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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
02 March 2023 (02.03.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/024500 A1

(51) International Patent Classification:
C12N 15/113 (2010.01) *C12P 19/34* (2006.01)

(21) International Application Number:
PCT/CN2022/082225

(22) International Filing Date:
22 March 2022 (22.03.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/CN2021/115029
27 August 2021 (27.08.2021) CN

(71) Applicant: **PEKING UNIVERSITY** [CN/CN]; No. 5 Yiheyuan Road, Haidian District, Beijing 100871 (CN).

(72) Inventors: **WEI, Wensheng**; Rm 205, No. 2 Integrated Science Research Building, Biomedical Pioneering Innovation Center, Peking University, No.5 Yiheyuan Road, Haidian District, Beijing 100871 (CN). **YI, Zongyi**; Rm 207, No. 2 Integrated Science Research Building, Biomedical Pioneering Innovation Center, Peking University, No.5 Yiheyuan Road, Haidian District, Beijing 100871 (CN).

(74) Agent: **BEIJING CAI HE LAW FIRM**; Room 1602, FuHai International Port, No. 17, Daliushu Road, Haidian District, Beijing 100081 (CN).

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

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

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: CONSTRUCTS AND METHODS FOR PREPARING CIRCULAR RNA

(57) Abstract: Provided are linear RNA precursors and constructs for preparing circular RNAs (circRNAs) comprising from the 5' end to the 3' end: (a) a first portion of an RNA element (such as an IRES), (b) an effector RNA sequence, and (c) a second portion of the RNA element, wherein the first portion of the RNA element and the second portion of the RNA element associate with each other to form a double-stranded region of at least 4 basepairs (bp) long, wherein the 5' end of the first portion of the RNA element and the 3' end of the second portion of the RNA element form a nick in the double-stranded region, and wherein the nick can be ligated by a RNA ligase. Also provided are methods of preparing circRNAs, and circRNAs prepared thereof.

WO 2023/024500 A1

CONSTRUCTS AND METHODS FOR PREPARING CIRCULAR RNA

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority benefit from International Patent Application No. PCT/CN2021/115029 filed on August 27, 2021, the content of which is incorporated herein by reference in its entirety.

SUBMISSION OF SEQUENCE LISTING ON ASCII TEXT FILE

[0002] The content of the following submission on ASCII text file is incorporated herein by reference in its entirety: a computer readable form (CRF) of the Sequence Listing (file name: 165392000741SEQLIST.TXT, date recorded: March 15, 2022, size: 10,097 bytes).

FIELD

[0003] The present application relates to constructs and methods for preparing circular RNA (circRNA) comprising an effector RNA (*e.g.*, coding RNA) sequence, circRNAs prepared thereof, and methods of use thereof.

BACKGROUND

[0004] In recent years, circular RNAs (circRNAs) with coding ability have been discovered, but there are still challenging in producing circular RNAs with coding ability accurately and efficiently *in vitro*. Current methods for generating circular RNAs mainly fall into two categories according to the enzyme used. The first category of methods leverage autocatalysis of a Group I intron to produce circular RNAs. The second category of methods use a T4 RNA ligase to produce circular RNAs. T4 RNA ligase 1-mediated ligation is a commonly used method for cyclizing RNA, but it does not accurately ligate the ends of a linear RNA due to insertions and deletions introduced during the ligation process. Furthermore, splint oligonucleotide is often required to bring the two ends of a linear RNA precursor together in order to allow efficient ligation of the linear RNA precursor into a circular RNA. The splint oligonucleotide has to be removed after ligation, and polymeric by-products may form.

[0005] There is need for methods of preparing circRNAs at high yields and without use of a splint oligonucleotide or introducing unnecessary sequences at the ends of a coding RNA.

BRIEF SUMMARY

[0006] The present application provides linear RNAs and constructs for preparing circRNAs by ligation, and methods for preparing circRNAs.

[0007] One aspect of the present application provides a linear RNA, comprising from the 5' end to the 3' end: (a) a first portion of a RNA element; (b) an effector RNA sequence; and (c) a second portion of the RNA element; wherein the first portion of the RNA element and the second portion of the RNA element associate with each other to form a double-stranded region of at least 4 basepairs (bp) long; wherein the 5' end of the first portion of the RNA element and the 3' end of the second portion of the RNA element form a nick in the double-stranded region; and wherein the nick can be ligated by a RNA ligase (*e.g.*, T4 RNA ligase 1 or T4 RNA ligase 2). In some embodiments, the double-stranded region is at least 6 bp, 8 bp, 10 bp, 12 bp, or more bp long. In some embodiments, the double-stranded region is about 6 bp long to about 25 bp long. In some embodiments, the double-stranded region has a minimum free energy of no more than about -8 kcal/mol (*e.g.*, no more than -9, -10, -11, -12, -13, -14, -15, -16 kcal/mol or lower) at 37°C and an ionic strength of 1.0M Na^+, *e.g.*, as predicted by RNAfold or Mfold.

[0008] In some embodiments according to any one of the linear RNAs described above, the RNA element is at least about 20 nt long, such as at least about 50, 100, 200, 300, 500, 600, 700, 800, 900, 1000 or more nt long.

[0009] In some embodiments according to any one of the linear RNAs described above, the double-stranded region comprises at least 2 bp 3' to the nick. In some embodiments, the double-stranded region comprises about any one of 2, 3, 4, 5, 6 or more basepairs 3' to the nick.

[0010] In some embodiments according to any one of the linear RNAs described above, the double-stranded region comprises at least 2 (*e.g.*, at least 4) bp 5' to the nick. In some embodiments, the double-stranded region comprises about any one of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more basepairs 5' to the nick.

[0011] In some embodiments according to any one of the linear RNAs described above, the first portion of the RNA element and the second portion of the RNA element are derived from a reference RNA element of N nucleotides, wherein N is an integer greater than 8; wherein the first portion of the RNA element comprises nucleotides X to N of the reference RNA element, wherein X is an integer greater than 1 and smaller than N; and wherein the second portion of the RNA element comprises nucleotides 1 to X-1 of the reference RNA element. In some embodiments, the

first portion of the RNA element and/or the second portion of the RNA element comprises a sequence that is exogenous to the reference RNA element. In some embodiments, the second portion of the RNA element comprises a first exogenous sequence at the 3' end, wherein the first portion of the RNA element comprises a second exogenous sequence that is complementary of the first exogenous sequence, and wherein the first exogenous sequence and the second exogenous sequence form basepairs flanking the 5' end of the nick. In some embodiments, the first exogenous sequence and/or the second exogenous sequence is at least 4 nt long. In some embodiments, the first exogenous sequence is GUUU.

[0012] In some embodiments according to any one of the linear RNAs described above, the RNA element is a naturally occurring RNA element or a derivative thereof. In some embodiments, the RNA element is a sgRNA. In some embodiments, the RNA element is an H/ACA box snoRNA.

[0013] In some embodiments according to any one of the linear RNAs described above, the effector RNA sequence is a coding RNA sequence. In some embodiments, the coding RNA sequence encodes a therapeutic polypeptide. In some embodiments, the therapeutic polypeptide is selected from the group consisting of an antigenic polypeptide (*e.g.*, an antigenic polypeptide of a pathogen, or a tumor antigen peptide), a functional protein (*e.g.*, an enzyme), a receptor protein (*e.g.*, a soluble receptor), and a targeting protein (*e.g.*, an antibody or antigen-binding fragment thereof).

[0014] In some embodiments according to any one of the linear RNAs described above, wherein the effector RNA sequence is a coding RNA sequence, the RNA element promotes translation of the coding RNA sequence. In some embodiments, the RNA element is an internal ribosomal entry site (IRES) or a portion thereof. In some embodiments, the IRES is derived from an IRES selected from the group consisting of Coxsackievirus B3 (CVB3) IRES, Enterovirus 71 (EV71) IRES, encephalomyocarditis virus (EMCV) IRES, picornavirus (PV) IRES, hepatitis C virus (HCV) IRES, adenovirus (AdV) IRES, human papillomavirus type 31 (HPV31) IRES, human herpesvirus (HHV) IRES, Rous sarcoma virus (RSV) IRES, classical swine fever virus (CSFV) IRES, FGF9 IRES, SLC7A1 IRES, and RUNX1 IRES. In some embodiments, the IRES is an IRES of a CVB3 virus or a derivative thereof. In some embodiments, the IRES of a CVB3 virus comprises the nucleotide sequence of SEQ ID NO: 1. In some embodiments, the first portion of the RNA element comprises nucleotides 386-749 of SEQ ID NO: 1, and wherein the second portion of the RNA element comprises nucleotides 1-385 of SEQ ID NO: 1. In some embodiments, the first portion of

the RNA element comprises the nucleotide sequence of SEQ ID NO: 3, and wherein the second portion of the RNA element comprises the nucleotide sequence of SEQ ID NO: 2. In some embodiments, the first portion of the RNA element comprises nucleotides 347-749 of SEQ ID NO: 1, and wherein the second portion of the RNA element comprises nucleotides 1-346 of SEQ ID NO: 1.

[0015] In some embodiments according to any one of the linear RNAs described above, wherein the effector RNA sequence is a coding RNA sequence, the linear RNA further comprising an in-frame 2A peptide (*e.g.*, T2A or P2A) coding sequence that is operably linked to the 3' end of the coding RNA.

[0016] In some embodiments according to any one of the linear RNAs described above, the effector RNA sequence is a sequence of a non-coding RNA selected from the group consisting of a guide RNA (gRNA), a deaminase-recruiting RNA (dRNA), a short interfering RNA (siRNA), a microRNA (miRNA), a short hairpin RNA (shRNA), and a long intervening non-coding (linc) RNA.

[0017] In some embodiments according to any one of the linear RNAs described above, the effector RNA sequence is at least about 50 nucleotides (nt) long, such as at least about any one of 60, 90, 120, 150, 200, 300, 400, 500, 600, 800, 1000, 1200, 1500 or more nt long. In some embodiments, the effector RNA sequence is about 50 nt to about 5000 nt long.

[0018] In some embodiments, there is provided a nucleic acid construct comprising a nucleic acid sequence encoding the linear RNA according to any one of the linear RNAs described above. In some embodiments, the nucleic acid construct comprises a T7 promoter operably linked to the nucleic acid sequence encoding the linear RNA.

[0019] Another aspect of the present application provides a method of preparing a circRNA, comprising: (a) contacting the linear RNA according to any one of the linear RNAs described above with a RNA ligase under conditions that allow ligation of the nick in the linear RNA to provide a circularized RNA product; and (b) isolating the circularized RNA product, thereby providing the circRNA. In some embodiments, the T4 RNA ligase is T4 RNA ligase 1. In some embodiments, the T4 RNA ligase is T4 RNA ligase 2. In some embodiments, the circRNA does not comprise nucleotide sequence that is exogenous to the linear RNA, such as insertion or deletion introduced by ligation. In some embodiments, the ligation does not require presence of a splint

oligonucleotide. In some embodiments, efficiency of the ligation is at least about 50% (*e.g.*, at least about 60^, 65%, 70%, 75%, 80%, 85%, 90% or higher).

[0020] In some embodiments according to any one of the methods of preparation described above, the method further comprises treating the circularized RNA product with RNAse R. In some embodiments, the method further comprises purifying the circularized RNA product. In some embodiments, the method further comprises obtaining the linear RNA by *in vitro* transcription of a nucleic acid construct comprising a nucleic acid sequence encoding the linear RNA.

[0021] Also provided is a circRNA prepared using the method according to any one of the methods of preparation described above.

[0022] In some embodiments, there is provided a circRNA formed by ligation of the nick in the linear RNA according to any one of the linear RNAs described above. In some embodiments, the circRNA is a circRNA vaccine.

[0023] In some embodiments, there is provided a method of treating or preventing a disease or condition in a subject in need thereof, comprising administering to the subject an effective account of any one of the circRNAs described above.

[0024] Further provided are compositions, kits and articles of manufacture for use in any one of the methods described above.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] FIG. 1 is a schematic showing production of a coding circular RNA using a T4 RNA ligase that ligates two portions of an IRES element in a linear RNA construct. The EGFP-coding RNA is the arch portion of the RNA transcript after *in vitro* transcription and ligation by a T4 RNA ligase. The nick in the linear RNA construct corresponds to the ligation site in the circular RNA product. Forward (F) and reverse (R) primers surrounding the ligation site were designed to detect ligation by reverse transcription followed by PCR amplification of a sequence encompassing the ligation site.

[0026] FIG. 2A shows the secondary structure of a wildtype CVB3 IRES as predicted by RNAfold. Each circle in the structure represents a nucleotide. The boxed region (also referred herein as the "engineered CVB3 IRES region") represents a region in the CVB3 IRES having highly stable double-stranded structures, which could promote formation of intramolecular double

strands. This region is enlarged in the right hand side of the figure. Linear RNA constructs with two engineered IRES split portions having split sites within Site 1 and Site 2, respectively, were designed for production of circular RNA via ligation of the engineered IRES split portions by a T4 RNA ligase.

[0027] FIG. 2B shows the predicted secondary structure of the engineered CVB3 IRES region in a first exemplary pair of engineered IRES split portions having the split site at Site 1 (nucleotide number 385) in a linear RNA transcript. The ligation site is marked with an arrow, and four nucleotides UUUG and its reverse complementary sequence were added in the engineered IRES split portions immediately neighboring the ligation site.

[0028] FIG. 2C shows the predicted secondary structure of the engineered CVB3 IRES region in a first exemplary pair of engineered IRES split portions having the split site at Site 2 (nucleotide number 346) in a linear RNA transcript. The ligation site is marked with an arrow, and four nucleotides UUUG and its reverse complementary sequence were added in the engineered IRES split portions immediately neighboring the ligation site.

[0029] FIG. 3 shows high-performance liquid chromatography (HPLC) analysis of a circular RNA sample produced by ligation of a linear RNA transcript of a construct based on the CVB3 IRES splits of FIG. 2B (*i.e.*, Site 1 ligation circular RNA) using T4 RNA ligase 1. In the top chromatograph, the circular RNA sample was not subject to RNAse R treatment. In the bottom chromatograph, the circular RNA sample was treated with RNAse R.

[0030] FIG. 4A shows expression EGFP in cells transfected with Site 1 and Site 2 ligated circular RNAs. FIG. 4B shows Western blot results detecting EGFP expression in cells after 24 hours of transfection with Site1 ligation circular RNA.

[0031] FIG. 5A shows PCR amplified products across the ligation junction using precursor RNA transcript and Site 1 ligation circular RNA samples. FIG. 5B shows Sanger sequencing of the ligation junction in Site 1 ligation circular RNA.

[0032] FIG. 6 shows a linear RNA precursor comprising split IRES portions flanking an effector RNA sequence. The split IRES portions reconstitute only a portion of a naturally occurring IRES.

[0033] FIG. 7 shows the secondary structure of an exemplary sgRNA and a potential split site.

[0034] FIG. 8 shows the secondary structure of an exemplary H/ACA box snoRNA bound to a target RNA, and a potential split site.

DETAILED DESCRIPTION

[0035] The present application provides linear RNA precursors and constructs for preparing circular RNAs (circRNAs) comprising an effector RNA (*e.g.*, coding RNA) sequence. The linear RNA precursors described herein comprise from the 5' end to the 3' end: a first split portion of an RNA element, an effector RNA (*e.g.*, coding RNA) sequence, and a second split portion of the RNA element, wherein the RNA element has a stable double-stranded region. The two split portions in the linear RNA precursor can associate with each other to reconstitute the RNA element, and form a nick in the stable double-stranded region. Ligation of the nick using an RNA ligase (such as T4 RNA ligase 1 or T4 RNA ligase 2) gives rise to a circRNA. In some embodiments, the RNA element is an internal ribosomal entry site (IRES), such as an IRES of a CVB3 virus. Also provided are methods of preparing circRNAs using the linear RNA precursors and constructs described herein. The methods allow *in vitro* production of circRNAs at high efficiency, accuracy, and without use of a splint oligonucleotide. Furthermore, split portions of the RNA elements used to promote intramolecular ligation in the methods described herein may serve functional roles, *e.g.*, promoting translation of a coding RNA sequence, thereby avoiding introducing additional ligation sequences that may have immunogenicity or reduce cargo size of the circRNA.

[0036] Accordingly, in some embodiments, there is provided a linear RNA comprising from the 5' end to the 3' end: (a) a first portion of an RNA element (such as an IRES), (b) an effector RNA (*e.g.*, coding RNA) sequence, and (c) a second portion of the RNA element, wherein the first portion of the RNA element and the second portion of the RNA element associate with each other to form a double-stranded region of at least 4 basepairs (bp) long, wherein the 5' end of the first portion of the RNA element and the 3' end of the second portion of the RNA element form a nick in the double-stranded region, and wherein the nick can be ligated by a RNA ligase (*e.g.*, T4 RNA ligase 2).

I. Definitions

[0037] Terms are used herein as generally used in the art, unless otherwise defined as follows.

[0038] The term "linear RNA" refers to a RNA molecule having a 5' end and a 3' end. A linear RNA may have secondary structures, including helices and loop regions.

[0039] The term "RNA element" refers to a RNA motif that folds into secondary structures, including double-stranded regions. In some embodiments, the RNA element is a cis-regulatory RNA element that regulates nucleic acid regions on the same molecule. The RNA element may be a naturally occurring RNA element (*e.g.*, IRES), or a man-made RNA element (*e.g.*, an aptamer).

[0040] The term "first portion" and "second portion" of an RNA element are used interchangeably herein with "'5' portion" and "3' portion", "first split portion" and "second split portion", or "first split" and "second split" of an RNA element to refer to engineered segments of a reference RNA element that can self-assemble into substantially the same secondary structures as the RNA element. In some embodiments, a reference RNA element can be divided or split at an internal position X into a first portion (i.e., 3' portion) containing nucleotides from position X to the end of the reference RNA element, and a second portion (i.e., 5' portion) containing nucleotides from position 1 to position X-1 of the reference RNA element. Position X is referred to as the "split site" in the reference RNA element. The first portion and/or the second portion may further contain exogenous sequences that are not present in the reference RNA element. The first portion and/or the second portion may also have deletions or substitutions with respect to the sequence in the reference RNA element. The exogenous sequence(s), deletion(s) and substitutions do not substantially alter (*e.g.*, do not reduce by more than 10%, 20%, 30%, 40%, or 50%) the functions and activities of the reconstituted RNA element by the first portion and the second portion of the RNA element.

[0041] The terms "polynucleotide," "nucleic acid," "nucleotide sequence," and "nucleic acid sequence" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof.

[0042] The terms "deaminase-recruiting RNA," "dRNA," "ADAR-recruiting RNA" and "arRNA" are used herein interchangeably to refer to an engineered RNA capable of recruiting an Adenosine Deaminase Acting on RNA (ADAR) to deaminate a target adenosine in an RNA. Exemplary dRNAs have been described, for example, in WO2021/008447, PCT/CN2021/071292, and PCT/CN2021/113290, the contents of which are incorporated herein by reference in their entirety.

[0043] As used herein, "guide RNA" and "gRNA" are used herein interchangeably to refer to RNA that is capable of forming a complex with a Cas protein and a target nucleic acid (*e.g.*, duplex DNA). A guide RNA may comprise a single RNA molecule or two or more RNA molecules

associated with each other via hybridization of complementary regions in the two or more RNA molecules. For example, a guide RNA for a Cas9 nuclease may comprise a crRNA and a tracrRNA, and a guide RNA for a Cas12a nuclease may only comprise a crRNA. The "crRNA" or "CRISPR RNA" comprises a guide sequence that has sufficient complementarity to a target sequence of a target nucleic acid (*e.g.*, duplex DNA), which guides sequence-specific binding of the CRISPR complex to the target nucleic acid. The "tracrRNA" or "trans-activating CRISPR RNA" is partially complementary to and base pairs with the crRNA, and may play a role in the maturation of the crRNA. A "single guide RNA" or "sgRNA" is an engineered guide RNA having both crRNA and tracrRNA fused to each other in a single molecule.

[0044] The term "therapeutic polypeptide" refers to a polypeptide having a therapeutic effect. A therapeutic polypeptide may be a naturally-occurring protein or an engineered functional variant thereof, including functional fragments, derivatives having one or more mutations (*e.g.*, insertion, deletion, substitution, *etc.*) to the amino acid sequence of the naturally-occurring protein, and fusion proteins comprising a naturally-occurring protein or fragment thereof. A therapeutic polypeptide may also be an engineered protein that does not have a naturally-occurring counterpart. Therapeutic polypeptide may have a single polypeptide chain or multiple polypeptide chains.

[0045] The term "antigenic polypeptide" refers to a polypeptide that can be used to trigger the immune system of a mammal to develop antibodies specific to the polypeptide or a portion thereof. Antigenic polypeptides described herein include naturally-occurring proteins, protein domains, and short peptide fragments derived from a naturally-occurring protein. An antigenic polypeptide may contain one or more known epitopes of a naturally-occurring protein. The antigenic polypeptide may comprise a carrier protein or multimerization protein that improves immunogenicity.

[0046] The term "functional protein" refers to a naturally-occurring protein, functional variants thereof, or an engineered derivative thereof that is functional in treating a genetic disease or condition. The disease or condition may be caused in whole or in part by a change, such as a mutation, in the wildtype, naturally-occurring protein corresponding to the functional protein.

[0047] The term "targeting protein" refers to a polypeptide that specifically binds to a target molecule. Targeting proteins described herein include both antibody-based and non-antibody based binding proteins or target-binding portions thereof.

[0048] The term "antibody" is used in its broadest sense and encompasses various antibody structures, including but not limited to monoclonal antibodies, polyclonal antibodies, multispecific antibodies (*e.g.*, bispecific antibodies), full-length antibodies and antigen-binding fragments thereof, so long as they exhibit the desired antigen-binding activity. The term "antigen-binding fragment" as used herein refers to an antibody fragment including, for example, a diabody, a Fab, a Fab', a F(ab')2, an Fv fragment, a disulfide stabilized Fv fragment (dsFv), a (dsFv)2, a bispecific dsFv (dsFv-dsFv'), a disulfide stabilized diabody (ds diabody), a single-chain Fv (scFv), an scFv dimer (bivalent diabody), a multispecific antibody formed from a portion of an antibody comprising one or more CDRs, a camelized single domain antibody, a nanobody, a domain antibody, a bivalent domain antibody, or any other antibody fragment that binds to an antigen but does not comprise a complete antibody structure.

[0049] As use herein, the terms "specifically binds," "specifically recognizing," and "is specific for" refer to measurable and reproducible interactions, such as binding between a target and a targeting moiety. For example, a targeting moiety that specifically recognizes a target (which can be an epitope) is a targeting moiety (*e.g.*, antibody) that binds this target with greater affinity, avidity, more readily, and/or with greater duration than its bindings to other molecules. In some embodiments, the extent of binding of a targeting moiety to an unrelated molecule is less than about 10% of the binding of the targeting moiety to the target as measured, *e.g.*, by a radioimmunoassay (RIA). In some embodiments, a targeting moiety that specifically binds a target has a dissociation constant (KD) of $\leq 10^{-5}$ M, $\leq 10^{-6}$ M, $\leq 10^{-7}$ M, $\leq 10^{-8}$ M, $\leq 10^{-9}$ M, $\leq 10^{-10}$ M, $\leq 10^{-11}$ M, or $\leq 10^{-12}$ M. In some embodiments, specific binding can include, but does not require exclusive binding. Binding specificity of the targeting moiety can be determined experimentally by methods known in the art. Such methods comprise, but are not limited to Western blots, ELISA, RIA, ECL, IRMA, EIA, BIACORETM and peptide scans.

[0050] The term "functional variant" of a reference protein refers to a variant polypeptide or polynucleotide derived from the reference protein or polynucleotide or a portion thereof, and the variant has substantially the same activity (*e.g.*, binding to a target or enzymatic activity) as the reference protein or polynucleotide. "Substantially the same activity" means an activity level that is at least about any one of 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more as the activity of the reference protein or polynucleotide.

[0051] The term "introducing" or "introduction" used herein means delivering one or more polynucleotides, such as circRNAs or one or more constructs including vectors as described herein, one or more transcripts thereof, to a host cell. The methods of the present application can employ many delivery systems, including but not limited to, viral, liposome, electroporation, microinjection and conjugation, to achieve the introduction of the circRNA or construct as described herein into a host cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids into mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding the circRNA of the present application to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (*e.g.* a transcript of a construct described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes for delivery to the host cell.

[0052] As used herein, "operably linked," when referring to a first nucleic acid sequence that is operably linked with a second nucleic acid sequence, means a situation when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter effects the transcription of the coding sequence. Likewise, the coding sequence of a signal peptide is operably linked to the coding sequence of a polypeptide if the signal peptide effects the extracellular secretion of that polypeptide. Generally, operably linked nucleic acid sequences are contiguous and, where necessary to join two protein coding regions, the open reading frames are aligned.

[0053] As used herein, "complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid by traditional Watson-Crick base-pairing. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (*i.e.*, Watson-Crick base pairing) with a second nucleic acid (*e.g.*, about 5, 6, 7, 8, 9, 10 out of 10, being about 50%, 60%, 70%, 80%, 90%, and 100% complementary respectively). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence form hydrogen bonds with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least about any one of 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of about 40, 50, 60, 70, 80, 100, 150, 200, 250 or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions.

[0054] As used herein, "treatment" or "treating" is an approach for obtaining beneficial or desired results including clinical results. For purposes of this application, beneficial or desired clinical results include, but are not limited to, one or more of the following: decreasing one more symptoms resulting from the disease, diminishing the extent of the disease, stabilizing the disease (*e.g.*, preventing or delaying the worsening of the disease), preventing or delaying the spread of the disease, preventing or delaying the occurrence or recurrence of the disease, delay or slowing the progression of the disease, ameliorating the disease state, providing a remission (whether partial or total) of the disease, decreasing the dose of one or more other medications required to treat the disease, delaying the progression of the disease, increasing the quality of life, and/or prolonging survival. Also encompassed by "treatment" is a reduction of pathological consequence of the disease. The methods of the present application contemplate any one or more of these aspects of treatment.

[0055] The terms "individual," "subject" and "patient" are used interchangeably herein to describe a mammal, including humans. In some embodiments, the individual is human. In some embodiments, the individual is a rodent, such as a mouse. In some embodiments, the individual suffers from a genetic disease or condition. In some embodiments, the individual suffers from a coronavirus infection. In some embodiments, the individual is at risk of contracting a coronavirus infection. In some embodiments, the individual is in need of treatment.

[0056] As is understood in the art, an "effective amount" refers to an amount of a composition sufficient to produce a desired therapeutic outcome (*e.g.*, stimulating the production of antibodies and improving immunity against one or more coronaviruses, reducing the severity or duration of, stabilizing the severity of, or eliminating one or more symptoms of a disease or condition). For therapeutic use, beneficial or desired results include, *e.g.*, decreasing one or more symptoms resulting from the disease (biochemical, histologic and/or behavioral), including its complications and intermediate pathological phenotypes presented during development of the disease, increasing the quality of life of those suffering from the disease, decreasing the dose of other medications required to treat the disease, enhancing effect of another medication, delaying the progression of the disease, and/or prolonging survival of patients. In some embodiments, an effective amount of the therapeutic agent may extend survival (including overall survival and progression free survival); result in an objective response (including a complete response or a partial response); relieve to some extent one or more signs or symptoms of the disease or condition; and/or improve

the quality of life of the subject. In some embodiments, an effective amount is a prophylactically effective amount, which is an amount of a composition sufficient to prevent or reduce the severity of one or more future symptoms of a disease or condition when administered to an individual who is susceptible and/or who may develop the disease or condition. For prophylactic use, beneficial or desired results include, *e.g.*, results such as eliminating or reducing the risk, lessening the severity of future disease, or delaying the onset of the disease (*e.g.*, delaying biochemical, histologic and/or behavioral symptoms of the disease, its complications, and intermediate pathological phenotypes presenting during future development of the disease).

[0057] As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms.

[0058] The present disclosure provides several types of compositions that are polynucleotide or polypeptide based, including variants and derivatives. These include, for example, substitutional, insertional, deletion and covalent variants and derivatives. The term "derivative" is synonymous with the term "variant" and generally refers to a molecule that has been modified and/or changed in any way relative to a reference molecule or a starting molecule.

[0059] As such, polynucleotides encoding peptides or polypeptides containing substitutions, insertions and/or additions, deletions and covalent modifications with respect to reference sequences, in particular, the polypeptide sequences disclosed herein, are included within the scope of this disclosure. For example, sequence tags or amino acids can be added to peptide sequences (*e.g.*, at the N-terminal or C-terminal ends). Sequence tags can be used for peptide detection, purification or localization. Alternatively, amino acid residues located at the carboxy and amino terminal regions of the amino acid sequence of a peptide or protein may optionally be deleted providing for truncated sequences. Certain amino acids (*e.g.*, C-terminal residues or N-terminal residues) alternatively may be deleted depending on the use of the sequence, as for example, expression of the sequence as part of a larger sequence that is soluble, or linked to a solid support.

[0060] The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free

from at least one other component with which they are naturally associated in nature and as found in nature.

[0061] As used herein, "expression" refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

[0062] The terms "polypeptide" or "peptide" are used herein to encompass all kinds of naturally occurring and synthetic proteins, including protein fragments of all lengths, fusion proteins and modified proteins, including without limitation, glycoproteins, as well as all other types of modified proteins (*e.g.*, proteins resulting from phosphorylation, acetylation, myristoylation, palmitoylation, glycosylation, oxidation, formylation, amidation, polyglutamylation, ADP-ribosylation, pegylation, biotinylation, *etc.*).

[0063] A "pharmaceutically acceptable carrier" refers to one or more ingredients in a pharmaceutical formulation, other than an active ingredient, which is nontoxic to a subject. A pharmaceutically acceptable carrier includes, but is not limited to, a buffer, excipient, stabilizer, cryoprotectant, tonicity agent, preservative, and combinations thereof. Pharmaceutically acceptable carriers or excipients have preferably met the required standards of toxicological and manufacturing testing and/or are included on the Inactive Ingredient Guide prepared by the U.S. Food and Drug administration or other state/federal government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in mammals, and more particularly in humans.

[0064] The term "package insert" is used to refer to instructions customarily included in commercial packages of therapeutic products, that contain information about the indications, usage, dosage, administration, combination therapy, contraindications and/or warnings concerning the use of such therapeutic products.

[0065] An "article of manufacture" is any manufacture (*e.g.*, a package or container) or kit comprising at least one reagent, *e.g.*, a medicament for treatment of a disease or condition (*e.g.*, coronavirus infection), or a probe for specifically detecting a biomarker described herein. In certain embodiments, the manufacture or kit is promoted, distributed, or sold as a unit for performing the methods described herein.

[0066] It is understood that embodiments of the invention described herein include "consisting" and/or "consisting essentially of" embodiments.

[0067] Reference to "about" a value or parameter herein includes (and describes) variations that are directed to that value or parameter per se. For example, description referring to "about X" includes description of "X".

[0068] As used herein, reference to "not" a value or parameter generally means and describes "other than" a value or parameter. For example, the method is not used to treat disease of type X means the method is used to treat disease of types other than X.

[0069] The term "about X-Y" used herein has the same meaning as "about X to about Y."

[0070] As used herein and in the appended claims, the singular forms "a," "an," or "the" include plural referents unless the context clearly dictates otherwise.

[0071] The term "and/or" as used herein a phrase such as "A and/or B" is intended to include both A and B; A or B; A (alone); and B (alone). Likewise, the term "and/or" as used herein a phrase such as "A, B, and/or C" is intended to encompass each of the following embodiments: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone).

II. Linear RNA and constructs

[0072] The present application provides linear RNA precursors (also referred herein as linear RNAs) and constructs for preparing circular RNAs (circRNAs) encoding polypeptides, such as therapeutic polypeptides. A linear RNA precursor described herein comprises two split portions of an RNA element flanking an effector RNA (*e.g.*, coding RNA) sequence. A circRNA can be obtained by ligating the ends of the split portions of the RNA element using a RNA ligase. The linear RNA may comprise any one of the first portions and second portions of the RNA elements described in Section A "RNA element" and any one of the effector RNA sequences described in Section B "Effector RNA sequence."

[0073] In some embodiments, there is provide a linear RNA, comprising from the 5' end to the 3' end: (a) a first portion of a RNA element; (b) an effector RNA sequence; and (c) a second portion of the RNA element; wherein the first portion of the RNA element and the second portion of the RNA element associate with each other to form a double-stranded region of at least 4 basepairs (bp) long (*e.g.*, about 6-25 bp long); wherein the 5' end of the first portion of the RNA element and the 3' end of the second portion of the RNA element form a nick in the double-

stranded region; and wherein the nick can be ligated by a RNA ligase. In some embodiments, the double-stranded region has a minimum free energy of no more than about -8 kcal/mol at 37°C and an ionic strength of 1.0M Na^+, *e.g.*, as predicted by RNAfold or Mfold. In some embodiments, the RNA element is at least about 20 nt (*e.g.*, at least 50 nt, or about 20-1000 nt) long. In some embodiments, the double-stranded region comprises at least 2 bp 3' to the nick. In some embodiments, the double-stranded region comprises at least 2 (*e.g.*, at least 4) bp 5' to the nick. In some embodiments, the RNA element is a naturally occurring RNA element or a derivative thereof. In some embodiments, the RNA element is a sgRNA. In some embodiments, the RNA element is an H/ACA box snoRNA. In some embodiments, the effector RNA sequence is about 50 nt to about 5000 nt long. In some embodiments, the effector RNA sequence is a sequence of a non-coding RNA selected from the group consisting of a guide RNA (gRNA), a deaminase-recruiting RNA (dRNA), a siRNA, a miRNA, a shRNA, and a long intervening non-coding (linc) RNA.

[0074] In some embodiments, there is provide a linear RNA, comprising from the 5' end to the 3' end: (a) a first portion of a RNA element; (b) a coding RNA sequence; and (c) a second portion of the RNA element; wherein the first portion of the RNA element and the second portion of the RNA element associate with each other to form a double-stranded region of at least 4 basepairs (bp) long (*e.g.*, about 6-25 bp long); wherein the 5' end of the first portion of the RNA element and the 3' end of the second portion of the RNA element form a nick in the double-stranded region; and wherein the nick can be ligated by a RNA ligase. In some embodiments, the double-stranded region has a minimum free energy of no more than about -8 kcal/mol at 37°C and an ionic strength of 1.0M Na^+, *e.g.*, as predicted by RNAfold or Mfold. In some embodiments, the RNA element is at least about 20 nt (*e.g.*, at least 50 nt, or about 20-1000 nt) long. In some embodiments, the double-stranded region comprises at least 2 bp 3' to the nick. In some embodiments, the double-stranded region comprises at least 2 (*e.g.*, at least 4) bp 5' to the nick. In some embodiments, the RNA element is a naturally occurring RNA element or a derivative thereof. In some embodiments, the RNA element promotes translation of the coding RNA sequence. In some embodiments, the coding RNA sequence is about 50 nt to about 5000 nt long. In some embodiments, the coding RNA sequence encodes a therapeutic polypeptide, such as an antigenic polypeptide, a functional protein, a receptor protein, or a targeting protein.

[0075] In some embodiments, there is provide a linear RNA, comprising from the 5' end to the 3' end: (a) a first portion of a RNA element; (b) an effector RNA (*e.g.*, coding RNA) sequence; and (c) a second portion of the RNA element; wherein the first portion of the RNA element and the second portion of the RNA element are derived from a reference RNA element of N nucleotides, wherein N is an integer greater than 8; wherein the first portion of the RNA element comprises nucleotides X to N of the reference RNA element, wherein X is an integer greater than 1 and smaller than N; and wherein the second portion of the RNA element comprises nucleotides 1 to X-1 of the reference RNA element; wherein the first portion of the RNA element and the second portion of the RNA element associate with each other to form a double-stranded region of at least 4 basepairs (bp) long (*e.g.*, about 6-25 bp long); wherein the 5' end of the first portion of the RNA element and the 3' end of the second portion of the RNA element form a nick in the double-stranded region; and wherein the nick can be ligated by a RNA ligase. In some embodiments, the double-stranded region has a minimum free energy of no more than about -8 kcal/mol at 37°C and an ionic strength of 1.0M Na^+, *e.g.*, as predicted by RNAfold or Mfold. In some embodiments, the RNA element is at least about 20 nt (*e.g.*, at least 50 nt, or about 20-1000 nt) long. In some embodiments, the double-stranded region comprises at least 2 bp 3' to the nick. In some embodiments, the double-stranded region comprises at least 2 (*e.g.*, at least 4) 5' to the nick. In some embodiments, the first portion of the RNA element and/or the second portion of the RNA element comprises a sequence that is exogenous to the reference RNA element. In some embodiments, the second portion of the RNA element comprises a first exogenous sequence at the 3' end, wherein the first portion of the RNA element comprises a second exogenous sequence that is complementary of the first exogenous sequence, and wherein the first exogenous sequence and the second exogenous sequence form base pairs flanking the 5' end of the nick. In some embodiments, the first exogenous sequence is GUUU. In some embodiments, the RNA element is a naturally occurring RNA element or a derivative thereof. In some embodiments, the effector RNA sequence is a coding RNA sequence. In some embodiments, the effector RNA sequence is about 50 nt to about 5000 nt long. In some embodiments, the RNA element promotes translation of the coding RNA sequence. In some embodiments, the coding RNA sequence encodes a therapeutic polypeptide, such as an antigenic polypeptide, a functional protein, a receptor protein, or a targeting protein.

[0076] In some embodiments, there is provided a linear RNA, comprising from the 5' end to the 3' end: (a) a first portion of an internal ribosomal entry site (IRES); (b) a coding RNA sequence; and (c) a second portion of the IRES; wherein the first portion of the IRES and the second portion of the IRES associate with each other to form a double-stranded region of at least 4 bp (*e.g.*, about 6-25 bp) long; wherein the 5' end of the first portion of the IRES and the 3' end of the second portion of the IRES form a nick in the double-stranded region; and wherein the nick can be ligated by a RNA ligase. In some embodiments, the double-stranded region has a minimum free energy of no more than about -8 kcal/mol at 37°C and an ionic strength of 1.0M Na^+, *e.g.*, as predicted by RNAfold or Mfold. In some embodiments, the double-stranded region comprises at least 2 bp 3' to the nick. In some embodiments, the double-stranded region comprises at least 2 (*e.g.*, at least 4) 5' to the nick. In some embodiments, the first portion of the IRES and the second portion of the IRES are derived from a reference IRES of N nucleotides, wherein N is an integer greater than 8; wherein the first portion of the IRES comprises nucleotides X to N of the reference IRES, wherein X is an integer greater than 1 and smaller than N; and wherein the second portion of the IRES comprises nucleotides 1 to X-1 of the reference IRES. In some embodiments, the first portion of the IRES and/or the second portion of the IRES comprises a sequence that is exogenous to the reference IRES. In some embodiments, the second portion of the IRES comprises a first exogenous sequence at the 3' end, wherein the first portion of the IRES comprises a second exogenous sequence that is complementary of the first exogenous sequence, and wherein the first exogenous sequence and the second exogenous sequence form base pairs flanking the 5' end of the nick. In some embodiments, the first exogenous sequence is GUUU. In some embodiments, IRES is derived from an IRES from a virus selected from the group consisting of CVB3, EV71, EMCV, PV, HCV, AdV, HPV31, HHV, RSV, and CSFV. In some embodiments, the IRES is derived from a human IRES, such as FGF9 IRES, SLC7A1 IRES or RUNX1 IRES. In some embodiments, the coding RNA sequence is about 50 nt to about 5000 nt long. In some embodiments, the coding RNA sequence encodes a therapeutic polypeptide, such as an antigenic polypeptide, a functional protein, a receptor protein, or a targeting protein.

[0077] In some embodiments, there is provided a linear RNA, comprising from the 5' end to the 3' end: (a) a first RNA portion comprising nucleotides 386-749 of a CVB3 IRES, wherein the nucleotide numbering is according to SEQ ID NO: 1; (b) a coding RNA sequence; and (c) a second RNA portion comprising nucleotides 1-385 of a CVB3 IRES; wherein the first RNA portion and

the second RNA portion associate with each other to form a double-stranded region comprising a nick formed by the 5' end of the first RNA portion and the 3' end of the second RNA portion; and wherein the nick can be ligated by a RNA ligase. In some embodiments, the first RNA portion and/or the second RNA portion comprises a sequence that is exogenous to the CVB3 IRES. In some embodiments, the second RNA portion comprises a first exogenous sequence at the 3' end, wherein the first RNA portion comprises a second exogenous sequence that is complementary of the first exogenous sequence, and wherein the first exogenous sequence and the second exogenous sequence form base pairs flanking the 5' end of the nick. In some embodiments, the first exogenous sequence is GUUU. In some embodiments, the first RNA portion comprises the nucleotide sequence of SEQ ID NO: 3, and wherein the second RNA portion comprises the nucleotide sequence of SEQ ID NO: 2. In some embodiments, the coding RNA sequence is about 50 nt to about 5000 nt long. In some embodiments, the coding RNA sequence encodes a therapeutic polypeptide, such as an antigenic polypeptide, a functional protein, a receptor protein, or a targeting protein.

[0078] In some embodiments, the linear RNA comprises, from 5' to 3' end, a first portion of IRES, an optional signal peptide, a coding RNA sequence, a sequence encoding a 2A peptide, and a second portion of IRES. In some embodiments, the linear RNA sequence comprises the nucleotide sequence of SEQ ID NO: 10 or 11.

[0079] In some embodiments, there is provided a nucleic acid construct comprising a nucleic acid sequence encoding any one of the linear RNAs described herein. In some embodiments, a T7 promoter is operably linked to the nucleic acid sequence encoding the linear RNA. In some embodiments, the T7 promoter comprises the sequence set forth in SEQ ID NO: 6. In some embodiments, the T7 promoter is capable of driving *in vitro* transcription.

[0080] In some embodiments, the nucleic acid construct is a plasmid. In some embodiments, the plasmids are obtained by cloning the sequence encoding the linear RNAs into a plasmid vector. Plasmids can be generated by techniques known in the art, such as Gibson cloning or cloning using restriction enzymes. In some embodiments, the plasmid vector includes an antibiotic expression cassette allowing antibiotic selection of bacteria expressing the plasmid. In some embodiments, the plasmids provided can be purified from bacteria and used for production of the linear RNA constructs. Any plasmid vector suitable for *in vitro* transcription of the linear RNA may be used. In some embodiments, the plasmids are linearized prior to *in vitro* transcription of the linear RNA.

In some embodiments, the recombinant plasmids are linearized by restriction enzyme digestion. In some embodiments, the recombinant plasmids are linearized by PCR amplification.

A. RNA element

[0081] The linear RNAs described herein comprise split portions of an RNA element that has extensive secondary structures, including double-stranded helical regions. Split portions (*i.e.*, first portion and second portion) can be engineered based on a reference RNA element by dividing the reference RNA element into a 5' portion (*i.e.*, the second portion in the linear RNA described herein) and a 3' portion (*i.e.*, the first portion in the linear RNA described herein) at a "split site", *i.e.*, a phosphodiester bond 5' to a nucleotide, in a stable double-stranded region of the reference RNA element. For example, if the split site is 5' to nucleotide X in an RNA element having N nucleotides, wherein X and N are integers and X is greater than 1 and smaller than N, then the first portion of the RNA element comprises nucleotides 1 to X-1 of the RNA element, and the second portion of the RNA element comprises nucleotides X to N of the RNA element. The split portions can be further engineered by introducing one or more exogenous sequence(s) into the first portion and/or the second portion of the RNA element, deleting one or more nucleotides in the first portion and/or the second portion of the RNA element, and/or substituting one or more nucleotides in the first portion and/or the second portion of the RNA element. The split portions of the RNA element in a linear RNA are capable of associating with each other to reconstitute the double-stranded region, while the 5' end of the first portion and the 3' end of the second portion give rise to a nick at the split site that can be ligated by a RNA ligase (*e.g.*, T4 RNA ligase 1 or T4 RNA ligase 2).

[0082] The split site in an RNA element can be chosen by first identifying a stable double-stranded region so that when the RNA element is split into two portions, which are placed at two ends of an effector RNA (*e.g.*, coding RNA) sequence in a linear RNA, the two portions can associate to form a double-stranded region with a nick. The low free energy of the double-stranded region promotes intramolecular ligation by a RNA ligase. Stable double-stranded regions in an RNA element can be identified computationally using known methods in the art. For example, RNAfold (Hofacker, I. L. Vienna RNA secondary structure server. Nucleic Acids Res. 31, 3429–3431 (2003)) and Mfold (M. Zuker, *Nucleic Acids Res.*, 31(13), 3406-15, (2003)) are web servers that can be used to computationally fold RNA molecules and predict secondary structures, including double-stranded regions, in RNA molecules. For example, FIG. 2A shows the secondary

structure of CVB3 IRES as predicted by RNAfold. A region with low free energy and double-stranded helical regions can be chosen, such as the boxed region of FIG. 2A, to engineer a split site within a double-stranded region, for example, at site 1 and site 2 as shown in FIG. 2A. The split site should not be adjacent to nucleotide positions that are known to affect the function of the RNA element. In some embodiments, a phosphodiester bond in the RNA element that can be easily accessed is chosen as the split site. In some embodiments, the double-stranded region has a high GC content, such as at least about any one of 40%, 50%, 60%, 70%, 80%, or higher GC basepairs.

[0083] In some embodiments, the split site is chosen close to the midpoint of the reference RNA element. In some embodiments, the first portion of the RNA element has substantially the same length as the second portion of the RNA element. For example, the length of the first portion of the RNA element differs from the length of the second portion of the RNA element by no more than about any one of 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5 or fewer nucleotides. In some embodiments, the length of the first portion of the RNA element differs from the length of the second portion of the RNA element by no more than about any one of 20%, 15%, 10%, 5%, or less.

[0084] In some embodiments, the split site is chosen in a double-stranded region that is at least 4 bp long, such as at least about any one of 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more bp long. In some embodiments, the double-stranded region is about any one of 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 bp long. In some embodiments, the double-stranded region is no more than about any one of 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, or 6 bp long. In some embodiments, the double-stranded region is about any one of 4-25, 4-15, 4-10, 6-25, 6-15, 6-10, 10-15, 10-20, or 10-25 bp. In some embodiments, the double-stranded region is about 6 bp to about 25 bp long. In some embodiments, the double-stranded region is part of a stem loop. In some embodiments, the double-stranded region is a helix. In some embodiments, the double-stranded region further comprises a bulge.

[0085] In some embodiments, the double-stranded region has a minimum free energy of no more than about -5 kcal/mol, such as no more than about any one of -5.5, -6, -6.5, -7, -7.5, -8, -8.5, -9, -9.5, -10, -10.5, -11, -11.5, -12, -12.5, -13, -14, -15 kcal/mol or lower, as predicted by RNAfold or Mfold. In some embodiments, the double-stranded region has a minimum free energy of about

any one of -15 to -5.5, -10 to -5.5, -8 to -5.5, -15 to -8, -12 to -8, or -10 to -8 kcal/mol. In some embodiments, the minimum free energy of the RNA element is no more than about any one of -50, -100, -150, -200, -250, -300 kcal/mol or lower.

[0086] In some embodiments, the RNA element is at least 20 nt long, such as at least about any one of 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 900 or more nt long. In some embodiments, the RNA element is no more than about any one of 1000, 900, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150, 100, 75, 50 or 20 nt long. In some embodiments, the RNA element is about any one of 20-50, 50-100, 20-200, 20-300, 20-500, 20-1000, 50-200, 50-1000, 100-1000, 100-500, 200-500, 200-1000, 300-1000, 300-800, 400-900, or 500-1000 nt long.

[0087] As shown in FIG. 1, in an exemplary linear RNA, the first portion of the RNA element flanks the 5' end of an effector RNA (*e.g.*, coding RNA), and the second portion of the RNA element flanks the 3' end of the coding RNA (*e.g.*, coding RNA). The first portion of the RNA element self-assemble with the second portion of the RNA element for form a double-stranded region with a nick at the split site, which serves as a substrate of a RNA ligase. In some embodiments, basepairs immediately adjacent to the 5' and/or 3' side of the nick promotes ligation of the nick by a RNA ligase. In some embodiments, the double-stranded region comprises at least 2 bp, such as 2, 3, 4, 5, 6, or more bp, 3' to the nick. In some embodiments, the double-stranded region comprises at least 2 bp, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, or more bp, 5' to the nick.

[0088] In some embodiments, the first portion of the RNA element and the second portion of the RNA element are derived from a reference RNA element of N nucleotides, wherein N is an integer greater than 8; wherein the first portion of the RNA element comprises nucleotides X to N of the reference RNA element, wherein X is an integer greater than 1 and smaller than N; and wherein the second portion of the RNA element comprises nucleotides 1 to X-1 of the reference RNA element. In some embodiments, the first portion of the RNA element consists of or consists essentially of nucleotides X to N of the reference RNA element, and wherein the second portion of the RNA element consists of or consists essentially of nucleotides 1 to X-1 of the reference element.

[0089] In some embodiments, compared to the reference RNA element, the RNA element has no exogenous sequence. In some embodiments, the RNA element reconstituted by ligating the 5'

end of the first portion of the RNA element and the 3' end of the second portion of the RNA element is identical to the reference RNA element.

[0090] In some embodiments, the RNA element after ligation of the nick is a portion of the reference RNA element. The portion contains the stable double-stranded region. The portion may or may not have the function of the reference RNA element.

[0091] In some embodiments, the RNA element after ligation of the nick contains all nucleotides of the reference RNA element. In some embodiments, the RNA element after ligation of the nick is the reference RNA element.

[0092] In some embodiments, compared to the reference RNA element, the RNA element comprises one or more exogenous sequences. In some embodiments, the first portion of the RNA element comprises a sequence that is exogenous to the reference RNA element. In some embodiments, the second portion of the RNA element comprises a sequence that is exogenous to the reference RNA element. In some embodiments, the first portion of the RNA element comprises a first sequence that is exogenous to the reference RNA element, and the second portion of the RNA element comprises a second sequence that is exogenous to the reference RNA element.

[0093] For example, exogenous sequences can be introduced to the first portion of the RNA element or the second portion of the RNA element in order to further stabilize the double-stranded region comprising the nick, thereby promoting ligation of the nick by a RNA ligase. In some embodiments, the first portion of the RNA element comprises a first exogenous sequence at the 3' end and the second portion of the RNA element comprises a second exogenous sequence that is complementary to the first exogenous sequence, wherein the first exogenous sequence and the second exogenous sequence form basepairs flanking the 5' end of the nick. In some embodiments, the first exogenous sequence and the second exogenous sequence are each at least 4 nt long, such as about any one of 4, 5, 6, 7, 8, or more nt long. In some embodiments, the first exogenous sequence is GUUU, and the second exogenous sequence is AAAC.

[0094] The RNA element can be a naturally occurring RNA element, such as a regulatory RNA element, a derivative thereof, or an artificial RNA element. In some embodiments, the RNA element is a cis-regulatory RNA element.

[0095] In some embodiments, the RNA element is a regulatory RNA element that promotes translation of the coding RNA sequence. In some embodiments, the RNA element is an internal ribosomal entry site (IRES). In some embodiments, the RNA element is a full-length IRES. In

some embodiments, the RNA element is a portion of an IRES comprising the stable double-stranded region. The portion of IRES may not be functional for promoting protein translation.

[0096] For example, as shown in FIG. 6, in an exemplary linear RNA, the first portion and the second portion are derived from a stem loop portion of a CVB3 IRES which contains a highly stable double-stranded region. The first portion and the second portion flank an effector RNA sequence, which may be a coding RNA or a non-coding RNA (such as a gRNA, a dRNA, a miRNA, a siRNA, a shRNA, or a lincRNA).

[0097] In some embodiments, the IRES is a viral IRES sequence. In non-limiting examples, the IRES sequence can be an IRES from a virus selected from the group consisting of CVB3, EV71, EMCV, PV, HCV, AdV, HPV31, HHV, RSV, and CSFV. *See*, for example, "Searching for IRES," RNA. 2006 Oct; 12(10): 1755–1785, which is incorporated herein by reference in its entirety. In some embodiments, the IRES sequence is a cellular IRES sequence. In some embodiments, the IRES is derived from a human IRES, such as FGF9 IRES, SLC7A1 IRES or RUNX1 IRES. Exemplary IRES sequences can be found at databases, such as reprod.njmu.edu.cn/cgi-bin/iresbase/view_eukaryote.php, and cobishss0.im.nuk.edu.tw/Human_IRES_Atlas/, which are incorporated herein by reference in their entirety.

[0098] In some embodiments, the RNA element is an IRES of a CVB3 virus or a derivative thereof. An exemplary sequence of a CVB3 IRES is SEQ ID NO: 1. The secondary structure of CVB3 IRES predicted by RNAfold is shown in FIG. 2. The CVB3 IRES can be split at two different split sites within stable double-stranded regions. Split site 1 is 5' to nucleotide 386, wherein the numbering is based on SEQ ID NO: 1. In some embodiments, the first portion of the RNA element comprises nucleotides 386-749 of SEQ ID NO: 1, and the second portion of the RNA element comprises nucleotides 1-385 of SEQ ID NO: 1. In some embodiments, the first portion of the RNA element comprises the nucleotide sequence of SEQ ID NO: 3, and the second portion of the RNA element comprises the nucleotide sequence of SEQ ID NO: 2. Split site 2 is 5' to nucleotide 347, wherein he numbering is based on SEQ ID NO: 1. In some embodiments, the first portion of the RNA element comprises nucleotides 347-749 of SEQ ID NO: 1, and the second portion of the RNA element comprises nucleotides 1-346 of SEQ ID NO: 1. In some embodiments, the first portion of the RNA element comprises the nucleotide sequence of SEQ ID NO: 5, and the second portion of the RNA element comprises the nucleotide sequence of SEQ ID NO: 4.

[0099] In some embodiments, the RNA element is a guide RNA, such as a single guide RNA (sgRNA) that is operable with a Cas nuclease. In some embodiments, the gRNA or sgRNA is for gene editing. Exemplary Cas nucleases include, but are not limited to, Cas9, Cas12a, Cas12b, Cas12c, Cas12d, Cas12f, Cas12g, Cas12h, Cms1, Cas12i, Cas12j, Cas12k and CasX. FIG. 7 shows a potential split site in a sgRNA. In some embodiments, wherein the RNA element is a gRNA (*e.g.*, sgRNA), the effector RNA encodes a Cas nuclease (*e.g.*, Cas9).

[0100] In some embodiments, the RNA element is a small nucleolar RNA (snoRNA) or a portion thereof. In some embodiments, the RNA element is an H/ACA box of a snoRNA. H/ACA box snoRNAs have a common secondary structure consisting of two hairpins and two single-stranded regions. FIG. 8 shows a potential split site in a H/ACA snoRNA. Exemplary H/ACA snoRNA sequences can be found in snoStrip Webserver at snostrip.bioinf.uni-leipzig.de. The H/ACA snoRNA may be used to convert a uridine into pseudouridine in a target RNA. H/ACA snoRNA sequences could be used for mRNA editing. *See*, for example, US8603457B2.

[0101] The first portion of the RNA element and the second portion of the RNA element serve as ligation sequences in the linear RNA. Other than the first portion and the second portion of the RNA element, the linear RNAs described herein have no additional sequences for ligation. In some embodiments, the linear RNA does not comprise a 5' ligation sequence at the 5' end, and a 3' ligation sequence at the 3' end, wherein the 5' ligation sequence and the 3' ligation sequence can be ligated to each other via a ligase (*e.g.*, T4 RNA ligase) in the presence of a splint oligonucleotide that hybridizes with the 5' ligation sequence and the 3' ligation sequence.

B. Effector RNA sequence

[0102] The linear RNA precursors and circRNAs described herein comprise an effector RNA sequence, which may be a coding RNA sequence or a non-coding RNA sequence. Exemplary non-coding RNAs include, but are not limited to, guide RNAs (gRNA, including single guide RNA or sgRNA), a deaminase-recruiting RNA (dRNA), a small RNA (such as a microRNA, a short hairpin RNA, or a small interfering RNA), or a long intervening non-coding RNA (lincRNA).

[0103] In some embodiments, the effector RNA sequence is at least about 50 nt long, such as at least about any one of 100, 150, 200, 300, 600, 900, 1200, 1500, 2000, 3000, 4000, 5000, or more nt long. In some embodiments, the effector RNA sequence is no more than about any one of 5000, 4000, 3000, 2000, 1500, 1200, 900, 600, 300, 200, 150, or 100 nt long. In some embodiments, the

effector RNA sequence is about any one of 50-100, 100-500, 500-1000, 1000-2000, 2000-5000, 50-5000, 100-5000, 100-3000, 500-5000, 500-2500, 2500-5000, or 1000-5000 nt long.

[0104] In some embodiments, the effector RNA sequence is a coding RNA sequence, which encode any polypeptide of interest. In some embodiments, the polypeptide is at least about 15 amino acids long, such as at least about any one of 20, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more amino acids long. In some embodiments, the polypeptide is no more than about any one of 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 50, or 20 amino acids long. In some embodiments, the polypeptide is about any one of 20-50, 50-100, 20-200, 20-500, 20-1000, 50-500, 50-1000, 100-500, 100-1000, 200-1000 or 500-1000 amino acids long.

[0105] In some embodiments, the coding RNA sequence encodes a therapeutic polypeptide. In some embodiments, the therapeutic polypeptide is an antigenic polypeptide, a functional protein, a receptor protein or a targeting protein (*e.g.*, an antibody).

[0106] In some embodiments, the coding RNA sequence encodes an antigenic polypeptide. A circRNA vaccine may be prepared using a linear RNA comprising a coding RNA sequence encoding an antigenic polypeptide. An antigenic polypeptide comprises at least one epitope recognizable by a T cell receptor (TCR). In some embodiments, the antigenic polypeptide is a full-length protein or a fragment thereof, or an antigenic fusion protein that can trigger an immune response in a subject. In some embodiments, the antigenic polypeptide is a short peptide of no more than 100 amino acids long. The antigenic polypeptide can be a naturally derived peptide fragment from a protein antigen containing one or more epitopes, or an artificially designed peptide with one or more natural epitope sequences, wherein a peptide linker may optionally be placed in between adjacent epitope sequences. In some embodiments, the antigenic polypeptide comprises a single epitope of an antigenic protein. In some embodiments, the antigenic polypeptide comprises about any one of 1, 2, 3, 4, 5, 10 or more epitopes from a single antigenic protein. In some embodiments, the antigenic polypeptide comprises epitopes from a plurality (*e.g.*, 2, 3, 4, 5, 10 or more) of different antigenic proteins. In some embodiments, the antigenic polypeptide comprises a Major Histocompatibility Complex (MHC) class I-restricted epitope. In some embodiments, the antigenic polypeptide comprises a MHC class II-restricted epitope. In some embodiments, the antigenic polypeptide comprises both MHC class I-restricted and MHC class II-restricted epitopes.

[0107] In some embodiments, the antigenic polypeptide is an antigenic protein or fragment thereof or a variant thereof from a pathogenic agent, such as a bacterium or a virus. In some

embodiments, the antigenic polypeptide is an antigenic protein or fragment of a coronavirus, such as SARS-CoV2, including variants thereof. In some embodiments, the antigenic polypeptide comprises a Spike (S) protein or a fragment thereof or a variant thereof of a coronavirus, such as SARS-CoV, MERS-COV, or SARS-CoV-2. CircRNA vaccines have been described, for example, in PCT/CN2021/074998, which is incorporated herein by reference in its entirety. The linear RNAs and constructs described herein may be used to prepare any one of the known circRNA vaccines in the art.

[0108] In some embodiments, the antigenic polypeptide is an antigenic protein or fragment thereof or a variant thereof of a self-antigen, such as an antigen involved in a disease or condition. In some embodiments, the antigenic polypeptide is a tumor antigen peptide. Tumor antigen peptide sequences are known in the art and can be found at public databases, such as the Cancer Antigenic Peptide Database (van der Bruggen P et al. (2013) "Peptide database: T cell-defined tumor antigens." *Cancer Immunity*. URL: caped.icp.ucl.ac.be). The coding RNA sequence in the linear RNA or circRNA described herein may encode any of the known tumor antigen peptides or combinations thereof. In some embodiments, the antigenic polypeptide comprises an epitope of a tumor associated antigen (TAA). In some embodiments, the antigenic polypeptide comprises an epitope of a tumor specific antigen. In some embodiments, the antigenic polypeptide comprises an epitope of a neoantigen, *i.e.*, newly acquired and expressed antigens present in tumor cells of an individual.

[0109] In some embodiments, the amino acid sequences of one or more epitope peptides are predicted based on the sequence of the antigen protein (including neoantigens) using a bioinformatics tool for T cell epitope prediction. Exemplary bioinformatics tools for T cell epitope prediction are known in the art, for example, see Yang X. and Yu X. (2009) "An introduction to epitope prediction methods and software" *Rev. Med. Virol.* 19(2): 77-96. In some embodiments, the sequence of the antigen protein is known in the art or available in public databases. In some embodiments, the sequence of the antigen protein (including neoantigens) is determined by sequencing a sample (such as a tumor sample) of the individual being treated.

[0110] In some embodiments, the antigenic polypeptide comprises a multimerization domain, such as a dimerization domain, a trimerization domain, or a domain that mediates formation of higher order multimers. In some embodiments, the multimerization domain is a trimerization domain. In non-limiting examples, the multimerization domain comprises a C-terminal Foldon

(Fd) domain of a T4 fibritin protein, wherein the C-terminal Foldon domain is the domain that mediates trimerization of the T4 fibritin protein. In another example, the multimerization domain comprises a GCN4-based isoleucine zipper (IZ) domain based on the trimerization domain of the GCN4 transcriptional activator from *Saccharomyces cerevisiae*. In some embodiments, the GCN4 IZ domain or T4 fibritin Fd domain can be modified to reduce their immunogenicity according to known techniques in the art. For example, the GCN4 IZ domain can be modified with *N*-linked glycosylation sites to reduce its immunogenicity (Sliepen et al. Immunosilencing a Highly Immunogenic Protein Trimerization Domain. The Journal of Biol. Chem. Vol. 290, No. 12, pp. 7436–7442).

[0111] In some embodiments, the antigenic polypeptide further comprises an immunogenic carrier protein. In some embodiments, the antigenic polypeptide comprises an epitope peptide conjugated to an immunogenic carrier protein. Exemplary immunogenic carrier proteins include, but are not limited to, tetanus toxoid (TT), diphtheria toxoid (DT), modified cross-reacting material of diphtheria toxin (CRM197), meningococcal outer membrane protein complex (OMPC), and Hemophilus influenzae protein D (HiD).

[0112] In some embodiments, the coding RNA sequence encodes a targeting protein. In some embodiments, the targeting protein is an antibody or an antigen-binding fragment thereof.

[0113] In some embodiments, the coding RNA sequence encodes an antibody. In some embodiments, the therapeutic polypeptide is a neutralizing antibody, i.e., an antibody that blocks an interaction between a protein and its binding partner. In some embodiments, the antibody inhibits activity of a protein, *e.g.*, by blocking binding of the protein to a binding partner. In some embodiments, the targeting protein is a therapeutic antibody. In some embodiments, the antibody is a checkpoint inhibitor, *e.g.*, an antibody inhibitor of CTLA-4, PD-1, or PD-L1. In some embodiments, the antibody specifically binds a cell surface antigen, such as a tumor antigen. Exemplary tumor antigens include, but are not limited to, glioma-associated antigen, carcinoembryonic antigen (CEA), β-human chorionic gonadotropin, alphafetoprotein (AFP), lectin-reactive AFP, thyroglobulin, RAGE-1, MN-CAIX, human telomerase reverse transcriptase, RU1, RU2 (AS), intestinal carboxyl esterase, mut hsp70-2, M-CSF, prostase, prostate-specific antigen (PSA), PAP, NY-ESO-1, LAGE-la, p53, prostein, PSMA, HER2/neu, survivin and telomerase, prostate-carcinoma tumor antigen-1 (PCTA-1), MAGE, ELF2M, neutrophil elastase, ephrinB2, CD22, insulin growth factor (IGF)-I, IGF-II, IGF-I receptor and mesothelin. In some

embodiments, the antibody specifically binds a target antigen on a pathogenic agent, such as a bacterium or a virus.

[0114] The antibody can be an antigen-binding fragment of an antibody, *e.g.*, a portion or fragment of an intact or complete antibody having fewer amino acid residues than the intact or complete antibody, which is capable of binding to an antigen or competing with the intact antibody (i.e., the intact antibody from which the antigen-binding fragment is derived) for binding to an antigen. Antigen-binding fragments can be prepared by recombinant DNA techniques, or by enzymatic or chemical cleavage of intact antibodies. Antigen binding fragments include, but are not limited to, Fab ', F (ab')2Fv, single chain Fv (scFv), single chain Fab, diabody (diabody), single domain antibody (sdAb, nanobody), camel Ig, Ig NAR, F (ab)'3Fragment, bis-scFv, (scFv)2Minibodies, diabodies, triabodies, tetradiabodies, disulfide stabilized Fv proteins ("dsFv"). In some embodiments, the neutralizing antibody can be a genetically engineered antibody, such as a chimeric antibody (*e.g.*, humanized murine antibodies), heteroconjugate antibody (*e.g.*, bispecific antibodies), or antigen-binding fragments thereof.

[0115] In some embodiments, the antibody is a neutralizing antibody that binds to a viral protein. In some embodiments, the antibody is a neutralizing antibody that binds to a receptor for a viral protein. In some embodiments, the antibody binds to a receptor that is required for viral entry into a cell (*e.g.*, an ACE2 receptor). In some embodiments, the antibody is a neutralizing antibody (nAb) that binds to the S protein of a coronavirus and prevents or reduces its ability to infect cells. In some embodiments, the coronavirus is SARS-CoV-2. In some embodiments, the nAb binds to a S protein comprising one or more mutations. In some embodiments, the nAb binds to a S protein or fragment thereof that comprises at least one point mutation in the S2 region, for example, a K986P, V987P, F817P, A892P, A899P or A942P mutation or combinations thereof. In some embodiments, the nAb binds to a S protein or fragment thereof that comprises at least one point mutation selected from A222V, E406W, K417N, K417T, N439K, L455N, E484K, Q493F, N501Y, A570D, D614G, P681H, A701V, T716I, S982A, or combinations thereof. In some embodiments the nAb is a monoclonal antibody (mAb), a functional antigen-binding fragment (Fab), a single-chain variable region fragment (scFv), or a single-domain antibody (a VHH or nanobody).

[0116] Exemplary nAbs for binding and neutralization of the S protein of SARS-CoV-2 have been described, for example, in Barnes, C.O. et al. SARS-CoV-2 neutralizing antibody structures

inform therapeutic strategies. Nature 588, 682–687 (2020), and Chinese Patent Application No. CN111690058A, the contents of which are herein incorporated by reference in their entirety.

[0117] In some embodiments, the coding RNA sequence encodes a targeting protein that is not an antibody. Examples of non-antibody-based targeting proteins include, but are not limited to, a lipocalin, an anticalin (artificial antibody mimetic proteins that are derived from human lipocalins), "T-body", a peptide (*e.g.*, a BICYCLE™ peptide), an affibody (antibody mimetics composed of alpha helices, *e.g.* an three-helix bundle), a peptibody (peptide-Fc fusion), a DARPin (**d**esignd **a**nkyrin **r**epeat **prote**i**ns**, engineered antibody mimetic proteins consisting repeat motifs), an affimer, an avimer, a knottin (a protein structural motif containing 3 disulfide bridges), a monobody, an affinity clamp, an ectodomain, a receptor ectodomain, a receptor, a cytokine, a ligand, an immunocytokine, and a centryin. *See*, for example, Vazquez-Lombardi, Rodrigo, et al. *Drug discovery today* 20.10 (2015): 1271-1283.

[0118] In some embodiments, the coding RNA sequence encodes a soluble receptor. Soluble receptors (sometimes referred to as soluble receptor decoys or "traps") can comprise all or a portion of the extracellular domain of a receptor protein. In some embodiments, a nucleotide sequence encoding all or a portion of the extracellular domain of a receptor protein is operably linked to a signal peptide for secretion from cells.

[0119] In some embodiments, the soluble receptor comprises an extracellular domain of a naturally occurring receptor. In some embodiments, the soluble receptor variant comprises an engineered variant of an extracellular domain of a naturally occurring receptor, such as a variant comprising one or more mutations in the extracellular domain. In some embodiments, the soluble receptor comprises one or more mutations that increase the affinity of the soluble receptor for its ligand compared to the affinity of the naturally occurring receptor for its ligand.

[0120] In some embodiments, the soluble receptor is a fusion protein comprising one or more additional protein domains operably linked to the extracellular domain of the receptor or a variant thereof. In some embodiments, the soluble receptor comprises an Fc domain of an immunoglobulin (Ig), *e.g.*, a human immunoglobulin. In some embodiments, the soluble receptor comprises an Fc domain of a human IgG1.

[0121] In some embodiments, the soluble receptor comprises the extracellular domain of a signaling receptor, and the soluble receptor can reduce or inhibit activity of the signaling pathway by blocking binding between the endogenous receptor and its ligand.

[0122] In some embodiments, the soluble receptor is a receptor that binds to a viral protein and/or that mediates viral entry. In some embodiments, soluble receptor is a soluble ACE2 receptor. In some embodiments, the therapeutic polypeptide is a soluble ACE2 receptor variant capable of binding to an S protein of a coronavirus. In some embodiments, the soluble ACE2 receptor variant binds to the receptor binding domain (RBD) of the S protein. In some embodiments, the ACE2 receptor variant is enzymatically active. In other embodiments, the ACE2 receptor variant is enzymatically inactive. In some embodiments, the soluble ACE2 receptor variant comprises the soluble extracellular domain of wild-type (WT) human recombinant ACE2 (APN01). In some embodiments, the soluble ACE2 receptor variant comprises one or more mutations in the extracellular domain of human ACE2. In some embodiments, the soluble ACE2 receptor variant is engineered via affinity maturation to have increased binding affinity to the RBD of the S protein. Soluble ACE2 receptor variants have been described, for example in Haschke M et al., Clin Pharmacokinet. 2013 Sep;52(9):783-92; Glasgow A et al., Proceedings of the National Academy of Sciences Nov 2020, 117 (45) 28046-28055; and Higuchi Y. et al., bioRxiv 2020. 09.16.299891, the contents of which are herein incorporated by reference in their entirety. In some embodiments, the soluble ACE2 receptor variant is a fusion protein, *e.g.*, a fusion of the extracellular ACE2 receptor domain to the Fc region of the human IgG1.

[0123] In some embodiments, the coding RNA sequence encodes a functional protein. In some embodiments, the coding RNA sequence is capable of being expressed by target cells (*e.g.*, human or mouse cells) for the production (and in certain instances, the secretion) of a functional enzyme or protein as disclosed, for example, in International Application No. PCT/US2010/058457 and WO2020237227, the contents of which are herein incorporated by reference in their entirety. In some embodiments, the therapeutic polypeptide can be engineered for secretion by operably linking a signal peptide to the amino terminus of the therapeutic polypeptide. For example, in some embodiments, upon the expression of one or more therapeutic polynucleotides by target cells, the production of a functional enzyme or protein in which a subject is deficient (*e.g.*, a urea cycle enzyme or an enzyme associated with a lysosomal storage disorder) may be observed.

[0124] In some embodiments, the coding RNA sequence encodes a protein such as IDUA, OTC, FAH, miniDMD, DMD, p53, PTEN, COL3A1, BMPR2, AHI1, FANCC, MYBPC3, ILRG2, or ARG1, wherein deficiency of the functional protein is associated with a disease or

disorder. In some embodiments, the coding RNA sequence a protein (*e.g.*, a lysosomal enzyme) wherein deficiency of the protein is associated with a lysosomal storage disorder.

[0125] In some embodiments, the coding RNA sequence encodes a protein (*e.g.*, an enzyme), wherein deficiency of the protein is associated with a metabolic disorder. In some embodiments, the therapeutic polypeptide comprises a urea cycle enzyme (*e.g.*, ARG1).

[0126] In some embodiments, the coding RNA sequence encodes a protein (*e.g.*, p53 or PTEN), wherein deficiency of the protein is associated with a cancer. In some embodiments, the therapeutic polypeptide comprises a tumor suppressor.

[0127] In some embodiments, the coding RNA sequence encodes a reporter protein, such as a fluorescent protein. Fluorescent proteins are well known to those skilled in the art, and include but are not limited to, green fluorescent proteins (GFPs), enhanced green fluorescent proteins (EGFPs), red fluorescent proteins (RFPs), and blue fluorescent proteins (BFPs).

[0128] In some embodiments, the coding RNA sequence encodes two or more polypeptides, such as two or more therapeutic polypeptides. In some embodiments, the coding RNA sequence encodes a therapeutic polypeptide and a reporter protein.

[0129] In some embodiments, the various domains or fragments in the polypeptide encoded by the coding RNA sequence may be fused to each other via a peptide linker. Flexible peptide linkers such as glycine linkers, glycine-serine linkers, and linkers containing other amino acids are known in the art (for example, suitable peptide linkers are described by Chen et al. in Fusion Protein Linkers: Property, Design and Functionality. Adv. Drug Deli Rev. 2013 October 15; 65(10): 1357–1369). Peptide linkers can also be designed by computation methods. The peptide linker can be of any length from 1 to 10, 10 to 20, 20 to 30, 30 to 40, 40 to 50, or greater than 50 amino acids.

[0130] In some embodiments, the coding RNA sequence is codon-optimized. A codon optimized sequence may be one in which codons in a polynucleotide encoding a polypeptide have been substituted in order to increase the expression, stability and/or activity of the polypeptide. Factors that influence codon optimization include, but are not limited to one or more of: (i) variation of codon biases between two or more organisms or genes or synthetically constructed bias tables, (ii) variation in the degree of codon bias within an organism, gene, or set of genes, (iii) systematic variation of codons including context, (iv) variation of codons according to their decoding tRNAs, (v) variation of codons according to GC %, either overall or in one position of the triplet, (vi) variation in degree of similarity to a reference sequence for example a naturally occurring

sequence, (vii) variation in the codon frequency cutoff, (viii) structural properties of mRNAs transcribed from the DNA sequence, (ix) prior knowledge about the function of the DNA sequences upon which design of the codon substitution set is to be based, and/or (x) systematic variation of codon sets for each amino acid. In some embodiments, a codon optimized polynucleotide may minimize ribozyme collisions and/or limit structural interference between the expression sequence and the IRES.

[0131] In some embodiments, the coding RNA sequence may encode or be operably linked to one or more additional elements that facilitate translation of the coding RNA sequence into a functional polypeptide. In some embodiments, the one or more additional elements are useful for monitoring translation of the coding RNA sequence.

[0132] In some embodiments, the coding RNA sequence encodes a polypeptide comprising a signal peptide (SP). In non-limiting examples, the signal peptide is the signal sequence and propeptide from human tissue plasminogen activator (tPA), the signal sequence from human IgE Immunoglobulin, or the signal peptide sequence of MHC I. In some embodiments, the signal peptide can facilitate secretion of the polypeptide encoded by the coding RNA sequence.

[0133] In some embodiments, the 3' end of the coding RNA sequence is operably linked to an in-frame 2A peptide coding sequence. In some embodiments, the coding RNA sequence does not comprise a stop codon at the 3' end. In some embodiments, the in-frame 2A peptide coding sequence replaces the stop codon. In some embodiments, the coding RNA sequence contains no stop codon and the number of nucleotides composing the coding RNA is a multiple of three. In some embodiments, the coding RNA sequence having no stop codon and the number of nucleotides composing the RNA being a multiple of three allow for rolling circle translation of the circRNA prepared using the linear RNA precursor. In some embodiments, the 2A peptide coding sequence allows for rolling circle translation of the circRNA prepared using the linear RNA precursor. In some embodiments, the 2A peptide allows cleavage of a polypeptide generated by rolling circle translation into monomeric polypeptide sequences. In non-limiting examples, the 2A peptide coding sequence encodes a P2A or T2A peptide, such as the sequence set forth in SEQ ID NO: 9 or 12.

[0134] In some embodiments, the coding RNA sequence comprises a nucleotide sequence encoding an affinity or identification tag. Exemplary tags include, but are not limited to, His tag, FLAG tag, SUMO tag, GST tag, and MBP tag.

[0135] In some embodiments, the 5' end of the coding RNA sequence is operably linked to a Kozak sequence. In some embodiments, the Kozak sequence functions as a protein translation initiation site. In some embodiments, the linear RNA comprises from the 5' end to the 3' end: a first portion of a RNA element (*e.g.*, IRES), a Kozak sequence, a coding RNA sequence, and a second portion of the RNA element (*e.g.*, IRES).

[0136] In some embodiments, the linear RNA further comprises a polyA or polyAC sequence disposed at the 3' end of the coding RNA sequence and at the 5' end of the second portion of the RNA element (*e.g.*, IRES). The internal polyA sequence or polyAC spacer may range from 1 to 500 nucleotides in length (*e.g.*, at least 20, 30, 40, 50, 60, 70, 80, 90, 120, 140, 160, 180, 200, 250, 300, 350, 400, 450, or 500 nucleotides). In some embodiments, the polyA sequence or polyAC sequence may range from 10-70, 20-60, or 30-60 nucleotides in length. In some embodiments, the linear RNA comprises no polyA sequence or polyAC sequence. Without being bound by any theory or hypothesis, an internal polyA sequence or a polyAC spacer added before IRES sequences in a circRNA can help to keep the functional second structure of IRES elements for efficient protein translation initiated by IRES. In some embodiments, the polyA sequence or polyAC spacer increases expression of the coding RNA.

II. Method of preparing circRNAs

[0137] The present application further provides methods of preparing circRNAs using any one of the linear RNAs or constructs described in Section II above. Intramolecular ligation of a linear RNA described herein by a RNA ligase provides a circRNA comprising a RNA element and an effector RNA (*e.g.*, coding RNA) sequence. In some embodiments, the circRNA is prepared by circularizing a linear RNA *in vitro*.

[0138] In some embodiments, the circRNA can be obtained by circularizing a linear RNA using a ligase such as a RNA ligase. In some embodiments, the linear RNA is circularized *in vitro*. In some embodiments, the linear RNA can be circularized by a T4 RNA ligase. In non-limiting examples, the linear RNA can be circularized by a ligase such as a T4 RNA ligase 1 (T4 Rnl1), and T4 RNA ligase 2 (T4 Rnl2). The linear RNAs described herein are circularized without the presence of a single stranded nucleic acid adaptor, *e.g.*, a splint oligonucleotide.

[0139] In some embodiments, there is provided a method of preparing a circRNA, comprising: (a) contacting any one of the linear RNAs described herein with a RNA ligase under conditions that allow ligation of the nick in the linear RNA to provide a circularized RNA product; and (b)

isolating the circularized RNA product, thereby providing the circRNA. In some embodiments, the ligation does not require presence of a splint oligonucleotide.

[0140] In some embodiments, the method described herein comprises circularizing a linear RNA *in vitro*, comprising: (a) contacting any one of the linear RNAs described herein with a RNA ligase under conditions that allow ligation of the nick in the linear RNA to provide a circularized RNA product; and (b) isolating the circularized RNA product, thereby providing the circRNA. In some embodiments, the ligation does not require presence of a splint oligonucleotide.

[0141] In some embodiments, there is provided a method of preparing a circRNA, comprising: (a) contacting any one of the linear RNAs described herein with a T4 RNA ligase 2 under conditions that allow ligation of the nick in the linear RNA to provide a circularized RNA product; and (b) isolating the circularized RNA product, thereby providing the circRNA. In some embodiments, the ligation does not require presence of a splint oligonucleotide.

[0142] A RNA ligase may be used to enzymatically link a 5'- phosphorylated end of a linear RNA described herein to the 3'-hydroxyl group of the linear RNA forming a new phosphodiester linkage. In an example reaction, a linear circular RNA is incubated at 25° C for 8 hour with 1-10 units of T4 RNA ligase (New England Biolabs, Ipswich, Mass.) according to the manufacturer' s protocol. The ligation reaction occurs without the need of a linear nucleic acid capable of basepairing with both the 5'- and 3'- region in juxtaposition that assists the enzymatic ligation reaction, *i.e.*, the ligation is not a splint ligation.

[0143] Ligation by T4 RNA ligase 1 may lead to insertion or deletion of nucleotides at the ligation site. In some embodiments, T4 RNA ligase 2 is used to ligate the linear RNA described herein so that no additional sequences are introduced to the circRNA.

[0144] The methods described herein has high intramolecular ligation efficiency. Ligation efficiency may be assessed using known methods in the art, such as by HPLC or agarose gel electrophoresis. In some embodiments, the ligation efficiency is at least about any one of 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or higher. In some embodiments, the method results in less than 5%, 4%, 3%, 2%, 1% or lower amount of polymeric circRNA products, *i.e.*, ligation of multiple copies of linear RNAs into a circular RNA.

[0145] In some embodiments, the method further comprises *in vitro* transcription of a nucleic acid construct encoding the linear RNA to obtain the linear RNA.

[0146] In some embodiments, the method further comprises treating the circularized RNA product with RNase R to digest the linear RNA transcripts.

[0147] In some embodiments, the method further comprises purifying the circRNA from the circRNA product. In non-limiting examples, the circRNA is purified by gel-purification or by high-performance liquid chromatography (HPLC). In some embodiments, agarose gel electrophoresis allows for simple and effective separation of circular RNA products from linear RNA precursor molecules and nicked circles. In some embodiments, the method comprises purifying the circular RNA by chromatography, such as HPLC. In some embodiments, the purified circular RNA can be stored at -80°C. In some embodiments, the step of isolating the circRNA comprises gel-purifying the circRNA. In some embodiments, the purified circRNA can be stored at -80°C.

III. Circular RNAs and methods of use

[0148] The present application further provides circRNAs and compositions prepared using any one of the methods of preparation, or any one of the linear RNAs or constructs described herein.

[0149] In some embodiments, the circRNA comprises an RNA element and an effector RNA sequence. In some embodiments, the effector RNA sequence is a sequence of a RNA molecule selected from the group consisting of a gRNA, a dRNA, a siRNA, a miRNA, a shRNA, a lincRNA and a coding RNA.

[0150] In some embodiments, the circRNA comprises an RNA element (*e.g.*, an IRES) and a coding RNA sequence encoding a therapeutic polypeptides, such as an antigenic polypeptide, a functional protein, a receptor protein, or a targeting protein.

[0151] In some embodiments, the circRNA is a circRNA vaccine comprising an RNA element (*e.g.*, an IRES) and a coding RNA sequence encoding an antigenic polypeptide. In some embodiments, the circRNA vaccine is a coronavirus vaccine, or a cancer vaccine.

[0152] In some embodiments, there is provided a cocktail composition comprising a plurality of circRNAs each comprising a coding RNA sequence encoding an antigenic polypeptide, a receptor protein of an infectious agent, or a targeting protein (*e.g.*, an antibody such as a neutralizing antibody). In some embodiments, the plurality of circRNA encode antigenic polypeptides that are different with respect to each other, such as different mutants of an antigenic polypeptide (*e.g.*, S protein or fragment thereof). In some embodiments, the plurality of

circRNA encode receptor proteins that are different with respect to each other, such as different mutants of a receptor protein (*e.g.*, ACE2). In some embodiments, the plurality of circRNA encode targeting proteins that are different with respect to each other, such as different antibodies (*e.g.*, neutralizing antibodies).

[0153] The circRNAs described herein may be used to treat or prevent a disease or condition in an individual, including, but not limited to genetic diseases (*e.g.*, hereditary genetic diseases, metabolic diseases and cancer), and infections (*e.g.*, viral infections such as coronavirus infections). In some embodiments, the circRNA is subject to rolling circle translation by a ribosome in the individual.

[0154] In some embodiments, there is provided a method of treating or preventing a disease or condition in an individual, comprising administering to the individual an effective amount of a circRNA comprising a coding RNA sequence encoding a functional protein. In some embodiments, the functional protein is an enzyme, a receptor, a ligand, a signaling molecule, or a transcription factor. In some embodiments, the disease or condition is a metabolic disease. In some embodiments, the disease or condition is a lysosomal storage disorder. In some embodiments, the disease or condition is a cancer.

[0155] The circRNAs described herein may be used for treating a genetic disease or condition that is associated with a mutation or deficiency in a naturally-occurring protein corresponding to the therapeutic polypeptide encoded by the circRNA. In some embodiments, the disease or condition is a disease or condition associated with insufficient levels and/or activity of a naturally-occurring protein corresponding to the therapeutic polypeptide. In some embodiments, the disease or condition is a hereditary genetic disease associated with one or more mutations in naturally-occurring protein corresponding to the therapeutic polypeptide. In some embodiments, the therapeutic polypeptide is a wildtype protein, or a functional variant thereof (*e.g.*, a functional fragment, fusion protein, or mutant).

[0156] In some embodiments, the therapeutic polypeptide can be any polypeptide that is capable of being expressed by target cells (*e.g.*, human or mouse cells) for the production (and in certain instances, the excretion) of a functional enzyme or protein as disclosed, for example, in International Application No. PCT/US2010/058457. In some embodiments, the therapeutic polypeptide can be engineered for secretion by operably linking a signal peptide to the amino terminus of the therapeutic polypeptide. For example, in some embodiments, upon the

expression of one or more therapeutic polynucleotides by target cells, the production of a functional enzyme or protein in which a subject is deficient (*e.g.*, a urea cycle enzyme or an enzyme associated with a lysosomal storage disorder) may be observed.

[0157] Examples of disease-associated mutations that may be treated by the methods of the present application include, but are not limited to, $TP53^{W53X}$ (*e.g.*, 158G>A) associated with cancer, $IDUA^{W402X}$ (*e.g.*, TGG>TAG mutation in exon 9) associated with Mucopolysaccharidosis type I (MPS I), $COL3A1^{W1278X}$ (*e.g.*, 3833G>A mutation) associated with Ehlers-Danlos syndrome, $BMPR2^{W298X}$ (*e.g.*, 893G>A) associated with primary pulmonary hypertension, $AHI1^{W725X}$ (*e.g.*, 2174G>A) associated with Joubert syndrome, $FANCC^{W506X}$ (*e.g.*, 1517G>A) associated with Fanconi anemia, $MYBPC3^{W1098X}$ (*e.g.*, 3293G>A) associated with primary familial hypertrophic cardiomyopathy, and $IL2RG^{W237X}$ (*e.g.*, 710G>A) associated with X-linked severe combined immunodeficiency. In some embodiments, the disease or condition is a cancer. In some embodiments, the disease or condition is a monogenetic disease. In some embodiments, the disease or condition is a polygenetic disease.

[0158] In some embodiments, the circRNA has a functional half-life of at least or at least about 20 hours, 24 hours, 30 hours, or 36 hours. In some embodiments, the circRNA has a duration of therapeutic effect in a human cell of at least or at least about 20 hours, 24 hours, 30 hours, or 36 hours. In some embodiments, the circRNA has a duration of therapeutic effect in a human cell greater than or equal to that of an equivalent linear RNA comprising the same expression sequence. In some embodiments, the circRNA has a functional half-life in a human cell greater than or equal to that of an equivalent linear RNA comprising the same expression sequence.

[0159] In some embodiments, there is provided a method of treating or preventing a disease or condition in an individual, comprising administering to the individual an effective amount of a circRNA vaccine comprising a coding RNA sequence encoding an antigenic polypeptide, or a cocktail composition comprising a plurality of circRNAs. In some embodiments, the antigenic polypeptide is a protein or a fragment thereof of an infectious agent, such as a virus, *e.g.*, a coronavirus. In some embodiments, the infectious agent is SARS-CoV-2.

[0160] In some embodiments, there is provided a method of treating or preventing a disease or condition in an individual, comprising administering to the individual an effective amount of a circRNA comprising a coding RNA sequence encoding a receptor protein. In some embodiments, the receptor protein is a receptor of an infectious agent, such as a virus, *e.g.*, a

coronavirus. In some embodiments, the receptor protein is a soluble receptor, such as a soluble ACE2 receptor.

[0161] In some embodiments, there is provided a method of treating or preventing a disease or condition in an individual, comprising administering to the individual an effective amount of a circRNA comprising a coding RNA sequence encoding a targeting protein, such as an antibody. In some embodiments, the targeting protein is a neutralizing antibody. In some embodiments, the targeting protein is a therapeutic antibody. In some embodiments, the targeting protein specifically binds an infectious agent, such as a virus, *e.g.*, a coronavirus.

[0162] In some embodiments, the present application provides circRNAs for use in treating or preventing a disease or condition in an individual.

[0163] In some embodiments, the present application provides use of a circRNA comprising a nucleic acid sequence encoding a therapeutic polypeptide for the manufacture of a medicament for treating or preventing a disease or condition in an individual.

[0164] In some embodiments, the circRNA is administered as naked circRNA, or as a pharmaceutical composition comprising a transfection agent. In non-limiting examples, the transfection agent is polyethylenimine (PEI) or a lipid nanoparticle (LNP). Other examples of lipidosomes that can be used to administer the circRNA composition for administration (*e.g.*, circRNA vaccine or pharmaceutical composition) include protamines, cationic nanoemulsions, modified dendrimer nanoparticles, protamine liposomes, cationic polymers, cationic polymer liposomes, polysaccharide particles, cationic lipid nanoparticles, cationic lipid-cholesterol nanoparticles, cationic lipid-cholesterol PEG nanoparticle, cationic lipid transfection reagents sold under the trademark LIPOFECTAMINE, nonliposomal transfection reagents sold under the trademark FUGENE, or any combination thereof can be used as the transfection agent.

[0165] In some embodiments, the liposome formulation may be influenced by, but not limited to, the selection of the cationic lipid component, the degree of cationic lipid saturation, the nature of the PEGylation, ratio of all components and biophysical parameters such as size. In some embodiments, the liposome formulation comprises a cationic lipid, a cholesterol and a PEGylated lipid. For example, a liposome formulation may comprise a cationic lipid, dipalmitoylphosphatidylcholine, cholesterol, and PEG-c-DMA. See, for example, Semple et al. Nature Biotech. 2010 28:172-176, herein incorporated by reference in its entirety. In some embodiments, liposome formulations may comprise from about 35 to about 45% cationic lipid,

from about 40% to about 50% cationic lipid, from about 50% to about 60% cationic lipid and/or from about 55% to about 65% cationic lipid. In some embodiments, the ratio of lipid to RNA in liposomes may be from about 5:1 to about 20: 1, from about 10:1 to about 25:1, from about 15:1 to about 30:1 and/or at least 30:1. Suitable liposome formulations have been described, for example, in WO2020237227, the contents of which are herein incorporated by reference in their entirety.

[0166] In some embodiments, the circRNA is not formulated with a transfection reagent. In some embodiments, the circRNA is delivered as naked RNA. In some embodiments, the circRNA is delivered by gene gun or by electroporation.

[0167] The circRNA composition for administration (*e.g.*, circRNA vaccine or pharmaceutical composition) can be administered to a subject by systemic injection into the vasculature, systemic injection into the lymph nodes, subcutaneous injection or depots, or by local injection.

[0168] In some embodiments, the circRNA may be formulated in a lipid nanoparticle such as those described in International Publication No. WO2012170930, herein incorporated by reference in its entirety.

[0169] In some embodiments, the synthetic nanocarriers may be formulated for controlled and/or sustained release of the circRNA described herein. As a non-limiting example, the synthetic nanocarriers for sustained release may be formulated by methods known in the art, described herein and/or as described in International Pub No. WO2010138192 and US Pub No. 20100303850, each of which is herein incorporated by reference in their entirety.

[0170] In some embodiments, the circRNA may be formulated for controlled and/or sustained release wherein the formulation comprises at least one polymer that is a crystalline side chain (CYSC) polymer. CYSC polymers are described in U.S. Patent No. 8,399,007, herein incorporated by reference in its entirety.

[0171] In some embodiments, the synthetic nanocarrier may be formulated for use as a vaccine. In some embodiments, the synthetic nanocarrier may encapsulate at least one circRNA, which encode at least one antigen. As a nonlimiting example, the synthetic nanocarrier may include at least one antigen and an excipient for a vaccine dosage form (see International Pub No. WO2011150264 and US Pub No. US201 10293723, each of which is herein incorporated by reference in their entirety). As another non-limiting example, a vaccine dosage form may include at least two synthetic nanocarriers with the same or different antigens and an excipient (see International Pub No. WO201

1150249 and US Pub No. US201 10293701, each of which is herein incorporated by reference in their entirety). The vaccine dosage form may be selected by methods described herein, known in the art and/or described in International Pub No. WO2011150258 and US Pub No. US20120027806, each of which is herein incorporated by reference in their entirety).

[0172] In some embodiments, the synthetic nanocarrier may comprise at least one circRNA, which encodes at least one adjuvant. As non-limiting example, the adjuvant may comprise dimethyldioctadecylammonium-bromide, dimethyldioctadecylammoniumchloride, dimethyldioctadecylammonium-phosphate or dimethyldioctadecylammoniumacetate (DDA) and an apolar fraction or part of said apolar fraction of a total lipid extract of a mycobacterium (See *e.g.*, U.S. Pat. No. 8,241,610; herein incorporated by reference in its entirety). In another embodiment, the synthetic nanocarrier may comprise at least one circRNA and an adjuvant. As a non-limiting example, the synthetic nanocarrier comprising and adjuvant may be formulated by the methods described in International Pub No. WO2011150240 and US Pub No. US20110293700, each of which is herein incorporated by reference in its entirety.

[0173] In some embodiments, the circRNA functions as an adjuvant. As an example, RNA-sensing in the cytoplasm can trigger innate immunity, and innate immune signaling is known to contribute to adaptive immunity by diverse routes. Thus, the circRNA encoding the antigenic polypeptide or a second circRNA (e.g., a circRNA that does not encode a polypeptide) can be used as an adjuvant for boosting the adaptive immune response to the antigenic polypeptide.

[0174] In some embodiments, the circRNA compositions of the present application may be administrated with other prophylactic or therapeutic compounds. As a non-limiting example, the prophylactic or therapeutic compound may be an adjuvant or a booster. As used herein, when referring to a prophylactic composition, such as a vaccine, the term "booster" refers to an extra administration of the prophylactic composition. A booster (or booster vaccine) may be given after an earlier administration of the prophylactic composition. The time of administration between the initial administration of the prophylactic composition and the booster may be, but is not limited to, 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 6 minutes, 7 minutes, 8 minutes, 9 minutes, 10 minutes, 15 minutes, 20 minutes 35 minutes, 40 minutes, 45 minutes, 50 minutes, 55 minutes, 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 13 hours, 14 hours, 15 hours, 16 hours, 17 hours, 18 hours, 19 hours, 20 hours, 21 hours, 22 hours, 23 hours, 1 day, 36 hours, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 10 days, 2 weeks, 3 weeks, 1

month, 2 months, 3 months, 4 months, 5 months, 6 months, 7 months, 8 months, 9 months, 10 months, 11 months, 1 year, 18 months, 2 years, 3 years, 4 years, 5 years, 6 years, 7 years, 8 years, 9 years, 10 years, 11 years, 12 years, 13 years, 14 years, 15 years, 16 years, 17 years, 18 years, 19 years, 20 years, 25 years, 30 years, 35 years, 40 years, 45 years, 50 years, 55 years, 60 years, 65 years, 70 years, 75 years, 80 years, 85 years, 90 years, 95 years or more than 99 years.

[0175] In some embodiments, the circRNA composition for administration (*e.g.*, circRNA vaccine or pharmaceutical composition) may be administered intranasally. For example, circRNA vaccines may be administered intranasally similar to the administration of live vaccines. In some embodiments, the circRNA may be administered intramuscularly or intradermally similarly to the administration of inactivated vaccines known in the art.

[0176] In some embodiments, the circRNA vaccine comprises an adjuvant, which may enable the vaccine to elicit a higher immune response. As a non-limiting example, the adjuvant could be a sub-micron oil-in-water emulsion, which can elicit a higher immune response in human pediatric populations (see *e.g.*, the adjuvant-containing vaccines described in US Patent Publication No. US20120027813 and US Patent No. US8506966, the contents of each of which are herein incorporated by reference in its entirety).

V. Compositions, kits and articles of manufacture

[0177] Further provided by the present application are compositions comprising any one of the linear RNAs, constructs, or circRNAs described herein. In some embodiments, there is provided a pharmaceutical composition comprising any one or plurality of circRNAs described herein, and a pharmaceutically acceptable carrier. Pharmaceutical compositions can be prepared by mixing the therapeutic agents described herein having the desired degree of purity with optional pharmaceutically acceptable carriers, excipients or stabilizers (Remington's Pharmaceutical Sciences 16th edition, Osol, A. Ed. (1980)), in the form of lyophilized formulations or aqueous solutions. Acceptable carriers, excipients, or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers, antioxidants including ascorbic acid, methionine, Vitamin E, sodium metabisulfite; preservatives, isotonicifiers (*e.g.* sodium chloride), stabilizers, metal complexes (*e.g.* Zn-protein complexes); chelating agents such as EDTA and/or non-ionic surfactants.

[0178] In some embodiments, the pharmaceutical composition is contained in a single-use vial, such as a single-use sealed vial. In some embodiments, the pharmaceutical composition is

contained in a multi-use vial. In some embodiments, the pharmaceutical composition is contained in bulk in a container. In some embodiments, the pharmaceutical composition is cryopreserved.

[0179] The present application further provides kits and articles of manufacture for use in any one of the methods for preparing circRNAs described herein. In some embodiments, the kit comprises a linear RNA or a construct encoding the linear RNA described herein, and instructions for preparing a circRNA. In some embodiments, the kit further comprises a RNA ligase, such as T4 RNA ligase 1, or T4 RNA ligase 2. In some embodiments, the kit further comprises a reverse transcriptase.

[0180] In some embodiments, there is provided a kit comprising any one of the circRNAs described herein and instructions for treating or preventing a disease or condition (*e.g.*, coronavirus infection or cancer). In some embodiments, the kit comprises instructions for administering the circRNA.

[0181] The kits of the invention are in suitable packaging. Suitable packaging includes, but is not limited to, vials, bottles, jars, flexible packaging (*e.g.*, sealed Mylar or plastic bags), and the like. Kits may optionally provide additional components such as buffers and interpretative information. The present application thus also provides articles of manufacture, which include vials (such as sealed vials), bottles, jars, flexible packaging, and the like.

[0182] The instructions relating to the use of the compositions generally include information as to dosage, dosing schedule, and route of administration for the intended treatment. The containers may be unit doses, bulk packages (*e.g.*, multi-dose packages) or sub-unit doses. For example, kits may be provided that contain sufficient dosages of the circRNA as disclosed herein to provide effective treatment of an individual or of many individuals. Additionally, kits may be provided that contain sufficient dosages of the circRNA to allow for multiple administrations to an individual (*e.g.*, initial vaccine administration and subsequent booster administration, in the case of a circRNA vaccine). Kits may also include multiple unit doses of the pharmaceutical compositions and instructions for use and packaged in quantities sufficient for storage and use in pharmacies, for example, hospital pharmacies and compounding pharmacies.

[0183] In some embodiments, the kit comprises a delivery system. The delivery system may be a unit dose delivery system. The volume of solution or suspension delivered per dose can be anywhere from about 5 to about 2000 microliters, from about 10 to about 1000 microliters, or from about 50 to about 500 microliters. Delivery systems for these various dosage forms can be syringes,

dropper bottles, plastic squeeze units, atomizers, nebulizers or pharmaceutical aerosols in either unit dose or multiple dose packages. In some embodiments, there is provided a delivery system of any one of the circRNAs described herein, comprising the circRNA and a device for delivering the circRNA.

[0184] All of the features disclosed in this specification may be combined in any combination. Each feature disclosed in this specification may be replaced by an alternative feature serving the same, equivalent, or similar purpose. Thus, unless expressly stated otherwise, each feature disclosed is only an example of a generic series of equivalent or similar features.

EXAMPLES

[0185] The invention will be more fully understood by reference to the following examples. They should not, however, be construed as limiting the scope of the invention. It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims.

Example 1. *In Vitro* circRNA production by ligation

[0186] Because circRNAs have no cap or poly(A) tails, Internal Ribosomal Entry Site (IRES) is required to start the translation process of circRNAs. Here we use rationally designed split IRES elements (also referred herein as "IRES splits" or "split IRES portions") to efficiently and accurately produce coding circular RNAs using a T4 RNA ligase. FIG. 1 shows an overview of the method for producing circRNAs using a construct comprising IRES splits. First, a RNA transcript (also referred to as "precursor" or "linear RNA precursor") is obtained by *in vitro* transcription of the construct, which contains the two IRES splits that associate with each other to form a stable intramolecular secondary structure having a nick across the two ends of the RNA transcript. The nick is subsequently ligated by a T4 RNA ligase to yield the circular RNA.

[0187] We analyzed the secondary structure of the wildtype CVB3 IRES (SEQ ID NO: 1) using RNAfold. We found that the wildtype CVB3 IRES contains highly stable double-stranded RNA structures, which may promote formation of intramolecular double strands. *See*, boxed region of FIG. 2A. We thus picked two sites in the double-stranded RNA regions to engineer pairs of split CVB3 IRES portions. In a first pair of CVB3 IRES splits (also referred herein as "Site 1 CVB3

IRES splits"), the 5' portion of the CVB3 IRES split contains nucleotides 1 to 385 of the wildtype CVB3 IRES (SEQ ID NO: 2), and the 3' portion of the CVB3 IRES split contains nucleotides 386 to 749 of the wildtype CVB3 IRES (SEQ ID NO: 3). *See*, FIG. 2B. In a second pair of CVB3 IRES splits (also referred herein as "Site 2 CVB3 IRES splits"), the 5' portion of the CVB3 IRES split contains nucleotides 1 to 346 of the wildtype CVB3 IRES (SEQ ID NO: 4), and the 3' portion of the CVB3 IRES split contains nucleotides 347 to 749 of the wildtype CVB3 IRES (SEQ ID NO: 5). *See*, FIG. 2C. To maintain the stability of the double-stranded structures, we inserted four nucleotides UUUG and its reverse complementary sequence (CTTT) immediately next to the split sites (boxed sequences in FIGS. 2B-2C).

[0188] Two constructs for producing circular RNA encoding EGFP are designed using the first pair and second pair of CVB3 IRES splits. Each construct comprises, from the 5' to the 3': a T7 promoter (SEQ ID NO: 6), the 3' portion of a CVB3 IRES split pair, EGFP-encoding sequence (SEQ ID NO: 7), a FLAG-tag encoding sequence (SEQ ID NO: 8), a P2A sequence (SEQ ID NO: 9), and the 5' portion of the CVB3 IRES split pair. The construct based on Site 1 CVB3 IRES splits (also referred herein as "Site 1 construct") has the DNA sequence of SEQ ID NO: 10. The construct based on Site 2 CVB3 IRES splits (also referred herein as "Site 2 construct") has the DNA sequence of SEQ ID NO: 11.

[0189] The two constructs were each subject to *in vitro* transcription to provide linear RNA precursors, which were ligated by T4 RNA ligase 1 or T4 RNA ligase 2 for 8 hours at 25°C. The circular RNA product from the Site 1 construct is referred herein as the Site 1 ligation circular RNA. The circular RNA product from the Site 1 construct is referred herein as the Site 2 ligation circular RNA. The ligation products were treated with RNase R to remove the linear RNA precursor. FIG. 3 shows HPLC chromatograms of a T4 RNA ligase 1-ligated Site 1 ligation circular RNA before and after RNAse R treatment. The cyclization efficiency was about 70%, and RNase R treatment degraded a lot of linear RNA products.

[0190] Next, we transfected each of the circular RNA products into HEK293T cells at 4μg circRNA per well in a twelve-well plate. As shown in FIG. 4A, EGFP expression was detected in cells transfected with Site 1 ligation circular RNA, but not in cells transfected with Site 2 ligation circular RNA. We reasoned that the absence of EGFP expression may be due to disruption of functional structures of the ligated IRES in Site 2 ligation circular RNA. Western blot analysis

showed that the EGFP expression level in cells transfected with Site 1 ligation circular RNA was much higher than the expression level of β-tubulin in the same cells (FIG. 4B).

[0191] Finally, we tested whether the ligation site was accurate, *i.e.*, free from insertion or deletion at the ligation site. We reverse transcribed the Site 1 ligation circular RNA, and PCR amplified the junction of the ligation site using forward and reverse primers as shown in FIG. 1. As shown in FIG. 5A, the PCR results confirmed that circular RNAs formed after ligation of the precursor RNA using T4 RNA ligase, while no circular RNAs were found in the precursor RNA that was not subject to T4 RNA ligase treatment. Using Sanger sequencing, we determined the sequences of the circular RNAs surrounding the ligated junction site in samples ligated by T4 RNA ligase 1 or T4 RNA ligase 2. As shown in FIG. 5B, T4 RNA ligase 2 accurately ligated the nick in the linear RNA precursor, but T4 RNA ligase 1 treatment resulted in a small number of insertions and deletions. Without being bound by theory, this may be due to the substrate requirement of T4 RNA ligase 2, which is double-stranded RNA.

SEQUENCE LISTING

SEQ ID NO: 1 Wildtype CVB3-IRES RNA sequence
UUAAAACAGCCUGUGGGUUGAUCCCACCCACAGGCCCAUUGGGCGCUAGCACUCU
GGUAUCACGGUACCUUUGUGCGCCUGUUUUAUACCCCCUCCCCCCAACUGUAACUU
AGAAGUAACACACACCGAUCAACAGUCAGCGUGGCACACCAGCCACGUUUUGAUC
AAGCACUUCUGUUACCCCGGACUGAGUAUCAAUAGACUGCUCACGCGGUUGAAGG
AGAAAGCGUUCGUUAUCCGGCCAACUACUUCGAAAAACCUAGUAACACCGUGGAA
GUUGCAGAGUGUUUCGCUCAGCACUACCCCAGUGUAGAUCAGGUCGAUGAGUCAC
CGCAUUCCCCACGGGCGACCGUGGCGGUGGCUGCGUUGGCGGCCUGCCCAUGGGG
AAACCCAUGGGACGCUCUAAUACAGACAUGGUGCGAAGAGUCUAUUGAGCUAGU
UGGUAGUCCUCCGGCCCCUGAAUGCGGCUAAUCCUAACUGCGGAGCACACACCCU
CAAGCCAGAGGGCAGUGUGUCGUAACGGGCAACUCUGCAGCGGAACCGACUACUU
UGGGUGUCCGUGUUUCAUUUUAUUCCUAUACUGGCUGCUUAUGGUGACAAUUGA
GAGAUCGUUACCAUAUAGCUAUUGGAUUGGCCAUCCGGUGACUAAUAGAGCUAU
UAUAUAUCCCUUUGUUGGGUUUAUACCACUUAGCUUGAAAGAGGUUAAAACAUU
ACAAUUCAUUGUUAAGUUGAAUACAGCAAA

SEQ ID NO: 2 5' portion of Site 1 CVB3-IRES split (nucleotides 1-385 of wildtype CVB3-IRES, with inserted nucleotide sequence underlined)
UUAAAACAGCCUGUGGGUUGAUCCCACCCACAGGCCCAUUGGGCGCUAGCACUCU
GGUAUCACGGUACCUUUGUGCGCCUGUUUUAUACCCCCUCCCCCCAACUGUAACUU
AGAAGUAACACACACCGAUCAACAGUCAGCGUGGCACACCAGCCACGUUUUGAUC
AAGCACUUCUGUUACCCCGGACUGAGUAUCAAUAGACUGCUCACGCGGUUGAAGG
AGAAAGCGUUCGUUAUCCGGCCAACUACUUCGAAAAACCUAGUAACACCGUGGAA
GUUGCAGAGUGUUUCGCUCAGCACUACCCCAGUGUAGAUCAGGUCGAUGAGUCAC
CGCAUUCCCCACGGGCGACCGUGGCGGUGGCUGCGUUGGCGGCCUGCCCAUGUUU

SEQ ID NO: 3 3' portion of Site 1 CVB3-IRES split (nucleotides 386-749 of wildtype CVB3-IRES, with inserted nucleotide sequence underlined)
GGGGAAACCCAAACAUGGGACGCUCUAAUACAGACAUGGUGCGAAGAGUCUAUU
GAGCUAGUUGGUAGUCCUCCGGCCCCUGAAUGCGGCUAAUCCUAACUGCGGAGCA
CACACCCUCAAGCCAGAGGGCAGUGUGUCGUAACGGGCAACUCUGCAGCGGAACC
GACUACUUUGGGUGUCCGUGUUUCAUUUUAUUCCUAUACUGGCUGCUUAUGGUG
ACAAUUGAGAGAUCGUUACCAUAUAGCUAUUGGAUUGGCCAUCCGGUGACUAAU
AGAGCUAUUAUAUAUCCCUUUGUUGGGUUUAUACCACUUAGCUUGAAAGAGGUU
AAAACAUUACAAUUCAUUGUUAAGUUGAAUACAGCAAA

SEQ ID NO: 4 5' portion of Site 2 CVB3-IRES split (nucleotides 1-346 of wildtype CVB3-IRES, with inserted nucleotide sequence underlined)
UUAAAACAGCCUGUGGGUUGAUCCCACCCACAGGCCCAUUGGGCGCUAGCACUCU
GGUAUCACGGUACCUUUGUGCGCCUGUUUUAUACCCCCUCCCCCCAACUGUAACUU
AGAAGUAACACACACCGAUCAACAGUCAGCGUGGCACACCAGCCACGUUUUGAUC
AAGCACUUCUGUUACCCCGGACUGAGUAUCAAUAGACUGCUCACGCGGUUGAAGG
AGAAAGCGUUCGUUAUCCGGCCAACUACUUCGAAAAACCUAGUAACACCGUGGAA
GUUGCAGAGUGUUUCGCUCAGCACUACCCCAGUGUAGAUCAGGUCGAUGAGUCAC
CGCAUUCCCCACGUUU

SEQ ID NO: 5 3' portion of Site 2 CVB3-IRES split (nucleotides 347-749 of wildtype CVB3-IRES, with inserted nucleotide sequence underlined)
GGGCGACCAAACGUGGCGGUGGCUGCGUUGGCGGCCUGCCCAUGGGGAAACCCAU
GGGACGCUCUAAUACAGACAUGGUGCGAAGAGUCUAUUGAGCUAGUUGGUAGUC
CUCCGGCCCCUGAAUGCGGCUAAUCCUAACUGCGGAGCACACACCCUCAAGCCAG
AGGGCAGUGUGUCGUAACGGGCAACUCUGCAGCGGAACCGACUACUUUGGGUGU
CCGUGUUUCAUUUUAUUCCUAUACUGGCUGCUUAUGGUGACAAUUGAGAGAUCG
UUACCAUAUAGCUAUUGGAUUGGCCAUCCGGUGACUAAUAGAGCUAUUAUAUAU
CCCUUUGUUGGGUUUAUACCACUUAGCUUGAAAGAGGUUAAAACAUUACAAUUC
AUUGUUAAGUUGAAUACAGCAAA

SEQ ID NO: 6 T7 promoter sequence
TAATACGACTCACTATAGG

SEQ ID NO: 7 EGFP-coding DNA sequence
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCT
GGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTG
CCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTAC
CCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTC
CAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT
GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCA
AGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCC
GCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACC
CCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCC
GCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGT
GACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG

SEQ ID NO: 8 FLAG tag-encoding DNA sequence
GATTACAAGGATGACGACGATAAG

SEQ ID NO: 9 P2A sequence
GCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACC
T

SEQ ID NO: 10 Site 1 Construct (Site 1 CVB3-IRES splits flanking EGFP-coding RNA)
TAATACGACTCACTATAGGGGAAACCCAAACATGGGACGCTCTAATACAGACATGG
TGCGAAGAGTCTATTGAGCTAGTTGGTAGTCCTCCGGCCCCTGAATGCGGCTAATCC
TAACTGCGGAGCACACACCCTCAAGCCAGAGGGCAGTGTGTCGTAACGGGCAACTC
TGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCATTTTATTCCTATACTGGCTGC
TTATGGTGACAATTGAGAGATCGTTACCATATAGCTATTGGATTGGCCATCCGGTGA
CTAATAGAGCTATTATATATCCCTTTGTTGGGTTTATACCACTTAGCTTGAAAGAGGT
TAAAACATTACAATTCATTGTTAAGTTGAATACAGCAAAACTAGTGCCACCATGGTG
AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGG
CGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT

ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTT
CGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTAT
ATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGA
GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCC
GCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGATTACAAGGATGACGACGA
TAAGGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTG
GACCTGGGTTTTTAAAACAGCCTGTGGGTTGATCCCACCCACAGGCCCATTGGGCGC
TAGCACTCTGGTATCACGGTACCTTTGTGCGCCTGTTTTATACCCCCTCCCCCAACTG
TAACTTAGAAGTAACACACACCGATCAACAGTCAGCGTGGCACACCAGCCACGTTT
TGATCAAGCACTTCTGTTACCCCGGACTGAGTATCAATAGACTGCTCACGCGGTTGA
AGGAGAAAGCGTTCGTTATCCGGCCAACTACTTCGAAAAACCTAGTAACACCGTGG
AAGTTGCAGAGTGTTTCGCTCAGCACTACCCCAGTGTAGATCAGGTCGATGAGTCAC
CGCATTCCCCACGGGCGACCGTGGCGGTGGCTGCGTTGGCGGCCTGCCCATGTTT

SEQ ID NO: 11 Site 1 Construct (Site 2 CVB3-IRES splits flanking EGFP-coding RNA)
TAATACGACTCACTATAGGGCGACCAAACGTGGCGGTGGCTGCGTTGGCGGCCTGC
CCATGGGGAAACCCATGGGACGCTCTAATACAGACATGGTGCGAAGAGTCTATTGA
GCTAGTTGGTAGTCCTCCGGCCCCTGAATGCGGCTAATCCTAACTGCGGAGCACACA
CCCTCAAGCCAGAGGGCAGTGTGTCGTAACGGGCAACTCTGCAGCGGAACCGACTA
CTTTGGGTGTCCGTGTTTCATTTTATTCCTATACTGGCTGCTTATGGTGACAATTGAG
AGATCGTTACCATATAGCTATTGGATTGGCCATCCGGTGACTAATAGAGCTATTATA
TATCCCTTTGTTGGGTTTATACCACTTAGCTTGAAAGAGGTTAAAACATTACAATTCA
TTGTTAAGTTGAATACAGCAAAACTAGTGCCACCATGGTGAGCAAGGGCGAGGAGC
TGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCAC
AAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCT
GAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCAC
CCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACG
ACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCA
AGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGG
GCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGC
AGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGC
GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCT
GCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACG
AGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCG
GCATGGACGAGCTGTACAAGGATTACAAGGATGACGACGATAAGGCTACTAACTTC
AGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTGGGTTTTTAAA
ACAGCCTGTGGGTTGATCCCACCCACAGGCCCATTGGGCGCTAGCACTCTGGTATCA
CGGTACCTTTGTGCGCCTGTTTTATACCCCCTCCCCCAACTGTAACTTAGAAGTAACA
CACACCGATCAACAGTCAGCGTGGCACACCAGCCACGTTTTGATCAAGCACTTCTGT

TACCCCGGACTGAGTATCAATAGACTGCTCACGCGGTTGAAGGAGAAAGCGTTCGTT
ATCCGGCCAACTACTTCGAAAAACCTAGTAACACCGTGGAAGTTGCAGAGTGTTTCG
CTCAGCACTACCCCAGTGTAGATCAGGTCGATGAGTCACCGCATTCCCCACGTTT

SEQ ID NO: 12 T2A peptide coding sequence
GAGGGCAGAGGAAGUCUUCUAACAUGCGGUGACGUGGAGGAGAAUCCCGGCCCU

SEQ ID NO: 13
CACCGACGCAACCGCCGGACGGGTACAAACCCCTTTGGGTTTGTACCCTGCGAGATT
ATGTCTGTAC

SEQ ID NO: 14
GTGGCTGCGTTGGCGGCCTGCCCATGTTTG

SEQ ID NO: 15
GTGGCTGCGTTGGCGGCCTGCCCATGTTTGGGGGGGGAAACCCAAACATGGGACGCT
CTAATACAGACATG

CLAIMS

What is claimed is:

1. A linear RNA, comprising from the 5' end to the 3' end:
   (a) a first portion of a RNA element;
   (b) an effector RNA sequence; and
   (c) a second portion of the RNA element;
   wherein the first portion of the RNA element and the second portion of the RNA element associate with each other to form a double-stranded region of at least 4 basepairs (bp) long;
   wherein the 5' end of the first portion of the RNA element and the 3' end of the second portion of the RNA element form a nick in the double-stranded region; and
   wherein the nick can be ligated by a RNA ligase.

2. The linear RNA of claim 1, wherein the double-stranded region is about 6 bp long to about 25 bp long.

3. The linear RNA of claim 1 or 2, wherein the double-stranded region has a minimum free energy of no more than about -8 kcal/mol at 37°C and an ionic strength of 1.0M Na+.

4. The linear RNA of any one of claims 1-3, wherein the RNA element is at least 20 nt long.

5. The linear RNA of any one of claims 1-4, wherein the double-stranded region comprises at least 2 bp 3' to the nick.

6. The linear RNA of any one of claims 1-5, wherein the double-stranded region comprises at least 2 bp 5' to the nick.

7. The linear RNA of any one of claims 1-6, wherein the first portion of the RNA element and the second portion of the RNA element are derived from a reference RNA element of N nucleotides, wherein N is an integer greater than 8; wherein the first portion of the RNA element comprises nucleotides X to N of the reference RNA element, wherein X is an integer greater than 1 and smaller than N; and wherein the second portion of the RNA element comprises nucleotides 1 to X-1 of the reference RNA element.

8. The linear RNA of claim 6, wherein the first portion of the RNA element and/or the second portion of the RNA element comprises a sequence that is exogenous to the reference RNA element.

9. The linear RNA of claim 8, wherein the second portion of the RNA element comprises a first exogenous sequence at the 3' end, wherein the first portion of the RNA element comprises a second exogenous sequence that is complementary of the first exogenous sequence, and wherein the first exogenous sequence and the second exogenous sequence form base pairs flanking the 5' end of the nick.

10. The linear RNA of any one of claims 1-9, wherein the RNA element is a naturally occurring RNA element or a derivative thereof.

11. The linear RNA of any one of claims 1-10, wherein the effector RNA sequence is a coding RNA sequence.

12. The linear RNA of claim 11, wherein the coding RNA sequence encodes a therapeutic polypeptide.

13. The linear RNA of claim 12, wherein the therapeutic polypeptide is selected from the group consisting of an antigenic polypeptide, a functional protein, a receptor protein, and a targeting protein.

14. The linear RNA of claim 12 or 13, wherein the RNA element promotes translation of the coding RNA.

15. The linear RNA of claim 14, wherein the RNA element is an internal ribosomal entry site (IRES) or a portion thereof.

16. The linear RNA of claim 15, wherein the IRES is derived from an IRES selected from the group consisting of Coxsackievirus B3 (CVB3) IRES, Enterovirus 71 (EV71) IRES, encephalomyocarditis virus (EMCV) IRES, picornavirus (PV) IRES, hepatitis C virus (HCV) IRES, adenovirus (AdV) IRES, human papillomavirus type 31 (HPV31) IRES, human herpesvirus (HHV) IRES, Rous sarcoma virus (RSV) IRES, classical swine fever virus (CSFV) IRES, FGF9 IRES, SLC7A1 IRES, and RUNX1 IRES.

17. The linear RNA of claim 16, wherein the IRES is an IRES of a CVB3 virus or a derivative thereof.

18. The linear RNA of claim 17, wherein the IRES of a CVB3 virus comprises the nucleotide sequence of SEQ ID NO: 1.

19. The linear RNA of claim 18, wherein the first portion of the RNA element comprises nucleotides 386-749 of SEQ ID NO: 1, and wherein the second portion of the RNA element comprises nucleotides 1-385 of SEQ ID NO: 1.

20. The linear RNA of claim 19, wherein the first portion of the RNA element comprises the nucleotide sequence of SEQ ID NO: 3, and wherein the second portion of the RNA element comprises the nucleotide sequence of SEQ ID NO: 2.

21. The linear RNA of claim 18, wherein the first portion of the RNA element comprises nucleotides 347-749 of SEQ ID NO: 1, and wherein the second portion of the RNA element comprises nucleotides 1-346 of SEQ ID NO: 1.

22. The linear RNA of any one of claims 11-21, further comprising an in-frame 2A peptide coding sequence that is operably linked to the 3' end of the coding RNA.

23. The linear RNA of any one of claims 1-9, wherein the RNA element is a sgRNA or a H/ACA box small nucleolar RNA (snoRNA).

24. The linear RNA of any one of claims 1-10 and 23, wherein the effector RNA sequence is a sequence of a non-coding RNA selected from the group consisting of a guide RNA (gRNA), a deaminase-recruiting RNA (dRNA), a siRNA, a miRNA, a shRNA, and a long intervening non-coding (linc) RNA.

25. The linear RNA of any one of claims 1-24, wherein the effector RNA sequence is about 50 to about 5000 nucleotides (nt) long.

26. A method of preparing a circRNA, comprising:
(a) contacting the linear RNA of any one of claims 1-25 with a RNA ligase under conditions that allow ligation of the nick in the linear RNA to provide a circularized RNA product; and

(b) isolating the circularized RNA product, thereby providing the circRNA.

27. The method of claim 26, wherein the RNA ligase is T4 RNA ligase 1.

28. The method of claim 26, wherein the RNA ligase is T4 RNA ligase 2.

29. The method of claim 28, wherein the circRNA does not comprise nucleotide sequence that is exogenous to the linear RNA.

30. The method of any one of claims 26-29, wherein the ligation does not require presence of a splint oligonucleotide.

31. The method of any one of claims 26-30, wherein efficiency of the ligation is at least about 50%.

32. The method of any one of claims 26-31, further comprising treating the circularized RNA product with RNAse R.

33. The method of any one of claims 26-32, further comprising purifying the circularized RNA product.

34. The method of any one of claims 26-33, further comprising obtaining the linear RNA by *in vitro* transcription of a nucleic acid construct comprising a nucleic acid sequence encoding the linear RNA.

35. A circRNA prepared using the method of any one of claims 26-34.

36. A circRNA formed by ligation of the nick in the linear RNA of any one of claims 1-25.

37. A nucleic acid construct comprising a nucleic acid sequence encoding the linear RNA of any one of claims 1-25.

38. The nucleic acid construct of claim 37, comprising a T7 promoter operably linked to the nucleic acid sequence encoding the linear RNA.

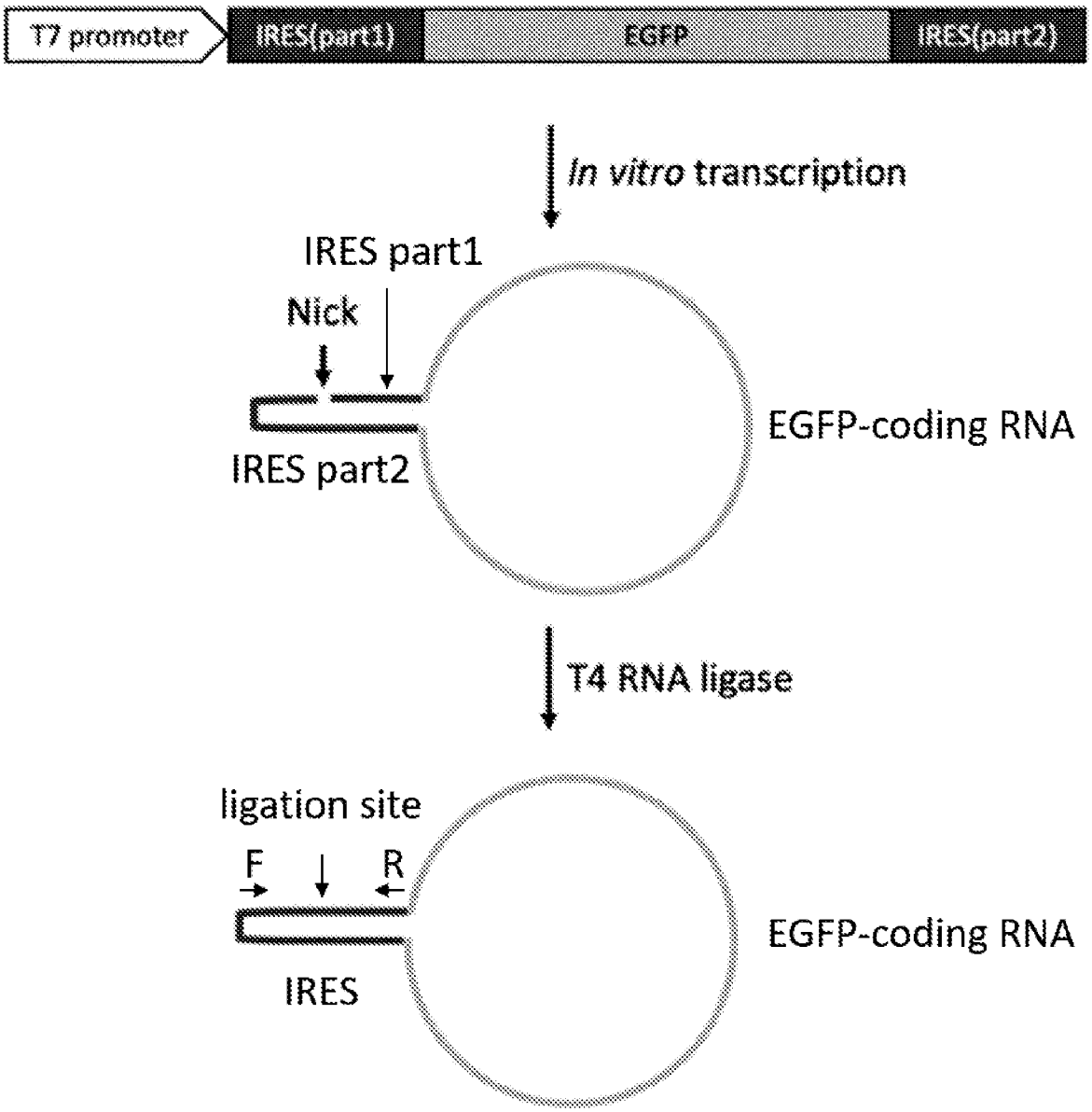

T7 promoter
IRES(part1)
EGFP
IRES(part2)
In vitro transcription
IRES part1
Nick
IRES part2
EGFP-coding RNA
T4 RNA ligase
ligation site
F
R
IRES
EGFP-coding RNA

FIG. 1

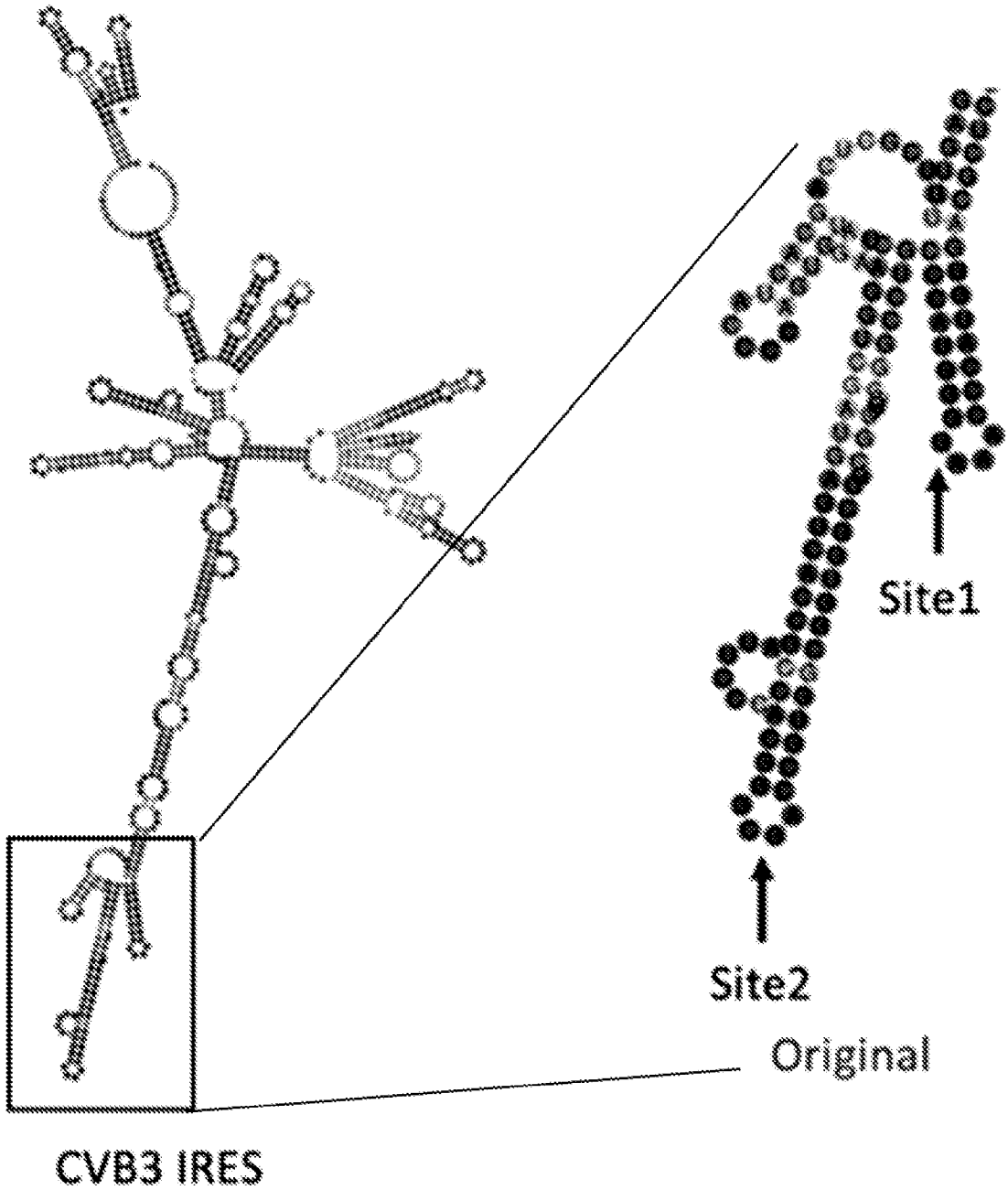
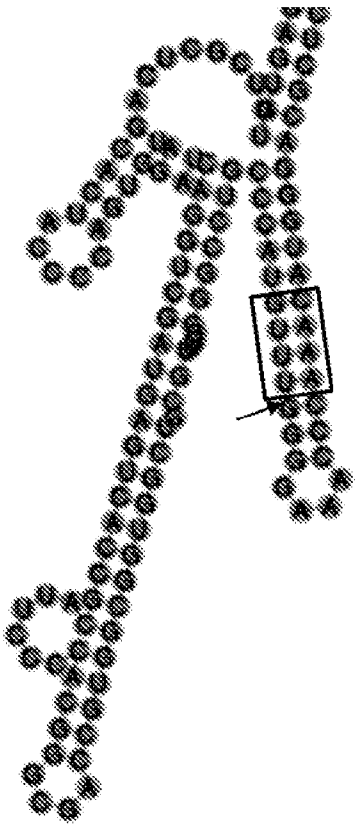
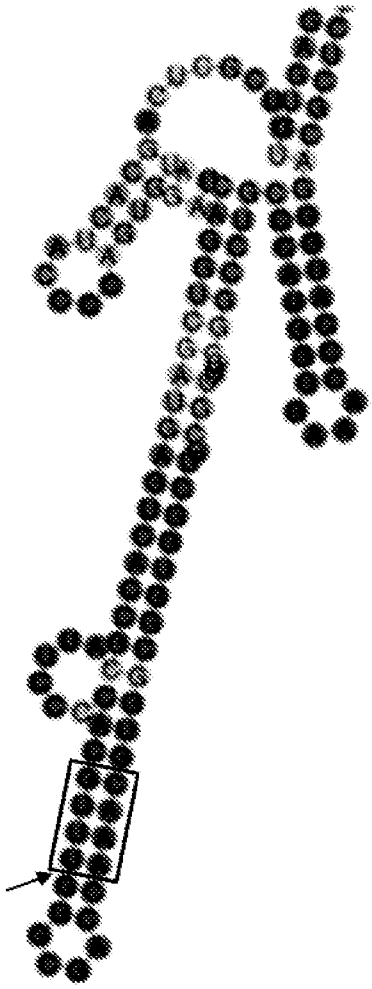


FIG. 2A



Site1

FIG. 2B



Site2

FIG. 2C

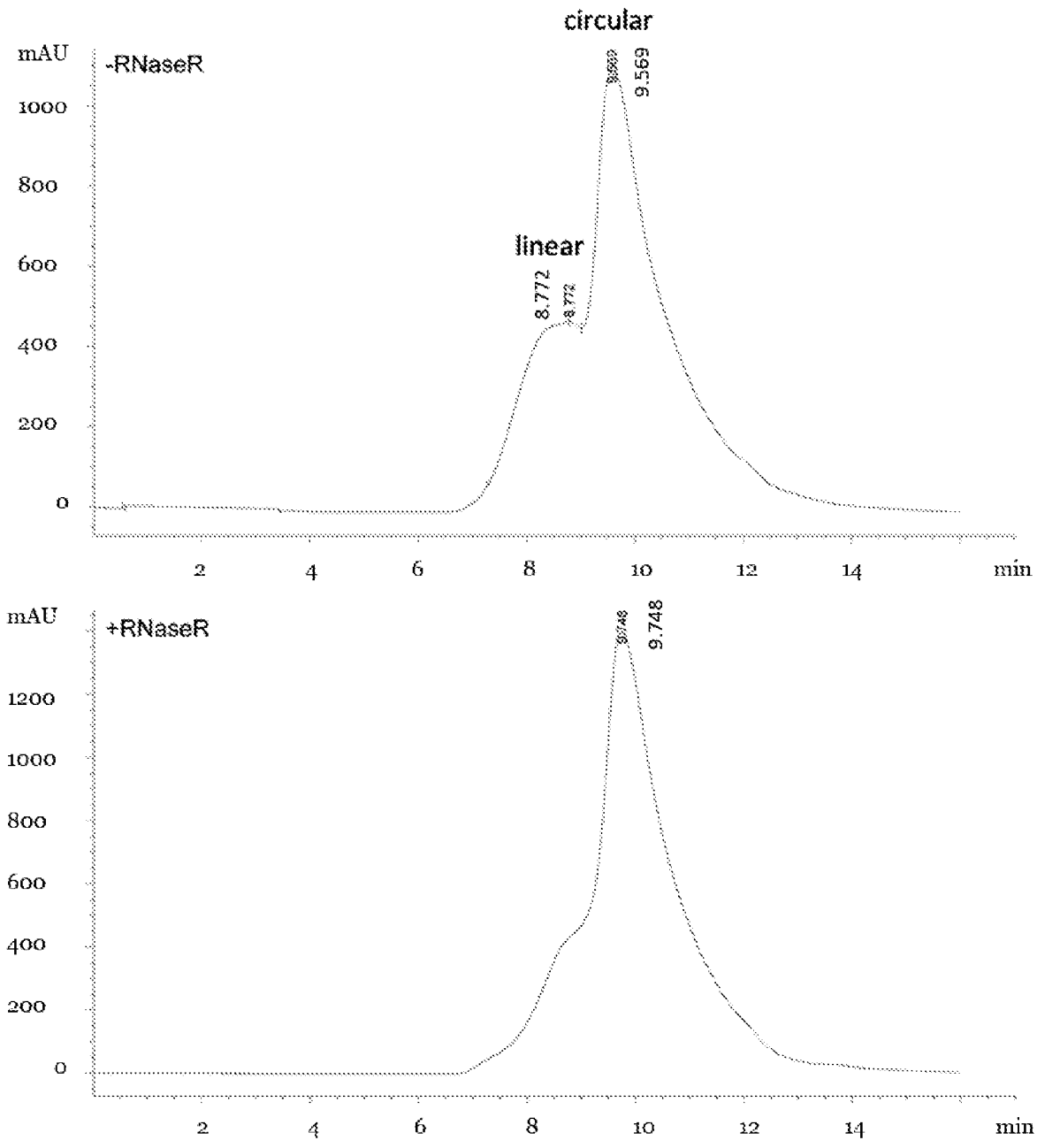

circular
mAU
-RNaseR
1000
800
600
400
200
0
9.569
linear
8.772
8.772
2
4
6
8
10
12
14
min
mAU
+RNaseR
1200
1000
800
600
400
200
0
9.748
2
4
6
8
10
12
14
min

FIG. 3

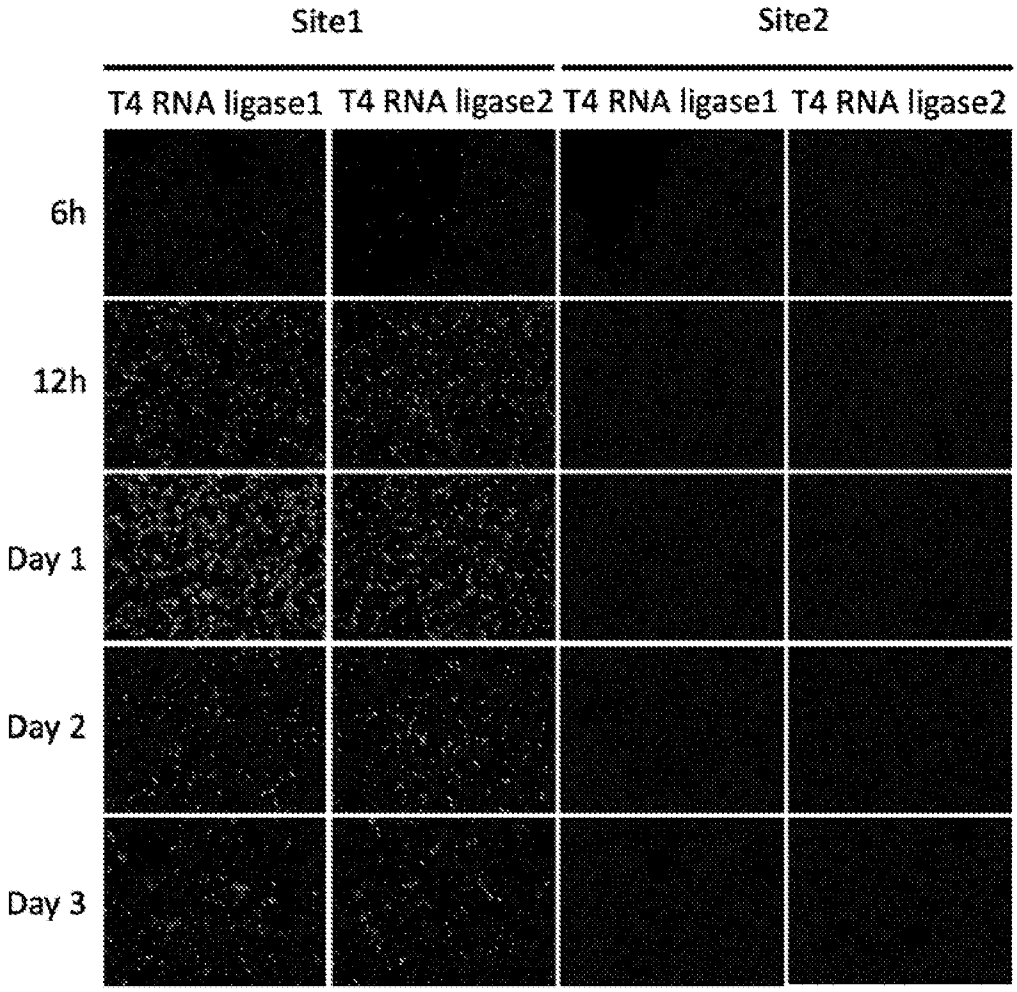


FIG. 4A

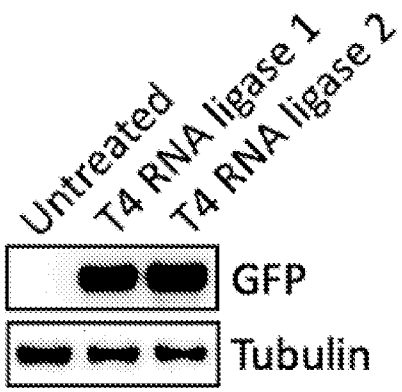


FIG. 4B

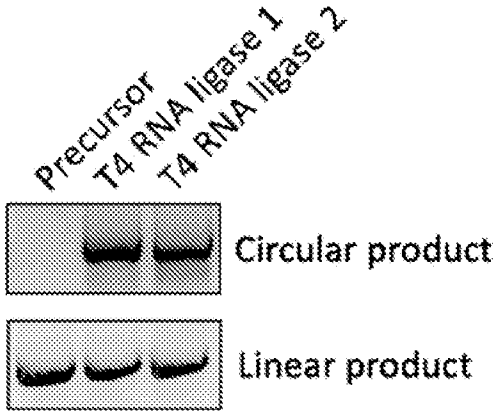


FIG. 5A

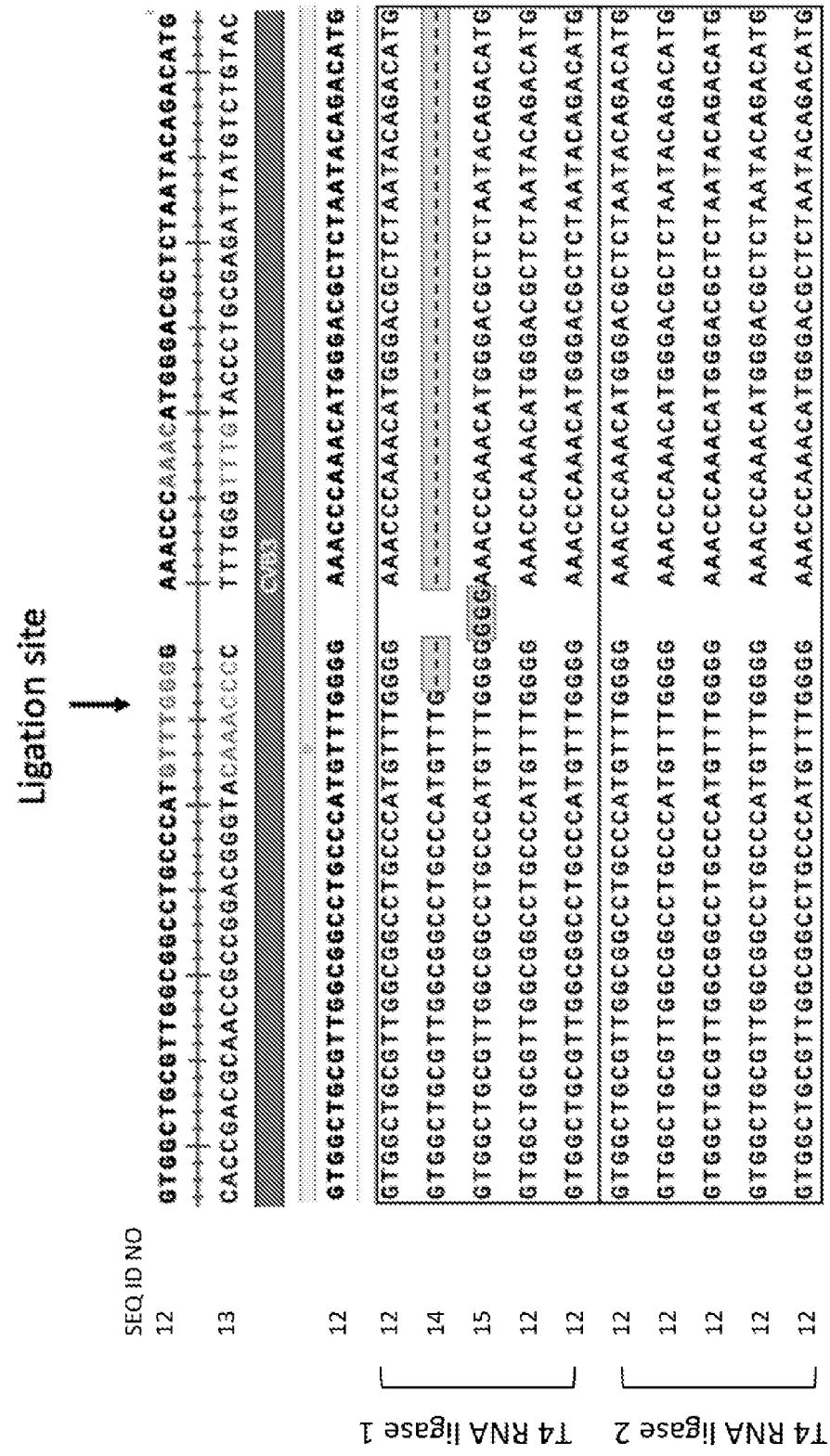

Ligation site
SEQ ID NO
12
13
12
12
14
15
12
12
12
12
12
12
12
T4 RNA ligase 1
T4 RNA ligase 2

FIG. 5B

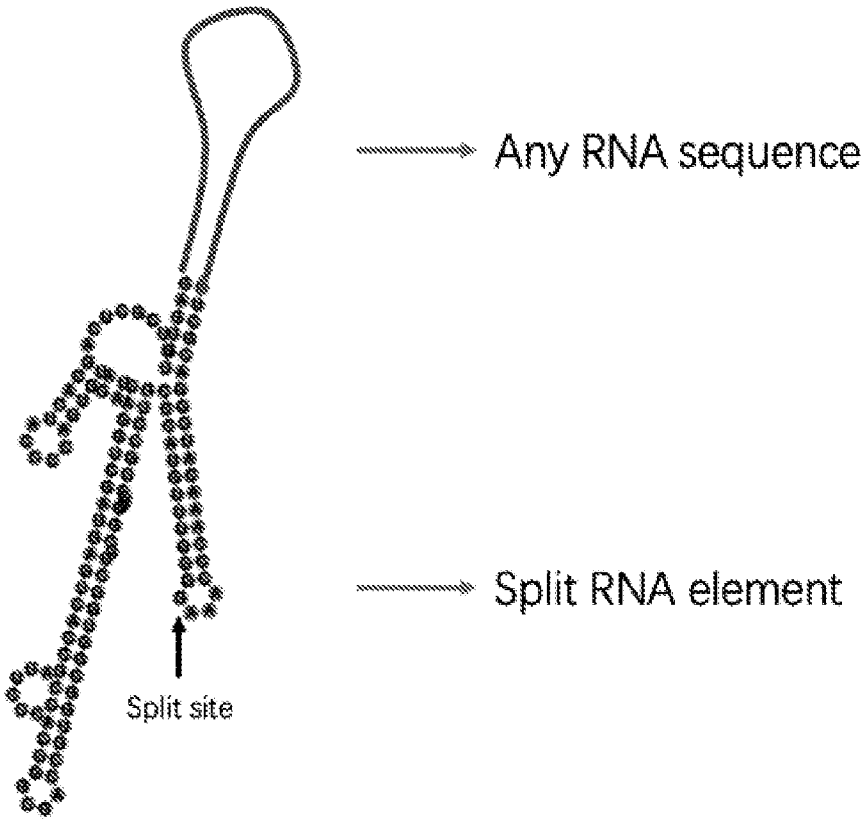

Any RNA sequence
Split RNA element
Split site

FIG. 6

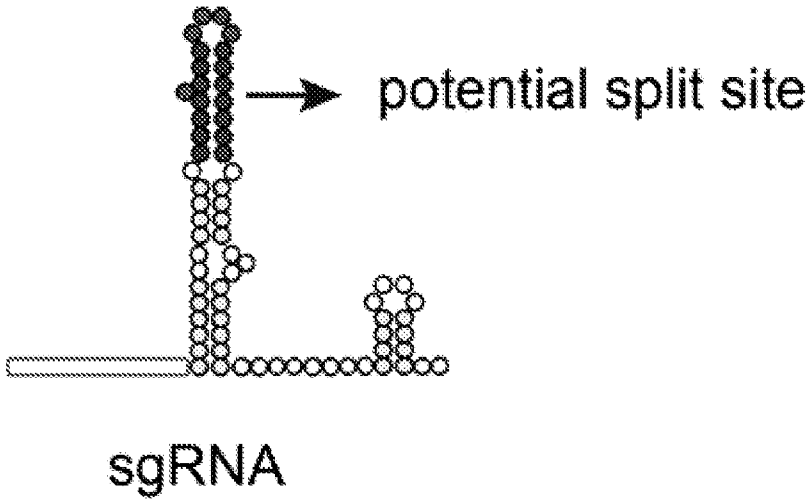

potential split site
sgRNA

FIG. 7

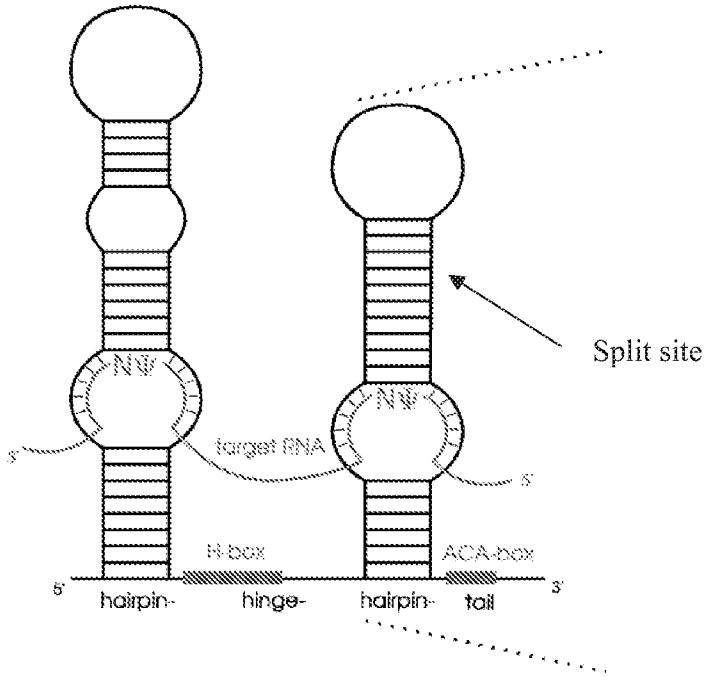

Split site
NΨ
NΨ
target RNA
5'
5'
5'
3'
B-box
ACA-box
hairpin-
hinge-
hairpin-
tail

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/082225

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/113(2010.01)i; C12P 19/34(2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N; C12P

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI;SIPOABS,CNABS,EPTXT;USTXT;WOTXT;CNTXT;CATXT;CNKI;WEB OF SCIENCE;PUBMED;WANFANG DATABASE:circular RNA?, circRNAs, linear RNA, precursor?, RNA element, IRES, nick, double-stranded region, RNA ligase, split, Genbank+EMBL+Retrieving system for biological sequence of Chinese Patent: sequence search on SEQ ID NOs:1-15.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2018237372 A1 (CORR UNIV CORNELL) 27 December 2018 (2018-12-27) see claims 11 and 20	35-36
Y	WO 2018237372 A1 (CORR UNIV CORNELL) 27 December 2018 (2018-12-27) see claims 11 and 20	1-34, 37-38
Y	CN 111321143 A (OCEAN UNIVERSITY OF CHINA) 23 June 2020 (2020-06-23) see the abstract and claim 1	1-34, 37-38
X	CN 112399860 A (MASSACHUSETTS INST TECHNOLOGY) 23 February 2021 (2021-02-23) see claim 1	35-36
X	US 5766903 A (UNIV TECHNOLOGY CORP) 16 June 1998 (1998-06-16) see claim 1	35-36
A	CN 101679962 A (RIKEN et al.) 24 March 2010 (2010-03-24) see the whole document	1-38
A	CN 111819280 A (TORAY IND INC) 23 October 2020 (2020-10-23) see the whole document	1-38

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 June 2022	**27 June 2022**
Name and mailing address of the ISA/CN **National Intellectual Property Administration, PRC** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088, China** Facsimile No. **(86-10)62019451**	Authorized officer **LI,Zhenpeng** Telephone No. **86-（010）-62411085**

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Zhang, ML et al. "A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis" *ONCOGENE*, Vol. 37, No. 13, 31 March 2018 (2018-03-31), ISSN: 0950-9232, pages 1805-1814, see the abstract	35-36

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/082225

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed:
      ☑ in the form of an Annex C/ST.25 text file.
      ☐ on paper or in the form of an image file.
   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).
      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2022/082225

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2018237372	A1	27 December 2018	EP	3642342	A1	29 April 2020
CN	111321143	A	23 June 2020	None			
CN	112399860	A	23 February 2021	BR	112020024292	A2	02 March 2021
				US	2020080106	A1	12 March 2020
				US	2021363540	A1	25 November 2021
				US	2021403944	A1	30 December 2021
				KR	20210018323	A	17 February 2021
				US	2021198688	A1	01 July 2021
				SG	11202011162R	A	30 December 2020
				JP	2021526792	A	11 October 2021
				US	2022025395	A1	27 January 2022
				AU	2019280583	A1	03 December 2020
				WO	2019236673	A1	12 December 2019
				CA	3100276	A1	12 December 2019
				IL	279157	D0	31 January 2021
				EP	3801639	A1	14 April 2021
US	5766903	A	16 June 1998	AU	7103796	A	19 March 1997
				CA	2229454	A1	06 March 1997
				EP	0846002	A1	10 June 1998
				WO	9707825	A1	06 March 1997
CN	101679962	A	24 March 2010	ES	2446040	T3	06 March 2014
				JP	2008278784	A	20 November 2008
				EP	2143792	A1	13 January 2010
				IL	201975	D0	01 August 2011
				CA	2685994	A1	20 November 2008
				AR	072034	A1	04 August 2010
				HR	P20140170	T1	25 April 2014
				US	2018073020	A1	15 March 2018
				MX	2009011101	A	11 December 2009
				WO	2008140126	A1	20 November 2008
				UA	101806	C2	13 May 2013
				MY	154495	A	30 June 2015
				US	2019292538	A1	26 September 2019
				CO	6241167	A2	20 January 2011
				RU	2009145502	A	20 June 2011
				NZ	581251	A	25 November 2011
				PT	2143792	E	03 April 2014
				KR	20100024407	A	05 March 2010
				US	2010137407	A1	03 June 2010
				SI	2143792	T1	30 May 2014
				CY	1115283	T1	04 January 2017
				BR	PI0811440	A2	21 October 2014
				AU	2008250075	A1	20 November 2008
				TW	200909577	A	01 March 2009
				PL	2143792	T3	31 July 2014
				IN	7243CHENP2009	A	12 March 2010
				DK	2143792	T3	07 April 2014
CN	111819280	A	23 October 2020	EP	3778886	A1	17 February 2021
				KR	20200136363	A	07 December 2020

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2022/082225

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
		CA	3094160	A1	03 October 2019
		TW	202003842	A	16 January 2020
		AU	2019242331	A1	10 September 2020
		IL	277346	D0	29 October 2020
		WO	2019189722	A1	03 October 2019
		PH	12020551487	A1	23 August 2021
		ZA	202005149	B	26 January 2022
		JP	2019213567	A	19 December 2019
		BR	112020017769	A2	05 January 2021
		US	2021024930	A1	28 January 2021
		RU	2020130258	A	04 May 2022
		JP	6631751	B1	15 January 2020