/or hydrophilic molecules. When a liposome is used to carry a nucleic acid such as RNA, the nucleic acid will be contained within the liposomal compartment in an aqueous phase.

Cationic Liposomes

Liposomes can be composed of cationic, anionic, and/or neutral lipids. As an important subclass of liposomes, cationic liposomes are liposomes that are made in whole or part from positively charged lipids, or more specifically a lipid that comprises both a cationic group and a lipophilic portion. In addition to the general characteristics profiled above for liposomes, the positively charged moieties of cationic lipids used in cationic liposomes provide several advantages and some unique structural features. For example, the lipophilic portion of the cationic lipid is hydrophobic and thus will direct itself away from the aqueous interior of the liposome and associate with other nonpolar and hydrophobic species. Conversely, the cationic moiety will associate with aqueous media and more importantly with polar molecules and species with which it can complex in the aqueous interior of the cationic liposome. For these reasons, cationic liposomes are increasingly being researched for use in gene therapy due to their favorability towards negatively charged nucleic acids via electrostatic interactions, resulting in complexes that offer biocompatibility, low toxicity, and the possibility of the large-scale production required for *in vivo* clinical applications. Cationic lipids suitable for use in cationic liposomes are listed herein below.

Lipid Nanoparticles

In contrast to liposomes and cationic liposomes, lipid nanoparticles (LNP) have a structure that includes a single monolayer or bilayer of lipids that encapsulates a compound in a solid phase. Thus, unlike liposomes, lipid nanoparticles do not have an aqueous phase or other liquid phase in its interior, but rather the lipids from the bilayer or monolayer shell are directly complexed to the internal compound thereby encapsulating it in a solid core. Lipid nanoparticles are typically spherical vesicles having a relatively uniform dispersion of shape and size. While sources vary on what size qualifies a lipid particle as being a nanoparticle, there is some overlap in agreement that a lipid nanoparticle can have

108

a diameter in the range of from 10 nm to 1000 nm. However, more commonly they are considered to be smaller than 120 nm or even 100 nm.

For lipid nanoparticle nucleic acid delivery systems, the lipid shell is formulated to include an ionizable cationic lipid which can complex to and associate with the negatively charged backbone of the nucleic acid core. Ionizable cationic lipids with apparent pKa values below about 7 have the benefit of providing a cationic lipid for complexing with the nucleic acid's negatively charged backbone and loading into the lipid nanoparticle at pH values below the pKa of the ionizable lipid where it is positively charged. Then, at physiological pH values, the lipid nanoparticle can adopt a relatively neutral exterior allowing for a significant increase in the circulation half-lives of the particles following i.v. administration. In the context of nucleic acid delivery, lipid nanoparticles offer many advantages over other lipid-based nucleic acid delivery systems including high nucleic acid encapsulation efficiency, potent transfection, improved penetration into tissues to deliver therapeutics, and low levels of cytotoxicity and immunogenicity.

Prior to the development of lipid nanoparticle delivery systems for nucleic acids, cationic lipids were widely studied as synthetic materials for delivery of nucleic acid medicines. In these early efforts, after mixing together at physiological pH, nucleic acids were condensed by cationic lipids to form lipid-nucleic acid complexes known as lipoplexes. However, lipoplexes proved to be unstable and characterized by broad size distributions ranging from the submicron scale to a few microns. Lipoplexes, such as the Lipofectamine® reagent, have found considerable utility for *in vitro* transfection. However, these first-generation lipoplexes have not proven useful *in vivo*. The large particle size and positive charge (imparted by the cationic lipid) result in rapid plasma clearance, hemolytic and other toxicities, as well as immune system activation.

In some aspects, nucleic acid molecules provided herein and lipids or lipid formulations provided herein form a lipid nanoparticle (LNP).

In other aspects, nucleic acid molecules provided herein are incorporated into a lipid formulation (i.e., a lipid-based delivery vehicle).

In the context of the present disclosure, a lipid-based delivery vehicle typically serves to transport a desired RNA to a target cell or tissue. The lipid-based delivery vehicle can be any suitable lipid-based delivery vehicle known in the art. In some aspects, the lipid-based delivery vehicle is a liposome, a cationic liposome, or a lipid nanoparticle containing a self-replicating RNA of the disclosure. In some aspects, the lipid-based delivery vehicle comprises a nanoparticle or a bilayer of lipid molecules and a self-replicating RNA of the disclosure. In some aspects, the lipid bilayer further comprises a neutral lipid or a polymer. In some aspects, the lipid formulation comprises a liquid medium. In some aspects, the formulation further encapsulates a nucleic acid. In some aspects, the lipid formulation further comprises a nucleic acid and a neutral lipid or a polymer. In some aspects, the lipid formulation encapsulates the nucleic acid.

The description provides lipid formulations comprising one or more self-replicating RNA molecules encapsulated within the lipid formulation. In some aspects, the lipid formulation comprises liposomes. In some aspects, the lipid formulation comprises cationic liposomes. In some aspects, the lipid formulation comprises lipid nanoparticles.

In some aspects, the self-replicating RNA is fully encapsulated within the lipid portion of the lipid formulation such that the RNA in the lipid formulation is resistant in aqueous

109

solution to nuclease degradation. In other aspects, the lipid formulations described herein are substantially non-toxic to animals such as humans and other mammals.

The lipid formulations of the disclosure also typically have a total lipid:RNA ratio (mass/mass ratio) of from about 1:1 to about 100:1, from about 1:1 to about 50:1, from about 2:1 to about 45:1, from about 3:1 to about 40:1, from about 5:1 to about 45:1, or from about 10:1 to about 40:1, or from about 15:1 to about 40:1, or from about 20:1 to about 40:1; or from about 25:1 to about 45:1; or from about 30:1 to about 45:1; or from about 32:1 to about 42:1; or from about 34:1 to about 42:1. In some aspects, the total lipid:RNA ratio (mass/mass ratio) is from about 30:1 to about 45:1. The ratio may be any value or subvalue within the recited ranges, including endpoints.

The lipid formulations of the present disclosure typically have a mean diameter of from about 30 nm to about 150 nm, from about 40 nm to about 150 nm, from about 50 nm to about 150 nm, from about 60 nm to about 130 nm, from about 70 nm to about 110 nm, from about 70 nm to about 100 nm, from about 80 nm to about 100 nm, from about 90 nm to about 100 nm, from about 70 to about 90 nm, from about 80 nm to about 90 nm, from about 70 nm to about 80 nm, or about 30 nm, about 35 nm, about 40 nm, about 45 nm, about 50 nm, about 55 nm, about 60 nm, about 65 nm, about 70 nm, about 75 nm, about 80 nm, about 85 nm, about 90 nm, about 95 nm, about 100 nm, about 105 nm, about 110 nm, about 115 nm, about 120 nm, about 125 nm, about 130 nm, about 135 nm, about 140 nm, about 145 nm, or about 150 nm, and are substantially non-toxic. The diameter may be any value or subvalue within the recited ranges, including endpoints. In addition, nucleic acids, when present in the lipid nanoparticles of the present disclosure, generally are resistant in aqueous solution to degradation with a nuclease.

In some aspects, the lipid formulations comprise a self-replicating RNA, a cationic lipid (e.g., one or more cationic lipids or salts thereof described herein), a phospholipid, and a conjugated lipid that inhibits aggregation of the particles (e.g., one or more PEG-lipid conjugates). The lipid formulations can also include cholesterol. In one aspect, the cationic lipid is an ionizable cationic lipid.

In the nucleic acid-lipid formulations, the RNA may be fully encapsulated within the lipid portion of the formulation, thereby protecting the nucleic acid from nuclease degradation. In some aspects, a lipid formulation comprising an RNA is fully encapsulated within the lipid portion of the lipid formulation, thereby protecting the nucleic acid from nuclease degradation. In certain aspects, the RNA in the lipid formulation is not substantially degraded after exposure of the particle to a nuclease at 37° C. for at least 20, 30, 45, or 60 minutes. In certain other aspects, the RNA in the lipid formulation is not substantially degraded after incubation of the formulation in serum at 37° C. for at least 30, 45, or 60 minutes or at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, or 36 hours. In some aspects, the RNA is complexed with the lipid portion of the formulation. One of the benefits of the formulations of the present disclosure is that the nucleic acid-lipid compositions are substantially non-toxic to animals such as humans and other mammals.

In the context of nucleic acids, full encapsulation may be determined by performing a membrane-impermeable fluorescent dye exclusion assay, which uses a dye that has enhanced fluorescence when associated with nucleic acid. Encapsulation is determined by adding the dye to a lipid formulation, measuring the resulting fluorescence, and comparing it to the fluorescence observed upon addition of a

110

small amount of nonionic detergent. Detergent-mediated disruption of the lipid layer releases the encapsulated nucleic acid, allowing it to interact with the membrane-impermeable dye. Nucleic acid encapsulation may be calculated as $E = (I_0 - I)/I_0$, where/and I_0 refers to the fluorescence intensities before and after the addition of detergent.

In some aspects, the present disclosure provides a nucleic acid-lipid composition comprising a plurality of nucleic acid-liposomes, nucleic acid-cationic liposomes, or nucleic acid-lipid nanoparticles. In some aspects, the nucleic acid-lipid composition comprises a plurality of RNA-liposomes. In some aspects, the nucleic acid-lipid composition comprises a plurality of RNA-cationic liposomes. In some aspects, the nucleic acid-lipid composition comprises a plurality of RNA-lipid nanoparticles.

In some aspects, the lipid formulations comprise RNA that is fully encapsulated within the lipid portion of the formulation, such that from about 30% to about 100%, from about 40% to about 100%, from about 50% to about 100%, from about 60% to about 100%, from about 70% to about 100%, from about 80% to about 100%, from about 90% to about 100%, from about 30% to about 95%, from about 40% to about 95%, from about 50% to about 95%, from about 60% to about 95%, from about 70% to about 95%, from about 80% to about 95%, from about 85% to about 95%, from about 90% to about 95%, from about 30% to about 90%, from about 40% to about 90%, from about 50% to about 90%, from about 60% to about 90%, from about 70% to about 90%, from about 80% to about 90%, or at least about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, or about 99% (or any fraction thereof or range therein) of the particles have the RNA encapsulated therein. The amount may be any value or subvalue within the recited ranges, including endpoints. The RNA included in any RNA-lipid composition or RNA-lipid formulation provided herein can be a self-replicating RNA.

Depending on the intended use of the lipid formulation, the proportions of the components can be varied, and the delivery efficiency of a particular formulation can be measured using assays known in the art.

In some aspects, nucleic acid molecules provided herein are lipid formulated. The lipid formulation is preferably selected from, but not limited to, liposomes, cationic liposomes, and lipid nanoparticles. In one aspect, a lipid formulation is a cationic liposome or a lipid nanoparticle (LNP) comprising:

- (a) an RNA of the present disclosure,
- (b) a cationic lipid,
- (c) an aggregation reducing agent (such as polyethylene glycol (PEG) lipid or PEG-modified lipid),
- (d) optionally a non-cationic lipid (such as a neutral lipid), and
- (e) optionally, a sterol.

In another aspect, the cationic lipid is an ionizable cationic lipid. Any ionizable cationic lipid can be included in lipid formulations, including exemplary cationic lipids provided herein.

Cationic Lipids

In one aspect, the lipid nanoparticle formulation comprises (i) at least one cationic lipid; (ii) a helper lipid; (iii) a sterol (e.g., cholesterol); and (iv) a PEG-lipid. In another aspect, the cationic lipid is an ionizable cationic lipid. In yet another aspect, the lipid nanoparticle formulation comprises (i) at least one cationic lipid; (ii) a helper lipid; (iii) a sterol

111

(e.g., cholesterol); and (iv) a PEG-lipid, in a molar ratio of about 40-70% ionizable cationic lipid:about 2-15% helper lipid:about 20-45% sterol; about 0.5-5% PEG-lipid. In a further aspect, the cationic lipid is an ionizable cationic lipid.

In one aspect, the lipid nanoparticle formulation consists of (i) at least one cationic lipid; (ii) a helper lipid; (iii) a sterol (e.g., cholesterol); and (iv) a PEG-lipid. In another aspect, the cationic lipid is an ionizable cationic lipid. In yet another aspect, the lipid nanoparticle formulation consists of (i) at least one cationic lipid; (ii) a helper lipid; (iii) a sterol (e.g., cholesterol); and (iv) a PEG-lipid, in a molar ratio of about 40-70% ionizable cationic lipid:about 2-15% helper lipid:about 20-45% sterol; about 0.5-5% PEG-lipid. In a further aspect, the cationic lipid is an ionizable cationic lipid.

In the presently disclosed lipid formulations, the cationic lipid may be, for example, N,N-dioleyl-N,N-dimethylammonium chloride (DODAC), N,N-distearyl-N,N-dimethylammonium bromide (DDAB), 1,2-dioleoyltrimethylammoniumpropane chloride (DOTAP) (also known as N-(2,3-dioleyloxy)propyl)-N,N,N-trimethylammonium chloride and 1,2-Dioleyloxy-3-trimethylaminopropane chloride salt), N-(1-(2,3-dioleyloxy)propyl)-N,N,N-trimethylammonium chloride (DOTMA), N,N-dimethyl-2,3-dioleyloxy)propylamine (DODMA), 1,2-Dilinoleyoxy-N,N-dimethylaminopropane (DLinDMA), 1,2-Dilinolenyloxy-N,N-dimethylaminopropane (DLenDMA), 1,2-di-y-linolenyloxy-N,N-dimethylaminopropane (γ -DLenDMA), 1,2-Dilinoleylcarbamoyloxy-3-dimethylaminopropane (DLin-C-DAP), 1,2-Dilinoleyoxy-3-(dimethylamino)acetoxypropane (DLin-DAC), 1,2-Dilinoleyoxy-3-morpholinopropane (DLin-MA), 1,2-Dilinoleoyl-3-dimethylaminopropane (DLinDAP), 1,2-Dilinoleylthio-3-dimethylaminopropane (DLin-S-DMA), 1-Linoleoyl-2-linoleyoxy-3-dimethylaminopropane (DLin-2-DMAP), 1,2-Dilinoleyoxy-3-trimethylaminopropane chloride salt (DLin-TMA-Cl), 1,2-Dilinoleyl-3-trimethylaminopropane chloride salt (DLin-TAP-Cl), 1,2-Dilinoleyoxy-3-(N-methylpiperazino)propane (DLin-MPZ), or 3-(N,N-Dilinoleylamino)-1,2-propanediol (DLinAP), 3-(N,N-Dioleylamino)-1,2-propanediol (DOAP), 1,2-Dilinoleyoxyo-3-(2-N,N-dimethylamino)ethoxypropane (DLin-EG-DMA), 2,2-Dilinoleyl-4-dimethylaminomethyl-[1,3]-dioxolane (DLin-K-DMA) or analogs thereof, (3aR,5s,6aS)—N,N-dimethyl-2,2-di((9Z,12Z)-octadeca-9,12-dienyl)tetrahydro-3aH-cyclopenta[d][1,3]dioxol-5-amine, (6Z,9Z,28Z,31Z)-heptatriaconta-6,9,28,31-tetraen-19-yl4-(dimethylamino)butanoate (MC3), 1,1'-(2-(4-(2-((2-(bis(2-hydroxydodecyl)amino)ethyl)(2-hydroxydodecyl)amino)ethyl)piperazin-1-yl)ethylazanediyI)diodecan-2-ol (C12-200), 2,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLin-K-C2-DMA), 2,2-dilinoleyl-4-dimethylaminomethyl-[1,3]-dioxolane (DLin-K-DMA), (6Z,9Z,28Z,31Z)-heptatriaconta-6,9,28,31-tetraen-19-yl 4-(dimethylamino)butanoate (DLin-M-C3-DMA), 3-((6Z,9Z,28Z,31Z)-heptatriaconta-6,9,28,31-tetraen-19-yloxy)-N,N-dimethylpropan-1-amine (MC3 Ether), 4-((6Z,9Z,28Z,31 Z)-heptatriaconta-6,9,28,31-tetraen-19-yloxy)-N,N-dimethylbutan-1-amine (MC4 Ether), or any combination thereof. Other cationic lipids include, but are not limited to, N,N-distearyl-N,N-dimethylammonium bromide (DDAB), 3P—(N—(N',N'-dimethylaminoethane)-carbamoyl)cholesterol (DC-Choi), N-(1-(2,3-dioleyloxy)propyl)-N-2-(sperminecarboxamido)ethyl)-N,N-dimethylammonium trifluoracetate (DOSPA), dioctadecylamidoglycyl carboxyspermine (DOGS), 1,2-dileoyl-sn-3-phosphoethanolamine (DOPE), 1,2-dioleoyl-3-

112

dimethylammonium propane (DODAP), N-(1,2-dimysteryloxyprop-3-yl)-N,N-dimethyl-N-hydroxyethyl ammonium bromide (DMRIE), and 2,2-Dilinoleyl-4-dimethylaminomethyl-11,31-dioxolane (XTC). Additionally, commercial preparations of cationic lipids can be used, such as, e.g., LIPOFFECTIN (including DOTMA and DOPE, available from GIBCO/BRL), and Lipofectamine (comprising DOSPA and DOPE, available from GIBCO/BRL).

Other suitable cationic lipids are disclosed in International Publication Nos. WO 09/086558, WO 09/127060, WO 10/048536, WO 10/054406, WO 10/088537, WO 10/129709, and WO 2011/153493; U.S. Patent Publication Nos. 2011/0256175, 2012/0128760, and 2012/0027803; U.S. Pat. No. 8,158,601; and Love et al., PNAS, 107(5), 1864-69, 2010, the contents of which are herein incorporated by reference.

The RNA-lipid formulations of the present disclosure can comprise a helper lipid, which can be referred to as a neutral helper lipid, non-cationic lipid, non-cationic helper lipid, anionic lipid, anionic helper lipid, or a neutral lipid. It has been found that lipid formulations, particularly cationic liposomes and lipid nanoparticles have increased cellular uptake if helper lipids are present in the formulation. (Curr. Drug Metab. 2014; 15(9):882-92). For example, some studies have indicated that neutral and zwitterionic lipids such as 1,2-dioleyl-sn-glycero-3-phosphatidylcholine (DOPC), Di-Oleoyl-Phosphatidyl-Ethanoalamine (DOPE) and 1,2-Di-Stearoyl-sn-glycero-3-PhosphoCholine (DSPC), being more fusogenic (i.e., facilitating fusion) than cationic lipids, can affect the polymorphic features of lipid-nucleic acid complexes, promoting the transition from a lamellar to a hexagonal phase, and thus inducing fusion and a disruption of the cellular membrane. (Nanomedicine (Lond). 2014 January; 9(1):105-20). In addition, the use of helper lipids can help to reduce any potential detrimental effects from using many prevalent cationic lipids such as toxicity and immunogenicity.

Non-limiting examples of non-cationic lipids suitable for lipid formulations of the present disclosure include phospholipids such as lecithin, phosphatidylethanolamine, lysolecithin, lysophosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, sphingomyelin, egg sphingomyelin (ESM), cephalin, cardiolipin, phosphatidic acid, cerebrosides, dicetylphosphate, distearoylphosphatidylcholine (DOPC), dioleoylphosphatidylcholine (DOPC), dipalmitoylphosphatidylcholine (DPPC), dioleoylphosphatidylglycerol (DOPG), dipalmitoylphosphatidylglycerol (DPPG), dioleoylphosphatidylethanolamine (DOPE), palmitoyloleoyl-phosphatidylcholine (POPC), palmitoyloleoyl-phosphatidylethanolamine (POPE), palmitoyloleoyl-phosphatidylglycerol (POPG), dioleoylphosphatidylethanolamine 4-(N-maleimidomethyl)-cyclohexane-1-carboxylate (DOPE-mal), dipalmitoyl-phosphatidylethanolamine (DPPE), dimyristoyl-phosphatidylethanolamine (DMPE), distearoyl-phosphatidylethanolamine (DSPE), monomethyl-phosphatidylethanolamine, dimethyl-phosphatidylethanolamine, dielaidoyl-phosphatidylethanolamine (DEPE), stearoyloleoyl-phosphatidylethanolamine (SOPE), lyso-phosphatidylcholine, dilinoleoylphosphatidylcholine, and mixtures thereof. Other diacylphosphatidylcholine and diacylphosphatidylethanolamine phospholipids can also be used. The acyl groups in these lipids are preferably acyl groups derived from fatty acids having C10-C24 carbon chains, e.g., lauroyl, myristoyl, palmitoyl, stearoyl, or oleoyl.

Additional examples of non-cationic lipids include sterols such as cholesterol and derivatives thereof. As a helper lipid,

113

cholesterol increases the spacing of the charges of the lipid layer interfacing with the nucleic acid making the charge distribution match that of the nucleic acid more closely. (J. R. Soc. Interface. 2012 Mar. 7; 9(68): 548-561). Non-limiting examples of cholesterol derivatives include polar analogues such as 5 α -cholestanol, 5 α -coprostanol, cholestryl-(2'-hydroxy)-ethyl ether, cholestryl-(4'-hydroxy)-butyl ether, and 6-ketocholestanol; non-polar analogues such as 5 α -cholestane, cholestenone, 5 α -cholestaneone, 5 α -cholestaneone, and cholestryl decanoate; and mixtures thereof. In some aspects, the cholesterol derivative is a polar analogue such as cholestryl-(4'-hydroxy)-butyl ether.

In some aspects, the helper lipid present in the lipid formulation comprises or consists of a mixture of one or more phospholipids and cholesterol or a derivative thereof. In other aspects, the neutral lipid present in the lipid formulation comprises or consists of one or more phospholipids, e.g., a cholesterol-free lipid formulation. In yet other aspects, the neutral lipid present in the lipid formulation comprises or consists of cholesterol or a derivative thereof, e.g., a phospholipid-free lipid formulation.

Other examples of helper lipids include nonphosphorous containing lipids such as, e.g., stearylamine, dodecylamine, hexadecylamine, acetyl palmitate, glycerol ricinoleate, hexadecyl stearate, isopropyl myristate, amphoteric acrylic polymers, triethanolamine-lauryl sulfate, alkyl-aryl sulfate polyethoxylated fatty acid amides, dioctadecyldimethyl ammonium bromide, ceramide, and sphingomyelin.

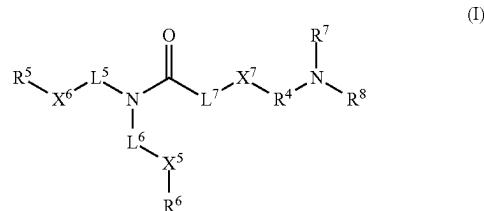
Other suitable cationic lipids include those having alternative fatty acid groups and other dialkylamino groups, including those, in which the alkyl substituents are different (e.g., N-ethyl-N-methylamino-, and N-propyl-N-ethyl-amino-). These lipids are part of a subcategory of cationic lipids referred to as amino lipids. In some embodiments of the lipid formulations described herein, the cationic lipid is an amino lipid. In general, amino lipids having less saturated acyl chains are more easily sized, particularly when the complexes must be sized below about 0.3 microns, for purposes of filter sterilization. Amino lipids containing unsaturated fatty acids with carbon chain lengths in the range of C14 to C22 may be used. Other scaffolds can also be used to separate the amino group and the fatty acid or fatty alkyl portion of the amino lipid.

In some embodiments, the lipid formulation comprises the cationic lipid with Formula I according to the patent application PCT/EP2017/064066. In this context, the disclosure of PCT/EP2017/064066 is also incorporated herein by reference.

In some embodiments, amino or cationic lipids of the present disclosure are ionizable and have at least one protonatable or deprotonatable group, such that the lipid is positively charged at a pH at or below physiological pH (e.g., pH 7.4), and neutral at a second pH, preferably at or above physiological pH. Of course, it will be understood that the addition or removal of protons as a function of pH is an equilibrium process, and that the reference to a charged or a neutral lipid refers to the nature of the predominant species and does not require that all of the lipid be present in the charged or neutral form. Lipids that have more than one protonatable or deprotonatable group, or which are zwitterionic, are not excluded from use in the disclosure. In certain embodiments, the protonatable lipids have a pKa of the protonatable group in the range of about 4 to about 11. In some embodiments, the ionizable cationic lipid has a pKa of about 5 to about 7. In some embodiments, the pKa of an ionizable cationic lipid is about 6 to about 7.

114

In some embodiments, the lipid formulation comprises an ionizable cationic lipid of Formula I:



or a pharmaceutically acceptable salt or solvate thereof, wherein R5 and R6 are each independently selected from the group consisting of a linear or branched C1-C31 alkyl, C2-C31 alkenyl or C2-C31 alkynyl and cholestryl; L5 and L6 are each independently selected from the group consisting of a linear C1-C20 alkyl and C2-C20 alkenyl; X5 is —C(O)O—, whereby —C(O)O—R6 is formed or —OC(O)— whereby —OC(O)O—R6 is formed; X6 is —C(O)O— whereby —C(O)O—R5 is formed or —OC(O)— whereby —OC(O)O—R5 is formed; X7 is S or O; L7 is absent or lower alkyl; R4 is a linear or branched C1-C6 alkyl; and R7 and R8 are each independently selected from the group consisting of a hydrogen and a linear or branched C1-C6 alkyl.

In some embodiments, X7 is S.

In some embodiments, X5 is —C(O)O—, whereby —C(O)O—R6 is formed and X6 is —C(O)O— whereby —C(O)O—R5 is formed.

In some embodiments, R7 and R8 are each independently selected from the group consisting of methyl, ethyl and isopropyl.

In some embodiments, L5 and L6 are each independently a C1-C10 alkyl. In some embodiments, L5 is C1-C3 alkyl, and L6 is C1-C5 alkyl. In some embodiments, L6 is C1-C2 alkyl. In some embodiments, L5 and L6 are each a linear C7 alkyl. In some embodiments, L5 and L6 are each a linear C9 alkyl.

In some embodiments, R5 and R6 are each independently an alkenyl. In some embodiments, R6 is alkenyl. In some embodiments, R6 is C2-C9 alkenyl. In some embodiments, the alkenyl comprises a single double bond. In some embodiments, R5 and R6 are each alkyl. In some embodiments, R5 is a branched alkyl. In some embodiments, R5 and R6 are each independently selected from the group consisting of a C9 alkyl, C9 alkenyl and C9 alkynyl. In some embodiments, R5 and R6 are each independently selected from the group consisting of a C11 alkyl, C11 alkenyl and C11 alkynyl. In some embodiments, R5 and R6 are each independently selected from the group consisting of a C7 alkyl, C7 alkenyl and C7 alkynyl. In some embodiments, R5 is —CH((CH₂)_pCH₃)₂ or —CH((CH₂)_pCH₃)((CH₂)_{p-1}CH₃), wherein p is 4-8. In some embodiments, p is 5 and L5 is a C1-C3 alkyl. In some embodiments, p is 6 and L5 is a C3 alkyl. In some embodiments, p is 7. In some embodiments, p is 8 and L5 is a C1-C3 alkyl. In some embodiments, R5 consists of —CH((CH₂)_pCH₃)((CH₂)_{p-1}CH₃), wherein p is 7 or 8.

In some embodiments, R4 is ethylene or propylene. In some embodiments, R4 is n-propylene or isobutylene.

In some embodiments, L7 is absent, R4 is ethylene, X7 is S and R7 and R8 are each methyl. In some embodiments, L7 is absent, R4 is n-propylene, X7 is S and R7 and R8 are each

115

methyl. In some embodiments, L7 is absent, R4 is ethylene, X7 is S and R7 and R8 are each ethyl.

In some embodiments, X7 is S, X5 is —C(O)O—, whereby —C(O)O—R6 is formed, X6 is —C(O)O— whereby —C(O)O—R5 is formed, L5 and L6 are each independently a linear C3-C7 alkyl, L7 is absent, R5 is

116

—CH((CH₂)_pCH₃)₂, and R6 is C7-C12 alkenyl. In some further embodiments, p is 6 and R6 is C9 alkenyl.

In some embodiments, the lipid formulation can comprise an ionizable cationic lipid selected from the group consisting of LIPID #1 to LIPID #8:

TABLE 5

LIPID #	STRUCTURE
1	
2	
3	

TABLE 5-continued

LIPID #	STRUCTURE
4	<p>Chemical structure of Lipid 4: A glyceride derivative. It features a long-chain fatty acid (stearic acid derivative) esterified to the 1-position of a glycerol backbone. The 2-position is esterified to a medium-chain fatty acid (oleic acid derivative). The 3-position is esterified to a trimethylammonium group ($\text{N}(\text{CH}_3)_3^+$). The structure shows the characteristic ester linkages (-O-C(=O)-) and the quaternary ammonium nitrogen atom.</p>
5	<p>Chemical structure of Lipid 5: A glyceride derivative. It features a long-chain fatty acid (stearic acid derivative) esterified to the 1-position of a glycerol backbone. The 2-position is esterified to a medium-chain fatty acid (oleic acid derivative). The 3-position is esterified to a diethylaminoethyl group ($\text{N}(\text{C}_2\text{H}_5)_2\text{CH}_2^-$). The structure shows the characteristic ester linkages and the diethylaminoethyl side chain.</p>
6	<p>Chemical structure of Lipid 6: A glyceride derivative. It features two medium-chain fatty acids (oleic acid derivatives) esterified to the 1 and 2 positions of a glycerol backbone. The 3-position is esterified to a diethylaminoethyl group ($\text{N}(\text{C}_2\text{H}_5)_2\text{CH}_2^-$). The structure shows the characteristic ester linkages and the diethylaminoethyl side chain.</p>

119

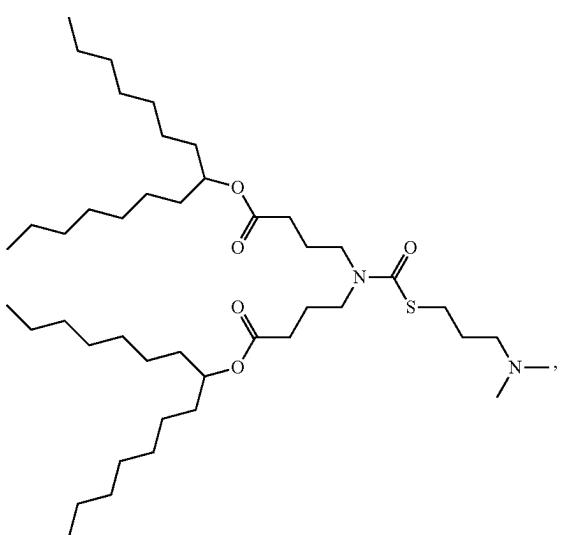
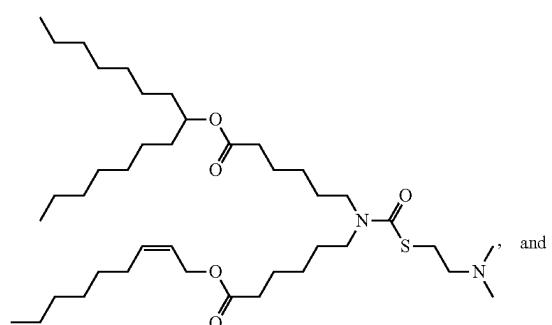
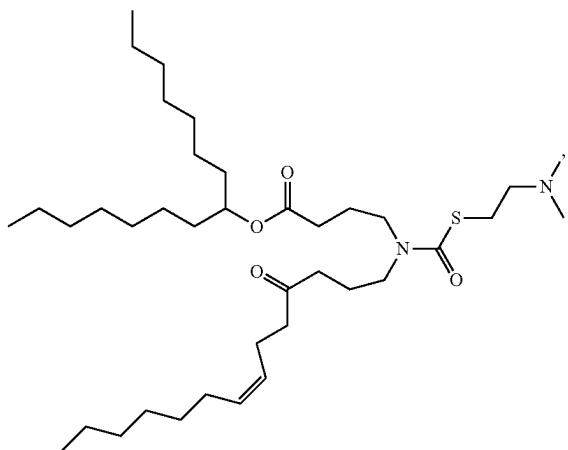
TABLE 5-continued

120

LIPID #	STRUCTURE
7	<p>The chemical structure of Lipid 7 is shown. It features two long, saturated fatty acid chains. The top chain is a straight chain of 16 carbons, ending in a 3-hydroxypropyl group (-CH(O)CH₂CH₃). This chain is esterified to the 2-position of a glycerol backbone. The glycerol backbone is further esterified at its 1-position to another fatty acid chain (16 carbons) and at its 3-position to a trimethylammonium cation (N+(CH₃)₃). The bottom chain is a branched chain of 16 carbons, ending in a 3-hydroxypropyl group (-CH(O)CH₂CH₃). This chain is also esterified to the 2-position of the glycerol backbone.</p>
8	<p>The chemical structure of Lipid 8 is shown. It features two long, saturated fatty acid chains. The top chain is a straight chain of 16 carbons, ending in a 3-hydroxypropyl group (-CH(O)CH₂CH₃). This chain is esterified to the 2-position of a glycerol backbone. The glycerol backbone is further esterified at its 1-position to another fatty acid chain (16 carbons) and at its 3-position to a diethylammonium cation (N+(CH₂CH₃)₂). The bottom chain is a branched chain of 16 carbons, ending in a 3-hydroxypropyl group (-CH(O)CH₂CH₃). This chain is esterified to the 2-position of the glycerol backbone.</p>

121

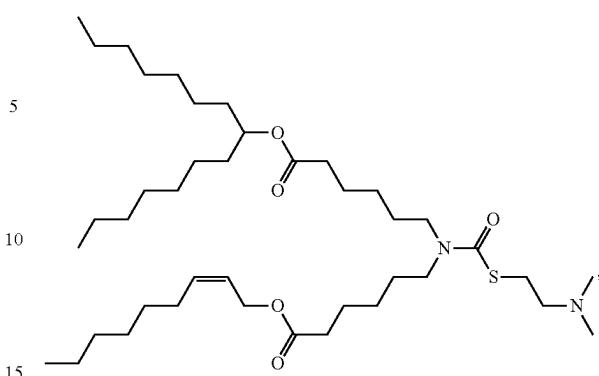
In some embodiments, the lipid formulation comprises an ionizable cationic lipid having a structure selected from



or a pharmaceutically acceptable salt thereof.

In some preferred embodiments, the lipid formulation comprises an ionizable cationic lipid having the structure

122



or a pharmaceutically acceptable salt thereof.

In embodiments, any one or more lipids recited herein may be expressly excluded.

20 In some aspects, the helper lipid comprises from about 2 mol % to about 20 mol %, from about 3 mol % to about 18 mol %, from about 4 mol % to about 16 mol %, about 5 mol % to about 14 mol %, from about 6 mol % to about 12 mol %, from about 5 mol % to about 10 mol %, from about 5 mol % to about 9 mol %, or about 2 mol %, about 3 mol %, about 4 mol %, about 5 mol %, about 6 mol %, about 7 mol %, about 8 mol %, about 9 mol %, about 10 mol %, about 11 mol %, or about 12 mol % (or any fraction thereof or the range therein) of the total lipid present in the lipid formulation.

The cholesterol or cholesterol derivative in the lipid formulation may comprise up to about 40 mol %, about 45 mol %, about 50 mol %, about 55 mol %, or about 60 mol % of the total lipid present in the lipid formulation. In some aspects, the cholesterol or cholesterol derivative comprises about 15 mol % to about 45 mol %, about 20 mol % to about 40 mol %, about 25 mol % to about 35 mol %, or about 28 mol % to about 35 mol %; or about 25 mol %, about 26 mol %, about 27 mol %, about 28 mol %, about 29 mol %, about 30 mol %, about 31 mol %, about 32 mol %, about 33 mol %, about 34 mol %, about 35 mol %, about 36 mol %, or about 37 mol % of the total lipid present in the lipid formulation.

45 In some aspects, the phospholipid component in the mixture may comprise from about 2 mol % to about 20 mol %, from about 3 mol % to about 18 mol %, from about 4 mol % to about 16 mol %, about 5 mol % to about 14 mol %, from about 6 mol % to about 12 mol %, from about 5 mol % to about 10 mol %, from about 5 mol % to about 9 mol %, or about 2 mol %, about 3 mol %, about 4 mol %, about 5 mol %, about 6 mol %, about 7 mol %, about 8 mol %, about 9 mol %, about 10 mol %, about 11 mol %, or about 12 mol % (or any fraction thereof or the range therein) of the total lipid present in the lipid formulation.

The percentage of helper lipid present in the lipid formulation is a target amount, and the actual amount of helper lipid present in the formulation may vary, for example, by ± 5 mol %.

60 A lipid formulation that includes a cationic lipid compound or ionizable cationic lipid compound may be on a molar basis about 30-70% cationic lipid compound, about 25-40% cholesterol, about 2-15% helper lipid, and about 0.5-5% of a polyethylene glycol (PEG) lipid, wherein the 65 percent is of the total lipid present in the formulation. In some aspects, the composition is about 40-65% cationic lipid compound, about 25-35% cholesterol, about 3-9%

helper lipid, and about 0.5-3% of a PEG-lipid, wherein the percent is of the total lipid present in the formulation.

The formulation may be a lipid particle formulation, for example containing 8-30% nucleic acid compound, 5-30% helper lipid, and 0-20% cholesterol; 4-25% cationic lipid, 4-25% helper lipid, 2-25% cholesterol, 10-35% cholesterol-PEG, and 5% cholesterol-amine; or 2-30% cationic lipid, 2-30% helper lipid, 1-15% cholesterol, 2-35% cholesterol-PEG, and 1-20% cholesterol-amine; or up to 90% cationic lipid and 2-10% helper lipids, or even 100% cationic lipid.

Lipid Conjugates

The lipid formulations described herein may further comprise a lipid conjugate. The conjugated lipid is useful in that it prevents the aggregation of particles. Suitable conjugated lipids include, but are not limited to, PEG-lipid conjugates, cationic-polymer-lipid conjugates, and mixtures thereof. Furthermore, lipid delivery vehicles can be used for specific targeting by attaching ligands (e.g., antibodies, peptides, and carbohydrates) to its surface or to the terminal end of the attached PEG chains (Front Pharmacol. 2015 Dec. 1; 6:286).

In some aspects, the lipid conjugate is a PEG-lipid. The inclusion of polyethylene glycol (PEG) in a lipid formulation as a coating or surface ligand, a technique referred to as PEGylation, helps to protect nanoparticles from the immune system and their escape from RES uptake (Nanomedicine (Lond). 2011 June; 6(4):715-28). PEGylation has been used to stabilize lipid formulations and their payloads through physical, chemical, and biological mechanisms. Detergent-like PEG lipids (e.g., PEG-DSPE) can enter the lipid formulation to form a hydrated layer and steric barrier on the surface. Based on the degree of PEGylation, the surface layer can be generally divided into two types, brush-like and mushroom-like layers. For PEG-DSPE-stabilized formulations, PEG will take on the mushroom conformation at a low degree of PEGylation (usually less than 5 mol %) and will shift to brush conformation as the content of PEG-DSPE is increased past a certain level (Journal of Nanomaterials. 2011; 2011:12). PEGylation leads to a significant increase in the circulation half-life of lipid formulations (Annu. Rev. Biomed. Eng. 2011 Aug. 15; 130:507-30; J. Control Release. 2010 Aug. 3; 145(3):178-81).

Examples of PEG-lipids include, but are not limited to, PEG coupled to dialkylloxypropyls (PEG-DAA), PEG coupled to diacylglycerol (PEG-DAG), PEG coupled to phospholipids such as phosphatidylethanolamine (PEG-PE), PEG conjugated to ceramides, PEG conjugated to cholesterol or a derivative thereof, and mixtures thereof.

PEG is a linear, water-soluble polymer of ethylene PEG repeating units with two terminal hydroxyl groups. PEGs are classified by their molecular weights and include the following: monomethoxypolyethylene glycol (MePEG-OH), monomethoxypolyethylene glycol-succinate (MePEG-S), monomethoxypolyethylene glycol-succinimidyl succinate (MePEG-S-NHS), monomethoxypolyethylene glycol-amine (MePEG-NH₂), monomethoxypolyethylene glycol-tresylate (MePEG-TRES), monomethoxypolyethylene glycol-imadazolyl-carbonyl (MePEG-IM), as well as such compounds containing a terminal hydroxyl group instead of a terminal methoxy group (e.g., HO-PEG-S, HO-PEG-S-NHS, HO-PEG-NH₂).

The PEG moiety of the PEG-lipid conjugates described herein may comprise an average molecular weight ranging from about 550 daltons to about 10,000 daltons. In certain aspects, the PEG moiety has an average molecular weight of from about 750 daltons to about 5,000 daltons (e.g., from about 1,000 daltons to about 5,000 daltons, from about 1,500 daltons to about 3,000 daltons, from about 750 daltons to

about 3,000 daltons, from about 750 daltons to about 2,000 daltons). In some aspects, the PEG moiety has an average molecular weight of about 2,000 daltons or about 750 daltons. The average molecular weight may be any value or subvalue within the recited ranges, including endpoints.

In certain aspects, the PEG can be optionally substituted by an alkyl, alkoxy, acyl, or aryl group. The PEG can be conjugated directly to the lipid or may be linked to the lipid via a linker moiety. Any linker moiety suitable for coupling the PEG to a lipid can be used including, e.g., non-ester-containing linker moieties and ester-containing linker moieties. In one aspect, the linker moiety is a non-ester-containing linker moiety. Exemplary non-ester-containing linker moieties include, but are not limited to, amido ($—C(O)NH—$), amino ($—NR—$), carbonyl ($—C(O)—$), carbamate ($—NHC(O)O—$), urea ($—NHC(O)NH—$), disulfide ($—S—S—$), ether ($—O—$), succinyl ($—(O)CCH_2CH_2C(O)—$), succinamidyl ($—NHC(O)CH_2CH_2C(O)NH—$), ether, as well as combinations thereof (such as a linker containing both a carbamate linker moiety and an amido linker moiety). In one aspect, a carbamate linker is used to couple the PEG to the lipid.

In some aspects, an ester-containing linker moiety is used to couple the PEG to the lipid. Exemplary ester-containing linker moieties include, e.g., carbonate ($—OC(O)O—$), succinoyl, phosphate esters ($—O—(O)POH—O—$), sulfonate esters, and combinations thereof.

Phosphatidylethanolamines having a variety of acyl chain groups of varying chain lengths and degrees of saturation can be conjugated to PEG to form the lipid conjugate. Such phosphatidylethanolamines are commercially available or can be isolated or synthesized using conventional techniques known to those of skill in the art. Phosphatidylethanolamines containing saturated or unsaturated fatty acids with carbon chain lengths in the range of C10 to C20 are preferred. Phosphatidylethanolamines with mono- or di-unsaturated fatty acids and mixtures of saturated and unsaturated fatty acids can also be used. Suitable phosphatidylethanolamines include, but are not limited to, dimyristoyl-phosphatidylethanolamine (DMPE), dipalmitoyl-phosphatidylethanolamine (DPPE), dioleoyl-phosphatidylethanolamine (DOPE), and distearoyl-phosphatidylethanolamine (DSPE).

In some aspects, the PEG-DAA conjugate is a PEG-didecyloxypropyl (C10) conjugate, a PEG-dilauryloxypropyl (C12) conjugate, a PEG-dimyristyloxypropyl (C14) conjugate, a PEG-dipalmityoxypropyl (C16) conjugate, or a PEG-distearyoxypropyl (C18) conjugate. In some aspects, the PEG has an average molecular weight of about 750 or about 2,000 daltons. In some aspects, the terminal hydroxyl group of the PEG is substituted with a methyl group.

In addition to the foregoing, other hydrophilic polymers can be used in place of PEG. Examples of suitable polymers that can be used in place of PEG include, but are not limited to, polyvinylpyrrolidone, polymethoxazoline, polyethyl-oxazoline, polyhydroxypropyl, methacrylamide, polymethacrylamide, and polydimethylacrylamide, polylactic acid, polyglycolic acid, and derivatized celluloses such as hydroxymethylcellulose or hydroxyethylcellulose.

In some aspects, the lipid conjugate (e.g., PEG-lipid) comprises from about 0.1 mol % to about 2 mol %, from about 0.5 mol % to about 2 mol %, from about 1 mol % to about 2 mol %, from about 0.6 mol % to about 1.9 mol %, from about 0.7 mol % to about 1.8 mol %, from about 0.8 mol % to about 1.7 mol %, from about 0.9 mol % to about 1.6 mol %, from about 0.9 mol % to about 1.8 mol %, from about 1 mol % to about 1.8 mol %, from about 1 mol % to

125

about 1.7 mol %, from about 1.2 mol % to about 1.8 mol %, from about 1.2 mol % to about 1.7 mol %, from about 1.3 mol % to about 1.6 mol %, or from about 1.4 mol % to about 1.6 mol % (or any fraction thereof or range therein) of the total lipid present in the lipid formulation. In other embodiments, the lipid conjugate (e.g., PEG-lipid) comprises about 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1.0%, 1.2%, 1.3%, 1.4%, 1.5%, 1.6%, 1.7%, 1.8%, 1.9%, 2.0%, 2.5%, 3.0%, 3.5%, 4.0%, 4.5%, or 5%, (or any fraction thereof or range therein) of the total lipid present in the lipid formulation. The amount may be any value or subvalue within the recited ranges, including endpoints.

The percentage of lipid conjugate (e.g., PEG-lipid) present in the lipid formulations of the disclosure is a target amount, and the actual amount of lipid conjugate present in the formulation may vary, for example, by ± 0.5 mol %. One of ordinary skill in the art will appreciate that the concentration of the lipid conjugate can be varied depending on the lipid conjugate employed and the rate at which the lipid formulation is to become fusogenic.

Mechanism of Action for Cellular Uptake of Lipid Formulations

In some aspects, lipid formulations for the intracellular delivery of nucleic acids, particularly liposomes, cationic liposomes, and lipid nanoparticles, are designed for cellular uptake by penetrating target cells through exploitation of the target cells' endocytic mechanisms where the contents of the lipid delivery vehicle are delivered to the cytosol of the target cell. (Nucleic Acid Therapeutics, 28(3):146-157, 2018). Prior to endocytosis, functionalized ligands such as PEG-lipid at the surface of the lipid delivery vehicle are shed from the surface, which triggers internalization into the target cell. During endocytosis, some part of the plasma membrane of the cell surrounds the vector and engulfs it into a vesicle that then pinches off from the cell membrane, enters the cytosol and ultimately enters and moves through the endolysosomal pathway. For ionizable cationic lipid-containing delivery vehicles, the increased acidity as the endosome ages results in a vehicle with a strong positive charge on the surface. Interactions between the delivery vehicle and the endosomal membrane then result in a membrane fusion event that leads to cytosolic delivery of the payload. For RNA payloads, the cell's own internal translation processes will then translate the RNA into the encoded protein. The encoded protein can further undergo posttranslational processing, including transportation to a targeted organelle or location within the cell or excretion from the cell.

By controlling the composition and concentration of the lipid conjugate, one can control the rate at which the lipid conjugate exchanges out of the lipid formulation and, in turn, the rate at which the lipid formulation becomes fusogenic. In addition, other variables including, e.g., pH, temperature, or ionic strength, can be used to vary and/or control the rate at which the lipid formulation becomes fusogenic. Other methods which can be used to control the rate at which the lipid formulation becomes fusogenic will become apparent to those of skill in the art upon reading this disclosure. Also, by controlling the composition and concentration of the lipid conjugate, one can control the liposomal or lipid particle size.

Lipid Formulation Manufacture

There are many different methods for the preparation of lipid formulations comprising a nucleic acid. (Curr. Drug Metabol. 2014, 15, 882-892; Chem. Phys. Lipids 2014, 177, 8-18; Int. J. Pharm. Stud. Res. 2012, 3, 14-20). The techniques of thin film hydration, double emulsion, reverse

126

phase evaporation, microfluidic preparation, dual asymmetric centrifugation, ethanol injection, detergent dialysis, spontaneous vesicle formation by ethanol dilution, and encapsulation in preformed liposomes are briefly described herein.

5 Thin Film Hydration

In Thin Film Hydration (TFH) or the Bangham method, the lipids are dissolved in an organic solvent, then evaporated through the use of a rotary evaporator leading to a thin lipid layer formation. After the layer hydration by an aqueous buffer solution containing the compound to be loaded, Multilamellar Vesicles (MLVs) are formed, which can be reduced in size to produce Small or Large Unilamellar vesicles (LUV and SUV) by extrusion through membranes or by the sonication of the starting MLV.

15 Double Emulsion

Lipid formulations can also be prepared through the Double Emulsion technique, which involves lipids dissolution in a water/organic solvent mixture. The organic solution, containing water droplets, is mixed with an excess of aqueous medium, leading to a water-in-oil-in-water (W/O/W) double emulsion formation. After mechanical vigorous shaking, part of the water droplets collapse, giving Large Unilamellar Vesicles (LUVs).

25 Reverse Phase Evaporation

The Reverse Phase Evaporation (REV) method also allows one to achieve LUVs loaded with nucleic acid. In this technique a two-phase system is formed by phospholipids dissolution in organic solvents and aqueous buffer. The resulting suspension is then sonicated briefly until the mixture becomes a clear one-phase dispersion. The lipid formulation is achieved after the organic solvent evaporation under reduced pressure. This technique has been used to encapsulate different large and small hydrophilic molecules including nucleic acids.

35 Microfluidic Preparation

The Microfluidic method, unlike other bulk techniques, gives the possibility of controlling the lipid hydration process. The method can be classified in continuous-flow microfluidic and droplet-based microfluidic, according to the way in which the flow is manipulated. In the microfluidic hydrodynamic focusing (MHF) method, which operates in a continuous flow mode, lipids are dissolved in isopropyl alcohol which is hydrodynamically focused in a microchannel cross junction between two aqueous buffer streams. Vesicles size can be controlled by modulating the flow rates, thus controlling the lipids solution/buffer dilution process. The method can be used for producing oligonucleotide (ON) lipid formulations by using a microfluidic device consisting of three-inlet and one-outlet ports.

50 Dual Asymmetric Centrifugation

Dual Asymmetric Centrifugation (DAC) differs from more common centrifugation as it uses an additional rotation around its own vertical axis. An efficient homogenization is achieved due to the two overlaying movements generated: the sample is pushed outwards, as in a normal centrifuge, and then it is pushed towards the center of the vial due to the additional rotation. By mixing lipids and an NaCl-solution a viscous vesicular phospholipid gel (VPC) is achieved, which is then diluted to obtain a lipid formulation dispersion. The lipid formulation size can be regulated by optimizing DAC speed, lipid concentration and homogenization time.

Ethanol Injection

The Ethanol Injection (EI) method can be used for nucleic acid encapsulation. This method provides the rapid injection of an ethanolic solution, in which lipids are dissolved, into an aqueous medium containing nucleic acids to be encap-

127

sulated, through the use of a needle. Vesicles are spontaneously formed when the phospholipids are dispersed throughout the medium.

Detergent Dialysis

The Detergent dialysis method can be used to encapsulate nucleic acids. Briefly lipid and plasmid are solubilized in a detergent solution of appropriate ionic strength, after removing the detergent by dialysis, a stabilized lipid formulation is formed. Unencapsulated nucleic acid is then removed by ion-exchange chromatography and empty vesicles by sucrose density gradient centrifugation. The technique is highly sensitive to the cationic lipid content and to the salt concentration of the dialysis buffer, and the method is also difficult to scale.

Spontaneous Vesicle Formation by Ethanol Dilution

Stable lipid formulations can also be produced through the Spontaneous Vesicle Formation by Ethanol Dilution method in which a stepwise or dropwise ethanol dilution provides the instantaneous formation of vesicles loaded with nucleic acid by the controlled addition of lipid dissolved in ethanol to a rapidly mixing aqueous buffer containing the nucleic acid.

Encapsulation in Preformed Liposomes

The entrapment of nucleic acids can also be obtained starting with preformed liposomes through two different methods: (1) A simple mixing of cationic liposomes with nucleic acids which gives electrostatic complexes called "Lipoplexes", where they can be successfully used to transfect cell cultures, but are characterized by their low encapsulation efficiency and poor performance in vivo; and (2) a liposomal destabilization, slowly adding absolute ethanol to a suspension of cationic vesicles up to a concentration of 40% v/v followed by the dropwise addition of nucleic acids achieving loaded vesicles; however, the two main steps characterizing the encapsulation process are too sensitive, and the particles have to be downsized.

Excipients

The pharmaceutical compositions disclosed herein can be formulated using one or more excipients to: (1) increase stability; (2) increase cell transfection; (3) permit a sustained or delayed release (e.g., from a depot formulation of the polynucleotide, primary construct, or RNA); (4) alter the biodistribution (e.g., target the polynucleotide, primary construct, or RNA to specific tissues or cell types); (5) increase the translation of encoded protein in vivo; and/or (6) alter the release profile of encoded protein in vivo.

The pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of associating the active ingredient (i.e., nucleic acid) with an excipient and/or one or more other accessory ingredients. A pharmaceutical composition in accordance with the present disclosure may be prepared, packaged, and/or sold in bulk, as a single unit dose, and/or as a plurality of single unit doses.

Pharmaceutical compositions may additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes, but is not limited to, any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, and the like, as suited to the particular dosage form desired.

In addition to traditional excipients such as any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, excipients of the present disclosure can include, without

128

limitation, liposomes, lipid nanoparticles, polymers, lipoplexes, core-shell nanoparticles, peptides, proteins, cells transfected with primary DNA construct, or RNA (e.g., for transplantation into a subject), hyaluronidase, nanoparticle mimics and combinations thereof.

Accordingly, the pharmaceutical compositions described herein can include one or more excipients, each in an amount that together increases the stability of the nucleic acid in the lipid formulation, increases cell transfection by 10 the nucleic acid, increases the expression of the encoded protein, and/or alters the release profile of encoded proteins. Further, the RNA of the present disclosure may be formulated using self-assembled nucleic acid nanoparticles.

Various excipients for formulating pharmaceutical compositions and techniques for preparing the composition are 15 known in the art (see Remington: The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro, Lippincott, Williams & Wilkins, Baltimore, Md., 2006; incorporated herein by reference in its entirety). The use of a conventional 20 excipient medium may be contemplated within the scope of the embodiments of the present disclosure, except insofar as any conventional excipient medium may be incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition.

The pharmaceutical compositions of this disclosure may further contain as pharmaceutically acceptable carriers substances as required to approximate physiological conditions, such as pH adjusting and buffering agents, tonicity adjusting agents, and wetting agents, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, and mixtures thereof. For solid compositions, conventional 30 nontoxic pharmaceutically acceptable carriers can be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like.

40 In certain embodiments of the disclosure, the RNA-lipid formulation may be administered in a time release formulation, for example in a composition which includes a slow release polymer. The active agent can be prepared with carriers that will protect against rapid release, for example a controlled release vehicle such as a polymer, microencapsulated delivery system, or a bioadhesive gel. Prolonged 45 delivery of the RNA, in various compositions of the disclosure can be brought about by including in the composition agents that delay absorption, for example, aluminum monostearate hydrogels and gelatin.

Methods of Inducing Immune Responses

Provided herein, in some embodiments, are methods of inducing an immune response in a subject. Any type of immune response can be induced using the methods provided herein, including adaptive and innate immune responses. In one aspect, immune responses induced using the methods provided herein include an antibody response, a cellular immune response, or both an antibody response and a cellular immune response.

60 Methods of inducing an immune response provided herein include administering to a subject an effective amount of any nucleic acid molecule provided herein. In one aspect, methods of inducing an immune response include administering to a subject an effective amount of any composition comprising a nucleic acid molecule and a lipid provided herein. In another aspect, methods of inducing an immune response include administering to a subject an effective amount of any

129

pharmaceutical composition comprising a nucleic acid molecule and a lipid formulation provided herein. In some aspects, nucleic acid molecules, compositions, and pharmaceutical composition provided here are vaccines that can elicit a protective or a therapeutic immune response, for example.

As used herein, the term "subject" refers to any individual or patient on which the methods disclosed herein are performed. The term "subject" can be used interchangeably with the term "individual" or "patient." The subject can be a human, although the subject may be an animal, as will be appreciated by those in the art. Thus, other animals, including mammals such as rodents (including mice, rats, hamsters and guinea pigs), cats, dogs, rabbits, farm animals including cows, horses, goats, sheep, pigs, etc., and primates (including monkeys, chimpanzees, orangutans and gorillas) are included within the definition of subject. As used herein, the term "effective amount" or "therapeutically effective amount" refers to that amount of a nucleic acid molecule, composition, or pharmaceutical composition described herein that is sufficient to effect the intended application, including but not limited to inducing an immune response and/or disease treatment, as defined herein. The therapeutically effective amount may vary depending upon the intended application (e.g., inducing an immune response, treatment, application *in vivo*), or the subject or patient and disease condition being treated, e.g., the weight and age of the subject, the species, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The term also applies to a dose that will induce a particular response in a target cell. The specific dose will vary depending on the particular nucleic acid molecule, composition, or pharmaceutical composition chosen, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, the tissue to which it is administered, and the physical delivery system in which it is carried.

Exemplary doses of nucleic acid molecules that can be administered include about 0.01 µg, about 0.02 µg, about 0.03 µg, about 0.04 µg, about 0.05 µg, about 0.06 µg, about 0.07 µg, about 0.08 µg, about 0.09 µg, about 0.1 µg, about 0.2 µg, about 0.3 µg, about 0.4 µg, about 0.5 µg, about 0.6 µg, about 0.7 µg, about 0.8 µg, about 0.9 µg, about 1.0 µg, about 1.5 µg, about 2.0 µg, about 2.5 µg, about 3.0 µg, about 3.5 µg, about 4.0 µg, about 4.5 µg, about 5.0 µg, about 5.5 µg, about 6.0 µg, about 6.5 µg, about 7.0 µg, about 7.5 µg, about 8.0 µg, about 8.5 µg, about 9.0 µg, about 9.5 µg, about 10 µg, about 11 µg, about 12 µg, about 13 µg, about 14 µg, about 15 µg, about 16 µg, about 17 µg, about 18 µg, about 19 µg, about 20 µg, about 21 µg, about 22 µg, about 23 µg, about 24 µg, about 25 µg, about 26 µg, about 27 µg, about 28 µg, about 29 µg, about 30 µg, about 35 µg, about 40 µg, about 45 µg, about 50 µg, about 55 µg, about 60 µg, about 65 µg, about 70 µg, about 75 µg, about 80 µg, about 85 µg, about 90 µg, about 95 µg, about 100 µg, about 125 µg, about 150 µg, about 175 µg, about 200 µg, about 250 µg, about 300 µg, about 350 µg, about 400 µg, about 450 µg, about 500 µg, about 600 µg, about 700 µg, about 800 µg, about 900 µg, about 1,000 µg, or more, and any number or range in between. In one aspect, the nucleic acid molecules are RNA molecules. In another aspect, the nucleic acid molecules are DNA molecules. Nucleic acid molecules can have a unit dosage comprising about 0.01 µg to about 1,000 µg or more nucleic acid in a single dose.

In some aspects, compositions provided herein that can be administered include about 0.01 µg, about 0.02 µg, about

130

0.03 µg, about 0.04 µg, about 0.05 µg, about 0.06 µg, about 0.07 µg, about 0.08 µg, about 0.09 µg, about 0.1 µg, about 0.2 µg, about 0.3 µg, about 0.4 µg, about 0.5 µg, about 0.6 µg, about 0.7 µg, about 0.8 µg, about 0.9 µg, about 1.0 µg, about 1.5 µg, about 2.0 µg, about 2.5 µg, about 3.0 µg, about 3.5 µg, about 4.0 µg, about 4.5 µg, about 5.0 µg, about 5.5 µg, about 6.0 µg, about 6.5 µg, about 7.0 µg, about 7.5 µg, about 8.0 µg, about 8.5 µg, about 9.0 µg, about 9.5 µg, about 10 µg, about 11 µg, about 12 µg, about 13 µg, about 14 µg, about 15 µg, about 16 µg, about 17 µg, about 18 µg, about 19 µg, about 20 µg, about 21 µg, about 22 µg, about 23 µg, about 24 µg, about 25 µg, about 26 µg, about 27 µg, about 28 µg, about 29 µg, about 30 µg, about 35 µg, about 40 µg, about 45 µg, about 50 µg, about 55 µg, about 60 µg, about 65 µg, about 70 µg, about 75 µg, about 80 µg, about 85 µg, about 90 µg, about 95 µg, about 100 µg, about 125 µg, about 150 µg, about 175 µg, about 200 µg, about 250 µg, about 300 µg, about 350 µg, about 400 µg, about 450 µg, about 500 µg, about 600 µg, about 700 µg, about 800 µg, about 900 µg, about 1,000 µg, or more, and any number or range in between, nucleic acid and lipid. In other aspects, pharmaceutical compositions provided herein that can be administered include about 0.01 µg, about 0.02 µg, about 0.03 µg, about 0.04 µg, about 0.05 µg, about 0.06 µg, about 0.07 µg, about 0.08 µg, about 0.09 µg, about 0.1 µg, about 0.2 µg, about 0.3 µg, about 0.4 µg, about 0.5 µg, about 0.6 µg, about 0.7 µg, about 0.8 µg, about 0.9 µg, about 1.0 µg, about 1.5 µg, about 2.0 µg, about 2.5 µg, about 3.0 µg, about 3.5 µg, about 4.0 µg, about 4.5 µg, about 5.0 µg, about 5.5 µg, about 6.0 µg, about 6.5 µg, about 7.0 µg, about 7.5 µg, about 8.0 µg, about 8.5 µg, about 9.0 µg, about 9.5 µg, about 10 µg, about 11 µg, about 12 µg, about 13 µg, about 14 µg, about 15 µg, about 16 µg, about 17 µg, about 18 µg, about 19 µg, about 20 µg, about 21 µg, about 22 µg, about 23 µg, about 24 µg, about 25 µg, about 26 µg, about 27 µg, about 28 µg, about 29 µg, about 30 µg, about 35 µg, about 40 µg, about 45 µg, about 50 µg, about 55 µg, about 60 µg, about 65 µg, about 70 µg, about 75 µg, about 80 µg, about 85 µg, about 90 µg, about 95 µg, about 100 µg, about 125 µg, about 150 µg, about 175 µg, about 200 µg, about 250 µg, about 300 µg, about 350 µg, about 400 µg, about 450 µg, about 500 µg, about 600 µg, about 700 µg, about 800 µg, about 900 µg, about 1,000 µg, or more, and any number or range in between, nucleic acid and lipid formulation.

In one aspect, compositions provided herein can have a unit dosage comprising about 0.01 µg to about 1,000 µg or more nucleic acid and lipid in a single dose. In another aspect, pharmaceutical compositions provided herein can have a unit dosage comprising about 0.01 µg to about 1,000 µg or more nucleic acid and lipid formulation in a single dose. A vaccine unit dosage can correspond to the unit dosage of nucleic acid molecules, compositions, or pharmaceutical compositions provided herein and that can be administered to a subject. In one aspect, vaccine compositions of the instant disclosure have a unit dosage comprising about 0.01 µg to about 1,000 µg or more nucleic acid and lipid formulation in a single dose. In another aspect, vaccine compositions of the instant disclosure have a unit dosage comprising about 0.01 µg to about 50 µg nucleic acid and lipid formulation in a single dose. In yet another aspect, vaccine compositions of the instant disclosure have a unit dosage comprising about 0.2 µg to about 20 µg nucleic acid and lipid formulation in a single dose.

A dosage form of the composition of this disclosure can be solid, which can be reconstituted in a liquid prior to administration. The solid can be administered as a powder. The solid can be in the form of a capsule, tablet, or gel. In

131

some embodiments, the pharmaceutical composition comprises a nucleic acid lipid formulation that has been lyophilized.

In a preferred embodiment, the dosage form of the pharmaceutical compositions described herein can be a liquid suspension of self-replicating RNA lipid nanoparticles described herein. In some embodiments, the liquid suspension is in a buffered solution. In some embodiments, the buffered solution comprises a buffer selected from the group consisting of HEPES, MOPS, TES, and TRIS. In some embodiments, the buffer has a pH of about 7.4. In some preferred embodiments, the buffer is HEPES. In some further embodiments, the buffered solution further comprises a cryoprotectant. In some embodiments, the cryoprotectant is selected from a sugar and glycerol or a combination of a sugar and glycerol. In some embodiments, the sugar is a dimeric sugar. In some embodiments, the sugar is sucrose. In some preferred embodiments, the buffer comprises HEPES, sucrose, and glycerol at a pH of 7.4. In some embodiments, the suspension is frozen during storage and thawed prior to administration. In some embodiments, the suspension is frozen at a temperature below about 70° C. In some embodiments, the suspension is diluted with sterile water during intravenous administration. In some embodiments, intravenous administration comprises diluting the suspension with about 2 volumes to about 6 volumes of sterile water. In some embodiments, the suspension comprises about 0.1 mg to about 3.0 mg self-replicating RNA/mL, about 15 mg/mL to about 25 mg/mL of an ionizable cationic lipid, about 0.5 mg/mL to about 2.5 mg/mL of a PEG-lipid, about 1.8 mg/mL to about 3.5 mg/mL of a helper lipid, about 4.5 mg/mL to about 7.5 mg/mL of a cholesterol, about 7 mg/mL to about 15 mg/mL of a buffer, about 2.0 mg/mL to about 4.0 mg/mL of NaCl, about 70 mg/mL to about 110 mg/mL of sucrose, and about 50 mg/mL to about 70 mg/mL of glycerol. In some embodiments, a lyophilized self-replicating RNA-lipid nanoparticle formulation can be resuspended in a buffer as described herein.

In some embodiments, the compositions of the disclosure are administered to a subject such that a self-replicating RNA concentration of at least about 0.05 mg/kg, at least about 0.1 mg/kg, at least about 0.5 mg/kg, at least about 1.0 mg/kg, at least about 2.0 mg/kg, at least about 3.0 mg/kg, at least about 4.0 mg/kg, at least about 5.0 mg/kg of body weight is administered in a single dose or as part of single treatment cycle. In some embodiments, the compositions of the disclosure are administered to a subject such that a total amount of at least about 0.1 mg, at least about 0.5 mg, at least about 1.0 mg, at least about 2.0 mg, at least about 3.0 mg, at least about 4.0 mg, at least about 5.0 mg, at least about 6.0 mg, at least about 7.0 mg, at least about 8.0 mg, at least about 9.0 mg, at least about 10 mg, at least about 15 mg, at least about 20 mg, at least about 25 mg, at least about 30 mg, at least about 35 mg, at least about 40 mg, at least about 45 mg, at least about 50 mg, at least about 55 mg, at least about 60 mg, at least about 65 mg, at least about 70 mg, at least about 75 mg, at least about 80 mg, at least about 85 mg, at least about 90 mg, at least about 95 mg, at least about 100 mg, at least about 105 mg, at least about 110 mg, at least about 115 mg, at least about 120 mg, or at least about 125 mg self-replicating RNA is administered in one or more doses up to a maximum dose of about 300 mg, about 350 mg, about 400 mg, about 450 mg, or about 500 mg self-replicating RNA.

Any route of administration can be included in methods provided herein. In some aspects, nucleic acid molecules, compositions, and pharmaceutical compositions provided

132

herein are administered intramuscularly, subcutaneously, intradermally, transdermally, intranasally, orally, sublingually, intravenously, intraperitoneally, topically, by aerosol, or by a pulmonary route, such as by inhalation or by nebulization, for example. In some embodiments, the pharmaceutical compositions described are administered systemically. Suitable routes of administration include, for example, rectal, vaginal, transmucosal, or intestinal administration; parenteral delivery, including intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, or intranasal. In particular embodiments, the intramuscular administration is to a muscle selected from the group consisting of skeletal muscle, smooth muscle and cardiac muscle. In some embodiments, the pharmaceutical composition is administered intravenously.

Pharmaceutical compositions may be administered to any desired tissue. In some embodiments, the self-replicating RNA delivered is expressed in a tissue different from the tissue in which the lipid formulation or pharmaceutical composition was administered. In preferred embodiments, self-replicating RNA is delivered and expressed in the liver.

In other aspects, nucleic acid molecules, compositions, and pharmaceutical compositions provided herein are administered intramuscularly.

In some aspects, the subject in which an immune response is induced is a healthy subject. As used herein, the term "healthy subject" refers to a subject not having a condition or disease, including an infectious disease or cancer, for example, or not having a condition or disease against which an immune response is induced. Accordingly, in some aspects, a nucleic acid molecule, composition, or pharmaceutical composition provided herein is administered prophylactically to prevent an infectious disease or cancer, for example. In other aspects, the subject in which an immune response is induced has cancer. The subject may suffer from any cancer or have any tumor, including solid and liquid tumors. In one aspect, the cancer is kidney cancer, renal cancer, urinary bladder cancer, prostate cancer, uterine cancer, breast cancer, cervical cancer, ovarian cancer, lung cancer, liver cancer, stomach cancer, colon cancer, rectal cancer, oral cavity cancer, pharynx cancer, pancreatic cancer, thyroid cancer, melanoma, skin cancer, head and neck cancer, brain cancer, hematopoietic cancer, leukemia, lymphoma, bone cancer, or sarcoma. Accordingly, a nucleic acid molecule, composition, or pharmaceutical composition provided herein can be administered therapeutically, i.e., to treat a condition or disease, such as cancer, after the onset of the condition or disease.

As used herein, the terms "treat," "treatment," "therapy," "therapeutic," and the like refer to obtaining a desired pharmacologic and/or physiologic effect, including, but not limited to, alleviating, delaying or slowing the progression, reducing the effects or symptoms, preventing onset, inhibiting, ameliorating the onset of a disease or disorder, obtaining a beneficial or desired result with respect to a disease, disorder, or medical condition, such as a therapeutic benefit and/or a prophylactic benefit. "Treatment," as used herein, includes any treatment of a disease in a mammal, particularly in a human, and includes: (a) preventing the disease from occurring in a subject, including a subject which is predisposed to the disease or at risk of acquiring the disease but has not yet been diagnosed as having it; (b) inhibiting the disease, i.e., arresting its development; and (c) relieving the disease, i.e., causing regression of the disease. A therapeutic benefit includes eradication or amelioration of the underlying disorder being treated. Also, a therapeutic

133

benefit is achieved with the eradication or amelioration of one or more of the physiological symptoms associated with the underlying disorder such that an improvement is observed in the subject, notwithstanding that the subject may still be afflicted with the underlying disorder. In some aspects, for prophylactic benefit, treatment or compositions for treatment, including pharmaceutical compositions, are administered to a subject at risk of developing a particular disease, or to a subject reporting one or more of the physiological symptoms of a disease, even though a diagnosis of this disease may not have been made. The methods of the present disclosure may be used with any mammal or other animal. In some aspects, treatment results in a decrease or cessation of symptoms. A prophylactic effect includes delaying or eliminating the appearance of a disease or condition, delaying or eliminating the onset of symptoms of a disease or condition, slowing, halting, or reversing the progression of a disease or condition, or any combination thereof.

Nucleic acid molecules, compositions, and pharmaceutical compositions provided herein can be administered once or multiple times. Accordingly, nucleic acid molecules, compositions, and pharmaceutical compositions provided herein can be administered one, two, three, four, five, six, seven, eight, nine, ten, or more times. Timing between two or more administrations can be one week, two weeks, three weeks, four weeks, five weeks, six weeks, seven weeks, eight weeks, nine weeks, weeks, ten weeks, 11 weeks, 12 weeks, 13 weeks, 14 weeks, 15 weeks, 16 weeks, 17 weeks, 18 weeks, 19 weeks, 20 weeks, 21 weeks, 22 weeks, 23 weeks, 24 weeks, 25 weeks, 26 weeks, 27 weeks, 28 weeks, 29 weeks, 30 weeks, 31 weeks, 32 weeks, 33 weeks, 34 weeks, 35 weeks, 36 weeks, 37 weeks, 38 weeks, 39 weeks, 40 weeks, 41 weeks, 42 weeks, 43 weeks, 44 weeks, 45 weeks, 46 weeks, 47 weeks, 48 weeks, 49 weeks, 50 weeks, 51 weeks, 52 weeks, or more weeks, and any number or range in between. In some aspects, timing between two or more administrations is one month, two months, three months, four months, five months, six months, seven months, eight months, nine months, ten months, 11 months, 12 months, 13 months, 14 months, 15 months, 16 months, 17 months, 18 months, 19 months, 20 months, 21 months, 22 months, 23 months, 24 months, or more months, and any number or range in between. In other aspects, timing between two or more administrations can be one year, two years, three years, four years, five years, six years, seven years, eight years, nine years, ten years, or more years, and any number or range in between. Timing between the first and any subsequent administration can be the same or different. In one aspect, nucleic acid molecules, compositions, or pharmaceutical compositions provided herein are administered once.

More than one nucleic acid molecule, composition, or pharmaceutical composition can be administered in the methods provided herein. In one aspect, two or more nucleic acid molecules, compositions, or pharmaceutical compositions provided herein are administered simultaneously. In another aspect, two or more nucleic acid molecules, compositions, or pharmaceutical compositions provided herein are administered sequentially. Simultaneous and sequential administrations can include any number and any combination of nucleic acid molecules, compositions, or pharmaceutical compositions provided herein. Multiple nucleic acid molecules, compositions, or pharmaceutical compositions that are administered together or sequentially can include transgenes encoding different antigenic proteins or fragments thereof. In this manner, immune responses against

134

different antigenic targets can be induced. Two, three, four, five, six, seven, eight, nine, ten, or more nucleic acid molecules, compositions, or pharmaceutical compositions including transgenes encoding different antigenic proteins or fragments thereof can be administered simultaneously or sequentially. Any combination of nucleic acid molecules, compositions, and pharmaceutical compositions including any combination of transgenes can be administered simultaneously or sequentially. In some aspects, administration is simultaneous. In other aspects, administration is sequential. Timing between two or more administrations can be one week, two weeks, three weeks, four weeks, five weeks, six weeks, seven weeks, eight weeks, nine weeks, weeks, ten weeks, 11 weeks, 12 weeks, 13 weeks, 14 weeks, 15 weeks, 16 weeks, 17 weeks, 18 weeks, 19 weeks, 20 weeks, 21 weeks, 22 weeks, 23 weeks, 24 weeks, 25 weeks, 26 weeks, 27 weeks, 28 weeks, 29 weeks, 30 weeks, 31 weeks, 32 weeks, 33 weeks, 34 weeks, 35 weeks, 36 weeks, 37 weeks, 38 weeks, 39 weeks, 40 weeks, 41 weeks, 42 weeks, 43 weeks, 44 weeks, 45 weeks, 46 weeks, 47 weeks, 48 weeks, 49 weeks, 50 weeks, 51 weeks, 52 weeks, or more weeks, and any number or range in between. In some aspects, timing between two or more administrations is one month, two months, three months, four months, five months, six months, seven months, eight months, nine months, ten months, 11 months, 12 months, 13 months, 14 months, 15 months, 16 months, 17 months, 18 months, 19 months, 20 months, 21 months, 22 months, 23 months, 24 months, or more months, and any number or range in between. In other aspects, timing between two or more administrations can be one year, two years, three years, four years, five years, six years, seven years, eight years, nine years, ten years, or more years, and any number or range in between. Timing between the first and any subsequent administration can be the same or different. Nucleic acid molecules, compositions, and pharmaceutical compositions provided herein can be administered with any other vaccine or treatment.

Following administration of the composition to the subject, the protein product encoded by the self-replicating RNA of the disclosure (e.g., an antigen) is detectable in the target tissues for at least about one to seven days or longer. For example, the protein product may be detectable in the target tissues at a concentration (e.g., a therapeutic concentration) of at least about 0.025-1.5 µg/ml (e.g., at least about 0.050 µg/ml, at least about 0.075 µg/ml, at least about 0.1 µg/ml, at least about 0.2 µg/ml, at least about 0.3 µg/ml, at least about 0.4 µg/ml, at least about 0.5 µg/ml, at least about 0.6 µg/ml, at least about 0.7 µg/ml, at least about 0.8 µg/ml, at least about 0.9 µg/ml, at least about 1.0 µg/ml, at least about 1.1 µg/ml, at least about 1.2 µg/ml, at least about 1.3 µg/ml, at least about 1.4 µg/ml, or at least about 1.5 µg/ml), for at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45 days or longer following administration of the composition to the subject.

In some embodiments, a pharmaceutical composition of the present disclosure is administered to a subject once per month. In some embodiments, a pharmaceutical composition of the present disclosure is administered to a subject twice per month. In some embodiments, a pharmaceutical composition of the present disclosure is administered to a subject three times per month. In some embodiments, a pharmaceutical composition of the present disclosure is administered to a subject four times per month.

Alternatively, the compositions of the present disclosure may be administered in a local rather than systemic manner,

135

for example, via injection of the pharmaceutical composition directly into a targeted tissue, preferably in a depot or sustained release formulation. Local delivery can be affected in various ways, depending on the tissue to be targeted. For example, aerosols containing compositions of the present disclosure can be inhaled (for nasal, tracheal, or bronchial delivery); compositions of the present disclosure can be injected into the site of injury, disease manifestation, or pain, for example; compositions can be provided in lozenges for oral, tracheal, or esophageal application; can be supplied in liquid, tablet or capsule form for administration to the stomach or intestines, can be supplied in suppository form for rectal or vaginal application; or can even be delivered to the eye by use of creams, drops, or even injection. Formulations containing compositions of the present disclosure complexed with therapeutic molecules or ligands can even be surgically administered, for example in association with a polymer or other structure or substance that can allow the compositions to diffuse from the site of implantation to surrounding cells. Alternatively, they can be applied surgically without the use of polymers or supports.

The self-replicating RNA, formulations thereof, or encoded proteins described herein may be used in combination with one or more other therapeutic, prophylactic, diagnostic, or imaging agents. By "in combination with," it is not intended to imply that the agents must be administered at the same time and/or formulated for delivery together, although these methods of delivery are within the scope of the present disclosure. Compositions can be administered concurrently with, prior to, or subsequent to, one or more other desired therapeutics or medical procedures. In general, each agent will be administered at a dose and/or on a time schedule determined for that agent. Preferably, the methods of treatment of the present disclosure encompass the delivery of pharmaceutical, prophylactic, diagnostic, or imaging compositions in combination with agents that may improve their bioavailability, reduce and/or modify their metabolism, inhibit their excretion, and/or modify their distribution within the body. As a non-limiting example, a self-replicating RNA of the disclosure may be used in combination with a pharmaceutical agent for immunizing or vaccinating a subject. In general, it is expected that agents utilized in combination with the presently disclosed self-replicating RNA and formulations thereof be utilized at levels that do not exceed the levels at which they are utilized individually. In some embodiments, the levels utilized in combination will be lower than those utilized individually. In one embodiment, the combinations, each or together may be administered according to the split dosing regimens as are known in the art.

Ranges: throughout this disclosure, various aspects can be presented in range format. It should be understood that any description in range format is merely for convenience and brevity and not meant to be limiting. Accordingly, the description of a range should be considered to have specifically disclosed all possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6, etc., as well as individual numbers within that range, for example 1, 2, 2.1, 2.2, 2.5, 3, 4, 4.75, 4.8, 4.85, 4.95, 5, 5.5, 5.75, 5.9, 5.00, and 6. This applies to a range of any breadth.

Example 1

This example describes characterization of self-replicating (STARR™) technology using firefly luciferase transgene

136

expression. In vitro transcripts were formulated with lipid nanoparticles (LNP) at a concentration of 0.1 mg/ml, and injected intramuscularly in both legs of female BALB/C mice (n=3) at a dose of 5 ug per leg. Expression of firefly luciferase (FLuc) was measured by IVIS Lumina LT Series III (PerkinElmer) by administering 100 ul of 1.5 mg Xenogen D-luciferin (PerkinElmer) in PBS via intraperitoneal injection ~10 min prior to the measurement. Six data points per group of mice were obtained at each time point (FIGS. 2A-2D).

Firefly luciferase (FLuc) expression was monitored from STARR™ Fluc, SINV FLuc, and mRNA FLuc up to day 28 by In Vivo Imaging System (IVIS). Enhanced levels and durations of transgene expression from STARR™ were observed. The expression from STARR™ Fluc peaked around day 3 to 7 and declined until day 22. FLuc expression from SINV FLuc also peaked on day 10, however, the expression was reduced at a significantly faster rate than STARR™ FLuc. Additionally, the expression on day 3 was significantly lower than STARR™ FLuc. FLuc expression from the conventional mRNA backbone was highest at day 1, the earliest time point in this study, and declined at a slightly faster rate than that of STARR™—Fluc (FIG. 2A). FIG. 2B shows that at 14 days post dosing, FLuc expression from STARR™ FLuc was higher than the other groups by about two orders of magnitude. FIG. 2D shows that the effect of the STARR™ backbone remained minimal throughout the experimental period (up to day 28), while prior administration of SINV replicon backbone resulted in a reduction of FLuc transgene expression by ~2 orders of magnitude.

A cancer vaccine substrate, TA STARR™, was constructed next with the STARR™ backbone that encodes AH1A5 epitope from gp70, an envelope glycoprotein of endogenous Murine leukemia virus. AH1 (SPSYVYHQF) (SEQ ID NO:110) is an H-2Ld-restricted antigen of gp70423-431, which is expressed in tumor cells such as the CT26 colorectal cancer cell line, but not expressed in most of the normal tissues. AH1-A5 is a mutated sequence with SPSYAHQF (SEQ ID NO:111) (the mutation underlined) with enhanced affinity to the T cell receptor (Slansky, et al., 2000, *Immunity* 13: 529-538). The open reading frame of the TA STARR™ subgenomic RNA contains a cassette with a signal peptide from the HLA class I antigen, gp70 sequence containing AH1A5 epitope, ovalbumin epitope (OVA323-339), and MHC class I trafficking signal (Kreiter, et al. 2008, *J Immunol* 180: 309-318). Three female BALB/c mice were intramuscularly injected with 10 ug of LNP formulated STARR™ transcripts, STARR™ FLuc or TA STARR™, on day 0 and day 7. On day 16, the spleens were harvested and the splenocytes were isolated. Splenocytes (2.5×10^5 cells) were incubated with or without AH1A5 (SPSYAYHQF) (SEQ ID NO:111), beta-gal peptide (TPH-PARIGL) (SEQ ID NO:112) at 1 ug/ml, and 1× Concanavalin A (Life Technologies). ELISpot detecting murine IFN-gamma (ImmunoSpot) was performed according to the manufacturer's instructions. As can be seen in FIG. 3, TA STARR™ elicited antigen-specific IFN-gamma responses.

BALB/c mice, 10 week-old female, were subcutaneously implanted in the right flank with 5×10^5 cells of CT26 cells in PBS. A day later, LNP formulated STARR™ RNA was injected intramuscularly in the left leg at a dose of 10 ug in 100 ul. The mice were administered another booster shot on day 8 with the same dose. For a group with combination treatment of anti-mouse PD1 (RMP1-14, BioXCell) and anti-mouse PDL1 (10F.9G2, BioXcell), the combined checkpoint inhibitor (100 ug each) was administered via

137

intraperitoneal injection in the right quadrant twice weekly for two weeks starting on day 3. For a group with the treatment of anti-mouse CTLA4 (9H10, BioXCell), 200 ug of the checkpoint inhibitor was administered in the same manner but starting on day 7. Five mice of the group with the combo treatment of TA STARR™ vaccine and the checkpoint inhibitors remained tumor-free on day 25, and were further challenged by subcutaneous implantation of CT26 (5x105 cells) in the right flank where the implantation site was slightly above the first implantation site. Naïve mice were used as a control group. The tumor growth was monitored for another 17 days (i.e. up to day 42 since the first CT26 implantation) before euthanization. FIGS. 4A-4F illustrate reduced tumor growth resulting from TA STARR™ vaccination and FIG. 5 shows prolonged protec-

138

tion resulting from treatment with the TA STARR™ vaccine in combination with checkpoint inhibitors.

5 Splenocytes from the combination treatment group with
TA STARR™ and anti-PD1/PDL1 were harvested for
tetramer staining with AH1 peptide. Splenocytes from the
control group with the LNP formulation buffer with the same
dosing schedule were used as a negative control. The
10 splenocytes (2×10^6 cells) were incubated with AH1
(H-2Ld)-tetramer (MBL) followed by appropriate fluores-
cent-labeled antibodies (Alexa Fluor 488 anti-CD8a (53-
6.7), Pacific Orange anti-CD4 (RM4-5), and Pacific Blue
anti-mouse CD3E (145-2C11), (eBioscience) and DRAQ7
(Invitrogen) by following the manufacturer's recommenda-
tion, and 500 K events were analyzed by ZE5 Cell Analyzer
(Bio-Rad). Results are shown in FIGS. 6A-6C.

TABLE 6

Transgene ORF nucleotide sequence						
mARM #	RNA back bone	transgene	Sequence			
2809 (SEQ ID NO: 84)	STARR TM	Fluc	AUGGAAAGAUGCCAAAAACAUUAAGAAGGGCCAGGCCAUUCUACC CACUCGAAGACGGGACCGCCGGCAGCAGCUGCACAAAGCAUGAA GGCGCUACGCCUUGGUGCCGGCACAUCCUUUACCGACGCACAU AUCGAGGUUGGACAUUACCUACGCGAGUACUUCGAGAUGGGUUC GGCUGGGAGAACUAUGAAGCGCUAUGGCUGAAUACAACAUACG GAUCGUGGUUGCGAGCAAUGCUUGCAGUUCUUCUCAUGCCGUG UUUGGGUGGCCUGUCAUCGGUGGCGUGGCCAGCUAACGACA UCUACAAACGAGCGCAGCUGCUGAACAGCAUGGGCAUCAGCCAGC CACCGUCGUAUUCGUGAGCAAGAAAGGGCUGCAAAGAACUCCUAC GUGCAAAAGAACGUACCGAUCAUACAAAAGAUCAUCAUCAUGGUA GCAAGACCCGACUACCCAGGGCUUCCAAGACAUACCUUCGUGAC UUCCCAUUUGGCCACCGGCUUCAACAGAUACGACUUCUGCCGAG ACGUUCGACCGGACAAACCAUCGCCUGUAUCAUGAACAGUAGUG GCAGUACCGGAUUGCACAGGGCGUAGCCUACCGCACCGCACC UUUGUGUCCGAUUCAGUCAUGCCCGCGUAGCCCAUCUUCGGCAAC AUCAUCCCACGACCCGCUAUCCUCAGCGUGGUUGGCCAUUAC GGCUGGGCGAUUACCCAGCGCUGGGCUACUAGUACUUCGGC GGUCUGGUCAUGUACCCGCUUCGAGGAGGAGCAUUAUUCUG UUGCAAGACAUUAAGAUUCAACUUCGCCCCUGCUGGUUGGCCAC UUAGCUUUCUGCUAAGAGCACUCUCAUGACAGAUACGACCUAAG CAACUUGCACGAGAUUCGCAAGCGGGGGCGCCUGACAGGAG GUAGGUGAGGCCGGUGGCAAAACCGCUUCCACCUACAGGCAUC AGGGCUACGCCUGACAGAAAACACAGGCCAUUCUGAUCACCC CGAAGGGGAGCAAGCCGGCAGGAGCAAGGGUGGCCUUC UUCGAGGUAAAGGGUGGUUGACUUUGGACACCGGUUAAGACAC UGAACCACGGCGCGAGCUGGUCCGGCCAGUAGCAUCAGGAG CGGCUACGUUAACAACCCGAGGCUACAAACGCUUCUCAUC GACGGCUGCGCUGCACAGCGGCAUCCCUACUGGGAGCAGGAGC AGCACUUCUCAUCGUGGACCCGGCUGAAGGUCCAGUAAUAC GGGUUACCCAGGUAGCCCGAGCCGAACUGGGAGAGCAU CAGGCAACACAUUUCGACGCCGGGGUGCCGGCCUGCCGAC AUGCCGGCGAGCUGCCGCCCGCAGUGCUGGUUGGGAAAC AACCAUGACCGAGAAGGAGAUCGUGGACAUUAUGGGCAG AACACCGCCAAGAAGCGCUGGGGGGUUGGUUGUUGU UGCCUAAAGGACUGACCGGCAAGUUGGACGCCGCAAGAU GAUUCUCAUUAAGGCCAAGAAGGGCGCAAGAUCGCCUG 2842 (SEQ ID NO: 85)	SINV replicon	Fluc	AUGGAAAGAUGCCAAAAACAUUAAGAAGGGCCAGGCCAUUCUACC CACUCGAAGACGGGACCGCCGGCAGCAGCUGCACAAAGCAUGAA GGCGCUACGCCUUGGUGCCGGCACAUCCUUUACCGACGCACAU AUCGAGGUUGGACAUUACCUACGCGAGUACUUCGAGAUGGGUUC GGCUGGGAGAACUAUGAAGCGCUAUGGGUGGUAAUACCAAC GAUCGUGGUUGCGAGGAGAAUAGCUUCAGUUCUCAUGCCGUG UUGGGUGGCCUGUCAUCGGUGGUUGGCGUGGUUGGCCAGCUA UCUACAAACGAGCGCAGCUGCUGAACAGCAUGGGCAUCAGCC CACCGUCGUAUUCGUGAGCAAGAAAGGGCUGCAAAGAAC GUGCAAAAGAACGUACCCAGGUACAUACAAAAGAUCAUCAUGG GCAAGACCGACUACCCAGGGCUUCCAAAGCAUGUACAC UUCCCAUUUGGCCACCGGCUUCAACGAGUACGACUUC AGCUGACCGGACAAACCAUCGCCUGAUCAUGAACAGUAGUG GCAGUACCGGAUUGCACAGGGCGUAGCCUACCGCACCG UUGUGUCCGAUUCAGUCAUGCCCGGACCCAUUCUGGCCA AUCAUCCCACGACCCGCUUCAUCAGCGUGGUUGGCC GGUUCGGCAUUGUACCCAGCGUUGGUCAUUGAUC CAGGCCUUUC

TABLE 6-continued

1782 (SEQ ID NO: 86)	mRNA (TEV-XbG)	Fluc	GGUCGUGCUAUGUACCGCUUCGAGGAGGCUAUUCUUGCGCAGC UUGCAAGACUAUAAGAUCAACUUCGCCUGCUGGGGCCACACUAU UUAGCUUCUUCGCUAAGAGCACUCUCAUCGACAAGUACGCCUAAG CAACUUGCACCGAACGGCCAGCGGGGGGCCUGCAAGGAG GUAGGUGAGGCUGGCCAAACGCUUCCACCUACCAGGCAUCCGAC AGGGCUACGCCUGACAGAACACAGGCCAUUUCUGAUCACCCC CGAAGGGGACGACAAGCCUGGCCAGUAGGAAGGGUGGCCUUU UUCGAGGCUAAGGUUGGGACUUGGCAAGGACACUGGGUG UGAACACGGCCGGCAGGCGACUGCCUGUCCGGCCCAUGAUCAUGAG CGGCUACGUUAACAACCCCGAGGUACAAACGCCUCAUCGACAAAG GACGGCUGGCCUGACAGCGGCAGACUCCUACUGGGAGGAGC AGCACUUCUCAUCGGGACUAGGUUGGGACUAGGUACAAUACAA GGGCUACCGGUAGCCAGCCAGCGAACUGGAGAGCAUCCUGUGCAA CACCCCAACAUUCUUCGACGCCGGGUCCGCCUGGCCGACGAG AUGCCGGCGAGCUGCCCGCCAGUGCUGUGCUGGGUACACGGUAA ACCAUGACCGAGAAGGAGAUCGGGACUAGGUUGGCCAGCCAGGUU ACAACCGCCAAGGAAGCUGCCGGUUGGUUGGUUCGUGGACGAGG UGCCUAAGGACUGACCGCAAGUUGGACGCCCAAGAUCCGGA GAUUCUCAUAAGGCAAGAAGGGGGCAAGAUCCGUGUAA
2847 (SEQ ID NO: 87)	STARR TM	KRAS epitope wt	AUGAAGUUGGGGUUGUGGGGGCGGGGUGUUGGCAAAAGGCC UUACAAUUGA
2862 (SEQ ID NO: 88)	SINV replicon	Empty	AUGGAUCCUAGACCCUACGCCCAUAGAUCCGACCAGCAAAACUCG AUGUACUUCGAGGAACUGA
3060 (SEQ ID NO: 89)	STARR TM	Signal peptide- gp70 with AH1A5- MITD	AUGAGAGUGACGCCCUAGAACCUUACUGCUUCUGCUUUGGGAG CUGUUGCUCUGACAGAGAACGGGUGGAUCUCUGAGCAGGUGAC CGGCCAGGGCUGUGCAUCGGGCCGUGCCAAAGACCCACCGUG CUGUGCAACACCCACCCAGAACGACCGAGCGACGGCAGCUACUACUGG CGCUCUCCACCGGCCACCCACGGCCUGCAGCACCGGCCUGACCC UUGCAUCAGCACCAUCCUGAACCCUGACCCAGCACUACUGGGUG CUGGUGGAGCUGUGGCCAGGGUGACCUACACAGCCACGCUACG CCUACACCAAGUUCGAGAGGAGGGCAAGUACAAAGAGGAGGCCU GAGCCUGACCCUGGCCUGCUGUGCCUGAGAACAAUUGGGGGC AUCGCCGCCGCCUGGGCACCGGCCACCCGGCCUGGUGGCCACCC AGCAGUUCAGCAGCUGCAGGCCAGCAUCGACGACCCUGUAGGA GGUGGAGAAGUCAUCCACUCCUGAGAACAGGAGGGCCUGGACCU UAGCAGGAGGGUGCUGCAGAACAGGAGGGCCUGGACCU UGAAGGAGGGCGCCUGCGCUGCCUGAAGGAGGAGGUGCUGCCU GAUCGCCGACCAACCCGGCCUGGUGACUUCGUGGCC

TABLE 6-continued

3061 (SEQ ID NO: 90)	STARR TM	Signal peptide- AH1A5 OVA- MITD	CUGGCGGUCCUCGCCGGGGGGUGAUUAGGAGCUGGGUCGCAGCUG UU AUG UGCAG AAG AAG UCA UC CGCG GAA AGGG AGG CUC UCAGG CUG CU UC UG CU AC AG UG GCU AG AG CUC UUA UAG CUG UAA
3076 (SEQ ID NO: 91)	STARR TM	Signal peptide- gp70 with AH1A5- MITD- FLAG	AUGAGAGUGACAGCCCCUAGAACCUUACUGCUUCUGCUUUGGGAG CUGUUGCUCUGACAGAGACAUGGGCUGGAUCUACCACAGCCCCAG CUACGCCUACCCACCAGAUUCAGCCGGCGUGGCACCGCGCCACCG GGAGGCUCCCUGAAGAACUGCCAGGGGGAGGAGGCUGGGCAGC AGAUACACGAGCCGGCGGGAGGGUGAUUCUGGGCAUUGUCGUG CCUGGCCGUCCUCGCCGGGGAGGGUGAUUCUGGGCAGCAGC GUUAUGUGCAAGAAAAGUCAUCGCCGGCAAAGGGAGGCUCCUAC CUCAGGCUGCUUCUGCUACAGUGGCCUAGAGCUCUUAUGGUUU UCAGCUGUAA
3068 (SEQ ID NO: 92)	STARR	Signal peptide- AH1A5 OVA- MITD- FLAG	AUGAGAGUGACAGCCCCUAGAACCUUACUGCUUCUGCUUUGGGAG CUGUUGCUCUGACAGAGACAUGGGCUGGAUCUACCACAGCCCCAG CUACGCCUACCCACCAGAUUCAGCCGGCGUGGCACCGCGCCACCG GGAGGCUCCCUGAAGAACUGCCAGGGGGAGGAGGCUGGGCAGC AGAUACACGAGCCGGCGGGAGGGUGAUUCUGGGCAUUGUCGUG CCUGGCCGUCCUCGCCGGGGAGGGUGAUUCUGGGCAGCAGC GUUAUGUGCAAGAAAAGUCAUCGCCGGCAAAGGGAGGCUCCUAC CUCAGGCUGCUUCUGCUACAGUGGCCUAGAGCUCUUAUGGUUU UCAGCUGGGCGGGAGGCAGCAGCACUACAAGGACGACGAUGACAAG AA
Transgene ORF amino acid sequence			
mARM #		transgene description	Sequence
2809, 2842, 1782 (SEQ ID NO: 93)	Fluc		MEDA KNIKKGPAPFYPLEDTAGEOLHKAMKRYALVPGTIAFTDAH IEVDITYAEYFEMS VR LAEAMKRYGLNTNHRIVVCSENSLQFFMPV LGALFIGVAVAPANDIYNERELLN SMG I S QPTVV F VSKGLQKILN VQKKLPIIQKIIIMDSKTDYQFQSMTFTVTSHLPPGFNEYDFVPE SFDRDKTIALIMNSSGSTGLPKGV ALP H TACVRF SHARDPIFGNQ IIPDTAILSVPFHGF GMFTTLGYLICGFRVLMYR FEEELFLRS LQDYKIQS ALLV PTLFSFFAKSTLIDKYDLSNLHEIASG GAPLSKE VGEAVAKRFHLPGIRQGYGLTETTSAILITPEGDDKPGAVGVVPF PEAKV DLD TGKTLGVNQR GEL CVRGPMIMS YVN NP EAT NALIDK DGWLHSGDIAYWDEDEHFFIVDRLKSLIKYKG YQVAPAELESILLQ HPNIFDAGVAGLPDD DAGELPAAVV L EH GKT MTE KEIVD YVASQV TTAKKL RG VV FV D E VP KGL T G K L D A R K I R E I L I K A K K G G K I A V *
2847 (SEQ ID NO: 94)	KRAS	epitope wt	MKLVVVGAGGVGKSALTI*
2862 (SEQ ID NO: 95)	empty		MDPRRYAPMIRPAKLDVLPRN*
3060 (SEQ ID NO: 96)	Signal peptide- gp70 with AH1A5- MITD		MRVTAPRTLLLWGAVALTETWAGSLSEVTGQGLCIGAVPKTHQV LCNTTQKTS DG SYYLAAP TGT WAC STG LTP CISTTILNLT TDYCV LVELWPRV TYHS PSYAYHOFERRAKYKREP VS LT ALL LG GLT MG IAAGVGTGTTALV ATQQF QQLQQA MHD DLKEVEKSITN LEK S L TS L SEV VLQNR RGL DLLFLKE GGLC AALKEE CCL YADHT GLV IVGIVAG LAVL AVV VIGAV VAA VMCR RKS SGK GGS YS QAA SAT V P R AL M CLS QL*

TABLE 6-continued

3061 (SEQ ID NO: 97)	Signal peptide- AH1A5 OVA- MITD	MRVTAPRTLLLLLGAVALTETWAGSYHSPSYAYHQFERGGGSGG GGSLKISQAVHAHAEINEAGREVIVGIVAGLAVLAVVIGAVVAA VMCRRKSSGGKGGSYSQAASATVPRALMCLSQL*
3076 (SEQ ID NO: 98)	Signal peptide- gp70 with AH1A5- MITD-FLAG	MRVTAPRTLLLLLGAVALTETWAGSLSEVTGQGLCIGAVPKTHQV LCNTTQKTSQDSYYLAAPTGTTWACSTGLTPCISTTILNLTTDYCV LVEWLPRVTHSPSYAYHQFERAKYKREPVSLTLALLLGGLTMGG IAAGVGTTALVATQQFQLQAAHDDLKEVEKSITNLKSLTSL SEVVLQNRRGLDLLFLKEGLLCAALKEECLLYADHTGLIVGIVAG LAVLAVVVIAGVVAAVMCRRKSSGGKGGSYSQAASATVPRALMCLSQLGGGSDYKDDDK*
3068 (SEQ ID NO: 99)	Signal peptide- AH1A5 OVA- MITD-FLAG	MRVTAPRTLLLLLGAVALTETWAGSYHSPSYAYHQFERGGGSGG GGSLKISQAVHAHAEINEAGREVIVGIVAGLAVLAVVIGAVVAA VMCRRKSSGGKGGSYSQAASATVPRALMCLSQLGGGSDYKDDDK*
whole RNA sequence		
mARM #	brief name	Sequence
2809 (SEQ ID NO: 100)	STARR TM Fluc	AUGGGCGGCCAUAGAGAGAACGCCAGACCAAUUACCUACCCAAA GGAGAAAGUUUCAGUUGAACUGCAGGAAGACAGCCAUCCUAGA GCUUUGCAGCGGAGCUUCCCGCAGUUUGAGGUAGAACGCAAGCAG UCACUGAUAAUGGACCAUGCUUAUGCCAGAGCCGUUUCUGCAUCUG UUCAUAGGAACUGAUCAACGGGAGGACCAUCCGACACGAUCCU GACAUJUGGAAGUGCCCGCCCGAGAAUGUAAUCUAAGCACAGAU AUCAUUUGUAUCUGCGAUGAGAUUGCGGAAGAUCCGGACAGAU GUUAAGUAUGCAACUAGCUGAAGAAAAACUGUAAGGAAUACU GAUAAGGAUUUGACAAGGAAAAGAUGGGAGCUGGCCGCUAUGA GCGACCCUGACGGGAAACUGAGCAUAUGUGCCUCCACGACGAGA GUCGUGUCGCUACGAAGGGCAAGUCGCUGUUUACCAGGAUGUAUAC GCCGUGCAGGCCACCAGCCUACCCAGGCAACAAGGGCAACAAGGGC UGAGGGUGGCCUACUGGAUCGGCUUCUGACACCACCCUUCAGUU CAAGAACCCUGCCGGCCUACCCAGCUACAGCACCAACUGGGC GACGAGACCCUGCCUACGGCAAGGCCAGGCUACUGCCUACAGCAGG ACGUGAUGGAGAGGGAGCCGGAGAGGAUGAGCAUCCUGAGGAAGA AUACCUUGAAGCCAGCAACAGCUGCUGUUCAGCGUGGGCAGGAC AUCUACACAGGAAGAGGGACCUUGCUCAGCAGCACCCUUGCC GCGGUUCCACUGGGCAAGGGACUACCCAGCUACAGCACCAACUG GACCAUCGUGAGCUGCGACGGCUACUGGGGAAGAGGAUCGCCAUC AGCAGACCCGGCCUACGGCAAGGCCAGGCUACUGCCUACAGCAGG ACGUGAUGGAGAGGGAGCCGGAGAGGAUGAGCAUCCUGAGGAAGA AUACCUUGAAGCCAGCAACAGCUGCUGUUCAGCGUGGGCAGGAC GCGGUUCCACUGGGCAAGGGACUACCCAGCUACAGCACCCUUG GACCAUCGUGAGCUGCGACGGCUACUGGGGAAGAGGAUCGCCAUC ACAGGGAGCCGGCUUCCUGCCUACGGCAAGGCCAGGCUACAGCAGG CGAGAGGGUGAGCUGCCUACGGCAAGGCCAGGCUACUGCCUACGG UGCGACCCGGCCUACGGCAAGGCCAGGCUACUGGGGAAGAGGAUC ACGUGAUGGAGAGGGAGCCGGAGAGGAUGAGCAUCCUGAGGAAGA CGGAGACCCAGGGGGCUAAGGGAACGGGAGAGGAUGGGCUACUG CCCUGGGGCCAGCCUUCGGCAGGUGGGCAAGGAGUACAGG AGGACCAAGGAAGACGAGAGGCCUUGGGCUGAGGGACAGGAC GGUGAUGGGCUGCUGCCUUCAGGGCAAGGAGUACAGGAC AUCUACAGGAGGGCCGACACCCAGACAUCAAGGUGAACAGCG ACUUCACAGCUUCGGCUGCCUACGGAUCCGCGACAAACCCUUG GAUCGGCUGAGGAGCCGAUCAGGAACAGGAUGCUGGGAGGAACAG GAGCCAGGCCACUGAUCAACGGCAGGAGAACAGGAC GGCUGCCGACGAGGCCAAGGAGGGAGGGAGGGCAGGAC GGCCGCCCCUGCCACCCUGCCUGCCAGCAGGGAGGGAC GAAGCCGACGUGGACCUAGCUGCCAGGAGGGCAGGGCGGGAGCG UGGAGACACCCAGGGGGCUAAGGGAACGGGAGGAC GGACAAAGAUCCGGCAGCUACGGCUGCCAGGAGGGCAACCCUUG AAGUGCGAGAACUGAGCUGACGGCAGGAGGGAGGGAGGGAG UCGUGAUCCCCACAGCGGCAGGAAGGGAGGUACGGCUGGGAGCC CUACCAAGGGCAAGGUGGUGCCUGGGCAGGGCAACCCUUG CAGGACUUCAGGCCUACGGGAGGGAGGGAGGGAGGGAG AGAGGGAGUUUCGGUAGAACGGUACCCUGACCAUAGCCACCCACGG CGGAGCCUGAACACCGACGAGGAUACUAACAGGCGAGGCC AGCGAGCACGACGGCAGGUACCGUAGCAGAACGGAGAG GGGUGAAGAAAGAGCUGGGAGACGGCCUGGGCAGGAC GGUGGACCCACCUUCCAGGAGGUACGGCUGGGAGGGAG AGACCCGGCCUCCUACAGGGGCCACCAUCCGGCGUACGGC UGCCCGGCAGGGAAAGAGCCGCAUCAAGAGGGCCGUGAC GAAAGACCUUGGGUGUCAGCGCCAAGAAAGAGAACUGCGCCGAG AUCAGGGACGUGAAGAAGAUGAAAGGGCUGGACGGUAGAC CCGUGGACAGCGUGCUGCUAGCGCUGCAAGCACCCGGAG CCUGUACAGCACGAGGCCUUCGGCAGGCCACCCUUG GCCUGAUCGCCAUCAUCAGGCCAAGGAAAGCCGUGUGCC ACCCCAAGCAGUGCCUUCUCAACAGAUGUGCCUGAAGGUG CUUCAACCACGAGAACUGCAGGCCAGGUGUUCACAAGAGCAUCAGC

TABLE 6-continued

TABLE 6-continued

2842
(SEQ
ID
NO: 101)

TABLE 6-continued

3060 STARRTM 3060
(SEQ gp70
ID
NO: 105)

TABLE 6-continued

TABLE 6-continued

CCAUCCUGUUUCAGCCCCAAGGGUGCCCGCUACAUCCACCCCAGGAA
 GUACCUGGUGGAGACCCCCACCCGUGGGACGAGACACCCGAGCCAAGC
 GCCGAGAACAGAGCACCGAGGGCACACCCGAGCAGCCACCCCUGA
 UCACCCAGGAGAGACAAGGGACCCGACCCGAGGAGCCAUCAUUAU
 CGAGGAAGAGGAAGAGGAGACGCAUCAGCCUGCUGAGCGAGGGCCC
 ACCCACCCAGGUGCUCAGCUGGGAGGGCGACAUCACAGGCCACCCCA
 GCGUGUCCAGCUCUCCAGCUGGAGCAUCCACACGCCAGCAGCAGUUGA
 CGUGGACAGCCUGAGCAUCCUGGACCCUGGGAGGGGCCAGGGUG
 ACCUCGGGCCACACAGCCGAGACAAACAGCUACUUCGCCAAGA
 GCAUGGGAGUUUCUGGCCAGGCCUGCCAGCUCUCCAGGACCGUUGU
 CAGGAACCCACCCACCCAGCUCUCCAGGACAGGAGCCAGGGCCAC
 GCUCGGCAGGGCAGGGCCUGCAGGACAGCCUGGGAGGAGCACCCAC
 CCGCGUGUAACAGGGUGAUACACAGGGAGGAACUGGAGGGCCUGAC
 ACCCAAGCAGGACCCCCAGCAGGUCCUGAGCAGGACAUAGCUGGUG
 UCCAACCCACCCGGUGAAGAGGGUGAUACACAGGGAGGAUUCG
 AGGCCUUCUGUGGCCAGCACAGGAGACGGUUCGAGCCGGGCCUA
 CAUCUUCAGCAGCGCACCCGGCAGGGACACCUCUGCAGCAAAAAGAGC
 GUGAGGCAGACGGCUGUGAGCGAGGGUGGUCCUGGAGAGGACCGAGC
 UGAAAUACAGCUACGCCAGGCGUCCAGGAGAAGGGAGAACU
 GCUCAGGAAGAAACUGCAGCUGAAACCCACCCAGCCAAACAGGAGC
 AGGUACCCAGAGCAGGAAGGGAGAAACAGAAGGGCAUCACCCCA
 GCGGAUCCUCCUGGGCCUGGGACACUACCUAAGGCGAGGGGCAA
 GGUGGAGUGUCAACAGGACCCUGCACCCGUGCCACUGUACAGCUCC
 AGCGUGAACACAGGCCUUCUCCAGGCCAACAGGGAGGGCCUGG
 GCAACCGCUAUGCUGAAGGGAGAACUUCCCACCGUGGGCAGCUACUG
 CAUCAUCCCCAGAUACGGCUCUCCUGGACAUUGGGAGCGGCC
 AGCUGCUGCCUGGACACCGCAGCUUCUGGCCGCAAGCUGAGGA
 GCUCUCCCCAAGAAACACAGCUACCCUGGAGCCACCAUACAGGAGGCC
 CGUGCCAGCGCAUCCAGAACACCCUGCAGAACUGCUGGCCGCU
 GCCACCAAGAGAACUGCAACUGACCCAGAUGAGGGAGCUGGCCG
 UGCUGGACAGCCUGCUUCAACGGUGAGUGCUUCAAGAAAACGC
 CUGCAACAAACAGGUACUGGGAGACCUUCAAGGAGAACCCCAUCAGG
 CUGAGCGAAGAGAACGGUGUACAUACCAACAGCUGUAAGGGCC
 CCAAGGGCCGUGGCCUGGUUCGCUAAGGACCCACACCCUGAACAUGC
 GCAGGACAUCCAAUGGACAGGUUCUGUGAUGGACCUAGAGGGAC
 GUGAAGGUGACACCCGGCACCAAGCACACCGAGGGAGGGCCAAGG
 UGCAGGUGUAUCCAGGGCUGACCCACUGGCCACCGCUACCCUG
 CGGCAUCCACAGGGAGCUGGGAGCGCUGAAGCCGGCUGCUG
 CCCAACUCCACACCCUGUUCGACAUAGGCGCAGGGACUCCGAC
 CCAUCAUCGCCAGCACUCCAGGCCGGCACUGCUGCUGGUAGAC
 CGAACUUCGCCAGCUUCGACAAAGAGGGAGGGACCCUACCCUG
 ACCGUCUCAUGAUCGGACCCUGGGAGGACCCUGGGCUGGGAGCGC
 UCACCCUGUAUCGAGGCGUGCCUUCGGCGAGACUGCUUCAACCU
 GCCCACCACAGAACUUAAGUUCGGCGCAUAGAUGAAAAGCGGA
 AUGUUCUGACCCUGUUCGUGUACACCCUGGAGACCCACAUUUGUGAUC
 CCAGCAGGGUGCUCGGGGAGAGGGCUGACCCGACCCCGUCCG
 CUUCAUCGGCACGACAAACUGUGAAGGGCGUGAAAAGCGAACAG
 CUGAUGGCGCACAGGUGCGCACCCUGGCGUAAACAGGAGGUGAAGA
 UCAUCGACGCCGGGGCGAGAAGGGCCCCUACUUCUGCGGG
 AUUCAUCUCCUGCGACAGCGUGACCCGACCCUGCUGGGGCC
 GACCCUGAAGAGGGCGUUCAGCUGGGCAAGCCACUGGCCGUG
 AGCAUGAGCACGACGAUGACAGGGGAGGGCCUGCACGAGGAAG
 CACCAAGGGGACAGGGGGGCAUCUGAGCAGCAGGGUGCAAGGCC
 GUGGAGAGCAGGUACAGGACGGGGCACAGCAUCAUCGUGAUGG
 CUAUGACCAACUGGCCAGCUCUGCUAAGGCGUUCUCCUACCU
 GGGGGCCCUUAACUCUCAACGGCUAACCUGAAGGGACUACGACA
 UAGCUAUGUCCCAAGGGCCGCAACCAUGAGAGUGACAGCCCUAG
 AACCUUACUGCUUCUGCUUUGGGAGCUGUUGCUCUGACAGAGACA
 UGGGCUUGGAUCUCUGAGCAGGGGACCCGGCCUGUGCAUCUG
 GCGCGUGCCCAAGACCCACAGGGGCGUGCGUACACACCAC
 GACCAAGCAGCGCACUACUACCCUGGGCGUCCACCCGGCAC
 UGGGCCUGACCCGACCCUACCCCGUCCUUGCAUCAGCACCA
 UGAACCGUACCCGACUACUGCGUGCCUGGGAGGAGAC
 GGUUGACCUACACAGCCCGACUACGGCCUACCCACAGUUCGAGAGG
 AGGGCCAAGUACAAAGAGGGAGCCGUGAGCCUGACCCUGGCC
 UGCUGGGCGGCCUGACAAAGGGGGCGACUCCGCGGGCGUGGG
 CGGCCACACCCCGUCCGGCAACCCAGCAGUUCAGCAGCUG
 GCCGCAUGCACGACCCGACUACGGGGAGGAGAGAC
 ACCUGGAGAACUCCUGACCCGUGAGCCUGACCCUGG
 CAGGAGGGGGCCUGGACCCUGCCUGGUACAGGGGGCGGCC
 GCCGCCUCAAGGAGGGAGUGCCUGGUACGCCGACCC
 UGGUGAUCCUGGGCAUUGUCGCUUGCCUGGGCGUCC
 GGUGAUUGGAGCUGUGUGCCAGCGUGUUUAUGUGCAGAAGAAA
 UCCGGCGGAAGGGGGAGGGCUCCUACUCUGAGGCGUGU
 UGCCUAGAGCUCUUAUGGUUUUAUCUGAGCUGUAACUC
 UUACUGGCAAAGGUGAUUGUACACCCCGAAAGAC
 ACACCCUCAUGUAUCAGGCCAACAUUUACAGGCCGG
 ACACCCUCAUGUAUCAGGCCAACAUUUACAGGCCGG
 UUAAAUAUAAAUGGUCCUGGGCUGGUACUAAUUGGG
 GCUGACCAACCAGAAACAUAAAUGGAAACAGCAGCA
 UUUGGCAAGC

TABLE 6-continued

TABLE 6-continued

AACACCGGCACCCUGAGGAACUACGACCCCCAGGAUCACCUGGUGC
 CCGUGAACAGGGCGCUGCCCCACGCCUGGUGCUGCACCAACGA
 GCACCCACAGAGCGACAUUCAGCUCCUUCGUGAGCAAGCUGAAAGC
 AGGACCGUGCGUGGGCAGAGAAGCGAGGUGGCCAGGCAAGA
 UGGUGGACUGGCGAGCGACAGGCCAGGGCAGGUCCAGGCAAGA
 GCUGGACCUCCGCAUCCCCGGCAGCUGGCCAGUACGACAUCAC
 UUCGUGAACGUACAGGACCCCCAUACAAGUACCAACCAAGCAG
 CGCAGGACCAACGGCAUCAAGCUGAGCAGCUGGCAAGAAGGGCUG
 CCUGCACUAGACCCCGAGGCCACCCUGUGAGCAUCCGUACCGC
 UACGCGACAGGCCAGGAGAGCAUCAUUGGCCAUUCGCCAGGC
 UGUUCAAGUUACAGCAGGGUGUGCAAACCCAAGAGCAGGCCUGGAGGA
 AACCGAGGUGUGCGUGUUCGUGUCAUCGGCUACGACGGGAAGGCAGG
 ACCCACAAACCCUACAAGCUGAGCAGCAGGCCACCCUGACAAACAUACA
 CCGCAGCAGGCCUGCACGCCAGGCCAGGCCAGGUACACCACGU
 GGUCAAGGGCGAUAUCGCCACCGCCACCGAGGGCGUGAUCAUACA
 GCUGCCAAACAGCAAGGGCAGGCCAGGGAGGGAGUGGCCAGGCC
 UGUAAGAAAGUUCCCGAGAGCUCUGGCCAGGCCAUUCGAGGU
 GGGCAAGGCCAGGUGUGAGGGCGGUACAGCACAUCAUCCAC
 GCCGUGGGCCCAACUUCAACAAGGUGAGCAGGAGGUAGGCAGA
 AGCAGCUGGCCAGAACGCCUACGAGGACAUCCGCAAGAUCUGUGAAGA
 CAUAACUACAAGAGCUGGCCAUCCACUGCUGACAGCAGGCCAUC
 UUCAGGGCAACAAAGGACAGGCCAGGCCAGGCCAUUCACUGC
 UCACCGCCUCCUGGACACCACCGAUGCCGACGUGGCCAUACUGCAG
 GGACAGAAGUGGGAGAUGACCCUGAAGGAGGGCCUGGCCAGGCC
 GAGGCCUGGGAAGAGAACUGCAUCAGCGACGACUCCAGCUGACCG
 AGCCCGACGCCAGCUGGGUGAGGGUGCACCCCAAGAGCUCCCUGGC
 CGGCAGGAAGGGCUACAGCACCCAGCAGGCCAGGCCAUUCAGCUAC
 CUGGAGGGCACCAAGUUCACCAGGCCGUAAAGGACAUUCGCCAGA
 UCAACCGCUAUGGGCCUGGGCAGGCCACAGGAGGGCCUGGCCAG
 CAUGUACAUCCUGGGGAGAGCAUUGCCAGCAUCAGGAGCAUGUC
 CCCUGGGAGGAAGGGCAGCACACCACCCAGCACCCUGGCC
 GCCUGUGCAUCCACGCUAUGACACCCGAGGGUGUGCCAGCGCUGA
 GGCAGCAGGCCAGGAGCACAGGCCAGGCCAGGCCAUUCAGGCCA
 CUGGAGGGCACCCAGGCCACAGGCCAGGCCAGGCCAUUCAGGCCA
 UCAACCGAGGAGAGGAGCACAGCAUCAGCCUGCAGGCCAGGCC
 ACCCACCAAGGUGCUGCAGGUGGGAGGCCACUCCAGGCCA
 GCGUGUCCAGCUCAGCUGGGAGCACCCACAGGCCAGGCCAUUC
 CGUGGACAGCCUGAGCAUCCUGGACACCCUGGAGGGCCAGCGUG
 ACCUCGGCGCACAGCGCCAGACAGCUACUUCGCCAAGA
 GCAUGGAGUUCUGGCCAGGCCAGGCCAGGCCAGGCCAGGU
 CAGGAACCCACCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCC
 ACCCACCAAGGUGCUGCAGGUGGGAGGCCACUCCAGGCCA
 GCGUGUCCAGCUCAGCUGGGAGCACCCACAGGCCAGGCCAUUC
 CGUGGACAGCCUGGAGCACCCAGGCCAGGCCAGGCCAUUCAGC
 ACCAGCAGGACCCCCAGCAGGCCAGGCCAGGCCAGGCCAGGCC
 UCAACCCACCCGGCGUGAACAGGGUGUACACCCAGGCCAGGCC
 AGGCUUUCUGGGCCAGCACAGGAGAGACGGUUCAGCAGGCC
 CAUCUUCAGCAGCGCACCCGGCAGGGACACCCAGGCCAGGCC
 GUGAGGCAGCCAGCUGAGCAGGUGGGUGCUGGAGAGGCCAGC
 UGAAAACAGCUACGCCAGGCCAGGCCAGGCCAGGCCAGGCC
 GUCAGGAAGAAACUGCAGCUACCCAGGCCAGGCCAGGCC
 AGGUACCCAGCAGGAGGGAGAACAGGUACAGGCCACCC
 GCGGAUUCUGCAGGGCUGGGACACUACCUGAAGGCCAGGCC
 GGUGGAGUGCUACAGGCCAGGCCAGGCCAGGCCAGGCC
 AGCGUGAACAGGCCUUCUCCAGGCCAGGCCAGGCCAGGCC
 GCAACCGCUAUGCUGAGGAGAACUCCCCACCGCUGGCCAGC
 CAUCAUCCCGAGUACAGGCCUACCCUGGACAUCCUGGAGGCC
 AGCUGCUGCCUGGACACCCAGGCCAGGCCAGGCCAGGCC
 GCUUCCCCAGAACACAGCUACCCUGGAGGCCACAGGCC
 CGUGCCAGGCCCAUCCAGAACACCCUGGAGGCCAGGCC
 GCCACCAAGGAGAACUGCAACGUGACCCAGAGGAGGCC
 UCGUGGACAGGCCUGCCUACAGGCCAGGCCAGGCCAGGCC
 CUGAGGAGAGAACAGGUAGGUAGACAUCAUCACAGC
 CCAAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCC
 GCAAGGACAUCCAAAGGGACAGGUUCUGUGAGGCCAGGCC
 GUGAAGGUGACCCGGCAGGACAGCACCCAGGCCAGGCC
 UGCAGGUGAUCCAGGCCUGCCACGGGCCACCCGGC
 CGGCAUCCACAGGGAGCUGGUAGGGGGCUGAAGGCC
 CCCAACUCCACACCCUGUUCGACAGGCCAGGCCAGGCC
 CCAUCAUCGCCAGCACUCCAGGCCAGGCCAGGCC
 CGAACAUCGCCAGCACUCCAGGCCAGGCCAGGCC
 ACCGUCUGAUGAUCCUGGAGGACCCGGCUGGU
 UCACCCUGAUCCAGGCCUGCCUACAGGCCAGGCC
 GCCACCAAGACCAAGUUAUGUUCGGCGU
 AUGUUCUGACCCUGUUCGUGAACACCC
 CCAGCAGGGUGCUGGGAGAGGCC
 CUUCAUCGCCAGCACAAUCUG
 GUGAAGGGCUGAAAAGGCCAGAAG

TABLE 6-continued

3067 STARRTM 3067
(SEQ gp70 -
ID FLAG
NO: 107)

TABLE 6-continued

TABLE 6-continued

3068 STARRTM 3068
(SEQ AH1A
ID 5-
NO: 108) FLAG

TABLE 6-continued

TABLE 6-continued

UCACCGCCUGGACACCACCGAUGC CGAC GUGGCCAU CUAC UGCAG
 GGA CAAGAAGUGGGGAUGACCCUGAAGGAGGGCGUGGCCAGCGG
 GAGGCGCUGGAAGAGAACUGCAUCAGCGACGACUCCAGCGUGACCC
 AGCCCCAGCGGAGCUGGUGAGGGGUGCACCCCAAGAGCUCCUUGC
 CGG CAGGAAGGGCUACAGCACCGACGGCAAGACCUUCAGCUAC
 CUGGAGGGCACCAAGUOCCACCAAGCCGCUAAGGACAU CGCCGAGA
 UCAACGCUAUGUGGCCGUGGCCACCGAGGGCAACGAGCAGGUG
 CAUGUACAUCCUGGGCAGAGCAUGUGCAUCAGGAGCAAGUG
 CCCUGGGAGGAAGGGAGGCCAGCACACCCACGACCCUGCC
 GCCUGUGCAUCCACCGCUAUGACACCCAGAGGGUGCAGCGCUGA
 GGCAGCAGGCCAGCAGCAUCACCGUGUGCAGCUCCUCCACUG
 CCAAUQUACAGGAUACCCGGCUGCAGAAGAAGUCCAGUGCAGCAGC
 CCAUCCUGUUCAGCCCAAGGUCCGCUAUCACCCACCCAGGAA
 GUACCUUGGGAGACCCCCACCCUGGAGCAGACCCAGGCAAGC
 GCGAGAACAGAGCACCGAGGGCACACCCAGCAGCCACCCUGA
 UCACCGAGGAGAGACAAGGACCCGGGAGCCAGGCAUCAUUAU
 CGAGGAAGAGGAAGAGGAGCAGCAUCAGCCUGAGCGACGGCCC
 ACCACCCAGGUGCUGCAGGGAGGGCGCAUCCACGGCCACCC
 GCGUGUCCAGCUCAGCUGGAGCAUCCACCGCCAGCGACUUCG
 CGUGGACAGCCUGAGCAUCCUGGA CACCCUGGAGGGCGCAGCG
 ACCUCCGGCGCACCCAGCGAGGACCAACAGCUACUUCGCGAAGA
 GCAUGGGAGUUCCUGGCCAGGGCGUGCCAGCUCCAGGGACCGU
 CAGGAACCCACCCACCCAGCUCCCAGGACAGGACCCAGGCG
 GCUCCCAGCAGGCCUGCAGCAGGACAGCCUGGUGAGCACCCAC
 CCGGGUGGAACAGGGUGAUACACAGGGAGGAACUGGAGGGCCUGAC
 ACCCAGCAGGACCCAGGGUGAGCAGGACUAGUGUGUG
 UCCAACCCACCCGGCGUGAACAGGGUGAUACACAGGGAGGAUUCG
 AGGCCUUCGUUGGCCAGCACAGGAGCGGUUCGAGCCGGCGCC
 CAUCUUCAGCAGGCCACCCGGCAGGGACACCUGCAGCAAAAGAGC
 GUGAGGCAGACCCUGCUGAGCGAGGGUGGUCCUGGAGAGGACCG
 UGAAAUACGUACGCCAGGGUGACCAGGAGAAGGGAGAAC
 GCUCAGGAAGAAACUGCAGCUGAACCCACCCAGCCACAGGAGC
 AGGUACAGAGCAGGAAGGGUGAGAACAUAGAAGGCAUACCCG
 GCGGAUCCUGCAGGGCUGGAGCACUACCUGAAGGCCAGGGCAA
 GGUGGAGUGCUACAGGACCCUGCACCCUGGCACUGUACAGCU
 AGCGUGAACAGGCCUUCUCCAGGCCAACGGUGGCCAGGAGGCC
 GCAACCGCUAUCUGAGGAGAACUCCACCGUGGCCAGCUACUG
 CAUCAUCCCGAGAUACGAGCCUACUCCAGGGACAUUGGGAGCG
 AGCUGCUGCCUGGAGCACCGCCAGCUUCUCCUGGCCAGCUGAG
 GCUUCCCAAGAACACAGCUACCCUGGAGGCCACCAUAGGAGCG
 CGUGCCAGGCCAUCCAGAACACCCUGCAGAACAGCGCUGGCC
 GCAACAGAGGAACAGCUGACGGCCAGAUGAGGGAGCUGCCCG
 UGCUGGACAGCCUGCCUUCAACUGGGAGUGCUUCAAGGAGAAC
 CUGCAACACAGAGUACUGGGAGACCUUCAAGGAGAACCCAUAG
 CUGACCCGAAGAGAACUGGGUGAACUACACCCAGCAGGAGGCC
 CCAAGGCCUCCUGUUCGUAGACAGGGAGAACCCACACCGUACAU
 GCAGGACAUCCCAUUGCAAGGUUCGUAGGGACCUAGAGGGAC
 GUGAAGGUGACACCCGGCACCAAGCACCCAGGAGGAGGCCAAG
 UGCAGGUGUAUCCAGGGCGUGACCCACUGCCACCGCCUACUG
 CGGCAUCCACAGGGAGCUGGGAGGCCUGAAGCCUGCUGCUG
 CCAACAUCCACACCCUGUUCGUAGACAGGCCAGGGACUUCGAG
 CCAUCAUCGCCAGCACUCCAGGCCAGCUGCUGCUGGAGAC
 CGAACUGCCAGCUUCGUAGAACAGGGCAGCCUACUGCCUG
 ACCGUCUCAUGAUCCUGGGAGAACUCCAGGGCUGGAGCGC
 UCACCCUGAUCCGAGGCGUGCCUUCGGCGAGAUCAUCUCCAC
 GCCCACCAGACAAAGUUCAGGUUCGGCGCUAUGAUGAAAGCG
 AUGUUCUGACCCUGUUCGUAGAACACCGUGUACACAUUUGUGA
 CCAGCAGGGUGCUGCGGGAGAGGGCUGACCCGGCAGCCCCUG
 CUUCAUCGCCAGCACAAUCUGCAAGGGCGUGAAAAGCGCAAG
 CUGAUGGCGACAGGUGGCCACCCUGGCGUAGAACUAGGAGG
 UCAUCGACGCCUGGGUGGGCGAGAAGGGCCCCUACUUCUGCG
 AUUCAUCUGCCAGCGUGACCCGGCAGCCUGCAGGGUGGCC
 GACCCCGUGAAGAGGGUGUUCAGCUGGGCAAGGCCACUGGCC
 ACGAUGAGCACGACGAUGACAGGGCGAGGGCCUGCAGGAGGA
 CACCAAGGGGAGAACAGGGUGGGCAUCCUGAGCAGCUGUG
 GUGGAGAGCAGGUACGAGACGGUGGGCAGGAGAACUACUG
 CUAUGACCAACUGCCAGCUCGUCAAGACUUCUCCAUCCUG
 GGGGGCCCUAUAACUCUCAUGGCUAACCUGAAGUGGACUAC
 UAGGUCAUGUCCGCAAGGCCACCAUAGAGAGUGACAGCC
 AACCUUACUGCUUCUGGUUUGGGAGCUGUUGCUCUGACAG
 UGGGUGGUACUUAACACAGCCCGAGCUAGCCUACCCAC
 AGAGGGGGGAGGAGGCGUCCGGGGAGGGCGUCCUGAAGAC
 CCAGGCCUGCACGCCAGGCCAGCGCAGAACACAGGGCG
 GAGGGUGAUCUGGGCAUUGUGCGCUGGCCUGGGCGUCC
 UGGUGAUUGGGAGCUGGGUGCGAGCUGUUAUGUG
 AUCCGGCGGAAGGGAGGCGUCCUACUCUCAAGGCG
 CUGGUACAGCUCUUAUGGUUUAUCUCAAGCUGGGCG
 GCGACUACAAGGAGCAGCAUGAAGAAGUAAACUG
 GCAAGGGUGAUUGUCACCCCCGAAAGACCAUAUUGUG
 ACACCC
 UCAGUAUCAGCCCAAACAUUAACAGGCCGGUGU
 AAAACCGCG

TABLE 6-continued

		UGGACGUGGUUAACAUCCCUGCUGGGAGGAUCAGCCGUAAAUAUA UAUUUGCUUGUGCUGGUACAUUUGUGGCCAUGUACGUGCUGAC CAACCAGAACAUAAUUGAAUACAGCAGCAAUUGGCAAGCUGCUUA CAUAGAACUCGGCGGAUUGCAUGCCGCCCCAAAUUUUUUUUUU AUUUUUUUUUUUUUUUUCCGAUUCGGAUUUUUGUUUUUAUUUCA AAAAAAAAAAAAAAAUUCUAGAAAAAAAUAAA AAAAAAAAAAAAAAAUAAA AAAAAAAAAAAAAAAUAAA AAAAAAAAAAAAAAAUAAA AAAAAAAAAAAAAAAUAAA non structural protein of SINV
mARM #		
2842 and 2862 (SBQ ID NO: 109)	SINV nsP1- 4 AA	MEKPVVNVDPQSPFVVQLQKSFPQFEVVAQQVTPNDHANARAFS HLASKLIELEVPTTATILDIGSAPARRMFSHQYHCVCPMRSPEPD DRMMKYASKLAEKACKITKNKLHEKIKDRLTVLDTPDAETPSILCFH NDVTCNMRAEY SVMQDVYINAPGTIYHQAMKGVRTLYWGFDTTQF MFSAMAGSYPAYNTNWADEKVLEARNIGLCSTKLSERGTGKLSIMR KKEKLPGSRVYFSVGSTLYPEHRSLSQSWHLPSPVFLNGKQSYTCR CDTVVSCEGYVVKKITISPGLGETGVYAVTHNSEGFLCKVTDTV KGERSFPPVCTYIPATICDQMTGIMATDPSDPAQKLLVGLNQRIV INGRTNRNTNTMQNYLLPIAQGFSKWAKERKDDLNEMKLGRTRER KLTYGCLWAFRTKKVHSFYRPPGTOTCVKVPASFSAPPMSVWTS LPMSSLRQKLKLALQPKKEEKKLQLQVSEELVMEAAKAFADEAQEEARAE KLREALPPLVADKGTEAAAABVCVEVGLQADIGAALVETPRGHVRI IPQANDRMIGQYIVVSPNSVLKNAKLAQAHPLADQVKIITHSGRSG RYAVEPYDAKVLMPAGGA伟WPPEFLALSESATLVYNEREFVNRLKY HIAMHGPANKTEEEQYKVTKAELAETEYVFVDVKRKCVKKEEASGL VLSGELTNPPYHELALEGLKTRPAVYKVETIGVIGTPGSGKSAII KSTVTARDLVTSGKENCREIEADVLRLRGMQITSKTVDSVMLNC HKAVEVLYVDEAFACHAGALLALIAIVRPRKKVLCGDPMQCGFFN MMQLKVFHNPEKDCITKTFYKISRRCQTQPVTAIVSTLHYDGKMK TTNPCKNNIEDITGATKPKPGDIILTCFRGWVKQLQDYPGHEVM TAAASQGLTRKGVYAVRQKVNPENLYAITSEHVNLLTRTEDRLVW KTLQGDPWIQLTNIPKGNFQATIEDWEAEHKGIIAAINSPTPRAN PFSCKTNVCWAKALEPILATAGVILTGQWSELPFQFADDKPHSAI YALDVICIKFFGMDLTSGLSKQSIPLTYHPADSARPVAHDNSPG TRKYGYDHAAELSRRFPVFLQAGKGTQLDLQTRTRVISAQHNL VPVNRNLPHALVPEYKEKQGPVEKFLNQFKHHSVLLVSEEKIEAP RKRIEWIAPIGIGIAGADKNYNLAQFGPPQARYDLVFNIIGTKYRNHH FQOCEDHAATLKTLSRSALNCLNPGGLVVKSYGYADRNSEDVVA LARKFVRVSAARPDCVSSNTEMYLIFRQLDNSTRQFTPHLNVC SSVYEGTRDGVAAPSYRTKRENIADCQEEAVVNAANPLGRPGEGV CRAIYKRWPPTSFTDSATETGSTARMTVCLGKVKVHAVGPDPRKHEPEA EALKLQLQNAHYAVADLVNEHNIKSVAIPLLSTGTYAAGKDRLEVS NCLTTALDRTDADVTIYCLDKKWRERIDAQLQKESVTELKDEME IDDELWVIIHPSCLKGRKFSTKGKLYSYFEGTKFHQAAKDMAEI KVLFPNDQESNEQLCAYIIGETMEIRKECPVDHPNSSPPKTLPC LCMYAMTPERVHLRSNNVKEVTCSSTFLPKHKIKNVQKVQCTKV VLFNPHTPAFPV PARKYIEVEPEQPTAPPAAQEAEAPEVVATPSPTAD NTSLDVDTSISLMDDSSEGSLFSSSGSDNSITSMDSWSSGPSSLE IVDRRQVVAADVHAQEPAPIPPRLKKMARLAARKEPTPPASNS SESLHLSFGGVSMSLGSIFDGETARQAAVQPLATGPTDVPMSSFGSF SDGEIDEALSRRVTESEPVLFGSFEPGEVNSIISRSRASVFPLRKQR RRRRSRRTEY*LTG VGGY IF STDTGPGHLQKKSVLQNQLTEPTLER NVLERIHAPVLDTSKEEQLKLRYQMPTEANKSRYQSRKVENQKAI TTERLLSGLRLYNSATDQPECYKITYPKPLYSSSV PANYSDPQFAV AVCNYYLHENYPTVASYQITDEYDAYLDMV DGTVA CLDTATFCPAK LRSYPKKHEYRAPNIRSAVPSAMQNTLQNVLIAATKRNCNVTQMRE LPTLDSATFNVECFRKYACNDEYWEEFARKPIRITTEFVITYAARL KGPKAAALFAKTYNLVPLQEVPMDFRVMMDMKRDVKVTPGTKHTEER PKVQVIQAAEPLATAYICGIHRELVRRLTAVLLPNIHTLFDMAED FDAIIAEHFQGDPVLETDIASFDKSQDDAMALTGLMILEDLGVDQ PLLDLIECAFGEISSTHLPTRFRKGAMMKS GMLTLPVNTVLNV VIASRVLEERLKTSCA AFIGDDNIIHGVS DKEMAERCATWLNE VKIIDAVIGERPPYFCGGFILQDSVISTACRVADPLKRLFKLGKPL PADDEQDEDRRRALDET KAWFRVGITGTLAVAVTTRYEVNDNITPV LLALR TFAQS KRAF QAIRGEIKHLYGGPK

Example 2

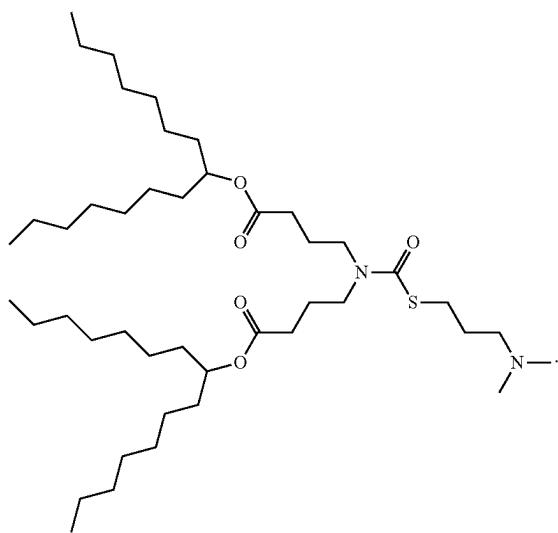
This example describes analysis of the immunogenicity of influenza hemagglutinin (HA) expressed from self-replicating RNA or mRNA.

Self-replicating RNA and mRNA vaccine constructs were designed to encode the full-length hemagglutinin (HA)

⁶⁰ protein from influenza virus A/California/07/2009 (H1N1) (SEQ ID NO:113 and 114). The mRNA vaccine construct encoding HA included a tobacco etch virus (TEV) 5' UTR and a *Xenopus* beta-globin (Xbg) 3' UTR. Both self-replicating RNA (SEQ ID NO:56; entire RNA mARM3039) and mRNA vaccine constructs (SEQ ID NO:116; entire RNA sequence mARM3038) were encapsulated in the same lipid

185

nanoparticle (LNP) composition that included four lipid excipients (an ionizable cationic lipid, 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), cholesterol, and PEG2000-DMG) dispersed in HEPES buffer (pH 8.0) containing sodium chloride and the cryoprotectants sucrose and glycerol. The N:P ratio of complexing lipid and RNA was approximately 9:1. The ionizable cationic lipid had the following structure:



Five female, 8-10 week old Balb/c mice were injected intramuscularly with 2 mg of mRNA or self-replicating RNA encoding HA. Mice were bled on days 14, 28, 42, and 56, followed by hemagglutination inhibition (HAI) assay using serially diluted sera. The reciprocal of the highest dilution of serum that caused inhibition of hemagglutination was considered the HAI titer, with a titer of 1/40 being protective against influenza virus infection and four-fold higher titers than baseline indicating seroconversion.

Results in FIG. 7 show that greater HAI titers were obtained with self-replicating RNA encoding HA as compared to mRNA encoding HA. HAI titers for the self-replicating RNA construct encoding HA were greater than HAI titers for the mRNA encoding HA at all time points beginning at day 28. In addition, protective HAI titers were seen for the self-replicating RNA construct encoding HA beginning at day 28 that were maintained for at least 56 days. By contrast, mRNA encoding HA showed protective HAI titers only at day 56 that were lower than HAI titers seen for the self-replicating RNA HA construct. At all other time points, HAI titers for the mRNA construct encoding HA were below the protective titer threshold, with an HAI titer that was comparable to injection with PBS control at day 28.

These results show that the self-replicating RNA construct encoding HA elicited protective HA antibody titers, with greater HAI titers as compared to the mRNA construct encoding HA.

Example 3

This example describes dsRNA production and luciferase expression for self-replicating RNA.

Several self-replicating RNA systems from different alphaviruses were tested for expression in vitro using either

186

green fluorescent protein (GFP) or firefly luciferase (Luc) as reporter genes. Initial transfection of cells with increasing amounts of self-replicating RNA resulted in expression of reporter genes at a lower dose compared to mRNA. However, as the amount of input self-replicating RNA increased, detectable expression of the reporter gene decreased.

Self-replicating RNA produces double stranded RNA (dsRNA) as an intermediate in the amplification process. Overproduction of dsRNA can suppress translation. To evaluate the effect of dsRNA production on transgene expression, dsRNA and the expression of reporter gene luciferase were measured simultaneously. HEK293 cells were transfected with 2 µg of replicon A (SEQ ID NO:115; entire RNA sequence mARM2826) or replicon B (SEQ ID NO:100, entire RNA sequence mARM2809) self-replicating RNA, or mRNA expressing Luc (SEQ ID NO:102, entire mRNA sequence mARM1782) using a commercial RNA transfection reagent. Untransfected cells (UTC) served as a control. dsRNA production (FIG. 8A) was quantified using immunohistochemical staining for dsRNA, followed by fluorescence quantification using a fluorescence scanner 24 hours after transfection. Luciferase expression (FIG. 8B) was assayed by measuring bioluminescence in parallel.

Replicon A produced a 3-fold higher level of dsRNA than replicon B 24 hrs after transfection (FIG. 8A). However, replicon B produced a 2.4-fold higher expression level of luciferase compared to replicon A. Furthermore, the level of luciferase expression from replicon A was equivalent to that observed for mRNA. Thus, even though replicon A had the ability to amplify the amount of replicon RNA and transcribed mRNA encoding luciferase, translation of the amplified mRNA was inhibited, consistent with overproduction of dsRNA inhibiting translation. Furthermore, higher levels of luciferase gene expression were seen for replicon RNA as compared to mRNA at 24, 48, and 72 hours after transfection of HEK293 cells (FIG. 9A). Self-replicating RNA with an expression cassette that included a luciferase reporter gene followed by an IRES and E3L also showed robust luciferase expression (FIGS. 9B, 9C; SEQ ID NOs: 118 and 119). Luciferase expression was also seen for a self-replicating RNA that expressed E3L from a first subgenomic promoter and a luciferase reporter gene from a second subgenomic promoter located 3' of the E3L open reading frame (not shown). Thus, not only did replicon RNA produce higher levels of luciferase gene expression compared to mRNA, but replicon RNA also showed increased duration of expression over a 72-hr period.

Example Sequences

Additional illustrative sequences are provided below, features of which are described in Table 7:

TABLE 7

SEQ ID NO	Description
SEQ ID NO: 72	nsP1-4 ORF, codon-optimized
SEQ ID NO: 73	5' UTR
SEQ ID NO: 74	5' UTR
SEQ ID NO: 75	5' UTR
SEQ ID NO: 76	3' UTR
60 SEQ ID NO: 77	Intergenic region between nsP1-4 ORF and antigenic protein ORF

187

TABLE 7-continued

SEQ ID NO	Description
SEQ ID NO: 78	Replicon sequence comprising SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 76, and SEQ ID NO: 77
SEQ ID NO: 79	nsP1-4 protein sequence
SEQ ID NO: 80	nsP1-4 protein sequence

188

TABLE 7-continued

SEQ ID NO	Description
SEQ ID NO: 81	nsP1-4 protein sequence
SEQ ID NO: 82	5' UTR (TEV)
SEQ ID NO: 83	3' UTR (Xbg)

SEQ ID NO: 72

ATGGAGAAAGTTACGTTGACATCGAGGAAGACAGCCCATTCCCTCAGAGCTTG
 CAGCGGAGCTTCCCGCAGTTGAGGTAGAACCCAAGCAGGTCACTGATAATGAC
 CATGCTAATGCCAGAGCGTTTCGCATCTGGCTCAAAACTGATCGAAACGGAGG
 TGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCGCCGAGAATGT
 ATTCTAAGCACAAGTATCATTGATCTGTCCGATGAGATGTGCCAGATCCGGA
 CAGATTGTATAAGTATGCAACTAAGCTGAAGAAAAACTGTAAGGAATAACTGA
 TAAGGAATTGGACAAGAAAATGAAGGAGCTGGCCGCCGTATGAGCAGCCCTGA
 CCTGGAAACTGAGACTATGTGCCCTCACGACGAGTCGTGCTACGAAGG
 GCAAGTCGTGTTTACCAAGGTATACGCCGTGAGGCCACCAGCTGAC
 CACCAGGCCAACAGGGCGTGAGGGTGGCTACTGGATCGCTCGACACCACA
 CCCTTCATGTTCAAGAACCTGGCCGCCCTACCCAGCTACAGCACCAACTGGG
 CCGACGAGACCGTGTGACCGCCAGGAACATCGCCTGTGCAGCAGCGACGTGA
 TGGAGAGGAGCCGAGAGGCATGAGCATCCTGAGGAAGAAATACCTGAAGGCC
 AGCAACAACGTGCTGTTCAGCGTGGCAGCACCATCTACCACGAGAAGAGGGAC
 CTGCTCAGGAGCTGGCACCTGCCAGCGTGTCCACCTGAGGGCAAGCAGAAC
 TACACCTGCAGGTGCGAGACCATCGTGAAGCTGCCACGGCTACGCCCTACAATG
 ATGCCCATCAGCCCCGGCTGTACGGCAAGCCCAGCGCTACGCCCTACAATG
 CACAGGGAGGGCTCCTGTGCTGCAAGGTGACGGACACCCCTGAACGGCGAGAGG
 GTGAGCTCCCCGTGTCACCTACGTGCCGCCACCCGTGCGACAGATGACCG
 GCATCCCTGGCCACCGACGTGAGCGCCGACGAGCCAGAAGCTGCTCGTGGCC
 TGAACACAGAGGATCGTGGTCAACGGCAGGACCCAGAGGAACACAAACACAATG
 AAGAACTACCTGCTGCCGTGGTGGCCAGGCTTCGCCAGGTGGGCAAGGAG
 TACAAGGAGGACCAGGAAGACGAGAGGCCCTGGCCTGAGGGACAGGCAGCT
 GGTGATGGGCTGCTGCTGGGCCCTCAGGCGGACAAGATCACCAGCATCTACAA
 GAGGCCCGACACCCAGACCATCATCAAGGTGAACAGCGACTTCACAGCTCGT
 GCTGCCAGGATCGGCAGCAACACCCCTGGAGATCGCCTGAGGACCCGGATCAG
 GAAGATGCTGGAGGAACACAAGGAGCCCAGCCACTGATCACCGCCAGGACGT
 GCAGGGAGGCAAGTGCCTGCCAGGAGGCCAGGGTGGAGGGAGGCCAGG
 AACCTGAGGGCCGCCCTGCCACCCCTGGCTGCCAGGTGGAGGAACCCACCC
 AAGCCGACGTGGACCTGATGCTGCAAGGAGGCCAGGGCCAGGAAGCGTGGAGA
 CCCAGGGCCTGATCAAGGTGACCAAGCTACGACGGCAGGACAAGATCGCAGC
 TAGGCCGTGCTGAGCCCACAGGCCGTGCTGAAGTCCGAGAAGCTGAGCTGCATC
 CACCCACTGGCCAGCAGGTGATCGTGTACCCACAGCGGGAGGAAGGGCAGG
 TACGCCGTGGAGCCCTACCAAGGCAAGGTGGTGGCTGCCGAGGGCCACGCCATC
 CCCGTGCAAGGACTTCCAGGCCCTGAGCGAGAGGCCACCATCGTGTACAACGAG

-continued

AGGGAGTTCGAACAGGTACCTGCACCATAATGCCACCCACGGCGGAGCCCTG
 AACACCGACGAGGAATACTACAAGACCGTGAAGCCCAGCGAGCACGACGGCGA
 GTACCTGTACGACATCGACAGGAAGCAGTGCCTGAAGAAAGAGCTGGTGACCGG
 CCTGGGACTGACCGGGCAGCTGGTGACCCACCCCTTCACGAGTTGCCTACGA
 GAGCCTGAGGACCAAGACCCGCCCTCCCTACCAAGGTGCCACCATCGGCGTGT
 CGCGTGCCTGGCAGCGAAAGAGCGGCATCATCAAGAGCGCGTACCCAAGA
 AAGACCTGGTGGTCAGCGCCAAAGAAAGAGAACTGCGCCGAGATCATCAGGGAC
 GTGAAGAAGATGAAAGGCTGGACGTGAACCGCGCACCGTGGACAGCGTGCTG
 CTGAACGGCTGCAAGCACCCCGTGGAGACCCCTGTACATCGACGGCTTCGCTT
 GCCACGCGGCCACCTGAGGGCCCTGATCGCCATCATCAGGCCAAGAAAGCCG
 TGCTGTGCGCGACCCCAAGCAGTGCCTTCTTCAACATGATGTGCCTGAAGGT
 GCACCTCAACCACGAGATCTGCACCCAGGTGTTCCACAAGAGCATCAGCGAGCG
 GTGCACCAAGAGCGTGACCGCGTGTGAGCACCCCTGTTACGACAAGAAAAT
 GAGGACCAACCAACCCCAAGGAGACCAAATCGTGTACGACACCCAGGCAGCA
 CCAAGCCCAAGCAGGACGACCTGATCCTGACCTGCTTCAGGGCTGGGTGAAGC
 AGCTGCAGATCGACTACAAGGGCAACGAGATCATGACCGCCGTGCCAGCGAG
 GCCTGACCAAGGAGGGCGTGTACGCCGTGAGGTACAAGGTGAACGAGAACCCAC
 TGTACGCTCCCCACCAGCGAGCACGTGAACCGTGTGCTGACCAAGGAGCGAG
 GGATCGTGTGAAAGACCCCTGGCCGGGACCCCTGGATCAAGACCCCTGACCGCCA
 AGTACCCCGCAACTTCACCGCCACCATCGAAGAGTGGCAGGCCAGCACGACG
 CCATCATGAGGCACATCTGGAGAGGCCGACCCACCGACGTGTTCCAGAACAA
 AGGCAACAGTGTGCTGGCCAAGGCCCTGGTGGCCGTGCTGAAGACCCGCCAG
 TCGACATGACCACAGAGCAGTGGAACACCGTGGACTACTCGAGACCGACAAGG
 CCCACAGCGCCAGATCGTGTGAACCAGCTGCGTGAGGTCTTCGGCTGG
 CCTGGACAGCGCCCTGTTCAAGGCCACCGTGCCTGACCATCAGGAACAA
 CCACTGGACACAGGCCAACATGTACGCCCTGAAACAAGGAGGTGG
 CAGGCAGCTGAGCAGGCGTACCCACAGCTGCCAGGGCGTGGCCACCGCAG
 GGTGTACGACATGAACACCGGACCCCTGAGGAACCTACGACCCAGGATCAACCT
 GGTGCCGTGAACAGGCCGCTGCCACGCCCTGGTGTGCTGCACCAACGAGCA
 CCCACAGAGCGACTTCAGCTCTTGTGAGCAAGCTGAAAGGCCAGGAGCGT
 GGTGTGGCGAGAAGCTGAGCGTGGCCCGAAGATGGTGGACTGGCTGAGCGA
 CAGGCCGAGGCCACCTCCGGCCAGGCTGGACCTCGGCATCCCCGGGACGT
 GCCCAAGTACGACATCATCTCGTGAACGTAGGACCCATACAAGTACCA
 TACCAAGCAGTGCAGGACCAACGCCATCAAGCTGAGCATGCTGACCAAGAAGGCC
 TGCCTGCACCTGAACCCGGAGGCACCTGCGTGAGCATCGCTACGGCTACGCC
 GACAGGCCAGCGAGAGCATATTGGGCCATGCCAGGCTGTTCAAGTTCA
 AGGTGTGCAAACCCAAGAGCAGCCCTGGAGGAACCGAGGTGCTGTTCGTGT
 ATCGGCTACGACCGGAAGGCCAGGACCCACAACCCCTACAAGCTGAGCAGCACC
 CTGACAAACATCACACCGGCAGCAGGCTGCACGAGGCCGGCTGCGCCCCAGC
 TACCAAGTGGTCAAGGGCGATATGCCACGCCACCGAGGGCGTGTACATCAAC

-Continued

GCTGCCAACAGCAAGGGCAGCCGGAGGCGGAGTGTGCGGCCCTGTACAAG
 AAGTCCCCGAGAGCTTCGACCTGCAGCCCCTCGAGGTGGCAAGGCAGGCTG
 GTGAAGGGCGCCGCTAAGCACATCATCCACGCCGTGGGCCCAACTTCAACAAG
 GTGAGCGAGGTGGAAGGCAGACAAGCAGCTGGCGAAGCCTACGAGAGCATCGC
 CAAGATCGTAACGACAATAACTACAAGAGCGTGCCATCCACTGCTCAGCAC
 CGGCATCTTCAGCGCAACAAGGACAGGCTGACCCAGAGCCTGAACCACCTGCT
 CACCGCCCTGGACACCACCGATGCCGACGTGGCATCTACTGCAGGGACAAGAA
 GTGGGAGATGACCCCTGAAGGGCCGTGGCCAGGCAGGGCGTGGAAAGAGA
 TCTGCATCAGCGACGACTCCAGCGTACCGAGCCGAGCCTGAGCTGGTGAGGG
 TGCAACCCAAGAGCTCCCTGGCCGGCAGGAAGGGCTACAGCACCAGCGACGGCA
 AGACCTTCAGCTACCTGGAGGGCACCAAGTTCACCCAGGCGCTAAGGACATCG
 CCGAGATCAACCGTATGTGGCCGTGGCCACCGAGGCAACGAGCAGGTGTGCA
 TGTACATCCTGGCGAGAGCATGTCCAGCATCAGGAGCAAGTGGCCGTGGAGG
 AAAGCGAGGCCAGCACACCACCCAGCACCCCTGCCCTGCCTGTGCATCCACGCTA
 TGACACCCGAGAGGGTGCAGGGCTGAAGGCCAGCAGGCCAGCAGATCACC
 GTGTGCAGCTCTTCCACTGCCCAAGTACAGGATCACGGCGTGCAGAAGATCC
 AGTGCAGCCAGCCCCTGTTCAAGGCCAAAGGTGGCCGCTACATCCACCCAG
 GAAGTACCTGGTGGAGACCCCACCCGTGGACGAGACACCCGAGCCAAGGCCGA
 GAACCAAGGCCAGGCCACCCAGAGCCCCTGATCACCGAGGAG
 AGACAAGGACCCGGACCCAGAGCCCCTGAGGAGAAGAGGAG
 AGCATCAGCCTGCTGAGCGACGGCCCCACCCAGGTGCTGCAGGTGGAGGCC
 GACATCCACGGCCCACCCAGCGTGTCCAGCTCAGCTGGAGCATCCCACACGCC
 AGCGACTTCGACGTGGACAGCCTGAGCATCTGGACACCCCTGGAGGGCGCCAGC
 GTGACCTCCGGCCCCACCGCCGAGACCAACAGCTACTTCGCAAGAGCATG
 GAGTTCCCTGGCCAGGCCGTGCGCTCCAGGACCGTGTTCAGGAACCCACCC
 ACCCAGCTCCAGGACCAGGACCCAGCCTGGCTCCAGCAGGCCGTGCAGCA
 GGACCAGCCTGGTGAGCACCCAGCAGGACCCAGCAGGTCCGTGAGCAGGAG
 AACTGGAGGCCCTGACACCCAGCAGGACCCAGCAGGTCCGTGAGCAGGAG
 GTCTGGTGTCCAACCCACCCGGCGTGAACAGGGTGTACCCAGGGAGGAATTG
 AGGCCTTCGTGGCCCAGCAACAGAGACGGTTCGACGCCGGCCTACATCTCA
 GCAGCGCACCCGGCAGGGACACCTGCAAGCAAAGAGCGTGGAGGCCAGCGT
 CTGAGCGAGGTGGTGTGGAGAGGAGCCAGGAGCTGGAAATCAGCTACGCCAGG
 CTGGACCGAGGAGAAGGGAGGAACGTGCTCAGGAAGAAACTGCAGCTGAACCC
 CCCAGCCAACAGGAGCAGGTACCAAGAGCAGGAAGGGTGGAGAACATGAAGGCC
 TCACCGCCAGGGGATCTGCAGGGCTGGGACACTACCTGAAGGCCAGGGCA
 AGGTGGAGTGCTACAGGACCCCTGCACCCCGTGCCACTGTACAGCTCCAGCGTGA
 ACAGGGCCTCTCCAGCCCCAAGGTGGCCGTGGAGGCCCTGCAACGCTATGCTGA
 AGGAGAACTTCCCCACCGTGGCCAGCTACTGCATCATCCCCAGTACGACGCC
 CCTGGACATGGTGGACGGCGCCAGCTGCTGCCCTGGACACCGCAGCTTCTGCC
 GCCAAGCTGAGGAGCTCCCCAAGAAACACAGCTACCTGGAGGCCACCATCAGG
 AGCGCCGTGCCAGCGCCATCCAGAACACCCCTGCAGAACGTGCTGGCCGTGCC

-continued

ACCAAGAGGAACGTCAACGTGACCCAGATGAGGGAGCTGCCGTGCTGGACAGC
 GCTGCCTCAACGTGGAGTGCTTCAGAAATACGCCCTGCAACAACGAGTACTGG
 GAGACCTCAAGGAGAACCCCATCAGGCTGACCGAAGAGAACGTGGTGAAC
 ATCACCAAGCTGAAGGGCCCCAAGGCCGTGCCCTGTTGCTAAGACCCACA
 CTGAACATGCTGACGGACATCCCAGGACAGCTGATGGACCTGAAGAG
 GACGTGAAGGTGACACCCGGACCAAGCACACCGAGGAGAGGCCAAGGTGCA
 GGTGATCCAGGGCGCTGACCCACTGGCACCCCTACCTGTGCGGCATCCACAG
 GGAGCTGGTGAGGCAGCTGAACGCCGTGCTGCCAACATCCACACCCGTT
 GACATGAGCGCCGAGGACTTCGACGCCATCATGCCGAGCATTCCAGGCCG
 GACTGCGTGCTGGAGACCGACATGCCAGCTCGACAAGAGCGAGGATGACGCT
 ATGGCCCTGACCCTCTGATGATCCTGGAGGACTGGCGTGGACGCCGAGCTG
 CTCACCCCTGATCGAGGTGCGCTTCGGCAGATCAGCTCCATCCACCTGCCACCA
 AGACCAAGTTCAAGTTCGGCGTATGATGAAAAGCGGAATGTTCTGACCCGTT
 CGTGAACACCGTGTCAACATTGATGCCAGCAGGTGCTGCGGAGAGGCT
 GACCGGAGCCCTGCGCTGCCATCGCGACGACAACATCGTAAGGGCGT
 GAAAAGCGACAAGCTGATGGCCGACAGGTGCGCCACCTGGCTGAACATGGAGG
 GAAGATCATCGACGCCGTGGTGGGCGAGAAGGCCCTACTCTGCCGGATT
 CATCCTGTGCGACAGCGTGACCGGACCGCCTGCAGGGTGGCGACCCCTGAA
 GAGGCTGTTCAAGCTGGCAAGCCACTGGCGCTGACGATGAGCACGAGATGA
 CAGCGGAGGGCCCTGACGAGGAAAGCACCAGGTGGAACAGGGTGGGATCC
 TGAGCGAGCTGTGCAAGGCCGTGGAGAGCAGGTACGAGACCGTGGGACCAGC
 ATCATCGTGTGGCTATGACCAACTGGCCAGCTCCGTCAAGAGCTTCTCTACC
 TGAGGGGGGCCCTATAACTCTACGGCTAA

SEQ ID NO: 73

ATGGGCGCGCATGAGAGAACGCCAGACCAATTACCTACCCAAA

SEQ ID NO: 74

GATGGGCGCGCATGAGAGAACGCCAGACCAATTACCTACCCAAA

SEQ ID NO: 75

GATAGGCGCGCATGAGAGAACGCCAGACCAATTACCTACCCAAA

SEQ ID NO: 76

ACTCGAGTATGTTACGTGCAAAGGTGATTGTCACCCCCGAAAGACCATATTGTG
 ACACACCCCTCAGTATCACGCCAACATTACGCCGGTGTCAAAACCGCG
 TGGACGTGGTTAACATCCCTGCTGGAGGATCAGCGTAATTATTATAATTGGCT
 TGGTGCCTGGCTACTATTGTGGCATGTACGTGCTGACCAACCAGAACATAATTG
 AATACAGCAGCAATTGCAAGCTGTTACATAGAACTCGCGGATTGGCATGC
 CGCCTTAAATTTTATTCTTTCTTCCGAATCGGATTTGTTTTA
 ATATTTCAAAAAAAAAAAAAAAATCTAGAAAAAAAAAAAAAA
 AAAAAAAAAAAAAAA
 AAAAAAAAAAAAAAA

195

-continued

CCTGAATGGACTACGACATAGTCTAGTCGCCAAGGCCACC

ATGGGCGCGCATGAGAGAACGCCAGACCAATTACCTACCCAAAATGGAGAAA
 GTTCACGTTGACATCGAGGAAGAACAGCCCATTCCCTAGAGCTTGAGCGGAGCT
 TCCCGCAGTTGAGGTAGAACCCAAGCAGGTCACTGATAATGACCAGCTAATG
 CCAGAGCGTTTCGCATCTGGCTCAAAACTGATCGAACCGGAGGTGGACCCATC
 CGACACGATCCTGACATTGAAAGTGCGCCGCCGCAGAATGTATTCTAACGAC
 AAGTATCATTGATCTGCCGATGAGATGTGGAAAGATCGGACAGATTGTATA
 AGTATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGG
 ACAAGAAAATGAAGGGAGCTGGCGCGTCATGAGCGACCCCTGACCTGGAAACTG
 AGACTATGCCCCCACGACGAGTCGTGCGCTACGAAGGGCAAGTCGCTG
 TTTACCAGGATGTATACTGCCGTCGACGGCCCACCAGCCTGTACCGGAGGCCAA
 CAAGGGCGTAGGGGTGGCTACTGGATCGGCTCGACACCACCCCTCATGTT
 AAGAACCTGGCGCGCCTACCCAGTACAGCACCAACTGGGCCAGGAC
 GTGCTGACGCCAGAACATGGCCTGTGCAGCGACGTGATGGAGAGGAGC
 CGGAGAGGCATGAGCATCCTGAGGAAGAAAATACCTGAAGGCCAGCAACAACGT
 GCTGTTCAGCGTGGCAGCACCATCTACCACGAGAACAGGGACCTGCTCAGGAG
 CTGGCACCTGCCAGCGTGTCCACCTGAGGGCAAGCAGAACACTACACCTGCAG
 GTGCGAGACCATCGTGAGCTGCGACGGCTACGTGGTGAAGAGGATGCCATCAG
 CCCGGCCTGTACGGCAAGCCAGCGGCTACGCCGCTACAATGCACAGGGAGGG
 CTTCCGTGCTGCAAGGTGACCGACACCCCTGAACGGCAGAGGGGTGAGCTCCC
 CGTGTGCACCTACGTGCCGCCACCTGTGCGACAGATGACGGCATCCTGCC
 ACCGACGTGAGGCCAGCACGACGCCAGAACAGCTGCTCGTGGCCTGAACCAAGG
 ATCGTGGTCAACGGCAGGACCCAGAGGAACACCAACACAATGAAGAACTACCTG
 CTGCCGTGGTGGCCAGGCTTCGCCAGGTGGGCAAGGAGTACAAGGGAGGAC
 CAGGAAGACGAGAGGCCCTGGGCTGAGGGACAGGCAGCTGGTATGGCTG
 CTGCTGGCCTTCAGGCGACAAGATCACCAGCATCTACAAGAGGCCGACAC
 CCAGACCATCATCAAGGTGAAACAGCAGCTCCACAGCTCGCTGCCAGGATC
 GGCAGCAACACCCCTGGAGATGCCCTGAGGACCCGGATCAGGAAGATGCTGGAG
 GAACACAAGGAGCCCAGCCACTGATCACCGCCAGGACGTGCAGGAGGCCAA
 GTGCGCTGCCAGCAGGGCAAGGAGGTGAGGGAGGCCAGGAACGGTGGAG
 CCTGCCACCCCTGGCTGCCAGCTGGAGGAACCCACCCCTGAGGCCAGCTGG
 ACCTGATGCTGAGGAGGCCGGCGCCAGGAAGCGTGGAGACACCCAGGGCTGA
 TCAAGGTGACCAAGCTACGACGGCAGGACAAGATCGGAGCTACGCCGTGCTGA
 GCCCACAGGCCGTGAGTCCGAGAACAGCTGGCAGGAGGGAGCTGGAG
 AGCAGGTGATCGTGTACCCACAGCGCAGGAAGGGCAGGTACGCCGTGGAGC
 CCTACCAAGGCAAGGTGGTGTGCCAGGGCACGCCATCCCGTGCAGGACT
 TCCAGGCCCTGAGCGAGAGGCCACCATCGTGTACAACGAGAGGGAGTTCGTGA
 ACAGGTACTGACCATATGCCACCCACGGCGAGGCCAGGAGCTGAACACCGAC
 AATACTACAAGACCGTGAAGGCCAGCGAGCACGCCAGTACCTGTACGACA
 TCGACAGGAAGCAGTGCCTGAAGAAAGAGCTGGTACCGGCCCTGGACTGACCG

196

SEQ ID NO: 77

SEQ ID NO: 78

-continued

GCGAGCTGGTGGACCCACCCCTCCACGAGTTGCCTACGAGAGCCTGAGGACCA
 GACCCGGCGCTCCCTACAGGTGCCACCATCGCGTGTACGGCGTGCCCCGCA
 GCGGAAAGAGCGGCATCATCAAGAGCGCCGTGACCAAGAAAGACCTGGTGGTC
 AGCGCCAAGAAGAGAACTGCGCCGAGATCATCAGGGACGTGAAGAAGATGAA
 AGGCCTGGACGTGAACGGCGCACCGTGGACACCGTGTGCTGCTGAACGGCTGCAA
 GCACCCCCGTGGAGACCCCTGTACATCGACGAGGCTTCGCTGCCACGCCGAC
 CTGAGGGCCCTGATGCCATCATCAGGCCAAGAAAGCCGTGCTGTGCGCGAC
 CCCAAGCAGTGGCGTCTTCAACATGATGTGCTGAAGGTGACTTCAACCACG
 AGATCTGCACCCAGGTGTTCCAAGAGCATCAGCAGGCGGTGACCAAGAGCG
 TGACCAGCGTGTGAGCACCCCTGTTCTACGACAAGAAAATGAGGACCACCAACC
 CCAAGGAGACCAAATCGTGATCGACACCACAGGCAGCACCAAGCCAAGCAG
 GACGACCTGATCCTGACCTGCTCAGGGCTGGTGAAGCAGCTGCGATCGAC
 TACAAGGGCAACGAGATCATGACCGCCGCTGCCAGCCAGGGCTGACCAAGGAAG
 GGCGTGTACGCCGTGAGGTACAAGGTGAACGAGAACCCACTGTACGCTCCCACC
 AGCGAGCACGTGAACGTGCTGCTGACCAGGACCGAGGACAGGATGTGTGGAAG
 ACCCTGGCCGGCGACCCCTGGATCAAGACCCCTGACGCCAAGTACCCGGCAAC
 TTCAACGCCACCATCGAAGAGTGGCAGGCCGAGCACGACGCCATCATGAGGCAC
 ATCCTGGAGAGGCCGACCCCTGAGTACTTCGAGACCGACAAGGCCAACGTGTG
 TGGGCCAAGGCCCTGGTGCCGTGCTGAAGACGCCGATCGACATGACCACA
 GAGCAGTGGAACACCGTGGACTACTTCGAGACCGACAAGGCCAACGCCGAG
 ATCGTGTGAACCAGCTGTGGTGAGGTTCTCGGCTGGACTGGACAGCGGCC
 TGTTCAAGCCCCCACCCTGCTGCACTGAGCATCAGGAACAACCAACTGGACAA
 GCCCCAGCCAAACATGTACGCCCTGAACAAGGAGGTGGTCAAGCAGCTGAGCA
 GCGGTACCCACAGCTGCCAGGGCGTGGCACCGCAGGGTGTACGACATGA
 ACACCGCACCCCTGAGGAACCTACGACCCAGGATCAACCTGGTGCCGTGAACA
 GCGGCTGCCAACGCCCTGGTGTGCAACGACACAGAGCAGACT
 TCAGCTCTTCTGTGAGCAAGCTGAAAGGCAGGACCGTGCTGGTGTGGCGAGA
 AGCTGAGCGTGGCCAGGCAAGATGGTGGACTGGCTGAGCGACAGGCCAGGCC
 CCTTCCGGCCAGGCTGGACCTCGGCATCCCCGGCAGCTGCCAAGTACGACA
 TCATCTTCTGTGAACGTCAGGACCCCATACAAGTACCACTTACGAGCTGCGA
 GGACCAAGCTGAGCATGCTGACCAAGAAGGCCCTGCCACCTGAA
 CCCGGAGGCACCTCGTGAGCATCGCTACGGCTACGCCAGGGCCAGCGA
 GAGCATATTGGCGCATGCCAGGCTGTTCAAGTTCAAGCAGGGTGTGCAA
 CAAGAGCAGCTGGAGGAACCGAGGTGCTGTTCAAGCTACGGCTACGACCG
 GAAGGCCAGGACCCACAACCCCTACAAGCTGAGCAGCACCCCTGACAAACATCTA
 CACCGCAGCAGGCTGCACGAGGCCGCTGCCACCCAGCTACCACTGGTCA
 GGGCGATATGCCACGCCACCGGAGGGCGTGTACCATCAACGCTGCCAACAGCAA
 GGGCCAGGCCGGAGGGCGAGTGTGCGGCCCTGTACAAGAAGTCCCCGAGAG
 CTTCGACCTGCAAGCCCATCGAGGTGGCAAGGCCAGGCTGGTGAAGGGCGCC
 TAAGCACATCATCCACGCCGTGGGCCCAACTCAACAAGGTGAGCGAGGTGGA

- continued

AGGCAGACAAGCAGCTGGCCGAAGCCTACGAGAGCATGCCAAGATCGTGAACG
 ACAATAACTACAAGAGCGTGGCATCCACTGCTCAGCACCGCCTTCAGCG
 GCAACAAGGACAGGCTGACCCAGAGCCTGAACCACCTGCTCACCGCCCTGGACA
 CCACCGATGCCACGTGGCATCTACTGCAGGGACAAGAAGTGGGAGATGACCC
 TGAAGGAGGCCGTGGCCAGGGGGAGGCCGTGGAAGAGATCTGCATCAGCGAC
 GACTCCAGCGTGAACCGAGCCCCAGCAGCTGGTGAGGGTGCACCCAAAGAGC
 TCCCTGGCGGAGGAAGGGCTACAGCACCGCAGCGACGGCAAGACCTTCAGCTAC
 CTGGAGGGCACCAGATTCCACCAAGGCCGCTAAGGACATCGCCAGATCAACGCT
 ATGTGGCCCGTGGCCACCGAGGCCAACGAGCAGGTGTGCATGTACATCCTGGC
 GAGAGCATGTCCAGCATCAGGAGCAAGTGCCCCGTGGAGGAAGCGAGGCCAG
 CAACACCCAGCACCCCTGCCCTGCTGTGCATCCACGCTATGACACCCGAGAGG
 GTGCAGCGGCTGAAGGCCAGCAGGCCAGCAGCATACCGTGTGCAGCTCCTTC
 CCACTGCCAAGTACAGGATCACCGCGTGCAGAAGATCCAGTGCAGCCAGCCC
 ATCCTGTTCAGCCAAAGGTGCCGCCTACATCCACCCAGGAAGTACCTGGTGG
 AGACCCACCCGTGGACGAGACACCCAGCCAAGGCCAGAGAACCCAGAGCACC
 GAGGGCACACCGAGCAGCCACCCCTGATCACCGAGGACGAGACAAGGACCCG
 GACCCAGAGCCATCATTATCGAGGAAGAGGAAGAGGACAGCAGCATCAGCCTGCT
 GAGCGACGGCCCCACCCACCAGGTGCTGCAGGTGGAGGCCAGATCCACGCC
 ACCCAGCGTGTCCAGCTCCAGCTGGAGCATCCACACGCCAGCGACTTCGACGT
 GGACAGCCTGAGCATCCTGGACACCCCTGGAGGGGCCAGCGTGCACCTCCGGCG
 CACCAGGCCGAGACCAACAGCTACTTCGCAAGAGCATGGAGTTCTGGCAG
 GCCCAGCTCCAGGACCGTGTTCAGGAACCCACCCACCCAGCTCCAG
 GACCAGGACCCAAAGCCTGGCTCCAGCAGGGCCTGCAGCAGGACCGCCTGGT
 GACCACCCACCCGGCGTGAACAGGGTGTACACCAGGGAGGAACCTGGAGGCC
 GACACCCAGCAGGACCCCCAGCAGGTGTGAGCAGGACTAGTCTGGTGTCAA
 CCCACCCGGCGTGAACAGGGTGTACACCAGGGAGGAATTGAGGCCCTCGG
 CCAGCAACAGAGACGGTTGACGCCGGCGCTACATCTTCAGCAGCGACACCG
 CCAGGGACACCTGCAGCAAAAGAGCGTGAGGCAGACCGTGTGAGCGAGGTGG
 TGCTGGAGAGGACCGAGCTGGAAATCAGCTACGCCACCCAGGTGGACCGAGGAG
 AGGAGGAACCTGCTCAGGAAGAAACTGCAGCTGAACCCACCCAGCCAACAGG
 AGCAGGTACCAAGAGCAGGAAGGTGGAGAACATGAAGGCCATCACCGCCAGGCG
 GATCCTGCAGGGCTGGACACTACCTGAAGGCCAGGGCAAGGTGGAGTGCTA
 CAGGACCCCTGCACCCCGTGCCTACTGAGCAGCTCCAGCGTGAACAGGGCCTCTCC
 AGCCCCAAGGTGGCGTGGAGGCCCTGCAACGCTATGCTGAAGGAGAACTTCCCC
 ACCGTGGCCAGCTACTGCATCATCCCCGAGTACGACGCCACCTGGACATGGTGG
 ACGGCGCCAGCTGCTGCCCTGGACACCGCCAGCTTCTGCCCGCAAGCTGAGGA
 GCTTCCCCAAGAACACAGCTACCTGGAGGCCACCATCAGGAGCGCCGTGCC
 GCGCCATCCAGAACACCCCTGCAGAACAGTGCTGCCAGCTGCCACCAAGAGGA
 GCAACAGTGAACCGAGATGAGGGAGCTGCCCGTGTGGACAGCGCTGCCCTCAACG
 TGGAGTGTCTCAAGAAATACGCCCTGCAACAACGAGTACTGGGAGACCTTCAGG
 AGAACCCCATCAGGCTGACCGAAGAGAACGTGGTGAACATCACCAAGCTGA

-continued

AGGGCCCCAAGGCCGCTGCCCTGTCGCTAACGACCCACAACCTGAACATGCTGC
 AGGACATCCCAATGGACAGGTTCGTGATGGACCTGAAGAGGGACGTGAAGGTGA
 CACCCGGCACCAAGCACACCGAGGGAGAGGCCAAGGTGCAGGTGATCCAGGCC
 GCTGACCCACTGGCCACCGCCTACCTGTGCGGCATCCACAGGGAGCTGGTGA
 CGGCTGAACGCCGTGCTGCCAACATCCACACCCGTGTCGACATGAGGCC
 AGGACTTCGACGCCATCATCGCGAGCACTTCAGCCGGCAGCTGCGTGC
 AGACCGACATGCCAGCTCGACAAGAGCGAGGATGACGCTATGCCCTGACCG
 CTCTGATGATCTGGAGGACCTGGCGTGACGCCAGCTGCTCACCGTAC
 GGCTGCCCTCGCGAGATCAGCTCCATCCACCTGCCACCAAGACCAAGTCAAG
 TCGCGCTATGATGAAAGCGGAATGTTCTGACCCGTGTCGACACC
 TCAACATTGTGATGCCAGCAGGGTGCTGCCAGGGCTGACCGCAGGCC
 GCGCTGCCCTCATCGCGACGACAACATCGTAAGGGCGTAAAAGCGACAAGC
 TGATGGCCGACAGGTGCCACCTGGCTGAACATGGAGGTGAAGAGATCATCGAC
 CGCGTGGGGCGAGAAGGCCCTACTTCTGCCGGATTCCATCGACAG
 CGTGACCGGCACCGCCTGCAGGGTGGCGACCCCTGAAGAGGCTGTTCAAGCT
 GCACGAGGAAAGCACCAGGTGGAACAGGGTGGCATCCTGAGCGAGCTGTGCA
 AGCCGTTGGAGGACAGGTACGAGACCGTGGGACCCAGCATCGTGTGCA
 TGACCACACTGCCAGCTCGTCAAGAGCTCTCCTACCTGAGGGGGCCCTAT
 AACTCTACGGCTAACCTGAATGGACTACGACATAGTCTAGTCCGCAAGGCC
 CCACCACTCGAGTATTTACCTGCAAAGGTGATTGTCACCCCGAAAGACCATA
 TTGTGACACACCCCTCAGTATCACGCCAACATTACAGCCGCGGTGTC
 CGCGTGGACGTGGTTAACATCCCTGCTGGGAGGATCAGCGTAATTATTATAATT
 GGCTTGGTGCTGGCTACTATTGTGGCATGTCGCTGACCAACCAGAAACATA
 ATTGAATACAGCAGCAATTGCAAGCTGCTTACATAGAACTCGCGCGATTGGC
 ATGCCGCTTAAAAAATTTTATTTTATTTCTTTCTTTCCGAATGGATTTGT
 TTTTAATTTCAAAAAAAAAAAAAAAAAAAAAAATCTAGAAAAAAAAAAA
 AAAAAAAAAAAAAAAAAAAAAA
 AAAAAAAAAAAAAA

SEQ ID NO: 79

MEKVHVDIEEDSPFLRALQRSFPQEVEAKQVTNDHANARAFSHLASKLIETEVDP
 SDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRILYKYATKLKKNCKEITDKELDK
 KMKELAAMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPTSLYHQANK
 GVRVAYWIGFDTPFMFKNLAGAYPSYSTNWAETVLTARNIGLCSSDVMERSRRG
 MSILRKKYLKPNSNNVLPSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRCETIVSC
 DGYVVKRIAISPGLYGKPGSYATMHREGFLCKVTDTLNGERVSFPVCTYVPATLC
 DQMTGILATDVSADDAQKLLVLGNQRIVVNNGRTQRNTNTMKNYLLPVVAQAFARW
 AKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPTQTIIKVNSDFHSFV
 LPRIGSNTLEIGLRTIRKMLEEHKEPSPLITAEDVQEAKCAADEAKEVREAEELRAA
 LPPLAADVEEPTLEADVDMLQEAAGAGSVETPRGLIKVTSYDGEDKIGSYAVLSPQA
 VLKSEKLSCHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAI
 PVQDFQALSESA

-continued

TIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIRKQCVKKEL
 VTGLGLTGEVLDPFFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKGSIIKSAVTKKD
 LVVSAKKENCAEIIRDVKKMGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAG
 TLRALIAIIRPKAVLCGDPKQCGFFNMCLKVHFNHEICTQVFHKSISRRTKSVTS
 VVSTLFYDKKMRTTNPKETKIVIDTTGSTPKQDDILITCFRGWVKQLQIDYKGNEI
 MTAASQGLTRKGVYAVRYKVNNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPW
 IKTLTAKYPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNANKANVCWAKALPVUL
 KTAGIDMTTEQWNTVDYFETDKAHSAEIVLNQLCVRFFGLDLSGLFSAPTVPLSIR
 NNHWNDNSPSPNMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRIN
 LVPVNRLRPHALVLHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKMDWLSDR
 PEATFRARLDLGIPGDVPKYDIIFVNVRTPYKYHHYQQCEDHAIKLSMLTKACLHL
 NPGGTCVISGYGADRASESIIGAIARLFKFSRVCKPKSSLTEETEVLFVFIGYDRKART
 HNPYKLSSTLTNIYTGSRLHEAGCAPSYHVVRCGIATATEGVIIINAANSKGQPGGGV
 CGALYKKFPESFDLQPIEVGKARLVKGAAKHIHAVGPNFNKVSEVEGDKQLAEAYE
 SIAKIVNDNNYKSVAPIPLLSTGIFSGNKDRLTQSLNHLLTADTTADVAIYCRDKKW
 EMTLKEAVARREAVEEICISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYL
 EGTKFHQAAKDIAEINAMWPVATEANEQVCMYILGESMSSIRSCKPVEESEASTPPST
 LPCLCIHAMTPERVQRLKASRPEQITVCSSFPLPKYRITGVQKIQCSPILFSPKVPAYI
 HPRKYLVETPPVDETPEPSAENQSTEQTPEQPPLITEDETRTRTPEPIIIEEEEDSISLLS
 DGPTHQVLQEADIHGPPSVSSSSWSPHASDFDVDSLISILDLEGASVTSGATSAETN
 SYFAKSMEFLARPVPAPRTVFNPVPHAPRTRTPSLAPSACSRSLVSTPPGVNRVIT
 REELEALTPSRTPSRSVSRTSLVSNPVGVRVITREEFEAFVAQQQRFDAGAYIFSSD
 TGQGHLQQKSVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKQLQNLPTPANRSR
 YQSRKVENMKAITARRILQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVA
 VEACNAMLKENPPTVASYCIIPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSY
 LEPTIRSAVPSAIQNTLQNVLAATKRNCNVTQMRELPUVLDAAFNVECPKKYACNN
 EYWETFKENPIRLTEENVVNYITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDLK
 RDVKVTPGTKHTEERPKVQVIQAADPLATAYLCGIHRELVRRLNAVLLPNIHTLFDM
 SAEDFDIAAEHFQPGDCVLETDIASFDKSEDDAMALTALMILEDLGVDAAELLTLEIA
 AFGEISSIONLPTKFKFGAMMKSGMPLTLPVNTVINIVIASRVLRERLTGSPCAAFIG
 DDNIVKGVKSDKLMADRCATWLNMEVKIIDAVVGEKAPYFCGGFILCDSVTGTACR
 VADPLKRLFKLGKPLAADDHEHDDRRRALHEESTRWNVRVGILSELCKAVESRYETV
 GTSIIVMAMTTLASSVKSFSYLRGAPITLYG

SEQ ID NO: 80

MPEKVHVDIEEDSPFLRALQRSPFQFVEAKQVTDNDHANARAFSHLASKLIENTEV
 PSDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKNCKEITDKELD
 KKMKEAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPTSLYHQAN
 KGVRVAYWIGFDTPFMFKNLAGAYPSYSTNWADETVLTARNIGLSSDVMERSRR
 GMSILRKKYLKPSNNVLFVGSTIYHEKRDLLRSWHLPSVFHRLGKQNYTCRCETIVS
 CDGYVVKRIAISPGLYGKPSGYATMHREGFLCCKVTDTLNGERVSFPVCTYVPATL
 CDQMTGILATDVSADDAQKLLVGLNQRIVVNGRTQRNTNTMKNYLLPVVAQAFAR

-continued

WAKEYKEDQEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPTQTIIKVNSDFHSF
 VLPRIGSNTELEIGLRTIRKMLEEHKEPSPLITAEDVQEAKCAADEAKEVREAELRA
 ALPPLAADVEEPTLEADVDLMLQEAGAGSVTPRGLIKVTSDGEDKIGSYAVLSPQ
 AVLKSEKLSICHIPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAI PVQDFQALSES
 ATIVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIRKQCVKK
 ELVTGLGLTGEVDPPHFESLRPAAPEYQVPTIGVYGPMSGKSGIIKSATVK
 KDLVVSACKENCAEIIRDVKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACH
 AGTLRALIAIIRPKKAVLCGDPKQCGFFNMCMCLVKVHFNHEICTQVFHKSISRRCTKSV
 TSVVSTLFYDKKMRTTNPKETKIVIDTTGSTPKQDDLILTCFRGWVKQLQIDYKGN
 EIMTAASQGLTRKGVYAVRYKVNEPNPLYAPTSEHVNVLLTRTEDRIWWTLAGDP
 WIKTLTAKYPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKA NCWAKALVPV
 LKTAGIDMTTEQWNNTVDYFETDKAHSAEIVLNQLCVRFFGLDDSGLFSAPTVPLSIR
 NHWDNSPSPNMYGLNKEVVRLSRRYQQLPRAVATGRVYDMNTGTLRNYDPRIN
 LPVPNRRLPHALVLHNEHPQSDSSVSKLKGRTVLUVGEKLSVPGKMDWLSDR
 PEATFRARLDLGIPGDVPKYDIIFVNVRTPYKYHYQQCEDHAIKLSMLTKACLHL
 NPGGTCVSIGYGYADRASESIIIGAIARLFKFSRVCKPKSSL EETEVLFVFIGYDRKART
 HNPYKLSSTLTNIYTGSRLHEAGCAPSYHVRGDIATATEGVIIINAANSKGQPGGGV
 CGALYKKFPESFDLQPIEVGKARLVKGAAKHIHAVGPNFNKVSEVEGDKQLAEAYE
 SIAKIVNDNNYKSVAPIPLLSTGIFSGNKDRLTQSLNHLLTADTTADVAIYCRDKKW
 EMTLKEAVARREAVEEICISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSYL
 EGTKFHQAQKIAEINAMWPVATEANEQVCMYILGESMSSIRSCKPVEESEASTPPST
 LPCLCIHAMTPERVQRKLKASRPEQITV CSSFPLPKYRITGVQKIQCSPILFSPKVPAYI
 HPRKYLVETPPVDETPEPSAENQSTE GTPEQPLITEDETRTRTPEPIIIEEEEEDSISLLS
 DGPTHQVLQVEADIHGPPSVSSSSWSIPHASDFDVDSL SILDLEGASVTSGATSAETN
 SYFAKSMEFLARPVPAPRTVFRNPPHPAPRTRTPSLAPSACSRSLVSTPPGVNRVIT
 REEELALTPSRTPSRSVSRSLVSNPPGVNRVITREEFEAFVAQQQRRFDAGAYIFSSD
 TGQGHLQQKSVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKQLQNLPTPANRSR
 YQSRKVENMKAITARRILQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKVA
 VEACNAMLKENPPTVASYCIIPPEYDAYLDMVDGASCCLDTASFCPAKLRSFPKKHSY
 LEPTIRSAVPSAIQNTLQNVLAATKRNCNVTQMREL PVLDSA AFNVECFKKYACNN
 EYWETFKENPIRLTEENVVNYITKLKGPKAAALFAKTHNLNMLQDIPMDRFVMDL
 RDVKVTPGTKHTEERPKVQVIQAADPLATAYLCGIHREL VRRLN A VLPN IHTLPDM
 SAEDFDAIIAEHFQPGDCVLET DIASFDKSEDDAMALTALMILED LGVDAELLT LIEA
 AFGEISSIHLPTKTKFGAMMKGMLTLPVNTVINIVIASRVLRERLTGSPCAAFIG
 DDNIVKGVKSDKLMADRCATWLNEVKIIDAVVGEKAPYFCGGFILCDSVTGTACR
 VADPLKRLFKLGKPLAADDEHDDRRRALHEESTRWNVRVGILSELCKAVESRYETV
 GTSIIIVMAMTTLASSVKSFSYLRGAPITLYG

SEQ ID NO: 81

MEKVHVDIEEDSPFLRALQRSFPQFEVAKQVTDNDHANARAFSHLASKLIETEVDP
 SDTILDIGSAPARRMYSKHKYHCICPMRCAEDPDRLYKYATKLKKNKEITDKELDK
 KMKELAAVMSDPDLETETMCLHDDESCRYEGQVAVYQDVYAVDGPTSLYHQANK

-continued

GVRVAYWIGFDTTPFMFKNLAGAYPSYSTNWADETVLTARNIGLCSSDVMERSRRG
 MSILRKKYLKPSNNVLPSVGSTIYHEKRDLLRSWHLPSVFHLRGKQNYTCRCETIVSC
 DGYVVKRIAISPGLYGKPSGYATMHREGFLCCKVTDTLNGERVSFPVCTYVPATLC
 DQMTGILATDVSAADAKLLVGLNQRIVVNGRQRNTNTMKNYLLPVVAQAFARW
 AKEYKEDEDERPLGLRDRQLVMGCCWAFRRHKITSIYKRPTQTIIKVNSDFHSFV
 LPRIGSNTLEIGLRTRIKMLEEHKEPSPLITAEDIQEAKCAADEAKEVREAEELRAL
 PPLAADFEEPTLEADVDLMLQEAGAGSVETPRGLIKVTSYAGEDKIGSYAVLSPQAV
 LKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAI PVQDFQALSESAT
 IVYNEREFVNRYLHHIATHGGALNTDEEYYKTVKPSEHDGEYLYDIRKQCVKKEL
 VTGLGLTGELEVDPFFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKD
 LVVSACKENCAEIIRDVKKMKGLDVNARTVDSVLLNGCKHPVETLYIDEAFACHAG
 TLRALIAIIRPKKAVLCGDPKQCGFFNMCLKVHFNHEICTQVFHKSISRRCTKSVTS
 VVSTLFYDKRMRTTNPKETKIVIDTTGSTKPKQDDLILTFCPRGWVKQLQIDYKGNEI
 MTAASQGLTRKGVYAVRYKVNEPNPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPW
 IKILTAKYPGNPTATIEEWQAEHDAIMRHILERPDPTDVFQNKAJVWAKALVPVLK
 TAGIDMTTEQWNNTVDYFETDKAHSAEIVLNQLCVRFFGLDSDGLFSAPTVPLSIRN
 NHWDNSPSPNMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTGTLRNYDPRINL
 VPVNRRRLPHALVLHHNEHPQSDFSSFVSKLKGRTVLVVGEKLSVPGKKVDWLSDQP
 EATFRARLDLGIPGDVPKYDIVFINVRTPYKYHYQOCEDHAIKLMSMLTKACLHLN
 PGGTCVSIGGYADRASESIIGAIARQFKFSRVCKPKSSHEETEVLFVFIGYDRKARTH
 NPYKLSSTLTNIYTGSRLHEAGCAPSYHVRGDIATATEGVIINAANSKGQPGGGVC
 GALYKKFPESFDLQPIEVGKARLVKGAAKHIHAVGPNFNKVSEVEGDQLAEAYES
 IAIKVNNDNNYKSVAIPLLSTGIFSGNKDRLTQSLNHLLTALDTTDADVAIYCRDKKWE
 MTLKEAVARREAVEEICISDDSSVTEPDAELVRVHPKSSLAGRKGYSTSDGKTFSTYLE
 GTKFHQAQKDIAEINAMWPVATEANEQVCMYILGESMSSIRSCKPVEESEASTPPSTL
 PCLCIHAMTPERVQLKASRPBQITVCSSFPLPKYRITGVQKIQCSQPILESPKPAYIH
 PRKYLVETPPVEETPESPAENQSTEGTPEQPALVNDACTRTRMPEPIIEEEEEDSISLL
 SDGPTHQVLQVEADIHGSPSVSSSSWISPHASDFDVDSLISLDTLGASVTGAVSAET
 NSYFARSMEFRARPVPAPRTVFRNPPHPAPRTTPPLAHSRASSRTSLVSTPPGVNRVI
 TREELEALTSPRAPSRSASRTSLVSNPPGVNRVITREEFEAFVAQQQ*RFDAGAYIFSS
 DTGQGHQQKSVRQTVLSEVVLERTELEISYAPRLDQEKEELLRKKLQLNPPTPANRS
 RYQSRRVENMKAITARRILQGLGHYLKAEGKVECYRTLHPVPLYSSSVNRAFSSPKV
 AVEACNAMLKENFPPTVASYCIIPEYDAYLDMVDGASCLDTASFCPAKLRSFPKKHS
 YLEPTIRSAVPSAIQNTLQNVLAATKRNCNTQMRELPLVLDAAFNCFCFKKYACN
 NEYWETFKENPIRLTEENVNVYITKLKGPKAALFAKTHNLNMLQDIPMDRFVMDL
 KRDVKVTPGKHTTEERPKVQVIQAADPLATADLCGIHRELVRLNAVLLPNIHTLFD
 MSAEDFDAIIIAEHFQPGDCVLETDIASFDKSEDDAMALTALMILEDLGVDAAELTLIE
 AAFGEISSIHLPTKTKFGAMMKSGMFLTFVNTRVINIVIASRVLRLTGSPCAAFI

-Continued

ACCACACCCCTCATGTTCAAGAACCTGGCCGGCGCTACCCCAGTACAGCACCAACTGGGCCACG
 AGACCGTGTGACCGCCAGGAACATCGGCCTGTGCAGCAGCAGTGATGGAGAGGAGCCGGAGA
 GGCATGAGCATCTGAGGAAGAAATACCTGAAGGCCAGCAACAACTGCTGTTCAAGCGTGGCAGC
 ACCATCTACCACGAGAAGAGGGACCTGCTCAGGAGCTGGCACCTGCCAGCGTGTCCACCTGAGG
 GGCAAGCAGAACTACACCTGCAGGTGCGAGACCATCGTGAGCTGCGACGGTACGTGGTAAAGAG
 GATGCCATCAGCCCCGGCTGTACGGCAAGGCCAGCGGCTACGCCGCTAACATGCACAGGGAGGG
 CTTCTGTGCTGCAAGGTGACCGACACCCCTGAACGGCGAGAGGGTGAGCTTCCCCTGTGACCTA
 CGTGGCCGCCACCCCTGTGCGACAGATGACCGCATTGGCACCGACGTGAGCGCCGACGACGC
 CCAGAAGCTGCTCGTGGCCTGAACCAGAGGATCGTGGTCAACGGCAGGACCCAGAGGAACACCA
 ACACAATGAAGAACTACCTGCTGCCGTGGTGGCCAGGCTTCGCCAGGTGGCAAGGAGTACA
 AGGAGGACCAGGAAGACGAGAGGCCCTGGCCTGAGGGACAGGCAGCTGGTATGGCTGCTG
 CTGGCCTTCAGGGCGACAAGATCACCAGCATCTACAAGAGGCCGACACCCAGACCATCATCAA
 GGTGAACAGCGACTTCCACAGCTTCGTGCTGCCAGGATCGCAGCAACACCCGGAGATCGCCT
 GAGGACCCGGATCAGGAAGATGCTGGAGGAACACAAGGAGGCCAGGCCACTGATCACGCCGAGG
 ACGTGCAGGAGGCCAAGTGCCTGCCAGGCCAGGGAGGTGAGGGAGGCCAGGAACCTGAG
 GGCGCCCTGCCACCCCTGGCTGCCAGTGAGGAACCCACCCCTGGAGGCCAGTGACCTGAT
 GCTGCAGGAGGCCGGCGCCGGAAGCGTGGAGACACCCAGGGCCTGATCAAGGTGACCAGCTACG
 ACGCGAGGACAAGATCGCAGCTACGCCGTGCTGAGCCCACAGGCCGTGCTGAAGTCCGAGAAG
 CTGAGCTGCATCCACCCACTGGCGAGCAGGTGATCGTGTACCCACAGCGCCAGGAAGGCAG
 GTACGCCGTGGAGCCCTACCACGGCAAGGTGGTGTGCCCCAGGGCACGCCATCCCGTGCAGGA
 CTTCCAGGCCCTGAGCGAGGCCACCATCGTGTACAACGAGAGGGAGTCTGTGAACAGGTACCT
 GCACCATATGCCACCCACGGCGAGCCCTGAACACCGACGAGGAATACTACAAGACCGTGAAGCC
 CAGCGAGCACGCGAGTACCTGTACGACAGGAAGCAGTGCAGTGAAGAAAGAGCTGG
 TGACCGGCTGGACTGACCGGGAGCTGGTGACCCACCCCTCCACGAGTTGCCCTACGAGAGCC
 TGAGGACAGACCCGCCCTCCACAGGTGCCACCATCGCGTGTACGGCGTGCCCGCAGCG
 GAAAGAGCGGCATCATCAAGAGCGCGTGACCAAGAAAGACCTGGTGGTACGCCAGAAAGAG
 AACTGCGCCGAGATCATCAGGGACGTGAAGAAGATGAAAGGCCTGGACGTGAACGCGCGACCGT
 GGACAGCGTGTGCTGAACGGCTGCAAGCACCCCGTGGAGACCCCTGTACATCGACAGGCCCTCG
 TTGCCACGCCGGACCCCTGAGGCCCTGATGCCCATCATCGGCCAGAAAGCCGTGCTGTGCG
 CGACCCCAAGCAGTGGCTTCTCAACATGATGTGCTGAAGGTGCACTTCAACCACGAGATCTGC
 ACCCAGGTGTTCCAAGAGCATCAGCAGGCCGTGACCAAGAGCGTACCGCGTGTGAGCACC
 CTGTTCTACGACAAGAAAATGAGGACCACCAACCCCAAGGAGACCAAAATCGTGTACGACACCA
 GGCAGCACCAAGGCCAAGCAGGACGACCTGATCCTGACCTGCTTCAGGGCTGGGTGAAGCAGCTG
 CAGATCGACTACAAGGGCAACGAGATCATGACCGCCGTGCCAGGCCAGGGCTGACCGAGGAAGGG
 CGTGTACGCCGTGAGGTACAAGGTGAACGAGAACCCACTGTACGCTCCACCGAGCACGTGAA
 CGTGTGCTGACCGAGGACGGAGGACAGGATCGTGTGAAAGACCCCTGGCGGCCAGCCCTGGATCA
 AGACCCCTGACGCCAAGTACCCCGCAACTTACCCGCCACCATCGAAGAGTGGCAGGCCAGCAG
 ACGCCATCATGAGGCACATCCTGGAGAGGCCAGCCACCGACGTGTTCCAGAACAGGCCAAC
 TGTGCTGGCCAAGGCCCTGGTGCCTGCTGAAGACCGCCGCATCGACATGACCAAGAGCAGT
 GGAACACCGTGGACTACTTCGAGACCGACAAGGCCACAGGCCAGGATCGTGTGAACCAAGCTGT
 GCGTGGAGTTCTCGGCCCTGGACCTGGACAGCGGCCCTGTTCAAGGCCAGGCCACTGAGCAT

-continued

CAGGAACAACCCTGGGACAACAGCCCCAGCCAAACATGTACGGCCTGAACAAGGAGGTGGTCA
 GGCAGCTGAGCAGGCCGTACCCACAGCTGCCAGGGCGTGCCCACCGCAGGGTGTACGACATG
 AACACCGGCACCCCTGAGGAACTACGACCCCAGGATCAACCTGGTGCCCCTGAACAGGCCGTGCC
 CACGCCCTGGTGTGACCACACAGAGCACCCACAGAGCGACTTCAGCTCCTCTGTGAGCAAGCTGA
 AAGGCAGGACCGTGCTGGTGTGGCGAGAACGCTGAGCGTGCCCGCAAGATGGTGGACTGGCTG
 AGCGACAGGCCGAGGCCACCTCCGGGCCAGGCTGGACCTGGCATCCCGGCACGTGCCAAG
 TAGGACATCATCTCGTGAACGTCAGGACCCATAAAGTACCAACATTACCAAGCAGTGCAGGACC
 ACGCCATCAAGCTGAGCATGCTGACCAAGAAGGCCCTGCCCTGACCTGAAACCCGGAGGCACCTGCG
 TGAGCATCGGCTACGGCTACGCCAGGGCCAGCGAGAGCATCATTGGGCCATGCCAGGCTGT
 TCAAGTTCAAGCAGGGTGTGCAAACCCAAGAGCAGCCTGGAGGAAACCGAGGTGCTTGTGTTCA
 TCGGCTACGACCGGAAGGCCAGGACCCACAACCCCTACAAGCTGAGCAGCACCTGACAAACATCT
 ACACCGGCAGCAGGCTGACGAGGCCGGCTGCCCTGGCCAGCTACACGTGGTCAGGGCGATATC
 GCCACCGCCACCGAGGGCGTGATCATCAACGCTGCCAACAGCAAGGGCCAGCCCGAGGCCAGT
 GTGCGGCCGCCCCGTACAAGAAGTTCCCGAGAGCTCGACCTGCAAGCCATCGAGGTGGCAAGGC
 CAGGCTGGTGAAGGGCGCCGCTAACGACATCATCCACGCCGTGGGCCAACCTCAACAAGGTGAG
 CGAGGTGGAAGGCAGACAAGCAGCTGGCGAACGCTACGAGAGCATGCCAACAGATCGTAACGACA
 ATAACATACAAGAGCGTGGCCATCCACTGCTCAGCACCGGATCTCAGCGGAAACAAGGACAGGC
 TGACCCAGAGCCTGAACCACCTGTCACCGCCCTGGACACCACCGATGCCACGTGGCCATCTACTG
 CAGGGACAAGAAGTGGAGATGACCTGAAGGAGGCCGTGGCCAGGCAGGCCGTGGAAGAG
 ATCTGCATCAGCGACACTCCAGCGTGACCGAGGCCGACGCCGAGCTGGTGGGGTGCACCCCAAG
 AGCTCCCTGGCCGGCAGGAAGGGCTACAGCACAGCGACGCCAAGACCTTCAGCTACCTGGAGGG
 CACCAAGTTCCACCAGGCCGCTAACGACATGCCGAGATCAACGCTATGTGGCCGTGGCCACCGA
 GGCCAAACGAGCAGGTGTGCATGTACATCCTGGCGAGAGCATGTCCAGCATCAGGAGCAAGTGCC
 CGTGGAGGAAAGCGAGGCCAGCACACCACCCAGCACCTGCCCTGTGCATCCACGCTATGAC
 ACCCGAGAGGGTGCAGCGCTGAAGGCCAGCAGGCCGAGCAGATCACCGTGTGCAGCTCTCCC
 ACTGCCAAGTAGCAGGATCACCGCGTGCAAGAAGATCCAGTGCAGGCCAGCCATCTGTTCACGCCA
 AAGGTGCCGCTTACATCCACCCAGGAAGTACCTGGTGGAGACCCACCCGTGGACGAGACACCC
 GAGCCAAGCGCCGAGAACCGAGCACCAGGCCAGGGCACACCCAGCAGGCCACCCCTGATCACCGAGGA
 CGAGACAAGGACCGGGACCCAGAGCCATCATTATCGAGGAAGAGGAAGAGGACAGCATCAGCC
 TGCTGAGCGACGCCACCCAGGTGCTGCCAGGTGGAGGCCGACATCCACGCCACCCAGCG
 TGTCCAGCTCCAGCTGGAGCATCCACGCCAGCGACTTCAGCTGGACGCCCTGAGCATCTGG
 ACACCCCTGGAGGGCGCCAGCGTGACCTCCGGGCCACAGCGCCGAGACCAACAGCTACTCGCCA
 AGAGCATGGAGTTCTGGCCAGGCCGTGCCAGCTCCAGGACCGTGTTCAGGAACCCACCCACC
 CAGCTCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCC
 GCACCCCAACCCGGCGTGAACAGGGTGTACCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCC
 ACCCCCAGCAGGCCGTGAGCAGGACTAGTCTGGTGTCCAACCCACCCGGCGTGAACAGGGTGTAC
 ACCAGGGAGGAATTGAGGCCCTCGTGAGGCCACCTGCAGCAAAAGAGCGTGGAGGCCAGGCCAGGCC
 TTCAAGCAGCGACACCGGCCAGGGACACCTGCAGCAAAAGAGCGTGGAGGCCAGGCCAGGCC
 GGTGGTGTGGAGAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCCAGGCC
 GAACTGCTCAGGAAGAAACTGCAGCTGAACCCACCCAGCCAACAGGAGCAGGTACCAAGCAG

-continued

GAAGGTGGAGAACATGAAGGCCATCACGCCAGGCGGATCTGCAGGGCTGGGACACTACCTGA
AGGCCGAGGGCAAGGTGGAGTGCTACAGGACCCCTGCACCCCGTGCACACTGTACAGCTCCAGCGTGA
ACAGGGCCTCTCCAGCCCCAAGGTGGCCCTGCAACGCTATGCTGAAGGAGAACTTCC
CCACCGTGGCCAGCTACTGCATCATCCCCGAGTACGACGCCACCTGGACATGGGACAGGGGCCA
GCTGCTGCCCTGGACACCAGCTGACCTCTGCCCGCCAAGCTGAGGAGCTTCCCAAGAACACAGCTA
CCTGGAGGCCACCATCAGGAGGCCGTGCCAGCGCATTCCAGAAACCCCTGCAGAACGTGCTGGC
CGCTGCCACCAAGAGGAACGTGACACGTGACCCAGATGAGGGAGCTGCCCTGCTGGACACCGCTG
CCTTCAACGTGGAGTGCTTAAGAAAATACGCTGCAACAACGAGTACTGGGAGACCTTCAGGAGA
ACCCCATCAGGCTGACCGAAGAGAACGTGGTAACATCACCAAGCTGAAGGGCCCAAGGCC
CTGCCCTGTTGCTAAGACCCACAACCTGAACATGCTGCAGGACATCCCAATGGACAGGTCGTGAT
GGACCTGAAGAGGGACGTGAAGGTGACACCCGGCCAAGCACACCGAGGAGAGGCCAAGGTG
CAGGTGATCCAGGCCGCTGACCCACTGCCACCGCTACCTGTGCGGCATCCACAGGGAGCTGGT
AGGCGGCTGAACGCCGTGCTGCCAACATCCACACCCCTGTTGACATGAGCGCCGAGGACTTC
GACGCCATCAGGCCGACTTCCAGCCGGCACTGCCACCGCTGGAGACCGACATGCCAGCTTC
GACAAGAGCGAGGATGACGCTATGCCCTGACCGCTCTGATGATCTGGAGGACCTGGCGTGG
CGCCGAGCTGCTCACCTGATCGAGGCTGCCCTCGCGAGATCAGCTCATCCACCTGCCAAC
ACCAAGTTCAAGTCGGCCTATGATGAAAAGCGGAATGTTCTGACCTGTTGTAACACCGTGA
TCAACATTGATGCCAGCAGGGTGTGCGGGAGAGGCTGACCCGAGCCCTGCGCTGCCCTCA
TCGGCGACGACAACATCGTAAGGGCGTGAAGGGCGAACAGCGACAAGCTGATGCCGACAGGTGCC
TGGCTGAACATGGAGGTGAAGATCATGACGCCGTGGTGGCGAGAAGGGCCCTACTTGTGCG
GGATTCATCCTGTGCGACAGCGTGACCGGACCGCTGCAGGGTGCGACCCCTGAAGAGGCTG
TTCAAGCTGGCAAGCCACTGGCGCTGACGATGAGCACCGACGATGACAGGGAGGGCCCTGCA
CGAGGAAAGCACCAGGTGGAACAGGGTGGCATCCTGAGCGAGCTGTGCAAGGCCGTGGAGAGC
AGGTACGAGACCGTGGGACCGCATCGTGTGGTATGACCACACTGCCAGCTCCGTCAAG
AGCTTCTCTACCTGAGGGGGCCCTATAACTCTACGGCTAACCTGAATGGACTACGACATAGT
CTAgccaccATGagcaagatctacatcgacgagcggagcaacgcggagatcgatgtgcaggccatcaagaccatcgcatcg
ggcgccaccgcggccagctgaccaggcagctgaacatggagaagcgggaggtaacaaggccctgtacgacctgcagagg
cgctatggtgactccagcgacatccctcccggtgttcatgaccaccggccgacaagccgcacgcgcacgctatggcc
acgtgatcatcgacgacgtgagcaggagaagtccatgagggaggaccacaagagacgttcacgcacgtgatccccgccaagaaga
tcatcgactggaaaggcgccaaccccggtgaccgtatcaacgactgtccagatcaccaggaggactggagctccggatcg
gagcgtggggcccaacagccccacccatcgccctgcgtggacatcgacggcagggtgtcgacaccgcacggcaagac
aagcgggacgccaagaacaacgcgcacgtggcgtggacaagctgtggctacgtatccggttcTAactcgagcta
gtgactgacttagatctgttaccactaaaccgcctcaagaacaccgcataaggtcttaagactataataccacttacact
acaaaatgtgtccccaaaaatgtagccattcgatctgtcttaataaaaagaaaagttttcacatttagAGCTCCGTCAAG
AGCTTCTCTACCTGAGGGGGCCCTATAACTCTACGGCTAACCTGAATGGACTACGACATAGT
CTAGCCACCATGGAAGATGCCAAAACATTAAGAAGGGCCAGCGCCATTCTACCCACTCGAAGAC
GGGACGCCGGCAGCTGCACAAAGCCATGAAGCCTACGCCCTGGTGGCCGACCATCGC
TTTACCGACGACATATCGAGGTGGACATTACCTACGCCAGTACTCGAGATGAGCGTTCGGCTGG
CAGAAGCTATGAAGCGCTATGGGCTGAATAACAAACCATCGGATCGTGGTGCAGCGAGAATAGCT
TGCAGTTCTCATGCCGTGTTGGTGGCCCTGTTATCGGTGTTGAGGAGCTGGCC
CTACAAACGAGCGCGAGCTGCTGAACAGCATGGCAGCGCCACCGCTGTTACCGTGGAGCTGGCC
CTACAAACGAGCGCGAGCTGCTGAACAGCATGGCAGCGCCACCGCTGTTACCGTGGAGCTGGCC

-continued

GAAAGGGCTGAAAAGATCCTCAACGTGAAAAGAAGCTACCGATCATACAAAAGATCATCATCAT
 GGATAGAACGACCGACTACCAGGGTCTCCAAGCATGTACACCTTCGTGACTTCCATTGCCACCC
 GGCTTCAACGAGTACGACTTCGTGCCGAGAGCTTCGACCGGGACAAAACCATGCCCTGATCATGA
 ACAGTAGTGGCAGTACCGATTGCCAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTCGCGATT
 CAGTCATGCCCGGACCCCATTTCGGCAACCAGATCATCCCCGACACCGCTATCCTCAGCGTGGTG
 CCATTTACCCACGGCTTCCGGCATGTTACCACGCTGGGCTACTTGATCTGCGGCTTCGGGCGTGC
 CATGTACCGCTTCGAGGAGGAGCTATTCTGCGCAGCTGCAAGACTATAAGATTCAATCTGCCCTG
 CTGGTGCCCACACTTTAGCTCTCGCTAAGAGCACTCTCATCGACAAGTACGACCTAACGAACTT
 GCACGAGATGCCAGCGCGGGCGCGCTCAGCAAGGAGGTAGGTGAGGCCGTGGCCAAACGC
 TTCCACCTACCAGGCATCCGACAGGGTACGCCCTGACAGAAACAACCAGCGCCATTCTGATCACCC
 CGAAGGGGACGACAAGCCTGGCGAGTAGGCAAGGTGGTGCCTTCTCGAGGCTAAGGTGGTG
 GACTGGACACCGTAAGACACTGGGTGTGAACCAGCGCGGAGCTGTGCGTCCGTGGCCCATG
 ATCATGAGCGGTACGTTAACAAACCCCGAGGCTACAAACGCTCTCATCGACAAGGACGGCTGGTG
 CACAGCGCGACATCGCCTACTGGGACGAGGACGAGCACTTTCATCGTGGACCGGCTGAAGTCC
 CTGATCAAATACAAGGGTACCAAGGTAGCCCCAGCGAACTGGAGAGCATCTGCTGCAACACCC
 AACATCTCGACGCCGGGTGCGCGCTGCCGACGACGATGCCGGAGCTGCCGCCAGTC
 GTCGTGCTGGACACCGTAAACCATGACCGAGAAGGAGATCGTGGACTATGTGGCCAGCCAGTT
 ACAACCGCCAAGAAGCTGCGCGGTGGTGTGTTCTGAGGAGCTAAAGGACTGACCG
 CAAGTTGGACGCCGCAAGATCCGAGATTCTCATTAAGCCAAGAAGGGCGCAAGATGCCGT
 GTAACTCGAGTATGTTACGTGCAAAGGTGATTGTCACCCCCGAAAGACCATATTGTGACACACCC
 CAGTATCACGCCAAACATTACAGCCGGTGTCAAAACCGCGTGGACGTGGTTAACATCCCTG
 TGGGAGGATCAGCGTAATTATTATAATTGGCTTGGTGTGGCTACTATTGTGGCCATGTACGTGCT
 GACCAACCAGAACATAATTGAATAACAGCAGCAATTGGCAAGCTGCTTACATAGAACCTCGCGCGA
 TTGGCATGCCGCTTAAATTATTATTATTCTTCTTCTTCCGAATGGATTGGTTAAATAT
 TTCAAAAAAAAAAAAAAAATctagAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 AAA
 (RNA sequence for STARR Fluc IRES-E3L)

SEQ ID NO: 118

AUGGGCGCGCAUGAGAGAAGCCCAGACCAUUACCUACCCAAAUGGAGAAA
 GUUCACGUUGACAUUCGAGGAAGACAGCCAUUCCUCAGAGCUUUGCAGCGGAG
 CUUCCCGCAGUUUGAGGUAGAAGCCAAGCAGGUCACUGAUAAUGACCAUGCUA
 AUGCCAGAGCGUUUUCGCAUCUGGCUUAAAACUGAUCGAAACGGAGGUGGA
 CCCAUCCGACACGAUCCUUGACAUUGGAAGUGCGCCGCCCGAGAAUGUAUU
 CUAAGCACAAGUAUCAUGUAUCUGUCCGAUGAGAUGUGCGGAAGAUCCGGA
 CAGAUUUGUAUAGUAUCACUAAGCUGAAGAAAACUGUAAGGAAUAAACU
 GAUAAGGAAUUGGACAAGAAAUGAAGGAGCUGGCCCGCUCAUGAGCGACC
 CUGACCUGGAAACUGAGACUAUGUGCCUCCACGACGAGCAGUGUGUCGCUAC
 GAAGGGCAAGUCGCUGUUUACAGGAUGUAUACGCCGUCGACGGCCCCACCAG
 CCUGUACCAACAGGCCAACAGGGCGUGAGGGUGGCCUACUGGAUCGGCUUCG
 ACACCAACACCCUCAUGUUCAAGAACCUUGGCCGGCGCUACCCAGCUACAGC
 ACCAACUGGGCCGACGAGACCGUGCUGACCGCCAGGAACAUCCGUGUGCAG

-Continued

CAGCGACGUGAUGGAGAGGAGCCGGAGAGGAUCAGCAUCCUGAGGAAGAAA
 UACCUGAAGCCCAGCAACAACGUGCUGUUUCAGCGUGGGCAGCACCAUCACCA
 CGAGAAGAGGGACCUGCUCAGGAGCUGGCCAGCGUGGUUCCACCUA
 GGGGCAAGCAGAACAUACACCUGCAGGUGCGAGACCAUCGUGAGCUGCGACGGC
 UACGUGGUGAAAGAGGAUCGCCAUCAGCCCCGCCUGUACGGCAAGGCCAGCG
 CUACGCCGUACAAUGCACAGGGAGGGCUUCCUGUGCUGCAAGGGACCGACA
 CCCUGAACGGCGAGAGGGUGAGCUUCCCCUGUGUGCACCUACGUGCCGCCACC
 CUGUGCGACCAGAUGACCGGGCAUCCUGGCCACCGACGUGAGCGCCGACGAC
 CCAGAACAGCUGCUCGUGGCCUGAACCAAGAGGAUCGUGGUACGGCAGGACCC
 AGAGGAACACCAACACAAUGAAGAACUACCUACUGCCGUGGUGGCCAGGCU
 UUCGCCAGGUGGCCAAGGAGUACAAGGAGGACCGAGGAAGACGAGAGGCCU
 GGGCUGAGGGACAGGCAGCUGGUGAUGGGCUGCUGCUGGGCUUCAAGGCG
 ACAAGAACCCAGCAUCUACAAGAGGCCGACACCCAGACCAUCAAGGUG
 AACAGCGACUUCACAGCUUCGUGCUGGCCAGGAUCGGCAGCACACCCUGGA
 GAUCGGCCUGAGGAACCCGGAUCAAGGAAGAUGCUGGAGGAACACAAGGAGGCC
 GCCCACUGAUACCGCCGAGGACGUGCAGGAGGCCAAGUGCGCUGGCCGACG
 GCCAAGGAGGUGAGGGAGGCCAGGAACUGAGGGCCCGCCUGCCACCCUGGC
 UGCCGACGUGGAGGAACCCACCCUGGAAGCCGACGUGGACCUAUGCUGCAGG
 AGGCCGGCGCCGAAGCUGGGAGACACCCAGGGCCUGAUCAAGGUGACCAGC
 UACGACGGCGAGGAACAGAUCCGAGCUACGCCGUGCUGAGCCCACAGGCC
 GCUGAACGUCCGAGAACGUGACUGCAUCCACCCACUGGCCGAGCAGGUGAUC
 UGAUCACCCACAGCGGCAGGAAGGGCAGGUACGCCGUGGAGCCUACACGGC
 AAGGUGGUGGCCAGGGCACGCCAUCCCGUGCAGGACUCCAGGCC
 GACCGAGAGGCCACCAUCGUGUACACGAGAGGGAGUUCGUGAACAGGUAC
 UGCACCAUAUCGCCACCCACGCCGGAGGCCUGAACACCGACGAGGAUAC
 AAGACCGUGAAGCCCAGCGAGCACGACGGCGAGUACCUACGACAG
 GAAGCAGUGCGUGAAGAAAGAGCUGGGAGCCUGGGACUGACCGCGAG
 CUGGUGGACCCACCCUCCACAGAGUUCGCCUACGAGAGGCCUGAGGACCA
 CGCCGCUCCCUACCAGGCCACCAUCGGCGUGUACGGCGUGGCCGGCG
 GAAAGAGCGGCAUCAAGAGCGCGUGACCAAGAAAGACCUUGGGUGAC
 GCGCAAGAAGAGAACUGCGCCGAGAUCAUCAGGGACGUGAAGAAUGAAG
 GCCUGGACGUGAACGGCGCACCGUGGACAGCGUGGUGCUGAACGGCUG
 CACCCCGUGGAGACCCUGUACACGACGAGGCCUUCGCUUGCCACGCC
 CCUGAGGGCCCUUGAUCGCCAUCAUCAGGCCAAGAAAGCCGUGCUGGCC
 ACCCAAGCAGUGCGGCUUCUCAACAGAUGUGCCUGAAGGUGCACU
 CACGAGAACUGCACCCAGGUGUUCCACAAGAGCAUCAGCAGGGGUG
 GAGCGUGACCGCGUGAGGACCCUGUUCACAGAAGAAAGAGGACCA
 CCAACCCCAAGGAGACCAAAUCGUGAUCGACACCACAGGCAGCAC
 AACGAGGACGACCCUGAUCGCCUGACCUGCUCAGGGCUGGGUGAAG
 GAUCGACUACAAGGGCAACGAGAACUAGACGCCGUGCCAGCCAGGG
 CCAGGAAGGGCGUGUACGCCGUGAGGUACAAGGUGAACGAGAACCC
 ACUGUAC

-continued

GCUCCCACCAGCGAGCACGUGAACGUGCUGCUGACCCAGGAGCAGGAG
 CGUGUGGAAGACCCUGGCCGGCGACCCUGGAUCAAGACCCUGACCGCCAAGU
 ACCCGGGCAACUUCACCGCCACCAUCGAAGAGUGGCAGGCCAGCACGCC
 AUCAUGAGGCACAUCUCCUGGAGAGGCCGACCCACCGACGUGUUCCAGAACAA
 GGCAACCGUGUGCCUGGGCAAGGCCUUGGUGCCGUGCUGAAGAGACCCGGCA
 UCGACAUGACCACAGAGCAGUGGAACACCGUGGACUACUUUCGAGACCGACAAG
 GCCCACAGCGCCGAGAUCGUGCUGAACAGCUGUGCGUGAGGUUCUUCGGCU
 GGACCUUGGACAGCGGCCUGUUUCAGCGCCCCACCGUGGCCACUGAGCAUCAGGA
 ACAACCACUGGGACAACAGCCCCAGCCAAACAUUGUACGGCCUGAACAGGAG
 GUGGUCAGGCAGCUGAGCAGGCCUACCCACAGCUGGCCAGGGCGUGGCCAC
 CGGCAGGGUGUACGACAUGAACACCGGACCCUGAGGAACUACGACCCAGGA
 UCAACCUGGUGCCGUGAACAGGCCUGGCCACGCCUUGGUGCUGCAC
 AACGAGCACCCACAGAGCGACUUCAGCUCCUUCGUGAGCAAGCUGAAAGGAG
 GACCGUGCUGGUUCGUGGGCGAGAAGCUGAGCGUGGCCAGGGCGUGGUGGAC
 UGGCUGAGCGACAGGCCGAGGCCACCUUCGGCCAGGCUGGACCUCCGGCAU
 CCCGGCGACGUGCCAAGUACGACAUCUUCGUGAACGUGACAGGACCCAU
 ACAAGUACCACCAUUUACCAGCAGUGCGAGGACACGCCAUCAAGCUGAGCAUG
 CUGACCAAGAAGGCCUGCCUGCCACCUACGUGGUACCCCGGAGGCACCCUGCG
 CGGUACGGCUACGCCACAGGCCAGCGAGAGCAUCAUUGCGCAUCGCCA
 GCGCUUCAAGUUCAGCAGGGUGUGCAAACCAAGAGCAGGCCUGGAGGAACC
 GAGGUGCUGUUUCGUGUCAUCGGCUACGACCGGAAGGCCAGGACCCACAACC
 CUACAAGCUGAGCAGCACCCUGACAAACAUUCACCCGGCAGCAGGCUGCAC
 AGGCCGGCUGGCCAGCUACCCACGUGGUACAGGGCGUAUCGCCACCGCC
 ACCGAGGGCGUGAUCAUCAACCCUGCAACAGCAAGGCCAGGCCAGGCG
 AGUGUGCGCGCCUGUACAAGAAGUUCCCGAGAGCUUCGACCCUGCG
 UCGAGGGUGGCAGGCCAGGCGUGGUGAAGGGCGCCUAAGCACAUCAUCCAC
 GCCGUGGGCCCCAACUCAACAAGGUGAGCGAGGUGGAAGGGGACAAGCAGCU
 GCGCGAAGCCUACGAGAGCAUCGCCAAGAUCGUGAACGACAAUAACUACAAGA
 GCGUGGCCAUCCCCACUGCUCACCGCAUCUUCAGCGGACAAGGGACAGG
 CUGACCCAGAGCCUGAACACCACUGCUCACCGCCUGGACACCCACCGAUGCG
 CGUGGCCAUCAUCUGCAGGGACAAGAAGUGGGAGAUGACCCUGAAGGGAGG
 GUGGCCAGGGCGGGAGGCGUGGAAGAGAACUGCAUCAGCGACGACUCCAGCG
 GACCGAGCCGACGCCAGCGUGGUGAGGGUGCACCCCAAGAGCUCGGCG
 GCAGGAAGGGCUACAGCACCGCGACGGCAAGACCUUCAGCUACCCUGGAGGG
 ACCAAGUUCCACCCAGGCCUAAGGACAUCGCCGAGAUCAACGCUAUGUGGCC
 CGUGGCCACCGAGGCCAACGAGCAGGUGUGCAUGUACAUCCUGGGCGAGAGCA
 UGUCCAGCAUCAGGAGCAAGUGCCCGUGGAGGAAAGCGAGGCCAGCACCCA
 CCCAGCACCCUGCCUGCCUGCAUCCACGCCAUGACACCCGAGAGGGUGCA
 CGGGCUGAAGGCCAGCAGGCCAGCAUCAGCUGUGCGAGCUUCCAC
 UGCCCAAGUACAGGAUCACCGCGUGCAGAAGAUCAGUGCAGGCCAUC

-continued

CUGUUCAGCCCCAAGGUGCCCCCUACAUCCACCCCAGGAAGUACCUGGUGGA
 GACCCCAACCGUGGACGAGACACCCGAGCCAAGCGCCGAGAACAGAGCACCG
 AGGGCACACCCGAGCAGCCACCCUGAUCACCGAGGACGAGACAAGGACCCGG
 ACCCCAGAGCCCAUCAUAUCGAGGAAGAGGAAGAGGACAGCAUCAGCCUGCU
 GAGCGACGGCCCCACCCACCAGGUGCUGCAGGUGGGGCCACAUCCACGGCC
 CACCCAGCGUGGUCCAGCUCCAGCUGGAGCAUCCCACAGCCAGCAGCACUUCGAC
 GUGGACAGCCUGAGCAUCCUGGACACCCUGGAGGGGCCAGCGUGACCUCGG
 CGCCACCAGCGCCGAGACCAACAGCUACUUCCAGCAAGAGCAUGGAGUCCUGG
 CCAGGCCGUGCCAGCUCCAGGACCGUGUUCAGGAACCCACCCAGCUC
 CCAGGACCAGGACCCAAAGCCUGGCUCCAGCAGGGCUGCAGCAGGACAGC
 CUGGUGACACCCACCCGGCGUGAACAGGGUGAUACACCAGGGAGGAUUCGAGGCC
 UGCCCCAACCCACCCGGCGUGAACAGGGUGAUACACCAGGGAGGAUUCGAGGCC
 UUCGUGGCCAGCAACAGAGACGGUUCGACGCCGGCCUACAUUCAGCAG
 CGACACCGGCCAGGGACACCUGCAGCAAAGAGCGUGAGGCAGGACUAGCUGG
 GCGAGGUGGUGCUGGAGAGGACCGAGCUGGAAACAGCUACGCCAGGCGUG
 GACCAAGGAGAAGGAGGAACUGCUCAGGAAGAACUGCAGCUGAACCCACCC
 AGCCAACAGGAGCAGGUACCAGAGCAGGAAGGUGGAGAACAGGAAGGCCAUC
 ACCGCCAGCGGAUCCUGCAGGCCUGGGACACUACCUGAAGGCCAGGGCAA
 GGUGGAGUGCACAGGACCCUGCACCCGUGGCCACUGUACAGCUCCAGCGUGA
 ACAGGGCCUUCUCCAGCCCCAAGGUGGCCUGGGAGGCCUGCAACGCUAUGCUG
 AAGGAGAACUCCCCACCGUGGCCAGCUACUGCAUCAUCCCGAGUACGACGC
 CUACUGGACAUGGUGGACGGGCCAGCUGCUGCCUGGACACCGCCAGCUUCU
 GCCCGCCAAGCUGAGGACCUUCCCAAGAACACAGCUACCUUGGAGGCCACC
 AUCAAGGAGCAGCCUGGCCAGCAGCAUCCAGAACACCCUGCAGAACGUGCUGG
 CGCUGCCACCAAGAGGAACUGCAACGUGACCCAGAUGAGGGAGCUGCCGUG
 UGGACAGCGCUGCCUUAACUGGGAGUGCUUCAAGAAAAUACGCCUGCAACAC
 GAGUACUGGGAGACCUUAAGGAGAACCCAUCAAGGUGACCCUGAAGAGAACGU
 GGUGAACUACUACCAAGCUGAAGGGCCCAAGGCCUGCCUGUUCGCUA
 AGACCCACAACCUGAACUGCAGGACAUCCAAUGGACAGGUUCGUGAUG
 GACCUGAAGAGGGACGUGAAGGUGACACCCGGCACCAAGCAGCACCCAGGAG
 GCGCAAGGUGCAGGUGAUCCAGGCCUGACCCACUGGCCACCGCCUACUGU
 GCGGCAUCCACAGGGAGCUGGUGAGGCCUGAACGCCUGCUGCUGCCAAAC
 AUCCACACCCUGUUCGACAUGAGCGCCGAGGACUUCGACGCCAUCAUCGCCA
 GCACUUCCAGCCCGGACUGCGUGCUGGAGACCGACAUCCAGCUUCGACA
 AGAGCGAGGAUGACGCUAUGGCCUGACCGCUCUGAUGAUCUGGAGGACUG
 GCGCUGGACGCCAGCUGCUCACCCUGAUCGAGGCUGCCUUCGGCGAGAAC
 CUCCAUCCACCUCCACCAAGACCAAGUUAAGUUCGGCGCUAUGAUGAAA
 GCGGAAUGUUCUGACCCUGUUCGUGAACACCGUGAUCAACAUUGUGAUCC
 AGCAGGGUGCUGCGGGAGAGGCUGACCGCAGCCCCUGCGCUGCCUUC
 CGACGACAACAUUGUGAAGGGCUGAAAAGCGACAAGCUGAUGGCCGACAGG

-continued

UGCGCCACCUGGCGUAACAUGGAGGUGAAGAUCAUCGACGCCGGGGCGA
 GAAGGCCCCUACUUUCUGCGGCGGAUUCAUCCUGUGCGACAGCGUGACCGCA
 CCGCCUGCAGGGUGGCCGACCCCCUGAAGAGGGCUGUCAAGCUGGGCAAGCCA
 CGGCCUGACGAUGAGCACGACGAUGACAGGGCGAGGGCCUGACGAGGA
 AACGACCAAGGGAAACAGGGUGGGCAUCCUGACGGACUGUGCAAGGCCUG
 GAGAGCAGGUACAGAGACCGUGGGCACCAGCAUCAUUCUGUGAUGGCCUAUGACAC
 ACUGGCCAGCUCCGUCAAGAGCUUCUCCUACCUGAGGGGGCCUAAUCUC
 UCUACGGCUAACUGAAUUGGACUACGACAUAGUCUAGUCCGCAAGGCCCA
 CCAUGGAAGAUGCCAAAACAUUAAGAAGGGCCAGCGCCAUCUACCCACUC
 GAAGACGGGACCGCCGGCGAGCAGCUGCACAAAGCCAUGAAGCGCUACGCCU
 GGUGCCGGCACCAUCGCCUUUACCGACGCACAUACUGAGGGACAUUACCU
 ACGCCGAGUACUUCGAGAUGAGCGUUCGGCUGGCAGCGAGAUAGCGCUA
 UGGGCUGAAUACAAACCAUCGGAUCUGGGUGUGCAGCGAGAAUAGCUUGCAG
 UUCUUCAUGCCCGUGUUGGGUGCCUGUUCAUCGGUGGGCUGUGGCCCCAGC
 UAACGACAUCUACAACGAGCGCAGCUGCUGAACAGCAUGGGCAUCAGCCAGC
 CCACCGUCGUAAUUCGUGAGCAAGAAAGGGCUGCAAAGAACCUACGUGCAA
 AAGAAGCUACCGAUCAUACAAAGAUCAUCAUCAUGGAUAGCAAGACCGACU
 ACCAGGGCUUCAAAGCAUGUACACCUUCGUGACUUCCAUUUGCCACCCGGC
 UUCAACGAGUACGACUUCGUGCCGAGAGCUUCGACCGGGACAAACCAUCGC
 CCUGAUCAUGAACAGUAGUGGGCAGUACCGGAUUGCCCAAGGGGUAGGCCUAC
 CGCACCGCACCCUUGGUCCGAUUCAGUCAUGCCCGGACCCAUUCGGC
 AACAGAUCAUCCCCGACACCGCUAUCCUCAGCGUGGGCAUUCACCG
 CUUCGGCAUGUUCACCACGCGUGGGCAGUCAUGACUUCGGCUUUCGGUG
 UCAUGUACCGCUUCGAGGAGGACUACUUCUGCGCAGCUUGCAAGACUAA
 GAUUCAUCUGCCCUGCUGGGCCCACACUAAAAGCUUCUGCUAAGAGCA
 CUCUCAUCGACAAGUACGACCUAAGCAACUUCGACAGAGAUUCGCCAGGG
 GCGCCGUCAGCAAGGAGGUAGGUGAGGCCAAGCGCUUCCACCUACC
 AGGCAUCGACAGGGCUACGGCUGACAGAAACAACAGCGCCAUUCUGAUCA
 CCCCCGAAGGGACGACAAGCCUGGCGCAGUAGGCAAGGGUGGGCCUUC
 GAGGCUAAGGGUGGGACUUGGACACCGGUAGACACUGGGUGUGAAC
 GCGCGAGCUGUGCGUCCGUGGCCAUGAUCAUGAGGGCUACGUUAAC
 CCCGAGGUACAAACCGCUCUCAUCGACAAGGACGGCUGGCUGACAGCG
 CAUCGCCUACUGGGACGAGGACGAGCACUUCUCAUCGUGGACGGCUG
 CCCUGAUAAAACAAGGGCUACCGGUAGGCGGACAGGGCAACUGGGAG
 CUGCUGCAACACCCAAACAUUCGACGCCGGGUCGCCGGCUGCCGACGA
 CGAUGCCGGCAGCUGCCGCCAGUCGUGCGUGCUGGAACACGGUAAA
 UGACCGAGAAGGAGAUCGUGGACUAUGGGCAAGCCAGGUUACA
 AACGUGCGCGGGGGUGGUUGGUUCGUGGACGAGGGUGCCUAAGG
 GCAAGUUGGACGCCCGCAAGAUCCGCGAGAUUCUCAUUAAGGCC
 GGCAAGAUCGCCGUGUAACUGAGCCGAAACGCAAUAGCCAAAA
 AAAA

-continued

ACAAAAAAACAAAAAAACCAAAAAACAAAACACAUAAAACAGCCUG
UGGUUGAUCCCACCCACAGGCCAUUGGCGCUAGCACUCUGGUUAUCAGGU
ACCUUUGUGCGCUCGUUUUAACCCCCUCCCCAACUGUAACUUAGAAGUAAC
ACACACCGAUCAACAGUCAGCGUGGCACACCAGCCACGUUUGAUCAAGCAC
UCUGUUACCCGGACUGAGUAUCAAAGACUGCUCACGCGGUUGAAGGAGAA
AGCGUUCGUUAUCGGCAACUACUUUCGAAAACCUAGUAACACCGUGGAAGU
UGCAGAGUGUUUCGUCAGCACUACCCCAGUGUAGAUCAGGUCGAUGAGUCAC
CGCAUUCCCCACGGCGACCGUGGCGUGGCUGCGUUGCGGCCUGCCAUGG
GGAAACCCAUHGGACGCUCUAAUACAGACAUGGGCGAAGAGUCUAUUGAGC
UAGUUGGUAGUCCUCGGCCCCUGAAUGCUGCUAAUCCUAACUGCGGAGCACA
CACCUCAAGCCAGAGGGCAGUGUGUCGUAAACGGCACUCUGCAGCGGAAC
GACUACUUUHGGUGUCCGUGUUUCAUUUUAUCCUUAUCUGGUGCUUAUGG
UGACAAUUGAGAGAUCGUUACAUUAGCUAUUGGUGGCCAUCCGGUGAC
UAAUAGAGCUAUUAUUAUCCUUUUGUUGGUUUUAACACUUAGCUUGAAA
GAGGUAAAACAUUACAUUCAUUGGUAAAGUUGAAUACAGCAAAUAGGCAA
GAUCUACUCAUCGACGAGGGAGCAACGCCAGAGAUCUGUGUGCGAGGCCAUCAAGA
CCAUCGGCAUCGAGGGCGCACGCCGCCAGCUGACCAGGCGACUGUAACAUG
GAGAAGCGGGAGGUGAACAGGCCUGUACGACCUGCAGAGGAGCGCUAUGG
UGUACUCCAGCGACGACAUCCUCCCCGGUGGUCAUGACCACCGAGGGCGAC
AAGCCGACGCCACGCUAUGGCGACGUGAUCAUCGACGACGUGAGCAGGGAA
GAAGGUCAUGAGGGAGGACCACAAGAGCUUCGACGACGUGAUCCCCGCCAAGA
AGAUCAUCGACUGGAAGGGCGCAACCCCGUGACCGUGAUCACAGGUACUGC
CAGAUCCACAGGAGGGACUGGAGCUUCCGAUCGAGAGCGUGGGGCCAGCAA
CAGCCCCACCUUCUACGCCUGCGUGGACAUCGACGGCAGGGUGUUCGACAAGG
CCGACGGCAAGAGCAAGCGGGACGCCAAGAACACGCCCAAGCUGGGCGUG
GACAAGCUGCUGGGCUACGUGAUCAUCCGUUCUAAACGUAGUUAACGUGCA
AAGGUGAUUGUACCCCCGAAAGACCAUAUUGUGACACACCCUAGUAUCAC
GCCAAACAUUACAGCCGGUGUCAAAACCGCGUGGACGUGGUUAACAU
CCUGCUGGGAGGAUCAGCCGUAAUUAUUAUUGGUUGGUGCUGGUACU
AUUGUGGCCAUUACGUGUGACCAACGAAACAUAAUUGAAUACAGCAGC
AAUUGGCAAGCUGCUUACAUAGAACUCGCCGAUUGGCAUGGCCUUAAA
UUUUUAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU
UUCAAAAAAAAAAAAAAAACAUUACUAGAAGAAGAAGAAGAAGAAGAAGAAGA
AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA
AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA

(RNA sequence for STARR Fluc IRES-E3L (short 3' UTR))
AUGGGCGGCCAUGAGAGAACGCCCAGACCAAUUACCUCACCCAAAUGGAGAAA
GUUACGUUGACAUCGAGGAAGACAGCCCCAUUCCUCAGAGCUUUGCAGCGGAG
CUUCCCGCAGUUUUGGGUAGAGCAGCAGGGCACUGUAUAUGACCAUGCUA
AUGCCAGAGCGUUUUCGCAUCUGGCUUCAAAACUGAUCGAAACGGAGGUGGA
CCCAUCCGACACGAUCCUUGACAUUUGGAAGUGGCAGCGCCCGCAGAAUGUAU

SEQ ID NO: 119

-continued

CUAAGCACAAGUAUCAUUGUAUCUGUCCGAUGAGAUGUGCGGAAGAUCCGGA
 CAGAUUGUUAAGUAUGCAACUAAGCUGAAGAAAAACUGUAAGGAAUAACU
 GAUAAGGAAUUGGACAAGAAAUGAAGGAGCUGGCCCGCUAUGAGCGACC
 CUGACCUGGAAACUGAGACUUAUGUGCCUCCACGACGAGUCGUGUCGUAC
 GAAGGGCAAGUCCUGUUUACAGGAUGUAUACGCCGUCGACGGCCCCACCA
 CCUGUACCAACAGGCAACAAGGGCGUGAGGGUGGCCUACUGGAUCGGCUUC
 ACACCAACACCCUUCAUGUUCAAGAACCUUGGCCGGGCCUACCCCAGCUACAGC
 ACCAACUGGCCGACGAGACCGUGCUGACCGCAGGAACAUCCGUGUGCAG
 CAGCGACGUGAUGGAGAGGAGCCGGAGAGGAUGAGCAUCCUGAGGAAGAAA
 UACCUAGGCCAGCAACAACGUGCUGUUCAGCGUGGGCAGCACCAUCUACCA
 CGAGAAGAGGGACCUGCUCAGGAGCUGGCACCUGCCAGCGUGUUCACCUGA
 GGGGCAAGCAGAACUACACCUGCAGGUGCGAGACCAUCUGAGCGUGCGACGG
 UACGUGGUGAAAGAGGAUCGCCAUCAGCCCCGGCUGUACGGCAAGGCCAGCG
 CUACGCCGCUACAAUGCACAGGGAGGGCUUCCUGUGCUGCAAGGUGACCGACA
 CCCUGAACGGCGAGAGGGUGAGCUUCCCCGUGUGCACCUACGUGCCGCCACC
 CUGUGCGACCAGAUGACCGGGCAUCCUGGCCACCGACGUGAGCGCCGACGAC
 CCAGAAGCUGCUCGUGGCCUGAACCAAGAGGAUCGUGGUACACGGCAGGACCC
 AGAGGAACACCAACACAAUGAAGAACUACCUUGCUGGCCGUGGGGCCAGGCU
 UUCGCCAGGUGGGCCAAGGAGUACAAGGAGGACCAGGAAGACGAGAGGCCU
 GGGCUGAGGGACAGGCAGCUGGUGAUGGGCUGCUGCUGGGCCUUCAGGCGG
 ACAAGAUCACCAAGCAUCUACAAGAGGCCGACACCCAGACCAUCUCAAGGUG
 AACAGCGACUUCACAGCUUCGUGCUGCCAGGAUCGGCAGCAACACCCUGGA
 GAUCGGCCUGAGGACCCGGAUCAAGGAAGAUGCUGGGAGGAACACAAGGAGGCC
 GCCCACUGAUCACCGCCGAGGACUGGCAGGAGGCCAGUGCGCUGCCGACGAG
 GCCAAGGAGGUGAGGGAGGCCAGGAACUGAGGGCCUCCUGCCACCCUGGC
 UGCCGACGUGGAGGAACCCACCCUGGAAGCCAGGGCAGUGGACCUAAGGUGACC
 AGGCCGGCGCCGAAGCGUGGGAGACACCCAGGGCCUGAUCAAGGUGACCAGC
 UACGACGGCGAGGACAAGAACUCCGAGCUACGCCGUGCUGAGCCCACAGGCC
 GCUGAAGUCCGAGAAGCUGAGCUGCAUCCACCCACUGGCCAGCAGGUGAUC
 UGAUCACCCACAGCGGCAGGAAGGGCAGGUACGCCGUGGAGCCCACAGGCC
 AAGGUGGUGCUGGCCAGGGCCACGCCAUCCCCGUGCAGGACUCCAGGCC
 GAGCGAGAGGCCACCAUCGUGUACACGAGAGGGAGUUCGUGAACAGGUAC
 UGCACCAUAUCGCCACCCACGGCGGAGGCCUGAACACCGACGAGGAUAC
 AAAGACCGUGAAGGCCAGCGAGCACGACGGCGAGUACCUUGUACGACAUC
 GAAGCAGUGCGUGAAGAAAGAGCUGGUGACCGGCCUGGGACUGACCGCGAG
 CUGGUGGACCCACCCUCCACGAGUUCGCCUACGCCAGAGGCCUGAGGACCA
 CGCCGCUCCCUACCGGCCCACCAUCGGCGUGUACGGCGUGGCCGGCAGCG
 GAAAGAGCGGCAUCAUCAAGAGCGCCGUGACCAAGAAAGACCUUGGUGG
 GCCAAGAAGAGAACUGCGCCGAGAUCAUCAGGGACGUGAAGAAGAUGAAAG
 GCCUGGACGUGAACGCGCGACCGUGGACAGCGUGCUGCUGAACGGCUGCAAG

-continued

CACCCCGUGGAGACCCUGUACAUCGACGGGCCUUUCGUUCCACGCCGGCAC
 CCUGAGGGCCCUGAUCGCCAUCAUCAGGCCAAGAAAGCCGUGCUGCGCG
 ACCCCAAGCAGUGCGGCUUCUCAACAUGAUGUGCCUGAAGGUGCACUUAAC
 CACGAGAACUGCACCCAGGUGUUCACAGAGCAUCAGCAGGCCGGCACCAA
 GAGCGUGACCAGCGUGUGAGCACCCUGUUCUACGACAAGAAAUGAGGACCA
 CCAACCCAAGGAGACCAAAUCGUGAUCGACACCACAGGCAGCACCAAGCCC
 AAGCAGGACGACCUGAUCCUGACCUGCUUCAGGGCUGGGUGAAGCAGCUGCA
 GAUCGACUACAAGGGCAACGAGAUCAUGACGCCGUGCCAGGCCAGGCCUGA
 CCAGGAAGGGCUGUACGCCGUGAGGUACAAGGUGAACGAGAACCCACUGUAC
 GCUCCCACCAGCGAGCACGUGAACGUGCUGCUGACCAGGAGCAGGAGAU
 CGUGUGGAAGACCCUGGCCGGGACCCUGGAUCAAGACCCUGACCAGGCCAAGU
 ACCCGGCAACUUCACCGCCACCAUCGAAGAGUGGAGGCCAGCACGCC
 AUCAUGAGGCACAUCCUGGAGAGGCCGACCCCCACCGACGUGUUCAGAACAA
 GGCAACGUGUGCUGGGCCAAGGCCUUGGUGCCGUGCUGAACGAGGCCGGCA
 UCGACAUGACCACAGAGCAGUGGAACACCGUGGACUACUUCGAGACGACAAG
 GCCCACAGCGCGAGAUCGUGCUGAACAGCUGUGCGUGAGGUUUCGGCCU
 GGACCUUGGACAGCGGCCUGUUCAGCGCCCCACCGUGGCCACUGAGCAUCAGGA
 ACAACCACUGGGACAACAGCCCCAGCCAAACAUGUACGCCUGAACAGGAG
 GUGGUCAGGAGCUGAGCAGGGGUACCCACAGCUGCCAGGCCGUGGCCAC
 CGGCAGGGUGUACGACAUGAACACCGGCACCCUGAGGAACUACGACCCCCAGGA
 UCAACCUGGUGCCGUGAACAGGGCUGCCAGGCCACGCCUGGUGCUGCAC
 AACGAGCACCCACAGAGCGACUUCAGCUCCUUCGUGAGCAAGCUGAAAGGCAG
 GACCGUGCUGGUUCGGCAGAGAAGCUGAGCGUGCCGGCAAGAUGGUGGAC
 UGGCUGAGCGACAGGCCGAGGCCACCUUCCGGCCAGGCUGGACCUCCGGCAU
 CCCCGGCGACGUGCCAAGUACGACAUCAUCUUCGUGAACGUGCAGGACCCAU
 ACAAGUACCACCAAUUACCAGCAGUGCGAGGACCACGCCAUCAAGCUGAGCAUG
 CUGACCAAGAAGGCCUGCCUGCACCUAGACCCCGGAGGCACCUGCGUGAGCAU
 CGGCUACGGCUACGCCGACAGGCCAGCGAGAGCAUCAUUGGCCAUCGCCA
 GCGCUGUUCAGGACAGCAGGGUGUGCAAACAGAGCAGGCCUGGAGGAACC
 GAGGUGCUGUUCGUGUCAUCGGCUACGACCGGAAGGCCAGGACCCACACCC
 CUACAAGCUGAGCAGCACCCUGACAAACAUCAACCCGGCAGGCCAGGAGCA
 AGGCCGGCUGGCCGGCCAGCUACCAACGUGGUACGGGCGAUUAUCGCCACCGCC
 ACCGAGGGCGUGAUCAUCAACCGCUGCCAACAGCAAGGCCAGGCCGGAGGCG
 AGUGUGCGCGCCUGUACAAGAAGUUCCCGAGAGCUUCGACCGCAGGCCA
 UCGAGGGGGCAAGGCCAGGCGUGGUGAAGGGCGCCGCUAAGCACAUCUCCAC
 GCCGUGGGCCCAACUCAACAAGGUGAGCGAGGUGGAAGGCCAGGCCAGGCG
 GCGCGAAGCCUACGAGAGCAUCGCCAAGAUCGUGAACGACAUAACUACAAGA
 GCGUGGCCAUCCACUGCUCAGCACCGCAUCUUCAGCGGAACAAGGACAGG
 CUGACCCAGAGGCCUGAACCCACCGCUCACCGCCUGGACACCACCGAUGGCCA
 CGUGGCCAUCAUCUGCAGGGACAAGAAGUGGGAGAUGACCCUGAAGGAGGCC
 GUGGCCAGGCCGGAGGCCUGGAGAGAUCUGCAUCAGCGACGACUCCAGCGU

-continued

GACCGAGCCGACGCCGAGCUGGUGAGGGUGCACCCCCAAGAGCUCCUGGCC
 GCAGGAAGGGCUACAGCACCAGCGACGGCAAGACCUUCAGCUACCUGGAGGG
 ACCAAGUUCCACCAGGCCUAAGGACAUCGCCAGAGAUCAACGCUAUGUGGCC
 CGUGGCCACCAGGCCAACGAGCAGGUGUGCAUGUACAUCCUGGGCGAGAGCA
 UGUCCAGCAUCAGGAGCAAGUGCCCCUGGGAGGAAGCGAGGCCAGCACCCA
 CCCAGCACCCUGCCUGCCUGCAUCCACGCUAUGACACCCGAGAGGGUGCA
 GCGGCUGAAGGCCAGCAGGCCAGCAGAUCAACCGUGUGCAGCUCCUCCAC
 UGCCCAAGUACAGGAUCACCGCGUGCAGAAGAUCCAGUGCAGGCCAUC
 CUGUUCAGCCAAAGGUGCCCUCUACAUCCACCCAGGAAGUACCUGGUGGA
 GACCCACCCGUGGAGCAGAGACACCCGAGCCAAGCGCCAGAACCAGAGCACCG
 AGGGCACACCCGAGCAGCCACCCUGAUCAUCAGGAGGACGAGACAAGGACCG
 ACCCCAGGCCAUCAUUAUCGAGGAAGAGGAAGAGGACAGCAUCAGCCUGCU
 GAGCGACGGCCCCACCCACCAGGUGUGCAGGUGGGGGGACAUCACGGCC
 CACCCAGCGUGGCCAGCAGGUGCAUCCACACGCCAGCGACUUCGAC
 GUGGACAGCCUGAGCAUCCUGGACACCCUGGAGGGGCCAGCGUGACCUCCGG
 CGCCACCAGCGCCGAGACCAACAGCUACUUUCGCAAGAGCAUGGAGUCCUGG
 CCAGGCCCGUGCCAGCUCCAGGACCGUGUUCAGGAACCCACCCACCGCUC
 CCAGGACAGGACCCAGCAGGACCCAGCAGGUCCAGGAGCAGGACCG
 CUGGUGAGCACCCACCCGGCGUGAACAGGGUGAUACACCAGGGAGGAACUGGA
 GGCCUGACACCCAGCAGGACCCAGCAGGUCCAGGAGCAGGACUAGUCUGG
 UGUCCAACCCACCCGGCGUGAACAGGGUGAUACACCAGGGAGGAUUCGAGG
 UUCGUGGCCAGCAACAGAGACGGGUUCGACGCCGGCCUACAUUUCAGCAG
 CGACACCGGCCAGGGACACCUGCAGCAAAGAGCGUGAGGCAGACCGUG
 GCGAGGUGGUGCUGGAGAGGACCGAGCUGGAAACAGCUACGCCAGGCUG
 GACCAGGAGAAGGAGGAACUGCUAGGAAGAACUGCAGCUGAACCCACCC
 AGCCAAACAGGAGCAGGUACAGAGCAGGAAGGGUGGAACAUGAAGGCCAUC
 ACCGCCAGGCCAUCCUGCAGGCCUGGGACACUACCUGAAGGCCAGGGCAA
 GGUGGAGUGCUACAGGACCCUGCACCCGUGGCCACUGUACAGCUCCAGCGUGA
 ACAGGGCCUUCUCCAGCCCCAAGGUGGCCUGGGAGGCCUGCAACGCUAUGC
 AAGGAGAACUUCCCCACCGUGGCCAGCUACUGCAUCAUCCCGAGUACGACGC
 CUACCUGGACAUGGUGGACGGGCCAGCUGCUGCCUGGACACCGCCAGCUUC
 GCCCGCCAAGCUGAGGAGCUUCCCCAAGAAACACAGCUACCCUGGAGGCCACC
 AUCAAGGAGCGCCUGGCCAGGCCAUCCAGAACACCCUGCAGAACCGUGCUG
 CGCUGGCCACCAAGGAAACUGAACGUGACCCAGAUGAGGGAGCUGCCGGUG
 UGGACAGCGCUGCCUUAACGUGGAGUGCUUCAAGAAAACGCCUGCAACAAAC
 GAGUACUGGGAGACCUUCAAGGAGAACCCAUAGGCUGACCGAAGAGAACGU
 GGUGAACUACAUCAACCAAGCUGAAGGGCCCAAGGCCUGCCUUCGCUA
 AGACCCACAACCUAACAUGCUGCAGGACAUCCCAUAGGACAGGUUCGUGAUG
 GACCUGAAGAGGGACGUGAAGGUGACACCCGGCACCAAGCACCCGAGGAGAG
 GCCCAAGGUGCAGGUGAUCCAGGCCUGACCCACUGGCCACCGCCUACCUGU

- continued

GCGGCAUCCACAGGGAGCUGGUGAGGCAGGCGCUGAACGCCGUGCCUGGCCAAC
 AUCCACACCCUGUUUCGACAUGAGCGCCGAGGAUCUGACGCCAUCAUCGCCGA
 GCACUUCCAGCCGGGACUGCGUGGUGGAGACCGACAUCGCCAGCUUCGACA
 AGAGCGAGGAUGACGCUAUGGCCUGACCGCUCUGAUGAUCUGGAGGACUG
 GCGUGGACGCCGAGCUGCUCACCCUGAUCGAGGCUGCCUUCGGCGAGAUCAG
 CUCCAUCACCUGCCCACCAAGACCAAGGUCAAGUUCGGCGCUAUGAUGAAAA
 GCGGAAUGUUCUGACCCUGUUCGUGAACCCUGAUCGAGGCUAACAUGUGAUCGCC
 AGCAGGGUGCUGCGGGAGGGCUGACCGCAGCCCCUGCGCUGCCUCAUCGG
 CGACGACAACAUUGUGAAGGGCUGAAAAGCGACAAGCUGAUGGCCGACAGG
 UGCGCCACCUGGUGAACAUGGAGGUGAAGAUCAUUGACGCCGUGGUGGGCGA
 GAAGGCCCUUCCGCGGGGAAUUCUGCGGGGAAUCAUCCUGUGCGACAGCGUGGCC
 CCGCUGCAGGGUGGCCGACCCCCUGAAGAGGCUGUUCAGCUGGGCAAGCCA
 CUGGCCGUGACGAUGAGCACGACUGACAGGGGGCCUGCGAGGAGGA
 AAGCACCAGGUGGAACAGGGUGGGCAUCCUGAGCGAGCUGUGCAAGGCCGUG
 GAGAGCAGGUACGAGACCGUGGGCACCAGCAUCAUCGUGAUGGCUAUGGACAC
 ACUGGCCAGCUUCGUCAAGAGCUUCCUACUUGAGGGGGCCCCUUAACUC
 UCUACGGCUAACUGAAUUGGACUACGACAUGUCUAGUCCGCAAGGCC
 CCAUGGAAGAUGCACAAACAUUAAGAAGGGCCAGCGCCAUCUACCCACUC
 GAAGACGGGACCGCCGGCGAGCAGCUGCACAAAGCCAUGAAGCGCUACGCC
 GGUGCCGGCACCAUCGCCUUUACCGACGCACAUACGAGGUGGACAUUACCU
 ACGCCGAGUACUUCGAGAUGAGCGUUCGGCUGGCAGAAGCUAUGAAGCGCUA
 UGGGCUGAAUACAAACCAUCGGAUCGUGGUGUGCAGCGAGAAUAGCUUGCAG
 UUUUCAUGCCGUGUUGGGUGGCCUGUUCAUCGGUGGGCUGUGGCCAGC
 UAACGACAUCUACAACGAGCCGAGCUGCUGAACAGCAUGGGCAUCAGCC
 CCACCGUGCUAUUCGUGAGCAAGAAAGGGCUGCAAAAGAUCCUAACGUGCAA
 AAGAAGCUACCGAUCAUACAAAGAUCAUCAUCAUGGAUAGCAAGACCGACU
 ACCAGGGCUUCAAAGCAUGUACCUUCGUGACUUCGCCAUUUGCCACCCGG
 UUCAACGAGUACGACUUCGUGCCGAGAGCUUCGACCGGACAAACCAUCGC
 CCUGAUAGAACAGUAGUGGAGUACCGGAUUGCCCAAGGGGUAGGCC
 CGCACCGCACCGCUUGUGUCCGAUUCAGUCAUGCCCGGACCCAUUCGGC
 AACCAGAUCAUCCCGACACCGCUAUCCUCAGCGUGGUGGCCAUUUCACCG
 CUUCGGCAUGUUCACCAACGCUUGGGCUACUUGAUCUGCGGCUUUCGGC
 UCAUGUACCGCUUCGAGGAGGAGCUAUUCUUGCGCAGCUUGCAAGACU
 GAUUCAAUCUGCCUGCUGGUGGCCACACUAUUAGCUUUCGCUAGAGCA
 CUCUCAUCGACAAGUACGACCUAAGCAACUUGCAGAGAUCGCCAGCGGG
 GCGCCGCUAGCAAGGAGGUAGGUGAGGCCGUGGCCAAACGCUUCCACCU
 AGGCACCGACAGGGCUACGCCUGACAGAAACAACCAGCGCCAUCUGAU
 CCCCCGAAGGGACGACAAGCCUGGCGAGUAGGCAAGGUGGUGGCC
 GAGGCUGAAGGUGGUGGCCACCCGGUAGACACUGGGUGUGAACCAGC
 GCGCGAGCUGUGCGUCCGUGGCCCAUGAUCAUGAGCGCUACGUUA
 CCCGAGGCUACAAACGCUUCUCAUCGACAAGGACGGCUGGCUG
 CACAGCGCGA

-continued

CAUCGCCUACUGGGACGAGGACGAGCACUUUUCAUCGUGGACCGGCUGAAGU
 CCCUGAUAAAACAAGGGCUACCAGGUAGCCCCAGCCGAACUGGAGAGCAUC
 CUGCUGCAACACCCCAACAUUUCGACGCCGGGUCGCCGGCUGCCCACGA
 CGAUGCCTGGCGAGCUGCCGCCAGUCGUGCUGUGGAACACGGUAAA
 UGACCGAGAAGGAGAUCCUGGACUAUGUGGCCAGCCAGGUUACAACGCCAAG
 AAGCUGCGGGGGUGUUGUUCGUGGACGAGGGUGCCUAAGGGACUGACCG
 GCAAGUUGGACGCCCAAGAUCCGCGAGAUUCUCAUUAAGGCCAAGAAGGGC
 GCAGAAGAUCGCCGUGAACUCGAGCCAAUCAGCGAAA
 ACAAAAAAAACAAAAAAACAAAAACACAUAAA
 ACAAAAAAAACAAAAAAACAAAAACACAUAAA
 UGGGUUGAUCCCCACCCAGGCCAUUGGGCGCUAGCACUCUGGUUAUCACGGU
 ACCUUUGUGGCCUGUUUAUACCCCCUCCCCAACUGUAACUUAGAAGUAAC
 ACACACCGAUCAACAGUCAGCGUGGCACACCAGCCACGUUUU
 UCUGUUACCCGGACUGAGUAUCAAAGACUGCUACGCCGUUGAAGGAGAA
 AGCGUUCGUUAUCGGCCAACUACUUCGAAAAACCUAGUAACACCGUGGAAGU
 UGCAGAGGUUUUCGCUCAGCACUACCCAGUGUAGAUCAGGUCAUGAGUAC
 CGCAUCCCCACGGCGACCGUGGCCUGCGUUGGCCUGCCAU
 GGAAACCCAUGGGACGCUAAUACAGACA
 UAGUUGGUAGUCUCCGGCCCCUGAAUGC
 CACCCUCAAGCCAGAGGGCAGUGUGU
 GACUACUUUGGUGUCCGUGUUU
 UGACAAUUGAGAGAU
 UAAUAGAGCU
 GAGGUAAA
 GAUCUACAU
 CCAUCGGCAUC
 GAGAAGCGGGAGG
 UGUACUCC
 AAGCCCC
 GAAGGU
 AGAUCA
 CAGAUCACC
 CAGCCCC
 CGACGGCAAG
 UACUAG
 UUU
 AAA
 AAA
 AAA

SEQUENCE LISTING

```

Sequence total quantity: 119
SEQ ID NO: 1      moltype = DNA  length = 18
FEATURE          Location/Qualifiers
misc_feature     1..18
note = Synthetic polynucleotide
source           1..18
mol_type = other DNA
organism = synthetic construct
modified_base    15
mod_base = OTHER
note = uracil
SEQUENCE: 1
gaggaaaactt aagatggg                                         18

SEQ ID NO: 2      moltype =   length =
SEQUENCE: 2
000

SEQ ID NO: 3      moltype =   length =
SEQUENCE: 3
000

SEQ ID NO: 4      moltype =   length =
SEQUENCE: 4
000

SEQ ID NO: 5      moltype = RNA  length = 129
FEATURE          Location/Qualifiers
misc_feature     1..129
note = Synthetic polynucleotide
source           1..129
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 5
tcaacacaaac atatacaaaa caaacgaatc tcaagcaatc aaggattcta cttctattgc 60
agcaatttaa atcatttctt ttaaagcaaa agcaattttc tgaaaatttt caccatttc 120
gaacgatag                                         129

SEQ ID NO: 6      moltype = RNA  length = 28
FEATURE          Location/Qualifiers
misc_feature     1..28
note = Synthetic polynucleotide
source           1..28
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 6
attattacat caaaaacaaaa agccgcca                                         28

SEQ ID NO: 7      moltype = RNA  length = 245
FEATURE          Location/Qualifiers
misc_feature     1..245
note = Synthetic polynucleotide
source           1..245
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 7
cttaaggggg cgctgcctac ggagggtggca gccatctcct tctcgccatc aagcttacca 60
tggtgccccca ggcctgtct ttggtcccgc tggctgtgtt cccccctctgc ttggcaagt 120
tccccatcta caccatcccc gacaagctgg ggcctgtggag ccccatcgac atccaccacc 180
tgtcctgccc caacaacacctc gtggctcgagg acgagggtctg caccacactg agcgggttct 240
cctac                                         245

SEQ ID NO: 8      moltype = RNA  length = 177
FEATURE          Location/Qualifiers
misc_feature     1..177
note = Synthetic polynucleotide
source           1..177
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 8
tgagtgtcgt acagcctcca ggccccccccc tcccgggaga gccatagttg tctcgaaac 60
cggtgagttac accggaaatttgc cggggaaagac tgggtccctt cttggataaa cccactctat 120
ggccggccat ttggcgtgc ccccgcaaga ctgcttagccg agtagtgtg gggtgcg                                         177

SEQ ID NO: 9      moltype = RNA  length = 89
FEATURE          Location/Qualifiers
misc_feature     1..89
note = Synthetic polynucleotide

```

-continued

```

source          1..89
               mol_type = other RNA
               organism = synthetic construct

SEQUENCE: 9
aattatttgt taaagaagta tattagtgtcattttccctc cgtttgctc agtttttc 60
ttctgtcaac cccacacgccc ttggcaca 89

SEQ ID NO: 10      moltype = RNA  length = 569
FEATURE          Location/Qualifiers
misc_feature     1..569
note = Synthetic polynucleotide
source           1..569
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 10
ctccctcccc ccccccctaac gttactggcc gaagccgctt ggaataaggc cggtgtgcgt 60
ttgtctatat gtatatttcc accatattgc cgtttttgg caatgtgagg gccccgaaac 120
ctggccctgt ctcttgacg agcatctca ggggtcttgc ceectctcgcc aaaggaaatgc 180
aaggctgtt gatatgtcgta aaggaaacgac ttccctcttgc agcttttgc agacaaacaa 240
cgtctgtgcgac gacccttgc aggcgcggg acccccccacc tggcgacagg tgccctctgcg 300
gccaaggcc acgtgtataa gatacacccctg caaaaggccg acaccccccac tgccacgtt 360
tgagttggat agttgtggaa agatgttataat ggctctccctt aagcttattt aacaaggggc 420
tgaaggatgc ccagaaggta ccccatgttgc tggatgtca tctggggctt cggtgacat 480
gtttagtgc tgtttagtgc aggttaaaaa acgtcttaggc ccccgaaacc acggggacgt 540
gttttacgtt tgtttagtgc aggttaaaaa acgtcttaggc ccccgaaacc acggggacgt 569

SEQ ID NO: 11      moltype = RNA  length = 44
FEATURE          Location/Qualifiers
misc_feature     1..44
note = Synthetic polynucleotide
source           1..44
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 11
gtcagcttc aaactctttt tttcttgcgtt gttgatttagtgc aata 44

SEQ ID NO: 12      moltype = RNA  length = 54
FEATURE          Location/Qualifiers
misc_feature     1..54
note = Synthetic polynucleotide
source           1..54
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 12
ctctcgctg agaaaaaaaaa tccacgaacc aatttctcaag caaccaggcag caccg 54

SEQ ID NO: 13      moltype = RNA  length = 52
FEATURE          Location/Qualifiers
misc_feature     1..52
note = Synthetic polynucleotide
source           1..52
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 13
acctgtgagg gttcgaaggaa agtagcgttgc tttttgttc ctagaggaag ag 52

SEQ ID NO: 14      moltype = RNA  length = 71
FEATURE          Location/Qualifiers
misc_feature     1..71
note = Synthetic polynucleotide
source           1..71
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 14
acacagaaaac attcgaaaaa acaaaatccc agtatcaaaa ttcttctttt tttttcatat 60
ttcgaaaaa c 71

SEQ ID NO: 15      moltype = RNA  length = 52
FEATURE          Location/Qualifiers
misc_feature     1..52
note = Synthetic polynucleotide
source           1..52
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 15
cagaaaaatt tgctacatttgc ttccacaaac ttcaaatattt attcattttt tt 52

SEQ ID NO: 16      moltype = RNA  length = 158
FEATURE          Location/Qualifiers

```

-continued

```

misc_feature      1..158
                  note = Synthetic polynucleotide
source           1..158
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 16
ctagtgactg actaggatct ggttaccact aaaccagcc caagaacacc cgaatggagt 60
ctctaagcta cataatacca acttacactt acaaaatgtt gtccccaaa atgttagccat 120
tcgttatctgc tcctaataaaa aagaagtt ctccacat                           158

SEQ ID NO: 17      moltype = RNA   length = 166
FEATURE          Location/Qualifiers
misc_feature     1..166
                  note = Synthetic polynucleotide
source           1..166
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 17
tgcaaggctg gccggaagcc cttgcctgaa agcaagatt cagcctggaa gagggcaaag 60
tggacgggg tggacaggag tggatgcgt aagatgttgt ttgaagctga tgggtgcag 120
ccctgcattt ctgagtcaat caataaagag ctttctttt acccat                           166

SEQ ID NO: 18      moltype = RNA   length = 143
FEATURE          Location/Qualifiers
misc_feature     1..143
                  note = Synthetic polynucleotide
source           1..143
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 18
acgcgcgaagc ctgcagccat gcgcacccac gcaccccggt gcctcctgcc tccgcgcagc 60
ctgcagcgggg agaccttgc cccgcacccag ccgtcctccct ggggtggacc ctgttaat 120
aaagattcac caagtttac gca                           143

SEQ ID NO: 19      moltype = RNA   length = 220
FEATURE          Location/Qualifiers
misc_feature     1..220
                  note = Synthetic polynucleotide
source           1..220
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 19
tagagcggca aacccttagct acactccata gctagtttct tttttttttt tttttttttt 60
ttttttttttttttttttt ctttctttt ctttctttt tttttttttt tttttttttt 120
tttgtggctc catcttagcc ctgtcacgg ctgtgtgt aagggtccgtg agccgcatga 180
ctgcagagag tgccgtactt ggtctctgtc cagatcatgt                           220

SEQ ID NO: 20      moltype = RNA   length = 170
FEATURE          Location/Qualifiers
misc_feature     1..170
                  note = Synthetic polynucleotide
source           1..170
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 20
acacatcaca accacaaccc tctcaggcta ccctgagaaaa aaaagacatg aagactcagg 60
actcatcttt tctgttgtgtg taaaatcaac accctaaggaa acacaaattt ctttaacat 120
ttgacttctt gtctctgtc tgcaattaat aaaaatggaa aagaatctac                           170

SEQ ID NO: 21      moltype = RNA   length = 110
FEATURE          Location/Qualifiers
misc_feature     1..110
                  note = Synthetic polynucleotide
source           1..110
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 21
gtctggacccctt cggttagccgtt tccctctgc cgtctggccccc cccaaacgggc 60
tcttgaccggccctt ggtctttgaa taaagtctgtt gttggcagca                           110

SEQ ID NO: 22      moltype = RNA   length = 123
FEATURE          Location/Qualifiers
misc_feature     1..123
                  note = Synthetic polynucleotide
source           1..123
                  mol_type = other RNA
                  organism = synthetic construct
SEQUENCE: 22
tagtgcatgc actggcacaa cgcgttgcggc ggtaagccaa tcgggtatac acggtcgtca 60

```

-continued

tactgcagac agggttcttc tacttgcaa gatagtctag agtagtaaaa taaatagttat aag	120 123
SEQ ID NO: 23 moltype = length = SEQUENCE: 23 000	
SEQ ID NO: 24 moltype = length = SEQUENCE: 24 000	
SEQ ID NO: 25 moltype = RNA length = 21 FEATURE Location/Qualifiers misc_feature 1..21 note = Synthetic polynucleotide source 1..21 mol_type = other RNA organism = synthetic construct SEQUENCE: 25 cacaaagagt aaagaagaac a	21
SEQ ID NO: 26 moltype = RNA length = 21 FEATURE Location/Qualifiers misc_feature 1..21 note = Synthetic polynucleotide source 1..21 mol_type = other RNA organism = synthetic construct SEQUENCE: 26 aacactaaaa gtagaagaaaa a	21
SEQ ID NO: 27 moltype = RNA length = 21 FEATURE Location/Qualifiers misc_feature 1..21 note = Synthetic polynucleotide source 1..21 mol_type = other RNA organism = synthetic construct SEQUENCE: 27 ctcagaaaaga taagatcagc c	21
SEQ ID NO: 28 moltype = RNA length = 21 FEATURE Location/Qualifiers misc_feature 1..21 note = Synthetic polynucleotide source 1..21 mol_type = other RNA organism = synthetic construct SEQUENCE: 28 aaccatatcgaa aagaaacccaa a	21
SEQ ID NO: 29 moltype = RNA length = 21 FEATURE Location/Qualifiers misc_feature 1..21 note = Synthetic polynucleotide source 1..21 mol_type = other RNA organism = synthetic construct SEQUENCE: 29 ctctaatcac caggagataaa a	21
SEQ ID NO: 30 moltype = RNA length = 21 FEATURE Location/Qualifiers misc_feature 1..21 note = Synthetic polynucleotide source 1..21 mol_type = other RNA organism = synthetic construct SEQUENCE: 30 gagagagatc ttaacaaaaaa a	21
SEQ ID NO: 31 moltype = RNA length = 21 FEATURE Location/Qualifiers misc_feature 1..21 note = Synthetic polynucleotide source 1..21 mol_type = other RNA organism = synthetic construct SEQUENCE: 31	

-continued

tgtgtaacaa caacaacaac a	21
SEQ ID NO: 32	moltype = RNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Synthetic polynucleotide
source	1..21
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 32	
ccgcagtagg aagagaaaac c	21
SEQ ID NO: 33	moltype = RNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Synthetic polynucleotide
source	1..21
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 33	
aaaaaaaaaa gaaatcataa a	21
SEQ ID NO: 34	moltype = RNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Synthetic polynucleotide
source	1..21
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 34	
gagagaagaa agaagaagac g	21
SEQ ID NO: 35	moltype = RNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Synthetic polynucleotide
source	1..21
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 35	
caattaaaaa tacttaccaa a	21
SEQ ID NO: 36	moltype = RNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Synthetic polynucleotide
source	1..21
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 36	
gcaaacagag taagcgaaac g	21
SEQ ID NO: 37	moltype = RNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Synthetic polynucleotide
source	1..21
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 37	
gcgaagaaga cgaacgcaaa g	21
SEQ ID NO: 38	moltype = RNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Synthetic polynucleotide
source	1..21
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 38	
ttaggactgt attgactggc c	21
SEQ ID NO: 39	moltype = RNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Synthetic polynucleotide
source	1..21
	mol_type = other RNA
	organism = synthetic construct

-continued

```

SEQUENCE: 39
atccatcgaa ttccggaaaaa g                                21

SEQ ID NO: 40      moltype = RNA  length = 21
FEATURE          Location/Qualifiers
misc_feature     1..21
                  note = Synthetic polynucleotide
source           1..21
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 40
aaaacaaaag ttccggaaaaa g                                21

SEQ ID NO: 41      moltype = RNA  length = 21
FEATURE          Location/Qualifiers
misc_feature     1..21
                  note = Synthetic polynucleotide
source           1..21
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 41
tttacatcaa ataagaaggc a                                21

SEQ ID NO: 42      moltype = RNA  length = 21
FEATURE          Location/Qualifiers
misc_feature     1..21
                  note = Synthetic polynucleotide
source           1..21
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 42
gggtggggagg tgagatttct                                21

SEQ ID NO: 43      moltype = RNA  length = 21
FEATURE          Location/Qualifiers
misc_feature     1..21
                  note = Synthetic polynucleotide
source           1..21
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 43
tgatttagaa actacaaaggc c                                21

SEQ ID NO: 44      moltype = RNA  length = 21
FEATURE          Location/Qualifiers
misc_feature     1..21
                  note = Synthetic polynucleotide
source           1..21
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 44
cattttcaa tttcataaaa c                                21

SEQ ID NO: 45      moltype = RNA  length = 21
FEATURE          Location/Qualifiers
misc_feature     1..21
                  note = Synthetic polynucleotide
source           1..21
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 45
ttacttttaa gcccaacaaa a                                21

SEQ ID NO: 46      moltype = RNA  length = 21
FEATURE          Location/Qualifiers
misc_feature     1..21
                  note = Synthetic polynucleotide
source           1..21
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 46
ggcgtgtgtg tgtgttgtt a                                21

SEQ ID NO: 47      moltype = RNA  length = 21
FEATURE          Location/Qualifiers
misc_feature     1..21
                  note = Synthetic polynucleotide
source           1..21
                  mol_type = other RNA

```

-continued

organism = synthetic construct

SEQUENCE: 47
gtggtaagg ggaagggttta g 21

SEQ ID NO: 48 moltype = RNA length = 21
FEATURE Location/Qualifiers
misc_feature 1..21
note = Synthetic polynucleotide
source 1..21
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 48 ttgtttttt ttgggttggt t 21

SEQ ID NO: 49 moltype = DNA length = 44
FEATURE Location/Qualifiers
misc_feature 1..44
note = Synthetic polynucleotide
source 1..44
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 49 atggcgccg catgagagaa gcccagacca attacctacc caaa 44

SEQ ID NO: 50 moltype = DNA length = 7482
FEATURE Location/Qualifiers
misc_feature 1..7482
note = Synthetic polynucleotide
source 1..7482
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 50 atggagaaag ttacatgtga catcgaggaa gacagccat tcctcagagc tttgcagcgg 60
agttccccgc agtttgaggat agaaggccaa caggtcactg ataatgcaca tgctaatgcc 120
agagcggtt ccgcattgtgc ttcaaaactg atcgaaacgg aggtggaccc atccgacacg 180
atcttgcaca ttggaaagtgc gcccccccgd agaatgtatt ctaagcacaa gtatcatgt 240
atctgtccga tgagatgtgc ggaagatccg gacagattgt ataagtatgc aactaagctg 300
aagaaaaact gtaaaaaat aactgtataag gaattggacaa agaaaaatgaa ggagctggcc 360
gcccgtcatga cgacccatgtga cctggaaact gagactatgt gcctccacga cgacgatcg 420
tgtcgctacg aaggccaaatg cgctgtttac caggatgtat acgcgcgtcg cggccccacc 480
agcctgtacc accaggccaa caaggccgtg agggctggct actggatcgg ctgcacacc 540
acacccttca ttgttcaaaagg cctggccggd gcttacccca gctacacgac caactggcc 600
gacgagaccc tgctgaccgc cggaaacatc ggctgtgtca gcagcgcacgt gatggagagg 660
agccggagag gcatgagcat cctggaaag aaatactgtaa agcccaacaa caacgtgtg 720
ttcagcgtgg gcaagccatctt accacacgag aagagggacc tgctcaggag ctggcacctg 780
cccaagcggtt tccacatgtga gggcaaggaa aactacacccgt gcaaggccgtcg gaccatcg 840
agtcgcacg gtcacgtgtt gaagaggatc gcttacgcgc cggccgtgtca cggcaagccc 900
agcggctacg cggcttacat gcaacggggg ggttctgtgt gtcacgtgtt gaccgcacacc 960
ctgaacccggc aggggtttagt cttccctgtt tgacccatcg tgccgcaccc cctgtgcgac 1020
caagatggcc gcatgttgcgc caccggacgtt agggccgcacg acggcccaaaa gtcgtctgt 1080
ggccctgttca aggatgtgtt ggttcaacggc aactacacccgt gcaaggccgtcg 1140
aactacatgc tgccctgtgtt gggccaggctt ttcggccagggt gggccaaaggaa gtacaaggag 1200
gaccaggaaag acggagaggcc cctggccgtg agggacaggcc agctgggtat gggctgtgc 1260
tggcccttca ggcggccacaa gatcaccaggc atctacaaga ggccgcacac ccagaccatc 1320
atcaaggatgtca acaggcactt ccacatgttgc tgctgcacca ggatggccgca caacaccctg 1380
gatcatggcc tgaggacccg gatcaggaaatgtgttggagg aacacaaaggaa gcccaccca 1440
ctgtatcccg ccggaggacgtt gcaaggaggcc aagtgcgtt ccgacggggc caaggagggt 1500
agggaggccg aggaaatgtt ggcggccctgtt ccacccctgtt ctgcgcacgtt ggaggaaaccc 1560
acccttggaaatg cggacgttgcg cttgtatgtt caggaggccg ggcggccaaatg cgtggagaca 1620
cccaaggggcc ttgtatccatgtt gacggatgttcc gacggccggg acaaatgttcc cagtcacg 1680
gtgtgtggcc cacaggccgtt gtcgtgttcc gagaatgttca gtcgtatcca cccactggcc 1740
gacccggatgg tggatgttcc ccacaggccg gggaaaggccca ggttacgcgtt ggagccctac 1800
cacggccatgg tggatgttcc ccacaggccgtt ggcacccctgtt tgccaggatcc ccaggccctg 1860
agccggatggcc ccacatgttgc gttacacggc agggatgttcc tgaacaggatcc cctgcaccc 1920
atccgcaccc acggccggatcc cttgttcc gacggatgttca actacaagac gtcgttcc 1980
agccggatggcc acggccgttcc cttgttcc gacggatgttca atccgcaccc 2040
ctgtgtggcc gtcgtgttcc gacggccgtt ccacccctgtt ccacaggccgtt gggccctac 2100
tacggatggcc tggaggccatgg accccctgtt cccttccatgg tggatgttcc cggccgtt 2160
ggccgtgttcc gcaaggccgtt gacggccgtt ccacccctgtt tgccaggatcc ccaggccctg 2220
gtgtgtggcc gcaaggaaatg gaaatgttcc gatcatca gggacgttcc gaaatgttcc 2280
ggccctgttcc gtaacggccgtt caccgttcc gacggatgttcc tgaacaggatcc caacccctgtt 2340
gtgtgtggcc tggatgttcc gacggccgtt ccacccctgtt ccacaggccgtt gggccctac 2400
atccgcaccc tcaaggccgtt gaaatgttcc gatcatca gggacgttcc gaaatgttcc 2460
ttcaacatgttca tggatgttcc ggttccatgg aaccacggatcc tggatgttcc ggttcc 2520
aagagcatca gcaaggccgtt caccgttcc gacggatgttcc tggatgttcc cctgttcc 2580
gacaagaaaa tgaggaccac caaccccaag gagacccaaatg tggatgttcc caccacggc 2640
agccaccaatg ccaaggccgtt cttgttcc gacggatgttcc tggatgttcc gggccctac 2700
ctgcagatgttca actacaaggatcc caacggatgttcc atgaccggccgtt cttgttcc gggccctac 2760
aggaaggccgtt tggatgttcc gggccctac gggccctac gggccctac 2820

-continued

agcgagcacg	tgaacgtgct	gctgaccagg	accgaggaca	ggatcgtgt	gaagacctg	2880
gcggccgacc	cctggatcaa	gaccctgacc	gccaagtacc	ccggcaactt	caccggcacc	2940
atcgaagagt	ggcaggccga	gcacgacgcc	atcatgaggc	acatcttgta	gaggcccgc	3000
cccaccgacg	tgttccagaa	caaggccaa	gtgtgtggg	ccaaggccct	ggtgcccgt	3060
ctgaagaccc	ccggcatcga	catgaccaca	gagcagttgg	acaccgtgg	ctacttcgg	3120
accgacaagg	cccacagcgc	cgagatcgt	ctgaaccagg	tgtgtgtgg	gttcttcgg	3180
ctggacttgg	acagcgccct	gttcagcgcc	cccaccgtgc	cactgtacat	cagaacaaac	3240
cactgggaca	acagccccag	cccaaaccat	tacggcctga	acaaggaggt	ggtcaggcag	3300
ctgagcaggc	ggttaccaca	gctgtccagg	gcctgtggcc	ccggcagggt	gtacgacatg	3360
aacacccggca	ccctggagaa	ctacgaccac	aggatcaacc	tggtgcggct	gaacaggcgg	3420
ctgccccccacg	ccctgtgtct	gcacccaa	gaggcaccac	agagcactt	cagtccttc	3480
gtgagcaagc	tgaaaaggcag	gaccgtgt	gtcgtggcg	agaagctgag	cgtgcccgg	3540
aagatgttgg	actgggtgg	cgacaggcc	gaggccaccc	tccggccac	gtctggactc	3600
ggcatcccg	cgcaagtcgc	caagtcac	atcatcttcg	tgaacgtcag	gaccctatac	3660
aagtaccacc	attaccagca	gtgcgaggac	cacggccatca	agctgacat	gctgaccaag	3720
aggcctgtcc	tgcacctgaa	ccccggaggc	acctgtgt	gcatcggt	cggtacggcc	3780
gacagggcoca	ggagggatc	cattggccgg	atcgccaggc	tgttcaagg	cagcagggt	3840
tgcaaaaccca	agagcactt	ggaggaaacc	tgggtgtgt	tctgtttcat	cggtacagac	3900
cggaaaggcoca	ggacccacaa	cccttacaa	ctgagcagca	ccctgacaaa	catctacacc	3960
ggcagcaggc	tgcaacgaggc	cggtgtgcgc	cccagctacc	acgtgtgt	ggggegatatc	4020
gcacccgcoca	ccggggcgt	gatcatca	gctgccaaca	gcaaggccca	gcceggaggc	4080
ggagtgtgtc	gcgcctgt	caagaagt	cccgagat	tegatcgca	gccccatcg	4140
gtggccaaagg	ccaggctgt	gaaggggcgc	gtaaggacca	tatccacgc	cgtggccccc	4200
aacttcaaca	aggtgagcga	ggtggaaagg	gacaaggcgc	tggcgaagc	ctacgagacg	4260
atcgccaaag	tctgtaaacg	caataactac	aagagctgt	ccatccact	gctcagcacc	4320
ggcatcttca	cgccgcaacaa	ggacaggct	accacagac	tgaaccac	gtcaccggcc	4380
ctggacacca	ccggatcgca	cggtggccat	tactgtgg	acaagaatg	ggagatgacc	4440
ctgaaggagg	ccgtggccag	gccccggag	gttgaagaga	tctgtcat	cgacgactc	4500
agcgtgaccc	acggccacgc	cgagctgtgt	agggtgcacc	ccaaggactc	cctggccggc	4560
aggaagggt	acagcaccac	cgacggcga	acccatcgat	acctggagg	ccacaagg	4620
caccaggccg	ctttaggacat	cgccgagat	aacgctat	ggccctgtgc	caccgaggcc	4680
aacggcagg	tgtcatgt	catectggc	gagagcat	ccagcat	gagaacgt	4740
cccggtgggg	aaagcgaggc	cagcacacca	ccccggaccc	tggccctgc	gtgcatccac	4800
getatcgatc	ccggagggtt	gcacgggtt	gacggccac	ggggccagca	gatcaccgt	4860
tgcagctct	tccactgtcc	caagtag	atcaccggc	tgcagaagat	ccagtgcagc	4920
cacccatcc	tggtttagccc	aaagggtgccc	gcttacatcc	accccaaggaa	gtacccgt	4980
gagacccccc	ccgtgtggac	gacaccggag	ccaagcgc	agaaccagag	caccgaggc	5040
acacccggac	acccacccct	gatccacgg	gacggacaca	ggacccggac	cccaaggccc	5100
atcattatcg	aaagagggaa	agaggacac	atcagccctgc	tgagcggc	ccccacccac	5160
caagggtgtc	aggtggaggc	cgacatcc	ggccacacca	gggtgtcc	ctccagctgg	5220
agcatccac	acggccacgc	cttcgacgt	gacagcctga	gcatcttgc	caccctggag	5280
ggcgccacgc	tgacccctccg	cgccaccagg	gcccggac	acagctact	cgccaaagac	5340
atgggatgtcc	tggccggcc	cggtggccat	cccgaggac	tgttgcgg	ccccacccac	5400
ccagctccca	ggaccaggac	cccaaggctg	gctccca	gggcctgc	caggaccac	5460
ctgggtggac	ccccacccgg	cgtgaacagg	gtgtatcc	ggggagaaact	ggggccctg	5520
acacccggac	ggaccccccac	cagggtccgt	agcaggacta	gtctgggt	caacccaccc	5580
ggcggtgac	gggtgtat	cggggagg	tccggggct	tctgtggcc	gcaacagaga	5640
cggttcacg	ccggggccct	catcttc	aggcaccac	ggccggggca	cctgcagca	5700
aagagcgtga	ggcagaccgt	gtcgacgt	gtgtgtgt	agaggacca	gttggaaatc	5760
agctacgccc	ccagggttgc	ccaggaggaa	gaggaaactc	tcaggaaagg	actcgactg	5820
aaccccccac	caggcaacac	gaggacgt	caagacggag	agggtggag	catgaaggcc	5880
atcaccgc	ggccgtatcc	gcagggtctg	ggacactacc	tgaaggcga	ggcaagggt	5940
gagtgtaca	ggaccctgca	ccccgtgca	ctgtacat	ccagcgt	caggccctc	6000
tcacccccc	agggtggccgt	ggggccctgc	aacgctatgc	tgaaggaa	cttccccac	6060
gtggccact	actgtcat	ccccggat	gacggccat	tggacat	ggacggcc	6120
agctgtgtcc	tggacaccgc	cattgtc	ccccggcaac	tgaggac	ccccaaagaaa	6180
cacagctacc	tggagccac	catcaggag	ggcggtccca	gcccac	ggccatcc	6240
cagaacgtgc	tggccgtgc	caccaagg	aactgtca	tgaccaggat	ggggggct	6300
cccggtgtcc	acagegtgc	cttcaacgt	gagtgttca	aaaaatacgc	ctgcacac	6360
gagtactgg	aggacccat	ggagaacccc	ataggctg	ccgaagagaa	cgtgtgaa	6420
tacatccacca	agctgtacgg	ccccaaaggc	gttgcctgt	tgcataagac	ccacaactg	6480
aacatgtgc	aggacatccc	aatggacagg	ttcgtgtat	acctgtacgg	ggacgtgt	6540
gtgacacccg	gcaccaagca	caccggagg	aggcccaagg	tgcgtgtat	ccaggccgt	6600
gacccactgg	ccacccgtca	cctgtggcc	atccacagg	agtcgtgt	ggccgtgaa	6660
ggcggtgtcc	tgcccaacat	ccacccctg	ttcgcacat	gcccggagga	cttcgcac	6720
atcatcgcc	agcacttca	gcccggccac	tgcgtgtgt	agaccgacat	cgccac	6780
gacaaggacg	aggatgtacg	tatggccct	accgtctgt	tgcacat	ggacccgttc	6840
gtggacacccg	agtcgtctac	cctgtacat	gtgcgttgc	gctggat	ctccatccac	6900
ctggccacca	agaccaatgt	caagtgttgc	gtatgtat	aaagcgtat	gttctgtacc	6960
ctgttgcgt	acaccgtgt	caacattgt	atgcgtac	gggtgtgt	ggagagggt	7020
accggcagcc	cctgtgtcc	gttcgttgc	gacgacaaca	tcgtgtacgg	cgtgtac	7080
gacaaggacg	tggccacac	tgggtggcc	tgcacat	ttcgtgtacgg	ttcgtgtac	7140
gccgtgggt	ggcggacgg	ccctactt	tgcggccgt	tcatctgt	cgacagcgt	7200
acccggcacc	cctgtgtcc	ggccgacccc	ctgaagacgc	tgttcaagg	ggcaaggcca	7260
ctggccgt	acgtatgtac	caacattgt	atgcgtac	gggtgtgt	ggagagggt	7320
aggtgttacca	gggtggccat	cctgtacat	gtgtgtac	ccgtgtggag	caggtac	7380
accgtggcc	ccagcatcat	gtgtgtac	atgaccacac	tggccac	cgtcaagac	7440
ttctcttacc	tgaggggggc	ccctataact	ctctacggct	aa		7482

-continued

SEQ ID NO: 51 moltype = AA length = 2493
 FEATURE Location/Qualifiers
 REGION 1..2493
 note = Synthetic polypeptide
 source 1..2493
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 51
 MEKVHVDIEE DSPFLRALQK SFPQFEVEAK QVTDNDHANA RAFSHLASKL IETEVDPDSI 60
 ILIDIGSAPAR RMYSKHKYHC ICPMRCAEDP DRLYKYATKL KKNCKEITDK ELDKKMKELA 120
 AVMSDPDLET ETMCLHDDES CRYEQVAVY QDVYAVDGPT SLYHQANKGV RVAYWIGFDT 180
 TPFMFKNLAG AYPSYSTNWA DETVLTARNI GLCSSDVMER SRRGMSILRK KYLKPSNNVL 240
 FSVGSTIYHE KRDLRLRSWHL PSVSHFLRGKQ NYTCRCEITV SCDFGYUVVKRI AISPGLYGKP 300
 SGYAATMHRE GFLCCKVDTD LINGERVSFPV CTVPATLKD QMTGILATDV SADDQKLLV 360
 GLNQRIIVNG RTQRNTNTM NYLLPVVAQQA FARWAKEKEYE DQEDEPLGI RDRQLVMGCC 420
 WAFRRHKITS IYKRDPDTQTI IKVNNSDFHSF VLPRIGSNTL EIGLRTRIK MLEEHKEPSP 480
 LTAAEDVQEE KCAADEAKEA REAEALRLAAL PPLAADVEEP TLEADVDLML QEAGAGGSVET 540
 PRGLIKVTSY DGEDKIGSYA VLSPQAVLKS EKLSCIHPLA EQVIVITHSG RKGRYAVEPY 600
 HGKVVPPEGH AIPVQDFQAL SESATIVYNE REFPVNRYLHH IATHGGALNT DEEYYKTVKP 660
 SEHDGEYLYD IDRKCQCVKKE LVTGLGLTGE LVDPPFHEFA YESLRTRPAA PYQVPTIGVY 720
 GVPGSGKSGI IKSAVTKKMD VVSALKENKA EIIRDVKKMK GLDVNARTVD SVLLNGCKHP 780
 VETLYIDEAF ACHAGTLRAL IAIIRPKKAV LCGDPKQCFC FNMMCLKVHF NHEICTQVFH 840
 KSISRRCTKS VTSVUSTLFY DKKMRTTNPK ETKIVIDTTG STPKQDDLI LTCFRGWVKQ 900
 LQIDYKGNEI MTAASQGLT RKGVYAVRYK VNENPLYAPT SEHVNVLTR TEDRIVWKT 960
 AGDPWIKLT AKYPGNFTAT IEEWQAEHDA IMRHILERPD PTDFVQNKKAN VCWAKALVPV 1020
 LKTAGIDMTT EQWNNTVDYFE TDKAHSAEIV LNQLCVRFFG LDLDSDGLFSA PTVPLSIRNN 1080
 HWNDNSPSPNM YGLNKEBVRQ LSRRYQPLPR AVATGRVYDM NTGTLRNYPD RINLVPVNRR 1140
 LPHALVLHNN EHPQSDFSSP VSKLKGRTVL VVGEKLSVPG KMVDWLSDRP EATFFRARLDL 1200
 GIPGDVVKYD IIFVNVRTPY KYHHYQQCED HAIKLMSMLT KACLHLPNGG TCVSIGYGYA 1260
 DRASESIIGA IARLFPKFSRV CKPKSSLEET EVLFVFIGYD RKAARTHNPYK LSSTLTNIYT 1320
 GSRSLHEAGCA PSYHVVRGDI ATATEGVIIIN AANSKGQPGG GVCAGLYKKF PESFDLQPIE 1380
 VGKARLVKGA AKHIIHAVGP NFNKVSSEVG DKQLAEAYES IAKIVNDNNY KSVAIPLLST 1440
 GIPSGGNKDRL TQSLNHLLTA LDTTDADVAI YCRDKKWEWT LKEAVARREA VEEICISDDS 1500
 SVTEPDAELV RVHPKSSLAG RKGYSTSDGK TFSVLEGTKF HQAAKDIAEI NAMWPVATEA 1560
 NEQVCMYILG ESMSSIRSKC PVEESEASTP PSTLPCLCIH AMTPERVQL KASRPEQITV 1620
 CSSFPLPKYR ITGVQKIQCS QPILFSPKVP AYIHPRKYLV ETTPVDETPE PSAENQSTEG 1680
 TPBQPLPLITE DETRTRTPEP IIIIEEEEDS ISLLSDGPTH QVLQVEADIH GPPSVSSSSW 1740
 SIPHASDFDV DSLSILDITLE GASVTSGATS AETNSYFAKS MEFLARPVPA PRTVFRNPPH 1800
 PAPRTRTPSL APSRACRSRTS LVSTPPGVNR VITREELEAL TPSRTPRSV SRSLVSNPP 1860
 GVNRVITREE PEAFVAQQQR RFDAGAYIFS SDTGQGHLOQ KSVRQTVLSE VVLERTELEI 1920
 SYAPRLDQEKK EELLRKKLQL NPTPANRSRY QSRKVENNMKA ITARRILQGL GHYLLKAEGKV 1980
 ECYRTLHPV LYSSSVNRAF SSPKVAEC NAMLKENFPVT VASYCIIPEY DAYLDMVGDGA 2040
 SCCLDTASFC PAKLRSFPKK HSYLEPTIRS AVPSAIQNTL QNVLAATKRN NCNVTMQREL 2100
 PVLDSSAFNV ECFKKYACNN EYWETFKENP IRLTEENVVN YITKLKGPKA AALPAKTHNL 2160
 NMLQDIPMDR FVMDLKRDKV VTPGTKHTEE RPKVQVQIAA DPLATAYLCG IHRELVRRLN 2220
 AVLLPNIHTL FDMSAEDFDA IIABHFQPGD CVLETDIASF DKSEDDAMAL TALMILEDLG 2280
 VDAELLTLIE AAFGEISSH LPTKTKFKFG AMMKSGMFLT LFVNTVINIV IASRVLRLERL 2340
 TGSPCAFIG DDNIVKGVKS DKLMAUDRCAT WLNEMEVIID AVVGEKAPYF CGGFILCDSV 2400
 TGTACRVADP LKRLFKLGKP LAADDEHDDD RRRALHEEST RWNRVGLSE LCKAVESRYE 2460
 TVGTSIIIVMA MTTLASSVKS FSYLRGAPIT LYG 2493

SEQ ID NO: 52 moltype = DNA length = 44
 FEATURE Location/Qualifiers
 misc_feature 1..44
 note = Synthetic polynucleotide
 source 1..44
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 52
 cctgaatggaa ctacgacata gtcttagtccg ccaaggccgc cacc 44

SEQ ID NO: 53 moltype = DNA length = 468
 FEATURE Location/Qualifiers
 misc_feature 1..468
 note = Synthetic polynucleotide
 source 1..468
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 53
 actcgagtat gttacgtgca aagggtatttc tcacccccc aaagaccata ttgtgacaca 60
 ccctcagtat cacggccaaa catttacago cgcgggtcata aaaaccgcgt ggacgtgggtt 120
 aacatccctg ctggggatcat cagccgtaat tatttaatgg ctggctactat 180
 tttttttttt ccgaatcgaa ccaaccagaa acataatttgaa atacagcagc aattggcaag 240
 ctgtttttttt ccgaatcgaa tttttttttt aatatttcaa aaaaaaaaaaaaaaaa 300
 aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaa 360
 aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaa 420
 aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaa 468

SEQ ID NO: 54 moltype = DNA length = 7485

-continued

FEATURE	Location/Qualifiers
misc_feature	1..7485 note = Synthetic polynucleotide
source	1..7485 mol_type = other DNA organism = synthetic construct
SEQUENCE: 54	
atggccgaga aggtgcacgt ggacatcgag gaggacagcc ctttccttag ggcctctcgag 60 aggagcttc cacagtccg agtggaggcc aagcaggtga ccgacaacga ccacgcacac 120 gccagggcct tcagccacct ggccagcaag ctgatcgaga ccgagggtgaa ccccaagcgac 180 accatctgg acatcgccag cgccccccg aggaaatgt acacaagca caatccac 240 tgcgtccg ccatcgagggt cgccgaggac cccgcaggcc tgcatacgta ccgcggccaa 300 ctgaagaaga actcgaaagg gatcacccgg aaggactgtt aacaaaaaat gaaggagctg 360 gcccgcgtga tgagcgaccc cgaccctggg acccgagacaa tgcgtctgca cgacgcgag 420 agctgcacgt acgaggggcca ggtggccgtc taccaggacg tgcgtccgt cgcggcccc 480 accaggctgt accaccaggc caaacaggc gttgggggtt cctactgtt cggcttcgg 540 accacaccct tcatgttcaa gaacctggcc ggccgcattcc ccaactacgac caacactgg 600 gcccgcgtga cctgtgtcac ccgcaggaaac atcgccgtt gcacgcgega cgtgtatgg 660 aggagccgga gaggcatgag catcttgagg aagaataacc tgaaggcccg caacaacgt 720 ctgttcacggc tggggcagcac catcttaccc gagaaggggg acctgcgtcag gactggcc 780 ctggcccgcc tggggccactt gaggggccaa cagaactaca cttcgagggtt cgagaccatc 840 gttggcgtcg acggctacgt ggtggaaagg atgcocata gccccccgtt gtacggcaag 900 cccagcggct acggcgctac aatgcacagg gagggtttcc tgcgtgtca ggtgaccgac 960 acccctgaacc gcgaggaggt ggcgttcccc gttggccactt acgtgcggccg caccctgtgc 1020 gaccatgtca cccggcatctt ggccacccgg tggagcggccg acgacgcccc gaactgttc 1080 gttggcgtga accagaggat cttgggttcaac ggcaggaccc agggaaacac caacaaatgt 1140 aagaactacc tgcgtccgtt ggtggcccg gctttccca ggtggggccaa ggagtacaag 1200 gaggaccagg aagacgaggag gcccctggcc ctggggacca ggcacgttgc gatgggttgc 1260 tgcgttgcggc tcaggcgcc caagatcacc agcatctaca agaggcccg caccaggacc 1320 atcatcaagg tgaacacgca cttccacccg ttcgtgtcgc ccaggatcgg cacaacacc 1380 ctggagatcg gcctggaggac cccggatcagg aagatgttgc aggaacacaa ggaggccagg 1440 ccactgtaca cccggcgaggc cttggcaggag gccaatgtcg ctggccggca ggccaaaggag 1500 gttggggagg cccggaggaa gggggccgc ctgcacccccc tggctccggc cttggggagg 1560 ccccccttgg aaggccgactt ggacccgtat ctggcaggagg ccggccgggg aaactgttgc 1620 acaccccgagg gctgtatcaa ggtgaccagg tacgacggcg aggacaagat cggcgttac 1680 gcccgtgtca gcccacaggc cttgggttcaag tccggaaagg tgcgtgtcat ccacccactt 1740 gcccggaggc tgatgtgtt caccacccgg ggcaggaaagg ggacgttgc cttggggcc 1800 taccacggc aggtgttgcgt gggccaggcc caecatcc cctgtccggg ctccggccgg 1860 ctgagcgaga gggccaccat cttgttacac gaggaggagg tgcgttacag gtactgtcac 1920 catatcgcca cccacggccgg agccctgttac ggcacggagg aataactacaa gacccgttgc 1980 ccccggcgcc acggacggcc gtatccgttc gacatcgaca ggaaggactg cttggggcc 2040 gacgtgttgc cccggcttgg actggccggc gacgttgcggg acgtgttgcggg acccacccctt 2100 gcctacgaga gcctggaggac cagacccggc gcttccatcc aggtggccac catccggcg 2160 tacggcgctc cccggcgaggc aaagaggccg atcatcaaga ggcggccgtac caagaaagac 2220 ctgggtgtca gggccaaagg agagaactgc ggcggatca tccggggactt gaaaggatgt 2280 aaaggccctgg acgtgttgcggc ggcacccgg tgcgttgc gacgttgc gacgttgcac 2340 ccctgtggaga ccctgtatcat cggcggggcc ttccgttcc acggccggcc ctggggggcc 2400 ctgatcgcca tcatcgaggc caagaaaggc gtgtgttgcg ggcacccca gcaactgtccg 2460 ttcttcacaa tgatgttgcgt gaaagggttgc ttcaccaaccg agatctgtcc cccagggttgc 2520 cacaagagca tcaggcgcc gttggccaaag acgtgttgc gacgttgc gacgttgcac 2580 tacagcaagaaa aatgttgcggc caccacccgg aaggagacca aatactgttgc acacccacca 2640 ggcagcaccac agcccaaggc ggacgttgcgt atctgttgc gtttccgggg ctgggttgaag 2700 caactgttgc tggactacaa gggcaacaggc atcatgttgc cccggccggc ccaggccggc 2760 accaggaaagg ggtgttgcgt cttggggaggc aagggttgcgtt gcaacttcc gttacgttcc 2820 accaggccggc acgtgttgcgtt cttgttgcgtt gggccgggg agatctgtcc gttggggagg 2880 ctggccggccg accccctgttac caagacccggt accgcgttac agccccggca cttccacccgg 2940 accatcgcca agtggccggc cggacgttac ggcacccatcg gccatcatcg ggcacccatcg gggaggcc 3000 gaccggccggc acgtgttccca gaaaggccggc aacgtgttgc gggccgggg cttggggcc 3060 gttgcgttgc cccggccggcgtt cccatgttgcg acaggacgtt gggccgggg cttggggcc 3120 gagaccggaca agggccggcc cggccggccgtt gttggggccgtt gggccgggg gggccggcc 3180 ggccctggacc tggacccggc cttgttgcgtt gcccggccggc tggccactgg catccgggg 3240 aaccactgggg acaacacggcc gggccggccgtt atgtgttgc gacgttgc gacgttgcac 3300 caactgttgc gggccggccgtt acatgttgcg gggccggccggc tggccggccggc tggggccgg 3360 atgaacacccg gcacccctgttgc gaaacttgcgtt cccaggatca acctgttgc gtcgttgc 3420 cggttgcggcc accggccctgttgc gttgttgcgtt gggccggccgtt gggccggccggc 3480 ttcgtgttgc gacgtgttgcggc cttgttgcgtt gggccggccgtt gggccggccggc 3540 ggcgcacccgg cccaggccggc catcttgcgtt gggccggccgtt gggccggccggc 3600 ctggccatcc cccggccgttgc gggccggccgtt gggccggccgtt gggccggccggc 3660 tacaagtacc accattacca gcaactgttgc gggccggccgtt gggccggccgtt gggccggccggc 3720 aagaaggccctt ggcgttgcgtt gggccggccgtt gggccggccgtt gggccggccggc 3780 ggccgttgcgtt gggccggccgtt gggccggccgtt gggccggccggc 3840 ggcgttgcgtt gggccggccgtt gggccggccgtt gggccggccggc 3900 gaccggccggc accggccggccgtt gggccggccgtt gggccggccggc 3960 accggccggc gggccggccgtt gggccggccgtt gggccggccggc 4020 atcgccaccgg ccacccggccgtt gggccggccgtt gggccggccggc 4080 ggccgttgcgtt gggccggccgtt gggccggccgtt gggccggccggc 4140 gagggttgcgtt gggccggccgtt gggccggccgtt gggccggccggc 4200 cccaacttca acaagggttgc gggccggccgtt gggccggccggc 4260 agcatcgcca agatgttgc gggccggccgtt gggccggccggc 4320	

-continued

accggcata	tcagcggcaa	caaggacagg	ctgaccagg	gcctgaacca	cctgtcacc	4380
gccctggaca	ccaccatgc	cgacgtggc	atctactgc	gggacaagaa	gtggagatg	4440
accctgaagg	aggccgtggc	caggcggag	gccgtggaa	agatctgc	cagcgacgc	4500
tccagcgtg	ccgagccga	cgccgagctg	gtgagggtg	accccaagag	ctccctggcc	4560
ggcaggaa	gotacagcac	cagcgacgc	aagaccttca	gtactctgg	gggcaccaag	4620
ttccaccagg	cgcctaagg	catcgccag	atcaacgc	tgtggccctg	ggccacccgg	4680
gccaacgac	aggtgtcat	gtatccctg	ggegagac	tgtccacat	caggagcaag	4740
tgccccgtgg	aggaaa	ggcagcaca	ccaccagg	ccctgcctg	cctgtcata	4800
cacgtatgc	cacccggag	ggtgacgg	ctgaaggcc	gcaggeccga	gcagatcacc	4860
gtgtgcag	cattccact	gccaagt	aggatcacc	cgctgcagaa	gatccagtgc	4920
agccagccca	tcctgtcag	cccaagg	cccgcctaca	tccacccag	gaagtactg	4980
gtggagaccc	caccgtgg	cgagacacc	gagccaag	cgagagaac	gagcaccgg	5040
ggcacacccg	agcagccacc	cctgtac	gaggacgaga	caaggacccg	gaccccaag	5100
cccatcatta	tcgagggaa	ggaaggagg	agcatcagg	tgtgtacg	cgccccacc	5160
caccagggtc	tcgagggtg	ggccgacat	cacggcc	ccagcg	cagtcacc	5220
tggagcatcc	cacacgccc	cgacttgc	gtggac	tgagcatct	ggacccctg	5280
gaggggcoca	gggtgac	cgggcc	agcggc	ccaacagct	cttgcacag	5340
agcatgggt	tcctggcc	gcccgtca	gcccgg	cgctgttag	gaacccaccc	5400
caccaggctc	ccaggacc	gaccca	ctggctcc	gcagggc	cagcagg	5460
agectgtg	gcacccacc	cgcg	gtgtgatc	ccagggagg	actggagg	5520
ctgacaccca	cgaggac	cagggt	gtgagc	ctagtcgt	gtccaa	5580
cccgccgt	acagggtg	caccagg	taccag	ggaatcg	ccttcgtgc	5640
agacgg	tcgtggcc	ctatcat	agcagc	acggcagg	acactgc	5700
caaaagagc	tgaggcag	cgtgtg	gagggtgt	tggagagg	cgactgtg	5760
atca	ccccagg	ggcagg	aaggagg	tgctcagg	gaaactgc	5820
ctgaac	cccccagg	cagg	gtgagg	ctactgt	gtccaa	5880
gcccattac	ccaggcc	cctgcagg	ctgg	acactgt	cgagg	5940
gtggagtgt	acaggacc	gcaccc	ccactgt	gtccac	gaacagg	6000
tttccac	ccaa	gtggagg	tgcac	tgctgaa	gaacttccc	6060
ac	gtactgt	caccc	tacgg	actctgg	gttggac	6120
gccc	gttgg	cgcc	tgcc	gtcgt	cttccc	6180
aa	acactgt	ac	gttgc	ccat	ccaga	6240
ctgc	gaa	ccat	ggc	actgt	ccgc	6300
ctgc	ccat	ccat	ggc	gttgc	ccat	6360
aac	ggagactt	caagg	ccat	ccat	tgacc	6420
aact	atc	caag	ggc	ggc	gaa	6480
ctg	acat	ccat	ggc	gttgc	ggc	6540
aa	gttgc	ccgg	ggc	actgt	ggc	6600
gctg	accc	ccat	ggc	gttgc	ggc	6660
aac	ccgt	ccat	ggc	gttgc	ggc	6720
gccc	atc	ccag	ggc	actgt	ggc	6780
ttcg	acaa	ccat	ggc	gttgc	ggc	6840
ggcg	tgat	ccat	ggc	actgt	ggc	6900
ca	ccat	ccat	ggc	gttgc	ggc	6960
acc	ctgt	ccat	ggc	gtat	ggc	7020
ctg	accgg	ccat	ggc	gtat	ggc	7080
ag	gacaa	ccat	ggc	gtat	ggc	7140
gac	tgat	ccat	ggc	gtat	ggc	7200
gt	ccat	ccat	ggc	gtat	ggc	7260
ca	actgt	ccat	ggc	gtat	ggc	7320
acc	ccat	ccat	ggc	gtat	ggc	7380
gac	ccat	ccat	ggc	gtat	ggc	7440
gag	ccat	ccat	ggc	gtat	ggc	7485

SEQ ID NO: 55 moltype = AA length = 2494
 FEATURE Location/Qualifiers
 REGION 1..2494
 note = Synthetic polypeptide
 source 1..2494
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 55
 MPEKVKVDIE EDSPFLRALQ RSFPQFEVEA KQVTDNDHAN ARAFSLASK LIETEVDPDSD 60
 TILDIGSAPA RRMYSKHKYH CICPMRCAD EDRLYKYATK LKKNCKEITD KELDKMKEL 120
 AAVMSDPDLE TETMCLHDDE SCRYPEQVAV YQDVYAVDGP TSLYHQANKG VRVAYWIGFD 180
 TTPFMFKNLA GAYPSYSTNW ADETIVLTARN IGLCSSDVMF RSRRGMSILR KKYLKPSSNNV 240
 LFSVGSTIYH EKRDLRLRSW LPSVFHLRGK QNYTCRCETI VSCDGYVVKR IAISPGLYGK 300
 PSGYAAATMHR EGFLCCKVKE TLNGERVSFP VCTYVPATLC DQMGTGILATD VSADDAQKLL 360
 VGLNQRIVVN GRTQRNTNTM KNYLLPVPVQ AFARWAKBYK EDQEDERPLG LRDRQLVMGC 420
 CWAFRRHKIT SIYKRPDTQT IIKVNSDFHS FVLPRIGSNT LEIGLRLTRIR KMLEEHHKEPS 480
 PLITAEDVQE AKCAADEAKE VREAEELRA LPPLAADVEE PTLEADVDM LQEAGAGSVE 540
 TPRGLIKVTS YDGEDKIGSY AVLSPQAVLK SEKLSCIHPL AEQVIVITHS GRKGRYAVEP 600
 YHGKVVPPEG HAIPVQDFQA LSESATIVYN EREFVNRYLH HIATHGGALN TDEEYYKTVK 660
 PSEHDGELEY DIDRKQCVKE ELVTGLGLTG ELVDPPFHET AYESLRTRPA APYQVPTIGV 720
 YGPGPSKGSG IIKSIAVTKKD LVVSAKKENC AEIIRDVKM KGLDVNARTV DSVLLNGCKH 780
 PVETLYIDEA FACHAGTLRA LIAIIRPKA VLCGDPKQCG FFNMMCLKVH FNHEICTQVF 840
 HKSISRRCTK SVTSVUSTLF YDKKMRTTNP KETKIVIDIT GSTKPKQDDL ILTCFRGWVK 900
 QLQIDYKGNE IMTAAASQGL TRKGVYAVRY KVNNENPLYAP TSEHVNVLTT RTEDRIVWKT 960
 LAGDPWIKTL TAKYPGNFTA TIEEWQAEHD AIMRHILERP DPTDVFQNKAVCWA
 LA 1020

-continued

VLIKTAGIDMT	TEQWNTVDYF	ETDKAHSABEI	VLNQLCVRFF	GLDLDGGLFS	APTVPLSIRN	1080
NHWDNSPSPN	MYGLNKEVVR	QLSRRYPQLP	RAVATGRVYD	MNTGTLRNYD	PRINLVPVNR	1140
RLPHALVLHH	NEHPQSDFSS	FVSKLKGRTV	LVVGEKLSVP	GKMVDWLSDR	PEATFRARLD	1200
LGIPGDVPKY	DIIFVNVRTP	YKYHHYQQCE	DHAIKLMSLT	KKACLHLNPG	GTCVSIGYGY	1260
ADRASESIIG	AIARLFKFSR	VCKPKSSLEE	TEVLFVFIGY	DRKARTHNPY	KLSSTLTNIY	1320
TGSRLHEAGC	APSYHVVRGD	IATATEGVII	NAANSKGQPG	GGVCGALYKK	FPESFDLQPI	1380
EVGKARLVKG	AAKHIHAVG	PNFNKVSEVE	GDQLAEEAYE	SIAKIVNDNN	YKSAIPLLS	1440
TGIFSGNKDR	LTQSLNHLLT	ALDDTDADVA	IYCRDKKWEM	TLKEAVARRE	AVEEICISDD	1500
SSVTEPDAEL	VRVHPKSSL	GRKGYSTSDG	KTFSYLEGTK	FHQAAKDIAE	INAMMPVATE	1560
ANEQVCMYIL	GESMSSIIRSK	CPVEESEAST	PPSTLPCLCI	HAMTPERVQR	LKASRPEQIT	1620
VCCSFPLPKY	RITGVQKIQC	SQILFSPKV	PAYIHPKYL	VETPPVDETP	EPSAENQSTE	1680
GTPEQPPLIT	EDETRTRTPE	PIIIIEEEEED	SISILSDGPT	HQVLQVEADI	HGPPSVSSSS	1740
WSIPIHASDFD	VDSLSDLTL	EGASVTSQAT	SABTNSYPAK	SMEFLARPVP	APRTVFRNPP	1800
HPAPTRRTPS	LAPSACSR	SLVSTPPGVN	RVTREELEA	LTPSRTFPSRS	VSRTSLVSNP	1860
PGVNRVITRE	EFEFAVFAQQQ	RRFDAGAYIF	SSDTGQGHLOQ	QKSVRQTCLS	EVVLERTELE	1920
ISYAPRLDQE	KEELLRKKLQ	LNPTPANRSR	YQSRKVENMK	AITARRILQG	LGHYLKAEGK	1980
VECYRTLHPV	FLYSSSVNRN	FSSPKVAECA	CNALMKENFP	TVASYCIYPE	YDAYLDMVDG	2040
ASCCLDASF	CPAKLSPFPK	KHSYLEPTIR	SAVPSAIQNT	LQNVLAATAK	RNCNVTQMRE	2100
LPVLDSSAFN	VECFKYYACN	NEYWETFKEN	PIRLTEENV	NYITKLKGPK	AAALFAKTHN	2160
LNMLQDIPMD	RFVMDLKRDV	KVTPGTKHT	ERPKVQVIQA	ADPLATAYLC	GIHRELVRR	2220
NAVLLPNIHT	LFDMSEADEFI	AIIAEHFQPG	DCVLETDI	FDKSEDDAMA	LTALMILEDL	2280
GVDAELLLTI	EEAFGEISSI	HLPTKTFKG	GAMMKSGMF	TLFVNTVINI	VIASRVLRER	2340
LTGSPCAAFI	GDDNIVKGVK	SDKLMADRC	TWLNMVEKII	DAVVGKAPY	FCGGILCDS	2400
VTGTACRVAD	PLKRLFKLGK	PLAADDEHDD	DRRRALHEES	TRWNNRVGILS	ELCKAVESRY	2460
ETVGTTSIIVM	AMTTLASSVK	SFSYLRGAPI	TLYG			2494

SEQ ID NO: 56	moltype = RNA	length = 9739				
FEATURE	Location/Qualifiers					
misc_feature	1..9739					
	note = Synthetic polynucleotide					
source	1..9739					
	mol_type = other RNA					
	organism = synthetic construct					
SEQUENCE: 56						
atgggcggcg	catgagagaa	gccccagacca	attacctacc	aaaaatggag	aaagtccacg	60
ttgacatcga	ggaagacagc	ccattccca	gagcttgc	gcggagctc	ccgcagtttgc	120
aggtagaaggc	caagcaggtc	actgataatg	accatgcta	tgccagagcg	ttttcgcatc	180
tggcttcaaa	actgtatccgaa	acggaggttg	accatccgaa	cacgatctt	gacattggaa	240
gtgcggccgc	ccgcagaatgg	tattctaaa	acaagtatca	ttgtatctgt	ccgatggat	300
gtgcggaaaga	tccggacaga	ttgtatcaatg	atgcaactaa	gctgaagaaa	aactgttaagg	360
aaataactga	taaggaattt	gacaagaaaa	tgaaggact	ggccgcgcgc	atgagcgacc	420
ctgacccatgg	aactgtact	atgtgcctcc	acqacgcac	gtcgtgtcgc	tacaaaggcc	480
aagtgcgtgt	tattccggat	gtatacgcgc	tcgacggccc	caccagctt	taccaccagg	540
ccaaacaagg	cgtgagggtt	gcctactgg	tcggcttcga	caccacaccc	ttcatgttca	600
agaacacctgc	cgccgcctac	cccagctaca	gcaccaactg	ggccgcgcgc	accgtctga	660
cccccggaa	cattccgcgt	tgcgtccatgg	acgtgtatgg	gaggagccgg	agaggccatg	720
gcatecttgc	ggaaaaatac	ctgaaggccca	gcaacaacgt	gtgttca	gtgggcagca	780
ccatcttcca	cgagaaggagg	gaccgtctca	ggagctggca	cctgcccc	gtgttccacc	840
tgagggggca	gcagaactac	acctgcgtgt	gcfagaccat	cgtgatgtc	gacggctacg	900
tgtgttcaag	gtatccatgc	agccccggcc	tgtatccgaa	gcccacggcc	tacccgccta	960
caatgcacag	gggggggttc	ctgtgttca	aggtgaccga	caccctgaat	ggcgagagg	1020
ttagcttccc	cgtgtgcacc	tacgttccgg	ccaccctgttgc	cgaccatgt	accggccatcc	1080
tggccaccgc	cgtgacgc	gacgcacccc	agaagctgt	cgtggccctg	aaccaggagga	1140
tctgtgttca	ccggcggacc	cagggaca	caaacaat	gaagaactac	ctgtgtcccg	1200
tggtggccca	ggctttcgcc	agggtggcc	aggatgtacaa	ggggaccac	gaagacgaga	1260
ggccccctgg	cgtggggac	aggcgatgg	tgtatggcttgc	ctgtgtggcc	ttcaaggccgc	1320
acaagatcac	cacgcac	aagaggcccg	acacccacac	catcatcaag	gtgaacacgc	1380
acttccacag	cttcgtgtc	cccaggatcg	gcaccaacac	cctggatgc	ggccgtgagga	1440
cccgatccat	gaagatgtcg	gagggacaca	aggacccac	cccactgtac	accggccgg	1500
acgtgcacgg	ggccaaatgc	gctggccacg	ggcccaatgg	ggtgaggggag	gccggaggaa	1560
tgaggccgc	cctgcacacc	ctgggtccgc	acgtggaggac	acccaccctg	gaagccgacg	1620
tggacctgtat	gtcgacggag	ggccggccgc	gaacccacgg	ggccgtatca		1680
aggtgacccat	ctacgcggc	gaggacaaga	tcggcagatc	cgccgtgtc	agccccacgg	1740
ccctgtgttca	gtccggacaa	gtcgatgtca	tccacccact	ggccggacgc	gtgtatgtca	1800
tcacccacat	cgccggacaa	ggcaggatcg	ctcgatggcc	ctaccacggc	aagggtgtcg	1860
tggccggacgg	ccacgcacatc	cccgtgcagg	acttccaggc	cctgtatgtc	agcgccacca	1920
tctgtatcaac	cgagggggat	ttcgatgtca	ggtacccgttgc	ccatatccgc	acccacggcg	1980
gagccctgtaa	caccgcacatc	gaataactaca	agacccgtga	gcccacggac	cacgacggcg	2040
agtacccatgt	gtcgatgtc	aggaactcg	gtcgatgtca	gaaaggccctg	gacgtgaa	2100
gactgacccg	cgagctgtgt	gaccoaccct	tccacgtgttgc	ccctactacg	agccctgatgg	2160
ccagaccgc	cgctccctac	cagggtccca	ccatccggcg	gtacggcg	cccgccacgc	2220
gaaaagacgc	catcatcaag	agcgcgcgtga	ccaaagaaa	cctgtatgtc	agcgccaa	2280
aagagaactcg	ccggcgacatc	atcaggac	tgaagaat	gaaaggccctg	gacgtgaa	2340
ccggccacccgt	ggacacgcgt	ctgtgtca	gtcgatgtca	ccccgtggag	accctgtaca	2400
tgcacggaccc	cttcgtgtc	cacgcgcgttgc	ccctgaggcc	cctgtatgtc	atcatcgac	2460
ccaagaaaacg	cgatgtgtc	ggcgcacccca	agcagtgcgg	cttcttcaac	atgtatgtcc	2520
tgaagggtgc	cttcacccat	gagatctgtc	cccaggatgttgc	ccacaagac	atcagcaggc	2580
gggtgcacca	gagcgtgtacc	agcgtgtcgt	gcaccctgttgc	ctacgacaa	aaaatggag	2640
ccaccaaccc	caaggagacc	aaaatgtca	tcgcacaccac	aggcagcacc	aagcccaac	2700

-continued

aggacgacct	gatcctgacc	tgcttcaggg	gctgggtgaa	gcagctgcag	atcgactaca	2760
aggcaacga	gatcatgacc	gccgtgcca	gccaggccct	gaccaggaag	ggcgttacg	2820
ccgtgaggt	caaggtgaa	gagaacccac	tgtacgctcc	caccagcgag	cacgtgaacg	2880
tgctgctgac	caggaccgag	gacaggatcg	tgtggaaagc	cctggccggc	gacccttgg	2940
taaaggccct	gaccgccaag	taccccgca	acttcacccgc	caccatcgaa	gagtggcagg	3000
ccgagcacga	ccggccatcg	aggccatcc	tggagaggc	cgaccggcc	gacgtgttcc	3060
agaacaaggc	caacgtgtc	tggccaagg	ccctgggtgc	egtgtgaaag	accggccgca	3120
tcgacatgac	cacagagcag	tggAACACCG	tggactactt	cgagaccgac	aaggcccaca	3180
gecccgagat	cgtgtgaa	cagctgtcg	tgagggttct	eggcctggac	ctggacageg	3240
gectgttcag	cgccccccacc	gtgecaact	gcatcaggaa	caaccactgg	gacaacagcc	3300
ccagccccaa	catgtacggc	ctgaacagg	aggtgttgc	cgacgtgcgc	aggcggtacc	3360
cacagctgcc	caaggccgtg	gcccacggc	gggtgtacg	catgaacacc	ggcaccccta	3420
ggaactacga	ccccggatc	aacctggtc	ccgtgaacag	ggggctgccc	cacggccctgg	3480
tgtgtcaca	caacgacac	ccacagagc	atttcagtc	tttcgtgacg	aagetgaaag	3540
gcaggaccgt	gtgttcgtg	ggcggagaag	tgagctgtgc	cgggcaatgt	gtggactggc	3600
tgagcgacag	gccccggcc	accttcggg	ccaggctgg	cctcggcata	cccgccgacg	3660
tgcccaagta	cgacatcata	ttctgtgaa	tcaggacccc	atacaagtac	caccattacc	3720
agcgtgcga	ggaccacgc	atcaagtcg	gcatgtgc	caagaaggcc	tgctgtcacc	3780
tgaacccccc	aggcacctgc	gtgacatcg	gtaacggct	cgccacacgg	gccaacgaga	3840
gcatcattgg	cgccatcgcc	aggctgtca	agttcagca	ggtgtgcaaa	cccaagagca	3900
gcgtggagga	aaccggatgt	ctgttcgtgt	tcatcggtc	cgaccggaa	gccaggaccc	3960
acaaccccta	caagctgacg	agcacatcc	caaactcgta	caccggcagc	aggctgcacg	4020
aggccggctg	cgccccccagc	taccacgtg	tcaggggcga	tatccggcc	gccaccgggg	4080
gegtgtatcat	caacgtgcc	aacagcaagg	gcccacggg	aggcggatgt	tgccggccccc	4140
tgtacaagaa	gttcccccgag	agtttcgacc	tgacggccat	cgagggtggc	aaggccaggc	4200
tggtgaaggg	cgcccgatgg	cacatcgatc	acggccgtgg	ccccaaacttc	aacaagggtga	4260
gchgagggtga	aggcgacaa	cagctggcc	aagcctacga	gacatcgcc	aagatcgta	4320
acgacaataaa	ctacaagagc	gtggcatacc	cactgtc	caccggcata	ttcagcggca	4380
acaaggacag	gtgtaccccg	agctgtca	cgccctgtgc	accaccgtat	4440	
ccgacgtggc	catttcgtgc	aggggacaa	agttggatgt	gaccctgtgc	gaggccgtgg	4500
ccaggccggga	ggccgtggaa	gagatctgc	tcagcgcac	ctccagcg	accgagcccg	4560
acgcccggact	ggtgagggtg	caccccaaga	gctccctggc	cgccaggaa	ggctacagca	4620
ccagcgcacgg	caagacatcc	agtcacccgt	agggcacca	gttccaccag	gccgctaagg	4680
acatcgccga	gatcaacgt	atgtggcc	tggccaccga	ggccaaacgg	cagggtgc	4740
tgtacatct	ggcgagacg	atgtcgc	tcaggacaa	gtgccccgt	gaggaaacgc	4800
aggccagcac	accacccaggc	accctggcc	gcttgtcat	ccacgtatg	acacccgaga	4860
gggtgcagcg	gtgttgcggc	agcaggcccg	agcagatc	cgtgtgc	tccttccac	4920
tgcccaagta	caggatcacc	ggcgtgcga	agatccatgc	cagccacccc	atccgttca	4980
gccccaaagg	gcccgcctac	atccacccca	ggaagtac	gttggagacc	ccacccgtgg	5040
acgagacacc	cgagccaa	gcccggaa	agagcacccg	gggcacaccc	gagcagccac	5100
ccctgtatcac	cgaggacgag	acaaggaccc	ggaccccgaga	gcccacatt	atcgaggaag	5160
aggaaggagg	cagcatcgcc	ctgtcgatc	acggcccccac	ccaccagggt	ctgcagggtg	5220
aggccggat	ccacggccca	cccagctgt	ccagctgtat	ctggagcatc	ccacacggca	5280
gchgactcga	ctggacacgc	ctgacatcc	tggacaccc	ggaggccggcc	agcgtgaccc	5340
ccggcgccac	cagcgccgag	accacagct	acttcgcca	gagcatggag	ttcctggcca	5400
ggcccggtcc	agctcccgagg	accgtgttca	ggaaacccacc	ccacccaggct	cccaaggcc	5460
ggaccccaagg	cctggctccc	agcaggccct	gcaacggagc	cagectgg	agcacccccc	5520
ccggcggtaa	cagggtgtac	accaggagg	aacttggagg	cctgacaccc	agcaggaccc	5580
ccagcaggtc	ctgtgacagg	actagtctgg	tgtccaaacc	acccggcg	aacagggtga	5640
tcaccatgg	ggaatccatcg	gccttcgtgg	cccaagccata	gagacgttgc	gacccggcg	5700
catacatctt	cagcagcgac	acggccagg	gacacccgt	gttggacacc	gttggccaga	5760
cgctgtcgat	cgagggtgtg	ctggagagg	ccqagctgg	aatcgtatc	gccccccaggc	5820
tggaccatgg	gaaggaggaa	ctgtcgcga	agaaactgc	gttgaacccc	accccaaggca	5880
acaggacgcg	gttaccatgc	aggaagggtg	agaaacatgc	ggccatcc	gccaggccga	5940
tcctgcgggg	cctgggacac	tacccgtgg	ccgaggggaa	gttggatgtc	tacaggaccc	6000
tgcaccccccgt	gcactgtac	atccatgc	ccgttgc	tttcccaac	cccaagggtgg	6060
ccgtggggcc	ctgcaacgt	atgtgttgc	agaacttccc	caccgtggcc	agctactgc	6120
tatccatccca	gtacgcacgc	tacccgtgg	ttgtggaccc	cgccatgtgc	tgccctggaca	6180
ccggccatgt	ctggccccc	aaegtggat	gcttcccaaa	gaaacacgc	tacctggagc	6240
ccacccatcg	gagcgcggcc	cccacggcc	tcccaacac	cctgcacaa	gttgcaccc	6300
ctggccacca	gaggactgtc	aacgtgaccc	agatgagg	gttgcaccc	ctggacagcg	6360
ctgccttcaa	ctggaggatgc	ttcaagaa	acgcctgc	caacgatc	ttggagaccc	6420
tcaaggatgt	ccccatcagg	ctgaccggaa	agaacgtgt	gaaatcata	accaagatgt	6480
aggggccccc	ggccgtgtcc	ctgttcgt	ccgttgcata	cctgtacatc	ctggaggacca	6540
tcccaatgg	cagggtgt	atggacccgt	agaggggatc	gttggatgt	ccggccacca	6600
agcacacccg	ggagaggccc	aagggtgcagg	tgatccaggc	cgctgaccc	ctggccaccc	6660
cctacatgt	ccgcgtatcc	aggggatctgg	tgaggccgg	gaacgcgtgc	ctgtgcacca	6720
acatccacac	cctgtcgac	atgagcttcg	aggacttcg	cgccatccatc	cccgagact	6780
tccagccccc	cgactgtgt	ctggagaccc	acatcgcc	cttcgacaa	agcggaggatg	6840
acgttatggc	cctgacccgt	ctgtatgtcc	tggaggaccc	gggcgtggac	gccgagctgc	6900
tcaccctgtat	cgagggtg	ttccggcg	taagcttccat	ccacccgtcc	accaagacca	6960
agttcaatgt	ccgcgtatcc	atgaaaagg	gtatgttcc	gaccctgtgc	gtgaacaccc	7020
tgatcaat	tgtgtatcgcc	agcagggtgt	tgccggggaa	gctgaccggc	agccccctgcg	7080
ctgccttcat	ccggcgtacgc	aaacatcgta	agggcgtgt	aagcgtacaa	ctgtatggcc	7140
acagggtgc	cacccatcg	aaatcgatgg	tgaagatcat	cgacgcgtgc	gtggggcgaga	7200
aggccccctt	cttctgcggc	ggattcatcc	tgtgcgacag	ctgtgcaccc	accgcctgc	7260
gggtggccga	ccccctgtaa	aggctgttca	agctggggaa	gccactggcc	gctgacatg	7320
agcacacgc	tgacaggccg	agggccctgc	acgaggaa	caccagggtt	aacagggtt	7380
gcacccctgt	cgagctgtgc	aaggccgtgg	agagcgtgt	cgagaccgt	ggcaccacgca	7440

-continued

```
SEQ ID NO: 57          moltype = RNA    length = 448
FEATURE                Location/Qualifiers
misc_feature           1..448
note = Synthetic polynucleotide
source                 1..448
mol_type = other RNA
organism = synthetic construct
```

SEQUENCE: 57 Organism = Synthetic construct

gataggccgc	gcatgagaga	agccccagacc	aattacctac	ccaaatagga	gaaagttcac	60
gttgcacatcg	aggaagacag	cccatcttc	agagcttttc	agcggagctt	cccgcaagt	120
gaggtagaaag	ccaagcagg	cactgataa	gaccatgcta	atggccagac	gttttcgcat	180
ctggcttcaa	aactgtcga	aaacgggggt	gaccatctc	acacgtctt	tgatcatgg	240
atagtctgca	tagtacattt	catctgacta	atactacaac	accaccacca	tgaatagagg	300
atcccttaatc	atgtctggcc	ggcccccctt	ccccggcccc	actgcctatgt	ggaggccg	360
gagaaggagg	caggccggcc	cggaagccgg	agtctactaac	ttcagctgc	tgaagcaggc	420
tggqacqtg	qaggqaiaacc	ctggacct				448

SEQ ID NO: 58 moltype = length =
SEQUENCE: 58
000

SEQ ID NO: 59 moltype = length =
SEQUENCE: 59
~~~

SEQ ID NO: 60 moltype = length =  
SEQUENCE: 60  
A

SEQ ID NO: 61 moltype = length =  
SEQUENCE: 61

SEQ ID NO: 62 moltype = length =  
SEQUENCE: 62

SEQ ID NO: 63 moltype = length =  
SEQUENCE: 63

-continued

000

```

SEQ_ID NO: 64      moltype = length =
SEQUENCE: 64
000

SEQ_ID NO: 65      moltype = length =
SEQUENCE: 65
000

SEQ_ID NO: 66      moltype = length =
SEQUENCE: 66
000

SEQ_ID NO: 67      moltype = DNA length = 30
FEATURE          Location/Qualifiers
misc_feature     1..30
note = Synthetic polynucleotide
source           1..30
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 67
atggactacg acatagtcta gtccgcgaag                                30

SEQ_ID NO: 68      moltype = DNA length = 16
FEATURE          Location/Qualifiers
misc_feature     1..16
note = Synthetic polynucleotide
source           1..16
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 68
atggactacg acatag                                         16

SEQ_ID NO: 69      moltype = DNA length = 15
FEATURE          Location/Qualifiers
misc_feature     1..15
note = Synthetic polynucleotide
source           1..15
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 69
atggactacg acata                                         15

SEQ_ID NO: 70      moltype = length =
SEQUENCE: 70
000

SEQ_ID NO: 71      moltype = length =
SEQUENCE: 71
000

SEQ_ID NO: 72      moltype = DNA length = 7482
FEATURE          Location/Qualifiers
misc_feature     1..7482
note = Synthetic polynucleotide
source           1..7482
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 72
atggagaaaatgttca gatcgaggaa gacagcccat tcctcagagc tttgcagcg 60
atgttcggccg agtttgaggt agaaggccaa caggtaactg ataatgacca tgctaattgc 120
agagcgtttt cgcatctggc ttccaaaactg atcgaaacgg aggtggacc atccgacacg 180
atcccttgaca ttggaaatgc gcccccccgcc agaaatgtatt ctaagcacaa gtatcatgt 240
atctgtccga tgagatgtgc ggaagatccg gacagattgt ataagtatgc aactaagctg 300
aagaaaaact gtaaggaaat aactgataag gaattggaca agaaaatgaa ggagctggcc 360
gcggcatgaa ggcggccctga cctggaaatg gagactatgt gcctccacga cgacgagtcg 420
tgtcgctacg aaggccaaatgc cgcgttttac caggatgtat acgcgcgtca cggccccacc 480
agccctgtacc accaggccaa caaggccgtg agggtggccct actggatccg cttegcacacc 540
acacccttca tggtaagaa cctggccggc gcctacccaa gtcacagcac caactgggcc 600
gacgagacccg tgctgaccgc caggaacatc ggcctgtca gcagcgcacgt gatggagagg 660
agccggagag gcatgagcat cctggaaatg aaataccatgaa agcccaacaa caacgtgtg 720
ttcagcgtgg gcagcaccat ctaccacgag aagaggccac tgctcaggag ctggcacctg 780
cccaagcgtgt tccacccatg gggcaagccg aactacaccat gcagggtcga gaccatctg 840
agctgcgacg gctacgtggt gaagaggatc gccatcagcc cccgcgtgt cggcaagccc 900
agccggctacg ccgcataat gcacaggag ggcttcctgt gctgcaaggt gaccgacacc 960
ctgaacggccg agagggttagt cttccctgt tgccacccat cctgtgcgac 1020
cagatgaccg gcatcctggc caccgacgtg agccgcacgt acgcccagaa gctgctctg 1080
ggcctgaacc agaggatcgt ggtcaacccgc aggaccacca cacaatgaaag 1140

```

-continued

---

|             |             |            |            |            |            |      |
|-------------|-------------|------------|------------|------------|------------|------|
| aactaacctgc | tccccgtgg   | ggcccaggct | ttcgccagg  | gggccaagga | gtacaaggag | 1200 |
| gaccaggaa   | acgagaggcc  | cctggccctg | agggacaggc | agctggat   | gggctgtgc  | 1260 |
| tgggcctca   | ggccgcacaa  | gatcaccago | atctacaaga | ggccgcacac | ccagaccatc | 1320 |
| atcaaggta   | acagcgactt  | ccacagcttc | gtgtgccc   | ggatccggc  | caacaccctg | 1380 |
| gagatcgcc   | tgaggacccg  | gatcaggaa  | atgtggagg  | aacacaagg  | gcccgccca  | 1440 |
| ctgtatcacg  | ccggaggacgt | cgaggaccc  | aagtgcgt   | ccgacggc   | caaggagggt | 1500 |
| agggaggccg  | aggaacttag  | ggccgcctg  | ccacccctg  | ctggcagct  | ggaggaaacc | 1560 |
| acccttggaa  | ccgacgttga  | cctgtatgt  | caggaggcc  | gcgcggaa   | cgtggagaca | 1620 |
| cccaaggggcc | tgatcaagg   | gaccgttcc  | gacggcgagg | acaagatcc  | cagtcacgc  | 1680 |
| gtgtgtggcc  | cacaggccgt  | gtgtgtcc   | gagaagctg  | gctgtcatca | cccactggcc | 1740 |
| gagcaggatg  | tcgtgtatc   | ccacaggcc  | aggaaaggcc | ggtacgcgt  | ggagccctc  | 1800 |
| cacggcaagg  | tggtgtgtcc  | cgaggccac  | gcatccccc  | tgcaggact  | ccaggccctg | 1860 |
| agcggagacg  | ccacccatgt  | gtacaaacgc | aggaggatc  | tgaacaggta | cttcacccat | 1920 |
| atcccccaccc | acggcgaggc  | cctgacac   | gacggaaat  | actacaaga  | cgtgaaggcc | 1980 |
| agcgagacg   | acggcgaggta | cctgtacac  | atcgacagg  | agcagtgcgt | gaagaaagag | 2040 |
| ctgtgtgacc  | gcctggact   | gaccggcg   | ctgggtgg   | cacccttca  | cgagtgcgc  | 2100 |
| tacgagaccc  | tgaggacccg  | accccccgt  | cccttacc   | tgcggccat  | cgcgctgtac | 2160 |
| ggcggtcccg  | gcagcgaaaa  | gagggcata  | atcaagagc  | ccgtgaccc  | gaaagactg  | 2220 |
| gtgtgtcaqg  | ccaaagaaa   | gaactgtcc  | gagatcatca | gggacgtgaa | gaagatgaaa | 2280 |
| ggcctggacg  | tgaacgcgc   | caccgtgg   | agegtgtgc  | tgaacggct  | caagcacccc | 2340 |
| gtggagaccc  | tgtatcatc   | cgaggccct  | gttgtccac  | ccggcacct  | gagggccctg | 2400 |
| atcgccatca  | tcaggccaa   | gaaaggctg  | ctgtgtgg   | accccaagca | gtgcggctt  | 2460 |
| ttaaacatga  | tgtgtgttgc  | ggtgcactt  | aaccacg    | tctgtacca  | ggtgttccac | 2520 |
| aagagcatca  | gcaggcggt   | caccaagago | gtgaccagc  | tgtgtgac   | cctgttctac | 2580 |
| gacaagaaaa  | tgaggaccac  | caacccca   | gagaccaa   | tctgtatc   | caccacaggc | 2640 |
| agcacaaggc  | ccaaaggagg  | cgacccatgt | ctgacccgt  | ctggggctg  | ggtgaaggag | 2700 |
| ctgcagatcg  | actacaagg   | caacaggatc | atgaccgc   | ctggcagcc  | gggcctgacc | 2760 |
| aggaaggccg  | tgtacgcgt   | gaggtaca   | gtgaacgaga | acccactgt  | cgctccacc  | 2820 |
| agcgagacg   | tgaacgtgt   | gtgtaccagg | accggaggac | ggatctgt   | gaagaccctg | 2880 |
| gcggcgacc   | cttgcgttac  | gaccctgac  | gccaaggatc | ccggcaactt | caccgcacc  | 2940 |
| atcgaaagat  | ggcaggcgca  | gcacgcgc   | atcatgg    | acatcttgc  | gaggcccgc  | 3000 |
| cccaaccgc   | tgttccagaa  | caaggccaa  | gtgtgtgg   | ccaaggcc   | ggtgcggctg | 3060 |
| ctgaagaccc  | cggcgatc    | catgacc    | gagcagtgg  | acaccgtg   | ctacttcgag | 3120 |
| accgcaccc   | ccccacagcg  | cgagatcgt  | ctgaaacc   | tgtgtgttgc | gttcttcggc | 3180 |
| ctggacccgt  | acaggccct   | gttgcgc    | ccaccgtgc  | cactgtac   | caggaaacac | 3240 |
| cactggac    | acagccccag  | cccaacat   | tacggcc    | acaaggagg  | ggtcaggcag | 3300 |
| ctgagcaggc  | ggttccac    | gctgccc    | ggccgtgg   | ccggcagg   | gtacgacat  | 3360 |
| aacaccggca  | ccctggagg   | ctacgacc   | aggatcaac  | tgtgtccctg | gaacaggccg | 3420 |
| ctgccccac   | ccctgtgtc   | gcacc      | gacccac    | ggacccac   | agagcactt  | 3480 |
| gtgagcaac   | tgaaaggcag  | gaccgtgt   | gtgtgtgg   | agaagctg   | cgtgcggc   | 3540 |
| aagatgtgg   | actggctg    | cgacggccc  | gaggccac   | tccggcc    | gctggactc  | 3600 |
| ggcatccccc  | gogacgttgc  | caatgtac   | atcatcttc  | tgaacgtc   | gacc       | 3660 |
| aagtaccacc  | attaccagca  | gtgcgg     | acgc       | atgtgttgc  | gtgtgttgc  | 3720 |
| aaggccttcc  | tgcacccgt   | ccccgggg   | acccgtgt   | gcatgttgc  | cggttaccc  | 3780 |
| gacagggcc   | gcgagacat   | cattggc    | atgcacccg  | tgttcaag   | caggagg    | 3840 |
| tgaaacc     | agaggccgt   | ggggaa     | gagggtgt   | tctgttcat  | cggttac    | 3900 |
| cggaaggcc   | ggggccaa    | ccccata    | ctgac      | ccctgtac   | catcttac   | 3960 |
| ggcagcaggc  | tgcacccgg   | cggtgc     | ccca       | gtgtgttgc  | gggcgtat   | 4020 |
| gccaccggca  | ccggggcgt   | gatcat     | gtgttgc    | gcaaggcc   | ggccgtgg   | 4080 |
| ggagtgtccg  | ggccgttgc   | caaga      | cccgat     | ccgttgc    | ggccatcg   | 4140 |
| gtggggcaagg | ccagggttgc  | gggggg     | gtaa       | ggccat     | gtgtgttgc  | 4200 |
| aacttcaaca  | agggtgac    | gggttgg    | ggac       | ggccgttgc  | ctatccac   | 4260 |
| atcgccaaga  | tgtgtac     | caataact   | aa         | ggccgttgc  | ctatccact  | 4320 |
| ggcatcttcc  | ggggcaacaa  | ggacagg    | accc       | ggccat     | gttgcaccc  | 4380 |
| ctggacacca  | ccgtatcg    | cggttgc    | ccat       | ggccat     | gttgcaccc  | 4440 |
| ctgaaggagg  | ccgttgc     | gggggg     | gtgtgttgc  | gatgttgc   | gttgcaccc  | 4500 |
| agcgttgc    | ggccgttgc   | cgatgttgc  | agggttgc   | ggccat     | gttgcaccc  | 4560 |
| aggaaggcc   | acaggccat   | cgacgg     | accc       | ggccat     | gttgcaccc  | 4620 |
| caccaggcc   | ctaaggacat  | cgccg      | gtgtgttgc  | ggccat     | gttgcaccc  | 4680 |
| aacggcagg   | tgtgtatc    | catcttgc   | gaggtat    | ggccat     | gttgcaccc  | 4740 |
| cccggtggagg | aaagcgaggc  | cagcac     | ccca       | ggccat     | gttgcaccc  | 4800 |
| gttatgtac   | ccggagggt   | gtgtgttgc  | ggccat     | ggccat     | gttgcaccc  | 4860 |
| tgcagcttcc  | tccctacttgc | caatgtac   | atcacc     | ggccat     | gttgcaccc  | 4920 |
| cagccatcc   | tgtttagcc   | aaagg      | gtgtgttgc  | ggccat     | gttgcaccc  | 4980 |
| gagacccac   | ccgttgc     | gacacc     | ccaa       | ggccat     | gttgcaccc  | 5040 |
| acaccggac   | ggccaccc    | gatcacc    | ggc        | ggccat     | gttgcaccc  | 5100 |
| atcattatcg  | aggaaagg    | agaggatc   | cc         | ggccat     | gttgcaccc  | 5160 |
| cagggtgtc   | agggtgg     | cgacatcc   | ggcc       | ggccat     | gttgcaccc  | 5220 |
| agcatccac   | acggccac    | cttcgac    | gtgtgttgc  | ggccat     | gttgcaccc  | 5280 |
| ggccgcac    | tgaccc      | cgcc       | gtgtgttgc  | ggccat     | gttgcaccc  | 5340 |
| atggagtcc   | tggccagg    | cgtgttgc   | ccca       | ggccat     | gttgcaccc  | 5400 |
| ccaggcc     | ggacc       | ccaa       | ggccat     | ggccat     | gttgcaccc  | 5460 |
| ctgggtg     | ggccac      | ggcc       | gtgtgttgc  | ggccat     | gttgcaccc  | 5520 |
| acacccac    | ggaccc      | cgat       | gtgtgttgc  | ggccat     | gttgcaccc  | 5580 |
| ggcgttgc    | gggtgttgc   | cgat       | gtgtgttgc  | ggccat     | gttgcaccc  | 5640 |
| cggttgcac   | ccggcc      | cat        | gtgtgttgc  | ggccat     | gttgcaccc  | 5700 |
| aagagcgt    | ggcagacc    | gtgtgttgc  | ggccat     | ggccat     | gttgcaccc  | 5760 |
| agctaccc    | ccaggcttgc  | ccagg      | gtgtgttgc  | ggccat     | gttgcaccc  | 5820 |
| aaaa        | caggcaac    | gaggatgt   | gag        | ggccat     | gttgcaccc  | 5880 |



-continued

-continued

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| gcctggagga  | aaccgagggtg | ctgttcgtgt  | tcatcggtca  | cgaccggaaag | gccaggaccc  | 3960 |
| acaaccccta  | caagctgagc  | agcacccctga | caaacatcta  | caccggcagc  | aggctgcacg  | 4020 |
| aggccggctg  | cgccccccagc | taccacgtgg  | tcagggggcga | tatcgcacc   | gccaccggagg | 4080 |
| gcgtgtatcat | caacgcgtcc  | aacagcaagg  | gcccagccgg  | aggcggagtg  | tgccggccccc | 4140 |
| tgtacaagaa  | gttcccccgag | agcttcgacc  | tgca        | cgagggtggc  | aaggccaggc  | 4200 |
| tggtaaggg   | cgccgctaa   | cacatccatcc | acgc        | cccaacttc   | aacaagggtga | 4260 |
| gcgagggtgg  | aggcgacaa   | cagctggcg   | aagoctacga  | gagcatcgcc  | aagatcgta   | 4320 |
| acgacaataa  | ctacaagagc  | gtggccatcc  | cactgctca   | caccggcata  | ttcagcgca   | 4380 |
| acaaggacag  | gtgtggccag  | agcttcgaa   | acctgtccac  | cgccctggac  | accaccgttg  | 4440 |
| ccgacgtgg   | catctactgc  | agggaaaga   | agtgggagat  | gaccctgtgg  | gaggccgtgg  | 4500 |
| ccaggcggga  | ggccgtggaa  | gagatctgc   | tcagcgac    | ctccagctg   | accgagcccg  | 4560 |
| acgcccggct  | ggtgagggtg  | cacccaaga   | getccctggc  | cgccagggaa  | ggctacagca  | 4620 |
| ccagcgacgg  | caagacccctc | agcttcgtgg  | aggggcaccat | gttccaccag  | gccgtcaagg  | 4680 |
| acatcgccga  | gatcaacgtc  | atgtggcccg  | tgcccacccg  | ggccaa      | cagggtgtca  | 4740 |
| tgtacatct   | gggcgagagc  | atgttcagca  | tcaggagca   | gtgcccgtg   | gaggaaagcg  | 4800 |
| aggccagcac  | accacccagc  | accctgcct   | gcctgtcat   | ccacgtatg   | acacccgaga  | 4860 |
| gggtgcacg   | gtgtgggg    | agcagggccc  | agcagatcc   | ctgtgtcagc  | tcctttccac  | 4920 |
| tgeccaagta  | caggatcacc  | ggcgtcaga   | agatccatgt  | cagccagccc  | atctgttca   | 4980 |
| gccccaaaggt | gcccgcctac  | atccccca    | ggaagatctt  | ggttggqacc  | ccacccgtgg  | 5040 |
| acgagacacc  | cgagcca     | gcccggaa    | agagcaccga  | gggcacaccc  | gagcagccac  | 5100 |
| ccctgtatcc  | cgaggacgg   | acaaggacc   | ggacccca    | gcccattt    | atcgaggaa   | 5160 |
| aggaagatc   | cagcatcgc   | ctgtgtgg    | acggccccc   | ccaccagggt  | ctgcagggtg  | 5220 |
| aggccgacat  | ccacggccca  | cccacgtgt   | ccagcttcc   | ctggagcatc  | ccacacggca  | 5280 |
| gcaacttgc   | cg          | tgagcatcc   | tggacacc    | ggagggggcc  | agcgtgaccc  | 5340 |
| ccggcgcac   | cagcgcgg    | accacacgt   | acttcgc     | gagcatgg    | ttcctggca   | 5400 |
| ggcccggtcc  | agctcccagg  | accgttca    | ggaacccacc  | ccacc       | ccaggacca   | 5460 |
| ggaccccaag  | ctgggttccc  | acggggcc    | gagcggac    | cgccgtgt    | agcacc      | 5520 |
| ccggcgtgaa  | cagggtgtc   | accagggg    | aactgggg    | cctgacaccc  | agcaggaccc  | 5580 |
| ccagcaggcc  | cg          | tggtgtgg    | tgttca      | accggcgtg   | aacagggtg   | 5640 |
| tcaccaggga  | gaaatccgg   | gccttcgtgg  | cccacca     | gagacgttgc  | gacgcggc    | 5700 |
| cctacatctt  | cagcagcgc   | acggccagg   | gacac       | cttccatc    | cccaagggtg  | 5760 |
| cctgtcttag  | cgagggtgt   | ctggaggg    | cc          | gatgttgc    | gccccccagg  | 5820 |
| tggaccaggaa | gaggaggaa   | ctgttcagg   | agaaactc    | gctgtaccc   | acc         | 5880 |
| acaggagcc   | gttccacagg  | aggaagggt   | gaaat       | ccatcc      | ggcaggcgg   | 5940 |
| tctgtcagg   | cttgggacac  | tac         | ttgggg      | gggttgg     | tacaggaccc  | 6000 |
| tgcacccctg  | gcacactgtac | agcttc      | cg          | tttctcc     | cccaagggtg  | 6060 |
| ccgtggggc   | ctgcaacgt   | atgttca     | gg          | acttccc     | caccgtgg    | 6120 |
| tcatcccgta  | gtacgaccc   | tac         | ttgg        | ttgg        | tgccctggaca | 6180 |
| ccggccatcc  | ctggccggcc  | aa          | gttcc       | tttt        | tac         | 6240 |
| ccaccatcc   | gagcgcgg    | cccacgt     | tcc         | tttt        | gtgtgtgg    | 6300 |
| ctgcccac    | gaggact     | aacgtgac    | tttt        | tttt        | cc          | 6360 |
| ctgccttca   | cggtggat    | tttca       | tttt        | tttt        | cc          | 6420 |
| tcaaggagaa  | ccccatcagg  | ctgac       | tttt        | tttt        | cc          | 6480 |
| aggccccc    | ggccgtg     | cttgc       | tttt        | tttt        | cc          | 6540 |
| tcccaatgg   | cagggtcg    | atggac      | tttt        | tttt        | cc          | 6600 |
| agcacacccg  | ggagggcc    | aa          | gttgc       | tttt        | cc          | 6660 |
| cctacccgt   | cgccatcc    | agggt       | tttt        | tttt        | cc          | 6720 |
| acatccac    | cctgtc      | atgac       | tttt        | tttt        | cc          | 6780 |
| tccagccccc  | cgactcg     | ctgg        | tttt        | tttt        | cc          | 6840 |
| acgctatggc  | cctgac      | ctgtat      | tttt        | tttt        | cc          | 6900 |
| tcacccgtat  | cgagggt     | tttt        | tttt        | tttt        | cc          | 6960 |
| agtgtca     | cg          | tttt        | tttt        | tttt        | cc          | 7020 |
| tgtcaacat   | tgtgtcg     | atgaaa      | tttt        | tttt        | cc          | 7080 |
| ctgccttcat  | cgccgac     | aa          | atcgt       | tttt        | cc          | 7140 |
| acagggtgc   | cac         | tttt        | tttt        | tttt        | cc          | 7200 |
| aggccccc    | tttgc       | tttt        | tttt        | tttt        | cc          | 7260 |
| gggtggcga   | ccccctg     | tttt        | tttt        | tttt        | cc          | 7320 |
| agcacacg    | tgacagg     | tttt        | tttt        | tttt        | cc          | 7380 |
| gcacccgt    | cgatgt      | tttt        | tttt        | tttt        | cc          | 7440 |
| tcatcgat    | gttcat      | aa          | atcg        | tttt        | cc          | 7500 |
| ggggccctat  | aactct      | tttt        | tttt        | tttt        | cc          | 7560 |
| ggccgcac    | actcgat     | tttt        | tttt        | tttt        | cc          | 7620 |
| tttgtacat   | ccctc       | tttt        | tttt        | tttt        | cc          | 7680 |
| ggacgtgtt   | aa          | atc         | tttt        | tttt        | cc          | 7740 |
| tggctactat  | tgtgg       | tttt        | tttt        | tttt        | cc          | 7800 |
| aattggcaag  | ctgtt       | tttt        | tttt        | tttt        | cc          | 7860 |
| tttat       | tttt        | tttt        | tttt        | tttt        | cc          | 7920 |
| aaaaaaa     | aaatctag    | aaaaaaa     | aaaaaaa     | aaaaaaa     | aaaaaaa     | 7980 |
| aaaaaaaa    | aaaaaaa     | aaaaaaa     | aaaaaaa     | aaaaaaa     | aaaaaaa     | 8038 |

SEQ ID NO: 79 moltype = AA length = 2493

FEATURE Location/Qualifiers

REGION 1..2493

note = Synthetic polypeptide

source 1..2493

mol\_type = protein

organism = synthetic construct

SEQUENCE: 79

MEKVHVDIEE DSPFLRALQR SFPQFEVEAK QVTDNDHANA RAFSHLASKL IETEVDPDS 60

-continued

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| IILDIGSAPAR | RMYSKHKYHC  | ICPMRCAEDP  | DRLYKYATKL  | KKNCKEITDK  | ELDKMMKELA  | 120  |
| AVMSDPDLE   | ETMCLHDDES  | CRYEGQAVY   | QDVYAVDGP   | SLYHQANKGV  | RVAYWIGFDT  | 180  |
| TPFMFKNLAG  | AYPYSYTNWA  | DETFLTARNI  | GLCSSDVMER  | SRRGMSILRK  | KYKLPNSNNVL | 240  |
| FSVGSTIYHE  | KRDLLRSWHL  | PSVFHRLRGKQ | NYTCRCETIV  | SCDGYYVKRI  | AISPGLYGKP  | 300  |
| SGYAATMHRE  | GFLCCKVTDFT | LNGERVSFPV  | CTYVPATLCD  | QMTGILATDV  | SADDAQKLLV  | 360  |
| GLNQRIVVNG  | RTQRNTNTMK  | NYLLPVVAQA  | FARWAKEYKE  | DQEDEPLGL   | RDRQLVMGCC  | 420  |
| WAFRRHKITS  | IYKRPTQT    | IKVNSDFHSF  | VLPTRIGSNTL | EIGLRTRIRK  | MLEEHKEPSP  | 480  |
| LITADEVQEA  | KCAADEAKEV  | REAAEELRAAL | PPLAADVEEP  | TLEADVDLML  | QEAGAGSVET  | 540  |
| PRGLIKVTSY  | DGEDKIGSYA  | VLSPOAVLKS  | EKLSCIHPLA  | EQVIVITHSG  | RKGRYAVEPY  | 600  |
| HGKVVVPEGH  | APIVQDFQAL  | SESATIVYNE  | REFVNRYLHH  | IATHGGALNT  | DEEYYKTVKP  | 660  |
| SEHDGEYLID  | IDRKQCVKKE  | LVTGLGLTGE  | LDLPPFFHFA  | YESLRTRPAA  | PYQVPTIGVY  | 720  |
| GVPGSGKSGI  | IKSAVTKKDL  | VVSAKKENCA  | EII RDVKMKM | GLDVNARTVD  | SVLLNGCKHP  | 780  |
| VEFLYIDEAF  | ACHAGTLRAL  | IAIIPRKKAV  | LCGDPKQCGF  | FNMMCCLKVH  | NHEICTQVFH  | 840  |
| KSISSRCTKS  | ITSVVSTLFY  | DKKMRRTNPK  | ETKIVIDTT   | STKPKQDDLI  | LTCFRGWVQ   | 900  |
| LQIDYKGNEI  | MTAAASQGLT  | RKGYVAVRYK  | VNEENPLYAPT | SEHVNVNLLTR | TEDRIVWKT   | 960  |
| AGDPWIKTLT  | AKYPGNFTAT  | IEEWQAEHDA  | IMRHILERPD  | PTDVFQNKA   | VCWAKALPV   | 1020 |
| LKTAGIDMTT  | BQWNTVDFYFE | TDKAHSAEIV  | LNLQCVRFQG  | LDLDSGLFSA  | PTVPLSIRNN  | 1080 |
| HWDNSPSPNM  | YLNKEVVRQ   | LSRRYSPQLP  | AVATGRVYDM  | NTGTLRNYPD  | RINLVPVNRR  | 1140 |
| LPHALVLHNN  | EHPQSDFSS   | VSKLKGRTV   | VVGEKLSVPG  | KMVDWLSDRP  | EATFRARLD   | 1200 |
| GIPGDVPKYD  | IIFVNVRTPY  | KYHYHQOCED  | HAIKLSMLTK  | KACLHLPNGG  | TCVSIGYGYA  | 1260 |
| DRASESIIIGA | IAARLFKFSR  | CCKPKSLEET  | EVLFVFIGYD  | RKARTHNPYK  | LSSTLTNIYT  | 1320 |
| GSRLHEAGCA  | PSYHVVRGDI  | ATATEGVII   | AANSKGQPGG  | GGVCGALYKK  | PESFDLQPIE  | 1380 |
| VGKARLVKGKA | AKHIIHAVGP  | NFNVKVSEVEG | DQKQLAEEAYS | IAKIVNDNNY  | KSVAIPLLST  | 1440 |
| GIFSGNKDR   | TQSLNHLLTA  | LDTTDADVAI  | YCRDKKKWEMT | LKEAVARREA  | VEEICISDDS  | 1500 |
| SVTEPDAELV  | RVHPKSSLAG  | RKGYSTSDGK  | TFSYLEGTKF  | HQAACKDIAEI | NAMNPVATEA  | 1560 |
| NEQVCMYIL   | ESMSSIRSKC  | PVEESEASTP  | PSTLPCLCI   | AMTPPERVQLR | KASRPEQITV  | 1620 |
| CSSFPPLPKYR | ITGVQKIQCS  | QPILFSPKV   | AYIHPRKYL   | ETPPVDETPE  | PSAENQSTEG  | 1680 |
| TPEQQPLITE  | DETRTRTPEP  | IIIIEEEEDS  | ISLLSDGPTH  | QVLQVEADIH  | GPPSVSSSSW  | 1740 |
| SIPHASDPDV  | DSLSDILDTL  | GAWSITSGAT  | AETNSYFAKS  | MEFLARPVPA  | PRTVFRNPPH  | 1800 |
| PAPRTTRTSL  | APSACRACRTS | LVSTPPGVNR  | VITREELAL   | TPSRTPSRSV  | SRTSLVSNPP  | 1860 |
| GVNRVITREE  | FEAFVAQQQR  | RFDAGYIIFS  | SDTGQGHLQQ  | KSVRQTVLSE  | VVLERTELEI  | 1920 |
| SYAPRLDQE   | EELLRKKLQL  | NPTPANRSRY  | QSRKVENMKA  | ITARRILQGL  | GHYKLAEGKV  | 1980 |
| ECYRTLHPV   | LYSSSVNRAF  | SSPKVAVEAC  | NAMLKENPFT  | VASYCIIPEY  | DAYLDMVGDGA | 2040 |
| SCCLDTASFC  | PAKLRSPFKK  | HSYLEPTIRS  | AVPSAIQNTL  | QNVLAATKR   | NCNVTQMREL  | 2100 |
| PVLDSSAFMV  | ECKFKYACNN  | EYWFTFKENP  | IRLTEENVN   | YITKLKGPKA  | AALFAKTHNL  | 2160 |
| NMLQDIPMDR  | FVMDLKRDKV  | VTPGTKHTE   | RPKVQVIQAA  | DPLATAYLCG  | IHRELVRLRN  | 2220 |
| AVLLPNIHTL  | FDMSAEDFDA  | IIAEHFQPGD  | CVLETDTIAS  | DKSEDDAMAL  | TALMILEDLG  | 2280 |
| VDAELLTLIE  | AAFGEEISSH  | LPTKTKFKFG  | AMMKSGMPLT  | LFVNTVINIV  | IASRVLRERL  | 2340 |
| TGSPCAFIG   | DDNIVKGVK   | DKLMADRCAT  | WLNMEVKIIID | AVVGEKAPYF  | CGGFILCDCSV | 2400 |
| TGTACRVADP  | LKRLFKLGKP  | LAADDEHDD   | RRRALHEEST  | RWNRVGILSE  | LCKAVESRYE  | 2460 |
| TVGTSIIVMA  | MTTLASSVKS  | FSYLRGAPIT  | LYG         |             |             | 2493 |

SEQ ID NO: 80  
 FEATURE moltype = AA length = 2494  
 REGION Location/Qualifiers  
 1..2494  
 note = Synthetic polypeptide  
 source 1..2494  
 mol\_type = protein  
 organism = synthetic construct

|              |             |             |              |             |             |      |
|--------------|-------------|-------------|--------------|-------------|-------------|------|
| SEQUENCE: 80 |             |             |              |             |             |      |
| MPEKVHVDIE   | EDSPFRLAQ   | RSFPQFEVEA  | KQVTDNDHAN   | ARAFSHLASK  | LIETEVDP    | 60   |
| TILDIGSAPA   | RRMYSKHKYH  | CICPMRCAED  | PDRLYKYATK   | LKKNC       | KEITDK      | 120  |
| AAVMSDPDLE   | ETMCLHDDE   | SCRYEGQAV   | YQDVYAVDGP   | TSLYHQANKG  | RVAYWIGFDT  | 180  |
| TPFMFKNLAG   | AYPYSYTNWA  | ADETFLTARNI | IGLCSSDVMER  | SRRGMSILRK  | KYKLPNSNNVL | 240  |
| FSVGSTIYHE   | KRDLLRSWHL  | PSVFHRLRGKQ | NYTCRCETIV   | SCDGYYVKRI  | AISPGLYGKP  | 300  |
| SGYAATMHRE   | GFLCCKVTDFT | LNGERVSFPV  | CTYVPATLCD   | QMTGILATDV  | SADDAQKLLV  | 360  |
| GLNQRIVVNG   | RTQRNTNTMK  | NYLLPVVAQA  | FARWAKEYKE   | DQEDEPLGL   | RDRQLVMGCC  | 420  |
| WAFRRHKITS   | IYKRPTQT    | IKVNSDFHSF  | VLPTRIGSNTL  | EIGLRTRIRK  | MLEEHKEPSP  | 480  |
| LITADEVQEA   | KCAADEAKEV  | REAAEELRAAL | PPLAADVEEP   | TLEADVDLML  | QEAGAGSVET  | 540  |
| PRGLIKVTSY   | DGEDKIGSYA  | VLSPOAVLKS  | EKLSCIHPLA   | EQVIVITHSG  | RKGRYAVEPY  | 600  |
| HGKVVVPEGH   | APIVQDFQAL  | SESATIVYNE  | REFVNRYLHH   | IATHGGALNT  | DEEYYKTVKP  | 660  |
| SEHDGEYLID   | IDRKQCVKKE  | LVTGLGLTGE  | LDLPPFFHFA   | YESLRTRPAA  | PYQVPTIGVY  | 720  |
| GVPGSGKSGI   | IKSAVTKKDL  | VVSAKKENCA  | EII RDVKMKM  | GLDVNARTVD  | SVLLNGCKHP  | 780  |
| VEFLYIDEAF   | ACHAGTLRAL  | IAIIPRKKAV  | LCGDPKQCGF   | FNMMCCLKVH  | NHEICTQVFH  | 840  |
| KSISSRCTKS   | ITSVVSTLF   | DKKMRRTNP   | ETKIVIDTT    | STKPKQDDLI  | LTCFRGWVQ   | 900  |
| LQIDYKGNEI   | MTAAASQGLT  | RKGYVAVRYK  | VNEENPLYAPT  | SEHVNVNLLTR | TEDRIVWKT   | 960  |
| LAGDPWIKTL   | AKYRPGVFTA  | TIEEWFVHED  | AIMRHILERPD  | PTDVFQNKA   | VCWAKALPV   | 1020 |
| VLKTAGIDMTT  | BQWNTVDFYFE | ETDKAHSAEIV | VLNQCVRFQ    | GLDLDGSLF   | APTVP       | 1080 |
| NHWNDNSPSPN  | YLNKEVVRQ   | LSRSPQLP    | AVATGRVYD    | MNTGTLRNYP  | PRINLVPVNR  | 1140 |
| RLPHALVLHNN  | EHPKQSDFSS  | FVSKLKGRTV  | LVVGKLSV     | GKMDWLSR    | PEATFRARLD  | 1200 |
| LGIPGDVPKY   | IIIFVNVRTP  | YKYHYYQQCE  | DHAIKLSMLT   | KKACLHLPNG  | GTCVSIGYGY  | 1260 |
| ADRASESIIIG  | AIARLFKFSR  | VCKPKSSLEE  | TEVLFVFIGY   | DRKARTHNPY  | KLSSTLTNIY  | 1320 |
| TGSRLHEAGC   | APSYHVVRGDI | IATATEGVII  | AANSKGQPGG   | GGVCGALYKK  | FPESFDLQPI  | 1380 |
| EVGKARLVKG   | AKHIIHAVVG  | PNFNKVSEVE  | DQKQLAEEAYS  | IAKIVNDNNY  | YKSVAIPLLS  | 1440 |
| TGIFSGNKDR   | LTQSLNHLLT  | ALDTTDADVA  | IYCRDKKKWEMT | TLKEAVARRE  | AVEEICISDD  | 1500 |
| SSVTEPDAEL   | RVVHPKSSLA  | GRKGYSTSDG  | KTFSYLEGTK   | FHQAAKDIAE  | INAMWPVATE  | 1560 |
| ANEQVCMYIL   | GESMSSIRSK  | CPVEESEAST  | PPSTLPCLCI   | HAMTPERVQR  | LKASRPEQIT  | 1620 |
| VCSSFPPLPKY  | RITGVQKIQC  | SQPILPSPKV  | PAYIHPRKYL   | VETPPVDETP  | EPSAENQSTE  | 1680 |
| GTPEQPLPLIT  | EDETRTRTPE  | PIIIEEEED   | SISLLSDGPT   | HQVLQVEADI  | HGPPSVSSSS  | 1740 |

-continued

|             |              |             |             |              |              |      |
|-------------|--------------|-------------|-------------|--------------|--------------|------|
| WSIPHASFDFD | VDSL SILDTL  | EGASVTSAGT  | SAETNSYFAK  | SMEFLARPVP   | APRTVFRNPP   | 1800 |
| HPAPRTRTPS  | LAPS RACSR   | SLVSTPPGVN  | RVTREELEA   | LTPSRTPSRS   | VSRTSLVSNP   | 1860 |
| PGVNRVITRE  | EFEAFVAQQQ   | RRFDAGAYIF  | SSDTGQGHLO  | QKS VRO QVLS | EV LERTELE   | 1920 |
| ISYAPRLDQE  | KEELLRKKLQ   | LNPT PANRSR | YOSRKVENMK  | AITARRILQG   | LGHYLKAEGK   | 1980 |
| VECYRTLHPV  | PLYSSSVNRA   | FSSPV KVAE  | CNAMLKENFP  | TVASYCIPE    | YDAYLDMVDG   | 2040 |
| ASCCLDTASF  | CPAKLRSPFK   | HKS YLEPTIR | SAVPSAIQNT  | LQNVLAAATK   | RNCNV TQMRE  | 2100 |
| LPVLD SAAFN | VECFKKYACN   | NEYWETFKEN  | PIRLTEENVV  | NYITKLKGPK   | AAALFAKTHN   | 2160 |
| LNMLQDIPMD  | RFVMDLKRD    | KVTPGT KHT  | ERPKVQVQIA  | ADPLATAYLC   | GIHRELVRRL   | 2220 |
| NAVLLPNIHT  | LF DMSAEDFD  | AIIAEHFQPG  | DCVLETDIAS  | FDKSEDDAMA   | LTALMILEDL   | 2280 |
| GVDAELLTLI  | EA AFGEISSI  | HLPTKTKF    | GAMMKSGMFL  | TLFVN TVINI  | VIASRV LER   | 2340 |
| LTGSPCAAFI  | GDDNIVKGVK   | SDKL MADC   | TA WLMEVKII | DAVVG EKAPY  | FCGG FILCD   | 2400 |
| VTGTACRVA   | D PLKRLFKLGK | PLAADDEHDD  | DRR RALHEES | TRW NRVGILS  | ELCK AVE SRY | 2460 |
| ETVGT SIIVM | AMTTLASSVK   | SFSYLRG API | TLYG        |              |              | 2494 |

SEQ ID NO: 81                    moltype = AA    length = 2492  
 FEATURE                         Location/Qualifiers  
 REGION                         1..2492  
 note = Synthetic polypeptide  
 source                         1..2492  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 81  
 MEKVHV DIEE DSPFLRALQR SFPQF EVEAK QVT DNDHANA RAFSHLASKL IETEV DPSDT 60  
 IL DIGSAPAR RMYSKHKYHC ICPMRCAEDP DR LYK YTAKL KKN CKE ITDK ELDKKM KELA 120  
 AVMSDPD LET ETMCLH DDES CRYEQV AVY QDVYAVD GPT SLYHQANKGV RVAYWIGF DT 180  
 TPFMFKNL AGYPSYST NW A DETVLT ARN GLC S SDVMER SRRGMSI LRK KYLKP SNV NL 240  
 FSVG STIY HE PSV HLRG KQ NYTCR C E TIV SCDG YV VKRI AIS PGLY GKP 300  
 SGYA AT MHRE GFL CCKV TD LGN ERV SFPV CTY VPAT LCD QMTG I LAT DV SADDA QK LL 360  
 GLN QRI VV NG RT QRN NT TM NY LL P VVA QJA FAR WAKE YKE DQ EDER PLGL RDR QL VMG CC 420  
 WA FRR HKITS IY KRP D QT I KV NSDF HSF VL PRIGS NTL EIG LR TR IRK MLEE HK EP SP 480  
 LITA AEDI QEA KCA ADE AK E REAE AL RAAL PPLA AD FE PP T LEAD V DML QEAG AGS VET 540  
 PRGLIK VTSY AGED KIG SYA VL SPQ A VL KS EK LSC I HPLA EQ VIVI TH SG RKG RY AVE PY 600  
 HG KV VV PEGH AIP VQD FQAL SESAT IVY NE REP VN RY LH IATH GGA LNT DEE YK KV KP 660  
 SEHD GEY LYD L VT GL GLT GE L VD PPF H E YES L RTR PA A PY QV PTI G VY 720  
 GP VPG SGS KGI IKS A VT KKD L VV SAK ENCA EI RDV K KM K GL DV N ART VD SV LL NG CK HP 780  
 VET LYI DE AF ACHAG T L RAL IAI IRP KK AV L CG DP K QCG F NM M CL KV HF NHEIC T QV FH 840  
 KSI SRR CT KS V IS V U ST LF Y DK R M RTT NPK E TK I V ID TT G ST KPK Q D L I LTC FRG W V K Q 900  
 LQID Y KG NEI MTAA S Q GLT RKG VY AV RY K VNE N PLY APT SEH VN V LL TR TED RIV W K TL 960  
 AGDP W KI L T AK Y PGN FT AT EEW QAE HDA IM RH ILE RP D PT DV F QN KAN VC WA K AL VP V 1020  
 LK TAG ID MTT B QW NT VD Y F E TD KA H SAE IV LN QLC V RFF G LD L D S G L F S A PTV PLS IR NN 1080  
 HW D N S P S P N M Y GLN K B V R Q LS R RY P Q L P R AV AT GRV Y DM NT GTL R NY DP RIN L VP V N RR 1140  
 LPH AL R L H N EHP QSD FSS F VS K L K G RT V L V V GE K L S V P G KK DV W L S D Q P EAT FRAR L DL 1200  
 GIP GDV PKY D IV F IN R PT Y K Y HH Q QC ED HAI K L S M L T K KAC L H N P G G TCV S I G Y G Y A 1260  
 DRASE S I I G A IAR QPK F S R V C KPK S SHEET E VL FV FIG Y D R K A R T H N P Y K LS S T L T N I Y T 1320  
 GSR L HEAG CA PSY H V V R G D I AT AT EG V I I N A A N S K G Q P G G G V C G A L Y K F P E S F D L Q P I E 1380  
 VG KAR L V KGA AK HII HAG VP NFN K VSE BE DG K Q L A E A Y E S I A K I V D N N Y K S V A I P L L S T 1440  
 G IF SGN K D RL T Q S L N H L L T A L D T D A D V AI Y CR D K K W E M T L K E A V A R R E A V E E I C I S D D S 1500  
 SV TEP D A E LV R VHP K S S L A G R KG Y ST S D G K T F S Y L E G T K F H Q A A K D I A E I N A M W P V A T E A 1560  
 NEQ VCM Y IL G ESM SSI RSK C PVE E S E A S T P P S T L P C L C I H A M T P E R V Q R L K A S R P E Q I T V 1620  
 C S S F P L P K Y R I T G V Q K I Q C S Q P I L F S P K V P A Y I H P R K Y L V E T P P V E E T P E S P A E Q S T E G 1680  
 T P E Q P A L V N V D A T R T R M P E P I I I E E E E D S I S L L S D G P H T Q L V Q V E A D I H G S P S V S S S W 1740  
 S I P H A S D F D V D S L S I L D T L D G A S V T S G A V S A E T N S Y F A R S M E F R A R P V A P R T V F R N P P H 1800  
 P A P R T R T P P L A H S R A S S R T S L V S T P P G V N R V I T R E E L A L T P S R A P S R S A S R T S L V S N P P 1860  
 G V N R V I T R E E F E A F V A Q Q C R FD A G A Y I F S S D T Q G H L Q Q K S V R Q T V L S E V V L E R T E L E I S 1920  
 Y A P R L D T Q E E K L R K K L Q L N P T P A N R S R V S R V E N M K A I T R R L Q G L G H Y L K A E G K V E 1980  
 C Y R T L H P V P L Y S S V N R A F S S P K V A E C A N A M L K E N F P T V A S Y C I I P E Y D A Y L D M V D G A S 2040  
 C C L D T A S F C P A K L R S F P K K H S Y L E P T I R S A V P S A I Q N T L Q N V L A A A T K R N C N V T Q M R E L P 2100  
 V L D S A A F N V E C F K K Y A C N N E Y W T F K E N P I R L T E E N V V N Y I T K L K G P K A A A L F A K T H N L N 2160  
 M L Q D I P M D R F V D M L K R D V K V T P G T K H T E R P K V Q V I Q A A D P L A T A D L C G I H R E L V R R L N A 2220  
 V L L P N I H T L F D M S A E D F D A I I A E H F Q P G D C V L E T D A S F D K S E D D A M A L T A L M I L E D L G V 2280  
 D A E L L T L I E A A F G E I S S I H L P T K T K F K G A M M K S G M F L T F V N T V I N I V I A S R V L R E R L T 2340  
 G S P C A A F I G D D N I V K G V K S D K L M A D R C A T W L N M E V K I I D A V V G E K A P Y F C G G F I L C D S V T 2400  
 G T A C R V A D P L K R L F K L G K P L A V D D E H D D R R R A L H E E S T R W N R V G I L P E L C K A V E S R Y E T 2460  
 V G T S I I V M A M T T L A S S V K S F S Y L R G A P I T L Y G 2492

SEQ ID NO: 82                    moltype = DNA    length = 146  
 FEATURE                         Location/Qualifiers  
 misc\_feature                 1..146  
 note = Synthetic polynucleotide  
 source                         1..146  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 82  
 aggaaactta agtcaacaca acatatacaa aacaaacgaa tctcaagcaa tcaaggattc 60  
 tacttctatt gcagcaattt aaatcatttc ttttaaagca aaagcaattt tctgaaaattt 120  
 ttccaccattt acgaaacgata gccacc 146

SEQ ID NO: 83                    moltype = DNA    length = 270  
 FEATURE                         Location/Qualifiers

-continued

---

misc\_feature            1..270  
                        note = Synthetic polynucleotide  
source                1..270  
                        mol\_type = other DNA  
                        organism = synthetic construct

SEQUENCE: 83  
actcgagcta gtgactgact aggtctgg taccactaaa ccagcctcaa gaacaccgga 60  
atggaggctc taagctacat aataccaact tacacttaca aaatgttgc ccccaaaatg 120  
tagccatttcg tatctgtcc taataaaaaag aaagtttctt cacatcttag aaaaaaaaaa 180  
aaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa 240  
aaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa 270

SEQ ID NO: 84            moltype = RNA length = 1653  
FEATURE                Location/Qualifiers  
misc\_feature           1..1653  
                        note = Synthetic polynucleotide  
source                1..1653  
                        mol\_type = other RNA  
                        organism = synthetic construct

SEQUENCE: 84  
atggaagatc cccaaaaacat taagaaggcc ccagcgcctat tctaccact cgaagacggg 60  
accgcggccg acgcgcgtc caaagccatg aagcgctacg ccctgggtcc cggcaccatc 120  
gcctttaaccg acgcacatcat cgagggtggac attacctacg ccgagttactt cgagatgagc 180  
gttcggctgg cagaagctat gaagcgctat gggctgaata caaaccatcg gatctgggt 240  
tgccggcggc atagcttgca gttcttcatg cccgtgttgc gtgcctgtt catcggtgt 300  
gtctggccc cagctaacgc catctacaac ggcgcggcgc tgctgaacag catgggcattc 360  
agccagcccc cccgtcgatt cgtgagcaag aaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa 420  
aagaagctac cgatcatataa aaagatcatc atcatggata gcaagaccga ctaccaggc 480  
ttccaaagca tgtacacctt cgtgacttcc catttgcac ccggcttcaa cgagtacgac 540  
ttctgtggccg agagcttcga ccgggacaaa accatcgccc tgatcatgaa cagtagttgc 600  
agtaccggat tgcccaagggg cgtggcccta ccgcaccgcg ccggctgtgt ccgattcagt 660  
catgccccgg accccatctt ccggcaaccag atcatccccg acaccgttat cctcagcgtg 720  
gtgccatttc accacggcgtt ccggatgttc accacgttgc gctacttgat ctggggctt 780  
cgggtctgtc tcatgttgc ctteggaggc gagctattct tgccgcgtt gcaagactat 840  
aagattcaat ctggccctgtc ggtggccaca ctatggatg tttcgctaa gagactctc 900  
atcgacaatg acgacctaag caacttgcac gagatcgcca gcccgggggc gcccgtcagc 960  
aaggaggttag gtgaggccgt ggccaaacgc ttccacccatc caggcatccg acagggttac 1020  
ggcctgacag aacaaccccg ccggccatctt atcacccccc aaggggacga caagctggc 1080  
gcagtaggca aggtgggtcc cttttcgtcc gctaagggttgc tgacttgaa caccggtaag 1140  
acactgggtt gtaaccaggc ccggcggctg tgccgtccgtt gccccatgtat catgacggc 1200  
tacgtaaaca accccggggc tacaacgcgtt ctcatcgaca aggacggctg gctgcacagc 1260  
ggcgacatcg cctactggga ccggccggcgtt cacttcttc tgccggaccc gctgaagtcc 1320  
ctgtatcaat acaagggtca ccaggtagcc ccggccgaac tgaggatgtat cctgtgtca 1380  
caccggccaa ttttcgtccg ccggggccgc ggccgtccgtt acggacatgc ccggcggctg 1440  
cccgcccgac tctgtgtgtt ggaacacggt aaaaccatgtg ccggagaaggaa gatctggac 1500  
tatgtggccca ggcgggttac aaccggccaa aqctggccgt tggtgttgtt gttcgtggac 1560  
gagggtgcata aaggactgac ccggcaagggtt gacggccgcg aqatccgcgaa gattcttatt 1620  
aaggccaaaga aggccggccaa gatccgtgtt taa 1653

SEQ ID NO: 85            moltype = RNA length = 1653  
FEATURE                Location/Qualifiers  
misc\_feature           1..1653  
                        note = Synthetic polynucleotide  
source                1..1653  
                        mol\_type = other RNA  
                        organism = synthetic construct

SEQUENCE: 85  
atggaagatc cccaaaaacat taagaaggcc ccagcgcctat tctaccact cgaagacggg 60  
accgcggccg acgcgcgtc caaagccatg aagcgctacg ccctgggtcc cggcaccatc 120  
gcctttaaccg acgcacatcat cgagggtggac attacctacg ccgagttactt cgagatgagc 180  
gttcggctgg cagaagctat gaagcgctat gggctgaata caaaccatcg gatctgggt 240  
tgccggcggc atagcttgca gttcttcatg cccgtgttgc gtgcctgtt catcggtgt 300  
gtctggccc cagctaacgc catctacaac ggcgcggcgc tgctgaacag catgggcattc 360  
agccagcccc cccgtcgatt cgtgagcaag aaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa 420  
aagaagctac cgatcatataa aaagatcatc atcatggata gcaagaccga ctaccaggc 480  
ttccaaagca tgtacacctt cgtgacttcc catttgcac ccggcttcaa cgagtacgac 540  
ttctgtggccg agagcttcga ccgggacaaa accatcgccc tgatcatgaa cagtagttgc 600  
agtaccggat tgcccaagggg cgtggcccta ccgcaccgcg ccggctgtgt ccgattcagt 660  
catgccccgg accccatctt ccggcaaccag atcatccccg acaccgttat cctcagcgtg 720  
gtgccatttc accacggcgtt ccggatgttc accacgttgc gctacttgat ctggggctt 780  
cgggtctgtc tcatgttgc ctteggaggc gagctattct tgccgcgtt gcaagactat 840  
aagattcaat ctggccctgtc ggtggccaca ctatggatg tttcgctaa gagactctc 900  
atcgacaatg acgacctaag caacttgcac gagatcgcca gcccgggggc gcccgtcagc 960  
aaggaggttag gtgaggccgt ggccaaacgc ttccacccatc caggcatccg acagggttac 1020  
ggcctgacag aacaaccccg ccggccatctt atcacccccc aaggggacga caagctggc 1080  
gcagtaggca aggtgggtcc cttttcgtcc gctaagggttgc tgacttgaa caccggtaag 1140  
acactgggtt gtaaccaggc ccggcggctg tgccgtccgtt gccccatgtat catgacggc 1200  
tacgtaaaca accccggggc tacaacgcgtt ctcatcgaca aggacggctg gctgcacagc 1260  
ggcgacatcg cctactggga ccggggccgtt cacttcttc tgccggaccc gctgaagtcc 1320

-continued

|             |              |             |              |            |             |      |
|-------------|--------------|-------------|--------------|------------|-------------|------|
| ctgtatcaat  | acaaggcgta   | ccaggttagcc | ccagccgaac   | tggagagcat | cctgctgcaa  | 1380 |
| caccggaaaca | tcttcgacgc   | cggggctgcc  | ggcctgccc    | acgacgatgc | cgccgagctg  | 1440 |
| cccccccgac  | tcgtctgtct   | ggAACACCGT  | aaaaccatga   | ccgagaagaa | gatctgtggac | 1500 |
| tatgtggccca | ggccaggtttac | aaccgcggaa  | aaactgtcgccg | gttgtgttgt | gttctgtggac | 1560 |
| gagggtggcta | aaggactgtac  | cggccaaatgt | gacgcggccga  | agatcccgaa | gattttcatt  | 1620 |
| aaggcgaaga  | agggccggcaaa | gatcgccgttg | taa          |            |             | 1653 |

```
SEQ ID NO: 86          moltype = RNA  length = 1653
FEATURE                Location/Qualifiers
misc_feature           1..1653
note = Synthetic polynucleot
source                 1..1653
mol_type = other RNA
organism = synthetic constru
```

SEQUENCE : 86

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| atggaaagatg | ccaaaaacat  | taagaagggc  | ccagcgccat  | tctacccact  | cgaagacggg  | 60   |
| accgcggcgc  | agcagctgca  | caaagccatg  | aagcgctacg  | ccctgggcc   | cggccccatc  | 120  |
| gcctttaccc  | acgcacatc   | cgagggtggac | attacactcg  | cccgactt    | cgagatgagc  | 180  |
| gttcggcttg  | cagaagctat  | gaaggcgct   | gggctgataa  | caaacatcg   | catcgctgtg  | 240  |
| tgcagcgaga  | atagcttgc   | gttcttcatg  | cccggttgg   | gtgcctgtt   | catcggtgt   | 300  |
| gtctgtggccc | cagctaaca   | catctacaa   | gagcgcgagc  | tgcgttggac  | catggccatc  | 360  |
| agccagccccc | ccgtcgat    | cggtggacag  | aaaagggtgc  | aaaagatct   | caacgtgca   | 420  |
| aaaagactac  | cgatcataca  | aaagatcatc  | atcatggata  | cgaaaggcca  | ctaccggggc  | 480  |
| ttccaaagca  | tgtacaccc   | cgtgacttcc  | catttggcac  | ccggcttcaa  | cgagtaacgc  | 540  |
| ttctgtccccc | agagcttcg   | ccgggacaaa  | accatcgccc  | tgatcatgaa  | caqtatgtgc  | 600  |
| agtagccatg  | tggccaaggg  | cgtagccct   | ccgcgcgcga  | ccgetgtgt   | ccgattcagt  | 660  |
| catggccgcg  | accccatctt  | cgccaaaccc  | atcatcccc   | acacccgtat  | cttcaggctg  | 720  |
| gtgccatttc  | accacggctt  | cggcattgtt  | accacgctgg  | gtctacttgc  | ctggggcttt  | 780  |
| cgggtctgtc  | tcatgtaccc  | cttgcggagg  | gagctatttc  | tgcgcgactt  | gcaagactat  | 840  |
| aagatccat   | ctggccctgt  | ggtgcocaca  | ctatggat    | tcttcgttgc  | gagactctc   | 900  |
| atcgcaaga   | acgaccta    | caacttgcac  | gagatcgcca  | ggggggggcc  | ggccctca    | 960  |
| aaggaggtag  | gtgaggccgt  | ggccaaacgc  | ttccacccat  | caggcatccg  | acagggctac  | 1020 |
| ggcctgacag  | aaacaaccag  | cgccatttct  | atcacccccc  | aaggggacga  | caagcctggc  | 1080 |
| cgacttaggc  | aggtgtgtgc  | ttcttcgtcg  | gctaagggtgg | ttggacttgc  | caccggtaa   | 1140 |
| acactgggtt  | tgaaccaggc  | cgccgacgtt  | tgctgtccgt  | ggcccatgtat | catgaggccc  | 1200 |
| tacgttaaca  | accccgaggc  | tacaaacgc   | ctcatcgaca  | aggacggctg  | gctgcacagc  | 1260 |
| ggcgacatcg  | cctactggg   | cgaggacgag  | cacttcttca  | tgcgtggaccc | gctgaagtcc  | 1320 |
| ctgtatcaat  | acaagggttc  | ccaggatggc  | ccagecgaa   | tggagacat   | cctgtctgca  | 1380 |
| caccccaaca  | tcttcgacgc  | cggggttcgc  | ggcttgcggcc | acagcactgc  | ccggcgactgt | 1440 |
| cccgccgcag  | tgcgtgtgt   | ggAACACCGT  | aaaaccatga  | ccgagaagga  | gatgtggac   | 1500 |
| tatgtggccca | ggccagggtac | aaccgcggag  | aaactgcgcg  | gtgggtgtgt  | gttcgtggac  | 1560 |
| gagggtggcc  | aaggactgac  | ccgcggatgg  | gacgcggccga | agatccgcga  | gatttcatt   | 1620 |
| aacggcaaaqa | aaaggccggca | gatccgcgtq  | taa         |             |             | 1653 |

SEQ ID NO: 87 moltype = RNA length = 57  
FEATURE Location/Qualifiers  
misc\_feature 1..57  
note = Synthetic polynucleotide  
source 1..57  
mol\_type = other RNA  
organism = synthetic construct

SEQUENCE: 87  
atgaagttagg tgattttttttt gacccaaaaat ttttttttttt gcccccttac aattttttttt 57

```
SEQ ID NO: 88          moltype = RNA    length = 66
FEATURE                Location/Qualifiers
misc_feature           1..66
source                 note = Synthetic polynucleotide
                      1..66
                      mol_type = other RNA
                      organism = svnthetic construct
```

SEQUENCE: 88  
atgggttctt gacgctacgc cccaaatgtatc cgaccaggcaa aactcgatgt actttccgagg 60  
aactqa 66

SEQ ID NO: 89 moltype = RNA length = 837  
FEATURE Location/Qualifiers  
misc\_feature 1..837  
note = Synthetic polynucleotide  
source 1..837  
mol\_type = other RNA  
organism synthetic polynucleotide

```

SEQUENCE: 89
atgagagatga cagccccctag aaccttactg ctctcgatggggagctgt tgctctgaca 60
gagacatggg ctggatctct gaggcgggtg accggccagg gcctgtcat cggcgccgtg 120
ccaaacaggccc accagggtgt gtgcacacc accccagaaga ccagcgacgg cagttactac 180
ctggccgttc ccacccggcac actggggggatc tgcagccggc gcctggggcc ttgcatacgg 240
accacccatttc tgaacctgac caccggactac ttgcgtgttggaggtgtg gcccagggtg 300
acctaccacca gccccagctac cgccatccac cagttcgaga ggagggccaa gtacaagagg 360

```

-continued

---

gagcccgtaga gcctgaccct ggccctgctg ctggggggcc tgacaatggg cggcatcgcc 420  
 gcccgcgtgg gcacccggcac caccgcctg gtggccacc agcagttca gcagctgcag 480  
 gcccgcgtgc acgacgaccc gaaggagggt gagaagtcca tcaccaactt ggagaatgtcc 540  
 ctgaccggcc tgagcgagggt ggtgtgcag aacaggagggg gcctggaccc gctgttctg 600  
 aaggaggggcg gcctgtgcgc cgccctgaag gaggagtgcg gcctgtacgc cgaccacacc 660  
 ggcctggtaga tctgtggcat tgcgtgttgc ctggccgtcc tgcgtgttgc ggtgatttgg 720  
 gctgtgttgc cagtcgttat tgcagaaga aagtcatcccg gccggaaagg aggctctac 780  
 ttcaggctgc ttctgtctac agtgcctaga gctttatgtt gtttatctca gctgtaa 837

**SEQ ID NO: 90** moltype = RNA length = 378  
**FEATURE** Location/Qualifiers  
**misc\_feature** 1..378  
**source** note = Synthetic polynucleotide  
 1..378  
**mol\_type** = other RNA  
**organism** = synthetic construct

**SEQUENCE: 90**  
 atgagagtga cagcccttag aaccttaactg cttctgctttt gggggagctgt tgctctgaca 60  
 gagacatggg ctggatctta ccacagcccc agctacgcctt accaccagggtt cgagaggggg 120  
 ggaggaggctt ccggggggagg aggtccctgtt aagatcagcc agggccgtca cgccgcccac 180  
 gccgagatca acggggcccg ccggggagggtt atcgtgggtca ttgtcgctgg cctggccgtc 240  
 ctggccgtgg tgggtgttgc agctgtgttgc gcagtcgttta tggcagaag aaagtcatcc 300  
 ggcggaaagg gaggctctta ctctcagggtt gcttcgttca cagtcgttgc agtctttatgt 360  
 tgtttatctca agtgcgttgc 378

**SEQ ID NO: 91** moltype = RNA length = 876  
**FEATURE** Location/Qualifiers  
**misc\_feature** 1..876  
**source** note = Synthetic polynucleotide  
 1..876  
**mol\_type** = other RNA  
**organism** = synthetic construct

**SEQUENCE: 91**  
 atgagagtga cagcccttag aaccttaactg cttctgctttt gggggagctgt tgctctgaca 60  
 gagacatggg ctggatctta gaggcgagggtt accggccagggtt gcctgttgcgtt cgccgcccgtt 120  
 cccaagacc accagggtgtt gtcaacacc acccagaaga ccagcgcaccc cagtcgttgc 180  
 ctggccgttc ccacccggcac cacctgggtt tgccgttgcgtt ccgttgcgtt ttgtcgatc 240  
 accaccatcc tgaacacttgc acggactactt tgccgtgttgcgtt gggagctgttgcgtt 300  
 acattaccaca gccccagtaga cgccttccaccac cagttcgagaa ggaggccaa gtacaagagg 360  
 gagcccgtaga gcttcgttgcgtt ggccttgcgtt ctggccgtcc tgacaatggg cggcatcgcc 420  
 gccggccgtgg gacccggcac caccggccctgtt gtggccacc accgttca gcagctgcag 480  
 gccggccatgc acggacgaccc taacggagggtt gagaagtccca tcaccaactt ggagaatgtcc 540  
 ctgaccaggcc tggcgagggtt ggtgtgcag aacaggagggg gcctggaccc gctgttctg 600  
 aaggaggggcg gcttcgttgcgtt ccggccgttgcgtt gggagctgttgcgtt ccgttgcgtt 660  
 ggcctggtaga tctgtggcat tgcgtgttgc ctggccgtcc tgcgtgttgc ggtgatttgg 720  
 gctgtgttgc cagtcgttat tgcagaaga aagtcatcccg gccggaaagg aggctctac 780  
 ttcaggctgc ttctgtctac agtgcctaga gctttatgtt gtttatctca gctggccggc 840  
 ggaggccagcg actacaagga cgcacgttgc aagttaa 876

**SEQ ID NO: 92** moltype = RNA length = 417  
**FEATURE** Location/Qualifiers  
**misc\_feature** 1..417  
**source** note = Synthetic polynucleotide  
 1..417  
**mol\_type** = other RNA  
**organism** = synthetic construct

**SEQUENCE: 92**  
 atgagagtga cagcccttag aaccttaactg cttctgctttt gggggagctgt tgctctgaca 60  
 gagacatggg ctggatctta ccacagcccc agtacacgcctt accaccagggtt cgagaggggg 120  
 ggaggaggctt ccggggggagg aggtccctgtt aagatcagcc agggccgtca cgccgcccac 180  
 gccgagatca acggggcccg ccggggagggtt atcgtgggtca ttgtcgctgg cctggccgtc 240  
 ctggccgtgg tgggtgttgc agtgcgttgcgtt gcagtcgttta tggcagaag aaagtcatcc 300  
 ggcggaaagg gaggctctta ctctcagggtt gcttcgttca cagtcgttgc agtctttatgt 360  
 tgtttatctca agtgcgttgc gactacaagg acgcacgttgc caagttaa 417

**SEQ ID NO: 93** moltype = AA length = 550  
**FEATURE** Location/Qualifiers  
**REGION** 1..550  
**source** note = Synthetic polypeptide  
 1..550  
**mol\_type** = protein  
**organism** = synthetic construct

**SEQUENCE: 93**  
 MDAKNIKKKG PAPFYPLEDG TAGEQLHKAM KRYALVPGTI AFTDAHIEVD ITYAEYFEMS 60  
 VRLAEAMKRY GLNTNHRIVV CSENLQFFM PVLGALFIGV AVAPANDIYN ERELLNSMGI 120  
 SQPTVVFSK KGLQKILNVQ KKLPIQKII IMDSKTDYQF PQSMYTFVTS HLPPGFNEYD 180  
 FVPESFDRDK TIALIMNSSG STGLPKGVAL PHRTACVRFS HARDPIFGNQ IIPDTAILSV 240  
 VPFHHGFGMF TTLGYLICGF RVVLMYRFEE ELFLRSLQDY KIQSALLVPT LFSFFAKSTL 300  
 IDKYDLISNLH EIASSGAPLS KEVGEAVAKR FHLPGIRQGY GLTETTSAIL ITPEGDDKPG 360

-continued

---

AVGVVVVPEFE AKVVDLDTGK TLGVNQRGEL CVRGPMIMSG YVNNPEATNA LIDKDGWLHS 420  
 GDIAYWDEDE HFFIVDRLKS LIKYKGYQVA PAELESILLQ HPNIFDAGVA GLPDDAGEL 480  
 PAAVVVLEHG KTMTEKEIVD YVASQVTTAK KLRRGGVVFDV EVPKGLTGKL DARKIREILI 540  
 KAKGGGKIAV 550

SEQ ID NO: 94 moltype = AA length = 18  
 FEATURE Location/Qualifiers  
 REGION 1..18  
 note = Synthetic polypeptide  
 source 1..18  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 94 MKLVVVGAGG VGKSALTI 18

SEQ ID NO: 95 moltype = AA length = 21  
 FEATURE Location/Qualifiers  
 REGION 1..21  
 note = Synthetic polypeptide  
 source 1..21  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 95 MDPRRYAPMI RPAKLDVLPR N 21

SEQ ID NO: 96 moltype = AA length = 278  
 FEATURE Location/Qualifiers  
 REGION 1..278  
 note = Synthetic polypeptide  
 source 1..278  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 96 MRVTAPRTLL LLLWGAVALT ETWAGSLSEV TGQGLCIGAV PKTHQVLCNT TQKTS DGSSYY 60  
 LAAPPTGTTWA CSTGLTPCIS TTILNLTTDY CVLVELWPRV TYHSPSYAH QFERRAKYKR 120  
 EPVSLTLALL LGGLTMGGIA AGVGTGTTAL VATQQFQQLQ AAMHDDLKEV EKSITNLEKS 180  
 LTSLSEVVLQ NRRLGLDPLLFL KEGGLCAALK EECCLYADHT GLVIVGIVAG LAVLAVVVIG 240  
 AVVAAVMCRR KSSGGKGGSY SQAASATVPR ALMCLSQL 278

SEQ ID NO: 97 moltype = AA length = 125  
 FEATURE Location/Qualifiers  
 REGION 1..125  
 note = Synthetic polypeptide  
 source 1..125  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 97 MRVTAPRTLL LLLWGAVALT ETWAGSYHSP SYAYHQFERG GGGSGGGGSSL KISQAVHAAH 60  
 AEINEAGREV IVGIVAGLAV LAVVVGAVV AAVMCRKKSS GGKGGSYSQA ASATVPRALM 120  
 CLSQL 125

SEQ ID NO: 98 moltype = AA length = 291  
 FEATURE Location/Qualifiers  
 REGION 1..291  
 note = Synthetic polypeptide  
 source 1..291  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 98 MRVTAPRTLL LLLWGAVALT ETWAGSLSEV TGQGLCIGAV PKTHQVLCNT TQKTS DGSSYY 60  
 LAAPPTGTTWA CSTGLTPCIS TTILNLTTDY CVLVELWPRV TYHSPSYAH QFERRAKYKR 120  
 EPVSLTLALL LGGLTMGGIA AGVGTGTTAL VATQQFQQLQ AAMHDDLKEV EKSITNLEKS 180  
 LTSLSEVVLQ NRRLGLDPLLFL KEGGLCAALK EECCLYADHT GLVIVGIVAG LAVLAVVVIG 240  
 AVVAAVMCRR KSSGGKGGSY SQAASATVPR ALMCLSQLGG 291

SEQ ID NO: 99 moltype = AA length = 138  
 FEATURE Location/Qualifiers  
 REGION 1..138  
 note = Synthetic polypeptide  
 source 1..138  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 99 MRVTAPRTLL LLLWGAVALT ETWAGSYHSP SYAYHQFERG GGGSGGGGSSL KISQAVHAAH 60  
 AEINEAGREV IVGIVAGLAV LAVVVGAVV AAVMCRKKSS GGKGGSYSQA ASATVPRALM 120  
 CLSQLGGGGS DYKDDDK 138

SEQ ID NO: 100 moltype = RNA length = 9690  
 FEATURE Location/Qualifiers

-continued

```

misc_feature          1..9690
source                note = Synthetic polynucleotide
                      1..9690
                      mol_type = other RNA
                      organism = synthetic construct
SEQUENCE: 100
atggggcgccg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg 60
ttgtacatcg ggaagacagc ccattcccta gagcttgcg cgggagcttc cgcagttt 120
aggtagaaacg caagcaggctc actgtataatg accatgtcaa tgccagagcg ttttcgcac 180
tggcttcaaa actgtatcgaa acggagggtt accccatccga cacgtacccctt gacattggaa 240
gtgcgcggccg cccgagaatg tattctaaacg acaagtatca ttgtatctgtt cccatgtatgat 300
gtgcgcggccg tccggacaga ttgtataatgt atgcataacta gtcgaaagaaaa aactgttaagg 360
aaataactga taaggaaatg gacaagaaaaaa ttggaggatgtt ggcggccgtc ataggcgacc 420
ctgacccctggaa aacttgagact atgtgcctcc acgacgacga gtctgttcgcg tacgaaggccc 480
aagtgcgtgtt ttaccaggat gtataccggc tcgcacggcc caccacgttcc taccaaccagg 540
ccaaatccggc cgtgagggtt gcctactggg ttcggcttgcg caccacccctt ctatgttgc 600
aaagaaacctggc cggccgttccccc ccacgtatca gcaaccaactg ggcggccggcggc accgtgtcgta 660
ccggccaggaa catcgccctg tgcagcagcg acgtgtatggg gaggaggccgg agaggcatgaa 720
gcataccttgcg gaagaaatcac ctgaagccca gcaacaaacgt gctgttcgcg gtggccggcaga 780
ccatcttacca cgagaagagg gacgttcgcg ggatgtggcc cctggccggcgttccatgttcc 840
tgaggggccaa gcaagactac acctgttgcgat ggcggacccat cccatgttgcg gacggcttgc 900
ttgtgttgcgaa gatcgccatc agccccggcc ttgtacggcaat gcccacggccg taccggcgta 960
caatgtccacag ggagggcttc ctgtgtctca aggtgaccga caccctgttgcg ggcgaggggg 1020
tgagcttccccc cgtgtgcacc tacgtgcctgg ccacccctgttgcg cccatgttgcg accggccatcc 1080
tggccacccga cgtgacggcc gacgacggcc acggatgttgcg cttggccgttccatgttcc 1140
ctgtgttgcgaa cggcggaggcc cccatgttgcg gacggatgttgcg cccatgttgcg ggcgttgcg 1200
tggggccca ggttttgcgcc aggtggggcca aggagtacaa ggaggaccag gaagacgaga 1260
ggcccttgcggg cctggggggc accggatgttgcg tttgtatgttgcg cttggccggcc ttccggccggc 1320
acaatgttgcg cccatgttgcg gggggccggc acggatgttgcg cccatgttgcg gacggatgttgcg 1380
atcccaacatc cttcgtgtcc cccatgttgcg gacggatgttgcg cccatgttgcg ggcgttgcg 1440
ccggatgttgcg gggggccggc cccatgttgcg gacggatgttgcg cccatgttgcg ggcgttgcg 1500
acgtgtccggg gggccaaatgc gtcgtccggc accggatgttgcg tttgtatgttgcg ggcgttgcg 1560
tggggccggcc cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg ggcgttgcg 1620
tggccatgtatc ctgtgtccggc cccatgttgcg gacggatgttgcg cccatgttgcg ggcgttgcg 1680
aggtgttgcgaa ctacgtccggc gggggccggc acggatgttgcg tttgtatgttgcg ggcgttgcg 1740
ccgtgttgcgaa gttccggacaaatgc gtcgtccggc cccatgttgcg gacggatgttgcg cccatgttgcg 1800
tccaccacatc cggcggaggcc cccatgttgcg gacggatgttgcg cccatgttgcg ggcgttgcg 1860
tggccggggcc cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg ggcgttgcg 1920
tcgtgttgcgaa ctgtgtccggc cccatgttgcg gacggatgttgcg cccatgttgcg ggcgttgcg 1980
tggccatgtatc ctgtgtccggc cccatgttgcg gacggatgttgcg cccatgttgcg ggcgttgcg 2040
agttatgttgcg cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg ggcgttgcg 2100
gactgtccggc cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg ggcgttgcg 2160
ccagaccggcc cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg ggcgttgcg 2220
gaaagagccgg catgtatcgaa acggatgttgcg cccatgttgcg gggggccggc acggatgttgcg 2280
aaagaaactgtc cggccggatcc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 2340
cgcgcggccgg cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc 2400
tcgtgttgcgaa ctgtgtccggc cccatgttgcg gacggatgttgcg cccatgttgcg ggcgttgcg 2460
ccaaatccggc cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg ggcgttgcg 2520
tgaagggttgcg cttcaaccac gagatgttgcg cccatgttgcg gggggccggc acggatgttgcg 2580
gggtggccaaatgc gacgtgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 2640
ccaccaacccatc cggccggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 2700
aggacgttgcgaa gatcgccatc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 2760
gatccgttgcgaa gatcgccatc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 2820
ccgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc 2880
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc 2940
tcaagacccatc gacgtgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3000
ccggccggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3060
agaacaaatgc gacgtgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3120
tcgtatgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc 3180
cggccggatccatc cggccggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3240
ccggccggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3300
ccatgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3360
cggccggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3420
ggaactacgtc cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3480
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3540
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3600
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3660
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3720
ccatgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3780
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3840
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3900
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 3960
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 4020
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 4080
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 4140
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 4200
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 4260
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 4320
tgcgtgttgcgaa cccatgttgcg gggggccggc acggatgttgcg tttgtatgttgcg gggggccggc acggatgttgcg 4380

```

-continued

---

|            |             |            |            |             |             |      |
|------------|-------------|------------|------------|-------------|-------------|------|
| acaaggacag | gctgaccacag | agcctgaacc | acctgtcac  | cgcctggac   | accaccatg   | 4440 |
| ccgacgtggc | catctactgc  | agggacaaga | agtgggagat | gaccctgaag  | gaggcggtgg  | 4500 |
| ccaggcggga | ggccgtggaa  | gagatctgc  | tcagcga    | ctccagctg   | accgagcccg  | 4560 |
| acgcccgtgt | ggtgagggtg  | cacccaaga  | gtccctggc  | cggcaggaag  | ggctacagca  | 4620 |
| ccagcga    | caagaccc    | tgacttgc   | ggggcaccaa | tttccacccag | ggcgcttaagg | 4680 |
| acatcgcga  | gatcaacgt   | atgtggcc   | tggccacccg | ggccaaacgg  | cagggtgtgc  | 4740 |
| tgtacatct  | ggcgagagc   | atgtcc     | tcaggagca  | tgccccctg   | gaggaaacgc  | 4800 |
| aggccagcac | accacc      | accctgc    | gcctgtcat  | ccacgtatg   | acaccgaga   | 4860 |
| gggtgcacg  | gtgtgggg    | agcaggccc  | ageagatc   | cggtgtc     | tcctcc      | 4920 |
| tgc        | ccaaagta    | caggatc    | ggcgtgc    | cagecagcc   | atcctgttca  | 4980 |
| gccc       | aaagtg      | gccc       | atccaccc   | ggaatgt     | ggtgagacc   | 5040 |
| acgagacacc | cgagca      | gccc       | gagaacc    | agagcaccg   | gggcacaccc  | 5100 |
| cctgttcc   | cgaggac     | aca        | ggagggacc  | gggccc      | gcccattt    | 5160 |
| aggaagagga | cacatc      | ctgtg      | agggcc     | ccacagg     | ctgcagg     | 5220 |
| aggccgacat | ccacgg      | cccac      | cgctgt     | ccagtc      | ccacacgg    | 5280 |
| gca        | ctgtcg      | ctg        | gacatc     | tggac       | ggggcgcc    | 5340 |
| ccggc      | ccac        | accac      | acttc      | ggacat      | ttcttggcc   | 5400 |
| ggcc       | ctgg        | acgt       | ccat       | ggatgtt     | ttcagg      | 5460 |
| ggac       | ccaa        | cttgc      | agcagg     | ggaa        | ccacccat    | 5520 |
| ccgg       | gtgt        | accagg     | aa         | ctgg        | ccaccc      | 5580 |
| ccag       | cagg        | actgt      | tttcc      | ccccc       | gggg        | 5640 |
| tcac       | ccgg        | ggaat      | ccct       | gggg        | gggg        | 5700 |
| cctacat    | catt        | ccat       | ggatgtt    | ggatgtt     | ggatgtt     | 5760 |
| ccgt       | gtgt        | cgagg      | ctgg       | ccat        | ggatgtt     | 5820 |
| tggac      | ccgg        | ctgtc      | gggg       | ggatgtt     | ggatgtt     | 5880 |
| acagg      | ggcc        | gtacc      | gggg       | ggatgtt     | ggatgtt     | 5940 |
| tct        | tcgg        | actt       | ccat       | ggatgtt     | ggatgtt     | 6000 |
| tgc        | accc        | gttca      | ggatgtt    | ggatgtt     | ggatgtt     | 6060 |
| ccgt       | gggg        | ctgtc      | ggatgtt    | ggatgtt     | ggatgtt     | 6120 |
| teat       | cccg        | gtac       | gggg       | ggatgtt     | ggatgtt     | 6180 |
| ccgc       | cagg        | gttcc      | gggg       | ggatgtt     | ggatgtt     | 6240 |
| ccacc      | atcg        | ggcc       | gggg       | ggatgtt     | ggatgtt     | 6300 |
| ctgc       | ccac        | gggg       | ggatgtt    | ggatgtt     | ggatgtt     | 6360 |
| ctgc       | cttc        | ctgg       | ggatgtt    | ggatgtt     | ggatgtt     | 6420 |
| taa        | gggg        | ccat       | ggatgtt    | ggatgtt     | ggatgtt     | 6480 |
| agg        | ccca        | gtac       | gggg       | ggatgtt     | ggatgtt     | 6540 |
| tcc        | ccat        | gggt       | gggg       | ggatgtt     | ggatgtt     | 6600 |
| ag         | accc        | gggg       | ggatgtt    | ggatgtt     | ggatgtt     | 6660 |
| cct        | actt        | ccgc       | gggg       | ggatgtt     | ggatgtt     | 6720 |
| acat       | ccac        | ctgt       | gggg       | ggatgtt     | ggatgtt     | 6780 |
| tcc        | aggcc       | cgact      | gggg       | ggatgtt     | ggatgtt     | 6840 |
| acg        | cata        | gttgc      | gggg       | ggatgtt     | ggatgtt     | 6900 |
| tcac       | ccct        | gttgc      | gggg       | ggatgtt     | ggatgtt     | 6960 |
| ag         | tca         | gttgc      | gggg       | ggatgtt     | ggatgtt     | 7020 |
| tgat       | caat        | gtgt       | gggg       | ggatgtt     | ggatgtt     | 7080 |
| ctgc       | catt        | ggatgtt    | gggg       | ggatgtt     | ggatgtt     | 7140 |
| acagg      | gtgc        | cactt      | gggg       | ggatgtt     | ggatgtt     | 7200 |
| agg        | ccct        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 7260 |
| gggt       | ggcc        | cccc       | gggg       | ggatgtt     | ggatgtt     | 7320 |
| agca       | cacg        | tgac       | gggg       | ggatgtt     | ggatgtt     | 7380 |
| gcat       | cctg        | ggatgtt    | gggg       | ggatgtt     | ggatgtt     | 7440 |
| tcat       | ctgt        | ggatgtt    | gggg       | ggatgtt     | ggatgtt     | 7500 |
| gggg       | ccct        | aactt      | gggt       | ggatgtt     | ggatgtt     | 7560 |
| ggcc       | ccac        | atgg       | gggg       | ggatgtt     | ggatgtt     | 7620 |
| cga        | agac        | ggcc       | gggg       | ggatgtt     | ggatgtt     | 7680 |
| ccgg       | caccat      | gc         | gggg       | ggatgtt     | ggatgtt     | 7740 |
| cgag       | atgt        | gggg       | gggg       | ggatgtt     | ggatgtt     | 7800 |
| gatc       | gtgt        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 7860 |
| catc       | gggt        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 7920 |
| catgg      | ccat        | gggg       | gggg       | ggatgtt     | ggatgtt     | 7980 |
| caac       | gtca        | aaag       | gggg       | ggatgtt     | ggatgtt     | 8040 |
| ctacc      | agggg       | ttccaa     | ggatgtt    | ggatgtt     | ggatgtt     | 8100 |
| cgagt      | acac        | ggatgtt    | gggg       | ggatgtt     | ggatgtt     | 8160 |
| cagt       | gtgc        | atgg       | gggg       | ggatgtt     | ggatgtt     | 8220 |
| ccgatt     | catt        | ggatgtt    | gggg       | ggatgtt     | ggatgtt     | 8280 |
| cctc       | aggc        | ggatgtt    | gggg       | ggatgtt     | ggatgtt     | 8340 |
| ctgc       | gggt        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 8400 |
| gca        | agact       | aat        | gggt       | ggatgtt     | ggatgtt     | 8460 |
| gagg       | catt        | ggatgtt    | gggg       | ggatgtt     | ggatgtt     | 8520 |
| ggcg       | ctgc        | aaag       | gggg       | ggatgtt     | ggatgtt     | 8580 |
| acagg      | gtac        | ggcc       | gggg       | ggatgtt     | ggatgtt     | 8640 |
| caagg      | cgtc        | aaac       | gggg       | ggatgtt     | ggatgtt     | 8700 |
| cacc       | gtgt        | ggatgtt    | gggg       | ggatgtt     | ggatgtt     | 8760 |
| catg       | ggcc        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 8820 |
| gctgc      | acag        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 8880 |
| gctga      | gttc        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 8940 |
| cctg       | ctgt        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 9000 |
| cgcc       | ggcc        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 9060 |
| gatc       | gtgg        | tttgc      | gggg       | ggatgtt     | ggatgtt     | 9120 |

---

-continued

---

|            |            |            |             |            |            |      |
|------------|------------|------------|-------------|------------|------------|------|
| gttcgtggac | gaggtgccta | aaggactgac | cggcaattt   | gacgccccga | agatccgcga | 9180 |
| gattcttatt | aaggccaaga | aggcggcaa  | gatcgccgt   | taactcgagt | atgttacgt  | 9240 |
| caaagggtat | tgtcaccccc | cgaaagacca | tattgtgaca  | caccctcagt | atcacgccta | 9300 |
| aacatttaca | gcgcgggt   | caaaaaccgc | gtggacgtgg  | ttaacatccc | tgctgggagg | 9360 |
| atcagccgtt | attattataa | ttgggttgg  | gtctggctact | atttgtggca | tgtacggtct | 9420 |
| gacaaccag  | aaacataatt | gaatacagca | gcaatttggc  | agctgtttac | atagaactcg | 9480 |
| cggcgatgg  | catgcgcct  | aaaaattttt | attttatttt  | ttcttttctt | ttccgaatcg | 9540 |
| gattttgtt  | ttaatattt  | aaaaaaaaaa | aaaaaaaaaa  | aaaaatctag | aaaaaaaaaa | 9600 |
| aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa  | aaaaaaaaaa | aaaaaaaaaa | 9660 |
| aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa  | aaaaaaaaaa | aaaaaaaaaa | 9690 |

---

SEQ ID NO: 101      moltype = RNA   length = 9773  
 FEATURE      Location/Qualifiers  
 misc\_feature      1..9773  
                   note = Synthetic polynucleotide  
 source      1..9773  
                   mol\_type = other RNA  
                   organism = synthetic construct

SEQUENCE: 101

|             |             |            |             |             |            |      |
|-------------|-------------|------------|-------------|-------------|------------|------|
| attgacggcg  | tagtacacac  | tattgaatca | aacagccgac  | caattgca    | accatcacaa | 60   |
| tggagaagcc  | agtagtaaac  | gttagacgt  | accccccag   | tccgtttgt   | gtgcaactgc | 120  |
| aaaaaaagctt | cccgcaattt  | gaggtagtt  | cacagcaggat | cactccaaat  | gaccatgcta | 180  |
| atgccagagc  | atttcgcat   | ctggccagta | aactaatcg   | gttgcgggtt  | cctaccacag | 240  |
| cgacgatctt  | ggacataggc  | agcgcacccg | ctcgtagaa   | gttttccgag  | caccagtatc | 300  |
| atttgttctt  | ccccatcgct  | agtccagaag | acccggacgg  | catgtgaaa   | tatgcca    | 360  |
| aactggcgga  | aaaaggcg    | aaagattcaa | acaagaactt  | gtatgagaag  | attaggatc  | 420  |
| tccggaccgt  | atttgatacg  | ccggatgtct | aaacaccatc  | gtctgttctt  | cacaacgt   | 480  |
| ttacctgc    | catgegtgcc  | gaatattccg | tcatgcgg    | cgtgtatata  | aacgtcccg  | 540  |
| gaactatcta  | tcatcaggct  | atgaaaggcg | tgccgaccct  | gtactggatt  | ggcttcgaca | 600  |
| ccaccaggat  | catgttctcg  | gtctggcag  | gttgcgtacaa | accacactggg | 660        |      |
| ccgacgagaa  | agtccctgaa  | gcgcgttaca | tccgactttt  | cagcacaaag  | ctgagtgaag | 720  |
| gttaggacagg | aaaattgtcg  | ataatgagga | agaaggagg   | gaagccccgg  | tcgcgggttt | 780  |
| atttctccgt  | aggatcgaca  | ctttatccg  | aaacacagac  | cagcttgcag  | agtcggatc  | 840  |
| ttccatcggt  | gttccacttg  | aatggaaagc | agtcgtatc   | ttccgcgtt   | gatacagtgg | 900  |
| ttagttcgga  | aggctatcgta | gttgcggat  | ttccggatgt  | gatacagtgg  | 960        |      |
| ccgtggata   | ccgcgttaca  | caacaatgcg | aggcgttctt  | gtatgc      | gttactgaca | 1020 |
| cagtaaaagg  | agaacgggt   | tcgttccctg | tgtgcacgt   | catcccgccc  | accatatgcg | 1080 |
| atcagatgac  | ttgttatatcg | gcccacggta | tatccctg    | cgtgcacaa   | aaacttctgg | 1140 |
| ttgggctcaa  | ccagcgtt    | gtcattaa   | actggactaa  | caggaaacacc | aaacccatgc | 1200 |
| aaaatttactt | tctggcgtt   | atagcaca   | ggttcagca   | atgggtca    | gagcgcagg  | 1260 |
| atgatcttga  | taacgagaaa  | atgctggta  | ctagagaac   | caagcttac   | tatggctgt  | 1320 |
| tgtggcggtt  | tccgtacta   | aaagtatc   | cgttttatcg  | cccacctg    | acgcagac   | 1380 |
| cgctaaaagt  | cccgatctt   | tttagcgctt | ttcccatgtc  | gtccgtatgg  | acgac      | 1440 |
| tgcccatgtc  | gtcgaggc    | aaatggaaac | ttggcattgc  | accaa       | gaggaaaaac | 1500 |
| tgctgcagg   | ctcgaggaa   | ttagtcatgg | aggccaaggc  | tgcttttgc   | gatgtcagg  | 1560 |
| aggaaggccg  | agcgagaa    | ctccggagaa | cacttccacc  | attagtgca   | gacaaggcga | 1620 |
| tcgaggc     | cgcaga      | gtctgcgg   | tggagggg    | ccagggggac  | atcgaggcag | 1680 |
| cattagttg   | aaccccg     | ggtcacgt   | ggataatacc  | tcaagcaat   | gaccgtatg  | 1740 |
| tcggacagta  | tatcggttgc  | tcgcca     | acttgcgt    | gaatgc      | ccgcacca   | 1800 |
| cgaccccgct  | agcagatc    | gttacatca  | taacacactc  | cggaaatc    | ggaaggatc  | 1860 |
| cgtcgaacc   | atacgacgt   | aaagtatcg  | tgcacgg     | aggcgttgc   | ccatggccag | 1920 |
| aattcttgc   | actgtgt     | agcgcacgt  | tgatgtacaa  | cgaaagag    | tttgcgtacc | 1980 |
| gcaaaactata | ccacatttgc  | atgcattgc  | ccgcca      | agaa        | tacagaag   | 2040 |
| aggttacaaa  | ggcgcacgt   | gtcgcacac  | agtcgtgtt   | tgacgttgc   | aagaagcgtt | 2100 |
| gcgttacaaa  | ggagaagac   | tcaggatcg  | tcccttcggg  | agaatcgacc  | aaccctccct | 2160 |
| atcatgac    | actcttgc    | ggactgt    | cccgac      | ggtccgttac  | aaagtgc    | 2220 |
| caataggat   | gtatggcaca  | ccgggg     | gtcaagtc    | tattatca    | tcaactgtca | 2280 |
| ccggc       | cacgaga     | tcttgcgtt  | agcggaa     | aaatgggtt   | tcgcgtt    | 2340 |
| tgctatgt    | gagggtt     | cagattatcg | cgtgcacat   | cgaaatcg    | atgtcaca   | 2400 |
| gtatgcacaa  | accgttagaa  | gtgtgttgc  | tttgcgttgc  | gttgcgttgc  | cacgcagg   | 2460 |
| cactacttgc  | cttgatttgc  | atcgatc    | ccgcacaa    | ggtagtacta  | tgccggag   | 2520 |
| ccatgtatgt  | ccggatctt   | aaatcgat   | acttacat    | caccctgaaa  | 2580       |      |
| aagacatata  | cacca       | ttctaca    | atatctcccg  | gtgtgcaca   | cagccgtt   | 2640 |
| caacttgc    | atcgacat    | cattacat   | gaaaatgt    | aaccacgaa   | ccgtgc     | 2700 |
| agaacatata  | aatcgat     | acagggcc   | caaagccgaa  | gcaaggggat  | atcatctg   | 2760 |
| catgtttccg  | ccgggttgg   | aaatcgat   | aaatcgact   | tccggat     | gaagta     | 2820 |
| cacccggcc   | cccaacaa    | aaatcgat   | ttccggat    | ttccggat    | ttccggat   | 2880 |
| atgaaaatcc  | actgtac     | atcacat    | agcatgt     | ttccggat    | ttccggat   | 2940 |
| aggacacgg   | actgttgg    | atccgttgc  | gtcgacccat  | gattaac     | ctcacta    | 3000 |
| taccttacgg  | aaacttgc    | gtctactat  | aggacttgg   | agctgtt     | acac       | 3060 |
| tttgcgtat   | aaacagcccc  | actccctgt  | ccatccgt    | cagctgt     | accaacgtt  | 3120 |
| gtctggcc    | actatggaa   | ccgactact  | ccacggccgg  | tatcgatctt  | accgggttgc | 3180 |
| agtggggc    | actgttccca  | cagtttgc   | atgacaa     | accatcg     | tttacgcct  | 3240 |
| tagacgtat   | tttgcgtt    | tttttgcgtt | tgactgttgc  | aaggcgttgc  | tttttgcgtt | 3300 |
| agagcatccc  | actaacgt    | cateccgc   | attcagc     | cgag        | ccgggtatc  | 3360 |
| acagcccg    | aaaccgc     | tatggtac   | atcacgc     | tgccggc     | ctctccgt   | 3420 |
| gatttccgt   | gttccgtt    | gtctggaa   | gcacaca     | tgccgg      | acggggag   | 3480 |
| ccagagtat   | ctctgcac    | cataacc    | tccgg       | ccgcaat     | cctcacgc   | 3540 |
| tagtcccg    | gtacaagg    | aagcaacc   | cccggtc     | aaaatctt    | aaccagt    | 3600 |

-continued

aacaccactc agtacttgt gtatcagagg aaaaaattga agtcggcgta aagagaatcg 3660  
aatggatcgc cccgattgca atagccgtgt cagataaga ctacaacctg gtttcgggt 3720  
ttccgcgcga ggcaeggtac gacctgtgt ttcatacatc tgaaactaaa tacagaacaa 3780  
accacttca gcagtgccaa gacatgcgg cgacccataaa aaccctttcg cgttcggccc 3840  
tgaattgcct taacccagga ggcacccctc ttgtgaagtc ctatggctac cgccgacccgca 3900  
acagttagga cgtagtccac gcttgcga aaaaattgt cagggtgtct gcagcgagac 3960  
cagatgtgt ctcaaqcaat acagaatgt acctgtttt ccacaaacta gacaaacgccc 4020  
gtacacggca attcaccccg caccatctga atgcgtgtat ttgcgtcgta tagatgggtta 4080  
caagatgttggtggggc ggcgcgtat accgcaccaa aaggggaaat attgtgtact 4140  
gtcaagagga agcgttgcc aacgcagcca atccgtggg tagaccaggc gaaggaggct 4200  
ggccgtccat ctataaacgt tggccgacca gttttaccca ttccggccacg gagacaggca 4260  
ccgcaagaatg gactgtgtc tttagaaaga aagtgttca cgcggctggc cctgtatcc 4320  
ggaaaggccccccca agaaacggaaa gccttggaaat tgcttacaaa ccgttccatc cgcgtggcc 4380  
acttagtaaa tgaacataac atcaagtcgt tcgcatttc actgttatct acaggcatt 4440  
acgcacggcc aaaagacccg cttgaatgt cacttaactg ttgcacaaac ggccgtatqaca 4500  
gaactgtacgc ggacgttaacccatatttgc ttgataaaatggggaaagaaatcgacg 4560  
ccgcattccca acttaaaggag tctgttacacg agtgcggaa tgaatgtatc gagatcgacg 4620  
atgagttatg atggatccat ccagacgtt gcttgcgggg aaaaagggtt ttcgtacta 4680  
caaaaggaaa attgatttcg tacttcgaaag gacccaaattt ccatacaagca gcaaaagaca 4740  
tggccggatg aaagggtccctt tttccataatc accggggaaat taatggaaatc ctgtgtgtcc 4800  
acatattggg tgagaccatc gaagacatc gggaaaatgtt ccggccgtac caataccctg 4860  
cgtagccccccccc gccccaaacg ttgcgtgtc ttgcgtatgt tgccatgtc ccagaaagggg 4920  
tccacagact tagaaagcaat aacgtcaag aagtgttcaatg atgctctcc acccccttc 4980  
ctaagccacaa aattaaatggat gttcagaagg ttctgtgtc gaaatgttcc ctgtttaatc 5040  
cgccatcccgccgcatttttccggccgtt agtacatcgtc agtgcggccaa ggcgttcc 5100  
cttctctgtc acaggccggag gaggccccccg aagttgttgc gacacccgtca ccatactacag 5160  
ctgataaacac ctcgttgcgt gtcacagaca ttcacttggg tatggatgac agtagcgaag 5220  
gctcaatctt ttgcgtgtc acggccatccgca accacttctat tactatgtg gacatgttgc 5280  
cgtagccggatc tagttacta gaggatgttcc accggggca ggtgtgttgc tgctacgttc 5340  
atggccatccca agagccgtcc cctattccca cggccaaatgggaaatgatc gccccttgg 5400  
cagccggcaag aaaagacccg actccacccg caagcaatag ctctgtgtcc ctccacccct 5460  
ctttttgtggg ggtatccatc tccctccgtat caattttcgaa cggggagacg gcccggccagg 5520  
cagccgttaca accccctggca acaggccccca cggatgttgc ttatgttttcc ggatctttt 5580  
ccgacggaga gattgtatgg ctgagccggca gagaactctgaa ttatatcgcc cggatcgtcc 5640  
gatcatttgc accggggcgaa gtgaaacttca ttatatcgcc cggatcgtcc gtagatcttc 5700  
cttctacggca gcaagatgttcc accggccggaa gcaaggccgg gcaaggccgg gttgtgttgc 5760  
taggttgggtt catatttcc acggacacacg gcccggggca ttggccaaatggggaaatcggtt 5820  
tgcagaaacca gtttacggaaa ccggatcttgc acggcaatgtt ctggggaaatggggaaatcggtt 5880  
cggtgtctca caccgtcgaaa gaggaaacac tcaaaactcgat gtagccatgt atgcccacgg 5940  
aagccaaacaa aagtgtgtac cagtctcgta aagttagaaaatccggaaatggcc ataaacactt 6000  
acgcactact gtcaggacta cggatgttcaatccggccatc accgtccggccatc gaaatgttata 6060  
agatcacttca tccggaaacca ttgtacttgc tgtaggttcc accgtccggccatc gaaatgttca 6120  
agttcggtgt agtctgtctgt aacaactatc tgcatgaaatggggaaatcgatc gtagccatctt 6180  
atcagattac tgacgtgtac gatgtttact tggatgttgc acggggaca gtcggctgtcc 6240  
tggacactgc aacccttcgtc cccgtatgg ttagaaatggggaaatggggaaatcgatc 6300  
gagccccggaa tattccggatc ggggttccatc cggatgttcaatccggccatc gaaatgttca 6360  
tcatgtccgc aactaaatggggaaatcgatc tcaacgttcc accgtccggccatc tccatccgg 6420  
actcagccgc accatgttcaatccggccatc ggggttccatc tcaacgttcc accgtccggccatc 6480  
aggaggttccg tccggaaacca atttaggttcaatccggccatc ggggttccatc tcaacgttcc 6540  
gactgttggggg ccctaaaggcc gccggccatc ttgcggaaatggggaaatcgatc 6600  
agaaggttccg ttttgcgtatgg acatggaaatggggaaatcgatc 6660  
gcacggaaaca cacagaagaa agacccggaaatggggaaatcgatc 6720  
cgactgttca ttttgcgttgcggg atttccgggg aattttgttcc accgtccggccatc tccatccgg 6780  
ttccaaacat tccatccggccatc ggggttccatc tcaacgttcc accgtccggccatc 6840  
aacacttca gcaaggccgc accgtccggccatc tcaacgttcc accgtccggccatc 6900  
aagacgacgc tatgggttcaatccggccatc ggggttccatc tcaacgttcc accgtccggccatc 6960  
caacttccggccatc ttttgcgtatgg acatggaaatggggaaatcgatc 7020  
gtactgtttaatccggccatc ggggttccatc tcaacgttcc accgtccggccatc tccatccgg 7080  
acacatgttcaatccggccatc ggggttccatc tcaacgttcc accgtccggccatc 7140  
gatgttccgc gtttgcgttccatc tcaacgttcc accgtccggccatc tcaacgttcc accgtccggccatc 7200  
tggctgttccatc tcaacgttcc accgtccggccatc tcaacgttcc accgtccggccatc 7260  
gtggagaccc acctttccgtc ttttgcgtatgg ttagaaatggggaaatcgatc 7320  
cgccggccgttccatc tcaacgttcc accgtccggccatc tcaacgttcc accgtccggccatc 7380  
acgacgacgc agacggaaatggggaaatcgatc 7440  
gagtaggttcaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 7500  
ttacacttccatc tcaacgttcc accgtccggccatc tcaacgttcc accgtccggccatc 7560  
ttagggggaaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 7620  
cttagggggaaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 7680  
ctgacttataatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 7740  
cgccatttcaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 7800  
gttacccgttccatc ttttgcgtatgg ttagaaatggggaaatcgatc 7860  
ttagaaatggggaaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 7920  
ttagggggaaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 7980  
ggcggatgttccatc ttttgcgtatgg ttagaaatggggaaatcgatc 8040  
ggctgttccatc ttttgcgtatgg ttagaaatggggaaatcgatc 8100  
ttagaaatggggaaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 8160  
ttagggggaaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 8220  
ttagggggaaatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 8280  
accggccatccgtatccggccatc ttttgcgtatgg ttagaaatggggaaatcgatc 8340

-continued

```
SEQ ID NO: 102          moltype = RNA    length = 2086
FEATURE                  Location/Qualifiers
misc_feature             1..2086
note = Synthetic polynucleotide
source                   1..2086
mol_type = other RNA
organism = synthetic construct
```

```
SEQ ID NO: 103          moltype = RNA    length = 8095
FEATURE                  Location/Qualifiers
misc_feature             1..8095
                           note = Synthetic polynucleotide
source                   1..8095
                           mol_type = other RNA
                           organism = synthetic construct
SEQUENCE: 103
atgggcggcg catgagagaaa gccccagacca attaccttacc caaaatggag aaagttcacg 60
ttgacatcg ggaagacagc ccattctta gagctttca gcggagcttc ccgcagttt 120
```

-continued

---

|             |             |             |            |            |             |      |
|-------------|-------------|-------------|------------|------------|-------------|------|
| aggtagaagc  | caagcaggtc  | actgataatg  | accatgtcaa | tgccagagcg | tttgcac     | 180  |
| tggcttcaa   | actgtcgaa   | acggagggtg  | accatcgaa  | cacgtcc    | gacattggaa  | 240  |
| gtgcgcggc   | ccgcagaatg  | tattctaagc  | acaagatca  | ttgtatctgt | ccgatgagat  | 300  |
| gtgcggaga   | tccggacaga  | ttgtataagt  | atgcaactaa | gtcgaagaaa | aactgttaagg | 360  |
| aaataactga  | taaggaaattg | gacaagaaaa  | tgaaggagct | ggccgcgc   | atgagcgacc  | 420  |
| ctgacactgga | aactgagact  | atgtgcctcc  | acgacgacga | gtcggtc    | tacgaaggggc | 480  |
| aagtgcgtgt  | ttacaggat   | gtatcgccg   | tcgacggccc | caccagctg  | taccaccagg  | 540  |
| ccaacaagg   | cgtgagggtg  | gcctactgtg  | tcggcttcga | caccacaccc | ttcatgttca  | 600  |
| agaaacctggc | ccggcactac  | cccgatcata  | gcaccaactc | ggccgacgg  | accgtgtc    | 660  |
| ccgcaggaa   | catcgccgt   | tcgcgcgt    | acgtgtatgg | gaggagccgg | agaggcatga  | 720  |
| gcacatcttg  | gaagaatac   | ctgaagccca  | gcacacaact | gtgttc     | gtggcagca   | 780  |
| ccatcttacca | cgagaagagg  | gacgtgtca   | ggagctggc  | cgtcc      | gttccacc    | 840  |
| tgaggggca   | cgagaactac  | acctgcgtgt  | gcgagacat  | cgtagtgc   | gacggctacg  | 900  |
| tgtgtgtgg   | gtatcgccat  | agccccggc   | tgtacggca  | gcccgcgg   | taaggcgtc   | 960  |
| caatgcacag  | ggagggttcc  | ctgtgtc     | agggtacgca | caccctgaa  | ggcgagagg   | 1020 |
| tgaggttccc  | ctgtgtcacc  | tacgtgc     | ccaccctgt  | cgaccatg   | accggcatcc  | 1080 |
| tggccacccg  | ctgtgtcgcc  | gaecgcgc    | agaagctgt  | ctgtggct   | aaccagagga  | 1140 |
| tctgtgttgc  | ccgcggacc   | cagaggaa    | acaacacaa  | gaagaactac | ctgtgc      | 1200 |
| tgtgtggccca | ggcttcgtcc  | aggtggc     | aggagatcaa | ggaggaccc  | gaagacgaga  | 1260 |
| ggcccttggg  | cttgaggggac | aggcagctgg  | tgtatggct  | ctgtggggc  | ttcaggcgg   | 1320 |
| acaagatcac  | cacgtatctac | aaggccccc   | acacccacac | catcatca   | gtgaacacg   | 1380 |
| acttccacag  | cttcgtgtc   | cccaggatc   | gcacacaa   | cgttggatc  | ggccgtgagg  | 1440 |
| cccgatcag   | gaagatgtcg  | gaggaaacaca | aggagcccg  | cccactgtc  | accggcagg   | 1500 |
| acgtgcagg   | ggccaaatgc  | gctgcgc     | aggccaagg  | ggtgagggg  | gccgaggaa   | 1560 |
| tgaggggcgc  | cttgccaccc  | ctgggtc     | acgtgggg   | acccaccc   | gaaggcgc    | 1620 |
| tggacctgt   | gtcgaggag   | gcccggcgc   | gaagcgtgg  | gacccagg   | ggccgtatc   | 1680 |
| aggtgaccag  | ctacgacggc  | gaggaaaga   | tcggcgtca  | cgccgtc    | agcccacagg  | 1740 |
| ccgtgtgt    | gtccgagaag  | ctgagtc     | tccacccact | ggccgagc   | gtatgtc     | 1800 |
| tcacccacag  | ccgcggagg   | ggcaggta    | ccgtggggc  | cttacacgg  | aagggtgtc   | 1860 |
| tgcggagg    | ccacgcacat  | cccggtcg    | acttccagg  | cgttgc     | agcgcacacca | 1920 |
| tctgtgtacaa | cgaggggg    | ttctgt      | ttgtac     | ccatatgc   | acccacgg    | 1980 |
| gagccctgaa  | caccgacgag  | gaatactaca  | agaccgtgaa | gcccacg    | cacgacgg    | 2040 |
| agtacgtgt   | cgacatcgac  | aggaagcgt   | tcgtgtaa   | agagctgt   | accggctt    | 2100 |
| gactgtgt    | cgagctgtgt  | gaccacccat  | tcgtgtgt   | ccctacacg  | agcgtgtgg   | 2160 |
| ccagacccgc  | cgctccat    | cagggtc     | ccatccgg   | gtacggcgt  | ccggcagcg   | 2220 |
| gaaaagacgg  | catcatca    | agcgcgtg    | ccaaagaa   | cctgtgt    | agcgcacaa   | 2280 |
| aagagaactg  | cgccgacat   | atcagggac   | tgaagaa    | gaaaggct   | gacgtgtac   | 2340 |
| cgccgacccgt | ggacacgtgt  | ctgtgt      | ggcgtgt    | ccccgtgg   | accctgtaca  | 2400 |
| tgcacgggc   | tttcgttgc   | cacgcggc    | ccctgtggc  | atcatcagg  | cc          | 2460 |
| ccaaagaacg  | cggtgtgtc   | ggcgtccca   | agcgtgtc   | tttcc      | atgtgtcc    | 2520 |
| tgaagggtca  | tttcaacc    | gagatctc    | cccagggtt  | ccacaagac  | atcagcagg   | 2580 |
| gtgtgtacaa  | gagcgtgt    | agcgtgtgt   | gcacccctgt | ctacgacaa  | aaaatgggg   | 2640 |
| ccaccaaccc  | caaggagac   | aaaatgt     | tcgacacac  | aggcgtac   | aagccaa     | 2700 |
| aggacgact   | gtatgttgc   | tgcttc      | gctgggtgaa | gagcgtc    | atcgtacta   | 2760 |
| agggcaacga  | gatcatgacc  | gccgtgca    | ccctgggt   | gaccaggaa  | ggcgtgtac   | 2820 |
| ccgtgtgt    | caagggtgt   | gagaaccc    | tgtacgtc   | caccagc    | cacgtgtac   | 2880 |
| tgtgtgtgt   | caggacc     | gacaggatc   | tgtggaa    | cgttgc     | gacccttgg   | 2940 |
| taaaggccct  | gaccgc      | tacccggc    | acttcaccc  | caccatcg   | gagtgtcagg  | 3000 |
| cccgacacga  | cgccatcat   | aggcacatc   | tggagagg   | cgacccccc  | gacgtgttcc  | 3060 |
| agaaacaggc  | caacgtgtc   | tggcaagg    | ccctgtgt   | cgatgtgt   | accggcgg    | 3120 |
| tgcacatgt   | cacagacgg   | tggacaccc   | tggactact  | cgagcgg    | aaggccaca   | 3180 |
| ggcccgagat  | ctgtgtgt    | cagctgt     | tgagggtt   | eggccgt    | ctggcagc    | 3240 |
| gcctgttc    | cgcccccc    | gtgcac      | gtcatc     | caaccact   | gacaacac    | 3300 |
| caacggccaa  | catgtgtgc   | ctgaa       | agggtgt    | cgacgtgt   | aggegggtac  | 3360 |
| cacagctgc   | cagggtgt    | gcccacgg    | gggtgtac   | catgacac   | ggcaccc     | 3420 |
| ggaactacca  | ccccggatc   | aacctgtc    | ccgtgtac   | ggccgtc    | ccacgcctt   | 3480 |
| tgtgtgttca  | caacgacac   | ccacagac    | acttcac    | ttcgtgt    | gac         | 3540 |
| gcaggacccgt | gttgtgtgt   | ggcggaga    | tgacgtgt   | ccggc      | ggccgttgc   | 3600 |
| tgagcgtgt   | ccccggcc    | accttccgg   | ccagggtgt  | cctcgg     | cccgccgac   | 3660 |
| tgcgtgtgt   | cgacatcat   | tgcgtgt     | tcatcggt   | ccacgg     | ggccggac    | 3720 |
| agcgtgtgt   | ggaccacg    | atcaag      | gtacgtgt   | gtacgtgt   | ccacattac   | 3780 |
| tgaacccccc  | aggcacttgc  | gtgacgtat   | gtacgggt   | ccgc       | ggccgttgc   | 3840 |
| gcacatgtt   | cgcccccc    | agggtgtt    | gtgtgt     | ccacgt     | ggccgttgc   | 3900 |
| gcctgtgt    | aaacgggtgt  | ctgtgtgt    | tcatcggt   | ccacgg     | ggccggac    | 3960 |
| acaaccccta  | caagctgt    | agcaccct    | caaacatct  | caccgg     | aggctgtac   | 4020 |
| aggccgggt   | cgcccccc    | taccacgtt   | tcagggcg   | tatcg      | ccacc       | 4080 |
| ccgtgtgtat  | caacggtgt   | aacagca     | ggccgg     | agg        | ggccgttgc   | 4140 |
| tgtacaagaa  | gttccggag   | agctgtgt    | tgacgtgt   | ccacgt     | ggccgttgc   | 4200 |
| tgtgtgtgt   | cgccgtgt    | catctact    | acgggtgt   | gggtgt     | ggccgttgc   | 4260 |
| gcgggtgt    | aggcgtgt    | cgatgtgt    | cccca      | actt       | ccaaagggt   | 4320 |
| acgacaataa  | ctacaagac   | gtggccat    | cactgtgt   | caccgg     | gttccacc    | 4380 |
| acaaggacac  | gtgtgtgt    | agcgtgtgt   | ccgtgtgt   | ccacgt     | ggccgttgc   | 4440 |
| ccgacgtgt   | ccgtgtgt    | catctact    | acgggtgt   | gggtgt     | ggccgttgc   | 4500 |
| ccaggcgggaa | ggccgtgt    | gagatctgt   | tcagcgt    | ccacgg     | ggccgttgc   | 4560 |
| acgcccgt    | gtgtgtgt    | cacccca     | gtccctgt   | ccgg       | ggccgttgc   | 4620 |
| ccagcgtac   | caagaccc    | agttac      | ccgtgtgt   | ccacgt     | ggccgttgc   | 4680 |
| acatcgccg   | gtatcaacgt  | atgtggcc    | tggccac    | ggcc       | ggccgttgc   | 4740 |
| tgtacatct   | gggcgtgt    | atgtcc      | tccag      | ggcc       | ggccgttgc   | 4800 |
| aggccacac   | accaccc     | accctgt     | ccctgtgt   | ccacgt     | ggccgttgc   | 4860 |

-continued

---

```

gggtgcagcg gctgaaggcc agcaggcccg agcagatcac cgtgtgcagc tccttcccac 4920
tgccaagta caggatcacc ggcgtcaga agatccatg cagccagccc atccgttca 4980
gccccaaaggt gcccgcctac atccacccca ggaagtagacct ggtggagacc ccacccgtgg 5040
acgagacacc cgagccaagc gccgagaacc agagcacccg gggcacaccc gagcagccac 5100
ccctgatcac cgaggacgag acaaggaccg ggaccccaga gcccacatt atcgaggaag 5160
aggaagagga cagcatcagc ctgtgcgcg acggcccccac ccacccgttg ctgcagggtg 5220
aggccgacat ccacggccca cccacgtgt ccagctccag ctggagcatc ccacacggca 5280
gcaacttcga cgtggacagc ctgagcatcc tggacacccct ggagggcgcc agcgtgaccc 5340
ccggcgcac accgcgcgag accaacaatg acttcgcacc gggatggag ttccctggcc 5400
ggcccggtcc agctcccgag accgtttca ggaacccaca ccacccatg cccaggacca 5460
ggaccccaag ctgggttccc agcaggccct gcaacggac cagccgttg agcaccaccc 5520
ccggcgtgaa cagggtgate accaggggagg aactggaggc cttgcacccca agcaggaccc 5580
caagcggcgt cgtgacgagg actagtctgg tgccaaccc accccggctg aacagggtga 5640
tcacccaggaa ggaatcccgag gccttcgtgg cccatggca gagaatggc gacggccgg 5700
cttacatctt cagcagcgc accggccagg gacacccatc gcaaaatggc gtgaggcaga 5760
ccgtgcttag cgaggtgtg ctggagggaa cggagcttca aatcagctac gcccccaaggc 5820
tgacccaggaa gaaggaggaa ctgtccatgg agaaaactggc gctgaacccc accccaggca 5880
acaggagcgc taccatggc aggaatgggg aacatggaa ggccatccac gccaggcgga 5940
tccctgcaggc cttgggacac tacctggagg ccggaggccaa ggtggatgtc tacaggaccc 6000
tgcaccccggt gcaactgtac agtccacccg tgaacaggccg cttctccatc cccaaagggtgg 6060
ccgtggaggc ctgcacccgtc atgtgttcaagg agaaacttccc caccgtggcc agtactgtca 6120
tcatcccgaa tgaatggcc tacctggca tggggatggc cggcagctgc tgcctggaca 6180
ccggccaggctt ctggcccgcc aatgtgggg gtttcccaaa gaaacacccg tacctggagc 6240
ccaccatcg gageggcgtg cccaggccca tccagaacac cttgcacaaat gtgctggccg 6300
ctggccaccaa gggaaactgc aacgttgcacc agatggggaa gtcggccgtg ctggacacccg 6360
ctggcccttcaa ctggccatccat ttcaaaaatc acggccatccaa caacggatgtac tggggagaccc 6420
tcaaggagaa ccccatcagg ctgacccggg agaaatgggtt gaaactatc accaaatgtca 6480
aggggcccaa ggccgtgtcc ctgttcgttca agacccaaatc cttgcacatg ctgcaggaca 6540
tcccaatgtt cagggtcggt atggatgtca agagggttgcgaa aggggttgc accccggacca 6600
agoacacccgaa ggagggccgg aaggatggcgg tgatccaggc cgatgcacca ctggccacccg 6660
cttacccgttccat cggccatccac aaggatggctt tgaggccggc gaacggccgtg ctgctggcc 6720
acatccaccc acgttccatc atgagccggc aggacttgcg ccggccatccatc gccggacact 6780
tccagcccccgg cgaactgtgt ctggagaccg acatccggccat cttgcacaaat agcggaggatg 6840
aegctatggc ctgtatgtatc tggggatgtc tggggatgtt gggggatggc gccggatgtc 6900
taccctgtat cggggatgtcc ttccggggaa tcaatccatc ccacccgtgg accaaagacca 6960
agttcaatgtt cggccgtatgtt ataaaaatggc gaatgttccat gaccctgttc gtgaaacccg 7020
tgatccatgtt tggatgttccat tggggatgttccat gtcggggaggg gtcggccggc agccctgtcg 7080
ctggcccttcat cggccatccatc aacatccgttca agggccgttca aagggccacatc ctgtatggcc 7140
acagggtgtcc caccatggc aacatgggg tggatgtatc cgacccgtgt gttggccgaga 7200
aggccccctt cttccgttccat gggatccatc ttggccggccat cttgcacccggc accggccgtca 7260
gggtggccgaa cccctgttca aggtgttca agctggggca gcaactggcc gctgacatgt 7320
agcacacccgaa tgacccgttccat aacggcccttccat aacggggaaatc caccagggtgg aacagggtgg 7380
gcatccctgttccat cggatgttccat aacggccgttccat gggatgttca agggccgttca ggcggccatc 7440
tcatctgtat ggttccatccatc aactccgttca gtcggccatccat ggttccatccatc tacctgggg 7500
ggggcccttat aactctctac ggctaacatc aatggacttac gacatgttccat agtccgcacca 7560
ggccgcaccat catggatgttccat gggccgggggggg tggggccaaa agccgttccat 7620
caatttgcactt cggatgttccat acgtgttccat aacggccatccat ccccccggaaa gccatatttgc 7680
tgacacccatccat cttccatccatc gcccaaatatc ttacccgttccat ggttccatccat accccgttccat 7740
cgtgggttacatccctgttccat gggatgttccat cttccatccatc ggttccatccatc tacctgggg 7800
ctactatgtt ggttccatccatc gtttccatccatc aactccgttccat gggccatccatc tacctgggg 7860
tggcaagcttccatccat gtttccatccatc aactccgttccat gggccatccatc tacctgggg 7920
attttttttttccatccatccat gtttccatccatc aatccgttccat gggccatccatc tacctgggg 7980
aaaaaaaaaaaa tctagaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa 8040
aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaa 8095

```

```

SEQ_ID NO: 104 moltype = RNA length = 8120
FEATURE Location/Qualifiers
misc_feature 1..8120
note = Synthetic polynucleotide
source 1..8120
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 104
atggacccgg tagtacacac tattgaatca aacagccgcac caattgcact accatcacaa 60
tggagaaggcc agtagtaaacat gtagacgttag acccccagag tccgtttgtc gtgcaactgc 120
aaaaaaaaatccat cccgcataatc tggatgttccat gggccatccat ccccccggaaa gccatatttgc 180
atggccatccat cttccatccatc aacttccatc gtttccatccat gggccatccat ccccccggaaa gccatatttgc 240
cgacccatccat cggatgttccat aacggccatccat ccccccggaaa gccatatttgc 300
atggatgttccat ccccccggaaa gggatgttccat gggccatccat ccccccggaaa gccatatttgc 360
aactggccggaa aaaaatgttccat gggccatccat ccccccggaaa gccatatttgc 420
tccggaccgttccatccat gggccatccat ccccccggaaa gccatatttgc 480
ttacccgttccat cggatgttccat gggccatccat ccccccggaaa gccatatttgc 540
gaactatccat cttccatccatc aatccgttccat gggccatccat ccccccggaaa gccatatttgc 600
ccacccatccat cttccatccatc gggccatccat ccccccggaaa gccatatttgc 660
ccgacccatccat cttccatccatc gggccatccat ccccccggaaa gccatatttgc 720
gtggacccatccat cttccatccatc gggccatccat ccccccggaaa gccatatttgc 780
atccatccatccat cttccatccatc aatccgttccat gggccatccat ccccccggaaa gccatatttgc 840
ttccatccatccat cttccatccatc gggccatccat ccccccggaaa gccatatttgc 900
tgagttccatccat cttccatccatc gggccatccat ccccccggaaa gccatatttgc 960

```

-continued

---

|             |             |            |             |            |            |      |
|-------------|-------------|------------|-------------|------------|------------|------|
| cgctgggata  | cgcgggttaca | cacaatageg | agggcttctt  | gctatgc当地  | gttactgaca | 1020 |
| cagtaaaagg  | agaacgggt   | tcgttccctg | tgtgcacgt   | catccc当地   | accatatgc  | 1080 |
| atcagatgac  | tggtataatg  | gccacggata | tatcacctga  | cgatgc当地   | aaaccttctt | 1140 |
| ttgggctca   | ccagcgaatt  | gtcattaacg | gttaggactaa | caggaacacc | aacaccatgc | 1200 |
| aaaatttacct | tctgcgatc   | atagcacaag | ggttcagc当地  | atgggctaa  | gagcgc当地   | 1260 |
| atgatcttga  | taacgagaaa  | atgctgggt  | ctagagaacg  | caagcttacg | tatggctgt  | 1320 |
| tgtggggctt  | tccactaag   | aaagtacatt | cgttttatcg  | cccacctgga | acgc当地     | 1380 |
| gcgtaaaagt  | cccacgttct  | tttagcgtt  | ttcccatgtc  | gtccgtatgg | acgc当地     | 1440 |
| tgcccatgtc  | gtgaggcag   | aaatggaaa  | tggcatttgc  | accaaagaag | gaggaaaaaa | 1500 |
| tgctcgagg   | ctcgaggaa   | ttatgtatcg | aggccagg    | tgcttttgag | gtgtcagg   | 1560 |
| aggaaggcc   | agcggagaag  | ctccgagaag | cacttccacc  | attagtgcc  | gacaaggc   | 1620 |
| tcgaggecgc  | cgcagaagtt  | gtctgc当地   | tggagggct   | ccaggeggac | atcgagg    | 1680 |
| cattagttga  | aacccggc    | ggtcacgtt  | ggataatacc  | tcaagcaat  | gaccgtat   | 1740 |
| tcggacagta  | tatcggttgc  | tcgccaact  | ctgtgtgaa   | gaatgc当地   | ctcgccac   | 1800 |
| cgcacccgt   | agcagatcag  | gttaagatca | taacacact   | cggaagatca | ggaaggta   | 1860 |
| cggcgaacc   | atacgcgt    | aaagtactga | tgccagcagg  | aggtgc当地   | ccatggcc   | 1920 |
| aattccatgc  | actgatgt    | agcgc当地    | tagtgc当地    | cgaaagagag | tttgtgaa   | 1980 |
| gcaaaactata | ccacattgc   | atgcattgc  | ccgccaagaa  | tacaaagag  | gagcgtaca  | 2040 |
| aggttacaaa  | ggcagacgtt  | gcagaaacag | agtacgtt    | tgacgtgac  | aagaaggc   | 2100 |
| gcgttaagaa  | ggaagaagcc  | tcaggcttgg | tcctctcg    | agaactgacc | aaccctcc   | 2160 |
| atcatgac    | actcttgc    | ggactgaa   | ccgc当地      | ggtccc当地   | aagggtgaaa | 2220 |
| caataggagt  | gataggcaca  | ccgggttgc  | gcaagg      | tgc当地      | tattatca   | 2280 |
| cggcagcaga  | tcttgc当地    | agcggaa    | aagaaaattt  | tcgc当地     | gaggcc     | 2340 |
| tgctaagact  | gagggttgc   | cagattacgt | cgaagac     | agatcgtt   | atgctca    | 2400 |
| gatgccacaa  | acccgtgaa   | gtctgtc    | ttgacgaa    | gttc当地     | cacgcagg   | 2460 |
| cactacttgc  | cttgattgt   | atcgc当地    | ccgc当地      | ggtagtacta | tgccgag    | 2520 |
| cacatgtatc  | cgatgttgc   | aatatgtc   | aactaaagg   | acatcttca  | caccctgaaa | 2580 |
| aagacatatg  | caccaagaca  | ttctaca    | atatctcc    | gggttgc    | cagccag    | 2640 |
| caagtatgt   | atcgacact   | cattacgt   | gaaagatgaa  | aaccac     | ccgtgca    | 2700 |
| agAACATGAA  | atcgatatt   | acaggcc    | caagccgaa   | gocagg     | atctcc     | 2760 |
| catgtttccg  | cggttgggtt  | aaatcgact  | tcccg       | ggat       | aatatg     | 2820 |
| cagccgcggc  | ctcacaaggg  | ctaacc     | aaagg       | tgccgttgc  | caaaaagg   | 2880 |
| atgaaaacc   | actgtacgc   | atcacat    | agcatgt     | cggttgc    | acccgc     | 2940 |
| aggacaggct  | actgtggaa   | accttgc    | gccc当地      | gccc当地     | ctca       | 3000 |
| taaccaaagg  | aaatctcg    | gtactatag  | aggactgg    | agctgaa    | aagggata   | 3060 |
| ttgctcaat   | aaacagcccc  | actccctgt  | ccaaatcc    | cagctgca   | accaacgtt  | 3120 |
| gctgggc当地   | agcattggaa  | ccgatact   | ccacgg      | tatcgat    | accgg      | 3180 |
| atgtggcgc   | actgttccca  | cagttgc    | atgaca      | accatcg    | atttac     | 3240 |
| tagacgtat   | ttgc当地      | ttttcg     | ggacttgc    | aagg       | actt       | 3300 |
| agagcatccc  | actaacgt    | catccgc    | attcagg     | ggccgtact  | cattgg     | 3360 |
| acagcccagg  | acccgc当地    | tatgg      | atcacc      | tgccgc当地   | ctctcc     | 3420 |
| gatttcccg   | gttc当地      | gctgg      | gac         | actat      | tgat       | 3480 |
| ccagaggat   | ctctgc      | cataacc    | tcccg       | ccgcaat    | cctc       | 3540 |
| tagtcccg    | gtacaagg    | aagaacc    | gccc当地      | aaaatctt   | aacc       | 3600 |
| aacaccact   | actacttgc   | gtatc      | agg         | atgttgc    | aaat       | 3660 |
| aatggatcgc  | cccgatgg    | ataggc     | cgat        | ccgaa      | acta       | 3720 |
| ttccgc当地    | ggcgc当地     | gac        | tcata       | ggactaa    | tac        | 3780 |
| accacttca   | gaoatgc     | gaccat     | cgac        | ccat       | ttcg       | 3840 |
| tgaattgc    | taacccagg   | ggcacc     | ctcg        | ttgt       | gact       | 3900 |
| acagtggaa   | ctgtgc当地    | gcttgc     | gaaat       | atgg       | ttgt       | 3960 |
| cgatgttgc   | ctcaaggat   | acagaaat   | actgtt      | ccgaa      | acta       | 4020 |
| gtacacgc当地  | attcacc     | cacat      | ttgc        | ttcg       | tat        | 4080 |
| caagagatgg  | atgtggagcc  | gccc当地     | accgc       | ccat       | ggat       | 4140 |
| gtcaagg     | agcactgttgc | aacgc      | atccg       | ttgg       | acttgc     | 4200 |
| gccc当地      | ctataacc    | ttccgc     | gacca       | tttacc     | ccgc       | 4260 |
| ccgcaaggat  | gactgttgc   | ctagg      | aaat        | ggat       | ccgt       | 4320 |
| ggaagcacc   | agaaggagaa  | gcctt      | tgct        | ccat       | gcagg      | 4380 |
| acttagtta   | tgaacata    | atca       | act         | tcgat      | tat        | 4440 |
| acgc当地      | aaaaggcc    | cttgc      | gat         | cact       | atct       | 4500 |
| gaactgacgc  | ggacgt      | atcttgc    | ttgt        | gaa        | atgc       | 4560 |
| ccgcacttca  | acttaagg    | tctgt      | gat         | gttgc      | acttgc     | 4620 |
| atgagatgt   | atggatcc    | ccagac     | gttgc       | ggagg      | aaaggat    | 4680 |
| aaaagg      | atgttatcg   | tacttgc    | gac         | ccat       | aaat       | 4740 |
| ttggcggat   | aaagg       | ttcc       | ggat        | ggaa       | atgttgc    | 4800 |
| acatatttgg  | tgagaccat   | gaag       | at          | ttgttgc    | ccatacc    | 4860 |
| cgtctagccc  | gccccaa     | ttgc当地     | ggat        | ggat       | accat      | 4920 |
| ttccac      | actaa       | aaatgt     | acttgc      | tttgc      | ccagg      | 4980 |
| ctaagg      | atata       | gttgc      | acttgc      | atgttgc    | ccat       | 5040 |
| cgcacacttcc | ccgc当地      | atgttgc    | acttgc      | atgttgc    | ccat       | 5100 |
| ctctcttgc   | acaggcc     | gagg       | ccat        | atgttgc    | ccat       | 5160 |
| ctgata      | actc        | ttgttgc    | acttgc      | atgttgc    | ccat       | 5220 |
| gtctactt    | ttcgat      | agcggat    | tact        | gttgc      | acttgc     | 5280 |
| cgtcagg     | acttgc      | gaggat     | gttgc       | ggat       | atgttgc    | 5340 |
| atgc当地      | acttgc      | ttgttgc    | atgttgc     | ggat       | atgttgc    | 5400 |
| cagcggcaag  | aaaaggcc    | acttgc     | ggat        | atgttgc    | ccat       | 5460 |
| ttttgttgg   | ggat        | ttcc       | ggat        | atgttgc    | ccat       | 5520 |
| cagcggat    | acc         | ttcc       | ggat        | atgttgc    | ccat       | 5580 |
| ccgacggaga  | gat         | ttcc       | ggat        | atgttgc    | ccat       | 5640 |
| gtcata      | ttat        | gttgc      | ccat        | atgttgc    | ccat       | 5700 |

---

-continued

---

|             |            |             |             |            |            |      |
|-------------|------------|-------------|-------------|------------|------------|------|
| ctctacgcaa  | gcagagacgt | agacgcaggag | gcaggaggac  | tgaatactga | ctaaccgggg | 5760 |
| taggtggta   | catatttcg  | acggacacag  | gccctggca   | cttgc当地    | aagtccgttc | 5820 |
| tgcagaacca  | gcttacagaa | ccgaccttgg  | agcgcaatgt  | cctgaaaga  | attcatgcc  | 5880 |
| cgggtctcga  | cacgtcgaaa | gaggaacaac  | tcaaactca   | gtaccagatg | atgcccac   | 5940 |
| aagccaacaa  | aagttagtac | cagtctcgt   | aagttagaaaa | ttagaaagcc | ataaccactg | 6000 |
| agcgactact  | gtcaggacta | cgactgtat   | actctggcc   | agatcgac   | gaatgtata  | 6060 |
| agatcaccta  | tccgaaacca | ttgtactca   | gtacgtacc   | ggcgaactac | tccgatcc   | 6120 |
| agttegttgt  | agctgtctgt | aacaactatc  | tgc当地       | ctatccgaca | gtacatctt  | 6180 |
| atcgattac   | tgacgagatc | gatgttact   | tggatatgg   | agacgggaca | gtcge      | 6240 |
| tggacactgc  | aaccttctgc | cccgactat   | ttagaaggta  | cccgaaaaaa | catgatata  | 6300 |
| gagccccggaa | tatccgact  | gcccgtccat  | cagcgatcga  | gaacacgct  | caaaatgtgc | 6360 |
| teattgcgc   | aactaaaaga | aatttgc当地   | tcacgc当地    | gcgtgactg  | ccaacactgg | 6420 |
| acttcagegc  | attcaactgc | gaatgttctt  | gaaaatatgc  | atgtatgac  | gagtatttgg | 6480 |
| aggagttcg   | tccgaaacca | attaggatta  | ccactgagg   | tgtcacc    | tatgtatgta | 6540 |
| gactgaaagg  | ccctaaggcc | gccc当地      | tttgc当地     | gtataattt  | gtcccat    | 6600 |
| aagaagtgc   | tatggataga | ttcgtcatgg  | atcgaaaag   | agacgtgaa  | gttacaccag | 6660 |
| gcacgaaaca  | cacagaaag  | agacggaa    | tacaatgtat  | acaacccca  | gaacccctgg | 6720 |
| cgactgat    | ttatcgccgg | attcaccggg  | aattatgtcg  | taggttac   | ggcgtctgc  | 6780 |
| ttccaaacat  | tcacacgctt | tttgc当地     | cgccggagg   | tttgc当地    | atcatagc   | 6840 |
| aacacttca   | gcaaggcgac | ccggta      | agacggat    | cgcat      | gacaaaagcc | 6900 |
| aagacgacgc  | tatggcttga | accggctg    | tgatcttgg   | ggacctgg   | gtggatca   | 6960 |
| cactactgc   | cttgatcg   | tgcccttgg   | gagaatatt   | atccaccat  | ctacctac   | 7020 |
| gtactegttt  | taaattcggg | gcatgat     | atccggat    | gttcc      | tacca      | 7080 |
| acactgtttt  | gaatgtcg   | atcgcc      | gactactag   | ggagcgg    | aaaacgt    | 7140 |
| gatgtcg     | cgttcat    | gacgaca     | tcatacat    | agtagtat   | gacaaagaaa | 7200 |
| tggctgt     | tgccg      | accgg       | tggat       | gatcat     | gcagtc     | 7260 |
| gtgagagacc  | accttactt  | tgccgg      | ttatcttgc   | agatccgg   | acttcc     | 7320 |
| cgtgcccgt   | ggccgatccc | ctgaaaagg   | tgttta      | ggtaaac    | ctcc       | 7380 |
| acgacgacg   | agacgaa    | agaagac     | ctctgt      | tga        | aaacaaag   | 7440 |
| gagtaggtat  | aaacggact  | ttagoat     | ccgtgac     | ccgtgac    | gtagaca    | 7500 |
| ttacac      | ctact      | tttgc当地     | tttgc当地     | ccaa       | agagca     | 7560 |
| tcagaggg    | aataa      | accat       | tc          | tc当地       | tacattt    | 7620 |
| ctgactata   | ctaca      | acc         | gtg         | cttgc      | gacg       | 7680 |
| cccaatgata  | tcacac     | acc         | acttcc      | gatgtt     | cataat     | 7740 |
| caggctgtt   | cattat     | ccgtt       | ccgtt       | ccgg       | acttcc     | 7800 |
| gtacttccg   | gga        | ccgg        | tg          | cc         | acttcc     | 7860 |
| aaccatttat  | ctag       | ccgg        | actt        | ttt        | tttgc当地    | 7920 |
| tgc         | cc         | ccgg        | actt        | ttt        | tttgc当地    | 7980 |
| ttaacat     | ttt        | aaaa        | aaaa        | aaaa       | aaaa       | 8040 |
| aaaaaaaaaa  | aaaa       | aaaa        | aaaa        | aaaa       | aaaa       | 8100 |
| aaaaaaaaaa  | aaaa       | aaaa        | aaaa        | aaaa       | aaaa       | 8120 |

---

SEQ ID NO: 105 moltype = RNA length = 8875  
 FEATURE Location/Qualifiers  
 misc\_feature 1..8875  
 note = Synthetic polynucleotide  
 source 1..8875  
 mol\_type = other RNA  
 organism = synthetic construct

SEQUENCE: 105

|            |            |            |            |            |           |      |
|------------|------------|------------|------------|------------|-----------|------|
| atggggccgc | catgagagaa | gcccagacca | attacctacc | aaaaatggag | aaagtccac | 60   |
| ttgacatcga | ggaagacagc | ccatccctca | gagcttgc   | gccc当地     | ccgc当地    | 120  |
| aggttaga   | ac         | aa         | actgata    | accatgt    | tgcc      | 180  |
| tggcttcaaa | actgtat    | acggagg    | tttgc当地    | tttgc当地    | tttgc当地   | 240  |
| gtgcgc     | ccgc       | ccgc       | tatttgc    | aca        | actgtat   | 300  |
| gtgcggaa   | tccg       | gac        | tttgc当地    | tttgc当地    | tttgc当地   | 360  |
| aaataactga | taagg      | aaatt      | gaca       | aaaa       | tgaagg    | 420  |
| ctgac      | cttgc      | actgt      | atgt       | cccg       | ccgt      | 480  |
| aagt       | ctgt       | actgt      | atgt       | cccg       | ccgt      | 540  |
| ccaa       | actt       | ccgt       | atgt       | cccg       | ccgt      | 600  |
| ccgc       | ccgc       | ccgc       | ccgt       | cccg       | ccgt      | 660  |
| ccgc       | ccgc       | ccgc       | ccgt       | cccg       | ccgt      | 720  |
| ccat       | ccat       | ccat       | ccgt       | cccg       | ccgt      | 780  |
| tgagg      | ggcc       | ggcc       | ccat       | cccg       | ccgt      | 840  |
| tgagg      | ggcc       | ggcc       | ccat       | cccg       | ccgt      | 900  |
| tgg        | ggcc       | ggcc       | ccat       | cccg       | ccgt      | 960  |
| caat       | ccat       | ccat       | ccgt       | cccg       | ccgt      | 1020 |
| tgagg      | ggcc       | ggcc       | ccat       | cccg       | ccgt      | 1080 |
| tgg        | ggcc       | ggcc       | ccat       | cccg       | ccgt      | 1140 |
| tgg        | ggcc       | ggcc       | ccat       | cccg       | ccgt      | 1200 |
| tgg        | ggcc       | ggcc       | ccat       | cccg       | ccgt      | 1260 |
| ggcc       | ggcc       | ggcc       | ccat       | cccg       | ccgt      | 1320 |
| acaag      | atcac      | cacat      | ccat       | cccg       | ccgt      | 1380 |
| acttcc     | acac       | acac       | ccat       | cccg       | ccgt      | 1440 |
| cccg       | cccg       | cccg       | ccat       | cccg       | ccgt      | 1500 |
| acgt       | ccat       | ccat       | ccgt       | cccg       | ccgt      | 1560 |
| tgagg      | ccgc       | ccgc       | ccat       | cccg       | ccgt      | 1620 |
| tgg        | ccgc       | ccgc       | ccat       | cccg       | ccgt      | 1680 |
| aggt       | ccgc       | ccgc       | ccat       | cccg       | ccgt      | 1740 |

-continued

---

|              |             |             |             |              |             |      |
|--------------|-------------|-------------|-------------|--------------|-------------|------|
| cgcgtgctgaa  | gtcccgagaag | ctgagactgca | tccaccact   | ggccgagcag   | gtgatcgta   | 1800 |
| tcaccccacag  | cgccaggaaag | ggcaggtacg  | ccgtggagcc  | ctaccacggc   | aagggtgtcg  | 1860 |
| tgccccgaggg  | ccacgcacatc | cccggtcagg  | acttccaggc  | cctgagcgag   | agcgcaccca  | 1920 |
| tcgtgtacaa   | cgagaggggag | ttcgtgaaca  | ggtacactgca | ccatactgccc  | acccacggcg  | 1980 |
| gagccctgaa   | caccgcacag  | gaatactaca  | agaccgtgaa  | gcccagcag    | cacgacggcg  | 2040 |
| agtacctgt    | cgacatcgac  | aggaagcagt  | gctgtgaagaa | agagctggtg   | acccggctgg  | 2100 |
| gactgacccg   | cgagctggtg  | gaccacccct  | tccacgagt   | cgccctacgg   | agcctgagga  | 2160 |
| ccagaccggc   | cgctccctac  | caggtgccc   | ccatcgccgt  | gtacggcggt   | cccgccagcg  | 2220 |
| gaaaagagcgg  | catcatcaag  | agcgcgtgca  | ccaagaaaaa  | cctgggtggc   | agcgcacaag  | 2280 |
| aagagaactg   | cgccggatc   | atcaggacg   | tgaagaagat  | gaaaggccgt   | gacgtgaacg  | 2340 |
| cgccgcacccgt | ggacacgtgt  | ctgtgtac    | gctgtcaacg  | ccccctggag   | accctgtaca  | 2400 |
| tcgacgaggc   | cttcgttgc   | cacgcggca   | ccctgaggggc | cctgtatcgcc  | atcatcgac   | 2460 |
| ccaaagaacg   | cgtgtgtgc   | ggccacccca  | agcagtgcgg  | cttcttcaac   | atgtatgtgc  | 2520 |
| tgaagggtgca  | ttcaaccac   | gagatctgca  | ccccagggtt  | ccaaagaaag   | atcagcaaggc | 2580 |
| ggtgtcaccaa  | gagcgtgacc  | agcgtgtca   | gcacccctgt  | stacgtacaag  | aaaatgagga  | 2640 |
| ccacccaaccc  | caaggagacc  | aaaatctgt   | tegacaccac  | aggcagcacc   | aagcccaagc  | 2700 |
| aggacgaccc   | gtatctgacc  | tgcttcaggc  | gctgggtgaa  | gcagctgcag   | atcgactaca  | 2760 |
| agggcaacag   | catgtgacc   | gcccgtc     | gcccggcc    | gaccggggct   | gaccggaaag  | 2820 |
| cctgtaggta   | caaggtgac   | gaaagggccac | tgtacgttcc  | caccgcgg     | cacgtgaaac  | 2880 |
| tgctgtgtac   | caggaccgg   | gacaggatcg  | tgtggaaagac | cctggccggc   | gacccttgg   | 2940 |
| taaaggaccc   | gaccgcca    | taccccgcc   | acttcacccg  | caccatgaa    | gagtggcagg  | 3000 |
| ccgagcgcac   | cgcccatcg   | aggcacaatc  | tggagaggcc  | cgcccccacc   | gacgtgttcc  | 3060 |
| agaacaaggc   | caacgtgtc   | tggcccaagg  | cccttggc    | cgtgtgtaa    | accggccggca | 3120 |
| tcgacatgac   | cacagagcag  | tggacacccg  | tggactactt  | cgagacggac   | aaggccacaca | 3180 |
| gcgcggagat   | cgtgtgtaa   | cagctgtcg   | tgagggttct  | cgccctggac   | ctggacacgc  | 3240 |
| gectgttcag   | cgccccccac  | gtgcactgt   | gcatcaggaa  | caaccactgg   | gaaacacagcc | 3300 |
| cacggccaaa   | catgtacggc  | ctgaacaaagg | atgtgggtcg  | cgacgtgtac   | aggcggttacc | 3360 |
| cacagctgcc   | caggccgtg   | gccaccggca  | gggtgtacga  | catgaacacc   | ggcaccctga  | 3420 |
| gaaactacgaa  | ccccggatc   | aacctggtd   | cctgtgaacag | gcccgtccc    | cacggccctgg | 3480 |
| tgtgtcacca   | caacgagcac  | ccacagagc   | acttcagatc  | cttcgtgagc   | aagctgaaag  | 3540 |
| gcaggaccgt   | gtgtgtcg    | ggcggagaag  | tgaggctgtc  | cgccggatgt   | gtggactggc  | 3600 |
| tgagcgtacag  | gccccggg    | accccccggg  | ccaggctgtg  | cctcggatc    | ccccggccgt  | 3660 |
| tgccccaga    | cgacatc     | ttcgtgtac   | tcaggacccc  | atacaagtac   | caccattacc  | 3720 |
| agcagtgtca   | ggaccacggc  | atcaactgt   | gcatgtgtac  | caagaaggcc   | tgccgtgacc  | 3780 |
| tgaacccccc   | aggccacccgt | gtgagcatcg  | gtactcggtt  | cgccggacagg  | ggccgggaga  | 3840 |
| gcatcatgg    | cgccatcgcc  | aggctgtca   | agttcagac   | gggtgtcaaa   | cccaagagca  | 3900 |
| gcctggagga   | aaccgggtg   | ctgttcgtgt  | tcatcggtt   | cgaccggaa    | gccaggaccc  | 3960 |
| acaaccccta   | caagctgtgc  | aggcaccctg  | caaacatctt  | caccggcagc   | agggtgcacg  | 4020 |
| aggccggctg   | cgccccccac  | taccacgtt   | tcagggtggc  | tatccggccacc | ggcaccggagg | 4080 |
| gcgtgtatcat  | caacgtgtcc  | aacagcaagg  | gccagccccc  | aggcggtgt    | tgccggccccc | 4140 |
| tgtacaagaa   | gttccccgg   | agcttcgacc  | tgcagcccat  | cgaggtgggc   | aaggccaggc  | 4200 |
| tgtgtgggg    | cgcccgat    | cacatcatc   | acggccgtgg  | cccccaactt   | aacaagggtt  | 4260 |
| gcggggatgg   | aggcgacaa   | cgatggccg   | aaagcttacga | gagcatcgcc   | aaagatcgta  | 4320 |
| acgacaataa   | ctacaagacg  | gttgcacatc  | cactgtctac  | caccggatc    | ttcaqcggtca | 4380 |
| acaaggacag   | gtgtacccag  | agcctgtacc  | acttgcgtac  | cgccctggac   | accaccgtat  | 4440 |
| ccgacgtggc   | cattactgt   | aggggacaaga | agtggggatg  | gaccctgtgg   | gaggccgtgg  | 4500 |
| ccaggccgg    | ggccgtggaa  | gagatctgt   | tcagcgacga  | cttcagctgt   | accggcccg   | 4560 |
| acggccgacgt  | gggtgggtg   | caccccaaga  | gtccctggc   | cgccggggag   | ggctacagac  | 4620 |
| ccagcgacgg   | caagaccc    | agcttacccgt | agggccacaa  | gttccaccag   | ggcgctaagg  | 4680 |
| acatcgccga   | gatcaacgt   | atgtggccgg  | tgccacccgg  | ggccaaacgg   | cagggtgtca  | 4740 |
| tgtatccat    | ggggcgagac  | atgtccagca  | tcaggacaa   | gtgccccgt    | gaggaaaggc  | 4800 |
| aggccacac    | accaccac    | accctggcc   | gtctgtcat   | ccacgttat    | acacccgaga  | 4860 |
| gggtgcacgc   | gtgtggcc    | agcaggcccg  | agcagatc    | ctgtgtc      | tccttccac   | 4920 |
| tgccccaga    | cagggttac   | ggctgtgtac  | agatccatgt  | cagccaccc    | atccctgtca  | 4980 |
| gcccaccaat   | ggccggctac  | atccacccca  | ggaagtaccc  | ggtggagacc   | ccacccgtgg  | 5040 |
| acgagacacc   | cgagccaa    | gcccggaccc  | agagacccca  | gggcacaccc   | gagcggccac  | 5100 |
| ccctgtatcat  | cgaggacgg   | acaaggaccc  | ggaccccaaga | gcccacatt    | atcgaggaag  | 5160 |
| aggaaggagg   | cagcatcg    | ctgtgtcg    | acggccccac  | ccaccagg     | ctgcagggtt  | 5220 |
| aggcccgacat  | ccacggccca  | cccaggctgt  | ceagctccat  | ctggaggatc   | ccacacggca  | 5280 |
| gcgacttgc    | cggtggac    | ctgacatcc   | tggacaccc   | ggaggccggcc  | agcgtgtac   | 5340 |
| ccggcgeccac  | cagcgccgg   | accacacgt   | acttgcacca  | gagcatggag   | ttcttggcca  | 5400 |
| ggcccggtcc   | actccatcc   | acccgttca   | ggacccaccc  | ccacccgt     | cccaggacca  | 5460 |
| ggaccccaag   | ctgggttccc  | agcaggccct  | gcagcaggac  | caggctgg     | agcaccaccc  | 5520 |
| ccggcgtgaa   | cagggtgtac  | acccgggg    | aaactggggc  | cctgacaccc   | agcaggaccc  | 5580 |
| ccagcaggcc   | cgtgtgtac   | actagtctgg  | tgtccaaacc  | acccggcggt   | aacagggttga | 5640 |
| tcaccaggga   | ggaattcgag  | gccttcgtt   | cccagcaaca  | gagacgg      | tgacggccgg  | 5700 |
| cttacatctt   | cagcaacgt   | accggccagg  | gacacccgt   | gcaaaaagac   | gtgaggccaga | 5760 |
| ccgtgtgtac   | cgagggtgtt  | ctggaggg    | ccgagcttgc  | aatcgtatc    | gccccccaggc | 5820 |
| tggaccagga   | gaaggaggaa  | ctgtgtac    | agaaaactgt  | gctgtacccc   | acccggacca  | 5880 |
| acaggagacg   | gttaccagac  | aggaagggtt  | agaacatgt   | ggccatcacc   | ggcaggccgg  | 5940 |
| tcctgtacgg   | cctgggacac  | tacctgtac   | ccgaggggca  | ggtggagtgc   | tacaggaccc  | 6000 |
| tgcacccatgt  | gacactgtac  | agctccatgt  | tgaacaggcc  | cttctccatc   | cccaagggtt  | 6060 |
| ccgtggaggc   | ctgcaacgt   | atgtgtac    | agaaacttcc  | caccgtggcc   | agctactgtca | 6120 |
| tcatccatgt   | gtacgtac    | tacccgtt    | gatgtgtac   | ccacacggcc   | ttccctggaca | 6180 |
| ccggccatgt   | ctggcccgcc  | aaactgt     | gcttccccc   | gaaacacac    | tacctgttgc  | 6240 |
| ccaccatctg   | gagcgccgt   | cccaggccca  | tccagaacac  | cctgtcgac    | gtgctggccg  | 6300 |
| ctggccacaa   | gagggactgc  | aacgtgaccc  | agatgaggaa  | gctgccccgt   | ctggacacgc  | 6360 |
| ctgccttcaa   | ctgtggatgtc | ttcaagaaat  | acgcctgtca  | caacgtgtac   | tggggaccc   | 6420 |
| tcaaggagaa   | ccccatcagg  | ctgaccgaa   | agaacgtgtt  | gaactacatc   | accaagctga  | 6480 |

-continued

|              |            |             |             |             |             |      |
|--------------|------------|-------------|-------------|-------------|-------------|------|
| aggggcccaa   | ggccgtgcc  | ctgttcgcta  | agacccacaa  | cctgaacatg  | ctgcaggaca  | 6540 |
| tcccaatgga   | caggttcgta | atggacctga  | agagggact   | gaaggtgaca  | cccgccacca  | 6600 |
| agcacacccg   | ggagagggcc | aaggtcagg   | tgatccaggc  | cgctgaccca  | ctggccaccc  | 6660 |
| cctacctgt    | ggccatccac | agggagctgg  | tgaggcggct  | gaacgcggc   | ctgtgcucca  | 6720 |
| acatccacac   | ctgttcgac  | atgacgcgg   | aggacttgc   | cgccatcatc  | ggcggact    | 6780 |
| tccagccgg    | cgactcgctg | ctggagaccg  | acatcgccag  | cttgcacaag  | agcgaggatg  | 6840 |
| acgctatggc   | ctgtacgct  | ctgtatcc    | ttggaggact  | gggcgtggac  | ggcgactgc   | 6900 |
| tcaccctgt    | cgaggctgcc | ttcggcgaga  | tcagctccat  | ccacctgccc  | accaagacca  | 6960 |
| agtcaagg     | gggcgttgc  | atgaaaagcg  | gaatgttcc   | gaccctgttc  | gtgaacaccc  | 7020 |
| tgatecaat    | tgtatcgcc  | agcagggttgc | tgccggagat  | gtgcacggc   | agccccctgcg | 7080 |
| ctgccttcat   | cgggcgtac  | aacatctgta  | agggcgtgaa  | aagcacaag   | ctgtggccg   | 7140 |
| acagggtgc    | oacatggctg | aacatggagg  | tgaagatcat  | cgacgcgtg   | gtgggggaga  | 7200 |
| aggcccccata  | cttcgtccgg | ggatccatcc  | ttgtgcacag  | cggtggggcc  | accgctgc    | 7260 |
| gggtggccg    | cccccttgc  | aggcttca    | agttggggcc  | gtcaactggc  | gtgcacgtg   | 7320 |
| agcacacgca   | tgacaggccg | agggccctgc  | acgaggaaag  | caccagggtt  | aacagggttg  | 7380 |
| gcacccctgt   | cccaagaccc | accagggt    | gtgcaacacc  | acccagaaga  | ccacgcacgg  | 7740 |
| cagctatgt    | ctggccgtcc | ccacccggc   | gtccgtca    | gagctttccc  | tacctgagg   | 7500 |
| ggggcccttat  | aactcttac  | ggctaaac    | aatggactac  | gacatagtt   | agtccgc     | 7560 |
| ggccgcacc    | atgagatgt  | cagecccttag | aaccttact   | cttctgttt   | ggggactgt   | 7620 |
| tgetctgaca   | gagacatggg | ctggatctct  | gagcggagg   | accggccagg  | gcctgtgc    | 7680 |
| cgccgcctgt   | cccaagaccc | accagggt    | gtgcaacacc  | acccagaaga  | ccacgcacgg  | 7740 |
| cagctatgt    | ctggccgtcc | ccacccggc   | gtccgtca    | gagctttccc  | tacctgagg   | 7800 |
| ttgcatcagc   | accaccatcc | tgaatctgac  | caccgtact   | tcgtgttgc   | ttggactgt   | 7860 |
| geccagggt    | acctaccaca | gccccagta   | cgccattacc  | cagttcgaga  | ggaggggccaa | 7920 |
| gtacaagagg   | gagccgtgt  | gcctgaccct  | ggccctgttgc | ctggggggcc  | tgacaatggg  | 7980 |
| cgccatcgcc   | gcccgggtgg | gcacccggc   | caccggcc    | gtggccaccc  | agcgttcca   | 8040 |
| gcacgtcg     | gcccgtatgc | acggacact   | gaagggttgc  | gagaagtcca  | tcaccaac    | 8100 |
| ggagaagtcc   | ctgaccagcc | tgagcgggt   | gtgtgtcg    | aacaggagg   | gtctggac    | 8160 |
| gtctgttctg   | aaggaggccg | gcctgtggc   | cgccttgc    | gaggagtgt   | gcctgtac    | 8220 |
| cgaccacacc   | ggccctgtgt | tcgtggcc    | tgcgtgttgc  | ctggccgttgc | tcgcgttgtt  | 8280 |
| gggtgatttgc  | gtgtgttgc  | cagtgttat   | tgccagaaga  | aagtatcc    | cgccggaaagg | 8340 |
| aggctctac    | tctcaagctg | cttctgtac   | agtgcctaga  | gtctttatgt  | gtttatctca  | 8400 |
| gtctgtaaact  | cgatgtatgt | acgtgc      | ttgttataat  | atgcactaa   | gtgttataat  | 8460 |
| tgacacacc    | tcaatgtac  | gccccaaat   | ttacagccgc  | gggtgtcaaaa | accgcgttga  | 8520 |
| ctgtgttataat | atccctgtgt | ggaggatcg   | cgtgttataat | tataatttgc  | ttgggtgt    | 8580 |
| ctactattgt   | ggccatgtac | gtgtgttgc   | accagaaaca  | taatttgcata | cagcagcaat  | 8640 |
| tggcaagctg   | cttacataga | actcgccgc   | attggcatgc  | cgccctttaaa | tttttattt   | 8700 |
| attttttttt   | ttcttttcc  | aatcgattt   | ttgttataat  | atttcaaaaa  | aaaaaaaaaa  | 8760 |
| aaaaaaaaaa   | tctagaaaaa | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | 8820 |
| aaaaaaaaaa   | aaaaaaaaaa | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaa      | 8875 |

SEQ ID NO: 106                  moltype = RNA    length = 8416  
 FEATURE                  Location/Qualifiers  
 misc\_feature                  1..8416  
 note = Synthetic polynucleotide  
 source                  1..8416  
 mol\_type = other RNA  
 organism = synthetic construct

SEQUENCE: 106

|              |              |              |              |             |            |      |
|--------------|--------------|--------------|--------------|-------------|------------|------|
| atgggcggcg   | catgagagaa   | gcccagacca   | attacctacc   | caaaatggag  | aaagttcacg | 60   |
| ttgacatgga   | ggaagacagc   | ccattctca    | gagtttgc     | gcccgggttc  | ccgcgttgc  | 120  |
| aggtagaa     | caagcagg     | actgataatg   | accatgttac   | tgccagacgc  | ttttccatc  | 180  |
| tggcttcaaa   | actgtatcgaa  | acggagggttgc | accatccgc    | cacgcatttgc | gacattggaa | 240  |
| gtgcgcggc    | ccgcagaatg   | tatttcaatgc  | acaatgtatca  | ttgtatccgt  | ccgtatgtat | 300  |
| gtgcggaga    | tccggacaga   | ttgtatcaatgt | atgcactaa    | gtgttataat  | aaatgttgc  | 360  |
| aaataactgt   | ttaaggatgt   | gacaaaaaaa   | tgaaaggact   | ggccgcggcc  | atgacgcacg | 420  |
| ctgacctgtt   | aactgtatgt   | atgtgcctcc   | acgacgcacg   | gtcggttgc   | tacgaagg   | 480  |
| aagtgcgtgt   | tttccatgttgc | gtatgcgttgc  | tcgacggcc    | caccagcttgc | taccaccagg | 540  |
| ccaaacaagg   | ctgtgttgc    | gttgcgttgc   | tcgtgttgc    | caccaccc    | ttcatgttca | 600  |
| agaaacccgtc  | ccgcgcctac   | cccaactaca   | gacccaacttgc | ggccgcacgg  | accgtgttgc | 660  |
| ccgcgcaggaa  | cattggccgt   | tcgcacgcac   | acgtgtatgg   | gaggagccgg  | agaggatcg  | 720  |
| gcacccctgt   | gttgcgttgc   | gttgcgttgc   | gttgcgttgc   | gttgcgttgc  | gttgcgttgc | 780  |
| ccatctacca   | cgagaagg     | gaccgttgc    | ggaggttgc    | ccgttgc     | gtgttccacc | 840  |
| tgaggggccaa  | cgagaactac   | acctgtgttgc  | gttgcgttgc   | gttgcgttgc  | gttgcgttgc | 900  |
| tgggtgttgc   | gttgcgttgc   | gttgcgttgc   | gttgcgttgc   | gttgcgttgc  | gttgcgttgc | 960  |
| caatgtatgt   | ggagggttgc   | ctgtgttgc    | gttgcgttgc   | gttgcgttgc  | gttgcgttgc | 1020 |
| tgtgttcc     | ctgtgttgc    | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1080 |
| ttggccatcc   | ctgtgttgc    | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1140 |
| tctgtgttca   | ccggcgttgc   | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1200 |
| tgggtgttgc   | ccggcgttgc   | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1260 |
| ggcccccgttgc | ctgtgttgc    | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1320 |
| acaagatcac   | cacgtatgt    | aagggttgc    | acacccacac   | catcatca    | gttgcgttgc | 1380 |
| acttccatcg   | cttcgttgc    | cccgaggatc   | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1440 |
| cccggtatcg   | gttgcgttgc   | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1500 |
| acgtgcgttgc  | ggccaaatgt   | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1560 |
| tgggtgttgc   | ctgtgttgc    | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1620 |
| tggacccatcg  | gttgcgttgc   | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1680 |
| aggtgttgc    | ctacgtatgt   | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1740 |
| ccgtgttgc    | gttgcgttgc   | tcgtgttgc    | tcgtgttgc    | tcgtgttgc   | tcgtgttgc  | 1800 |

-continued

---

|               |              |             |             |             |              |      |
|---------------|--------------|-------------|-------------|-------------|--------------|------|
| tcacccacag    | cggcaggaag   | ggcaggtacg  | ccgtggagcc  | ctaccacggc  | aagggtggcg   | 1860 |
| tgcccgggg     | ccacgcccc    | cccggtcagg  | acttccaggc  | cctgagcgag  | agcgccacca   | 1920 |
| tctgttacaa    | cgagagggag   | ttctgtaca   | ggtacctgc   | ccatatccgc  | accacggcg    | 1980 |
| gagccctgaa    | caccgcgag    | gaatactaca  | agaccgtgaa  | gcccgagcgag | cacgcacggcg  | 2040 |
| agtacctgta    | cgacatcgac   | aggaacgtgt  | gctgtgaagaa | agagctgg    | accggccctgg  | 2100 |
| gactgacccg    | cgacgtgggt   | gaccaccc    | tccacgagt   | cgccatcgag  | agcctgagga   | 2160 |
| ccagaccggc    | cgctccctac   | cagggtccca  | ccatcgccgt  | gtacggcg    | cccgccagcg   | 2220 |
| gaaagagcgg    | catcatcaag   | agcgcgctga  | ccaagaaaaga | cctgggtggc  | agcgccaaaga  | 2280 |
| aaagagaactg   | cgccgcgatc   | atcaggggc   | tgaagaagat  | gaaaggocgt  | gacgtgaacg   | 2340 |
| cgcgccacgt    | ggacagcggt   | ctgtgtac    | gctgcgaagg  | cccctgggg   | accctgttaca  | 2400 |
| tgcacgaggc    | cttcgttgc    | cacgcggca   | cctgtggggc  | ctgtatccgc  | atcatcgac    | 2460 |
| ccaagaaaagc   | ctgtgtgtgc   | ggcgcaccca  | agcagtgcgg  | tttcttcaac  | atgatgtgcc   | 2520 |
| tgaagggtgca   | cttcaacacc   | gagatctgc   | cccagggtgt  | ccacaagagc  | atcaqcgac    | 2580 |
| gtgtgcaccaa   | gagggtgacc   | aggctgtgt   | gtacggcgat  | aaaatgagga  | 2640         |      |
| ccaccaaccc    | caaggagacc   | aaaaatcgta  | tcgcacacc   | aggcagcacc  | aagcccaagc   | 2700 |
| aggacgaccc    | gatcctgacc   | tgcttcagg   | gctgggtgaa  | gcagctgcag  | atcgactaca   | 2760 |
| aggggcaacga   | gatcatgacc   | gcccgtgc    | gccaggccct  | gaccaggaa   | ggcgtgtacg   | 2820 |
| ccgtgaggac    | caagggtgac   | gagaacccac  | tgtacgtctc  | caccagcgag  | cacgtgaacg   | 2880 |
| tgtctgtgac    | caggaccgg    | gacaggatcg  | tgtggaaac   | cctggccggc  | gaccctgg     | 2940 |
| tcaagaccc     | gaccgcacag   | tacccggca   | acttcaccgc  | caccatcgaa  | gagttggcagg  | 3000 |
| ccgagcgcac    | cgcccatcatc  | aggcacatc   | tggagaggcc  | cgacccccc   | gacgtgttcc   | 3060 |
| agaacaacgc    | caacgtgtc    | tggggcaagg  | ccttgggtgc  | cgtgtgaag   | accgcggcga   | 3120 |
| tgcacatgac    | cacagacgac   | tggaaacacc  | tggactactt  | cgagacccac  | aaggccacaa   | 3180 |
| gcccgcgat     | ctgtgtgaac   | cagetgtgeg  | tgaggttctt  | cgccctggac  | ctggacagcg   | 3240 |
| gtctgttcag    | cgccccccac   | gtgcactgt   | gcatcaggaa  | caaccaactgg | gacaacagcc   | 3300 |
| ccagccccaa    | catgtacggc   | ctgaacagg   | agggtgttgc  | gcagctgcag  | aggcggttacc  | 3360 |
| cacagctgcc    | caggcccggt   | gcccacggca  | gggtgtacga  | catgaacacc  | ggcacccttg   | 3420 |
| ggaactacga    | cccccaggatc  | aacctgggt   | ccgtgaacag  | goggctgccc  | cacgcctgg    | 3480 |
| tgtgtgcacca   | caacgcgac    | ccacagacgc  | acttcagtc   | cttcgtgtac  | aagctgaaag   | 3540 |
| gcaggacccgt   | gtgtgtgtgt   | ggcggagaag  | tgagcgtgc   | cgggcaagatg | gtggacttgc   | 3600 |
| tgagcgcacag   | ggcccgaggcc  | accccccgtt  | ccaggcttgc  | cctcggcatc  | cccgccgacg   | 3660 |
| tgcaccaagta   | cgacatcatc   | ttctgtaca   | tcaggacccc  | atacaagttac | caccattttac  | 3720 |
| agcagtgtgt    | ggaccacacgc  | atcaagctg   | gcatgtgtac  | caagaaggcc  | tgcctgtacc   | 3780 |
| tgaacccccc    | aggccatgttgc | gtgacgtgt   | gtacggctt   | cgccgacagg  | gcccacgg     | 3840 |
| gcatcatcttgg  | cgccatcgcc   | aggctgttca  | agttcagcag  | ggtgtgc     | ccaaagagca   | 3900 |
| gcctggggaa    | aaccgggtg    | ctgttcgtgt  | tcatcggtt   | cgacccggaa  | gccaggaccc   | 3960 |
| acaaccccta    | caagctgagc   | agcacccctg  | caaacatcta  | caccggcagc  | aggctgcacg   | 4020 |
| aggcccggtgt   | cgccccccac   | taccacgtgt  | tcaggggcgtt | tatcgccacc  | gcccacgggg   | 4080 |
| gctgtatcat    | caacgtgtcc   | aacagcagg   | gcccacgggg  | aggccggatgt | tgcggccccc   | 4140 |
| tgtacaagaa    | gttcccccgg   | agttcgtacc  | tgagccccc   | cgaggtgggc  | aaggccaggc   | 4200 |
| tggtgaagg     | cgccgctaa    | cacatcatc   | acgcgcgtgg  | cccccaacttc | aacaagggtg   | 4260 |
| ggcgggtgt     | aggcgacaa    | cagctggcc   | aagcttacg   | gagcatcgcc  | aagatctgt    | 4320 |
| acgacaataa    | ctacaagac    | gtggccatcc  | cactgtctc   | caccggcatc  | ttcagcggca   | 4380 |
| acaaggacac    | gtgtacccat   | agcctgtacc  | acctgtctac  | cgccctggac  | accaccgtat   | 4440 |
| ccgacgtgtcc   | catctactgc   | agggacaaga  | agtgggagat  | gaccctgaag  | gaggccgtgg   | 4500 |
| ccaggcgggaa   | ggccgtggaa   | gagatctgc   | tcagcgcac   | ctccagctgt  | accgagcccg   | 4560 |
| acgcccggat    | gtgtgtgtgt   | cccccaaga   | gtccctgtgc  | cgccggagg   | ggctacacga   | 4620 |
| ccagcgtacgg   | caagacgttc   | agtcaccc    | ggccacccaa  | gttccaccc   | ggccgttaagg  | 4680 |
| acatcgccga    | gatcaacgtc   | atgtggcc    | tggccaccc   | ggccaaacgg  | caggtgtgc    | 4740 |
| tgtacatctt    | gggcggacac   | atgttcac    | tcaggacac   | gtgccccgt   | gaggaaacgg   | 4800 |
| aggccacccac   | accacccac    | acccctcc    | gtctgtgc    | caacgttatg  | acacccggaa   | 4860 |
| gggtgcacgc    | gtgtacggcc   | agcaggcccg  | agcagatcac  | cggtgtac    | tccttccac    | 4920 |
| tgcaccaagta   | caggatcacc   | ggcgtgcaga  | agatccat    | cagccagccc  | atcctgttca   | 4980 |
| gccccaaatgt   | gcccgtctac   | atccacccca  | ggaatgttac  | ggtggagacc  | ccacccgtgg   | 5040 |
| acgagacacc    | cgagccaa     | gcccgtac    | agagcacc    | gggcacaccc  | gagcgcac     | 5100 |
| cctgtatcat    | cgaggacgg    | acaaggaccc  | ggacccggaa  | ggccatcatt  | atcgaggaa    | 5160 |
| aggaagggaa    | cagcatc      | ctgtgtac    | acggccccc   | ccaccagggt  | tgcagggtgg   | 5220 |
| aggccgacat    | ccacccggcc   | cccaacgtgt  | ccagctcc    | ctggagcatc  | ccacaccc     | 5280 |
| gctgtatcat    | cggtggacac   | ctgacatcc   | tggtggacat  | ggggccggcc  | agcgtgact    | 5340 |
| ccggccgtac    | cgccgtcc     | accaacgtt   | acttcgttca  | gagcatgg    | ttcctggcc    | 5400 |
| ggccctgtcc    | agctcccg     | accgttgt    | ggaacccacc  | ccacccagct  | cccaggacca   | 5460 |
| ggacccaaatgt  | cttggctccc   | agcaggccgt  | gcagcaggac  | cagcgttgc   | agcaccaccc   | 5520 |
| ccggcgtgt     | cagggtgtac   | accagggggg  | aacttgggg   | cctgacatcc  | agcaggaccc   | 5580 |
| ccaggcgttc    | cggtggacac   | actgtctgg   | tgtccaccc   | acccctgggt  | aacagggttgc  | 5640 |
| tcaccaggaa    | ggaattcg     | gccttgcgtt  | cccagcaaca  | gagacgggt   | gacgcccgg    | 5700 |
| cctacatctt    | cagcagcgac   | accggccagg  | gacacctgc   | gcaaaaagac  | gtgaggcaga   | 5760 |
| cctgtgtgt     | cgagggtgt    | ctggagacac  | ccqaggtgt   | aatcgtatc   | gccccccaggc  | 5820 |
| tggaccaggaa   | gaggaggaa    | ctgtgtac    | agaaacttgc  | gctgaacccc  | acccccacca   | 5880 |
| acaggacac     | gttacccat    | aggaacgtgt  | agaaacatg   | ggccatcacc  | ggcaggccga   | 5940 |
| tcctgtgtgt    | cctggacac    | tacctgtac   | ccgaggggca  | ggtggaggt   | tacaggaccc   | 6000 |
| tgcacccctgt   | gccaactgtac  | agctccat    | tgaacaggcc  | tttctccat   | cccaagggttgc | 6060 |
| ccgtgtgtgt    | ctgcaacgt    | atgtgtac    | ccatgttcc   | caccgttgc   | agctactgc    | 6120 |
| tcatcccgat    | gtacgtac     | tacccgttcc  | ttctgttcc   | cgccatgttgc | tgcttgc      | 6180 |
| ccggccatgtt   | ctggccgtcc   | aagctgttcc  | gttccaccc   | gaaacacacg  | tacctggac    | 6240 |
| ccaccatcgat   | gagcgtgtgt   | cccacgttca  | tccagaacac  | cctgtcaaa   | gtgtgtggcc   | 6300 |
| ctggccaccaaa  | gaggaaactgc  | aacgtgtac   | agatgaggaa  | gctggccgt   | ctggacagcg   | 6360 |
| ctgccttcaatgt | ctgtgtgtgt   | ttcaagaaat  | acgcctgttca | caacgtgtac  | tgggagaccc   | 6420 |
| tcaaggaaatgt  | ccccatcagg   | ctgacgtac   | agaaacgtgt  | gaactacatc  | accaagctg    | 6480 |
| aggccccccaa   | ggccgtgtcc   | ctgttcgttca | agacccacaa  | cctgtacat   | ctgcaggac    | 6540 |

-continued

---

|              |             |             |             |             |             |      |
|--------------|-------------|-------------|-------------|-------------|-------------|------|
| tcccaatgga   | cagggttcgtg | atggacactga | agagggacgt  | gaagggtgaca | cccgccacca  | 6600 |
| agcacaccga   | ggagaggccc  | aagggtcagg  | tgatccaggc  | cgctgaccca  | ctggccacccg | 6660 |
| cctacctgtg   | cggcacccac  | agggagctgg  | tgaggcgccgt | gaacgcccgt  | ctgctgcucca | 6720 |
| acatccacac   | cctgttcgtac | atgagcgcgg  | aggacttgcg  | cgccatcatc  | gccgagact   | 6780 |
| tccagccccg   | cgactcgctg  | ctggagacgg  | acatcgccag  | cttcgacaaag | agcgaggatg  | 6840 |
| acgetatggc   | cctgacccgt  | ctgatgatcc  | tggaggaccc  | gggcgtggac  | gccgagctgc  | 6900 |
| tcaccctgtat  | cgaggctgccc | ttccggcaga  | tcagctccat  | ccacctgcgg  | accaagacca  | 6960 |
| agttcacatgt  | cgggcgatgt  | atgaaaagcg  | aatgttccct  | gaccctgttc  | gtgaacacccg | 7020 |
| tgtacacat    | tgtgtatcgcc | agcagggtgc  | tgccggagag  | gctgacccggc | agccccctgg  | 7080 |
| ctgccttcat   | cggcacgcac  | aacatcgta   | aggcgctgaa  | aagcacaag   | ctgatggccg  | 7140 |
| acaggcgccgc  | cacctggctg  | aacatgggg   | tgaagatcat  | cgacgcccgt  | gtggggcgaga | 7200 |
| aggcccccta   | ttctcgccgc  | ggattcatcc  | tgtgcacag   | cgtgacccggc | accgcctgc   | 7260 |
| gggtggccgg   | ccccctgtgg  | aggtgttca   | agctggccaa  | gccaactggc  | gtgacatgt   | 7320 |
| agcaacgacga  | tgacaggcg   | aggggccgtc  | acgaggaaag  | caccagggtt  | aacagggttg  | 7380 |
| gcatctcgtag  | cgagctgtgc  | aggggcggtg  | agagcaggta  | cgagacccgt  | ggcaccacga  | 7440 |
| tcatecgatgt  | ggctatgacc  | acactggca   | gctccgtcaa  | gagettctcc  | tacctgaggg  | 7500 |
| ggggccctata  | aactctctac  | ggctaacctg  | aatggactac  | gacatagtt   | agtccgcaca  | 7560 |
| ggccgcaccacc | atagagatgt  | cageccctag  | aaccttactg  | cttctgtt    | ggggagctgt  | 7620 |
| tgtctgtacaa  | gagacatggg  | ctggatctt   | ccacagcccc  | actgtcgctt  | accaccaagt  | 7680 |
| cgagagggggg  | ggaggagggt  | ccgggggagg  | aggctccctg  | aatgtcggcc  | aggccgtgca  | 7740 |
| cgccgcaccac  | gcccggatgt  | acgaggccgg  | ccggggagggt | atcgccggca  | tttgtcgctg  | 7800 |
| cctggccgtgc  | ctcgccgtgg  | tgggtggatc  | agctgtggtc  | cgactgtta   | tgtcgagaag  | 7860 |
| aaagtcatcc   | ggccggaaagg | gagggtccct  | ctctcaggct  | gcttcgtcta  | cagtgcctag  | 7920 |
| agctttagt    | tggttatctc  | agctgtaaac  | tcgagtatgt  | taclgtcaaa  | gtgtattgtc  | 7980 |
| acccccccgaa  | agaccattat  | gtgacacacc  | ctcagatata  | cgcccaaaaca | tttacagccg  | 8040 |
| cgggtgtcaaa  | aaccgggtgg  | acgtgttgc   | atccctgtt   | ggggaggatca | gcccgttaatt | 8100 |
| tttataattgg  | cttgggtgtc  | gtctacttgg  | atgtgttgc   | tttgcgtacc  | aaccggaaac  | 8160 |
| ataattgaat   | acagcagca   | tttggaaatgt | gtttatcatag | aactcgccgc  | gattggcatg  | 8220 |
| ccgccttaaa   | atttttttt   | tttcttttcc  | gaatcgatt   | ttgtttttaa  | 8280        |      |
| tatttcaaaa   | aaaaaaaaaa  | atcttagaaaa | aaaaaaaaaa  | aaaaaaaaaa  | 8340        |      |
| aaaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | 8400        |      |
| aaaaaaaaaaa  | aaaaaaa     |             |             |             | 8416        |      |

---

SEQ ID NO: 107      moltype = RNA   length = 8914  
 FEATURE      Location/Qualifiers  
 misc\_feature      1..8914  
 note = Synthetic polynucleotide  
 source      1..8914  
 mol\_type = other RNA  
 organism = synthetic construct

SEQUENCE: 107

|              |              |             |             |             |             |      |
|--------------|--------------|-------------|-------------|-------------|-------------|------|
| atgggcggc    | catgagagaa   | gcccagacca  | attacctacc  | caaaatggag  | aaagttcacg  | 60   |
| ttgacatcgaa  | ggagacacgc   | ccatcttc    | gagctttcga  | ggcgagttc   | ccgcgttgc   | 120  |
| aggtagaaacg  | caagcagggtc  | actgataatg  | accatgtta   | tgccagacgc  | tttgcgtatc  | 180  |
| tggcttcaaa   | actgtatcgaa  | acggagggtgg | accatccga   | cacgatcc    | gacattggaa  | 240  |
| gtgcgcggc    | ccgcacaaat   | tatcttaa    | gcaagatata  | ttgtatctgt  | ccgtatgtat  | 300  |
| gtcgccaga    | atcgccatgt   | ttgtataatg  | atgcaactaa  | gtgtggaaa   | aactgttaagg | 360  |
| aaataactgt   | taaggaatttgc | gacaagaaaa  | tgaaggacgt  | ggccgcggcgt | atgagcgcacc | 420  |
| ctgaccttgc   | aaactgagact  | atgtgcctcc  | acgacgacga  | gtcgctgtgc  | tacgaagggc  | 480  |
| aagtgcgtgt   | ttaccaggat   | gtatacgcgc  | tcgacggccc  | caccagocgt  | taccaccagg  | 540  |
| ccaaacaagggg | cgtgggggt    | gcctactgg   | tcggcttcga  | caccacacc   | tttgcgttca  | 600  |
| agaaacctggc  | ccggcgttac   | cccgatctac  | gcaccaactg  | ggccgcggag  | accgtgtcgt  | 660  |
| ccgcgcaggaa  | catcggtctg   | tgcagcagcg  | acgtgtatgg  | gaggagccgg  | agaggcatga  | 720  |
| gcatctcgat   | gaagaaatac   | ctgaaggccca | gcaacaacgt  | gtgtttcagc  | gtgggcagca  | 780  |
| ccatcttacca  | cgagaaggagg  | gacctgtca   | ggagctggca  | cctgcaccc   | gtgtttccacc | 840  |
| tgaggggccaa  | cgaaqaactac  | acctgcgtgt  | gcgagacccat | cgtgtatgtc  | gacggctacg  | 900  |
| ttgtgttgc    | gatgcacatc   | agccccggcc  | tgtacggca   | gcccagccgc  | tacgcgccta  | 960  |
| caatgtcacat  | ggagggttcc   | ctgtgtctca  | aggtgtacca  | caccctgtt   | ggcgagagg   | 1020 |
| tgagtttccc   | cgtgtgcacc   | taclgttgc   | ccaccctgtt  | cgacccatgt  | accggcatcc  | 1080 |
| tgggccacccg  | cgtgtgcgc    | gacgtacggcc | agaaggtgt   | ctgtggccctg | aaccaggagga | 1140 |
| tcgtgttcaaa  | cggcaggacc   | cagaggaaaca | ccaacacaat  | gaagaactac  | ctgtgcggc   | 1200 |
| ttgtgttgc    | ggcttgcgttca | agggtggccca | aggagttacaa | ggaggacccag | gaagacgaga  | 1260 |
| ggccccctggg  | cgtgggggtt   | aggcgttgc   | tgatggggctg | ctgtgtggcc  | ttcaggccgc  | 1320 |
| acaagatcac   | cacgtatctac  | aaaggccccc  | acacccacac  | catcatca    | gtgtacacgc  | 1380 |
| acttccacat   | cttcgtgtc    | cccaggatgc  | gcagcaacac  | cctggagatc  | ggccgttgcgg | 1440 |
| cccggtatcg   | gaagatgtc    | gaggaacaca  | aggagccca   | cccactgtc   | accggccgg   | 1500 |
| acgtgcacgg   | ggccaaatgt   | ggtgtcgacg  | aggccaaatgt | gggtgagggt  | gccgaggaaac | 1560 |
| tgaggggccgc  | cgtgcaccc    | ctgggtgc    | actgtggagg  | accacccctgt | gaagccgcacg | 1620 |
| tgggacttgc   | gtgtcaggat   | gcccggccgc  | gaacgtgtca  | gacacccagg  | ggccgtatca  | 1680 |
| aggtgaccat   | ctacgcacggc  | gaggacaaga  | tcggcagata  | cgccgtgtc   | agcccacagg  | 1740 |
| ccgtgtatcgaa | gtccgaga     | ctgagctca   | tccacccact  | ggccgagcag  | gtgtatgtca  | 1800 |
| tcacccacat   | cggcaggaaat  | ggcgttgcgt  | ctaccacggc  | aagggtgtc   | 1860        |      |
| tgcccgagg    | ccacgcacatc  | cccgtgc     | acttccaggc  | cctgtggcgt  | agccaccca   | 1920 |
| tcgtgtatca   | cgagagggtt   | ttctgttca   | ggtacctgtca | ccatatcc    | acccacggcg  | 1980 |
| gagccctgtaa  | caccgcacgg   | gaataactaca | agaccgtgaa  | gcccagccgc  | cacgcggcg   | 2040 |
| agtacccatgt  | cgacatcgac   | aggaagacgt  | gcgtgtaa    | agagctgtt   | accggccgtt  | 2100 |
| gactgaccgg   | cgagctgtgt   | gacccaccc   | tcacacgt    | cgccctacgt  | agccctgagg  | 2160 |
| ccagacccgc   | cgctccat     | cagggtccca  | ccatcgccgt  | gtacggcg    | ccggcagcgt  | 2220 |
| gaaaagagccg  | catcatca     | agcgcctgt   | ccaagaaaga  | cctgggtgtc  | agcgcacaa   | 2280 |

-continued

---

|             |             |             |             |             |              |      |
|-------------|-------------|-------------|-------------|-------------|--------------|------|
| aagagaactg  | cgccgagatc  | atcagggacg  | tgaagaagat  | gaaaggcctg  | gacgtgaacg   | 2340 |
| cgcgcaccgt  | ggacagcgtg  | ctgctgaacg  | gctgcaagca  | ccccgtggag  | accctgtaca   | 2400 |
| tcgacgaggc  | cttcgttgc   | cacgccggca  | ccctgagggc  | cctgatccgc  | atcatcaggc   | 2460 |
| ccaagaaaagc | cgtgtgtgc   | ggcgcaccca  | agcagtgcgg  | tttcttcaac  | atgatgtgcc   | 2520 |
| tgaagggtca  | cttcaaccac  | gagatctgc   | cccagggtgt  | ccacaagagc  | atcagcaggc   | 2580 |
| ggtgtcaccaa | gagcgtgacc  | agcgtgtgc   | gcacccctgt  | ctacgacaag  | aaaatgagga   | 2640 |
| ccacccaacc  | caaggagacc  | aaaatctgt   | tcgacaccac  | aggcagaccc  | aagcccaacg   | 2700 |
| aggacgaccc  | gatcctgacc  | tgcttcaggg  | gctgggtgaa  | gcagctgcag  | atcgactaca   | 2760 |
| agggcaacga  | gatcatgacc  | gcccgtgcca  | gcccagggtct | gaccaggaaag | ggcgtgtacg   | 2820 |
| ccgtgaggat  | caagggtgac  | gagaacccac  | tgtacgtctc  | caccaggcg   | cacgtgaacg   | 2880 |
| tgtgtgtgc   | caggaccgg   | gacaggatcg  | tgtggaaagac | cttgcggggc  | gacccttgg    | 2940 |
| tcaagaccc   | gaccgccaag  | tacccggca   | acttcaccgc  | caccatcgaa  | gagtggcagg   | 3000 |
| ccgagcacga  | cccccaccc   | aggcacatcc  | ttggagaggcc | cgaccccccac | gacgtgttcc   | 3060 |
| agaacaaggc  | caacgtgtgc  | tggggcaagg  | ccctgggtgc  | cgtgtcaag   | accgcgcggca  | 3120 |
| tcgacatgac  | cacagacgac  | tggaaacaccg | tggactactt  | cgagacccgc  | aaggccacaca  | 3180 |
| gcccggagat  | cgtgtgaac   | cagctgtgeg  | tgagggtctt  | cggcctggac  | ctggacagcg   | 3240 |
| gcctgttcag  | cgccccccac  | gttgcactca  | gcatcaggaa  | caaccaactgg | gacaacagcc   | 3300 |
| ccagccccaa  | catgtacggc  | ctgaacaaa   | agggtgttcg  | cgacgtgacg  | aggcggttacc  | 3360 |
| cacagctcc   | caggccgtg   | gcccacggca  | gggtgtacga  | catgaacacc  | ggcaccctga   | 3420 |
| ggaactacga  | ccccaggatc  | aacctgtgc   | ccgtgaacag  | geggctgccc  | caegccctgg   | 3480 |
| tgtgtcacca  | caacggacac  | ccacagacgc  | acttcagctc  | tttcgtgagc  | aagctgaaag   | 3540 |
| gcaggacccgt | gtctgtgtgc  | ggcgcagaac  | tgagcgtgc   | cggcgaatgt  | gtggacttgc   | 3600 |
| tgagcagac   | gcccggggc   | acccctccgg  | ccaggctgga  | cctcggcata  | cccgccgacg   | 3660 |
| tgcccaagta  | cgacatcatc  | ttcgtgaacg  | tcaggacccc  | atacaagtac  | caccattacc   | 3720 |
| agcaagtgcg  | ggaccacgc   | atcaagctca  | gcatgtgtac  | caagaaggcc  | tgcctgcacc   | 3780 |
| tgaacccccc  | aggcacctgc  | gtgagcatcg  | gttgcgggtc  | cgccgcacagg | gcccgcggaa   | 3840 |
| gcatcatttt  | cgccatccgc  | aggcttca    | agttcagcag  | ggtgtcaaaa  | cccaagagca   | 3900 |
| gcctggagga  | aaccgggtg   | ctgttctgt   | tcatcggcta  | cgaccggaaag | gcccggaccc   | 3960 |
| acaaacctca  | caagctgacg  | agcaccctga  | caaacatcta  | caccggcagc  | aggctgcacg   | 4020 |
| aggccggctg  | cgccccccac  | taccacgtgc  | tcagggggca  | tatcgcaccc  | gcccacccggg  | 4080 |
| gcgtgtatcat | caacgtgtcc  | aacacgca    | gcccggccgc  | aggccggatgt | tgccggccccc  | 4140 |
| tgtacaagaa  | gttcccccgag | agtttcgacc  | tgcagcccat  | cgagggtggc  | aaggccaggc   | 4200 |
| tgtgtgaaggg | cgccgcttaag | cacatcatcc  | acggcgttgg  | cccccaacttc | aacaagggt    | 4260 |
| gcgagggtgg  | aggcgacac   | cagetggc    | aaggcttacta | gagcatcgcc  | aggatcgta    | 4320 |
| acgacaataa  | ctacaagacg  | gtggccatcc  | caactgtcg   | caccggcata  | ttcagcggca   | 4380 |
| acaaggacac  | gtgtacccag  | agcttgcacc  | cccccctggac | accaccgtat  | 4440         |      |
| ccgacgtggc  | catctactgc  | agggacaaga  | agtgggagat  | gaccctgaa   | gaggccgtgg   | 4500 |
| ccaggccgg   | ggccgtggaa  | gagatctcg   | tcagcgcac   | ctccagctgt  | accgagcccg   | 4560 |
| acggccgacgt | gtgtgggtg   | caccccaaga  | gtcccttgc   | cgccggggaa  | ggctacacga   | 4620 |
| ccagcgtaccc | caagaccc    | agttacttgc  | ggccgttgc   | aggccacca   | gttccaccc    | 4680 |
| acatcgccg   | gatcaacgct  | atgtggcccg  | tggccacccg  | ggccaaacgag | caggtgtgc    | 4740 |
| tgtacatctt  | ggggcggagc  | atgttcacca  | tcaggagca   | gtgccccctg  | gagggaaagcg  | 4800 |
| aggccacccac | accaccac    | acccttgcct  | ccacgtctat  | acaccggaga  | 4860         |      |
| gggtgcacgc  | gtgttgcac   | agcagccccc  | agcagatcac  | cgtgtcaacg  | tccttccac    | 4920 |
| tgcccaagta  | caggatcacc  | ggcgtcaga   | agatccatgt  | cagccacccc  | atcctgttca   | 4980 |
| gcccggaaat  | gcccgttac   | atccccccca  | ggaagttactt | gtgttgcaccc | ccacccgtgg   | 5040 |
| acggacaccc  | cgagccaa    | gcccggac    | agagcaccga  | gggcacaccc  | gagcagccac   | 5100 |
| cctgtgtatcc | cgaggacgg   | acaaggaccc  | ggacccccc   | gcccacatt   | atcgaggaa    | 5160 |
| aggaagggaa  | cagcatcgc   | ctgtgtgc    | acggccccac  | ccaccagggt  | ctgcagggtt   | 5220 |
| aggccgacat  | ccacggccca  | cccaactgt   | ccagctccat  | ctggagaccc  | ccacacggca   | 5280 |
| gcgacttgc   | cggtggac    | ctgacatcc   | ctgacacttgc | ggagggccct  | agcgtgtaccc  | 5340 |
| ccggccgtac  | cagcgtcc    | accaacagct  | acttcgcac   | gagcatggag  | ttccctggca   | 5400 |
| ggccccgtcc  | agetcggcagg | accgtgttca  | ggaacccacc  | ccacccagct  | cccaggacca   | 5460 |
| gacccggaaat | cttgggtcc   | agcagggtt   | gacacgggt   | cgccgttgc   | agcaccggcc   | 5520 |
| ccggcgttgc  | cgagggtatc  | acaggggagg  | aactggggatc | cctgacaccc  | agcaggaccc   | 5580 |
| ccacggcgtt  | cgtgtacgg   | actgtctgg   | tgttcaaccc  | acccggcg    | aacagggt     | 5640 |
| tcaccaggga  | ggaattcgag  | gtcttgcgt   | cccaacca    | gagacgggtt  | gacccggcg    | 5700 |
| cctacatctt  | cagcgtcgac  | accccccagg  | gacacctgc   | gaaaaagagc  | gtgaggcaga   | 5760 |
| cctgtgttgc  | cgagggtgtt  | ctggagatgg  | ccgagcttgc  | aatcagatcc  | gcccccaaggc  | 5820 |
| tggaccagga  | gaaggaggaa  | ctgttcagg   | agaaaaactgc | gtgttgcaccc | acccggacca   | 5880 |
| acaggagcc   | gttccaccc   | aggaagggtt  | agaacatgaa  | ggccatcacc  | gcccgggacca  | 5940 |
| tcctgcagg   | cttgggacac  | tacccatgt   | cccgaggccaa | gttgggtgc   | tacaggaccc   | 6000 |
| tgcaccccg   | gtccactgtac | agtcctccagg | tgaaacaggcc | cttccatcc   | cccaagggttgg | 6060 |
| cctgtgttgc  | ctgcaactgt  | atgttcagg   | agaaacttcc  | caccgttgc   | agctactgtca  | 6120 |
| tcatcccttgc | gtacgtcgcc  | tacccatgg   | ttgttgcaccc | tggttgcaccc | tgccttgcaca  | 6180 |
| ccggccgttcc | ctggccgttcc | aagcttgc    | gttggaccc   | cccgacgttgc | tgccttgcaca  | 6240 |
| ccacccatcg  | gagccgtgt   | cccaactgt   | ccctgacaa   | cctgcacac   | gtgttgcggc   | 6300 |
| ctggccatcc  | ggggacatgc  | aacgttgc    | agatggggaa  | gtgttgc     | ctggccaccc   | 6360 |
| ctgccttcc   | cgtggatgt   | ttaaagaaat  | acccctgtca  | caacgtatc   | ttggggaccc   | 6420 |
| tcaaggagaa  | ccccatcagg  | ctgaccgttgc | agaacgttgc  | gaactacatc  | accaagctga   | 6480 |
| aggggcccca  | ggccgtgtcc  | ctgttgc     | agacccacaa  | cctgacatgc  | ctgcaggacca  | 6540 |
| tcccaatgtt  | cagggtgtgt  | atggacatgt  | agagggacgt  | gaaggttgc   | cccgccacca   | 6600 |
| agcacacccg  | ggggaggccc  | aaagggtc    | tgatccagg   | cgctgacca   | ctggccaccc   | 6660 |
| cctacatgtt  | cggtgtatcc  | aggaggatgt  | tgagggttgc  | gaacggccgt  | ctgttgc      | 6720 |
| acatccacac  | cttgcgttgc  | atgagcgttgc | aggacttgc   | cgccatcata  | gcccggacact  | 6780 |
| tccagccccc  | cgactgtgt   | ctggggaccc  | acatcgccat  | tttcgttgc   | agcggaggat   | 6840 |
| acgctatggc  | cctgtgtgt   | ctgtgtatcc  | ttggaggaccc | gggcgttgc   | gccgagctgc   | 6900 |
| tcaccctgt   | cgagggtgtcc | ttccggcg    | taagcttccat | ccacccgtcc  | accaagacca   | 6960 |
| agttcaatgtt | cggtgtatcc  | atggaaatggc | gaatgttcc   | gaccctgttgc | gtgaaaccc    | 7020 |

-continued

---

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| tgtcaacat   | tgtgatcgcc  | agcagggtgc  | tgccggagag  | gctgaccggc  | agccctgcg   | 7080 |
| ctgccttcat  | cggcgacgac  | aacatcgta   | agggcgtga   | aagcgacaag  | ctgatggccg  | 7140 |
| acaggtgcgc  | cacctggctg  | aacatggagg  | tgaagatcat  | cgacgcccgt  | gtggcgaga   | 7200 |
| agggcccccta | cttctgcggc  | ggattcatcc  | tgtgcgacag  | cgtgaccggc  | accgcctgca  | 7260 |
| gggtggccga  | ccccctgaag  | aggcttca    | agctggcaaa  | gccactggcc  | gctgacatg   | 7320 |
| agcacgacg   | tgacaggcgg  | aggccctgc   | acgagggaaa  | caccaggttg  | aacagggtgg  | 7380 |
| gcatctctag  | cgagctgtgc  | aaggccgtgg  | agagcaggta  | cgagaccgtg  | ggcaccagca  | 7440 |
| tcatectgtat | ggctatgacc  | acactggcca  | gctccgtcaa  | gagcttctcc  | tacctgaggg  | 7500 |
| ggggcccttat | aactcttctac | ggctaacctg  | aatggactac  | gacatagtct  | agtccgc当地   | 7560 |
| ggccgc当地    | atagagatgt  | cageccctact | aaccttactg  | tttctgttt   | ggggagctgt  | 7620 |
| tgtctgacaa  | gagacatgggg | ctggatctct  | gagcgagggt  | accggccagg  | gcctgtgc当地  | 7680 |
| cgccgc当地    | ccaaagaccc  | accagggtct  | gtgcaacacc  | accaggaaaga | ccagcgacgg  | 7740 |
| caigactac   | ctggccgctc  | ccacccggc   | cacttggcc   | tgcagcacc   | gcctgacccc  | 7800 |
| ttgcatacgc  | accaccatcc  | tgaacatcg   | caccgacta   | tgcgtgttgg  | tgtagctgtg  | 7860 |
| gcccagggt   | acctaccaca  | gccccagcta  | cgccctaccac | cagtc当地     | ggaggggccaa | 7920 |
| gtacaagagg  | gagccgtga   | gcctgaccct  | gccc当地      | ctgggeggcc  | tgacaatggg  | 7980 |
| cgccatcgcc  | gccc当地      | gcacccggc   | cacccccc    | gtggccacc   | agcagttcca  | 8040 |
| gcagctgc当地  | gccc当地      | acgacgac    | tgacagggtg  | gagaagggtt  | tgacatgtca  | 8100 |
| ggagaagtgcc | ctgaccaggc  | tgagcgggt   | ggtgctc当地   | aacaggagg   | gcctgaccc   | 8160 |
| gctgttctg   | aaggaggccg  | gcctgtgc当地  | cgccctga    | gaggagg     | gctgtacgc   | 8220 |
| cgaccacacc  | ggccctgt    | tcgtggcc    | tgc当地       | tcggcc      | tcgccc当地    | 8280 |
| ggtgatttgt  | tgctgttgc当地 | cagetttat   | tgccaga     | aagtcttca   | ggggaaagg   | 8340 |
| aggcttctac  | tctcaggctg  | cttctgtac   | atgtc当地     | tttcttatgt  | gtttatctca  | 8400 |
| gctgggccc   | ggaggcagcg  | actacaagga  | cgacgatgac  | aagttaactc  | gagttatgtt  | 8460 |
| cgtgcaaaagg | tgatttgtc   | ccccccgaa   | accatattgt  | gacacacc    | cagtatc     | 8520 |
| ccaaacacat  | tacagccgc当地 | gtgtaaaaaa  | ccgc当地      | gtggtaaca   | tccctgtc当地  | 8580 |
| gaggatcgc当地 | cgttattat   | atatttgtt   | tggtgtc当地   | tactattgt   | gccatgtac   | 8640 |
| tgtgaccaa   | ccagaacat   | aattgaatac  | tgacgcaatt  | ggcaaggtc当地 | ttacatagaa  | 8700 |
| ctcgccgca   | ttggcatg    | gccttaaaa   | tttttattt   | tttttctt    | tctttccg    | 8760 |
| atcggatttt  | tttttataa   | tttcaaaaaa  | aaaaaaaaaa  | aaaaaaaaat  | ctagaaaaaa  | 8820 |
| aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | 8880 |
| aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaa        |             |             | 8914 |

---

SEQ ID NO: 108      moltype = RNA   length = 8455  
 FEATURE      Location/Qualifiers  
 misc\_feature      1..8455  
 note = Synthetic polynucleotide  
 source      1..8455  
 mol\_type = other RNA  
 organism = synthetic construct

SEQUENCE: 108

|            |             |            |             |            |            |      |
|------------|-------------|------------|-------------|------------|------------|------|
| atgggc当地   | catgagagaa  | gcccagacca | attacatcacc | caaaatggag | aaagttc    | 60   |
| ttgacatcg  | gaaagacagc  | ccatcttca  | gagctttc    | gcggagotc  | ccgc当地     | 120  |
| aggtagaa   | caagcagg    | actgataatg | accatgtta   | tgccagacg  | tttgc当地    | 180  |
| tggcttcaa  | actgatcg    | acggagg    | accatccg    | ca         | gatc当地     | 240  |
| gtcgccccc  | ccgcaaaatg  | tatctta    | acaatgtatc  | tgtatctgt  | ccgatgagat | 300  |
| gtcgccg    | aga         | ttgtatctgt | atgc当地      | gtgaa      | aaactgtaa  | 360  |
| aaataactg  | taaggaa     | ttgtatctgt | atgc当地      | gtgaa      | aaactgtaa  | 420  |
| ctgacctg   | aaactgagact | atgtgc     | cc          | gtcg       | gtc当地      | 480  |
| aagtgc当地   | tttccaggat  | gtatcgc当地  | tc当地        | caccagoc   | gttaccagg  | 540  |
| ccaaacaagg | ctgtgggt    | cc         | actgtg      | cc         | tttgc当地    | 600  |
| agAACCTG   | ccgc当地      | cc         | cc          | cc         | tttgc当地    | 660  |
| ccgc当地     | ccatcg      | cc         | cc          | cc         | cc         | 720  |
| gcatctcg   | gaagaaat    | ctgaa      | cc          | cc         | cc         | 780  |
| ccatcttca  | ccgagaagg   | cc         | cc          | cc         | cc         | 840  |
| tgaggggc   | ccgaaat     | cc         | cc          | cc         | cc         | 900  |
| ttgtgaa    | cc          | cc         | cc          | cc         | cc         | 960  |
| caatgcac   | cc          | cc         | cc          | cc         | cc         | 1020 |
| tgagtttcc  | cc          | cc         | cc          | cc         | cc         | 1080 |
| tgaggcc    | cc          | cc         | cc          | cc         | cc         | 1140 |
| tcgtgtca   | ccgc当地      | cc         | cc          | cc         | cc         | 1200 |
| ttgtggcc   | cc          | cc         | cc          | cc         | cc         | 1260 |
| ggcccttgg  | cc          | cc         | cc          | cc         | cc         | 1320 |
| acaagatc   | cc          | cc         | cc          | cc         | cc         | 1380 |
| acttccac   | cc          | cc         | cc          | cc         | cc         | 1440 |
| cccgatc    | cc          | cc         | cc          | cc         | cc         | 1500 |
| acgtgc当地   | cc          | cc         | cc          | cc         | cc         | 1560 |
| tgagggcc   | cc          | cc         | cc          | cc         | cc         | 1620 |
| tgagctgtat | cc          | cc         | cc          | cc         | cc         | 1680 |
| aggtgacc   | cc          | cc         | cc          | cc         | cc         | 1740 |
| ccgtgtca   | cc          | cc         | cc          | cc         | cc         | 1800 |
| tcacccac   | cc          | cc         | cc          | cc         | cc         | 1860 |
| tgcccg     | cc          | cc         | cc          | cc         | cc         | 1920 |
| tcgtgtac   | cc          | cc         | cc          | cc         | cc         | 1980 |
| gagccctg   | cc          | cc         | cc          | cc         | cc         | 2040 |
| agtacctgt  | cc          | cc         | cc          | cc         | cc         | 2100 |
| gactgacc   | cc          | cc         | cc          | cc         | cc         | 2160 |
| ccagccccc  | cc          | cc         | cc          | cc         | cc         | 2220 |
| gaaaagagcg | cc          | cc         | cc          | cc         | cc         | 2280 |

-continued

---

|             |             |             |              |             |             |      |
|-------------|-------------|-------------|--------------|-------------|-------------|------|
| aagagaactg  | cgccgagatc  | atcagggacg  | tgaagaagat   | gaaaggcctg  | gacgtgaacg  | 2340 |
| cgcgcaccgt  | ggacagcgtg  | ctgctgaacg  | gctgcaagca   | ccccgtggag  | accctgtaca  | 2400 |
| tcgacgaggc  | cttcgttgc   | cacgccggca  | ccctgagggc   | cctgatccgc  | atcatcaggc  | 2460 |
| ccaagaaaagc | cgtgtgtgc   | ggcgcaccca  | agcagtgcgg   | tttcttcaac  | atgatgtgcc  | 2520 |
| tgaagggtca  | cttcaaccac  | gagatctgc   | cccagggtgt   | ccacaagagc  | atcagcaggc  | 2580 |
| ggtgtcaccaa | gagcgtgacc  | agcgtgtgc   | gcacccctgt   | ctacgacaag  | aaaatgagga  | 2640 |
| ccacccaacc  | caaggagacc  | aaaatctgt   | tcgacaccac   | aggcagaccc  | aagcccaacg  | 2700 |
| aggacgaccc  | gatcctgacc  | tgcttcaggg  | gctgggtgaa   | gcagctgcag  | atcgactaca  | 2760 |
| agggcaacga  | gatcatgacc  | gcccgtgcca  | gcccagggtct  | gaccaggaaag | ggcgtgtacg  | 2820 |
| ccgtgaggatc | caagggtgac  | gagaacccac  | tgtacgtctc   | caccaggcgag | cacgtgaacg  | 2880 |
| tgtgtgtgc   | caggaccgg   | gacaggatcg  | tgtggaaagac  | cttgcggggc  | gacccttgg   | 2940 |
| tcaagaccc   | gaccgccaag  | tacccggca   | acttcaccgc   | caccatcgaa  | gagtggcagg  | 3000 |
| ccgagcacga  | cccccaccc   | aggcacatcc  | ttggagaggcc  | cgaccccccac | gacgtgttcc  | 3060 |
| agaacaaggc  | caacgtgtgc  | tggggcaagg  | ccctgggtgc   | cgtgtcaag   | accgcgcggca | 3120 |
| tcgacatgac  | cacagacgac  | tggaaacaccg | tggactactt   | cgagacccgc  | aaggccacaca | 3180 |
| gcccggagat  | cgtgtgaac   | cagetgtgeg  | tgaggttctt   | cggcctggac  | ctggacagcg  | 3240 |
| gcctgttcag  | cgccccccac  | gttgcactca  | gcatcaggaa   | caaccaactgg | gacaacagcc  | 3300 |
| ccagccccaa  | catgtacggc  | ctgaacaaa   | agggtgttcg   | cgacgtgacg  | aggcggttacc | 3360 |
| cacagctgcc  | caggccgtg   | gcccacggca  | gggtgtacga   | catgaacacc  | ggcaccctga  | 3420 |
| ggaactacga  | ccccaggatc  | aacctgtgc   | ccgtgaacag   | geggctgccc  | caegccctgg  | 3480 |
| tgtgtcacca  | caacgacgac  | ccacagacgc  | acttcagctc   | tttcgtgagc  | aagctgaaag  | 3540 |
| gcaggacccgt | gtctgtgtgc  | ggcgcagaac  | tgagcgtgc    | cggcgaatgt  | gtggacttgc  | 3600 |
| tgagcagac   | gcccggggc   | acccctccgg  | ccaggctgga   | cctcggcatac | cccgccgacg  | 3660 |
| tgcccaagta  | cgacatcatc  | ttcgtgaacg  | tcaggacccc   | atacaagtac  | caccattacc  | 3720 |
| agcaagtgcg  | ggaccacgc   | atcaagctca  | gcatgtgtac   | caagaaggcc  | tgcctgcacc  | 3780 |
| tgaacccccc  | aggcacctgc  | gtgagcatcg  | gttgcgggtc   | cgccgcacagg | gcccgcggaa  | 3840 |
| gcatcatcgg  | cgccatcgcc  | aggctgtca   | agttcagcag   | ggtgtcaaaa  | cccaagagca  | 3900 |
| gcctggagga  | aaccgggtg   | ctgttctgt   | tcatcggcta   | cgaccggaaag | gcccggaccc  | 3960 |
| acaaccccta  | caagctgacg  | agcaccctga  | caaacatcta   | caccggcagc  | aggctgcacg  | 4020 |
| aggccggctg  | cgccccccac  | taccacgtgc  | tcagggggcga  | tatcgcaccc  | gcccacccggg | 4080 |
| gcgtgtatcat | caacgtgtcc  | aacacgcaagg | gcccggccgc   | aggccggatgt | tgccggccccc | 4140 |
| tgtacaagaa  | gttcccccgag | agtttcgacc  | tgcagcccat   | cgagggtggc  | aaggccaggc  | 4200 |
| tgtgtgaaggg | cgccgcttaag | cacatcatcc  | acggcgttgc   | cccccaacttc | aacaagggtg  | 4260 |
| gcgagggtgg  | aggcgacacg  | cagetggc    | aaggcttacga  | gagcatcgcc  | aggatcgta   | 4320 |
| acgacaataa  | ctacaagacg  | gtggccatcc  | caactgtcg    | caccggcatac | ttcagcggca  | 4380 |
| acaaggacacg | gtgtacccag  | agcttgcacc  | cccccctggac  | accaccgtatc | 4440        |      |
| ccgacgtggc  | catctactgc  | agggacaaga  | agtgggagat   | gaccctgaa   | gaggccgtgg  | 4500 |
| ccaggccgg   | ggccgtggaa  | gagatctgc   | tcagcgcacg   | ctccagctgt  | accgagcccg  | 4560 |
| acggccgacgt | gtgtgggtgc  | caccccaaga  | gctccctggc   | cgccggggaa  | ggctacacga  | 4620 |
| ccagcgtaccc | caagacccctc | agttacttgc  | aggccaccaa   | gttccaccc   | ggccgttaagg | 4680 |
| acatcgccg   | gatcaacgct  | atgtggcccg  | tggccacccg   | ggcccaacgag | cagggtgtca  | 4740 |
| tgtacatctc  | gggcgcacgc  | atgttcacgc  | tcaggagca    | gtgccccctgt | gaggaaaggcg | 4800 |
| aggccacgcac | accaccacgc  | accctggcc   | ccacgcatac   | acacccgaga  | 4860        |      |
| gggtgcacgc  | gtgtacggcc  | agcagccccc  | agcagatcac   | cgtgtcaacg  | tccttccac   | 4920 |
| tgcccaagta  | caggatcacc  | ggcgtcaga   | agatccatgt   | cagccacccc  | atcctgttca  | 4980 |
| gccccaaagg  | gcccgttac   | atccccccca  | ggaagtaactt  | gttgtggaccc | ccacccgtgg  | 5040 |
| acggacaccc  | cgagccaaacg | gcccgcacgg  | agagcaccgg   | gggcacaccc  | gagcagccac  | 5100 |
| cctgtgtatc  | cgaggacgg   | acaaggaccc  | ggacccccc    | gcccacatt   | atcgaggaa   | 5160 |
| aggaagagga  | cagcatcgc   | ctgtgtgc    | acggccccac   | ccaccagggt  | ctgcagggtt  | 5220 |
| aggccacatc  | ccacggccca  | cccaactgt   | ccagctccatc  | ctggagacatc | ccacacgcga  | 5280 |
| gcgacttgc   | cggtggacgc  | ctgacatcc   | ctgacaccc    | ggaggccggcc | agcgtgtaccc | 5340 |
| ceggcgtccac | cagcgtccgg  | accaacagct  | acttcgcac    | gagcatggag  | ttccctggca  | 5400 |
| ggcccggtcc  | agetcggcagg | accgtgttca  | ggaacccacc   | ccacccaggt  | cccaggacca  | 5460 |
| gaccccaagg  | cctgggtcc   | agcggggcc   | gcaacgggg    | cagcgttgc   | agcacccccac | 5520 |
| ccggcgtgaa  | cagggtgtac  | acaggggagg  | aactggggac   | cctgacaccc  | agcaggaccc  | 5580 |
| ccacggcgtt  | cgtgtacgg   | actgtctgg   | tgttcaaccc   | acccggcggt  | aacagggtg   | 5640 |
| tcaccaggga  | ggaattcgag  | gccttcgtgg  | cccagcaaca   | gagacgggtc  | gacgcggcg   | 5700 |
| cctacatctt  | cagcagcgc   | accccccagg  | gacacctgc    | gaaaaaagac  | gtgaggccaga | 5760 |
| cctgtgtgc   | cagggtgtgc  | ctggagggaa  | ccgagcttgc   | aatcagtcac  | gcccccaaggc | 5820 |
| tggaccagga  | gaaggaggaa  | ctgttcaggaa | agaaaaactgc  | gtgtacccc   | accccaaggcc | 5880 |
| acaggagcag  | gtaccagacg  | aggaagggtgg | agaacatgaa   | ggccatcacc  | ggcaggcgga  | 5940 |
| tcctgcagg   | cttgggacac  | tacccatgc   | ccgaggggca   | gttgtggatgc | tacaggaccc  | 6000 |
| tgcacccctt  | gcccactgtac | agtcctccagg | tgaacaggggc  | cttccatcc   | cccaagggtgg | 6060 |
| cctgtgtgc   | ctgcacatgt  | atgttcagg   | agaaacttccc  | caccgtggcc  | agctactgtca | 6120 |
| tcatccctca  | gtacgcacgc  | tacccatgc   | ttgtggaccc   | cgccagctgc  | tgcctggaca  | 6180 |
| ccgcccacgtt | ctggcccgcc  | aagctgagg   | gttcccccac   | gaaacacacg  | tacctggac   | 6240 |
| ccacccatcg  | gagcgtcggt  | ccccatcgcc  | tccagaaacac  | cctgcacac   | gtgtggcccg  | 6300 |
| ctggccacca  | gaggacatgc  | aacgtgacc   | agatgaggaa   | gttgtggcc   | ctggccaccc  | 6360 |
| ctgccttca   | cgtggagtgc  | tccaaagaaat | acgcctgc     | caacgatc    | tgggagaccc  | 6420 |
| tcaaggagaa  | ccccatcagg  | ctgaccgaa   | agaacgtgtt   | gaactacatc  | accaagctga  | 6480 |
| agggcccccac | ggccgtgccc  | ctgttcgcta  | agacccacaa   | cctgaacatc  | ctgcaggac   | 6540 |
| tcccaatgtt  | cagggtcggt  | atggacatgc  | agagggacgt   | gaatgttgc   | cccgccacca  | 6600 |
| agcacacccg  | ggagggcc    | aaagggtc    | tgatccagg    | cgctgacca   | ctggccaccc  | 6660 |
| cctacatgt   | cggtgtccac  | aggaggatgt  | tgaggccgt    | gaacgcgttgc | ctgtgtccca  | 6720 |
| acatccacac  | cctgttcgac  | atgagcgcgg  | aggacttgc    | cgccatcatac | gcccgcac    | 6780 |
| tccagccccc  | cgactgtcg   | ctggagaccc  | acatcgcc     | ttcgcacaa   | agcggaggat  | 6840 |
| acgctatggc  | cctgacccgt  | ctgtgtaccc  | tggaggaccc   | gggcgtggac  | gccgagctgc  | 6900 |
| tcaccctgt   | cgagggtgc   | ttcggcgaga  | tcaatgttccat | ccacccgtcc  | accaagacca  | 6960 |
| agttcaatgtt | cggtgtatgt  | atggaaaggcg | gaatgttcc    | gaccctgttgc | gtgaacaccc  | 7020 |

-continued

|              |             |             |             |             |             |      |
|--------------|-------------|-------------|-------------|-------------|-------------|------|
| tgatcaacat   | tgtgatcgcc  | agcagggtgc  | tgccggagag  | gctgaccggc  | agccctgcg   | 7080 |
| ctgccttcat   | cggcgcac    | aacatcgta   | agggcgtga   | aagcgacaag  | ctgatggccg  | 7140 |
| acaggtgcgc   | cacctggctg  | aacatggagg  | tgaagatcat  | cgacgccgtg  | gtggcgaga   | 7200 |
| aggcccccta   | cttctgcggc  | ggattcatcc  | tgtgcgacag  | cgtgaccggc  | accgcctgca  | 7260 |
| gggtggccga   | ccccctgaag  | aggctttca   | agctgggca   | gccactggcc  | gctgacatg   | 7320 |
| agcacgcacg   | tgacaggcgg  | agggccctgc  | acgagggaaac | caccagggtgg | aacagggtgg  | 7380 |
| gcatccttag   | cgagctgtgc  | aaggccgtgg  | agagcaggta  | cgagaccgtg  | ggcaccagca  | 7440 |
| tcatectgtat  | ggctatgacc  | acactggcca  | gctccgtcaa  | gagcttctcc  | tacctgaggg  | 7500 |
| ggcccccata   | aactcttcac  | ggctaaccctg | aatggactct  | gacatagtct  | agtccgcaca  | 7560 |
| ggccgcacc    | atgagactga  | cagccccatg  | aaccttactg  | cttctgtt    | ggggagctgt  | 7620 |
| tgotctgaca   | gagacatggg  | ctggatcta   | ccacagcccc  | agctacggct  | accaccaatg  | 7680 |
| cgagaggggg   | ggaggaggct  | ccggggagg   | aggctccctg  | aagatcagcc  | aggeccgtgca | 7740 |
| cgcgcgcac    | gcccgcac    | acgaggccgg  | ccggggagggt | atcgccggca  | ttgtcgctgg  | 7800 |
| cetggcgctc   | tcgcgcgtgg  | tggtgatgg   | agctgtgttc  | gcagctgtta  | tgtgcagaag  | 7860 |
| aaagtcatcc   | ggcggaaagg  | gaggctcta   | ctctcaggct  | gcttcgtcta  | cagtgcctag  | 7920 |
| agctttagt    | tgtttatctc  | agctggccgg  | cggaggccgc  | gactacaagg  | acgacgtatg  | 7980 |
| caagtaaacact | cgagatgtt   | acgtgcacaa  | gtgattgtca  | ccccccaaaa  | gaccatattt  | 8040 |
| tgacacaccc   | tcatgtatc   | gccccaaat   | ttacagccgc  | ggtgtcaaaa  | accgcgtgga  | 8100 |
| cgtggttaac   | atccctgtct  | ggaggatca   | ccgtaattat  | tataattggc  | ttgtcgctgg  | 8160 |
| ctactattgt   | ggccatgtac  | gtgctgacca  | accagaaaca  | taattgaata  | cagcagcaat  | 8220 |
| tggcaagctg   | ttcatataga  | actcgcggcg  | atggcatgc   | ccgcctaaaa  | tttttattt   | 8280 |
| attttttctt   | ttctttccg   | aatcgattt   | tgtttttat   | atttcaaaaa  | aaaaaaaaaa  | 8340 |
| aaaaaaaaaa   | tctagaaaaaa | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | 8400 |
| aaaaaaaaaa   | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaa      | 8455 |

SEQ ID NO: 109            moltype = AA length = 2512  
 FEATURE                    Location/Qualifiers  
 REGION                    1..2512  
 note = Synthetic polypeptide  
 source                    1..2512  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 109  
 MEKPVVNVDV DPQSPFVVQL QKSFPQFEVVAQQVTPNDHA NARAFSHLAS KLIELEVPTT 60  
 ATILDIGSAP ARRMFSEHQY HCVCPMRSPB DPDPRMMKYAS KLAEKACKIT NKNLHEKIKD 120  
 LRTVLDTPTDA ETPSLCFHND VTCNMRAEYS VMQDVYINAP GTIYHQAMKG VRTLYWIGFD 180  
 TTQFMFSAMA GSYPAYNTNW ADEKVLEARN IGLCSTKLSE GRTGKLSIMR KKELKPGSRV 240  
 YFSVGSTLTPY EHRSLSQSWH LPSVFHLNGK QSYTCRCDTV VSCEGYVVKK ITISPGITGE 300  
 TVGYAVTHNS EGFLLCKHNS TVKGVERSPF VCTYIPATIC DQMTGIMATD ISPDDAQKLL 360  
 VGLNQRIVIN GRTNRNTNTM QNYLLPIIAQ GFSKWAKERK DDLDNEKMLG TRERKLTYGC 420  
 LWAFRTKVKH SFYRPPGTQT CVKVPASFSF FPMSSVWTT LPMSLRQKLK LALQPKKEEK 480  
 LLQVSEELVM EAKAAFEDAQ EEARAELKR ALPPLVADKG IEAAAEEVVC EVEGLQADIGA 540  
 ALVETPRGHV RIQPQANDRM ILQYIVVSPN SVLKNKALAP AHPLADQVKI ITHSGRSGRY 600  
 AVEPYDAKVL MPAGGAVPWP EFLALSESAT LVYNEREFVN RKLYHTAMHG PAKNTEEEQY 660  
 KVTKAELAET EYVFDVDKKR CVKKEEASGL VLSGELENTPP YHELALEGLK TRPAVPYKVE 720  
 TIGVIGTPGS GKSAAIKSTV TARDLVTSGK KENCREIEAD VLRLRGMQIT SKTVD SVMLN 780  
 GCHKAVEEVLY DEAFACHAG ALLALIAIVR PRKKVVLCGD PMQCGFFNMM QLKVFHNHPE 840  
 KDICTKTFYK YISRRCQPV TAIYSTLHYD GKMKTTPCK KNIEIDITGA TKPKPGDIIIL 900  
 TCFRGWVQKL QIDYPGHEVM TAAASQGLTR KGVYAVRKV NENPLYAITS EHVNVLLTRT 960  
 EDRLVWKTQG QDPWIKQLTN IPKGNFQATI EDWEAEHKGN IAAAINSPTPR ANPFSCKTNV 1020  
 CWAKALEPIL ATAGIVLTGC QWSLEFPQFA DDKPHSAIY LDVIC1KFEG MDLTSGLFSK 1080  
 QSIPLTYHPA DSARPVAHWI NSPGRTRKYGY DHAIAAELSR RFPVFQLAGK GTQLDLQTGR 1140  
 TRVISQAQHNL VPVNRLNPHA LVPEYKEKQP GPVEKFLNQF KHHSVLLVSE EKIEAPRKRI 1200  
 EWIAPIGIAQ ADKNYNLAFG FPPQARYDLV FINIGTKYRN HHFQQCEDHA ATLKTLSRSA 1260  
 LNCLNPGGTL VVKSYVYADR NSEDVVTAL RKFVRVSAAR PDCVSSNTEM YLIFRQLDNS 1320  
 RTRQFTPWHY NCVISSVYEG TRDGVGAAPS YRTKRENIAD CQEAEVNAA NPLGRPEGV 1380  
 CRAIYKRWPT SFTDSATETG TARMTVCLGK KVIHAVGPDF RKHPEAEALK LLQNAYHAVA 1440  
 DLVNEHNIKS VAIPLLSTG YAAAGKDRLEV SLNCLTTALD RTDADTVIYC LDKKWKERID 1500  
 AALQLKESVT ELKDEDMEID DELVWVHPDS CLKGRKGFT TKGKLYSYFE GTKFHQAOKD 1560  
 MAEIKVLFPN DQESNBQLCA YILGETMEAI REKCPVDHNP SSSPPKTLPC LCMYAMTPER 1620  
 VHRLRSNNVK EVTVCSSTPL PKHKKKNVQK VQCTKVVLFN PHTPAFVPAR KYIEVPEQPT 1680  
 APPAQAEAAP EVVATPSPST ADNTSLDVT ISLDMDDSSE GSLFSSFSGS DNSITSMDSW 1740  
 SSGPSSLEIV DRRQVWVADW HAVQEPAPIP PPRLLKKMARL AAARKEPTPP ASNSSESLSHL 1800  
 SFGGVSMMSLG SIFDGETARQ AAVOPLATGP TDVPMFSFGSF SDGEDELRLR RVTESEPVLF 1860  
 GSFEPGEVNS IISSRASVSP PLRKQRRRRR SRRTEYLTGV GGYIFSTDG PGHLQKKSVL 1920  
 QNQLTEPTLE RNVLERIHAP VLDTSKEEQL KLYQMMPTE ANKSRYOSRK VENQKAITTE 1980  
 RLLSGLRLYNN SATDQPECYK ITYPKPLYSS SVPANYSDPQ FAVAVCNLYN HENYPTVASY 2040  
 QITDEYDYLN MDDGTVACL DTATFCPAKL RSYPKKHEYR APNIRSAVPS AMQNTLQNVL 2100  
 IAATKRNCNV TQMRELPTLD SATFNVECFR KYACNDEWEY EFARKPIRIT TEFVFTAYVAR 2160  
 LKGPKAAALF AKTYNLVPLQ EVPMDRFVMD MKRDVKVTPG TKHTEERPKV QVIQAAEPLA 2220  
 TAYLCGITHRE LVRRLLTAVLL PNIHTLFDMG AEDFDAIIAE HFKQGDGVLE TDIASFDSQ 2280  
 DDAMALTGLM ILEDLGVDQP LLDLIECARG EISSTHLPITG TRFKFGAMMK SGMLTFLVN 2340  
 TVLNNVVIASR VLEERLKTSL CAAFIGDDNI IHGVVSDKEM AERCATWLNM EVKIIDAVIG 2400  
 ERPPYFCGGF ILQDSTVSTA CRVADPLKRL FKLGKPLPAD DEQDEDRRAA LLDETKAWFR 2460  
 VGITGTЛАVA VTTRYEVNDI TPVLLALRTF AQSXRFAQAI RGEIKHLYGG PK 2512

SEQ ID NO: 110            moltype = AA length = 9  
 FEATURE                    Location/Qualifiers  
 REGION                    1..9

-continued

---

```

source          note = Synthetic polypeptide
               1..9
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 110
SPSYVYHQF                                              9

SEQ ID NO: 111      moltype = AA  length = 9
FEATURE          Location/Qualifiers
REGION           1..9
source          note = Synthetic polypeptide
               1..9
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 111
SPSYAYHQF                                              9

SEQ ID NO: 112      moltype = AA  length = 9
FEATURE          Location/Qualifiers
REGION           1..9
source          note = Synthetic polypeptide
               1..9
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 112
TPHPARIGL                                              9

SEQ ID NO: 113      moltype = RNA  length = 1701
FEATURE          Location/Qualifiers
misc_feature     1..1701
source          note = Synthetic polynucleotide
               1..1701
               mol_type = other RNA
               organism = synthetic construct
SEQUENCE: 113
atgaaggcta tcctgggtg gctgctctac acctttgc caagccaatgc tgacaccctg  60
tgtattggct accatgccaa caaacgcaca gacacagtgg acacagtgtt ggagaagaat 120
gtgacagtgta cccactctgt gaacctgtt gaggacaaac acaatggccaa actgtgtaaa 180
ctgagggggag tggctccact gcacccggc aagtgtaaaca ttgctggctg gattctggc 240
aaccctgagt gtgagtcctc gagcacagcc tccctctggt cttacatgtt ggagcaccca 300
tcctctgaca atggcactt ttaccctgga gacttcattt actatgagga actgagggaa 360
caacttctt ctgtgtcctc ctggagattt ttccaaagac ctccctctgg 420
ccaaaccatg acgaacaaca gggagtgaca gcacccgttc cacatgtgg agccaagtcc 480
tttacaaaga acctgtattt gctgtgaag aaggccaact cttacccaaa actgagcaag 540
tcctacatca atgacaaggg caaggagggt ctggtgctgt ggggcatttccca ccaccaagc 600
acccctgtc accaacaatgc ccttcattttc aatgtgtttt gttgggttcc 660
agcagataca gcaagaagtt caagecttag attgccccatca gaccaaaaatgg ggggatcatg 720
gagggcagga tgaactacta ctggaccctg gtgaaacctt gagacaatgtt accctttgg 780
gtcacaggca acctgtgg gccaagatgtt gctttgttca tggagggaa tggctggctt 840
ggcatcatca ttctggacac acctgtccat gactgttaaca ccactgttca gacaccaaag 900
ggagccatca acacccctt gccatccatca aacatccaccat caatccatcat tggcaagtt 960
ccaaaatatg tcaagagcac caaaatggaa ctggcttacat gactgaggaa catcccaagc 1020
atccagagca ggggactgtt tggaccatt gctggcttca ttgagggggg ctggacacggg 1080
atgggtggat gtgttatgg ctaccacca cagaatgttca aagggtttttt ctatgtgtct 1140
gaccgttata gcaaccggaa tgccatgtt gatgttaccatca acaagggttca ctctgttatt 1200
gagaatgtt acacccatgtt cacacgtt ggcacccatgtt tcaaccatgtt ggagaaggagg 1260
attgagaacc tgaacaagaa ggtggatgtt ggcttcctgg acatctggac ctacaatgtt 1320
gaactgttgg tgctgttggaa gaatgttgggg accctgtggat accatgttca caatgttgg 1380
aacctctatg agaagggttgg gatgttggat gatgttggat gatgttggat tggcaatggc 1440
tgttttgtt gtttaccatca gatgttggat gatgttggat gatgttggat gatgttggat 1500
tatgtactacc caaaatactt tgaggaggat aaactgttca gggaggat gatgttggat 1560
aaattggaga gacccaggat ttaccatca ctggccatctt acagccatgtt ggcacccatgtt 1620
ctgggtgttgg tggttggat gggccatctt agtggatgtt gatgttggat gatgttggat 1680
cagtgcggat tctgttccatca a 1701

SEQ ID NO: 114      moltype = AA  length = 566
FEATURE          Location/Qualifiers
REGION           1..566
source          note = Synthetic polypeptide
               1..566
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 114
MKAILVVLLY TFATANADTL CIGYHANNST DTVDTVLEKN VTVTHSVNLL EDKHNGLCK  60
LRGVAPLHLG KCNIAGWILG NPECESLSTA SSSWYIVETP SSDNGTCYPG DFIDYEELRE 120
QLSSVSSFER FEIFPKTSSW PNHDSNKVT AACPHAGAKS FYKNLIIWVK KGNSYPKLSK 180
SYINDKGKEV LVLWGIHHPS TSADQQSLYQ NADAYVFVGS SRYSKKFKPE IAIRPKVRDQ 240
EGRMNNYYWTL VEPGDKITFE ATGNLNVVPRY AFAMERNAGS GIIISDTPVH DCNTTCQTPK 300
GAINTSLPFQ NIHPITIGKC PKYVKSTKL LATGLRNIPS IQSRGLFGAI AGFIEGGWTG 360

```

-continued

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| MVDGWYGYHH | QNEQGSGYAA | DLKSTQNAID | EITNKVNSVI | EKMNTQFTAV | GKEFNHLEKR | 420 |
| IENLNKKVDD | GFLDIWTYNA | ELLVLLENER | TLDYHDSNVK | NLYEKVRSQL | KNNAKEIGNG | 480 |
| CFFFYHKCDN | TCMESVKNGT | YDYPKYSEEA | KLNREEIDGV | KLESTRIYQI | LAIYSTVASS | 540 |
| LVLVVSLGAI | SFWMCSNGSL | QCRICI     |            |            |            | 566 |

SEQ ID NO: 115      moltype = RNA length = 9911  
 FEATURE      Location/Qualifiers  
 misc\_feature      1..9911  
 note = Synthetic polynucleotide  
 source      1..9911  
 mol\_type = other RNA  
 organism = synthetic construct

SEQUENCE: 115

```

gataggccgc gcatgagaga agccccagacc aattacctac ccaaataatggaa gaaagtccac 60
gttgacatcg aggaagacag cccatccctc agagcttgc acgggagctt cccgcaggtt 120
gaggtagaag ccaagcagggt cactgataat gaccatgctta atgcccagagc gtttcgcatt 180
ctggcttcaa aactgatcga aacggagggt gaccctgg acacgatctt tgacatttgg 240
atagtccatc tagtacattt catctgacta taatcacaaac accacccacca tgaatagagg 300
attctttaac atgctccggcc gccccccctt cccggcccccc actgcattgtt ggaggcccg 360
gagaaggagg caggcgccgc cgaaaaaggcg agctactaaac ttccgccttc tgaagccaggc 420
tggagacgtg gaggagaacc ctggacatcg gaaaaggccatc gttgcacatcg aggaagacag 480
ccattccctc agagcttgc acgggagctt cccgcagggtt gaggtagaaag ccaagcagg 540
cactgataat gaccatgctta atgccaggac gtttcgcatt ctggcttcaa aactgatcga 600
aacggagggt gaccatcccg acacgatctt tgacatttggaa agtgcgcggc cccgcaggaaat 660
gtattctaaag cacaagtatc atttgtatcg tccgcattgtt tggcgccaaatc atccggacac 720
atttgttataag tatgcacta acgtcaagaa aactgttaag gaaataactcg ataaaggaaat 780
ggacaagaaa atgaaggacg tcgcggccgtt catggccgc acggccctgg aaactcgagac 840
tatgtgcctc cacgcgcacg agtgcgtgtcg ctacgaagggg caagtcgtgtt tttaccagg 900
tgtatcacccg gttgcggcgc cgacaaatgtt ctatcaccaaa gccaataagg gagtttaggt 960
cgccatctgg ataggatcttgc acacccacccctt tttatgtttt aagaacttgg ctggaggcata 1020
tccatcatac ttatccaaact gggccgcacca aaccgtgttta acggccgttta acataggccct 1080
atgcagctct gacgttatgg acgggttcacg tagagggatg tccattctta gaaaaggatc 1140
tttggaaacca tccaacaaatg ttcttatttc tggggctcgg accatctacc acggagaagag 1200
ggacttacttgc agaggatggc acctgcggccg ttttttttttccatcg ttttttttttccatcg 1260
cacatgttgg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 1320
cagtccaggc ctgtatggga agccctcagg ctatgtgtgtt acgtatgcacc gcggaggatt 1380
cttgcgtgtgc aaagtgcacag acacattggaa cggggagagg gtctcttttcccggtgtgcac 1440
gtatgtgcacatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 1500
ggacgcacgcg cccaaactgc tggggggctt ccacccatcg atatgtgtgtca acgggtgcac 1560
ccagagaaac accaataccatc tggggggctt ccacccatcg atatgtgtgtca acgggtgcac 1620
taggtggggca aaggatataa aggaagatca agaagatgaa aggccacttag gactacgaga 1680
tagacagtttgc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 1740
taagcgcggc gataccaaaaatccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 1800
ggccaggatata ggcgttacaca cattttggat cggggcttggaa acaagaatca ggaaaatgtt 1860
agaggagcac aaggagccgtt cacccatccatc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 1920
cgccatctgg gaggatgttgc agtgcgtgttgc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 1980
tttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2040
gggtggggcc ggttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2100
cgaggacaatcg atccgttgc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2160
attatcttgc accatccctc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2220
aggccatctgg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2280
accatccctc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2340
gttgcgttacatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2400
agaatattac aaaactgttc acggccatcg gacccatcg gacccatcg gacccatcg gacccatcg 2460
caggaaacccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2520
ggatcccttc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2580
ccaaatccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2640
aaacccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2700
tataaggggac gtcggatggaaatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2760
gttgcgttacatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2820
tcatgttgc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2880
cgccatctgg aacccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 2940
cgatgttgc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3000
ttccgttgc ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3060
taagatttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3120
tttggggccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3180
ggcgttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3240
tggggccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3300
ggccatctgg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3360
gttgcgttacatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3420
gaggccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3480
tttggggccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3540
atggggccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3600
ccaaatccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3660
tttggggccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3720
gttgcgttacatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3780
gcttgcgttacatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3840
aaacccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3900
ccacccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg ttttttttttccatcg 3960

```

-continued

---

|                           |                         |                                      |      |
|---------------------------|-------------------------|--------------------------------------|------|
| cggggaaaag ttgtccgtcc     | caggaaat ggttactgg      | ttgtcagacc ggctgaggc                 | 4020 |
| tacccatcaga gtcggctgg     | atttaggcatt cccagggtat  | gtgccaat atgacataat                  | 4080 |
| atttgttaat gtgaggaccc     | catataaata ccatcaat     | cagcagtgt aagaccatgc                 | 4140 |
| cattaagctt agcatgttga     | ccaagaaago ttgtctcgat   | ctgaatecccg gcggaaactg               | 4200 |
| tgtcagcata ggttatgtt      | acgtcgacag ggccagcgaa   | agcatatcg gtgtatagc                  | 4260 |
| gcggcagttc aagtttccc      | gggtatgc aaaggaaatcc    | tcacttgaag agacggaaat                | 4320 |
| tctgtttgtt ttcattgggt     | acgatcgaa ggcccgtac     | cacaatcctt acaagcttc                 | 4380 |
| atcaacccctg accaacattt    | atacagggtt cagactccac   | gaagccggat gtgcacctc                 | 4440 |
| atatacatgtt gtgcgggggg    | atattgcac ggcacccggaa   | ggagtgtatataa ataatgtc               | 4500 |
| taaacagcaa ggacaaatcg     | gccccgggtt gtgcgggggg   | ctgttataaga aattccggaa               | 4560 |
| aaaccttgcat ttacacgcg     | tcgaaggtagg aaaaagcgca  | ctggtaaag gtgcagctaa                 | 4620 |
| acatatcatt catgcgttag     | gaccaaactt caacaaatgt   | tcggagggtt aaggtgacaa                | 4680 |
| acagttggca gaggtttatg     | tgactatcgta taagattgtt  | aacgataaca attacaatgc                | 4740 |
| atgtcgattt ccactgtgt      | tgactatcgat cttttccggg  | aacaaagatc gactaaacca                | 4800 |
| atcatttgcac cattttgttgc   | ccacccgtat cttttccggg   | ccatataactg                          | 4860 |
| caggggacaag aaatggaaa     | tgactctaa ggaagcgtg     | gtctggagag aagcgttgg                 | 4920 |
| ggagatatgc atatccgcac     | actcttcgtt gacagaacat   | gtatcgacgc tggtgagggt                | 4980 |
| gcacatccaaag agtttcttgg   | cttgcggggaa gggctacagg  | acaagegttgc gcaaaactt                | 5040 |
| ctcatatgg gaaggacca       | atgttccaca ggcggccaa    | gatatacgaa aatattatgc                | 5100 |
| catgtggccc gttcaacgg      | aggccaaatga gcaggatgtc  | atgtatatcc tcggagaaag                | 5160 |
| catgagcagt attagggtc      | aatggccctg cgaaggatgtc  | gaagccttca caccacccat                | 5220 |
| cacgtctgcg tgcgtgtca      | tccatgcctt gactccggaa   | agatgtacgc gcctaaaagc                | 5280 |
| ctcacgttca gaacaaatcc     | tcgtgtgtc tatcttccca    | tcggcaatgtt atagaatcac               | 5340 |
| ttgtgtgcgt aagatccat      | gctcccgatc tatattgttc   | tcaccggaaag tgccctgcgt               | 5400 |
| tatttcattca aggaagtatc    | tcgtggaaatcc accacccgtt | gacgagactc cggaggccat                | 5460 |
| ggcagagaaac caatccacag    | agggtggacaccat          | tgaacaaacca coacttataa ccggaggatga   | 5520 |
| gaccaggactt aacacgcgt     | ccggccatcat             | catcgaagag gaagaagagg atagcataag     | 5580 |
| tttgctgtca gatggccgg      | cccaccaggat             | gtcgcaatgc gaggccgacca ttacccggcc    | 5640 |
| gcctctgtta tctatgtcat     | cctgttccat              | tcctcatgc tccgacttgc atgtgacag       | 5700 |
| tttattccat ctgtacaccc     | ttggggggatc tagcgtgttgc | agcggggccaa cgtcagccgt               | 5760 |
| gactaactct tacttgc        | agatgttgcgat            | gtttcttgcg ccacccgttgc ctgcgcctcg    | 5820 |
| aacagttttc aggaacccctc    | cacatcccgat             | tcggcgacaa agaacacccgt cacttgcacc    | 5880 |
| cagcaggggc tgctcgagaa     | ccagcgttgc              | ttccaccccg ccaggcgttgc atagggtgtat   | 5940 |
| caactagatgg gagctcgagg    | cgcttaccatc             | gtcactcgact cttcgttgcgat cgggttgc    | 6000 |
| aaccaggctt gtctccaaacc    | cgccggcgtt              | aaatagggtt attacaagag aggaggttga     | 6060 |
| ggcggttgcgtt gacacaaca    | aatgtcggtt              | tgatcgccgtt gcatacatct tttctccgt     | 6120 |
| caccggctaa gggcattttac    | aaacaaaatc              | agtaaggca acgggtctat ccgaagtgtt      | 6180 |
| tttgcgttgcg aggcaatttg    | agatgttgc               | tgccccggcc ctcgaccaag aaaaagaaga     | 6240 |
| attactacgc aagaaatccat    | agtttacatcc             | acacacccgtt aacagaacca gataccatgc    | 6300 |
| caagaaatgg gagaacatca     | aaaggccat               | agcttagactt attctgcac gcttggggca     | 6360 |
| ttatgttgcg aagacacata     | ttttttgcat              | tttgcgttgcg attttttgcg acagggttgc    | 6420 |
| ttcatctgtt gtgaaacccgt    | ccttgcgttgc             | cccccgggtt gcagttggaaag cctgttacgc   | 6480 |
| catgttggaa gagaactttc     | tcactgtgttgc            | ttcttactgtt attattccat agtacatgtc    | 6540 |
| cttatttgcg atgggttgc      | gagtttgc                | ctgtttttgcg actgttgcgtt tttgccttc    | 6600 |
| aaagctgcgc agcttccaa      | agaaaacactc             | ctattttggaa cccacaatcc gatccggcgt    | 6660 |
| gccttcacgc atccacaa       | cgccggccat              | cgccggccatccaa aaagaaatgc            | 6720 |
| caatgttcaagc caatgttgc    | aaatgttgc               | attttttgcg attttttgcg attttttgcg     | 6780 |
| cttcaagaaa tactgtgttgc    | ataatgttgc              | tttggggaaatc tttaaaatcc accccatccat  | 6840 |
| gtttactgtt gaaaacgtgg     | taaatttgc               | ttttttttgcg aatgttgcgttcc gatccggcgt | 6900 |
| tcttttttgcg aagacacata    | attttttgc               | ttttttttgcg aatgttgcgttcc gatccggcgt | 6960 |
| aatggactta aagagatgg      | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7020 |
| caaggatcgat gtgttgc       | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7080 |
| cccgagatcg ttgttgc        | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7140 |
| ttatgttgcgtt gaaatggat    | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7200 |
| tctggaaactt gacatcggt     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7260 |
| gtttatgttgcg ttgttgc      | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7320 |
| tttcggggaa atttcatca      | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7380 |
| gtatgtatgc ttgttgc        | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7440 |
| aaggatcgat ttgttgc        | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7500 |
| caatatcgat aagggatcg      | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7560 |
| gaatatggaa gtcacatca      | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7620 |
| agggtttttt ttgttgcgtt     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7680 |
| aaggctgtt gaaatggat       | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7740 |
| aaggccatcgat cgatcgat     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7800 |
| caaggccatcgat gatcgat     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7860 |
| tactctatgtt gacatcgat     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7920 |
| cgccatccatcgat gatcgat    | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 7980 |
| tgccaaaaac ataaatggaa     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8040 |
| cgacgcgttgc tccatccatcgat | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8100 |
| cgacgcacat atcgatcgat     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8160 |
| ggcagaaatcgat gatcgat     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8220 |
| gaatatggaa gtcacatca      | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8280 |
| cccagcttcaac gacatcgat    | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8340 |
| caccgttgcgtt ttcgttgc     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8400 |
| accgatcgat caaaatggaa     | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8460 |
| catgttacacc ttgttgcgtt    | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8520 |
| cgagatcgat gacccggacca    | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8580 |
| atttgcgttgc ggcgttgc      | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8640 |
| cgaccatccatcgat gatcgat   | ttttttttgcg             | ttttttttgcg aatgttgcgttcc gatccggcgt | 8700 |

-continued

|                                                                          |      |
|--------------------------------------------------------------------------|------|
| tcaccacggc ttcggcatgt tcaccacgt gggacttgc atctggggct ttccgggtcg          | 8760 |
| gctcatgtac cgcttcgagg aggagctatt cttgcgcgcg ttgcagaact ataagattca        | 8820 |
| atctgcctg ctgggtcccc cactatttg cttcttcgcg aagagactc tcatacgacaa          | 8880 |
| gtacgaccta agcaacttgc acgagatcg cagcgggggg gcgcgcgtca gcaaggaggt         | 8940 |
| agggtgagcc gtggccaaac gcttcaccc accaggcatc cgacaggctc acggcctgac         | 9000 |
| agaaacaacc agcgccatc tgatcacccc cgaaggggac gacaaggctg ggcgcgttag         | 9060 |
| caagggttgtt cccttttcg aggtaagggt gttggacttgc gacacccgtg agacacttgg       | 9120 |
| tgtgaaccag cgccggcgacg tgtgcgtccg tggccccatg atcatgagcg gctacgttaa       | 9180 |
| caaccccgag gotacaaccg ctctccatcg caaggacggc tggctgcaca gcggcgacat        | 9240 |
| cgcctacttgc gacggagacg agcacttgc catcggtggac cggctgaaat ccctgtatcaa      | 9300 |
| atacaagggc taccatgtac ccccgccga acttggagacg atccctgtgc aacacccaa         | 9360 |
| catcttcgac gccggggctcg cggcgtgcg cgacgcacat gccggcgagc tgccgcgc          | 9420 |
| agtgcgttgtt ctggaaacacg gtaaaccatc gaccggaaatc gagatcggtt actatgtggc     | 9480 |
| caggcagggtt caaacccggc ayaagtcgac cgggtgggtt gtgttcgtgg acgggtggc        | 9540 |
| taaaggactg accggcaagt tggacggccg caagatccgc gagatctca ttaaggccaa         | 9600 |
| gaaggggcgc aagatgcgcg tgtaaggcgc gccgtttaaa cggccggcct taattaagta        | 9660 |
| acgatacagc acgaatttgcg aagctgttca catagaactc gccggcgatg gcatgccc         | 9720 |
| ttaaaaatttt tattttttt tttccgttac ggatttttt tttatattt                     | 9780 |
| aaaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa | 9840 |
| aaaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa | 9900 |
| aaaaaaaaaaaa a                                                           | 9911 |

SEQ ID NO: 116 moltype = RNA length = 2117  
 FEATURE Location/Qualifiers  
 misc\_feature 1..2117 note = Synthetic polynucleotide  
 source 1..2117 mol\_type = other RNA organism = synthetic construct  
 SEQUENCE: 116  

|                                                                       |      |
|-----------------------------------------------------------------------|------|
| agggaaactta agtcaacaca acatatacaa aacaaacgaa tctcaagcaa tcaaggattc    | 60   |
| taatttctatt gcagaattt aaatcatttc ttttaaagca aaagcaattt tctgaaaattt    | 120  |
| ttcaccattt acgaacgata gcccacatga aggctatctt ggtgggtctg ctctacaccc     | 180  |
| ttgcacacagc acatgtgtac accctgtgttac ttggcttacca acatggaaacaa agcaca   | 240  |
| cagtggacac agtggggatc aagaatgtga cagtggccca ctctgtgaaac ctgtggagg     | 300  |
| acaaacacaa tggcaactg tgtaaactga ggggagtgcc tccactgcac ctggggcaagt     | 360  |
| gtAACATTGC TGGCTGGATT CTGGGCAACC CTGAGTGTGA GTCCCTGAGC ACAGCCTCCT     | 420  |
| CCTGGCTTA CATTGGGAG ACACCATCTT CTGACAATGCG CACTTGTTCAC CCTGGAGACT     | 480  |
| TCAATTGACTA TGAGGAACTG AGGGAAACAC TTTCTCTGTG GTCTCTTCTT GAGAGTTG      | 540  |
| AGATTTTCC AAAGACCTC TCCTGGCAA ACCATGACAG CAACAAGGGA GTGACAGCAG        | 600  |
| CCTGTCCACA TGCTGGAGCC AAGTCCTCTT ACAAGAACCTT GATTGGCTG GTGAAGAAGG     | 660  |
| GCAACTCTA CCCAAACACTG AGCAAGTGC ACATCAATGCA CAAGGGCAAG GAGGTGCTGG     | 720  |
| TGCTGTGGGG ATCCACACAC CCAAGCACCT CTGTCGACCA CTGCTCCCTC TACCAAGATG     | 780  |
| CTGACGCGTA TGTGTTTGTG GGCTCCAGCA GATACAGCAA GAAGTCAAG CCTGAGATTG      | 840  |
| CCATCAGACCC AAAGGTGAGG GATCAGGAGG GCAGGATGAA CTACTACTGG ACCCTGGTGG    | 900  |
| AACCTGGAGA CAAGGATACC TTGGGGATCA CAGGCAACCTT GGTGGTGCCA AGATATGCC     | 960  |
| TTGCTATGGAGA GGGCTCTGGCA TCTACATCTC TGACACACCT GTCCATGACT             | 1020 |
| GTAAACACAC TTGTCAGACA CCAAAGGGAG CCATCAACAC CTCCCTGCCA TTCCAGAACAA    | 1080 |
| TCCACCCAAAT CACCATTGGC AAGTGTCCAA AATATGTCAA GAGCACCAAA CTGAGACTGG    | 1140 |
| CTACAGGACT GAGGAACATC CCAAGCATCC AGAGCAGGGG ACTGTTTGA GCCATTGCTG      | 1200 |
| GTCTCATGGA TGGGGGGCTGG AACGGGGATG TTGGATGGCTG GTATGGCTAC CACCACAGA    | 1260 |
| ATGAACACGGG CTCTGGCTAT GCTGTCGACCG TGGGGGGCTGG AACGGGGATG TTGGATGGCTG | 1320 |
| TTACCAACAA GGTGAACTCT GTGATTGAGA AGATGAACAC CCAGTTCACA GCAGTGGCA      | 1380 |
| AGGAGTTCAC CCACTGGAG AAGGGGGATG AGAACCTGAA CAAGAAGGTG GATGATGGCT      | 1440 |
| TCCTGGACAT CGGGACATAC AATGTCGACG TGCTGGTGTG TTGGAGAAAT GAGGGACCC      | 1500 |
| TGGACTACCA TGACAGCAAT GTGAAGAACCTT TCTATGAGAA GGTGGGGAGC CAACCTAAAA   | 1560 |
| ACAATGCCAA GGAGATTGGC AATGGCTTT TTGAGTTCTA CCACAAGTGT GACAACACTT      | 1620 |
| GTATGGAGTC TGTGAAGATA GGCACCTATG ACTACCCAAA ATACTCTGAG GAGGCTAAAC     | 1680 |
| TGAACAGGGA GGAGGGATG GGAGTGAAT TGGAGAGCAG CAGGATTAC CAGATCTGG         | 1740 |
| CCATCTCAC CACCGTGGCC AGCACCTGG TGCTGGTGTG GAGCCTGGC GCCATCAGCT        | 1800 |
| TCTGGATGTG CAGCAACGGC AGCTTGCAGT GCAGGATCTG CATCTAAACT CGAGCTAGT      | 1860 |
| ACTGACTAGG ATCTGGTTAC CACTAAACCA GCTCTCAAGAA CACCCGAATG GAGTCCTCAA    | 1920 |
| GCTACATAAT ACCAAACTAC ACTTACAAA TGTTGTCCCC CAAATGTAG CCATTCTGTAT      | 1980 |
| CTGCTCTAA TAAAGAGAAA TTGTTCTCAC ATTCTAGAAA AAAAAGAAAA AAAAAGAAAAA     | 2040 |
| AAAAAAAAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAAA   | 2100 |
| AAAAAAAAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAAA   | 2100 |
| AAAAAAAAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAA AAAAAGAAAAA   | 2117 |

SEQ ID NO: 117 moltype = DNA length = 10508  
 FEATURE Location/Qualifiers  
 misc\_feature 1..10508 note = Synthetic polynucleotide  
 source 1..10508 mol\_type = other DNA organism = synthetic construct  
 SEQUENCE: 117  

|                                                                       |     |
|-----------------------------------------------------------------------|-----|
| atggggccgcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcaccg   | 60  |
| ttgacatcgaa ggaagacagg ccattccctca gagctttgcg gcggagctc ccgcagttt     | 120 |
| aggtagaaacg caagcagggtc actgtataatg accatgtctaa tgccagagcg ttttcgc    | 180 |
| ttggcttcaaa actgtatcgaa acggaggtgg accatcccgaa cagcatccctt gacattggaa | 240 |

-continued

---

|              |             |            |            |             |              |      |
|--------------|-------------|------------|------------|-------------|--------------|------|
| gtgcggccgc   | ccgcagaatg  | tattctaago | acaagtatac | ttgttatctgt | ccgtatggagat | 300  |
| gtgcggaga    | tccggacaga  | ttgtataagt | atgcaactaa | gctgaagaaa  | aactgttaagg  | 360  |
| aaataactga   | taaggaattt  | gacaagaaaa | tgaaggagct | ggccgcggc   | atgagcgacc   | 420  |
| ctgacacctga  | aactgagact  | atgtgcctc  | acgacgacga | gtcgatcg    | tacgaaggc    | 480  |
| aagtgcgtgt   | tttaccaggat | gtatacgcgg | tcgacggccc | caccagctg   | taccaccagg   | 540  |
| ccaacaaggg   | cgtgagggtg  | gctactgg   | tcggcttgc  | caccacaccc  | ttcatgttca   | 600  |
| agaaacctgc   | ccggcgcctac | ccccactaca | gacccaactg | ggccgacgg   | accgtgtca    | 660  |
| ccgcaggaa    | catcgccctg  | tgccagcagc | acgtatgg   | gaggagccgg  | agaggatgt    | 720  |
| gcatctcg     | gaagaataac  | ctgaagccca | gcaacaact  | gtgttcagc   | gtgggcagca   | 780  |
| ccatctacca   | cgagaaggagg | gacactgtca | ggagctggc  | cctggccagc  | gtgttccacc   | 840  |
| ttagggggca   | cgagaactac  | acctgcgtt  | tcgagacat  | gtgtgtc     | gacggctacg   | 900  |
| tggtaaagag   | gatcgccatc  | agccccggc  | tgtacggca  | gcccagcgg   | tacgcgtct    | 960  |
| caatgcacag   | ggagggttcc  | ctgtgtc    | tggttgc    | caccctg     | ggcagagagg   | 1020 |
| ttagcttccc   | cgtgtgcacc  | taatgttgc  | ccacccctgt | tcgacat     | accggcatcc   | 1080 |
| tggccacccg   | cgtgaggcgc  | gacgacgccc | agaagctgt  | ctgtggc     | aaccagagga   | 1140 |
| tcgtggtaa    | ccgcaggacc  | cagaggaaca | ccaacacaat | gaagaactac  | ctgctgc      | 1200 |
| tggtgccca    | ggcttccgc   | agggtggcc  | aggatatacc | ggaggacccag | gaagacgaga   | 1260 |
| ggcccccggg   | cctgaggagc  | aggcagctgg | tgatgggtc  | ctgtggggc   | ttcaggcgc    | 1320 |
| acaagatcac   | cagcatctac  | aaagggccgg | acacccacac | catatca     | gtgaacacq    | 1380 |
| acttccacag   | cttcgtgt    | ccccaggatc | gcagcaacac | cctggagatc  | ggcctgagga   | 1440 |
| cccggtatcg   | gaagatgtc   | gaggaaacac | aggagccgg  | cccaactgt   | accggccgg    | 1500 |
| acgtgcagg    | ggccaaatgtc | gctggcgc   | aggccaaagg | gttggggag   | ggcggaggaa   | 1560 |
| tgaggccgc    | cctgcaccc   | ctggctgc   | acgtgggg   | acccaccc    | gaagccgacg   | 1620 |
| tggacacctgat | gctgcaggag  | gccggcgc   | gaagcgtgg  | gacacc      | ggcctgatca   | 1680 |
| aggtgaccag   | ctacgcggc   | gaggacaaga | tcggcagct  | cccccgt     | agccccacagg  | 1740 |
| cctgtgtc     | gttccggaa   | ctgagctgc  | tcacccact  | ggccgac     | gtgtatgt     | 1800 |
| tcacccacag   | ccgcaggaa   | ggcaggatc  | ccgtggggc  | ctaccacgc   | aagggtgtc    | 1860 |
| tgcccgaggg   | ccacgcac    | ccccgtc    | acttccagg  | cctgagc     | agcgcacca    | 1920 |
| tcgtgtacaa   | cgagagggg   | tttgtaaca  | gttacgtca  | ccatatc     | acccacggc    | 1980 |
| gagccctgaa   | caccgcacg   | gaataacta  | agacccgtg  | gcccag      | cacgcacgg    | 2040 |
| agtacactgt   | cgacatcgac  | aggacatgt  | cgctgt     | ggcgttgc    | accggcttgc   | 2100 |
| gactgacccg   | cgagctgtgt  | gacccaccc  | tccacgt    | ccgcctac    | agccgtgagga  | 2160 |
| ccagaccgc    | cgctccctac  | cagggtcc   | ccatcggt   | gtacggcgt   | ccggcagcg    | 2220 |
| gaaagagccg   | catcatca    | agcgcgtgt  | ccaagaaa   | cctgtgtc    | agcgcacaa    | 2280 |
| aaagagaactg  | cggcgcac    | atcaggac   | tgaagaa    | gaaaggc     | gtacgt       | 2340 |
| cggcgcacccgt | ggacacg     | ctgtgt     | gctgt      | ccccgtt     | accctgt      | 2400 |
| tcgacgcggc   | cttcgttgc   | cacgcggc   | ccctgagg   | ctgtatc     | atcatc       | 2460 |
| ccaaagaaacg  | cgtgtgtc    | ggcggaccc  | aggcgttgc  | tttccat     | atgtatgt     | 2520 |
| tgaagggtca   | ttcaaccac   | gagatgtc   | ccccgtt    | ccacaag     | atcagc       | 2580 |
| ggtgcaccaa   | gagcgtgt    | agcgtgt    | gtcaccctgt | ctacgac     | aaaatgagga   | 2640 |
| ccaccaaccc   | eaaggagacc  | aaaatgt    | tcgacacc   | aggcag      | aagccaa      | 2700 |
| aggacgac     | gttccgttgc  | tgcttgc    | ggctgggt   | gacgtgt     | atcacta      | 2760 |
| agggcacacg   | gatcatgacc  | ggcgtgtc   | ggcggggc   | tttccat     | atgtatgt     | 2820 |
| cctgtgttgc   | caagggt     | gagaaccc   | tgtacgt    | caccac      | acgtgt       | 2880 |
| tgctgtgtac   | cgaggac     | gacaggatc  | tgtgg      | gac         | ccctgttgc    | 2940 |
| taaaggaccc   | gaccgcac    | tacccgg    | acttcaccc  | caccatcg    | gagtgccagg   | 3000 |
| ccgagcaca    | cgccatcat   | aggcatac   | ttggagg    | ccgaccc     | gacgtgttcc   | 3060 |
| agaacaaggc   | caacgtgt    | ttggc      | ccctgt     | cgatgt      | accggcc      | 3120 |
| tcgacatgc    | cacagac     | tggacac    | tggactt    | cgagac      | aaggccaca    | 3180 |
| ggccgcagat   | cgtgtgt     | cagctgt    | tgagg      | ctggc       | ctggacag     | 3240 |
| gctgttgc     | cccccc      | ctggc      | catcgagg   | caaccact    | ggacaa       | 3300 |
| ccagccaa     | catgtac     | ctgaa      | agggtgt    | cgacgt      | aggcggtac    | 3360 |
| cacagctgc    | caggcgt     | gccac      | gggtgt     | catgaa      | ggcaccc      | 3420 |
| gaaatcgac    | ccccggatc   | aacctgt    | ccgtgt     | gcccgt      | caaccc       | 3480 |
| tgctgcacca   | caacgac     | ccacagac   | acttcag    | cttcgt      | gacgtaa      | 3540 |
| gcagggccgt   | gtgtgtgt    | ggcgg      | tgacgtgt   | ccggc       | gtgtgttgc    | 3600 |
| tgagcgtac    | ccccgg      | acctt      | ccaggct    | gttgc       | ccggcgttgc   | 3660 |
| tgcccaatgt   | cgacatc     | tttgt      | tcaggatcc  | ataca       | accattacc    | 3720 |
| agcagtgc     | ggacac      | atca       | tcgtgt     | gacgtgt     | ccatgggg     | 3780 |
| tgaacccccc   | aggacac     | gtgtgt     | gtacgt     | ccgtgt      | ccacgg       | 3840 |
| gcatcattgg   | cgccatc     | agggtt     | atgtc      | gtgtgt      | cccaagagca   | 3900 |
| gcotggag     | accggatgt   | ctgtgt     | tcatcggt   | cgaccgg     | ggcaggaccc   | 3960 |
| acaacccct    | caagctgt    | agcacc     | tcacat     | caccatcg    | ccacgg       | 4020 |
| aggccggcgt   | cccccc      | tacacc     | tcagg      | ggcgg       | ccacgg       | 4080 |
| gcgtgtat     | caacgtgt    | aacagca    | ggcc       | ccacgg      | aggcggatgt   | 4140 |
| tgtacaagaa   | gttcccg     | aggttgc    | tcacgtt    | ccagg       | ggcggcc      | 4200 |
| tgtgttgc     | ccgcgtt     | cacat      | tcggttgc   | cccca       | ccacgg       | 4260 |
| gcaaggatgt   | ggccgtt     | acgcgtt    | tcggatcc   | ggcc        | ccacgg       | 4320 |
| acgacata     | tttaccat    | tcggatcc   | tcgtgt     | ccat        | ccatgg       | 4380 |
| acaaggac     | gtgtgttgc   | agcctgt    | tcgtgt     | ccac        | ccacgg       | 4440 |
| ccgacgtgtc   | catctact    | agggaca    | agttgg     | ccac        | ggggcgttgc   | 4500 |
| ccaggcgg     | ggccgtt     | gagat      | tcacgt     | ccac        | ggggcgttgc   | 4560 |
| acgcgtgt     | ggccgtt     | cagtcgtt   | tcggatcc   | ccac        | ggggcgttgc   | 4620 |
| caagacat     | tttaccat    | tcggatcc   | tcgtgt     | ccat        | ccatgg       | 4680 |
| acatgcgt     | gtatca      | atgtgg     | tcagg      | ggcc        | ccacgg       | 4740 |
| tgttacat     | ggggcgtt    | atgttgc    | tcagg      | gttgc       | ccatgg       | 4800 |
| aggccac      | accaccc     | tcacgtt    | tcacgt     | ccac        | ccacgg       | 4860 |
| gggtgcac     | ggccgtt     | accctgc    | tcacgtt    | ccac        | ccacgg       | 4920 |
| tgcccaatgt   | caggatc     | ggcgtt     | tcacgtt    | ccac        | ccacgg       | 4980 |

-continued

---

|             |            |            |            |            |            |      |
|-------------|------------|------------|------------|------------|------------|------|
| gccccaaaggt | gcccgcctac | atccacccca | ggaagtaccc | ggtggagacc | ccacccgtgg | 5040 |
| acgagacacc  | cgagccaagc | gccgagaacc | agagcacccg | gggcacaccc | gaggccac   | 5100 |
| cctgtatcac  | cgaggacgag | acaaggaccc | ggacccca   | gcccattt   | atcgaggaa  | 5160 |
| aggaagagga  | cagcatcagc | ctgtgtac   | acggccccc  | ccaccagg   | ctgcagg    | 5220 |
| aggccgacat  | ccacggccca | ccacgtgt   | ccacgtt    | ctggagatc  | ccacaccc   | 5280 |
| gcgacttca   | cgtggacagc | ctgacatcc  | tggacacc   | ggaggggcc  | agcgtgac   | 5340 |
| ccggcgccac  | cagcggcag  | accaacgt   | acttcgca   | gagcatgg   | ttctggcc   | 5400 |
| ggcccggtcc  | agctccagg  | accgtttca  | ggaacccacc | ccacccag   | cccaggac   | 5460 |
| ggacccaaag  | cttggctcc  | acggggct   | gcagcagg   | cagctgg    | agcacccc   | 5520 |
| ccggcggt    | cagggtatc  | acagggg    | aactgggg   | cctgacacc  | agcaggac   | 5580 |
| ccacgggtc   | cgtgacagg  | actagtct   | tgtccaa    | acccgg     | aacagggt   | 5640 |
| tcaccaggga  | ggaattcgag | gecttctgg  | cccagcaaca | gagacgg    | gacgcgg    | 5700 |
| ctacatctt   | cagcagcag  | acccgg     | gacaccc    | gcaaaa     | agcgtg     | 5760 |
| cctgtgttag  | cgagggtgt  | ctggagag   | ccgagctt   | aatcag     | tcgaccc    | 5820 |
| tgaccggaa   | gaaggaggaa | ctgtcg     | agaaaact   | gctgaa     | acccagg    | 5880 |
| acaggagcag  | gtaccagagc | aggaagg    | agaacat    | ggccat     | ccgcagg    | 5940 |
| tcctgcagg   | cttgggacac | tacctg     | ccgagg     | ggtgg      | tacagg     | 6000 |
| tgeaccccg   | gcccactgt  | agctcc     | ccacgt     | tgaacagg   | cttcc      | 6060 |
| ccctggaggc  | ctgcaacgt  | atgtca     | agaactt    | caccgt     | agctact    | 6120 |
| tcatccccga  | gtacgacg   | tacctgg    | tgggg      | cgccag     | tgcctgg    | 6180 |
| ccgcgcgtt   | ctgccccggc | aaggtg     | gttcccc    | gaaacac    | tacctgg    | 6240 |
| ccacccatcg  | gagcggcg   | cccagc     | tccaga     | acac       | ctgcagg    | 6300 |
| ctgcccacaa  | gaggacatc  | aacgtg     | agatgg     | gctgg      | ctggac     | 6360 |
| ctgccttca   | ctggggatc  | ttcaagaa   | acgcct     | caac       | gagta      | 6420 |
| taaaggaa    | ccccatcagg | ctgacc     | gaaacgt    | gaaact     | accac      | 6480 |
| aggggcccaa  | ggccgcgtcc | ctgtcg     | gaccc      | cctgaa     | ctgcagg    | 6540 |
| tcccaatgg   | cagggtcg   | atggac     | agaggg     | gatgg      | cccg       | 6600 |
| agcacacccg  | ggagaggccc | aagg       | tgatcc     | cgctg      | ctggcc     | 6660 |
| cctacccgt   | gggcaccc   | aggggatc   | tgagg      | ggcgc      | ctgctg     | 6720 |
| acatccac    | cgttgc     | atgagc     | aggactt    | cgccat     | gcccag     | 6780 |
| tccagccccc  | cgtgcgt    | ctggag     | ccatc      | cttc       | cttc       | 6840 |
| acgctatggc  | cctgac     | ctgtat     | tggagg     | gggcgt     | gccgag     | 6900 |
| tcaccctgt   | cgagggtgc  | ttcggc     | tcagctt    | ccac       | ctgc       | 6960 |
| agttcaatgt  | atggatgt   | atggaa     | gaaat      | gatgtt     | gaccctgt   | 7020 |
| tgtatcaat   | tgtatcg    | tgagg      | tgccgg     | gatgg      | gtgg       | 7080 |
| ctgccttcat  | gggcgc     | aacatcg    | aggcg      | gttgc      | ggcgg      | 7140 |
| acagggtgc   | cacctgg    | aatcgg     | tgaagat    | cgac       | ccgc       | 7200 |
| aggcccttat  | cttgcgg    | ggattc     | tgtgc      | cgatgg     | accgc      | 7260 |
| gggtggccg   | ccccctga   | agggtt     | tca        | gccc       | gtgc       | 7320 |
| agcacacccg  | tgacaggc   | agggc      | acgagg     | gatgg      | ggcc       | 7380 |
| gcatcctgag  | cgagctgt   | aaggc      | agagc      | cgagat     | gttgc      | 7440 |
| tcatcgat    | ggctatc    | acatcg     | ggcc       | gttcc      | tac        | 7500 |
| gggccttat   | aactctac   | ggcta      | aatgg      | gatct      | agccac     | 7560 |
| agcaagatct  | acatcgac   | gccc       | gagatc     | tgtgc      | gggg       | 7620 |
| atcggcatcg  | ggggcc     | cgccccc    | ctgacc     | gatgt      | gtgg       | 7680 |
| gaggtaaca   | aggccctgt  | cgatcg     | aggagc     | gttgt      | tatc       | 7740 |
| atccctcccc  | gggtgttcat | gaccac     | ggcc       | gatgtt     | ccgc       | 7800 |
| gacgtatca   | tccagc     | ggagg      | aatgtt     | ccat       | gggg       | 7860 |
| gacgacgt    | tccccccaa  | gaagat     | gactt      | ggcc       | ccat       | 7920 |
| atcaacgt    | atcgatgt   | caccagg    | gactt      | gatgt      | ccat       | 7980 |
| cccaagca    | ccccccat   | ctacgct    | gtgg       | gatcg      | gggg       | 8040 |
| ggcgacggc   | agagca     | ggaccc     | aaac       | acttgc     | cgatgg     | 8100 |
| ctgtgggt    | acgtgtat   | ccgtt      | actcg      | gtgact     | aggatct    | 8160 |
| taccatcaa   | ccagcttca  | gaaccc     | atgg       | gatct      | aat        | 8220 |
| tacacttca   | aatatgttca | ccccaa     | atgg       | gatctt     | taataaaaa  | 8280 |
| aaagtttttt  | cacattct   | agtc       | cttgc      | ctac       | gggg       | 8340 |
| taactctca   | cggttac    | gaatgg     | cgat       | ccat       | ggaa       | 8400 |
| aaaaacatta  | agaaggcc   | agccat     | tacc       | ggac       | ggcc       | 8460 |
| cagctgcaca  | aaaggcat   | ggttac     | ctgg       | ccat       | cttac      | 8520 |
| gacatcatcg  | agggtgacat | taccc      | gatgt      | atgt       | ccat       | 8580 |
| gaagctatga  | agcgtatgg  | gttgc      | accat      | ccat       | gggg       | 8640 |
| agtttgcgt   | tcttcat    | cggttgg    | gttgc      | tgc        | ccat       | 8700 |
| gctaaccaca  | tctaca     | ggcg       | acttgc     | tgg        | ccat       | 8760 |
| gtctgtatcg  | ttagca     | agggt      | atgtt      | ccat       | ggat       | 8820 |
| atcatacaaa  | agatcat    | catggat    | aaagg      | ccat       | gggg       | 8880 |
| taacaccc    | tgacttcc   | tttgc      | ggctt      | caac       | gttgc      | 8940 |
| agcttcgacc  | ggggaaa    | categ      | ccat       | gtat       | ggcc       | 9000 |
| cccaaggccg  | tagcc      | ccat       | ggcc       | ccat       | ccat       | 9060 |
| cccatcttc   | gcaacc     | ccat       | acc        | ccat       | ccat       | 9120 |
| cacggcttc   | gcatgtt    | cacgtt     | tactt      | gttgc      | ccat       | 9180 |
| atgttaccgt  | tcgagg     | gttac      | ccgc       | atgt       | ccat       | 9240 |
| gcctctgg    | tgcccc     | atcat      | gtat       | ccat       | ccat       | 9300 |
| gacctaagca  | acttgc     | gtat       | ggcc       | ccat       | ccat       | 9360 |
| gaggccgtgg  | ccaaac     | ccat       | ggcc       | ccat       | ccat       | 9420 |
| acaaccaggc  | ccattt     | caccc      | gggg       | ccat       | ccat       | 9480 |
| gtgtgtcc    | tcttc      | taagg      | gttgc      | ccgtt      | ccat       | 9540 |
| aaccaggcgc  | gca        | ccat       | ggcc       | ccat       | ccat       | 9600 |
| cccgaggct   | caaac      | catc       | ggcc       | ccat       | ccat       | 9660 |
| tactggac    | aggac      | ccat       | ggcc       | ccat       | ccat       | 9720 |

---

-continued

|             |             |             |            |              |             |       |
|-------------|-------------|-------------|------------|--------------|-------------|-------|
| aagggttacc  | aggttagcccc | agccgaactg  | gagagcatcc | tgctgcaca    | cccaacatc   | 9780  |
| ttcgacccg   | gggtcgccgg  | cctgcccgc   | gacgatccgc | gcgagctgcc   | cgccgcagt   | 9840  |
| gtcgtgttgg  | aacacggtaa  | aaccatgacc  | gagaaggaga | tcgtggacta   | tgtggccagc  | 9900  |
| cagggttaca  | ccgccaagaa  | gctgcccgg   | ggtgttgcgt | tcgtggacga   | ggtgccctaaa | 9960  |
| ggactgaccg  | gcaagttgg   | ccccccgcaag | atccgcgaga | tttcttattaa  | ggccaagaag  | 10020 |
| ggccggcaaga | tgcgggtgt   | actcgagat   | tttacgtgca | aagggtatgg   | tcacccccc   | 10080 |
| aaagaccata  | tttgtacaca  | ccctcgtat   | cacgccccaa | catttacagc   | cgccgtgtca  | 10140 |
| aaaacccgcgt | ggacgtgg    | aacatccctg  | ctgggaggat | cagccgtaat   | tattataatt  | 10200 |
| gggtttgggtc | ttggctactat | tgtggccat   | taactgtgtc | ccaaaccaggaa | acataatgt   | 10260 |
| atacagcgc   | aattggcaag  | ctgtttatcat | agaactcgcg | gctggggca    | tgccgcctt   | 10320 |
| aaatttttat  | ttttttttt   | ccqaaatcgga | ttttgtt    | aatatttcaa   |             | 10380 |
| aaaaaaaaaa  | aaaaaaaaaa  | aaatcttagaa | aaaaaaaaaa | aaaaaaaaaa   | aaaaaaaaaa  | 10440 |
| aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa | aaaaaaaaaa   | aaaaaaaaaa  | 10500 |
| aaaaaaaaaa  |             |             |            |              |             | 10508 |

SEQ ID NO: 118                    moltype = RNA   length = 11075  
 FEATURE                            Location/Qualifiers  
 misc\_feature                    1..11075  
 note = Synthetic polynucleotide  
 source                            1..11075  
 mol\_type = other RNA  
 organism = synthetic construct

SEQUENCE: 118

|              |              |              |              |              |              |      |
|--------------|--------------|--------------|--------------|--------------|--------------|------|
| atgggcggcg   | catgagagaa   | gcccagcca    | attacctacc   | aaaaatggag   | aaagttcacg   | 60   |
| ttgacatcga   | ggaagacagc   | ccatccctca   | gagcttgc     | ggggaggttc   | ccgcagtttgc  | 120  |
| aggttagaagc  | caagcagggtc  | actgtataatg  | accatgtctaa  | tgccagagcg   | tttccgcata   | 180  |
| tggcttcaaa   | actgtatcgaa  | acggagggttgc | accatccgaa   | cacgcatttgc  | gacattggaa   | 240  |
| gtgcgcgcgc   | ccgcagaatg   | tattctaago   | acaagtatca   | ttgtatctgt   | ccgcgtggat   | 300  |
| gtgcggaaaga  | tccggacaca   | ttgtataatg   | atgcactaa    | gtcgaagaaa   | aactgttaagg  | 360  |
| aaataactgt   | taaggaaatgg  | gacaagaaa    | tgaaaggact   | ggccgcggc    | atgacgcgac   | 420  |
| ctgacccgttgc | aaactgagact  | atgtgccttc   | acgcacgcga   | gtcgtgtcgc   | tacgaagggc   | 480  |
| aagtgcgtgt   | tttacccaggat | gtatacgcgc   | tcgacggccc   | caccacgttgc  | taccacccagg  | 540  |
| ccaaacaagg   | cgtaggggttgc | gcctacttgc   | tcggcttcgc   | caccacaccc   | ttcatgttca   | 600  |
| agaacacctgc  | cgccgccttc   | cccgacttgc   | gacccaaactgc | ggccgcacgg   | accgcgttgc   | 660  |
| ccgcaggaa    | catcgccgtc   | tgacgcgcg    | acgtgtatgg   | gaggagccgg   | agggccatgc   | 720  |
| gcacatctcg   | gaagaaat     | ctgaaggcca   | gcaacaactgt  | gtgttcagc    | gtgggcagca   | 780  |
| ccatcttacca  | cgagaagagg   | gacgtgtca    | ggagctggca   | cctgccc      | gtgttccacc   | 840  |
| tggggccaa    | cgagaactac   | acccgtcgat   | ggcagaccat   | cgtagtgc     | gacggctacg   | 900  |
| tggtgaagag   | gatcgccat    | agccccggc    | tgtacggca    | gcccagcgc    | tacgcccgt    | 960  |
| caatgcacac   | ggagggttgc   | ctgtgtca     | agggtacccg   | caccctgaa    | ggcgagaggg   | 1020 |
| tgagcttccc   | cggtgtcacc   | taatgtccgc   | ccaccctgttgc | cgaccatgt    | accggcatcc   | 1080 |
| tggccacccg   | cgtagcgcgc   | gacgcacggc   | agaagctgcgt  | cgtagggcttgc | aaccagaggaa  | 1140 |
| tctgtgttca   | cgccaggacc   | gaggacaca    | ccaaacaaat   | gaagaactac   | ctgtgtccgc   | 1200 |
| tggtggccca   | ggcttttcgc   | agggtggca    | aggagttacaa  | ggaggaccag   | gaagacgaga   | 1260 |
| ggccccctggg  | cttgaggggac  | aggcagctgg   | tgatgggtgc   | ctgtgtggcc   | ttcaggccgc   | 1320 |
| acaagatcac   | cagcatctac   | aagaggcccg   | acacccacac   | catcatca     | gtgaacacgc   | 1380 |
| acttccacag   | cttcgtgtc    | cccaggatgtc  | cgacgcacac   | cctgtggatc   | ggccgtggat   | 1440 |
| cccggtatcg   | gaagatgtgc   | gaggacaca    | aggagcccg    | ccactgtac    | accgcggagg   | 1500 |
| acgtgcagg    | ggccaaatgtc  | getgcgcgc    | aggccaaagg   | gttgcgggg    | gccggaaac    | 1560 |
| tggggccgc    | cttgccaccc   | cttgcgtgc    | acgtggggac   | accacccctg   | gaagccgcac   | 1620 |
| tggacccgtat  | gtgtcgat     | ggccgcgcgc   | aaaggctgg    | gacccacccgt  | ggccgttgc    | 1680 |
| agggttccac   | ctacgcggc    | gaggacaaga   | tcggcgttgc   | ccgcgtgtc    | agcccacagg   | 1740 |
| ccgtgtgttca  | gtcccgaaag   | ctgagctca    | tccacccact   | ggccggac     | gtgatgttgc   | 1800 |
| taaccacacag  | cgccaggacaa  | ggcgtgttgc   | ccgtggatgc   | cttacacggc   | aagggtggcg   | 1860 |
| tgcctggagg   | ccacgcacatc  | ccgtgttgc    | acttccaggc   | cctgtggatgc  | agccgttgc    | 1920 |
| tctgttgcac   | cgaggatgg    | ttctgttgc    | ttgtgttgc    | ccatatgttgc  | acccacggcg   | 1980 |
| gagccctgttca | caccgcacgg   | gaataactaca  | agaccgttgc   | gacccacccgt  | ggccgttgc    | 2040 |
| agtacccgttca | cgacatcgac   | aggaaacgt    | cgctgttgc    | ggccgttgc    | accggccctgg  | 2100 |
| gactgtccgg   | cgactgtgttgc | gacccacccgt  | tccacgtatc   | cgccgttgc    | agccgttgc    | 2160 |
| ccagacccgc   | cgctccctac   | cagggttgc    | ccatccgttgc  | gtacggccgt   | cccgccacgc   | 2220 |
| gaaagagccgc  | catcatca     | agcgcgttgc   | ccaagaaaga   | cctgtgttgc   | agccgttgc    | 2280 |
| aaagagaactgc | cgccgcacatc  | atcagggttgc  | tgaaagaaat   | gaaaggccgt   | gacgtgttgc   | 2340 |
| cgccgcacgcgt | ggacgcgttgc  | ctgtgttgc    | gctgttgc     | ggccgttgc    | accggccctgg  | 2400 |
| tcgacgttgc   | cttcgtgtc    | tcgcgttgc    | ccgtgttgc    | ccgtgttgc    | atcatcgac    | 2460 |
| ccaaagaaacgc | cggtgtgttgc  | ggcgtgttgc   | aggcgttgc    | tttcttca     | atgtgttgc    | 2520 |
| tgaaggttgc   | cttcaaccac   | gagatgttgc   | cccagggttgc  | ccacaagac    | atcaggccgc   | 2580 |
| gggtgcacca   | ggacgcgttgc  | agcgtgttgc   | tgatgggtgc   | ccatccgttgc  | ggccgttgc    | 2640 |
| ccaccaaccc   | caaggaggacc  | aaaatgttgc   | tcgacacccat  | aggcgttgc    | aaagcccaacgc | 2700 |
| aggacgttgc   | gatcttgc     | tgcttcagg    | gttgcgttgc   | ggcgttgc     | atcgttgc     | 2760 |
| agggcaacgc   | gatcatgttgc  | ggcgtgttgc   | ggcgtgttgc   | gaccaggaa    | ggcgtgttgc   | 2820 |
| ccgttgcagg   | caagggttgc   | gagaacccac   | tgtacgttgc   | caccaggc     | cacgttgc     | 2880 |
| tgtgtgttgc   | caggaccgttgc | gacccatgttgc | tgatgggtgc   | ccgtgttgc    | ggccgttgc    | 2940 |
| tcaagacccgtt | gaccgcacatc  | taccccgcc    | acttccacgc   | caccatgttgc  | gatgttgc     | 3000 |
| cccgacgttgc  | cgccatccat   | aggcgttgc    | tggaggttgc   | ccacccatgc   | gacgtgttgc   | 3060 |
| agaacaaggc   | caacgtgttgc  | tgggttgc     | ccctgtgttgc  | cgatgttgc    | accgcgttgc   | 3120 |
| tcgacatgttgc | cacagacgttgc | tggacacacc   | tggactactt   | cgagaccgc    | aaggcccaacgc | 3180 |
| gtggccggat   | cggtgttgc    | caggtgttgc   | tgatgggttgc  | ccggccgttgc  | ctggacacgc   | 3240 |
| gcctgttgc    | cgccccccacc  | gtgttgc      | gcatcgttgc   | gaccaggaa    | gacaacacgc   | 3300 |
| ccagcccaaa   | catgtacccgc  | ctgaaaca     | agggtgttgc   | cgatgttgc    | aggccgttgc   | 3360 |

-continued

---

|             |             |            |             |             |             |           |      |
|-------------|-------------|------------|-------------|-------------|-------------|-----------|------|
| cacagctgcc  | caggccgtg   | gccaccggca | gggtgtacga  | catgaacacc  | ggcacccctga | 3420      |      |
| ggaactacga  | ccccaggatc  | aacctggtgc | ccgtgaacag  | cgccgtccc   | cacgcctgg   | 3480      |      |
| tgtgtccaca  | caacgagcac  | ccacagagcg | acttcagctc  | ttcgttgagc  | aagctgaaag  | 3540      |      |
| gcaggaccgt  | gttggctgt   | ggcgagaago | tgagcgtgcc  | cgccaagatg  | gtggactggc  | 3600      |      |
| tgagcgtac   | gccccggcc   | accttccgg  | ccaggctgg   | cctcggcatc  | cccgccgacg  | 3660      |      |
| tgcctaagta  | cgacatcata  | ttctgt     | tcaggacacc  | atacaagtac  | caccattacc  | 3720      |      |
| agcaagtgcg  | ggaccacgca  | atcaagctga | gcatgtgcac  | caagaaggcc  | tgccctgcacc | 3780      |      |
| tgaacccccc  | aggcacctgc  | gtgagcatcg | gctacggctc  | cgcccgacagg | gccagcgaga  | 3840      |      |
| gcatcatgg   | ccccatcgcc  | aggctgttc  | agttcagtag  | ggtgtgcaaa  | cccaagagca  | 3900      |      |
| gcctgggg    | aaccgggtg   | ctgttgc    | tcatecgctc  | cgacccggaa  | gccaggaccc  | 3960      |      |
| acaaaccccta | caagctgagc  | agcacctga  | caaacatcta  | caccggcaac  | aggctgcacg  | 4020      |      |
| aggccggctg  | cgccccccage | taceacgtgg | tcaggggctg  | tatcgcacc   | gccaccgagg  | 4080      |      |
| gegtgtatcat | caacgctgcc  | aacacgcaaa | ggccagccccc | aggccggatg  | tgccggcccc  | 4140      |      |
| tgtacaagaa  | gttccccgg   | agettcagac | tgcaaggccat | cgagggtggc  | aaggccaggc  | 4200      |      |
| tgggtgaagg  | cgccgctaa   | ccccatcc   | acgcccgtgg  | ccccaaactc  | aacaagggtg  | 4260      |      |
| gcgaggtgga  | aggcgacaa   | cagetggcc  | aagccctacg  | gagcatcgcc  | aagatcgta   | 4320      |      |
| acgacaataa  | stacaaagacg | gtggccatc  | cactgtctcg  | caccggcatc  | ttcagcggca  | 4380      |      |
| acaaggacag  | gtcgaccac   | agcctgcac  | acctgtctca  | cgccctggac  | accaccgtat  | 4440      |      |
| ccdacgtg    | catctactgc  | aggacaaga  | agtgggatq   | gaccctgaa   | gaggccgtgg  | 4500      |      |
| ccaggccgg   | ggccgtggaa  | gagatctgc  | tcagcgacg   | ctccagcgtg  | accgagcccg  | 4560      |      |
| acgcccgt    | gttgggtg    | cacccaaga  | gttccctggc  | cgccaggaa   | ggctacagca  | 4620      |      |
| ccagcgacgg  | caagaccc    | agcttccgt  | aggggccacca | gttccaccag  | cccgctaaagg | 4680      |      |
| acatcgcc    | gatcaacgt   | atgtggcc   | tggccacccg  | ggcccaacgg  | cagggttgca  | 4740      |      |
| tgtacatct   | ggcgagagc   | atgtccagca | tcaggagca   | gtgcccgtg   | gaggaaagcg  | 4800      |      |
| agggccagac  | accacccac   | acccgtcc   | gctgtgtcat  | ccacgctatg  | acacccgaga  | 4860      |      |
| gggtgcagcg  | gttggaggcc  | agcaggccc  | agcagatcac  | cggtgtcagc  | tccttccac   | 4920      |      |
| tgcctaagta  | caggatcacc  | ggcgtcaga  | agatccatcg  | cagccagccc  | atccgttca   | 4980      |      |
| gccccaaagg  | gccccctac   | atccacccca | ggaagtac    | gttggagacc  | ccaccctgtgg | 5040      |      |
| acgagacacc  | cgagccaa    | gcccggaa   | agacgacccg  | gggcacaccc  | gagcagccac  | 5100      |      |
| cctgtatcac  | cgaggacgag  | acaaggaccc | ggacccccc   | gcccacat    | atcgaggaa   | 5160      |      |
| aggaagagcg  | cagcatcagc  | ctgtgt     | acggccccc   | ccacccagg   | ctgcagggtg  | 5220      |      |
| agggccgacat | ccacggccca  | cccacgtgt  | ccagctcc    | ctggagcatc  | ccacacggca  | 5280      |      |
| gcaactgtc   | cggtggacag  | ctgagcatc  | tggacaccc   | ggagggcgcc  | agcgtgacat  | 5340      |      |
| ccggcgeccac | accaccaatg  | acttgc     | gacca       | gagcatggag  | ttccctggca  | 5400      |      |
| ggccgtgtcc  | agctccccc   | accgttca   | ggaaacccac  | ccacccag    | cccaggacca  | 5460      |      |
| ggaccccaag  | cctggctcc   | agcaggcc   | gcacgtgg    | cagcctgtg   | agcacccac   | 5520      |      |
| ccggcgtgaa  | cagggtgatc  | accagggagg | aactggaggc  | cctgacaccc  | agcaggaccc  | 5580      |      |
| ccagcggtc   | cggtggacag  | actagtctgg | tgccaaaccc  | accggcgctg  | aacagggtg   | 5640      |      |
| tcaccaggga  | ggaattcgag  | gccttgc    | ccca        | gagacggttc  | gacgcccgg   | 5700      |      |
| cctacatctt  | cagcagcgac  | accggccagg | gacacgtca   | gcaaaaagac  | gtgaggcaga  | 5760      |      |
| ccgtgtcg    | cgaggtgt    | ctggaggg   | ccgagctgg   | aatcagctac  | gccccccaggc | 5820      |      |
| tggaccagg   | gaaggaggaa  | ctgtcgagg  | agaaaactcg  | gctgaacccc  | accccaaccc  | 5880      |      |
| acaggacgg   | tatccacg    | aggaatgtgg | agaacatgaa  | ggccatcacc  | gccaggcgga  | 5940      |      |
| tccctgcagg  | cctggacac   | tacctgaa   | ccgaggggca  | gttggagtgc  | tacaggaccc  | 6000      |      |
| tgcacccctg  | gccactgtac  | agtcctcg   | tgaacaggcc  | cttctcc     | cccaagggtgg | 6060      |      |
| ccgtggacgc  | ctgcaacgt   | atgtcgaa   | agaaactccc  | caccgtggcc  | agctactgca  | 6120      |      |
| tcatecccg   | gtacgacgc   | tacccgtgg  | tggggacgg   | cgccagctgc  | tgccctggaca | 6180      |      |
| ccgcccgtt   | ctggccccc   | aagctgagg  | gttccccc    | aaaaacacgc  | tacctggac   | 6240      |      |
| ccaccatcg   | gagccgtgt   | cccacgtcc  | tccagaacac  | cctgca      | gagaac      | 6300      |      |
| ctggccacca  | gaggatcgac  | aactgtcc   | ggatgggg    | gctgcccgtg  | ctggacagcg  | 6360      |      |
| ctgccttcaa  | cggtggatg   | ttaaagaaat | acgcctcgaa  | caacgat     | tggggagacc  | 6420      |      |
| tcaaggagaa  | ccccatcagg  | ctgacc     | agaacgtgg   | gaactacatc  | accaagctg   | 6480      |      |
| agggcccca   | ggccgtg     | ctgttgc    | agacccaa    | cctgacat    | ctgcaggaca  | 6540      |      |
| tcccaatg    | cggttcgt    | atggac     | agggggatg   | gaaatgg     | cccgccacca  | 6600      |      |
| agcacaccc   | gggaggccc   | aagggtcagg | tgatcc      | cgctgacca   | ctggccaccc  | 6660      |      |
| cctacatgt   | cggtatcc    | agggatcg   | tgatcc      | ccacctgccc  | accaagacca  | 6720      |      |
| acatccac    | cctgttc     | atgagcgc   | aggacttc    | cgccat      | ccgcgacact  | 6780      |      |
| ccagccccc   | cgactcg     | ctggagac   | acatcgcc    | cttcgac     | agcggat     | 6840      |      |
| acgttatg    | cgttgc      | ctgtat     | ggggat      | ttggggat    | ggccgat     | 6900      |      |
| tcaccctgt   | cgaggatcg   | ttccgg     | ccatcgat    | ccacctgccc  | accaagacca  | 6960      |      |
| agttcaat    | cggtcgat    | atgaaa     | aggcg       | gatgttcc    | gaccctgtc   | 7020      |      |
| tgtacat     | gtgtcg      | agcagggt   | tgccgg      | gatgtcc     | gtgacaccc   | 7080      |      |
| ctgccttca   | cggtcgac    | aaatcg     | gtggcg      | aaagc       | ccgcgttgc   | 7140      |      |
| acagggtgc   | cacccgt     | aaatcg     | ggggat      | ccacat      | ccgcgttgc   | 7200      |      |
| aggcccccta  | tttgcggc    | ggattcat   | tgtgc       | gacac       | ccgcgttgc   | 7260      |      |
| gggtggccg   | ccccctg     | aggctgtt   | agctgg      | ccact       | ccgcgttgc   | 7320      |      |
| agcacacg    | tgacaggcc   | aggccct    | acgagg      | aa          | ccgcgttgc   | 7380      |      |
| gatctcg     | cgagctgt    | agggcg     | agagc       | ccat        | ccgcgttgc   | 7440      |      |
| tcatctgt    | ggtatcg     | acactg     | ccat        | ccgcgttgc   | ccat        | ccgcgttgc | 7500 |
| ggggccctat  | aactctctac  | ggcta      | aatgg       | act         | ccgcgttgc   | 7560      |      |
| ggccgcacc   | atggagat    | ccaaaaacat | taaga       | agggc       | ccagccat    | 7620      |      |
| cgaagacgg   | accgcgg     | agcagct    | caa         | aggcc       | ccgcgttgc   | 7680      |      |
| cgccaccat   | cccttacc    | acgcacat   | cgagg       | ttgc        | ccgcgttgc   | 7740      |      |
| cgagatg     | gttcggct    | caga       | acgt        | ccat        | ccgcgttgc   | 7800      |      |
| gatctgtt    | tgca        | atgctt     | gat         | ccat        | ccgcgttgc   | 7860      |      |
| catcggtgt   | gtctgg      | cagta      | ac          | cat         | ccgcgttgc   | 7920      |      |
| catggc      | atcc        | cgat       | ccat        | ccat        | ccgcgttgc   | 7980      |      |
| caacgtgc    | aaga        | acgt       | ccat        | aa          | ccgcgttgc   | 8040      |      |
| ctaccagg    | gttccaa     | aggca      | tgtac       | ccat        | ccgcgttgc   | 8100      |      |

-continued

|             |              |             |             |             |              |       |
|-------------|--------------|-------------|-------------|-------------|--------------|-------|
| cagatgtggc  | ttcggtcccg   | agagcttgcg  | ccggggacaaa | accatcgccc  | tgatcatgaa   | 8160  |
| cagttagtgc  | agtaccggat   | tgcccaaggg  | cgtagcccta  | ccgcacccgca | ccgcttgtgt   | 8220  |
| ccgatttgc   | catggccggc   | accacatctt  | ccgcaaccag  | atcatccccc  | acacacgtat   | 8280  |
| cctcagctgt  | gtggcatttc   | accacgggtt  | ccgcatgttc  | accacgtgg   | getacttgat   | 8340  |
| ctggcggtt   | cggggtgtgc   | tcatgttaccc | cttcgaggagg | gagactatctt | tgccgagctt   | 8400  |
| gcaagactat  | aagattcaat   | ctggccctgt  | ggtgccacaca | ctatttagct  | tettgcgtaa   | 8460  |
| gaggacttc   | atcgacaaatgt | acgacactaa  | caacttgcac  | gagatgcggca | ggccgggggg   | 8520  |
| gcccgtcgc   | aaggaggtag   | gtgaggccgt  | ggccaaacgc  | ttccacccat  | caggcatccg   | 8580  |
| acagggttac  | ggccgtacac   | aaacaaaccac | cgccatcttc  | atccccccg   | aggggacgca   | 8640  |
| caaggcttgc  | gcagtaggca   | aggtggtgc   | cttcttcgag  | gctaagggtgg | tggacttgg    | 8700  |
| caccggtaag  | acaactgggt   | tgaaccacgg  | cgccgagctg  | tgccgtccgt  | ggcccatat    | 8760  |
| catggaggcc  | tacgttaaca   | accggaggc   | tacaaacgct  | ctcatcgaca  | aggacggctg   | 8820  |
| gtgcacacgc  | gggcacatc    | cctactggga  | cgaggacagcc | catttcata   | tcgtggaccg   | 8880  |
| gctgaagtc   | ctgatcaaat   | acaagggtt   | ccaggttgc   | ccagccgaa   | tggagagcat   | 8940  |
| cctgtcgca   | caccccaaca   | tcttcgacgc  | cggggttgc   | ggccctgcgg  | acgacgtatc   | 9000  |
| cgccgagct   | cccgccgcag   | tctgtgtgt   | ggggatgttgc | ggccctgcgg  | acgacgtatc   | 9060  |
| gatcggtgc   | tatgtggc     | gccaagggtt  | aaaccgcac   | aactgcgcgc  | ttgtgtgtgt   | 9120  |
| gttcgtggac  | gagggtgcct   | aaggactgac  | cgccaaagt   | gacgcccgc   | agatccgcga   | 9180  |
| gatttcatt   | aaggcacaaga  | aggggccgaa  | gatcgcgtgt  | taactcgacg  | cgaaaacgc    | 9240  |
| atagccgaa   | aacaaaaaa    | aaaaaaaac   | aaaaaaaaac  | aaaaaaaaac  | aaacacatta   | 9300  |
| aaacagctt   | tgggggtatc   | ccacccacac  | ggccattttgg | cgcttagact  | ctgttatcatc  | 9360  |
| ggtaaccttg  | tgcgtctgtt   | ttataccccc  | tcccccacact | gttaacttga  | agtaacacac   | 9420  |
| accgtatcac  | agtacgcgt    | gcacaccacg  | cacgttttgc  | tcaagacactt | ctgttacccc   | 9480  |
| ggactgtgat  | tcaatagact   | gctacgcgg   | tgttaaggaga | aaggctgtgt  | tateccgcct   | 9540  |
| actacttcga  | aaaacatgtt   | aacaccgtgg  | aagttgcaga  | tgttttgtct  | cagactacc    | 9600  |
| ccagtgtaga  | tcaggtcgat   | gagtccaccc  | atccccacac  | gggcacccgt  | ggccgtggctg  | 9660  |
| cgttggggc   | ctgcccattgg  | ggaaacccat  | gggacgtct   | aatacagaca  | ttgtgcgaag   | 9720  |
| agtcttattg  | gctatgtgt    | agtcttccgg  | ccctgtatgt  | cggtatccat  | taactcgccg   | 9780  |
| gcacacacc   | tcaaggccaga  | ggggcagtgt  | tgttaacggg  | caactctgc   | ggccgaaaccga | 9840  |
| ctactttggg  | tgtccgtgtt   | tcatttttat  | cttatactgg  | ctgttatgg   | tgacataatgg  | 9900  |
| gagatcgta   | ccatatagct   | attggattgg  | ccatccgggt  | actaatagag  | ctattatata   | 9960  |
| tcccttgg    | gggtttatac   | cacttagctt  | gaaagagggt  | aaaacattac  | aattttatgt   | 10020 |
| taatgttgaat | acagcaaaaat  | ggcaagatc   | tacatcgacg  | acggggacaa  | cgcccgagatc  | 10080 |
| gtgtcgagg   | ccatcaagac   | cattggccat  | gagggcgccca | ccggccgc    | ctgtcaggac   | 10140 |
| cagctgaaca  | tggagaacgg   | gggggtgttac | aaggccctgt  | acgacccgt   | gaggacgcgt   | 10200 |
| atgggtgtact | ccagcgacgc   | cattccccc   | cggtgttca   | tgaccacccg  | ggccgacacaa  | 10260 |
| cccgacccgc  | acgtatgtcc   | cgacgtgtatc | atcgacgacg  | tgagccggg   | gaagtccatc   | 10320 |
| agggaggacc  | acaagagctt   | cgacgacgt   | atccccgc    | agaagatcat  | cgacttggaa   | 10380 |
| ggcgccaaacc | ccgtgaccgt   | gatcaacgc   | tactgcaga   | tcaccaggag  | ggacttggac   | 10440 |
| ttccggatcg  | agaggctgtgg  | ccccacaa    | agccccac    | tctacatgt   | ctgtggacatc  | 10500 |
| gacggcagg   | tgttgcacaa   | ggccgcacgc  | aagacaaac   | gggacgc     | gaacaacgc    | 10560 |
| gccaagotgg  | ccgtggacaa   | gtgtcggtt   | taactgtatc  | tccgggttca  | aacgtatgtt   | 10620 |
| acgtcgaag   | gtgattgtca   | ccccccggaaa | gaccatattg  | tgacacaccc  | tcagatcatc   | 10680 |
| gccccaaatc  | ttacagccgc   | gggtgtcaaaa | accgcgttga  | cgtgtttaac  | atccctgtgt   | 10740 |
| ggaggatcg   | ccgtatattt   | tataatttgc  | ttgtgtgttgc | ctactatgt   | ggccatgtac   | 10800 |
| tgctgtacca  | accagaaacaa  | taatgtatc   | cagccacat   | tgccgaact   | cttacataga   | 10860 |
| actcgccgc   | attggcatgc   | cgcctttaaa  | tttttatttt  | attttttctt  | ttctttccg    | 10920 |
| aatcgattt   | tgttttataat  | atttcaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | tctagaaaaaa  | 10980 |
| aaaaaaaaaa  | aaaaaaaaaa   | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa  | aaaaaaaaaa   | 11040 |
| aaaaaaaaaa  | aaaaaaaaaa   | aaaaaaaaaa  | aaaaaaa     | aaaaaaa     | aaaaaaa      | 11075 |

```
SEQ ID NO: 119          moltype = RNA    length = 10851
FEATURE                  Location/Qualifiers
misc_feature             1..10851
note = Synthetic polynucleotide
source                   1..10851
mol_type = other RNA
organism = synthetic construct
```

```

SEQUENCE: 119
atgggcggcg catgagagaa gcccagacca attacccatt caaaatggag aaagttcacg 60
ttgacatcg ggaagacgc ccatttccta gagcttgcg cgggacgttc ccgcaggttg 120
aggtaaagc caaggcaggc actgtataatg accatgttac tgccagcgc ttgcgcatac 180
tggtcttcaa actgtatcgaa acggaggcttgg acccatccga cagatcctt gacatggaa 240
gtgcgcggcg ccgcagaatg tatttctaagc acaagtatca ttgtatctgt ccgtatggat 300
gtgcggaaaga tccggacaga ttgtataatg atgcaactaa gtcgaaagaaaa aactgttaagg 360
aaataacttg taaggatgtt gacaagaaaaa tgaaggagct ggccgcgtc atgagcgacc 420
ctgcacccgtt aactgtggact atgtgcctt acgcacgacg gtctgttcgca taaggaggcc 480
aagtctgtt ttaccaggat gtatccgcg tcgacggccc caccacccgtt taccacccgg 540
ccaacaaggcg ctgtgggggtt gcctacttgcg tcggcttcga caccacaccc ttcatgttca 600
agaacccgtgc cggcgccctac cccagctaca gcaaccaactg ggccgacgag accgtgtctga 660
ccggcaggagg catcgccctg tgccagacgc acgtgtatggaa gaggacccgg agaggatctga 720
gtatctgttgg gaagaatatact tcgtggccca gcaacaacgt gctgttcage gtggggcagca 780
ccatcttacca cgagaaaggagg gacccgttgcg ggacgttgcg cctggccacgc gtgttccacc 840
tgaggggcggaa gcagaactac acctgtgggtt gcgagacccat cgtgtatgtc gacggctacg 900
tggtgtaaagag gatcgccatc agccccggcc ttgtatggccaa gcccacggc tacggcccta 960
caatgtccacag ggagggttcc ctgtgttgcgaa aggtgtacccga caccctgttgcg cggcaggatggg 1020
tgtagcttccccc cgtgtgcaccatc taatgtggcccg ccacccgttgcg cggacccatgc accggccatcc 1080
tgccggccggc cgtgtggccgac gacggccggc agaaatgttgcgtt cgtggccctg aaccaggaggaa 1140
tcgtgttcaatc cggcaggacc cagaggaaaca ccaacacaaat gaaatgttgcgtt cgtgtggccgg 1200

```

-continued

---

|              |            |             |             |             |             |      |
|--------------|------------|-------------|-------------|-------------|-------------|------|
| tgggtggccca  | ggctttcgcc | aggtgggcca  | aggagtacaa  | ggaggaccag  | gaagacgaga  | 1260 |
| ggcccctggg   | cctgagggac | aggcagctgg  | tgatgggctg  | ctgctggggc  | ttcaggccgc  | 1320 |
| acaagatcac   | cagcatctac | aagaggcccc  | acacccagac  | catcatcaag  | gtgaacagcg  | 1380 |
| acttccacag   | cttcgtctg  | cccaggatcg  | gcagcaacac  | cctggagatc  | ggcctgagga  | 1440 |
| cccggtatcg   | gaagatgtc  | gaggaaacaca | aggagccca   | cccactgtac  | accggccgagg | 1500 |
| acgtgcagga   | ggccaagtgc | getgcgacg   | aggccaaggaa | ggtgaggggag | gcccggaaac  | 1560 |
| tgagggccgc   | cctgccaccc | ctgggtcg    | acgtggagga  | acccaccctg  | gaagccgacg  | 1620 |
| tggacctgtat  | gtcgaggag  | gcccggccg   | gaagcgtggaa | gacaccagg   | ggcctgatca  | 1680 |
| aggtgaccag   | ctacgacggc | gaggacaaga  | teggcageta  | cgccgtgtcg  | agcccaacagg | 1740 |
| ccgtgctgaa   | gtccggaaag | ctgagatcg   | tcacccact   | ggccggagcg  | gtgatcgta   | 1800 |
| tcacccacaa   | cgccggaaag | ggcggatcg   | ccgtggacgc  | ctaccaacggc | aagggtgtcg  | 1860 |
| tgcccgagggg  | ccacgcatac | cccgtgcagg  | acttccaggc  | cctgagcgag  | agcgcaccca  | 1920 |
| tcgtgtacaa   | cgagggggag | ttcgtaaca   | ggtagctgc   | ccatatcgcc  | acccacggcg  | 1980 |
| gagccctgaa   | caccggcag  | gaatactaca  | agacccgtgaa | gcccggcg    | cacgacggcg  | 2040 |
| agtacctgtat  | cgacatcgac | aggaaggact  | gcgtgaaagaa | agagctgtg   | accggccctgg | 2100 |
| gactgacccg   | cgagctgtg  | gaccaccc    | tccacgagg   | cgccctacgag | agcctgagga  | 2160 |
| ccagacccgc   | cgctccctac | cagggtccca  | ccatcgccgt  | gtacggcg    | cccggccagcg | 2220 |
| gaaagagggc   | catcatcaag | agcgcgtgt   | ccaagaaaaga | cctgggtgtc  | agcgcacaga  | 2280 |
| aagagaactcg  | cgccggatcg | atcaggacg   | tgaagaqat   | gaaaggccgt  | gacgtgaacg  | 2340 |
| cgcgacccgt   | ggacagcgtg | ctgctgaa    | gctgcaagca  | ccccgtggag  | accctgtaca  | 2400 |
| tcgacgaggc   | cttcgttgc  | caecggccg   | ccctgagggc  | cctgtatcgcc | atcatcaggc  | 2460 |
| ccaagaaage   | cggtgtgtc  | ggcggccccc  | agcagtgtcg  | tttcttcaac  | atgtgtgccc  | 2520 |
| tgaagggtc    | cttcacccac | gagatctgc   | cccaaggatgt | ccacaagagc  | atcaggaggc  | 2580 |
| ggtgtcaccaa  | gagcgtgacc | agcgtgtcg   | gcacccctgtt | ctacgacaag  | aaaatgagga  | 2640 |
| ccaccaaccc   | caaggagacc | aaaatcgta   | tegacaccac  | aggcagcaccc | aagcccaacg  | 2700 |
| aggacgacact  | gttcgttgc  | tgcttcagg   | gctgggtgaa  | gaogctgc    | atcgtactaca | 2760 |
| aggggcaacga  | gatcatgacc | gcccgtcc    | gcccggcc    | gaccggaaag  | ggcgtgtacg  | 2820 |
| ccgtgaggt    | caaggtgaac | gagaacccac  | tgtacgtcc   | caccagcgg   | cacgtgaacg  | 2880 |
| tgtgtgtgc    | caggaccgg  | gacaggatcg  | tgtggaaagac | cctggccggc  | gacccttgga  | 2940 |
| tcaagaccc    | gaccggcc   | taccccgcc   | acttcacccg  | caccatcgaa  | gagtggcagg  | 3000 |
| ccgagcagcga  | ggccatcat  | aggcatacc   | tggagggcc   | cgacccccc   | gacgtgttcc  | 3060 |
| agaaacaaggc  | caacgtgtc  | tggggcaagg  | ccctgggtc   | cggtgtgaa   | accggccggca | 3120 |
| tcgacatgc    | cacagagcag | tggaaacacc  | tggactactt  | cgagaccgac  | aaggccacaca | 3180 |
| gecccgagat   | cggtgtgaa  | cagetgtcg   | tgggttctt   | ccggcttgac  | ctggacagcg  | 3240 |
| gctgtgttgc   | cgcccccc   | gtgcactga   | gtatcgaa    | caaccactgg  | gacaacagcc  | 3300 |
| ccagccaaaa   | catgtacggc | ctgaaacaagg | aggtggtc    | gcagctgagc  | aggcggtacc  | 3360 |
| cacagctgcc   | caggccgtg  | gccaccggc   | gggtgtacg   | catgaacacc  | ggcacccctga | 3420 |
| ggaactacac   | ccccggatc  | aaccctgtgc  | ccgtgtac    | gcccgtcc    | cacgccttgg  | 3480 |
| tgtgtcacca   | caacgacgac | ccacagcgg   | atttcagtc   | cttcgtgac   | aagctgaaag  | 3540 |
| gcaggaccgt   | gtgtgtgt   | ggcggagaag  | tgacgtgtc   | ccggcaagat  | gtggacttgc  | 3600 |
| tgagcgacag   | ggccggggcc | accttccgg   | ccaggctgg   | cctcgccatc  | ccccggcgacg | 3660 |
| tgcccaagta   | cgacatcatc | ttcgtgtac   | tcaggacccc  | atacaatgtac | caccattacc  | 3720 |
| agcagtgtcg   | gggacccgtc | atcaatgtcg  | gtacgtgtac  | caagaaggcc  | tgcctgtcacc | 3780 |
| tgaacccccc   | aggcacctgc | tggtacatcg  | gctacggct   | ccggcggagg  | gccagcgaga  | 3840 |
| gcatcattgg   | ggccatcgcc | aggctgttca  | agttcagc    | gtgtgtcaaa  | cccaagagca  | 3900 |
| gcctgggg     | accgggggt  | ctgtgtgt    | tcatcggtt   | cgaccggaa   | gccaggaccc  | 3960 |
| acaaccccta   | caagctgac  | agcacccat   | caaacatcta  | cacccggc    | aggctgcacg  | 4020 |
| aggccgggtc   | cgcccccc   | taccacgtg   | tcagggggca  | tatcgccacc  | gccaccgg    | 4080 |
| gcgtgtatcat  | caacgcgtcc | aacagcaagg  | gcaagccccc  | aggcggagtg  | tgcggccccc  | 4140 |
| tgtacaagaa   | gttcccccgg | agcttcgtcg  | tgacgtccat  | cgagggtggc  | aaggccaggc  | 4200 |
| tgtgtggaa    | cgccgtgt   | caatcatcc   | acggcgtgg   | cccccaactt  | aacaagggtg  | 4260 |
| geggagggtg   | aggcgacaa  | cagctggcc   | aaagctacg   | gagcatggcc  | aagatcgta   | 4320 |
| acgacaataa   | ctacaagagc | gtggccatcc  | cactgtcg    | cacccggatc  | ttcagcgcc   | 4380 |
| acaaggacag   | gtgtggcc   | agcttcgtac  | acctgtcg    | cccccgtgg   | accaccgt    | 4440 |
| ccgacgtggc   | catctactgc | agggacaaga  | atgtgggat   | gaccctgt    | gaggccgtgg  | 4500 |
| ccagccggga   | ggccgtggaa | gagatctgc   | tcacgacg    | ctccagcg    | accqagcccg  | 4560 |
| acgcccgg     | gtgtgtgt   | caccccaaga  | gtccctggc   | cgccggagaa  | ggctacacgca | 4620 |
| ccagccgg     | caagaccc   | agtcacccgt  | agggccacaa  | gttccaccag  | ggccgttaagg | 4680 |
| acatcgccgt   | gtacaaatcg | atgtggcc    | tggccacc    | ggccaaacg   | cagggtgtca  | 4740 |
| tgtacatct    | ggggcggagc | atgtccagca  | tcaggacaa   | tgccggccgt  | gaggaaagcg  | 4800 |
| aggccagac    | accaccac   | accctgcct   | gtctgtcat   | ccacgtatg   | acaccggaga  | 4860 |
| gggtgtacgt   | gtgtgggg   | agcagatcg   | cggtgtc     | tcctcc      | ccat        | 4920 |
| tgcccaagta   | caggatcacc | ggcgtgtac   | agatccat    | cacccggatc  | accctgtca   | 4980 |
| gcccggatc    | atcccccc   | atcccccc    | ggaaatctt   | gttggggaccc | ccaccctgtgg | 5040 |
| acgagacacc   | cgagcaacg  | gcccggaa    | agagcaccga  | gggcacaccc  | gagcggccac  | 5100 |
| ccctgtatcatc | cgaggacgag | acaaggaccc  | ggacccac    | ccacccatatt | atcgaggaa   | 5160 |
| aggaaggagg   | cagcatcgac | ctgtgtgg    | acggcccc    | ccacccat    | ctgcagggtt  | 5220 |
| aggccggatc   | ccacggccca | cccacgtgt   | ccatgtgt    | ccacacggatc | ccacacggcc  | 5280 |
| gcgtgtatcg   | cggtgtac   | ctgacatcc   | ttggacaccc  | ggaggccccc  | agcgtgtac   | 5340 |
| ccggccgcac   | cagcgccg   | accacacgt   | acttcgtccaa | gagcatgg    | ttcctggcc   | 5400 |
| ggcccggtcc   | agctcccg   | accgtgttca  | ggaaacccac  | ccacccat    | cccaggacca  | 5460 |
| ggaccccaag   | ctgggttcc  | agcaggcc    | cgacgggg    | caggctgt    | agcacc      | 5520 |
| ccggcggtcc   | cagggtgtac | accagggg    | aactgggg    | cctgacaccc  | agcaggaccc  | 5580 |
| ccagcggtc    | cgtgtgtac  | actgtctgg   | tgtccaa     | accctggcgt  | aacagggtg   | 5640 |
| tcaccagggt   | ggaatccgt  | gccttgcgt   | cccagcaaca  | gagacgg     | gagccggcg   | 5700 |
| cctacatctt   | cagcagcgac | accggccagg  | gacacctgt   | gcaaaa      | agacgt      | 5760 |
| cgtgtgtac    | cgagggtgt  | ctggaggg    | ccgagctgt   | aatcgtatc   | gccccaggc   | 5820 |
| tggaccagg    | gaaggaggaa | ctgtgtgt    | ccgtgtgt    | agaaaactgt  | gctgtacccc  | 5880 |
| acaggagg     | gtaccacgac | aggaaagggt  | agaacatgt   | gaa         | ggccatcc    | 5940 |

-continued

---

|             |             |            |             |            |             |       |
|-------------|-------------|------------|-------------|------------|-------------|-------|
| tcctgcaggc  | cctgggacac  | tacctgaagg | ccggaggccaa | ggtggagtgc | tacaggacc   | 6000  |
| tgcaccccg   | gccactgtac  | agctccagcg | tgaacaggc   | cttctccagc | cccaagggtgg | 6060  |
| ccgtggagc   | ctgcaacgct  | atgctgaagg | agaacttccc  | caccgtggcc | agctactgca  | 6120  |
| tcateccccg  | gtacgacg    | tacctggaca | tggtgaggc   | cgccagctgc | tgcctggaca  | 6180  |
| ccgcccagg   | ctggcccgg   | aagctggagg | gttccccc    | aaaacacagc | taccctggagc | 6240  |
| ccacccatcg  | gagcggcgt   | cccagccca  | tccagaacac  | cctgcagaa  | gtgctggccg  | 6300  |
| ctgcccacca  | gaggaaactgc | aacgtgaccc | agatgaggga  | gctgccccgt | ctggacagcg  | 6360  |
| ctgccttcaa  | ctggggatgc  | ttcaagaat  | acgcctgc    | caacgagtac | tgggagac    | 6420  |
| tciaaggagaa | ccccatcagg  | ctgacgca   | agaacgtgg   | gaactacatc | accaagctga  | 6480  |
| aggggcccca  | ggccgtcgcc  | ctgttgc    | gatccacaa   | cctgaacatg | ctgcaggaca  | 6540  |
| tcccaatgg   | cagggtcg    | atggacgt   | agagggact   | gaaggtgaca | ccggccacca  | 6600  |
| agcacacccg  | ggagaggccc  | aagggtcagg | tgatccaggc  | cgctgacca  | ctggcacccg  | 6660  |
| ctacatgtg   | cgccatccac  | aggggactgg | tgaggcggt   | gaacgcccgt | ctgtgc      | 6720  |
| acatccacac  | ctgttgcac   | atgagcggc  | aggacttcga  | cgccatcata | gcccggact   | 6780  |
| tccagccccc  | cgactgcgt   | ctggagac   | acatcgcc    | cttcgacaa  | agcgaggatg  | 6840  |
| acgctatgg   | ctgtaccgt   | ctgatgatc  | tggaggact   | gggcgtgg   | gcccggatgc  | 6900  |
| taaccatgt   | cgaggctg    | tccggcg    | tcagctccat  | ccacctgccc | accaagacca  | 6960  |
| agttcaagg   | cgccgcata   | atgaaaag   | gaatgttcc   | gaccctgtc  | gtgaacac    | 7020  |
| tgtcaacat   | tgtgtcg     | acgagggt   | tgcgggag    | gctgacccgg | agccctgc    | 7080  |
| ctgccttcat  | cggcacgac   | aacatcg    | agggcgtgaa  | aagcgacaa  | ctgatggcc   | 7140  |
| acagggtgc   | cacccgtgc   | aaatgggg   | tgaagatcat  | cgacgccc   | gtgggggaga  | 7200  |
| aggggcccca  | cttctcg     | ggattctca  | tgtgcac     | cggtgc     | accgectgca  | 7260  |
| gggtggccg   | ccccctcg    | agggttca   | agctggccaa  | gcaactgg   | gctgacatg   | 7320  |
| agcacgacg   | tgacaggcgg  | agggccctg  | acgagggaa   | caccagg    | aacagggtgg  | 7380  |
| gcatctgt    | cgagctgt    | aaggcgtg   | agagcaggta  | cgagacccgt | ggcacccag   | 7440  |
| tcatctgt    | ggcttatg    | acactggcc  | gctccgtca   | gagcttctc  | tacctgagg   | 7500  |
| ggggccctat  | aactcttac   | ggctaact   | aatggactac  | gacatgt    | agtccgc     | 7560  |
| ggccgcacc   | atggagatg   | ccaaaaacat | taagaagg    | ccagcgocat | tctacc      | 7620  |
| cgaagacgg   | accgcggc    | agcagctg   | aaagccat    | aagcgtac   | ccctgtgc    | 7680  |
| cgccacccat  | gccttacc    | acgcacat   | cgagggtg    | attactac   | ccgactt     | 7740  |
| cgagatg     | gttcggcgt   | cagaact    | aaagcgt     | gggctgata  | caaaccatc   | 7800  |
| gatcgtgt    | tgcagcg     | atagctg    | gttctt      | ccctgtt    | gtgcctgt    | 7860  |
| catcgggt    | gtgtggcc    | cagtaac    | catctac     | gagcgcg    | tgctgaa     | 7920  |
| catgggatc   | agccgatcc   | ccgtgtt    | cgtgagca    | aaagggtgc  | aaaagatc    | 7980  |
| caacgtg     | aaaagatc    | cgatata    | aaagatcat   | atcatg     | gcaagac     | 8040  |
| ctaccaggc   | ttccaaagc   | tgtacac    | ctgtactt    | catttgoc   | ccggcttca   | 8100  |
| cgagatg     | ttcgtgccc   | agagttc    | ccgggacaa   | accatcg    | cc          | 8160  |
| caagtgtgc   | agttccggat  | tggccaa    | cgtagcc     | ccgcaccc   | ccgtt       | 8220  |
| ccgattcgt   | catgcgc     | accatctt   | ccgcaac     | atcatcc    | acaccgt     | 8280  |
| cctcagcg    | gtgccattt   | accacgg    | ccgcacat    | accacgtt   | accacgtt    | 8340  |
| ctggegtt    | cggtcg      | tcatgtac   | cttcgagg    | gagctattt  | tgcgc       | 8400  |
| gcaagactat  | aaggatca    | ctgcctgt   | ggtgccc     | ccat       | tttgc       | 8460  |
| gagcacttc   | atcgcac     | ttccaa     | ttttgt      | tttgc      | tttgc       | 8520  |
| gcccgtc     | aaaggatg    | gttggcc    | ggccaa      | ttccac     | caggcat     | 8580  |
| acagggtac   | ggcgtac     | aaacaacc   | cgccattt    | atcacc     | cc          | 8640  |
| caagcgtgc   | cgacttgg    | aggtgt     | cttctc      | ggcttgc    | cc          | 8700  |
| cacccgtt    | acacttgg    | tgaacc     | cg          | ggcgttgc   | cc          | 8760  |
| catgagccgc  | tacgttac    | accccg     | gat         | ccat       | cc          | 8820  |
| gctgcac     | ggcgtac     | cttact     | gg          | ggagc      | cc          | 8880  |
| gctgtgt     | ttgttac     | acaagg     | gttgc       | cc         | ggcc        | 8940  |
| cctgtgt     | cccccac     | tcttc      | gggggt      | cc         | ggcgttgc    | 9000  |
| ccggcag     | cccgcc      | tcgtgt     | gg          | aaacac     | cc          | 9060  |
| gatcgtgt    | tatgtgg     | gcaagg     | tttgc       | aa         | ctgttgc     | 9120  |
| gttcgtgt    | gagggtgt    | aaggact    | ccgg        | tttgc      | tttgc       | 9180  |
| gatttcatt   | aaggccaa    | agggg      | tttgc       | tttgc      | tttgc       | 9240  |
| atagccaa    | aaaaaaa     | aaaaaa     | tttgc       | tttgc      | tttgc       | 9300  |
| aaacagct    | ttgggtt     | ccacc      | tttgc       | tttgc      | tttgc       | 9360  |
| ggatcctt    | tcgcgtt     | ttat       | tttgc       | tttgc      | tttgc       | 9420  |
| accgatca    | atcgt       | ccgt       | tttgc       | tttgc      | tttgc       | 9480  |
| ggactgtgt   | tcaatag     | gttgc      | tttgc       | tttgc      | tttgc       | 9540  |
| actacttgc   | aaaatct     | tttgc      | tttgc       | tttgc      | tttgc       | 9600  |
| ccagtgt     | tcagg       | tttgc      | tttgc       | tttgc      | tttgc       | 9660  |
| cgttgg      | tcggccat    | ggaa       | tttgc       | tttgc      | tttgc       | 9720  |
| agtctat     | ttgttgc     | tttgc      | tttgc       | tttgc      | tttgc       | 9780  |
| gcacac      | tcaagg      | tttgc      | tttgc       | tttgc      | tttgc       | 9840  |
| ctacttgg    | tgtccgt     | tttgc      | tttgc       | tttgc      | tttgc       | 9900  |
| gagatcg     | ccatata     | tttgc      | tttgc       | tttgc      | tttgc       | 9960  |
| tcccttgg    | ttgttgc     | tttgc      | tttgc       | tttgc      | tttgc       | 10020 |
| taagtgt     | aaat        | tttgc      | tttgc       | tttgc      | tttgc       | 10080 |
| gtgtgc      | ccatca      | tttgc      | tttgc       | tttgc      | tttgc       | 10140 |
| cagctg      | ttgg        | tttgc      | tttgc       | tttgc      | tttgc       | 10200 |
| atgtgt      | ccagg       | tttgc      | tttgc       | tttgc      | tttgc       | 10260 |
| cccgac      | ccgtat      | tttgc      | tttgc       | tttgc      | tttgc       | 10320 |
| agggagg     | acaagg      | tttgc      | tttgc       | tttgc      | tttgc       | 10380 |
| ggcgca      | ccgtgt      | tttgc      | tttgc       | tttgc      | tttgc       | 10440 |
| ttccggat    | agagcgt     | tttgc      | tttgc       | tttgc      | tttgc       | 10500 |
| gacggc      | tttt        | tttgc      | tttgc       | tttgc      | tttgc       | 10560 |
| gccaagg     | tttt        | tttgc      | tttgc       | tttgc      | tttgc       | 10620 |
| aagctt      | cataga      | tttgc      | tttgc       | tttgc      | tttgc       | 10680 |

---

-continued

|                                                                   |                   |
|-------------------------------------------------------------------|-------------------|
| tttcttttct tttccgaatc ggattttgtt ttaatattt caaaaaaaaaaaaaaaa      | aaaaaaaaaaa 10740 |
| aaaaaaaaatcta gaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa | 10800             |
| aaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa a | 10851             |

What is claimed is:

1. A nucleic acid molecule comprising:

- (i) a first polynucleotide encoding one or more viral replication proteins, wherein the first polynucleotide is codon-optimized as compared to a reference polynucleotide of SEQ ID NO:54 encoding the one or more viral replication proteins; and
- (ii) a second polynucleotide comprising a first heterologous transgene encoding a first antigenic protein or a fragment thereof; and
- (iii) a 5' untranslated region (UTR) comprising the sequence of SEQ ID NO: 73 or SEQ ID NO:74.

2. The nucleic acid molecule of claim 1, further comprising a 3' untranslated region (UTR).

3. The nucleic acid molecule of claim 1, wherein the first antigenic protein is a viral protein, a bacterial protein, a fungal protein, a protozoan protein, a parasite protein, or a tumor protein.

4. The nucleic acid molecule of claim 3, wherein the viral protein is an orthomyxovirus protein, a paramyxovirus protein, a picornavirus protein, a flavivirus protein, a filovirus protein, a rhabdovirus protein, a togavirus protein, an arterivirus protein, a bunyavirus protein, an arenavirus protein, a reovirus protein, a bornavirus protein, a retrovirus protein, an adenovirus protein, a herpesvirus protein, a polyomavirus protein, a papillomavirus protein, a poxvirus protein, or a hepadnavirus protein.

5. The nucleic acid molecule of claim 3, wherein the first antigenic protein is an influenza virus protein, a respiratory syncytial virus (RSV) protein, a human immunodeficiency virus (HIV) protein, a hepatitis C virus (HCV) protein, a cytomegalovirus (CMV) protein, a Lassa Fever Virus (LFV) protein, an Ebola Virus (EBOV) protein, a *Mycobacterium* protein, a *Bacillus* protein, a *Yersinia* protein, a *Streptococcus* protein, a *Pseudomonas* protein, a *Shigella* protein, a *Campylobacter* protein, a *Salmonella* protein, a *Plasmodium* protein, or a *Toxoplasma* protein.

6. The nucleic acid molecule of claim 3, wherein the tumor protein is a kidney cancer, renal cancer, urinary bladder cancer, prostate cancer, uterine cancer, breast cancer, cervical cancer, ovarian cancer, lung cancer, liver cancer, stomach cancer, colon cancer, rectal cancer, oral cavity cancer, pharynx cancer, pancreatic cancer, thyroid cancer, melanoma, skin cancer, head and neck cancer, brain cancer, hematopoietic cancer, leukemia, lymphoma, bone cancer, or sarcoma protein.

7. The nucleic acid molecule of claim 1, wherein the second polynucleotide comprises at least two heterologous transgenes.

8. The nucleic acid molecule of claim 7, wherein a second heterologous transgene encodes a second antigenic protein or a fragment thereof, an immunomodulatory protein, or a reporter protein.

9. The nucleic acid molecule of claim 7, wherein the second polynucleotide further comprises a sequence encoding a 2A peptide, an internal ribosomal entry site (IRES), a subgenomic promoter, or a combination thereof, located between heterologous transgenes.

10. The nucleic acid molecule of claim 7, wherein the first and second heterologous transgenes encode viral proteins,

bacterial proteins, fungal proteins, protozoan proteins, parasite proteins, tumor proteins, immunomodulatory proteins, reporter proteins, or any combination thereof.

11. The nucleic acid molecule of claim 1, wherein the first polynucleotide is located 5' of the second polynucleotide.

12. The nucleic acid molecule of claim 11, further comprising an intergenic region located between the first polynucleotide and the second polynucleotide.

13. The nucleic acid molecule of claim 12, wherein the intergenic region comprises a sequence having at least 85% identity to the sequence of SEQ ID NO:77.

14. The nucleic acid molecule of claim 1, wherein the nucleic acid molecule is

(a) a DNA molecule; or

(b) an RNA molecule, wherein T is substituted with U.

15. The nucleic acid molecule of claim 14, wherein the DNA molecule further comprises a promoter located 5' of the 5' UTR, wherein the promoter is a T7 promoter, a T3 promoter, or an SP6 promoter.

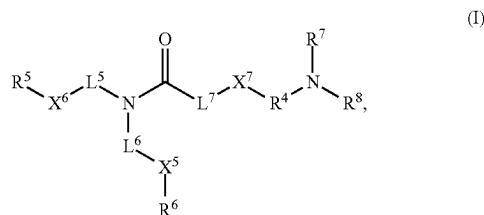
16. The nucleic acid molecule of claim 14, wherein the RNA molecule is a self-replicating RNA molecule.

17. The nucleic acid molecule of claim 14, wherein the RNA molecule further comprises a 5' cap having a Cap 1 structure, a Cap 1 ("<sup>7M</sup>A) structure, a Cap 2 structure, or a Cap 0 structure.

18. A pharmaceutical composition comprising the nucleic acid molecule of claim 1 and a lipid formulation.

19. The pharmaceutical composition of claim 18, wherein the lipid formulation comprises an ionizable cationic lipid.

20. The pharmaceutical composition of claim 19, wherein (a) the ionizable cationic lipid has a structure of Formula I:

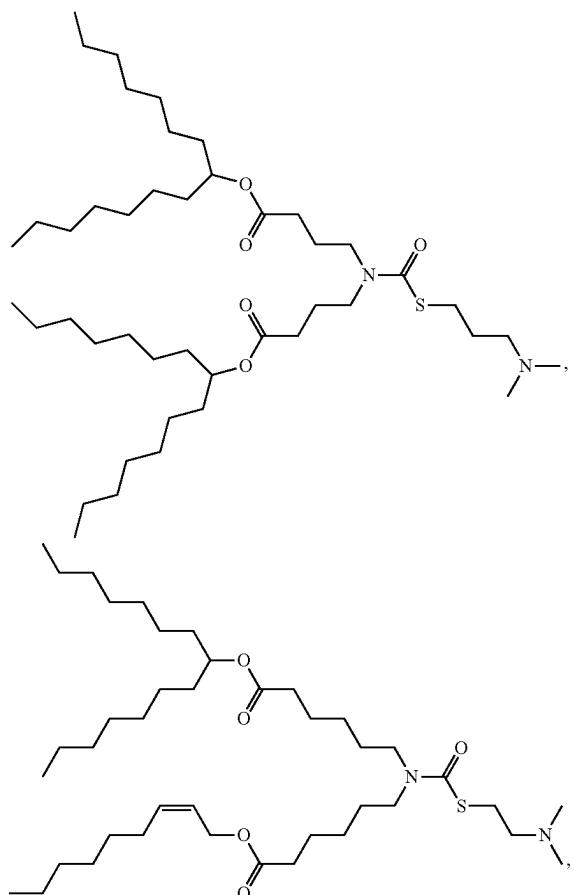


or a pharmaceutically acceptable salt thereof,

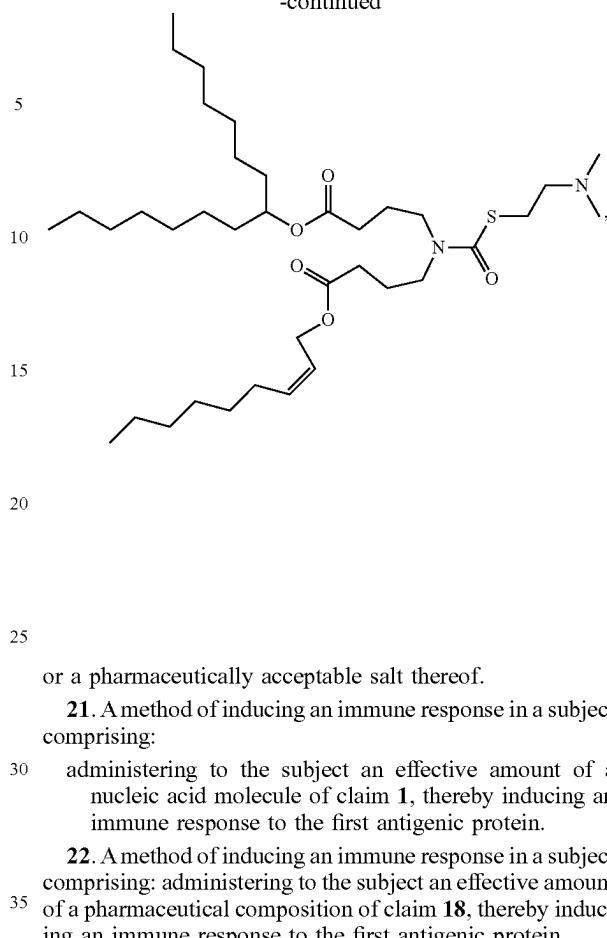
wherein R<sup>5</sup> and R<sup>6</sup> are each independently selected from the group consisting of a linear or branched C<sub>1</sub>-C<sub>31</sub> alkyl, C<sub>2</sub>-C<sub>31</sub> alkenyl or C<sub>2</sub>-C<sub>31</sub> alkynyl and cholesterol; L<sup>5</sup> and L<sup>6</sup> are each independently selected from the group consisting of a linear C<sub>1</sub>-C<sub>20</sub> alkyl and C<sub>2</sub>-C<sub>20</sub> alkenyl; X<sup>5</sup> is —C(O)O—, whereby —C(O)O—R<sup>6</sup> is formed or —OC(O)— whereby —OC(O)O—R<sup>6</sup> is formed; X<sup>6</sup> is —C(O)O— whereby —C(O)O—R<sup>6</sup> is formed or —OC(O)— whereby —OC(O)O—R<sup>5</sup> is formed; X<sup>7</sup> is S or O; L<sup>7</sup> is absent or lower alkyl; R<sup>4</sup> is a linear or branched C<sub>1</sub>-C<sub>6</sub> alkyl; and R<sup>7</sup> and R<sup>8</sup> are each independently selected from the group consisting of a hydrogen and a linear or branched C<sub>1</sub>-C<sub>6</sub> alkyl; or

**347**

(b) the ionizable cationic lipid has a structure of

**348**

-continued



or a pharmaceutically acceptable salt thereof.

**21.** A method of inducing an immune response in a subject comprising:

30 administering to the subject an effective amount of a nucleic acid molecule of claim 1, thereby inducing an immune response to the first antigenic protein.

**22.** A method of inducing an immune response in a subject comprising: administering to the subject an effective amount of a pharmaceutical composition of claim 18, thereby inducing an immune response to the first antigenic protein.

\* \* \* \* \*