



US 20240165144A1

(19) **United States**

(12) **Patent Application Publication**  
Wheeler et al.

(10) **Pub. No.: US 2024/0165144 A1**

(43) **Pub. Date: May 23, 2024**

(54) **COMPOSITIONS AND METHODS FOR MODULATING INNATE IMMUNE SIGNALING PATHWAYS**

(71) Applicant: **Dana-Farber Cancer Institute, Inc.**, Boston, MA (US)

(72) Inventors: **Douglas Wheeler**, Boston, MA (US); **Matthew Meyerson**, Boston, MA (US)

(21) Appl. No.: **17/769,517**

(22) PCT Filed: **Oct. 16, 2020**

(86) PCT No.: **PCT/US20/56157**

§ 371 (c)(1),

(2) Date: **Apr. 15, 2022**

**Related U.S. Application Data**

(60) Provisional application No. 62/916,153, filed on Oct. 16, 2019.

**Publication Classification**

(51) **Int. Cl.**

**A61K 31/7076** (2006.01)

**C12N 9/22** (2006.01)

**C12Q 1/6874** (2018.01)

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

CPC ..... **A61K 31/7076** (2013.01); **C12N 9/22**

(2013.01); **C12Q 1/6874** (2013.01); **C12N**

**2310/20** (2017.05); **C12Q 2600/106** (2013.01);

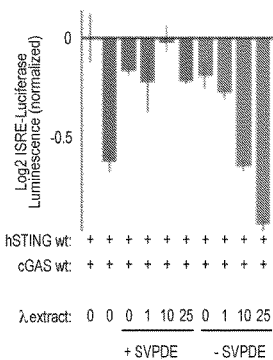
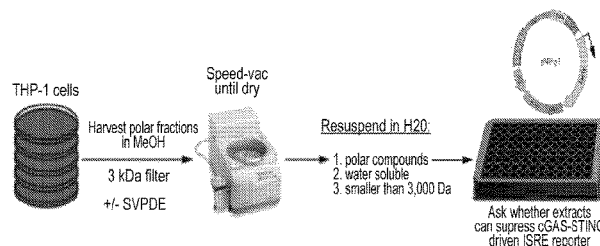
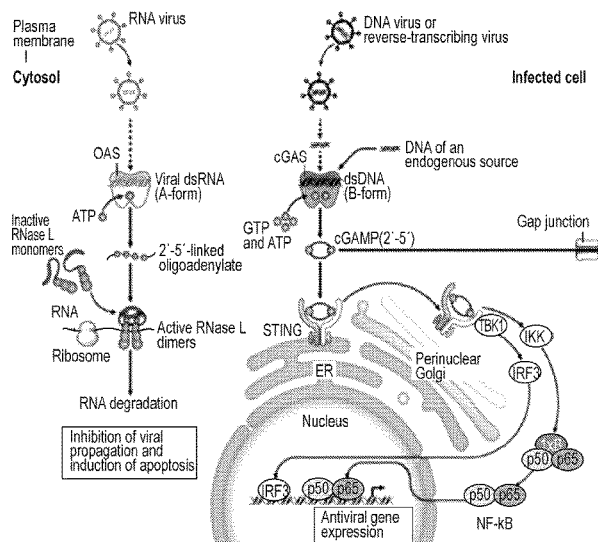
**C12Q 2600/158** (2013.01)

(57)

**ABSTRACT**

The present invention provides novel compositions and methods based on the discovery of the mechanisms associated with STING signaling. Compositions identified find use in treatment of diseases where STING signaling is implicated in the pathology.

**Specification includes a Sequence Listing.**



Dose(s) of cell extracts from THP-1 cells

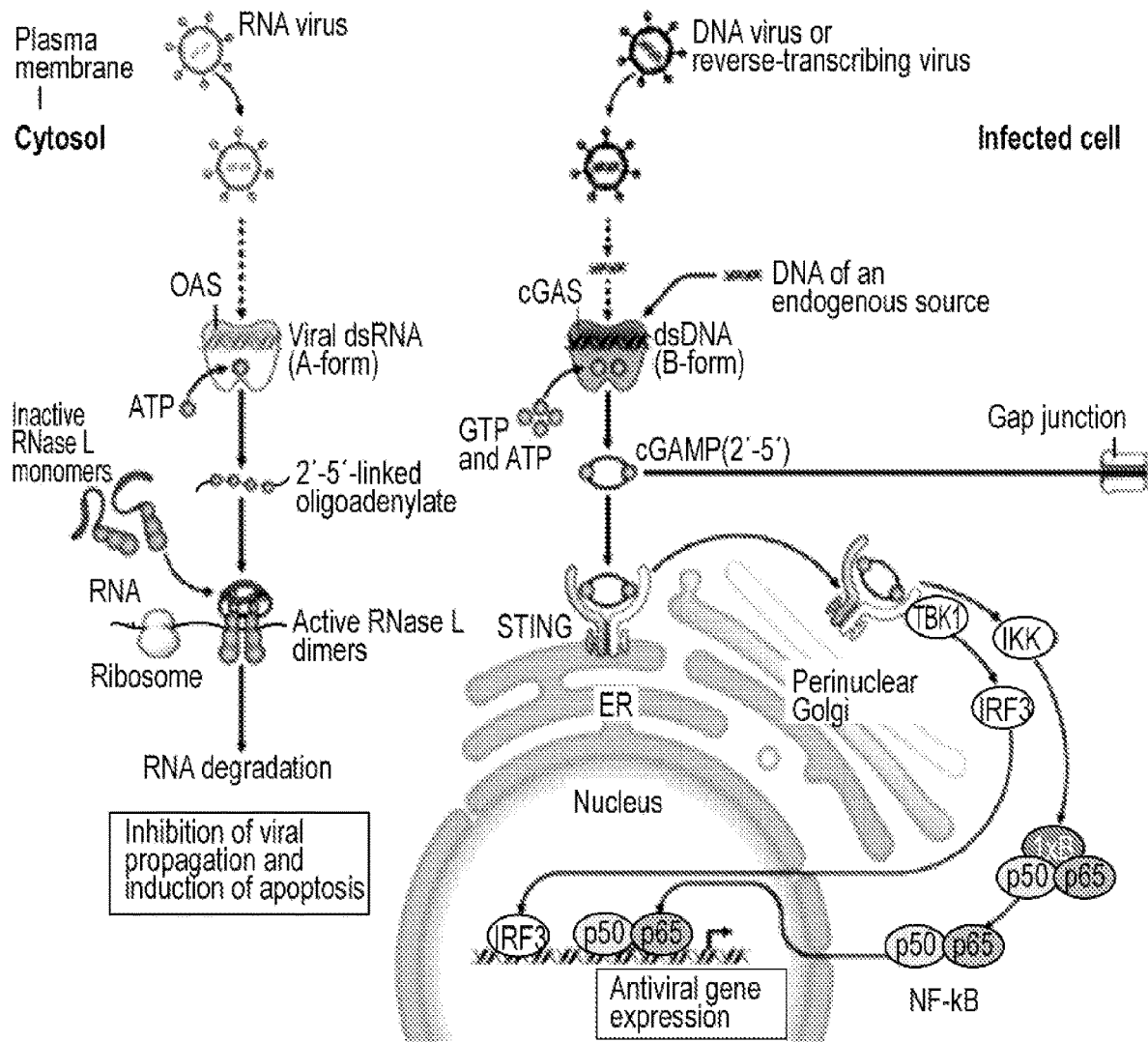


FIG. 1A

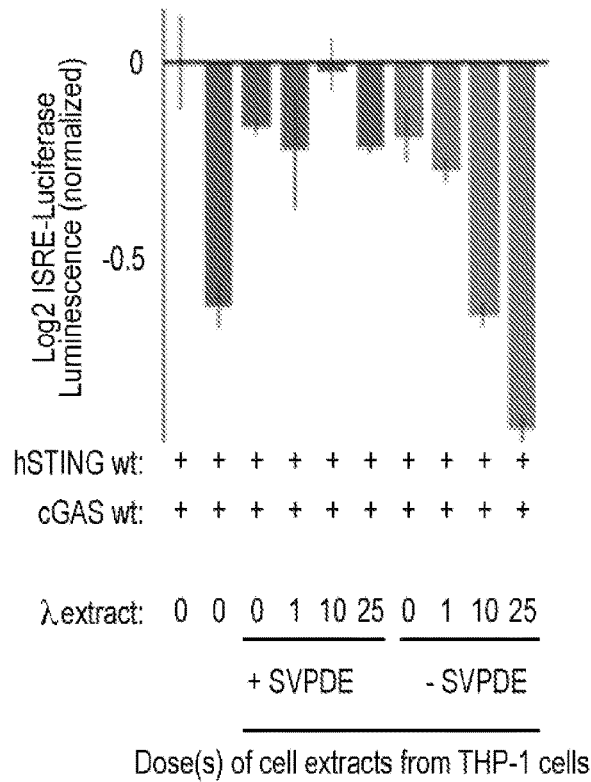
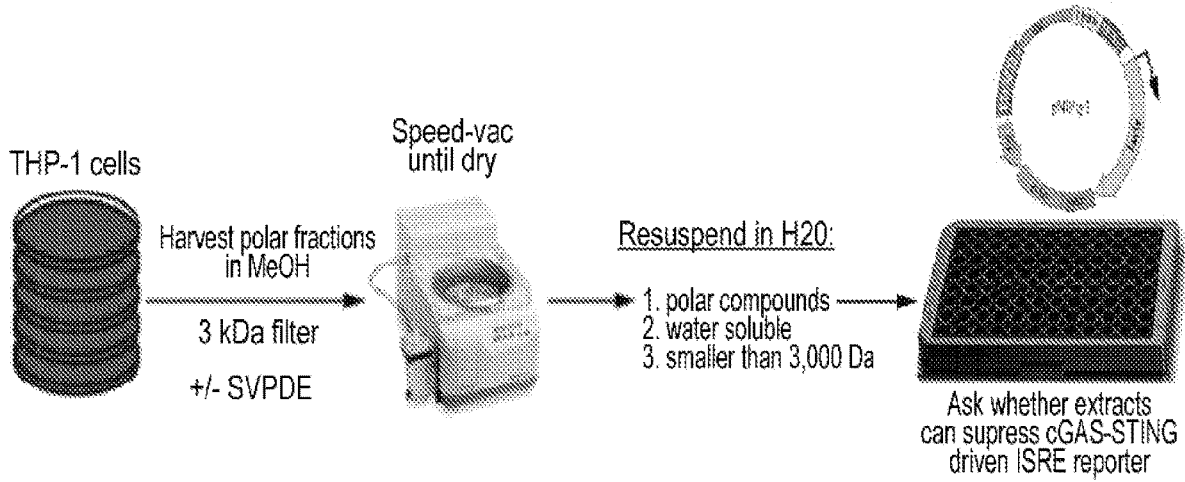
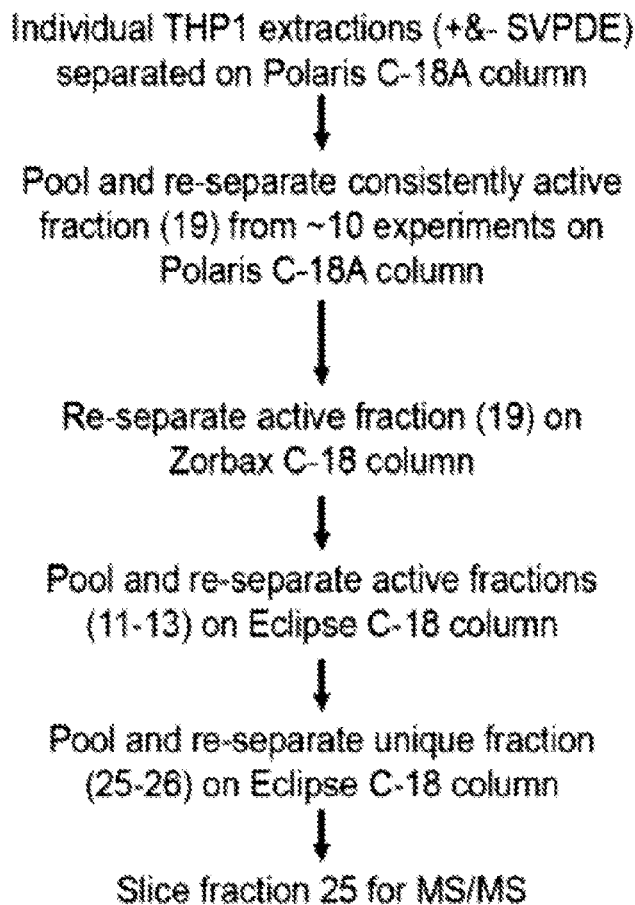
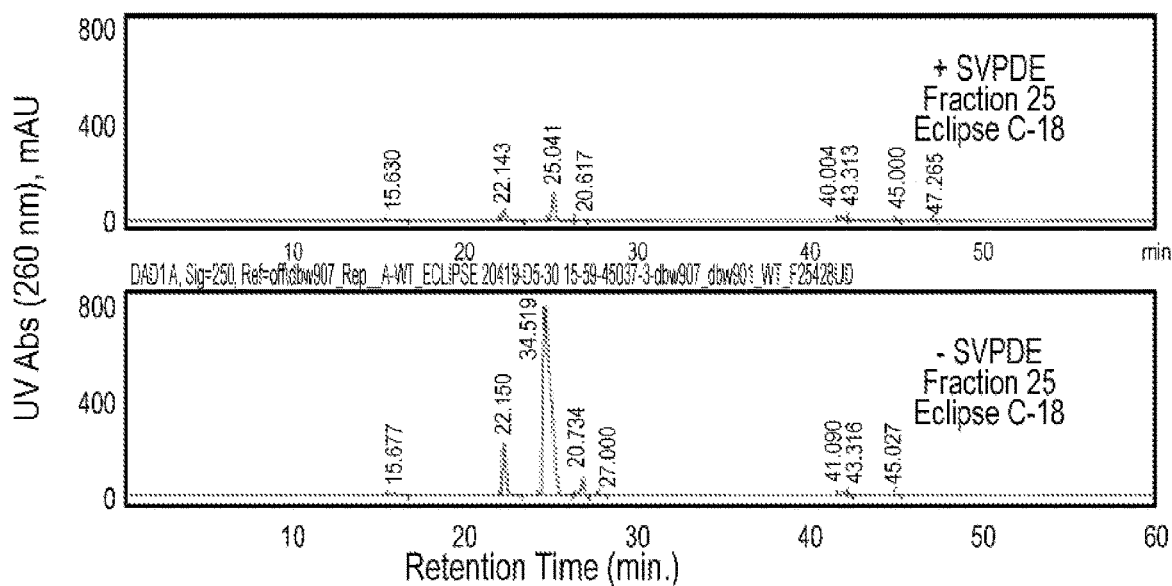


FIG. 1B



**FIG. 2**



**FIG. 3**

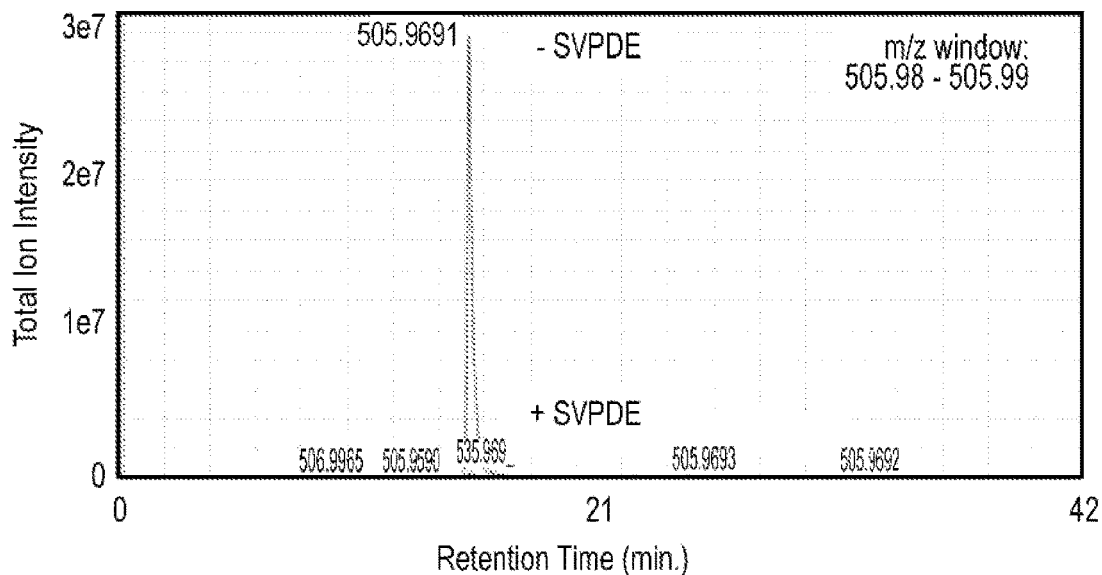


FIG. 4

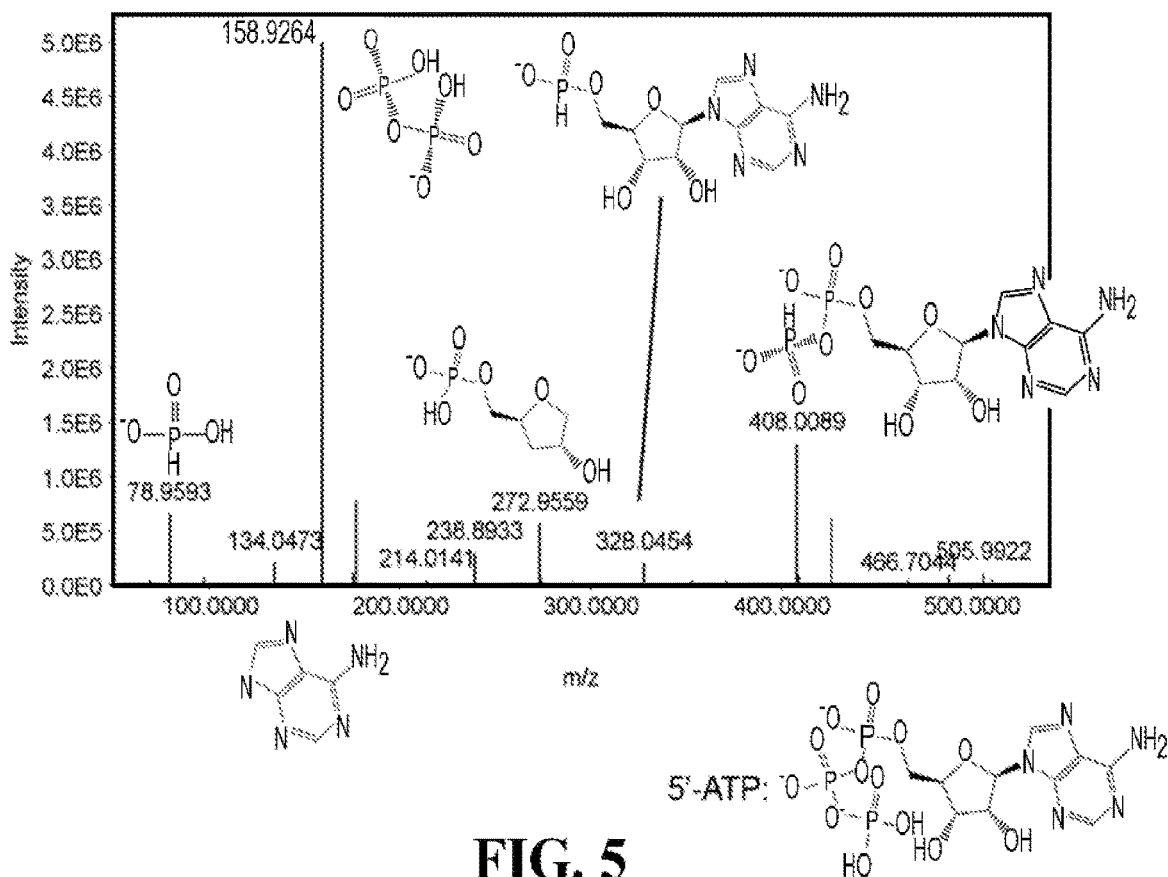


FIG. 5

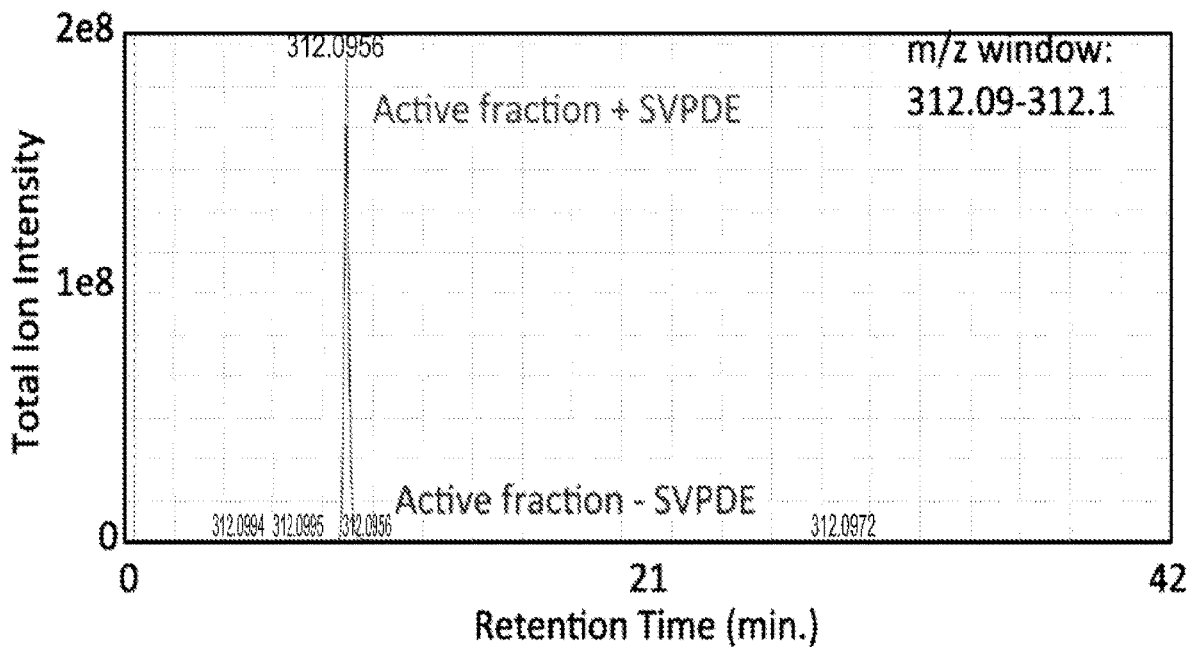


FIG. 6A

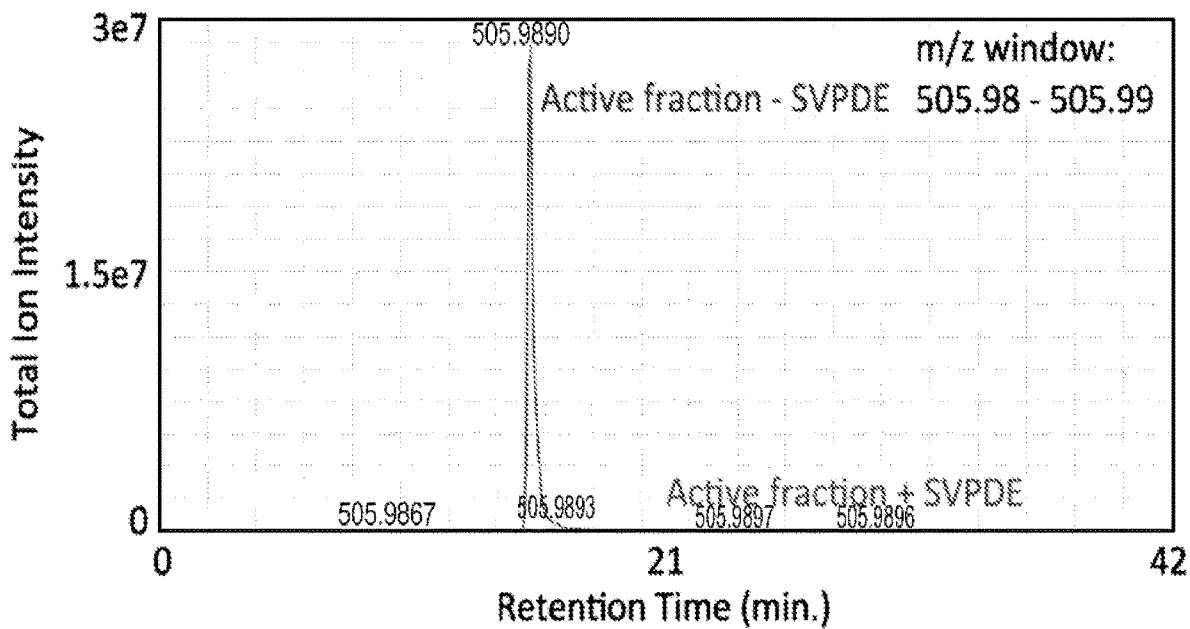


FIG. 6B

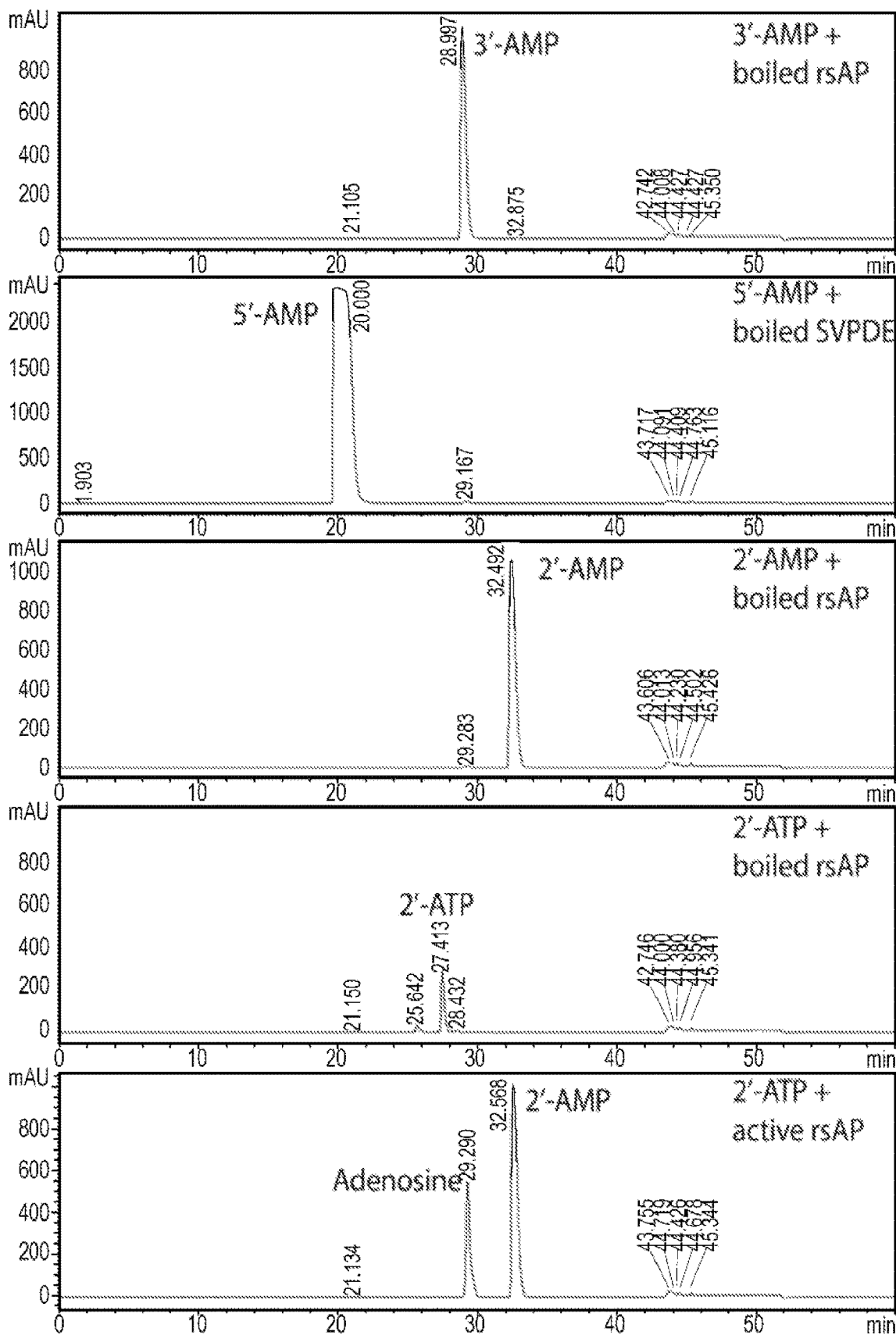


FIG. 7

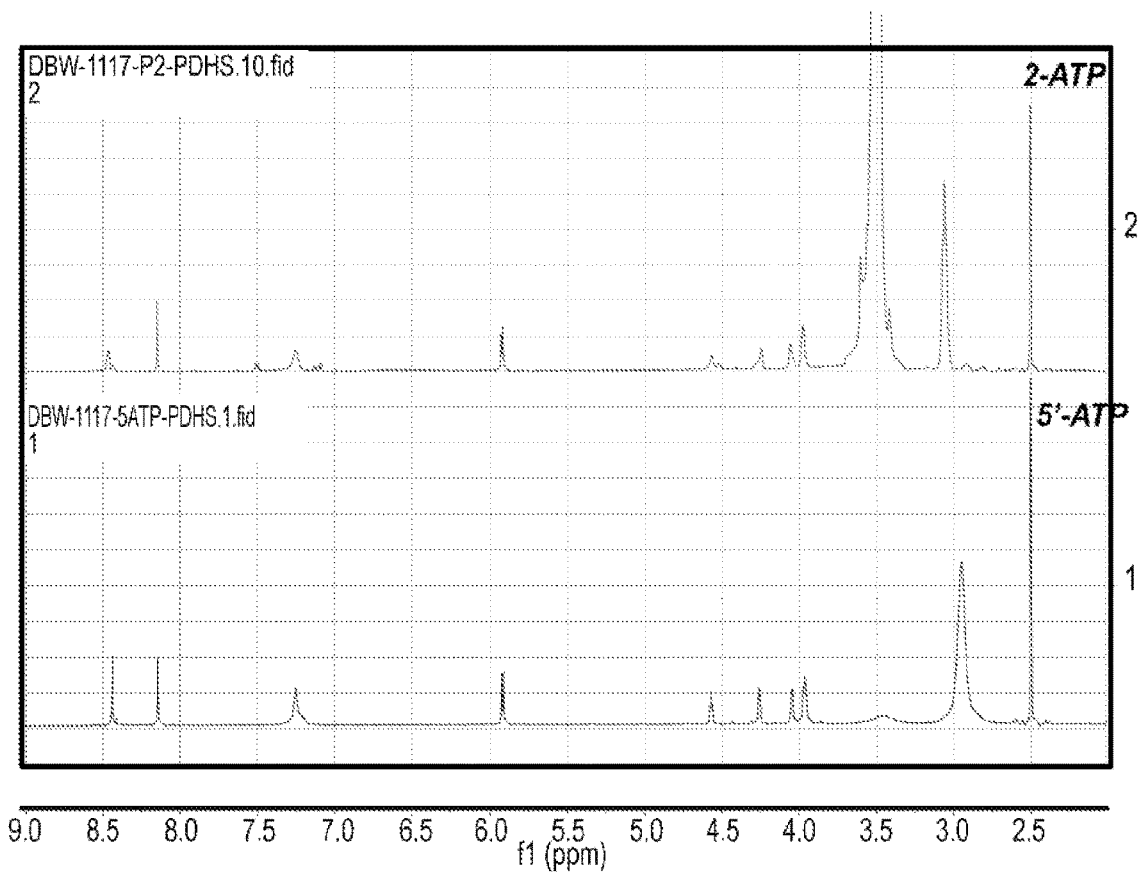


FIG. 8A

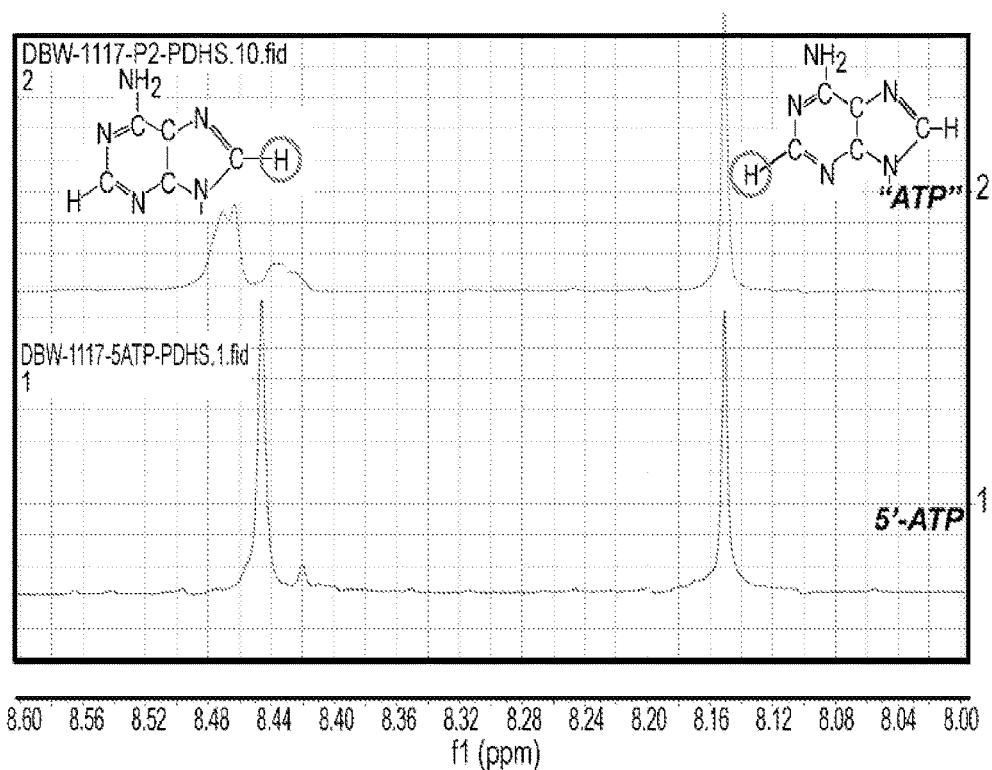


FIG. 8B



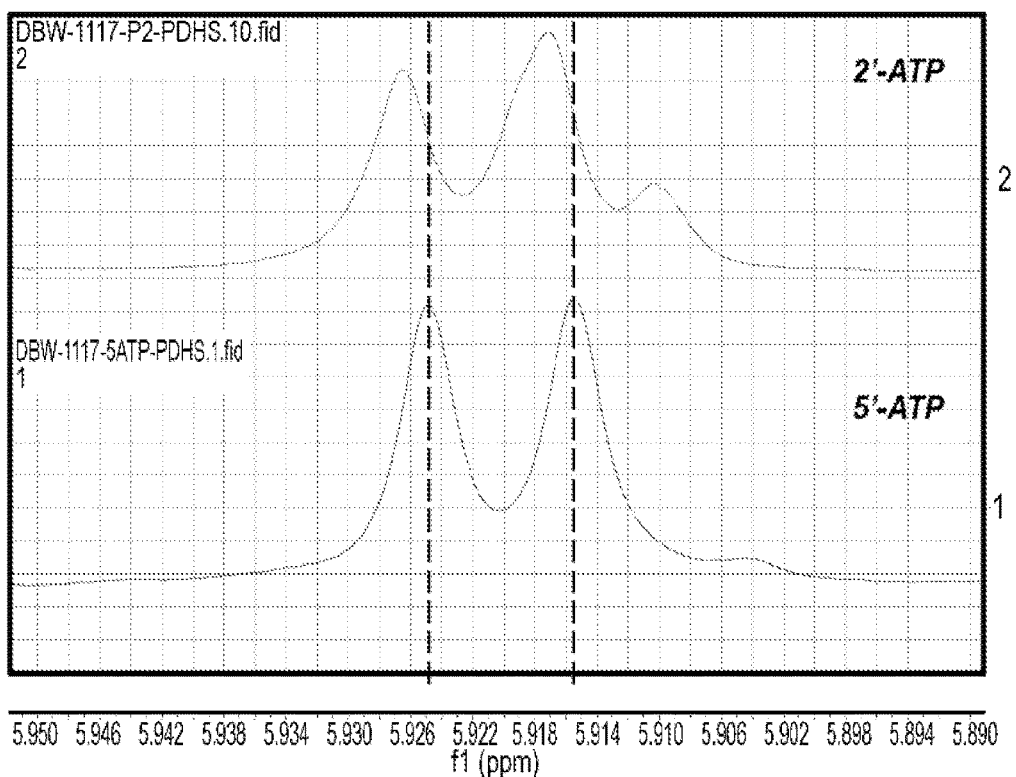


FIG. 8C

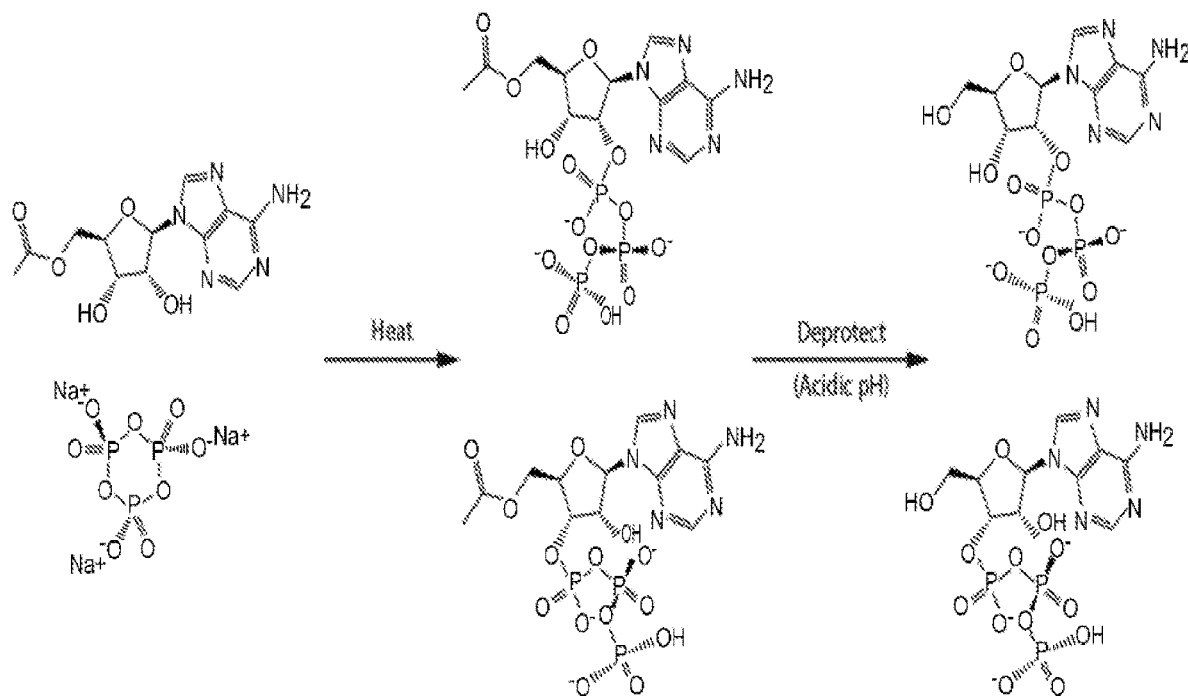
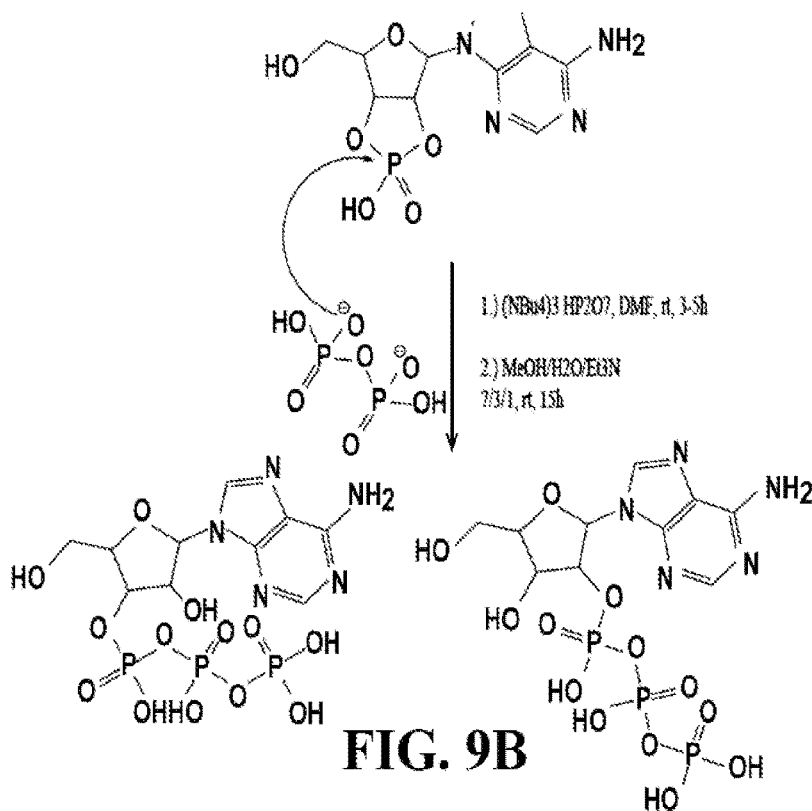
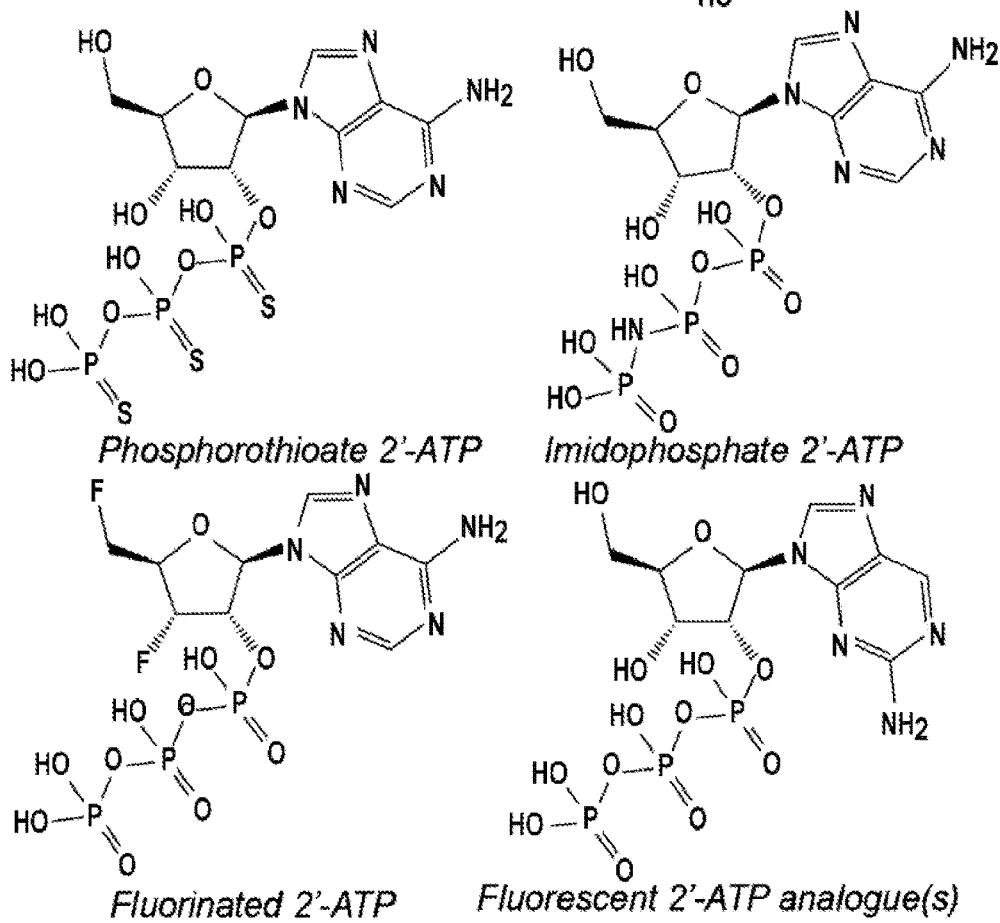


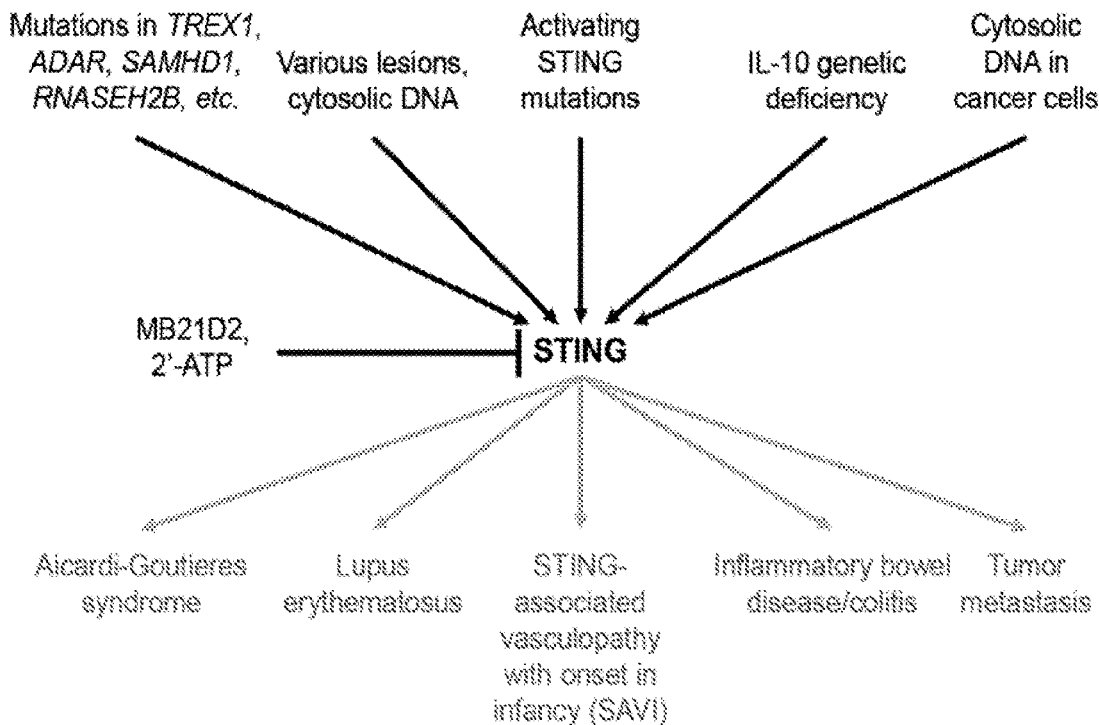
FIG. 9A



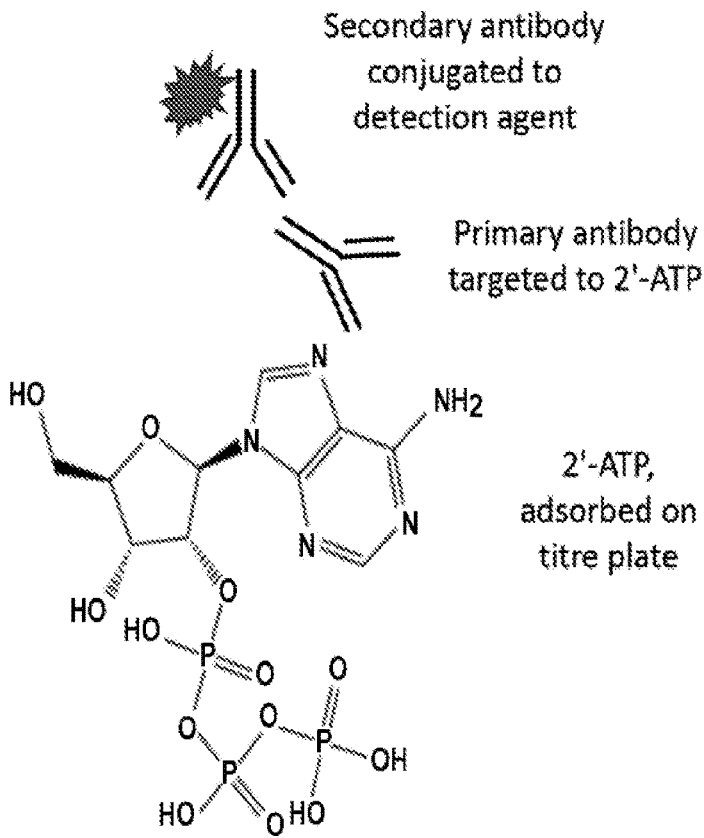
**FIG. 9B**



**FIG. 9C**



**FIG. 10**



**FIG. 11**

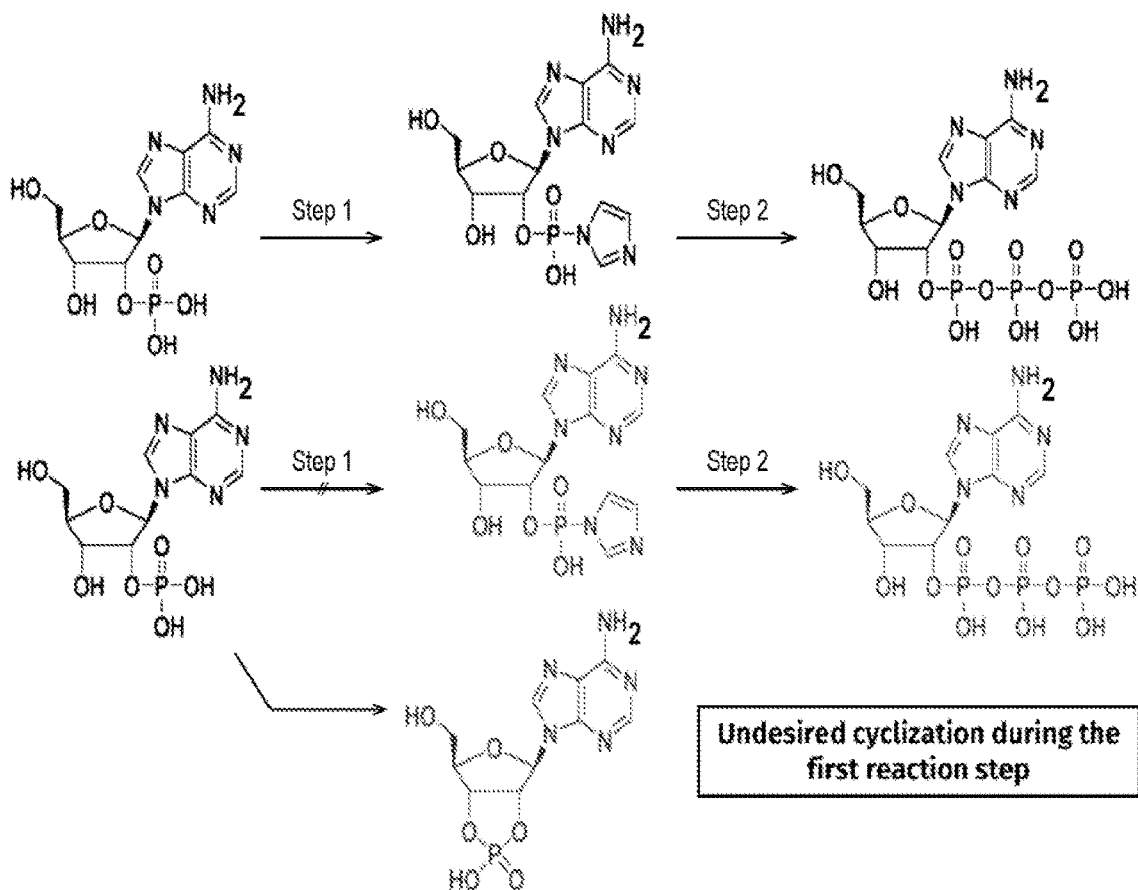
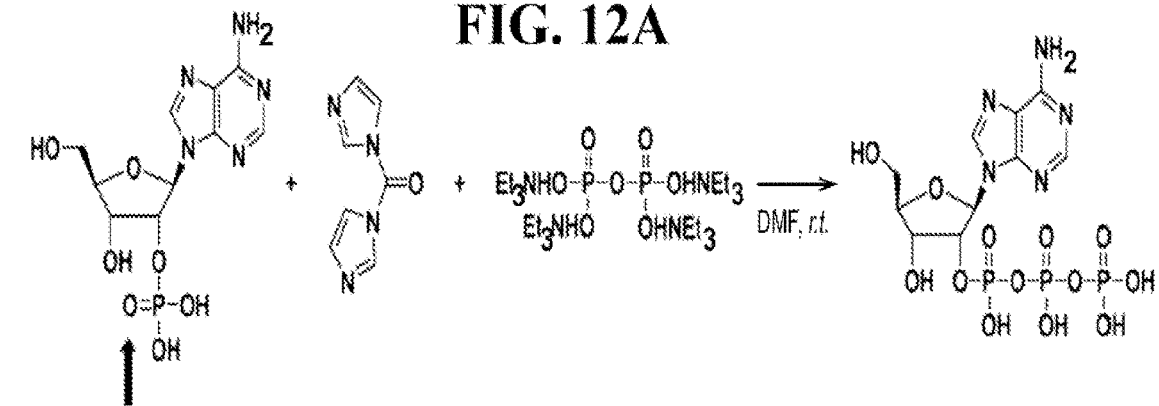


FIG. 12A



Batch#	Method	Observed Cyclisation (% HPLC)	Used 2'AMP (μmol)	Yield (μmol)	Yield (%)
FK020-070	Two-step approach	100	576	Not detected	Not detected
FK020-078*	One-pot	80	288	49	17
FK021-020	One-pot	85	576	56	10

FIG. 12B

## COMPOSITIONS AND METHODS FOR MODULATING INNATE IMMUNE SIGNALING PATHWAYS

### CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims the benefit of U.S. Provisional Application No. 62/916,153, filed Oct. 16, 2019. The entire contents of the above-identified application are hereby fully incorporated herein by reference.

### STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

**[0002]** This invention was made with government support under Grant No. CA197568 awarded by the National Institutes of Health. The government has certain rights in the invention.

### REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

**[0003]** The contents of the electronic sequence listing (“BROD\_5000US\_ST25.txt”; Size is 8 Kilobytes and it was created on Oct. 16, 2020) is herein incorporated by reference in its entirety.

### TECHNICAL FIELD

**[0004]** The subject matter disclosed herein is generally directed to compositions and methods targeting STING signaling.

### BACKGROUND

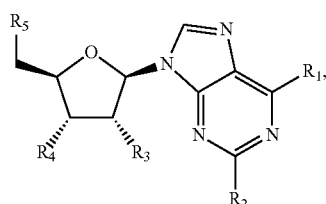
**[0005]** Compositions and methods for modulating the Stimulator of Interferon Genes (STING) signaling activity, which has therapeutic applications in various disease states would be an advance in the art.

**[0006]** Citation or identification of any document in this application is not an admission that such a document is available as prior art to the present invention.

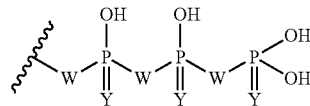
### SUMMARY

**[0007]** In certain example embodiments, an agent for inhibiting a Stimulator of Interferon Genes (STING) signaling pathway is provided. The agent in an aspect is a 2'-ATP derivative, or a 3'-ATP derivative.

**[0008]** In embodiments, the agent is a 2'-ATP derivative is according to the formula:



wherein  $R_1$  and  $R_2$  are independently H or  $NH_2$ ,  $R_3$  is



wherein Y is O or S, and W is O or NH, and  $R_4$  and  $R_5$  are independently OH or F.

**[0009]** A method of inhibiting STING/IFN $\beta$  signaling is disclosed comprising administering to a subject in need thereof a therapeutically effective amount of a STING signaling antagonist. In an aspect, the STING signaling antagonist is 2'-ATP. Administration of the STING signaling antagonist via a delivery vehicle comprising liposomes, lipid particles, or nanoparticles.

**[0010]** In embodiments, the methods comprise administering the agents herein to a subject that suffers from an interferonopathy or auto-inflammatory disease. In an aspect, the subject suffers from Aicardi-Goutieres syndrome, Lupus erythematosus, STING-associated vasculopathy with onset in infancy (SAVI), inflammatory bowel disease, colitis, or another disease process where STING signaling drives pathology. The agent can comprise a small molecule, small molecule degrader, genetic modifying agent, antibody, antibody fragment, antibody-like protein scaffold, aptamer, protein, or any combination thereof.

**[0011]** In embodiments, the agent is a genetic modifying agent. The method of claim 13, wherein the genetic modifying agent is a TALEN, a Zn-finger nuclease, a CRISPR-Cas system, or a meganuclease.

**[0012]** In embodiments, the genetic modifying agent is a Class 1 or Class 2 CRISPR-Cas system. In an aspect, the Class 2 CRISPR-Cas system is a Type II, Type V or Type VI CRISPR-Cas system. The Type II CRISPR-Cas system can comprise Cas9. The Type V CRISPR-Cas system can be a Cas12a, Cas12b, Cas12c, or Cas12d system in embodiments, or may be a Type VI CRISPR-Cas system that is a Cas13a, Cas13b, Cas13c or Cas13d system. In embodiments, the Cas protein is a dCas protein fused to a functional domain, which may be a nucleotide deaminase domain. The nucleotide deaminase can be a cytidine deaminase or an adenosine deaminase. The method of claim 20, wherein the dCas is a Type V or type VI.

**[0013]** Methods disclosed herein may further comprise detecting, prior to an administering step, the presence of innate immunity pathway activation, wherein the one or more agents are administered only if innate immunity pathway activation is detected. In embodiments, the detecting innate immunity activation comprise detecting mutations in one or more of TREG1, ADAR, STING, or other similar genes involved in innate immune signaling. Detecting can be performed in certain embodiments by sequencing, amplification, hybridization, or CRISPR-based detection, and/or detecting innate immunity pathway activation comprises biochemical detection of STING activation.

**[0014]** The agents herein can be administered in a delivery vehicle comprising liposomes, lipid particles, or nanoparticles. In an aspect, the agent comprises a small molecule, small molecule degrader, genetic modifying agent, antibody, antibody fragment, antibody-like protein scaffold, aptamer, protein, or any combination thereof. The agent may be any of the genetic modifying agents disclosed herein, including a TALEN, a Zn-finger nuclease, a CRISPR-Cas system as detailed herein, or a meganuclease.

[0015] Kits for detecting levels of 2'-ATP, 3'-ATP or derivatives thereof in a sample comprising a first binding molecule that bind 2'-ATP, 3'-ATP or derivatives thereof and a labeled binding molecule that binds the first binding molecule are provided. In an aspect, the kit further comprising a solid substrate capable of absorbing 2'ATP, 3'ATP or derivatives thereof.

[0016] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of example embodiments.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0017] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0018] FIG. 1A-FIG. 1B—FIG. 1A cGAS is an innate immune signaling enzyme that generates cyclic GMP-AMP in response to cytosolic DNA, image adapted from (7); FIG. 1B. Polar extracts from THP-1 cells treated with inactive snake venom phosphodiesterase (SVPDE) but not active SVPDE can suppress cGAS-STING signaling. Workflow at top shows method to measure the effects of cell extracts on interferon signaling; chart shows extracts only from cell extracts not treated with SVPDE.

[0019] FIG. 2—Charts two purification approaches used for chromatographic analysis.

[0020] FIG. 3—Fraction 25 from purification step 4 is chromatographically homogenous

[0021] FIG. 4—MS/MS analysis of the purified peak from FIG. 3 compared to the control fraction revealed this activity to have a predicted mass of 506.9969 Da.

[0022] FIG. 5—m/z 505.9891 is chemically similar to 5'-Adenosine Triphosphate

[0023] FIG. 6A-6B—Snake Venom PDEI breaks down the ATP product to Adenosine. FIG. 6A m/z window 312.09-312.1; FIG. 6B m/z window 505.98-505.99.

[0024] FIG. 7—Partial degradation with recombinant shrimp alkaline phosphatase (rsAP) of purified material reveals 2'-AMP but not 3'-AMP as a breakdown product of novel 2'-ATP. Upper panels are 2'-, 3'- and 5'-AMP, with the fourth panel showing the purified material appearing as a major peak when treated with inactive rsAP, but when treated with active rsAP the breakdown products that appear are adenosine and 2'-AMP (bottom panel), indicating that the purified molecule is 2'-ATP.

[0025] FIG. 8A-8C—FIG. 8A <sup>1</sup>H spectra of purified active 2'-ATP vs. 5'-ATP appear similar; FIG. 8B chemical shifts for the adenine protons are not identical; FIG. 8C the <sup>1</sup> carbon proton is shifted downfield from 5'-ATP.

[0026] FIG. 9A-9C—FIG. 9A One synthetic route proposed by using 5' protected Acetyl-O-Adenosine and trimetaphosphate; FIG. 9B another proposed synthetic route with synthesis of 2'-ATP and 3'ATP using 2',3'cAMP as a starting material; FIG. 9C Exemplary 2'-ATP derivatives.

[0027] FIG. 10—Exemplary diseases that can be targeted with therapeutics that inhibit STING, including the STING inhibitor 2'-ATP or mimetic molecules.

[0028] FIG. 11—Exemplary approach to the design and creation of an assay for detection and measurement of the MB21D2 product 2'-ATP.

[0029] FIG. 12A-12B 2'ATP synthesis. FIG. 12A Initial two-step proposed synthesis of 2'ATP proposed synthesis of 2'ATP produced undesired cyclization during the first reaction step. FIG. 12B a one-pot 2'ATP synthesis.

[0030] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

#### DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

##### General Definitions

[0031] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4<sup>th</sup> edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F. M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.); PCR 2: A Practical Approach (1995) (M. J. MacPherson, B. D. Hames, and G. R. Taylor eds.); Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.); Antibodies A Laboratory Manual, 2<sup>nd</sup> edition 2013 (E. A. Greenfield ed.); Animal Cell Culture (1987) (R. I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew et al. (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton et al., Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2<sup>nd</sup> edition (2011).

[0032] As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

[0033] The term “optional” or “optionally” means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0034] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0035] The terms “about” or “approximately” as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier “about” or “approximately” refers is itself also specifically, and preferably, disclosed.

[0036] As used herein, a “biological sample” may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a “bodily fluid”. The present invention encompasses embodiments

wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

**[0037]** The terms “subject,” “individual,” and “patient” are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

**[0038]** Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to “one embodiment,” “an embodiment,” “an example embodiment,” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases “in one embodiment,” “in an embodiment,” or “an example embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

**[0039]** All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

#### Overview

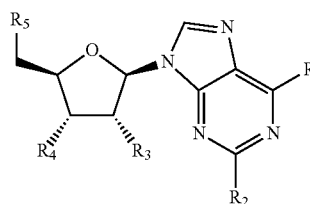
**[0040]** Embodiments disclosed herein provide compositions and methods for modulating the Stimulator of Interferon Genes (STING) signaling activity, which has therapeutic applications in various disease states. Inhibition of the STING signaling pathway may be therapeutically beneficial in the context diseases of auto-inflammation and auto-immunity. Thus, embodiments disclosed herein are directed to the use of 2' ATP/3' ATP and mimetics thereof in inhibiting the STING signaling pathway.

**[0041]** Further descriptions of therapeutic agents for use in these and other embodiments, as well as methods of treatment, diagnosis and screening are described in further detail below.

#### Therapeutic Agents for Inhibiting STING Signaling Pathway

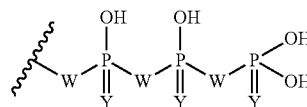
**[0042]** Agents disclosed herein can be utilized for inhibiting a Stimulator of Interferon Genes (STING) signaling pathway. In particular embodiments, agents are small molecules, including a 2'-ATP derivative, or a 3'-ATP derivative. An agent for inhibiting a Stimulator of Interferon Genes (STING) signaling pathway, wherein the agent is a 2'-ATP, a 3'-ATP, or a derivative thereof, is provided.

**[0043]** The small molecule may comprise a 2'-ATP derivative according to the formula



**[0044]** wherein  $R_1$  and  $R_2$  are independently H or  $\text{NH}_2$ ,

**[0045]**  $R_3$  is



wherein Y is O or S, and W is O or NH, and

**[0046]**  $R_4$  and  $R_5$  are independently OH or F.

**[0047]** In particular embodiments, the 2'-ATP derivative is phosphorothioate 2'-ATP, imidophosphate 2'-ATP, fluorinate 2'-ATP, or a fluorescent 2'-ATP analog, see, e.g. FIG. 9C. See, e.g. Shafiee M, Gosselin G, Imbach J L, Eriksson S, Maury G. Synthesis of new fluorescent nucleoside analogues and application to the study of human deoxycytidine kinase, *Nucleosides Nucleotides*, 1999, vol. 18 (pg. 717-719); see also, Li Y F, Soni P B, Liu L F, Zhang X, Liotta D C, Lutz S. Synthesis of fluorescent nucleoside analogs as probes for 2'-deoxyribonucleoside kinases, *Bioorg. Med. Chem. Lett.*, 2010, vol. 20 (pg. 841-843), incorporated herein by reference.

**[0048]** In embodiments, the mimetic may comprise guanosine instead of adenosine, e.g. 2'-GTP. The molecule can be other 2' nucleoside triphosphates, which may have similar chemical properties.

**[0049]** Methods of synthesizing 2'-ATP, 3'-ATP, derivatives and mimetics thereof include utilization of chemical synthesis and enzymatic synthesis. In particular embodiments, the synthesis is chemical and the chemical synthesis can generally follow one of two reaction synthesis schemes, as detailed in FIG. 9A-9B. Preferred synthesis is according to a one-pot synthesis as detailed herein and depicted in FIG. 12B.

**[0050]** In an aspect, the method of synthesizing 2'-ATP comprises heating 5' protected acetyl-O-Adenosine and trimetaphosphate. Optimization of the heating times and temperature can be optimized, and may include checking for yield of the protected 5' product comprising the triphosphate group at the 2' location on the ribose and subsequently deprotecting the 5' product under acidic pH to generate 2'-ATP.

**[0051]** In an aspect, a method of synthesizing 2'-ATP is provided comprising reacting 2'-cAMP with Tris(tetra-n-butylammonium) Hydrogen Pyrophosphate,  $(\text{NBu}_4)_3\text{HP}_2\text{O}_7$ , and dimethylformamide (DMF) for about 2 to 6 hours, or about 3 to 5 hours. The next step is subsequently reacting with methanol (MeOH), water ( $\text{H}_2\text{O}$ ) and triethylamine ( $\text{Et}_3\text{N}$ ) at a ratio of about 7:3:1 to product 2'-ATP. In an aspect the step comprises reacting for about 10 to about 20 hours to produce 2'-ATP, adding the triphosphate functionality at the 2' carbon of the ribose sugar.

**[0052]** In embodiments, the enzymatic production comprises production of 2'-ATP, 3'-ATP, or derivatives or mimetics, as described herein.

#### STING Signaling

**[0053]** A wide range of hematopoietic and nonhematopoietic cells express STING. STING is an endoplasmic reticulum protein in mammalian cells that recruits and activates the cytosolic kinases IKK and TBK1, which activate the transcription factors NF- $\kappa$ B and IRF3, respectively. NF- $\kappa$ B and IRF3 then enter the nucleus and function together to induce IFNs and other cytokines. Cyclic-GMP-AMP (cGAMP) binds to and activates STING to trigger the downstream signaling cascades. Wang et al., (2013). The presence of DNA in the cytoplasm of mammalian cells is a danger signal that triggers the host immune responses such as the production of type-I interferons (IFN) through the production of cGAMP, which binds to and activates the adaptor protein STING. As described in Sun et al. (2013), a cGAMP synthase (cGAS) belongs to the nucleotidyltransferase family, and overexpression of cGAS activated the transcription factor IRF3 and induced IFN $\beta$  in a STING-dependent manner.

**[0054]** Methods of inhibition may further comprise detecting, prior to the administering step, the presence of innate immunity pathway activation with administration effected only if innate immunity pathway activation detected. Detecting innate immunity activation may comprise detecting mutations in one or more of TREX1, ADAR, STING, or other similar genes involved in innate immune signaling, and/or biochemical detection of STING activation.

**[0055]** STING is encoded by TMEM173 gene, with variants impacting human health. See, e.g. Patel et al., *Genes & Immunity* volume 20, pages 82-89(2019), incorporated by reference, with Table 2 somatic TMEM173 mutations in human cancer tissues in Table 2, incorporated by reference. STING signaling pathway has been characterized, including in Barber, *Nat Rev Immunol.* 2015; 15(12): 760-770, doi: 10.1038/nru3921, incorporated herein by reference in its entirety. TREX1 prevents immune activation by depleting damaged DNA, with TREX1 mutations are associated with an inflammatory and autoimmune diseases which are apparently independent such as Aicardi-Goutieres syndrome (AGS), systemic lupus erythematosus (SLE), familial chilblain lupus (FCL), cryofibrinogenemia, and retinal vasculopathy with cerebral leukodystrophy (RVCL). See, e.g.

Hosseini et al, *Genetics*, TREX1 Mutations can include D18N allele. See, e.g. Grieve et al., *PNAS* Apr. 21, 2015 112 (16) 5117-5122, doi: 10.1073/pnas.1423804112; Rice et al., *Neurology*, 12:12 1159-1169 (2013), doi: 10.1016/S1474-4422(13)70258-8; Li et al., *Nucleic Acids Research*, Volume 45, Issue 8, 5 May 2017, Pages 4619-4631, doi:10.1093/nar/gkx178. Similarly, ADAR mutations are implicated in Aicardi-Goutieres Syndrome. See, e.g. Fisher et al., *RNA Biol.* 2017; 14(2): 164-170; doi: 10.1080/15476286.2016.1267097. Further ADAR mutations have been identified. Hou et al., *Acta Dermato-Venereologica*, Volume 87, Number 1, January 2007, pp. 18-21(4); doi: 10.2340/00015555-0168; Savva et al., *Genome Biology* volume 13, Article number: 252 (2012); Schmelzer et al., *European Journal of Paediatric Neurology*, Volume 22, Issue 1, January 2018, pp. 186-189; doi:10.1016/j.ejpn.2017.11.003.

#### Therapeutic Agents

**[0056]** Agents for use in the methods disclosed herein for modulating activity of STING/IFN $\beta$  signaling are provided herein.

#### Protein Binding Agent

**[0057]** In embodiments, agents for use in the methods disclosed herein for modulating activity of STING/IFN $\beta$  signaling may comprise protein binding agents. In certain embodiments, an "agent" can refer to a protein-binding agent that permits modulation of activity of proteins or disrupts interactions of proteins and other biomolecules, such as but not limited to disrupting protein-protein interaction, ligand-receptor interaction, or protein-nucleic acid interaction. The terms "fragment," "derivative," and "analog" when referring to polypeptides as used herein refers to polypeptides which either retain substantially the same biological function or activity as such polypeptides. An analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide. Such agents include, but are not limited to, antibodies ("antibodies" includes antigen-binding portions of antibodies such as epitope- or antigen-binding peptides, paratopes, functional CDRs; recombinant antibodies; chimeric antibodies; humanized antibodies; nanobodies; tribodies; midibodies; or antigen-binding derivatives, analogs, variants, portions, or fragments thereof), protein-binding agents, nucleic acid molecules, small molecules, recombinant protein, peptides, aptamers, avimers and protein-binding derivatives, portions or fragments thereof.

**[0058]** In certain embodiments, the agent is capable of inhibiting or blocking STING signaling. Such agents may also be referred to as STING inhibitors or antagonists and can inhibit either the expression. In some embodiments, STING is suppressed by 2'-ATP or a derivative or mimetic thereof. In some embodiments, STING is inhibited, e.g., by a DNA targeting agent (e.g., CRISPR system, TALE, Zinc finger protein) or an RNA targeting agent (e.g., inhibitory nucleic acid molecules). In some embodiments, STING activity is inhibited. In certain embodiments, the antagonist is an antibody or fragment thereof. In certain embodiments, the antibody is specific for STING. In certain other example embodiments, the agent is 2'-ATP for use in inhibiting STING signaling.

#### Antibodies

**[0059]** In some embodiments, the agent may be an antibody or fragment thereof. The term "antibody" (e.g., anti-



STING antibody) is used interchangeably with the term “immunoglobulin” herein, and includes intact antibodies, fragments of antibodies, e.g., Fab, F(ab')<sub>2</sub> fragments, and intact antibodies and fragments that have been mutated either in their constant and/or variable region (e.g., mutations to produce chimeric, partially humanized, or fully humanized antibodies, as well as to produce antibodies with a desired trait, e.g., enhanced binding and/or reduced FcR binding). The term “fragment” refers to a part or portion of an antibody or antibody chain comprising fewer amino acid residues than an intact or complete antibody or antibody chain. Fragments can be obtained via chemical or enzymatic treatment of an intact or complete antibody or antibody chain. Fragments can also be obtained by recombinant means. Exemplary fragments include Fab, Fab', F(ab')<sub>2</sub>, Fabc, Fd, dAb, VHH and scFv and/or Fv fragments.

**[0060]** In some embodiments, the antibody is a humanized or chimeric antibody. “Humanized” forms of non-human (e.g., murine) antibodies are chimeric antibodies that contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a hypervariable region of the recipient are replaced by residues from a hypervariable region of a non-human species (donor antibody) such as mouse, rat, rabbit or nonhuman primate having the desired specificity, affinity, and capacity. In some instances, FR residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, humanized antibodies may comprise residues that are not found in the recipient antibody or in the donor antibody. These modifications are made to further refine antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the hypervariable regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin sequence. The humanized antibody optionally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin.

**[0061]** Antibodies may act as agonists or antagonists of the recognized polypeptides. For example, the present invention includes antibodies which disrupt receptor/ligand interactions either partially or fully. The invention features both receptor-specific antibodies and ligand-specific antibodies. The invention also features receptor-specific antibodies which do not prevent ligand binding but prevent receptor activation. Receptor activation (i.e., signaling) may be determined by techniques described herein or otherwise known in the art. For example, receptor activation can be determined by detecting the phosphorylation (e.g., tyrosine or serine/threonine) of the receptor or of one of its downstream substrates by immunoprecipitation followed by western blot analysis. In specific embodiments, antibodies are provided that inhibit ligand activity or receptor activity by at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 60%, or at least 50% of the activity in absence of the antibody.

**[0062]** The invention also features receptor-specific antibodies which both prevent ligand binding and receptor activation as well as antibodies that recognize the receptor-ligand complex. Likewise, encompassed by the invention are neutralizing antibodies which bind the ligand and prevent binding of the ligand to the receptor, as well as

antibodies which bind the ligand, thereby preventing receptor activation, but do not prevent the ligand from binding the receptor. Further included in the invention are antibodies which activate the receptor. These antibodies may act as receptor agonists, i.e., potentiate or activate either all or a subset of the biological activities of the ligand-mediated receptor activation, for example, by inducing dimerization of the receptor. The antibodies may be specified as agonists, antagonists or inverse agonists for biological activities comprising the specific biological activities of the peptides disclosed herein. The antibody agonists and antagonists can be made using methods known in the art. See, e.g., PCT publication WO 96/40281; U.S. Pat. No. 5,811,097; Deng et al., *Blood* 92(6):1981-1988 (1998); Chen et al., *Cancer Res.* 58(16):3668-3678 (1998); Harrop et al., *J. Immunol.* 161(4):1786-1794 (1998); Zhu et al., *Cancer Res.* 58(15):3209-3214 (1998); Yoon et al., *J. Immunol.* 160(7):3170-3179 (1998); Prat et al., *J. Cell. Sci.* III (Pt2):237-247 (1998); Pitard et al., *J. Immunol. Methods* 205(2):177-190 (1997); Liautard et al., *Cytokine* 9(4):233-241 (1997); Carlson et al., *J. Biol. Chem.* 272(17):11295-11301 (1997); Taryman et al., *Neuron* 14(4):755-762 (1995); Muller et al., *Structure* 6(9):1153-1167 (1998); Bartunek et al., *Cytokine* 8(1):14-20 (1996).

**[0063]** The antibodies as defined for the present invention include derivatives that are modified, i.e., by the covalent attachment of any type of molecule to the antibody such that covalent attachment does not prevent the antibody from generating an anti-idiotypic response. For example, but not by way of limitation, the antibody derivatives include antibodies that have been modified, e.g., by glycosylation, acetylation, pegylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other protein, etc. Any of numerous chemical modifications may be carried out by known techniques, including, but not limited to specific chemical cleavage, acetylation, formylation, metabolic synthesis of tunicamycin, etc. Additionally, the derivative may contain one or more non-classical amino acids.

#### Genetic Modifying Agents

**[0064]** In certain embodiments, the one or more modulating agents may be a genetic modifying agent. The genetic modifying agent may comprise a CRISPR system, a zinc finger nuclease system, a TALEN, a meganuclease or RNAi system. In some embodiments, a polynucleotide of the present invention described elsewhere herein can be modified using a genetic modifying agent (e.g., one or more genes are selected from TREX1, ADAR, STING).

#### CRISPR-Cas Modification

**[0065]** In some embodiments, a polynucleotide of the present invention described elsewhere herein can be modified using a CRISPR-Cas and/or Cas-based system.

**[0066]** In general, a CRISPR-Cas or CRISPR system as used in herein and in documents, such as International Patent Publication No. WO 2014/093622 (PCT/US2013/074667), refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated (“Cas”) genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a “direct repeat” and a tracrRNA-

processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a “spacer” in the context of an endogenous CRISPR system), or “RNA(s)” as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). See, e.g., Shmakov et al. (2015) “Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems”, *Molecular Cell*, DOI: [dx.doi.org/10.1016/j.molcel.2015.10.008](https://doi.org/10.1016/j.molcel.2015.10.008).

**[0067]** CRISPR-Cas systems can generally fall into two classes based on their architectures of their effector molecules, which are each further subdivided by type and subtype. The two classes are Class 1 and Class 2. Class 1 CRISPR-Cas systems have effector modules composed of multiple Cas proteins, some of which form crRNA-binding complexes, while Class 2 CRISPR-Cas systems include a single, multi-domain crRNA-binding protein.

**[0068]** In some embodiments, the CRISPR-Cas system that can be used to modify a polynucleotide of the present invention described herein can be a Class 1 CRISPR-Cas system. In some embodiments, the CRISPR-Cas system that can be used to modify a polynucleotide of the present invention described herein can be a Class 2 CRISPR-Cas system.

#### Class 1 CRISPR-Cas Systems

**[0069]** In some embodiments, the CRISPR-Cas system that can be used to modify a polynucleotide of the present invention described herein can be a Class 1 CRISPR-Cas system. Class 1 CRISPR-Cas systems are divided into types I, II, and IV. Makarova et al. 2020. *Nat. Rev. 18*: 67-83, particularly as described in FIG. 1. Type I CRISPR-Cas systems are divided into 9 subtypes (I-A, I-B, I-C, I-D, I-E, I-F1, I-F2, I-F3, and I-G). Makarova et al., 2020. Class 1, Type I CRISPR-Cas systems can contain a Cas3 protein that can have helicase activity. Type III CRISPR-Cas systems are divided into 6 subtypes (III-A, III-B, III-E, and III-F). Type III CRISPR-Cas systems can contain a Cas10 that can include an RNA recognition motif called Palm and a cyclase domain that can cleave polynucleotides. Makarova et al., 2020. Type IV CRISPR-Cas systems are divided into 3 subtypes. (IV-A, IV-B, and IV-C). Makarova et al., 2020. Class 1 systems also include CRISPR-Cas variants, including Type I-A, I-B, I-E, I-F and I-U variants, which can include variants carried by transposons and plasmids, including versions of subtype I-F encoded by a large family of Tn7-like transposon and smaller groups of Tn7-like transposons that encode similarly degraded subtype I-B systems. Peters et al., *PNAS* 114 (35) (2017); DOI: [10.1073/pnas.1709035114](https://doi.org/10.1073/pnas.1709035114); see also, Makarova et al. 2018. *The CRISPR Journal*, v. 1, n5, FIG. 5.

**[0070]** The Class 1 systems typically comprise a multi-protein effector complex, which can, in some embodiments, include ancillary proteins, such as one or more proteins in a complex referred to as a CRISPR-associated complex for antiviral defense (Cascade), one or more adaptation proteins (e.g., Cas1, Cas2, RNA nuclease), and/or one or more accessory proteins (e.g., Cas 4, DNA nuclease), CRISPR

associated Rossmann fold (CARF) domain containing proteins, and/or RNA transcriptase.

**[0071]** The backbone of the Class 1 CRISPR-Cas system effector complexes can be formed by RNA recognition motif domain-containing protein(s) of the repeat-associated mysterious proteins (RAMPs) family subunits (e.g., Cas 5, Cas6, and/or Cas7). RAMP proteins are characterized by having one or more RNA recognition motif domains. In some embodiments, multiple copies of RAMPs can be present. In some embodiments, the Class I CRISPR-Cas system can include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more Cas5, Cas6, and/or Cas 7 proteins. In some embodiments, the Cas6 protein is an RNase, which can be responsible for pre-crRNA processing. When present in a Class 1 CRISPR-Cas system, Cas6 can be optionally physically associated with the effector complex.

**[0072]** Class 1 CRISPR-Cas system effector complexes can, in some embodiments, also include a large subunit. The large subunit can be composed of or include a Cas8 and/or Cas10 protein. See, e.g., FIGS. 1 and 2. Koonin E V, Makarova K S. 2019. *Phil. Trans. R. Soc. B* 374: 20180087, DOI: [10.1098/rstb.2018.0087](https://doi.org/10.1098/rstb.2018.0087) and Makarova et al. 2020.

**[0073]** Class 1 CRISPR-Cas system effector complexes can, in some embodiments, include a small subunit (for example, Cas11). See, e.g., FIGS. 1 and 2. Koonin E V, Makarova K S. 2019. *Origins and Evolution of CRISPR-Cas systems*. *Phil. Trans. R. Soc. B* 374: 20180087, DOI: [10.1098/rstb.2018.0087](https://doi.org/10.1098/rstb.2018.0087).

**[0074]** In some embodiments, the Class 1 CRISPR-Cas system can be a Type I CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-A CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-B CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-C CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-D CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-E CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-F1 CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-F2 CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-F3 CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a subtype I-G CRISPR-Cas system. In some embodiments, the Type I CRISPR-Cas system can be a CRISPR Cas variant, such as a Type I-A, I-B, I-E, I-F and I-U variants, which can include variants carried by transposons and plasmids, including versions of subtype I-F encoded by a large family of Tn7-like transposon and smaller groups of Tn7-like transposons that encode similarly degraded subtype I-B systems as previously described.

**[0075]** In some embodiments, the Class 1 CRISPR-Cas system can be a Type III CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-A CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-B CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-C CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-D CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a

subtype III-E CRISPR-Cas system. In some embodiments, the Type III CRISPR-Cas system can be a subtype III-F CRISPR-Cas system.

**[0076]** In some embodiments, the Class 1 CRISPR-Cas system can be a Type IV CRISPR-Cas system. In some embodiments, the Type IV CRISPR-Cas system can be a subtype IV-A CRISPR-Cas system. In some embodiments, the Type IV CRISPR-Cas system can be a subtype IV-B CRISPR-Cas system. In some embodiments, the Type IV CRISPR-Cas system can be a subtype IV-C CRISPR-Cas system.

**[0077]** The effector complex of a Class 1 CRISPR-Cas system can, in some embodiments, include a Cas3 protein that is optionally fused to a Cas2 protein, a Cas4, a Cas5, a Cas6, a Cas7, a Cas8, a Cas10, a Cas11, or a combination thereof. In some embodiments, the effector complex of a Class 1 CRISPR-Cas system can have multiple copies, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14, of any one or more Cas proteins.

#### Class 2 CRISPR-Cas Systems

**[0078]** The compositions, systems, and methods described in greater detail elsewhere herein can be designed and adapted for use with Class 2 CRISPR-Cas systems. Thus, in some embodiments, the CRISPR-Cas system is a Class 2 CRISPR-Cas system. Class 2 systems are distinguished from Class 1 systems in that they have a single, large, multi-domain effector protein. In certain example embodiments, the Class 2 system can be a Type II, Type V, or Type VI system, which are described in Makarova et al. "Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants" *Nature Reviews Microbiology*, 18:67-81 (February 2020), incorporated herein by reference. Each type of Class 2 system is further divided into subtypes. See Markova et al. 2020, particularly at Figure. 2. Class 2, Type II systems can be divided into 4 subtypes: II-A, II-B, II-C1, and II-C2. Class 2, Type V systems can be divided into 17 subtypes: V-A, V-B1, V-B2, V-C, V-D, V-E, V-F1, V-F1(V-U3), V-F2, V-F3, V-G, V-H, V-I, V-K (V-U5), V-U1, V-U2, and V-U4. Class 2, Type IV systems can be divided into 5 subtypes: VI-A, VI-B1, VI-B2, VI-C, and VI-D.

**[0079]** The distinguishing feature of these types is that their effector complexes consist of a single, large, multi-domain protein. Type V systems differ from Type II effectors (e.g., Cas9), which contain two nuclear domains that are each responsible for the cleavage of one strand of the target DNA, with the HNH nuclease inserted inside the Ruv-C like nuclease domain sequence. The Type V systems (e.g., Cas12) only contain a RuvC-like nuclease domain that cleaves both strands. Type VI (Cas13) are unrelated to the effectors of Type II and V systems and contain two HEPN domains and target RNA. Cas13 proteins also display collateral activity that is triggered by target recognition. Some Type V systems have also been found to possess this collateral activity with two single-stranded DNA in *in vitro* contexts.

**[0080]** In some embodiments, the Class 2 system is a Type II system. In some embodiments, the Type II CRISPR-Cas system is a II-A CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-B CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-C1 CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-C2 CRISPR-

Cas system. In some embodiments, the Type II system is a Cas9 system. In some embodiments, the Type II system includes a Cas9.

**[0081]** In some embodiments, the Class 2 system is a Type V system. In some embodiments, the Type V CRISPR-Cas system is a V-A CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-B1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-B2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-C CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-D CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-E CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F1 (V-U3) CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F3 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-G CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-H CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-I CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-K (V-U5) CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U4 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system includes a Cas12a (Cpf1), Cas12b (C2c1), Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), and/or Cas14.

**[0082]** In some embodiments the Class 2 system is a Type VI system. In some embodiments, the Type VI CRISPR-Cas system is a VI-A CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-B1 CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-B2 CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-C CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-D CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system includes a Cas13a (C2c2), Cas13b (Group 29/30), Cas13c, and/or Cas13d.

#### Specialized Cas-Based Systems

**[0083]** In some embodiments, the system is a Cas-based system that is capable of performing a specialized function or activity. For example, the Cas protein may be fused, operably coupled to, or otherwise associated with one or more functional domains. In certain example embodiments, the Cas protein may be a catalytically dead Cas protein ("dCas") and/or have nickase activity. A nickase is a Cas protein that cuts only one strand of a double stranded target. In such embodiments, the dCas or nickase provide a sequence specific targeting functionality that delivers the functional domain to or proximate a target sequence. Example functional domains that may be fused to, operably coupled to, or otherwise associated with a Cas protein can be or include, but are not limited to a nuclear localization signal (NLS) domain, a nuclear export signal (NES) domain, a translational activation domain, a transcriptional activation domain (e.g. VP64, p65, MyoD1, HSF1, RTA, and SETT/9),

a translation initiation domain, a transcriptional repression domain (e.g., a KRAB domain, NuE domain, NcoR domain, and a SID domain such as a SID4X domain), a nuclease domain (e.g., FokI), a histone modification domain (e.g., a histone acetyltransferase), a light inducible/controllable domain, a chemically inducible/controllable domain, a transposase domain, a homologous recombination machinery domain, a recombinase domain, an integrase domain, and combinations thereof. Methods for generating catalytically dead Cas9 or a nickase Cas9 (WO 2014/204725, Ran et al. Cell. 2013 Sep. 12; 154(6):1380-1389), Cas12 (Liu et al. Nature Communications, 8, 2095 (2017), and Cas13 (International Patent Publication Nos. WO 2019/005884 and WO2019/060746) are known in the art and incorporated herein by reference.

**[0084]** In some embodiments, the functional domains can have one or more of the following activities: methylase activity, demethylase activity, translation activation activity, translation initiation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity, molecular switch activity, chemical inducibility, light inducibility, and nucleic acid binding activity. In some embodiments, the one or more functional domains may comprise epitope tags or reporters. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporters include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and auto-fluorescent proteins including blue fluorescent protein (BFP).

**[0085]** The one or more functional domain(s) may be positioned at, near, and/or in proximity to a terminus of the effector protein (e.g., a Cas protein). In embodiments having two or more functional domains, each of the two can be positioned at or near or in proximity to a terminus of the effector protein (e.g., a Cas protein). In some embodiments, such as those where the functional domain is operably coupled to the effector protein, the one or more functional domains can be tethered or linked via a suitable linker (including, but not limited to, GlySer linkers) to the effector protein (e.g., a Cas protein). When there is more than one functional domain, the functional domains can be same or different. In some embodiments, all the functional domains are the same. In some embodiments, all of the functional domains are different from each other. In some embodiments, at least two of the functional domains are different from each other. In some embodiments, at least two of the functional domains are the same as each other.

**[0086]** Other suitable functional domains can be found, for example, in International Patent Publication No. WO 2019/018423.

#### Split CRISPR-Cas Systems

**[0087]** In some embodiments, the CRISPR-Cas system is a split CRISPR-Cas system. See e.g., Zetche et al., 2015. Nat. Biotechnol. 33(2): 139-142 and International Patent

Publication WO 2019/018423, the compositions and techniques of which can be used in and/or adapted for use with the present invention. Split CRISPR-Cas proteins are set forth herein and in documents incorporated herein by reference in further detail herein. In certain embodiments, each part of a split CRISPR protein are attached to a member of a specific binding pair, and when bound with each other, the members of the specific binding pair maintain the parts of the CRISPR protein in proximity. In certain embodiments, each part of a split CRISPR protein is associated with an inducible binding pair. An inducible binding pair is one which is capable of being switched “on” or “off” by a protein or small molecule that binds to both members of the inducible binding pair. In some embodiments, CRISPR proteins may preferably split between domains, leaving domains intact. In particular embodiments, said Cas split domains (e.g., RuvC and HNH domains in the case of Cas9) can be simultaneously or sequentially introduced into the cell such that said split Cas domain(s) process the target nucleic acid sequence in the algae cell. The reduced size of the split Cas compared to the wild type Cas allows other methods of delivery of the systems to the cells, such as the use of cell penetrating peptides as described herein.

#### DNA and RNA Base Editing

**[0088]** In some embodiments, a polynucleotide of the present invention described elsewhere herein can be modified using a base editing system. In some embodiments, a Cas protein is connected or fused to a nucleotide deaminase. Thus, in some embodiments the Cas-based system can be a base editing system. As used herein, “base editing” refers generally to the process of polynucleotide modification via a CRISPR-Cas-based or Cas-based system that does not include excising nucleotides to make the modification. Base editing can convert base pairs at precise locations without generating excess undesired editing byproducts that can be made using traditional CRISPR-Cas systems.

**[0089]** In certain example embodiments, the nucleotide deaminase may be a DNA base editor used in combination with a DNA binding Cas protein such as, but not limited to, Class 2 Type II and Type V systems. Two classes of DNA base editors are generally known: cytosine base editors (CBEs) and adenine base editors (ABEs). CBEs convert a C•G base pair into a T•A base pair (Komor et al. 2016. Nature. 533:420-424; Nishida et al. 2016. Science. 353; and Li et al. Nat. Biotech. 36:324-327) and ABEs convert an A•T base pair to a G•C base pair. Collectively, CBEs and ABEs can mediate all four possible transition mutations (C to T, A to G, T to C, and G to A). Rees and Liu. 2018. Nat. Rev. Genet. 19(12): 770-788, particularly at FIGS. 1b, 2a-2c, 3a-3f, and Table 1. In some embodiments, the base editing system includes a CBE and/or an ABE. In some embodiments, a polynucleotide of the present invention described elsewhere herein can be modified using a base editing system. Rees and Liu. 2018. Nat. Rev. Genet. 19(12):770-788. Base editors also generally do not need a DNA donor template and/or rely on homology-directed repair. Komor et al. 2016. Nature. 533:420-424; Nishida et al. 2016. Science. 353; and Gaudeli et al. 2017. Nature. 551:464-471. Upon binding to a target locus in the DNA, base pairing between the guide RNA of the system and the target DNA strand leads to displacement of a small segment of ssDNA in an “R-loop”. Nishimasu et al. Cell. 156:935-949. DNA bases within the ssDNA bubble are modified by the enzyme

component, such as a deaminase. In some systems, the catalytically disabled Cas protein can be a variant or modified Cas can have nickase functionality and can generate a nick in the non-edited DNA strand to induce cells to repair the non-edited strand using the edited strand as a template. Komor et al. 2016. *Nature*. 533:420-424; Nishida et al. 2016. *Science*. 353; and Gaudeli et al. 2017. *Nature*. 551:464-471. **[0090]** Other Example Type V base editing systems are described in International Patent Publication Nos. WO 2018/213708, WO 2018/213726, and International Patent Applications No. PCT/US2018/067207, PCT/US2018/067225, and PCT/US2018/067307, each of which is incorporated herein by reference.

**[0091]** In certain example embodiments, the base editing system may be an RNA base editing system. As with DNA base editors, a nucleotide deaminase capable of converting nucleotide bases may be fused to a Cas protein. However, in these embodiments, the Cas protein will need to be capable of binding RNA. Example RNA binding Cas proteins include, but are not limited to, RNA-binding Cas9s such as *Francisella novicida* Cas9 (“FnCas9”), and Class 2 Type VI Cas systems. The nucleotide deaminase may be a cytidine deaminase or an adenosine deaminase, or an adenosine deaminase engineered to have cytidine deaminase activity. In certain example embodiments, the RNA base editor may be used to delete or introduce a post-translation modification site in the expressed mRNA. In contrast to DNA base editors, whose edits are permanent in the modified cell, RNA base editors can provide edits where finer, temporal control may be needed, for example in modulating a particular immune response. Example Type VI RNA-base editing systems are described in Cox et al. 2017. *Science* 358: 1019-1027, International Patent Publication Nos. WO 2019/005884, WO 2019/005886, and WO 2019/071048, and International Patent Application Nos. PCT/US20018/05179 and PCT/US2018/067207, which are incorporated herein by reference. An example FnCas9 system that may be adapted for RNA base editing purposes is described in International Patent Publication No. WO 2016/106236, which is incorporated herein by reference.

**[0092]** An example method for delivery of base-editing systems, including use of a split-intein approach to divide CBE and ABE into reconstitutable halves, is described in Levy et al. *Nature Biomedical Engineering* doi.org/10.1038/s41441-019-0505-5 (2019), which is incorporated herein by reference.

#### Prime Editors

**[0093]** In some embodiments, a polynucleotide of the present invention described elsewhere herein can be modified using a prime editing system. See e.g. Anzalone et al. 2019. *Nature*. 576: 149-157. Like base editing systems, prime editing systems can be capable of targeted modification of a polynucleotide without generating double stranded breaks and does not require donor templates. Further prime editing systems can be capable of all 12 possible combination swaps. Prime editing can operate via a “search-and-replace” methodology and can mediate targeted insertions, deletions, all 12 possible base-to-base conversion and combinations thereof. Generally, a prime editing system, as exemplified by PE1, PE2, and PE3 (Id.), can include a reverse transcriptase fused or otherwise coupled or associated with an RNA-programmable nickase and a prime-editing extended guide RNA (pegRNA) to facilitate direct

copying of genetic information from the extension on the pegRNA into the target polynucleotide. Embodiments that can be used with the present invention include these and variants thereof. Prime editing can have the advantage of lower off-target activity than traditional CRISPR-Cas systems along with few byproducts and greater or similar efficiency as compared to traditional CRISPR-Cas systems.

**[0094]** In some embodiments, the prime editing guide molecule can specify both the target polynucleotide information (e.g., sequence) and contain a new polynucleotide cargo that replaces target polynucleotides. To initiate transfer from the guide molecule to the target polynucleotide, the PE system can nick the target polynucleotide at a target site to expose a 3' hydroxyl group, which can prime reverse transcription of an edit-encoding extension region of the guide molecule (e.g. a prime editing guide molecule or peg guide molecule) directly into the target site in the target polynucleotide. See e.g. Anzalone et al. 2019. *Nature*. 576: 149-157, particularly at FIGS. 1b, 1c, related discussion, and Supplementary discussion.

**[0095]** In some embodiments, a prime editing system can be composed of a Cas polypeptide having nickase activity, a reverse transcriptase, and a guide molecule. The Cas polypeptide can lack nuclease activity. The guide molecule can include a target binding sequence as well as a primer binding sequence and a template containing the edited polynucleotide sequence. The guide molecule, Cas polypeptide, and/or reverse transcriptase can be coupled together or otherwise associate with each other to form an effector complex and edit a target sequence. In some embodiments, the Cas polypeptide is a Class 2, Type V Cas polypeptide. In some embodiments, the Cas polypeptide is a Cas9 polypeptide (e.g. is a Cas9 nickase). In some embodiments, the Cas polypeptide is fused to the reverse transcriptase. In some embodiments, the Cas polypeptide is linked to the reverse transcriptase.

**[0096]** In some embodiments, the prime editing system can be a PE1 system or variant thereof, a PE2 system or variant thereof, or a PE3 (e.g. PE3, PE3b) system. See e.g., Anzalone et al. 2019. *Nature*. 576: 149-157, particularly at pgs. 2-3, FIGS. 2a, 3a-3f, 4a-4b, Extended data FIGS. 3a-3b, 4,

**[0097]** The peg guide molecule can be about 10 to about 200 or more nucleotides in length, such as 10 to/or 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, or 200 or more nucleotides in length. Optimization of the peg guide molecule can be accomplished as described in Anzalone et al. 2019. *Nature*. 576: 149-157, particularly at pg. 3, FIG. 2a-2b, and Extended Data FIGS. 5a-c.

### CRISPR Associated Transposase (CAST) Systems

**[0098]** In some embodiments, a polynucleotide of the present invention described elsewhere herein can be modified using a CRISPR Associated Transposase (“CAST”) system. CAST system can include a Cas protein that is catalytically inactive, or engineered to be catalytically active, and further comprises a transposase (or subunits thereof) that catalyze RNA-guided DNA transposition. Such systems are able to insert DNA sequences at a target site in a DNA molecule without relying on host cell repair machinery. CAST systems can be Class 1 or Class 2 CAST systems. An example Class 1 system is described in Klompe et al. Nature, doi:10.1038/s41586-019-1323, which is incorporated herein by reference. An example Class 2 system is described in Strecker et al. Science. 10/1126/science.aax9181 (2019), and PCT/US2019/066835 which are incorporated herein by reference.

### Guide Molecules

**[0099]** The CRISPR-Cas or Cas-Based system described herein can, in some embodiments, include one or more guide molecules. The terms guide molecule, guide sequence and guide polynucleotide refer to polynucleotides capable of guiding Cas to a target genomic locus and are used interchangeably as in foregoing cited documents such as International Patent Publication No. WO 2014/093622 (PCT/US2013/074667). In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. The guide molecule can be a polynucleotide.

**[0100]** The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay (Qui et al. 2004. BioTechniques. 36(4)702-707). Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible and will occur to those skilled in the art.

**[0101]** In some embodiments, the guide molecule is an RNA. The guide molecule(s) (also referred to interchangeably herein as guide polynucleotide and guide sequence) that are included in the CRISPR-Cas or Cas based system can be any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of

complementarity, when optimally aligned using a suitable alignment algorithm, can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting examples of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at [www.novocraft.com](http://www.novocraft.com)), ELAND (Illumina, San Diego, CA), SOAP (available at [soap.genomics.org.cn](http://soap.genomics.org.cn)), and Maq (available at [maq.sourceforge.net](http://maq.sourceforge.net)).

**[0102]** A guide sequence, and hence a nucleic acid-targeting guide, may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

**[0103]** In some embodiments, a nucleic acid-targeting guide is selected to reduce the degree secondary structure within the nucleic acid-targeting guide. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A. R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and G M Church, 2009, Nature Biotechnology 27(12): 1151-62).

**[0104]** In certain embodiments, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In certain embodiments, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

**[0105]** In certain embodiments, the crRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem loop.

**[0106]** In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27 to 30 nt, e.g., 27, 28, 29, or 30 nt, from 30 to 35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

**[0107]** The “tracrRNA” sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, the degree of complementarity between the tracrRNA sequence and crRNA sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and crRNA sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin.

**[0108]** In general, degree of complementarity is with reference to the optimal alignment of the sca sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm and may further account for secondary structures, such as self-complementarity within either the sca sequence or tracr sequence. In some embodiments, the degree of complementarity between the tracr sequence and sca sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher.

**[0109]** In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and tracr RNA can be 30 or 50 nucleotides in length. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it being advantageous that off target is 100% or 99.9% or 99.5% or 99% or

99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

**[0110]** In some embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e., an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence. Where the tracr RNA is on a different RNA than the RNA containing the guide and tracr sequence, the length of each RNA may be optimized to be shortened from their respective native lengths, and each may be independently chemically modified to protect from degradation by cellular RNase or otherwise increase stability.

**[0111]** Many modifications to guide sequences are known in the art and are further contemplated within the context of this invention. Various modifications may be used to increase the specificity of binding to the target sequence and/or increase the activity of the Cas protein and/or reduce off-target effects. Example guide sequence modifications are described in International Patent Application No. PCT US2019/045582, specifically paragraphs [0178]-[0333], which is incorporated herein by reference.

Target Sequences, PAMs, and PFSs

Target Sequences

**[0112]** In the context of formation of a CRISPR complex, “target sequence” refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise RNA polynucleotides. The term “target RNA” refers to an RNA polynucleotide being or comprising the target sequence. In other words, the target polynucleotide can be a polynucleotide or a part of a polynucleotide to which a part of the guide sequence is designed to have complementarity with and to which the effector function mediated by the complex comprising the CRISPR effector protein and a guide molecule is to be directed. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

**[0113]** The guide sequence can specifically bind a target sequence in a target polynucleotide. The target polynucleotide may be DNA. The target polynucleotide may be RNA. The target polynucleotide can have one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc. or more) target sequences. The target polynucleotide can be on a vector. The target polynucleotide can be genomic DNA. The target polynucleotide can be episomal. Other forms of the target polynucleotide are described elsewhere herein.

**[0114]** The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-

coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmic RNA (scRNA). In some preferred embodiments, the target sequence (also referred to herein as a target polynucleotide) may be a sequence within an RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

#### PAM and PFS Elements

**[0115]** PAM elements are sequences that can be recognized and bound by Cas proteins. Cas proteins/effector complexes can then unwind the dsDNA at a position adjacent to the PAM element. It will be appreciated that Cas proteins and systems that include them that target RNA do not require PAM sequences (Marraffini et al. 2010. *Nature*. 463:568-571). Instead, many rely on PFSs, which are discussed elsewhere herein. In certain embodiments, the target sequence should be associated with a PAM (protospacer adjacent motif) or PFS (protospacer flanking sequence or site), that is, a short sequence recognized by the CRISPR complex. Depending on the nature of the CRISPR-Cas protein, the target sequence should be selected, such that its complementary sequence in the DNA duplex (also referred to herein as the non-target sequence) is upstream or downstream of the PAM. In the embodiments, the complementary sequence of the target sequence is downstream or 3' of the PAM or upstream or 5' of the PAM. The precise sequence and length requirements for the PAM differ depending on the Cas protein used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). Examples of the natural PAM sequences for different Cas proteins are provided herein below and the skilled person will be able to identify further PAM sequences for use with a given Cas protein.

**[0116]** The ability to recognize different PAM sequences depends on the Cas polypeptide(s) included in the system. See e.g., Gleditsch et al. 2019. *RNA Biology*. 16(4):504-517. Table 3 (from Gleditsch et al. 2019) below shows several Cas polypeptides and the PAM sequence they recognize.

TABLE 3

Example PAM Sequences	
Cas Protein	PAM Sequence
SpCas9	NGG/NRG
SaCas9	NGRRT or NGRRN
NmeCas9	NNNGATT
CjCas9	NNNNRYAC
StCas9	NNAGAAW
Cas12a (Cpf1) (including LbCpf1 and AsCpf1)	TTTV
Cas12b (C2c1)	TTT, TTA, and TTC
Cas12c (C2c3)	TA

TABLE 3-continued

Example PAM Sequences	
Cas Protein	PAM Sequence
Cas12d (CasY)	TA
Cas12e (CasX)	5'-TTCN-3'

**[0117]** In a preferred embodiment, the CRISPR effector protein may recognize a 3' PAM. In certain embodiments, the CRISPR effector protein may recognize a 3' PAM which is 5'H, wherein H is A, C or U.

**[0118]** Further, engineering of the PAM Interacting (PI) domain on the Cas protein may allow programming of PAM specificity, improve target site recognition fidelity, and increase the versatility of the CRISPR-Cas protein, for example as described for Cas9 in Kleinstiver B P et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. 2015 Jul. 23; 523(7561):481-5. doi: 10.1038/nature14592. As further detailed herein, the skilled person will understand that Cas13 proteins may be modified analogously. Gao et al, "Engineered Cpf1 Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: http://dx.doi.org/10.1101/091611 (Dec. 4, 2016). Doench et al. created a pool of sgRNAs, tiling across all possible target sites of a panel of six endogenous mouse and three endogenous human genes and quantitatively assessed their ability to produce null alleles of their target gene by antibody staining and flow cytometry. The authors showed that optimization of the PAM improved activity and also provided an on-line tool for designing sgRNAs.

**[0119]** PAM sequences can be identified in a polynucleotide using an appropriate design tool, which are commercially available as well as online. Such freely available tools include, but are not limited to, CRISPRFinder and CRISPR-RTarget. Mojica et al. 2009. *Microbiol.* 155(Pt. 3):733-740; Atschul et al. 1990. *J. Mol. Biol.* 215:403-410; Biswass et al. 2013 *RNA Biol.* 10:817-827; and Grissa et al. 2007. *Nucleic Acid Res.* 35:W52-57. Experimental approaches to PAM identification can include, but are not limited to, plasmid depletion assays (Jiang et al. 2013. *Nat. Biotechnol.* 31:233-239; Esvelt et al. 2013. *Nat. Methods.* 10:1116-1121; Kleinstiver et al. 2015. *Nature.* 523:481-485), screened by a high-throughput in vivo model called PAM-SCNAR (Patanayak et al. 2013. *Nat. Biotechnol.* 31:839-843 and Leenay et al. 2016. *Mol. Cell.* 16:253), and negative screening (Zetsche et al. 2015. *Cell.* 163:759-771).

**[0120]** As previously mentioned, CRISPR-Cas systems that target RNA do not typically rely on PAM sequences. Instead such systems typically recognize protospacer flanking sites (PFSs) instead of PAMs. Thus, Type VI CRISPR-Cas systems typically recognize protospacer flanking sites (PFSs) instead of PAMs. PFSs represents an analogue to PAMs for RNA targets. Type VI CRISPR-Cas systems employ a Cas13. Some Cas13 proteins analyzed to date, such as Cas13a (C2c2) identified from *Leptotrichia shahii* (LSHCas13a) have a specific discrimination against G at the 3' end of the target RNA. The presence of a C at the corresponding crRNA repeat site can indicate that nucleotide pairing at this position is rejected. However, some Cas13



proteins (e.g., LwaCas13a and PspCas13b) do not seem to have a PFS preference. See e.g., Gleditsch et al. 2019. *RNA Biology*. 16(4):504-517.

**[0121]** Some Type VI proteins, such as subtype B, have 5'-recognition of D (G, T, A) and a 3'-motif requirement of NAN or NNA. One example is the Cas13b protein identified in *Bergeyella zoohelcum* (BzCas13b). See e.g., Gleditsch et al. 2019. *RNA Biology*. 16(4):504-517.

**[0122]** Overall Type VI CRISPR-Cas systems appear to have less restrictive rules for substrate (e.g., target sequence) recognition than those that target DNA (e.g., Type V and type II).

#### Zinc Finger Nucleases

**[0123]** In some embodiments, the polynucleotide is modified using a Zinc Finger nuclease or system thereof. One type of programmable DNA-binding domain is provided by artificial zinc-finger (ZF) technology, which involves arrays of ZF modules to target new DNA-binding sites in the genome. Each finger module in a ZF array targets three DNA bases. A customized array of individual zinc finger domains is assembled into a ZF protein (ZFP).

**[0124]** ZFPs can comprise a functional domain. The first synthetic zinc finger nucleases (ZFNs) were developed by fusing a ZF protein to the catalytic domain of the Type IIS restriction enzyme FokI. (Kim, Y. G. et al., 1994, Chimeric restriction endonuclease, *Proc. Natl. Acad. Sci. U.S.A.* 91, 883-887; Kim, Y. G. et al., 1996, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.* 93, 1156-1160). Increased cleavage specificity can be attained with decreased off target activity by use of paired ZFN heterodimers, each targeting different nucleotide sequences separated by a short spacer. (Doyon, Y. et al., 2011, Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods* 8, 74-79). ZFPs can also be designed as transcription activators and repressors and have been used to target many genes in a wide variety of organisms. Exemplary methods of genome editing using ZFNs can be found for example in U.S. Pat. Nos. 6,534,261, 6,607,882, 6,746,838, 6,794,136, 6,824,978, 6,866,997, 6,933,113, 6,979,539, 7,013,219, 7,030,215, 7,220,719, 7,241,573, 7,241,574, 7,585,849, 7,595,376, 6,903,185, and 6,479,626, all of which are specifically incorporated by reference.

#### Sequences Related to Nucleus Targeting and Transportation

**[0125]** In some embodiments, one or more components (e.g., the Cas protein and/or deaminase) in the composition for engineering cells may comprise one or more sequences related to nucleus targeting and transportation. Such sequence may facilitate the one or more components in the composition for targeting a sequence within a cell. In order to improve targeting of the CRISPR-Cas protein and/or the nucleotide deaminase protein or catalytic domain thereof used in the methods of the present disclosure to the nucleus, it may be advantageous to provide one or both of these components with one or more nuclear localization sequences (NLSs).

**[0126]** In some embodiments, the NLSs used in the context of the present disclosure are heterologous to the proteins. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV

(SEQ ID NO:1) or PKKKRKVEAS (SEQ ID NO:2); the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID NO:3)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO:4) or RQRRNELKRSP (SEQ ID NO:5); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPGYGGGGQYFAK-PRNQGGY (SEQ ID NO: 6); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKD-EQILKRRNV (SEQ ID NO:7) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO:8) and PPKKARED (SEQ ID NO:9) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO:10) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO:11) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO:12) and PKQKKRK (SEQ ID NO:13) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO:14) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO:15) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO:16) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO:17) of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the DNA-targeting Cas protein in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the CRISPR-Cas protein, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the nucleic acid-targeting protein, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of nucleic acid-targeting complex formation (e.g., assay for deaminase activity) at the target sequence, or assay for altered gene expression activity affected by DNA-targeting complex formation and/or DNA-targeting), as compared to a control not exposed to the CRISPR-Cas protein and deaminase protein, or exposed to a CRISPR-Cas and/or deaminase protein lacking the one or more NLSs.

**[0127]** The CRISPR-Cas and/or nucleotide deaminase proteins may be provided with 1 or more, such as with, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous NLSs. In some embodiments, the proteins comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g., zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypep-

tide chain from the N- or C-terminus. In preferred embodiments of the CRISPR-Cas proteins, an NLS attached to the C-terminal of the protein.

**[0128]** In certain embodiments, the CRISPR-Cas protein and the deaminase protein are delivered to the cell or expressed within the cell as separate proteins. In these embodiments, each of the CRISPR-Cas and deaminase protein can be provided with one or more NLSs as described herein. In certain embodiments, the CRISPR-Cas and deaminase proteins are delivered to the cell or expressed with the cell as a fusion protein. In these embodiments one or both of the CRISPR-Cas and deaminase protein is provided with one or more NLSs. Where the nucleotide deaminase is fused to an adaptor protein (such as MS2) as described above, the one or more NLS can be provided on the adaptor protein, provided that this does not interfere with aptamer binding. In particular embodiments, the one or more NLS sequences may also function as linker sequences between the nucleotide deaminase and the CRISPR-Cas protein.

**[0129]** In certain embodiments, guides of the disclosure comprise specific binding sites (e.g. aptamers) for adapter proteins, which may be linked to or fused to a nucleotide deaminase or catalytic domain thereof. When such a guide forms a CRISPR complex (e.g., CRISPR-Cas protein binding to guide and target), the adapter proteins bind and the nucleotide deaminase or catalytic domain thereof associated with the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective.

**[0130]** The skilled person will understand that modifications to the guide which allow for binding of the adapter+nucleotide deaminase, but not proper positioning of the adapter+nucleotide deaminase (e.g. due to steric hindrance within the three-dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and in some cases at both the tetra loop and stem loop 2.

**[0131]** In some embodiments, a component (e.g., the dead Cas protein, the nucleotide deaminase protein or catalytic domain thereof, or a combination thereof) in the systems may comprise one or more nuclear export signals (NES), one or more nuclear localization signals (NLS), or any combinations thereof. In some cases, the NES may be an HIV Rev NES. In certain cases, the NES may be MAPK NES. When the component is a protein, the NES or NLS may be at the C terminus of component. Alternatively or additionally, the NES or NLS may be at the N terminus of component. In some examples, the Cas protein and optionally said nucleotide deaminase protein or catalytic domain thereof comprise one or more heterologous nuclear export signal(s) (NES(s)) or nuclear localization signal(s) (NLS(s)), preferably an HIV Rev NES or MAPK NES, preferably C-terminal.

#### Templates

**[0132]** In some embodiments, the composition for engineering cells comprise a template, e.g., a recombination template. A template may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. In some embodiments, a recombination template is designed to serve as a

template in homologous recombination, such as within or near a target sequence nicked or cleaved by a nucleic acid-targeting effector protein as a part of a nucleic acid-targeting complex.

**[0133]** In an embodiment, the template nucleic acid alters the sequence of the target position. In an embodiment, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

**[0134]** The template sequence may undergo a breakage mediated or catalyzed recombination with the target sequence. In an embodiment, the template nucleic acid may include sequence that corresponds to a site on the target sequence that is cleaved by a Cas protein mediated cleavage event. In an embodiment, the template nucleic acid may include a sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Cas protein mediated event, and a second site on the target sequence that is cleaved in a second Cas protein mediated event.

**[0135]** In certain embodiments, the template nucleic acid can include a sequence which results in an alteration in the coding sequence of a translated sequence, e.g., one which results in the substitution of one amino acid for another in a protein product, e.g., transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introducing a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a non-sense mutation. In certain embodiments, the template nucleic acid can include a sequence which results in an alteration in a non-coding sequence, e.g., an alteration in an exon or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control element, e.g., a promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

**[0136]** A template nucleic acid having homology with a target position in a target gene may be used to alter the structure of a target sequence. The template sequence may be used to alter an unwanted structure, e.g., an unwanted or mutant nucleotide. The template nucleic acid may include a sequence which, when integrated, results in decreasing the activity of a positive control element; increasing the activity of a positive control element; decreasing the activity of a negative control element; increasing the activity of a negative control element; decreasing the expression of a gene; increasing the expression of a gene; increasing resistance to a disorder or disease; increasing resistance to viral entry; correcting a mutation or altering an unwanted amino acid residue conferring, increasing, abolishing or decreasing a biological property of a gene product, e.g., increasing the enzymatic activity of an enzyme, or increasing the ability of a gene product to interact with another molecule.

**[0137]** The template nucleic acid may include a sequence which results in a change in sequence of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more nucleotides of the target sequence.

**[0138]** A template polynucleotide may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In an embodiment, the template nucleic acid may be 20+/-10, 30+/-10, 40+/-10, 50+/-10, 60+/-10, 70+/-10, 80+/-10, 90+/-10, 100+/-10, 110+/-10, 120+/-10, 130+/-10, 140+/-10, 150+/-10, 160+/-10, 170+/-10, 180+/-10, 190+/-10, 200+/-10, 210+/-10, of 220+/-10 nucleotides in length. In an embodiment, the template nucleic acid may be 30+/-20, 40+/-20, 50+/-20, 60+/-20, 70+/-20, 80+/-20,

90+/-20, 100+/-20, 110+/-20, 120+/-20, 130+/-20, 140+/-20, 150+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/-20, 200+/-20, 210+/-20, of 220+/-20 nucleotides in length. In an embodiment, the template nucleic acid is 10 to 1,000, 20 to 900, 30 to 800, 40 to 700, 50 to 600, 50 to 500, 50 to 400, 50 to 300, 50 to 200, or 50 to 100 nucleotides in length.

**[0139]** In some embodiments, the template polynucleotide is complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a template polynucleotide might overlap with one or more nucleotides of a target sequences (e.g. about or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more nucleotides). In some embodiments, when a template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

**[0140]** The exogenous polynucleotide template comprises a sequence to be integrated (e.g., a mutated gene). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include polynucleotides encoding a protein or a non-coding RNA (e.g., a microRNA). Thus, the sequence for integration may be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated may provide a regulatory function.

**[0141]** An upstream or downstream sequence may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000.

**[0142]** An upstream or downstream sequence may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000.

**[0143]** In certain embodiments, one or both homology arms may be shortened to avoid including certain sequence repeat elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm may be shortened to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

**[0144]** In some methods, the exogenous polynucleotide template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The exogenous polynucleotide template of the disclosure can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996).

**[0145]** In certain embodiments, a template nucleic acid for correcting a mutation may be designed for use as a single-stranded oligonucleotide. When using a single-stranded oligonucleotide, 5' and 3' homology arms may range up to

about 200 base pairs (bp) in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length.

**[0146]** Suzuki et al. describe in vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration (2016, Nature 540:144-149).

#### TALE Nucleases

**[0147]** In some embodiments, a TALE nuclease or TALE nuclease system can be used to modify a polynucleotide. In some embodiments, the methods provided herein use isolated, non-naturally occurring, recombinant or engineered DNA binding proteins that comprise TALE monomers or TALE monomers or half monomers as a part of their organizational structure that enable the targeting of nucleic acid sequences with improved efficiency and expanded specificity.

**[0148]** Naturally occurring TALEs or “wild type TALEs” are nucleic acid binding proteins secreted by numerous species of proteobacteria. TALE polypeptides contain a nucleic acid binding domain composed of tandem repeats of highly conserved monomer polypeptides that are predominantly 33, 34 or 35 amino acids in length and that differ from each other mainly in amino acid positions 12 and 13. In advantageous embodiments the nucleic acid is DNA. As used herein, the term “polypeptide monomers”, “TALE monomers” or “monomers” will be used to refer to the highly conserved repetitive polypeptide sequences within the TALE nucleic acid binding domain and the term “repeat variable di-residues” or “RVD” will be used to refer to the highly variable amino acids at positions 12 and 13 of the polypeptide monomers. As provided throughout the disclosure, the amino acid residues of the RVD are depicted using the IUPAC single letter code for amino acids. A general representation of a TALE monomer which is comprised within the DNA binding domain is  $X_{1-11}-(X_{12}X_{13})-X_{14-33}$  or  $_{34}$  or  $_{35}$ , where the subscript indicates the amino acid position and X represents any amino acid.  $X_{12}X_{13}$  indicate the RVDs. In some polypeptide monomers, the variable amino acid at position 13 is missing or absent and in such monomers, the RVD consists of a single amino acid. In such cases the RVD may be alternatively represented as  $X^*$ , where X represents  $X_{12}$  and (\*) indicates that  $X_{13}$  is absent. The DNA binding domain comprises several repeats of TALE monomers and this may be represented as  $(X_{1-11}-(X_{12}X_{13})-X_{14-33}$  or  $_{34}$  or  $_{35})_z$ , where in an advantageous embodiment, z is at least 5 to 40. In a further advantageous embodiment, z is at least 10 to 26.

**[0149]** The TALE monomers can have a nucleotide binding affinity that is determined by the identity of the amino acids in its RVD. For example, polypeptide monomers with an RVD of NI can preferentially bind to adenine (A), monomers with an RVD of NG can preferentially bind to thymine (T), monomers with an RVD of HD can preferentially bind to cytosine (C) and monomers with an RVD of NN can preferentially bind to both adenine (A) and guanine (G). In some embodiments, monomers with an RVD of IG can preferentially bind to T. Thus, the number and order of the polypeptide monomer repeats in the nucleic acid binding domain of a TALE determines its nucleic acid target specificity. In some embodiments, monomers with an RVD of NS can recognize all four base pairs and can bind to A, T, G or C. The structure and function of TALEs is further described in, for example, Moscou et al., Science 326:1501 (2009);

Boch et al., *Science* 326:1509-1512 (2009); and Zhang et al., *Nature Biotechnology* 29:149-153 (2011).

**[0150]** The polypeptides used in methods of the invention can be isolated, non-naturally occurring, recombinant or engineered nucleic acid-binding proteins that have nucleic acid or DNA binding regions containing polypeptide monomer repeats that are designed to target specific nucleic acid sequences.

**[0151]** As described herein, polypeptide monomers having an RVD of HN or NH preferentially bind to guanine and thereby allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In some embodiments, polypeptide monomers having RVDs RN, NN, NK, SN, NH, KN, HN, NQ, RG, KH, RH and SS can preferentially bind to guanine. In some embodiments, polypeptide monomers having RVDs RN, NK, NQ, HH, KH, RH, SS and SN can preferentially bind to guanine and can thus allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In some embodiments, polypeptide monomers having RVDs HH, KH, NH, NK, NQ, RH, RN and SS can preferentially bind to guanine and thereby allow the generation of TALE polypeptides with high binding specificity for guanine containing target nucleic acid sequences. In some embodiments, the RVDs that have high binding specificity for guanine are RN, NH, RH and KH. Furthermore, polypeptide monomers having an RVD of NV can preferentially bind to adenine and guanine. In some embodiments, monomers having RVDs of H\*, HA, KA, N\*, NA, NC, NS, RA, and S\* bind to adenine, guanine, cytosine and thymine with comparable affinity.

**[0152]** The predetermined N-terminal to C-terminal order of the one or more polypeptide monomers of the nucleic acid or DNA binding domain determines the corresponding predetermined target nucleic acid sequence to which the polypeptides of the invention will bind. As used herein the monomers and at least one or more half monomers are “specifically ordered to target” the genomic locus or gene of interest. In plant genomes, the natural TALE-binding sites always begin with a thymine (T), which may be specified by a cryptic signal within the non-repetitive N-terminus of the TALE polypeptide; in some cases, this region may be referred to as repeat 0. In animal genomes, TALE binding sites do not necessarily have to begin with a thymine (T) and polypeptides of the invention may target DNA sequences that begin with T, A, G or C. The tandem repeat of TALE monomers always ends with a half-length repeat or a stretch of sequence that may share identity with only the first 20 amino acids of a repetitive full-length TALE monomer and this half repeat may be referred to as a half-monomer. Therefore, it follows that the length of the nucleic acid or DNA being targeted is equal to the number of full monomers plus two.

**[0153]** As described in Zhang et al., *Nature Biotechnology* 29:149-153 (2011), TALE polypeptide binding efficiency may be increased by including amino acid sequences from the “capping regions” that are directly N-terminal or C-terminal of the DNA binding region of naturally occurring TALEs into the engineered TALEs at positions N-terminal or C-terminal of the engineered TALE DNA binding region. Thus, in certain embodiments, the TALE polypeptides described herein further comprise an N-terminal capping region and/or a C-terminal capping region.

**[0154]** An exemplary amino acid sequence of a N-terminal capping region is:

```
(SEQ ID NO: 1)
MDPIRSRTSPSPARELLSGPQPDGVQPTADRGVSPAG
GPLDGLPARRTMSRTRLPSPPAPSPAFSADSFSDLLRQDFPSL
FNTSLFDSLPPFGAHHTEAATGEWDEVQSGLRADAPPPTM
RVAVTAARPPRAKPAPRRRAAQPSDASPAQVDLRTLGYSQ
QQQEKI KPKVRS TVAQHHEALVGHGFTHAHIVALSQHPAAL
GTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTV
AGELRGPPQLDTGQLLKI AKRGGV TAVEAVHAWRNALTGA
PLN
```

**[0155]** An exemplary amino acid sequence of a C-terminal capping region is:

```
(SEQ ID NO: 2)
RPALESIVAQLSRPDPALALNDHLVALACLGRPA
LDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRVLG
FFQCHSHPAQAFDDAMTQFGMSRHGLLQLFRRVGVTELEAR
SGTLPPASQRWDRILQASGMKRAKPSPTSTQTPDQASLHFAF
DSLERDL DAPSPMHEGDQTRAS
```

**[0156]** As used herein the predetermined “N-terminus” to “C terminus” orientation of the N-terminal capping region, the DNA binding domain comprising the repeat TALE monomers and the C-terminal capping region provide structural basis for the organization of different domains in the d-TALEs or polypeptides of the invention.

**[0157]** The entire N-terminal and/or C-terminal capping regions are not necessary to enhance the binding activity of the DNA binding region. Therefore, in certain embodiments, fragments of the N-terminal and/or C-terminal capping regions are included in the TALE polypeptides described herein.

**[0158]** In certain embodiments, the TALE polypeptides described herein contain a N-terminal capping region fragment that included at least 10, 20, 30, 40, 50, 54, 60, 70, 80, 87, 90, 94, 100, 102, 110, 117, 120, 130, 140, 147, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260 or 270 amino acids of an N-terminal capping region. In certain embodiments, the N-terminal capping region fragment amino acids are of the C-terminus (the DNA-binding region proximal end) of an N-terminal capping region. As described in Zhang et al., *Nature Biotechnology* 29:149-153 (2011), N-terminal capping region fragments that include the C-terminal 240 amino acids enhance binding activity equal to the full length capping region, while fragments that include the C-terminal 147 amino acids retain greater than 80% of the efficacy of the full length capping region, and fragments that include the C-terminal 117 amino acids retain greater than 50% of the activity of the full-length capping region.

**[0159]** In some embodiments, the TALE polypeptides described herein contain a C-terminal capping region fragment that included at least 6, 10, 20, 30, 37, 40, 50, 60, 68, 70, 80, 90, 100, 110, 120, 127, 130, 140, 150, 155, 160, 170,

180 amino acids of a C-terminal capping region. In certain embodiments, the C-terminal capping region fragment amino acids are of the N-terminus (the DNA-binding region proximal end) of a C-terminal capping region. As described in Zhang et al., *Nature Biotechnology* 29:149-153 (2011), C-terminal capping region fragments that include the C-terminal 68 amino acids enhance binding activity equal to the full-length capping region, while fragments that include the C-terminal 20 amino acids retain greater than 50% of the efficacy of the full-length capping region.

**[0160]** In certain embodiments, the capping regions of the TALE polypeptides described herein do not need to have identical sequences to the capping region sequences provided herein. Thus, in some embodiments, the capping region of the TALE polypeptides described herein have sequences that are at least 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical or share identity to the capping region amino acid sequences provided herein. Sequence identity is related to sequence homology. Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs may calculate percent (%) homology between two or more sequences and may also calculate the sequence identity shared by two or more amino acid or nucleic acid sequences. In some preferred embodiments, the capping region of the TALE polypeptides described herein have sequences that are at least 95% identical or share identity to the capping region amino acid sequences provided herein.

**[0161]** Sequence homologies can be generated by any of a number of computer programs known in the art, which include but are not limited to BLAST or FASTA. Suitable computer programs for carrying out alignments like the GCG Wisconsin Bestfit package may also be used. Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

**[0162]** In some embodiments described herein, the TALE polypeptides of the invention include a nucleic acid binding domain linked to the one or more effector domains. The terms “effector domain” or “regulatory and functional domain” refer to a polypeptide sequence that has an activity other than binding to the nucleic acid sequence recognized by the nucleic acid binding domain. By combining a nucleic acid binding domain with one or more effector domains, the polypeptides of the invention may be used to target the one or more functions or activities mediated by the effector domain to a particular target DNA sequence to which the nucleic acid binding domain specifically binds.

**[0163]** In some embodiments of the TALE polypeptides described herein, the activity mediated by the effector domain is a biological activity. For example, in some embodiments the effector domain is a transcriptional inhibitor (i.e., a repressor domain), such as an mSin interaction domain (SID), SID4X domain or a Krüppel-associated box (KRAB) or fragments of the KRAB domain. In some embodiments, the effector domain is an enhancer of transcription (i.e., an activation domain), such as the VP16, VP64 or p65 activation domain. In some embodiments, the nucleic acid binding is linked, for example, with an effector domain that includes but is not limited to a transposase, integrase, recombinase, resolvase, invertase, protease, DNA

methyltransferase, DNA demethylase, histone acetylase, histone deacetylase, nuclease, transcriptional repressor, transcriptional activator, transcription factor recruiting, protein nuclear-localization signal or cellular uptake signal.

**[0164]** In some embodiments, the effector domain is a protein domain which exhibits activities which include but are not limited to transposase activity, integrase activity, recombinase activity, resolvase activity, invertase activity, protease activity, DNA methyltransferase activity, DNA demethylase activity, histone acetylase activity, histone deacetylase activity, nuclease activity, nuclear-localization signaling activity, transcriptional repressor activity, transcriptional activator activity, transcription factor recruiting activity, or cellular uptake signaling activity. Other preferred embodiments of the invention may include any combination of the activities described herein.

#### Meganucleases

**[0165]** In some embodiments, a meganuclease or system thereof can be used to modify a polynucleotide. Meganucleases, which are endodeoxyribonucleases characterized by a large recognition site (double-stranded DNA sequences of 12 to 40 base pairs). Exemplary methods for using meganucleases can be found in U.S. Pat. Nos. 8,163,514, 8,133,697, 8,021,867, 8,119,361, 8,119,381, 8,124,369, and 8,129,134, which are specifically incorporated herein by reference.

#### RNAi

**[0166]** In certain embodiments, the genetic modifying agent is RNAi (e.g., shRNA). As used herein, “gene silencing” or “gene silenced” in reference to an activity of an RNAi molecule, for example a siRNA or miRNA refers to a decrease in the mRNA level in a cell for a target gene by at least about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, about 99%, about 100% of the mRNA level found in the cell without the presence of the miRNA or RNA interference molecule. In one preferred embodiment, the mRNA levels are decreased by at least about 70%, about 80%, about 90%, about 95%, about 99%, about 100%.

**[0167]** As used herein, the term “RNAi” refers to any type of interfering RNA, including but not limited to, siRNAi, shRNAi, endogenous microRNA and artificial microRNA. For instance, it includes sequences previously identified as siRNA, regardless of the mechanism of down-stream processing of the RNA (i.e. although siRNAs are believed to have a specific method of in vivo processing resulting in the cleavage of mRNA, such sequences can be incorporated into the vectors in the context of the flanking sequences described herein). The term “RNAi” can include both gene silencing RNAi molecules, and also RNAi effector molecules which activate the expression of a gene.

**[0168]** As used herein, a “siRNA” refers to a nucleic acid that forms a double stranded RNA, which double stranded RNA has the ability to reduce or inhibit expression of a gene or target gene when the siRNA is present or expressed in the same cell as the target gene. The double stranded RNA siRNA can be formed by the complementary strands. In one embodiment, a siRNA refers to a nucleic acid that can form a double stranded siRNA. The sequence of the siRNA can correspond to the full-length target gene, or a subsequence thereof. Typically, the siRNA is at least about 15-50 nucleotides in length (e.g., each complementary sequence of the

double stranded siRNA is about 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, preferably about 19-30 base nucleotides, preferably about 20-25 nucleotides in length, e.g., 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length).

**[0169]** As used herein “shRNA” or “small hairpin RNA” (also called stem loop) is a type of siRNA. In one embodiment, these shRNAs are composed of a short, e.g. about 19 to about 25 nucleotide, antisense strand, followed by a nucleotide loop of about 5 to about 9 nucleotides, and the analogous sense strand. Alternatively, the sense strand can precede the nucleotide loop structure and the antisense strand can follow.

**[0170]** The terms “microRNA” or “miRNA” are used interchangeably herein are endogenous RNAs, some of which are known to regulate the expression of protein-coding genes at the posttranscriptional level. Endogenous microRNAs are small RNAs naturally present in the genome that are capable of modulating the productive utilization of mRNA. The term artificial microRNA includes any type of RNA sequence, other than endogenous microRNA, which is capable of modulating the productive utilization of mRNA. MicroRNA sequences have been described in publications such as Lim, et al., *Genes & Development*, 17, p. 991-1008 (2003), Lim et al *Science* 299, 1540 (2003), Lee and Ambros *Science*, 294, 862 (2001), Lau et al., *Science* 294, 858-861 (2001), Lagos-*Quintana* et al, *Current Biology*, 12, 735-739 (2002), Lagos-*Quintana* et al, *Science* 294, 853-857 (2001), and Lagos-*Quintana* et al, *RNA*, 9, 175-179 (2003), which are incorporated herein by reference. Multiple microRNAs can also be incorporated into a precursor molecule. Furthermore, miRNA-like stem-loops can be expressed in cells as a vehicle to deliver artificial miRNAs and short interfering RNAs (siRNAs) for the purpose of modulating the expression of endogenous genes through the miRNA and or RNAi pathways.

**[0171]** As used herein, “double stranded RNA” or “dsRNA” refers to RNA molecules that are comprised of two strands. Double-stranded molecules include those comprised of a single RNA molecule that doubles back on itself to form a two-stranded structure. For example, the stem loop structure of the progenitor molecules from which the single-stranded miRNA is derived, called the pre-miRNA (Bartel et al. 2004. *Cell* 116:281-297), comprises a dsRNA molecule.

#### Immunoassays

**[0172]** Immunoassay methods are based on the reaction of an antibody to its corresponding target or analyte and can detect the analyte in a sample depending on the specific assay format. To improve specificity and sensitivity of an assay method based on immunoreactivity, monoclonal antibodies are often used because of their specific epitope recognition. Polyclonal antibodies have also been successfully used in various immunoassays because of their increased affinity for the target as compared to monoclonal antibodies. Immunoassays have been designed for use with a wide range of biological sample matrices. Immunoassay formats have been designed to provide qualitative, semi-quantitative, and quantitative results.

**[0173]** Quantitative results may be generated through the use of a standard curve created with known concentrations of the specific analyte to be detected. The response or signal from an unknown sample is plotted onto the standard curve,

and a quantity or value corresponding to the target in the unknown sample is established.

**[0174]** Numerous immunoassay formats have been designed. ELISA or EIA can be quantitative for the detection of an analyte/biomarker. This method relies on attachment of a label to either the analyte or the antibody and the label component includes, either directly or indirectly, an enzyme. ELISA tests may be formatted for direct, indirect, competitive, or sandwich detection of the analyte. Other methods rely on labels such as, for example, radioisotopes ( $I^{125}$ ) or fluorescence. Additional techniques include, for example, agglutination, nephelometry, turbidimetry, Western blot, immunoprecipitation, immunocytochemistry, immunohistochemistry, flow cytometry, Luminex assay, and others (see *ImmunoAssay: A Practical Guide*, edited by Brian Law, published by Taylor & Francis, Ltd., 2005 edition).

**[0175]** Exemplary assay formats include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay, fluorescent, chemiluminescence, and fluorescence resonance energy transfer (FRET) or time resolved-FRET (TR-FRET) immunoassays. Examples of procedures for detecting biomarkers include biomarker immunoprecipitation followed by quantitative methods that allow size and peptide level discrimination, such as gel electrophoresis, capillary electrophoresis, planar electrochromatography, and the like.

**[0176]** Methods of detecting and/or quantifying a detectable label or signal generating material depend on the nature of the label. The products of reactions catalyzed by appropriate enzymes (where the detectable label is an enzyme; see above) can be, without limitation, fluorescent, luminescent, or radioactive or they may absorb visible or ultraviolet light. Examples of detectors suitable for detecting such detectable labels include, without limitation, x-ray film, radioactivity counters, scintillation counters, spectrophotometers, colorimeters, fluorometers, luminometers, and densitometers.

**[0177]** Any of the methods for detection can be performed in any format that allows for any suitable preparation, processing, and analysis of the reactions. This can be, for example, in multi-well assay plates (e.g., 96 wells or 384 wells) or using any suitable array or microarray. Stock solutions for various agents can be made manually or robotically, and all subsequent pipetting, diluting, mixing, distribution, washing, incubating, sample readout, data collection and analysis can be done robotically using commercially available analysis software, robotics, and detection instrumentation capable of detecting a detectable label.

#### Single Cell Sequencing

**[0178]** In certain embodiments, the invention involves single cell RNA sequencing (see, e.g., Kalisky, T., Blainey, P. & Quake, S. R. *Genomic Analysis at the Single-Cell Level*. Annual review of genetics 45, 431-445, (2011); Kalisky, T. & Quake, S. R. *Single-cell genomics*. *Nature Methods* 8, 311-314 (2011); Islam, S. et al. *Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq*. *Genome Research*, (2011); Tang, F. et al. *RNA-Seq analysis to capture the transcriptome landscape of a single cell*. *Nature Protocols* 5, 516-535, (2010); Tang, F. et al. *mRNA-Seq whole-transcriptome analysis of a single cell*. *Nature Methods* 6, 377-382, (2009); Ramsköld, D. et al. *Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells*. *Nature Biotechnology* 30, 777-782, (2012); and Hashimshony, T., Wagner, F., Sher, N.

& Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Reports*, Cell Reports, Volume 2, Issue 3, p666-673, 2012).

**[0179]** In certain embodiments, the invention involves plate based single cell RNA sequencing (see, e.g., Picelli, S. et al., 2014, “Full-length RNA-seq from single cells using Smart-seq2” *Nature protocols* 9, 171-181, doi: 10.1038/nprot.2014.006).

**[0180]** In certain embodiments, the invention involves high-throughput single-cell RNA-seq. In this regard reference is made to Macosko et al., 2015, “Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets” *Cell* 161, 1202-1214; International patent application number PCT/US2015/049178, published as WO2016/040476 on Mar. 17, 2016; Klein et al., 2015, “Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells” *Cell* 161, 1187-1201; International patent application number PCT/US2016/027734, published as WO2016168584A1 on Oct. 20, 2016; Zheng, et al., 2016, “Haplotyping germline and cancer genomes with high-throughput linked-read sequencing” *Nature Biotechnology* 34, 303-311; Zheng, et al., 2017, “Massively parallel digital transcriptional profiling of single cells” *Nat. Commun.* 8, 14049 doi: 10.1038/ncomms14049; International patent publication number WO2014210353A2; Zilionis, et al., 2017, “Single-cell barcoding and sequencing using droplet microfluidics” *Nat Protoc.* January; 12(1):44-73; Cao et al., 2017, “Comprehensive single cell transcriptional profiling of a multicellular organism by combinatorial indexing” *bioRxiv preprint first posted online Feb. 2, 2017*, doi: dx.doi.org/10.1101/104844; Rosenberg et al., 2017, “Scaling single cell transcriptomics through split pool barcoding” *bioRxiv preprint first posted online Feb. 2, 2017*, doi: dx.doi.org/10.1101/105163; Rosenberg et al., “Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding” *Science* 15 Mar. 2018; Vitak, et al., “Sequencing thousands of single-cell genomes with combinatorial indexing” *Nature Methods*, 14(3):302-308, 2017; Cao, et al., *Comprehensive single-cell transcriptional profiling of a multicellular organism.* *Science*, 357(6352):661-667, 2017; and Gierahn et al., “Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput” *Nature Methods* 14, 395-398 (2017), all the contents and disclosure of each of which are herein incorporated by reference in their entirety.

**[0181]** In certain embodiments, the invention involves single nucleus RNA sequencing. In this regard reference is made to Swiech et al., 2014, “In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9” *Nature Biotechnology* Vol. 33, pp. 102-106; Habib et al., 2016, “Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons” *Science*, Vol. 353, Issue 6302, pp. 925-928; Habib et al., 2017, “Massively parallel single-nucleus RNA-seq with DroNc-seq” *Nat Methods*. 2017 October; 14(10):955-958; and International patent application number PCT/US2016/059239, published as WO2017164936 on Sep. 28, 2017, which are herein incorporated by reference in their entirety.

**[0182]** In certain embodiments, the invention involves the Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq) as described. (see, e.g., Buenrostro, et al., Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* 2013; 10

(12): 1213-1218; Buenrostro et al., Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486-490 (2015); Cusanovich, D. A., Daza, R., Adey, A., Pliner, H., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. & Shendure, J. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*. 2015 May 22; 348(6237): 910-4. doi: 10.1126/science.aab1601. Epub 2015 May 7; US20160208323A1; US20160060691A1; and WO2017156336A1).

#### Modulation and Modulating Agents

**[0183]** The term “modulate” broadly denotes a qualitative and/or quantitative alteration, change or variation in that which is being modulated. Where modulation can be assessed quantitatively—for example, where modulation comprises or consists of a change in a quantifiable variable such as a quantifiable property of a cell or where a quantifiable variable provides a suitable surrogate for the modulation—modulation specifically encompasses both increase (e.g., activation) or decrease (e.g., inhibition) in the measured variable. The term encompasses any extent of such modulation, e.g., any extent of such increase or decrease, and may more particularly refer to statistically significant increase or decrease in the measured variable. By means of example, modulation may encompass an increase in the value of the measured variable by at least about 10%, e.g., by at least about 20%, preferably by at least about 30%, e.g., by at least about 40%, more preferably by at least about 50%, e.g., by at least about 75%, even more preferably by at least about 100%, e.g., by at least about 150%, 200%, 250%, 300%, 400% or by at least about 500%, compared to a reference situation without said modulation; or modulation may encompass a decrease or reduction in the value of the measured variable by at least about 10%, e.g., by at least about 20%, by at least about 30%, e.g., by at least about 40%, by at least about 50%, e.g., by at least about 60%, by at least about 70%, e.g., by at least about 80%, by at least about 90%, e.g., by at least about 95%, such as by at least about 96%, 97%, 98%, 99% or even by 100%, compared to a reference situation without said modulation. Preferably, modulation may be specific or selective, hence, one or more desired phenotypic aspects of an immune cell or immune cell population may be modulated without substantially altering other (unintended, undesired) phenotypic aspect(s).

**[0184]** As used herein, “modulating” or “to modulate” generally means either reducing or inhibiting the expression or activity of, or alternatively increasing the expression or activity of a target or antigen. In particular, “modulating” or “to modulate” can mean either reducing or inhibiting the activity of, or alternatively increasing a (relevant or intended) biological activity of, a target or antigen as measured using a suitable *in vitro*, cellular or *in vivo* assay (which will usually depend on the target involved), by at least 5%, at least 10%, at least 25%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or more, compared to activity of the target in the same assay under the same conditions but without the presence of an agent. An “increase” or “decrease” refers to a statistically significant increase or decrease respectively. For the avoidance of doubt, an increase or decrease will be at least 10% relative to a reference, such as at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, at least

98%, or more, up to and including at least 100% or more, in the case of an increase, for example, at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 6-fold, at least 7-fold, at least 8-fold, at least 9-fold, at least 10-fold, at least 50-fold, at least 100-fold, or more. “Modulating” can also involve effecting a change (which can either be an increase or a decrease) in affinity, avidity, specificity and/or selectivity of a target or antigen. “Modulating” can also mean effecting a change with respect to one or more biological or physiological mechanisms, effects, responses, functions, pathways or activities in which the target or antigen (or in which its substrate(s), ligand(s) or pathway(s) are involved, such as its signaling pathway or metabolic pathway and their associated biological or physiological effects) is involved. Again, as will be clear to the skilled person, such an action as an agonist or an antagonist can be determined in any suitable manner and/or using any suitable assay known or described herein (e.g., in vitro or cellular assay), depending on the target or antigen involved.

**[0185]** Modulating can, for example, also involve allosteric modulation of the target and/or reducing or inhibiting the binding of the target to one of its substrates or ligands and/or competing with a natural ligand, substrate for binding to the target. Modulating can also involve activating the target or the mechanism or pathway in which it is involved. Modulating can for example also involve effecting a change in respect of the folding or confirmation of the target, or in respect of the ability of the target to fold, to change its conformation (for example, upon binding of a ligand), to associate with other (sub)units, or to disassociate. Modulating can for example also involve effecting a change in the ability of the target to signal, phosphorylate, dephosphorylate, and the like.

**[0186]** The term “agent” broadly encompasses any condition, substance or agent capable of modulating one or more phenotypic aspects of a cell or cell population as disclosed herein. Such conditions, substances or agents may be of physical, chemical, biochemical and/or biological nature. The term “candidate agent” refers to any condition, substance or agent that is being examined for the ability to modulate one or more phenotypic aspects of a cell or cell population as disclosed herein in a method comprising applying the candidate agent to the cell or cell population (e.g., exposing the cell or cell population to the candidate agent or contacting the cell or cell population with the candidate agent) and observing whether the desired modulation takes place.

**[0187]** Agents may include any potential class of biologically active conditions, substances or agents, such as for instance antibodies, proteins, peptides, nucleic acids, oligonucleotides, small molecules, or combinations thereof, as described herein.

**[0188]** The methods of phenotypic analysis can be utilized for evaluating environmental stress and/or state, for screening of chemical libraries, and to screen or identify structural, syntenic, genomic, and/or organism and species variations. For example, a culture of cells, can be exposed to an environmental stress, such as but not limited to heat shock, osmolarity, hypoxia, cold, oxidative stress, radiation, starvation, a chemical (for example a therapeutic agent or potential therapeutic agent) and the like. After the stress is applied, a representative sample can be subjected to analysis, for example at various time points, and compared to a control, such as a sample from an organism or cell, for

example a cell from an organism, or a standard value. By exposing cells, or fractions thereof, tissues, or even whole animals, to different members of the chemical libraries, and performing the methods described herein, different members of a chemical library can be screened for their effect on immune phenotypes thereof simultaneously in a relatively short amount of time, for example using a high throughput method.

**[0189]** A further aspect of the invention relates to a method for identifying an agent capable of modulating one or more phenotypic aspects of a cell or cell population as disclosed herein, comprising: a) applying a candidate agent to the cell or cell population; b) detecting modulation of one or more phenotypic aspects of the cell or cell population by the candidate agent, thereby identifying the agent. The phenotypic aspects of the cell or cell population that is modulated may be a gene signature or biological program specific to a cell type or cell phenotype or phenotype specific to a population of cells (e.g., an inflammatory phenotype or suppressive immune phenotype). In certain embodiments, steps can include administering candidate modulating agents to cells, detecting identified cell (sub)populations for changes in signatures, or identifying relative changes in cell (sub) populations which may comprise detecting relative abundance of particular gene signatures.

**[0190]** Aspects of the present disclosure relate to the correlation of an agent with the spatial proximity and/or epigenetic profile of the nucleic acids in a sample of cells. In some embodiments, the disclosed methods can be used to screen chemical libraries for agents that modulate chromatin architecture epigenetic profiles, and/or relationships thereof.

**[0191]** In some embodiments, screening of test agents involves testing a combinatorial library containing a large number of potential modulator compounds. A combinatorial chemical library may be a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical “building blocks” such as reagents. For example, a linear combinatorial chemical library, such as a polypeptide library, is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (for example the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

**[0192]** In certain embodiments, the present invention provides for gene signature screening. The concept of signature screening was introduced by Stegmaier et al. (Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nature Genet.* 36, 257-263 (2004)), who realized that if a gene-expression signature was the proxy for a phenotype of interest, it could be used to find small molecules that effect that phenotype without knowledge of a validated drug target. The signatures or biological programs of the present invention may be used to screen for drugs that reduce the signature or biological program in cells as described herein. The signature or biological program may be used for GE-HTS. In certain embodiments, pharmacological screens may be used to identify drugs that are selectively toxic to cells having a signature.

**[0193]** The Connectivity Map (cmap) is a collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules and



simple pattern-matching algorithms that together enable the discovery of functional connections between drugs, genes and diseases through the transitory feature of common gene-expression changes (see, Lamb et al., *The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease*. *Science* 29 Sep. 2006: Vol. 313, Issue 5795, pp. 1929-1935, DOI: 10.1126/science.1132939; and Lamb, J., *The Connectivity Map: a new tool for biomedical research*. *Nature Reviews Cancer* January 2007: Vol. 7, pp. 54-60). In certain embodiments, Cmap can be used to screen for small molecules capable of modulating a signature or biological program of the present invention in silico.

#### Genes and Polypeptides

**[0194]** All gene name symbols refer to the gene as commonly known in the art. The examples described herein that refer to the mouse gene names are to be understood to also encompass human genes, as well as genes in any other organism (e.g., homologous, orthologous genes). The term, homolog, may apply to the relationship between genes separated by the event of speciation (e.g., ortholog). Orthologs are genes in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution. Gene symbols may be those referred to by the HUGO Gene Nomenclature Committee (HGNC) or National Center for Biotechnology Information (NCBI). Any reference to the gene symbol is a reference made to the entire gene or variants of the gene. The signature as described herein may encompass any of the genes described herein.

#### Diseases

**[0195]** It will be understood by the skilled person that treating as referred to herein encompasses enhancing treatment, or improving treatment efficacy. Treatment may include inhibition of an inflammatory response, tumor regression as well as inhibition of tumor growth, metastasis or tumor cell proliferation, or inhibition or reduction of otherwise deleterious effects associated with the tumor.

**[0196]** Use of products, such as 2'-ATP or its mimetics described in (1) for the purpose of inhibiting STING/IFN $\beta$  signaling in diseases where STING is central to pathogenesis are contemplated, including diseases such as: Aicardi-Goutieres Syndrome (8-10), Inflammatory Bowel Disease (13), Lupus Erythematosus (11), and cancer, including for preventing the spread of metastatic disease (14-15).

**[0197]** A method of identifying candidates for treatment with inhibitors of the STING pathway are provided and may comprise detecting one or more mutations in a gene signature for innate immunity genes such as TREX1, ADAR, STING and/or gene signature for STING activation, thereby identifying the subject as a candidate for STING signaling inhibition.

**[0198]** Efficaciousness of treatment is determined in association with any known method for diagnosing or treating the particular disease. The invention comprehends a treatment method comprising any one of the methods or uses herein discussed.

**[0199]** The phrase "therapeutically effective amount" as used herein refers to a sufficient amount of a drug, agent, or compound to provide a desired therapeutic effect.

**[0200]** Therapy or treatment according to the invention may be performed alone or in conjunction with another therapy, and may be provided at home, the doctor's office, a clinic, a hospital's outpatient department, or a hospital. Treatment generally begins at a hospital so that the doctor can observe the therapy's effects closely and make any adjustments that are needed. The duration of the therapy depends on the age and condition of the patient, the stage of the cancer, and how the patient responds to the treatment. Additionally, a person having a greater risk of developing an inflammatory response (e.g., a person who is genetically predisposed or predisposed to allergies or a person having a disease characterized by episodes of inflammation) may receive prophylactic treatment to inhibit or delay symptoms of the disease.

#### Administration

**[0201]** It will be appreciated that administration of therapeutic entities in accordance with the invention will be administered with suitable carriers, excipients, and other agents that are incorporated into formulations to provide improved transfer, delivery, tolerance, and the like. A multitude of appropriate formulations can be found in the formulary known to all pharmaceutical chemists: Remington's *Pharmaceutical Sciences* (15th ed, Mack Publishing Company, Easton, PA (1975)), particularly Chapter 87 by Blaug, Seymour, therein. These formulations include, for example, powders, pastes, ointments, jellies, waxes, oils, lipids, lipid (cationic or anionic) containing vesicles (such as Lipofectin™), DNA conjugates, anhydrous absorption pastes, oil-in-water and water-in-oil emulsions, emulsions carbowax (polyethylene glycols of various molecular weights), semi-solid gels, and semi-solid mixtures containing carbowax. Any of the foregoing mixtures may be appropriate in treatments and therapies in accordance with the present invention, provided that the active ingredient in the formulation is not inactivated by the formulation and the formulation is physiologically compatible and tolerable with the route of administration. See also Baldrick P. "Pharmaceutical excipient development: the need for preclinical guidance." *Regul. Toxicol Pharmacol.* 32(2):210-8 (2000), Wang W. "Lyophilization and development of solid protein pharmaceuticals." *Int. J. Pharm.* 203(1-2):1-60 (2000), Charman W N "Lipids, lipophilic drugs, and oral drug delivery-some emerging concepts." *J Pharm Sci.* 89(8):967-78 (2000), Powell et al. "Compendium of excipients for parenteral formulations" *PDA J Pharm Sci Technol.* 52:238-311 (1998) and the citations therein for additional information related to formulations, excipients and carriers well known to pharmaceutical chemists.

**[0202]** The medicaments of the invention are prepared in a manner known to those skilled in the art, for example, by means of conventional dissolving, lyophilizing, mixing, granulating or confectioning processes. Methods well known in the art for making formulations are found, for example, in Remington: *The Science and Practice of Pharmacy*, 20th ed., ed. A. R. Gennaro, 2000, Lippincott Williams & Wilkins, Philadelphia, and *Encyclopedia of Pharmaceutical Technology*, eds. J. Swarbrick and J. C. Boylan, 1988-1999, Marcel Dekker, New York.

**[0203]** Administration of medicaments of the invention may be by any suitable means that results in a compound concentration that is effective for treating or inhibiting (e.g., by delaying) the development of a disease. The compound is

admixed with a suitable carrier substance, e.g., a pharmaceutically acceptable excipient that preserves the therapeutic properties of the compound with which it is administered. One exemplary pharmaceutically acceptable excipient is physiological saline. The suitable carrier substance is generally present in an amount of 1-95% by weight of the total weight of the medicament. The medicament may be provided in a dosage form that is suitable for administration. Thus, the medicament may be in form of, e.g., tablets, capsules, pills, powders, granulates, suspensions, emulsions, solutions, gels including hydrogels, pastes, ointments, creams, plasters, drenches, delivery devices, injectables, implants, sprays, or aerosols.

**[0204]** The agents disclosed herein (e.g., STING signaling agonists or antagonists) may be used in a pharmaceutical composition when combined with a pharmaceutically acceptable carrier. Such compositions comprise a therapeutically-effective amount of the agent and a pharmaceutically acceptable carrier. Such a composition may also further comprise (in addition to an agent and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. Compositions comprising the agent can be administered in the form of salts provided the salts are pharmaceutically acceptable. Salts may be prepared using standard procedures known to those skilled in the art of synthetic organic chemistry.

**[0205]** The term “pharmaceutically acceptable salts” refers to salts prepared from pharmaceutically acceptable non-toxic bases or acids including inorganic or organic bases and inorganic or organic acids. Salts derived from inorganic bases include aluminum, ammonium, calcium, copper, ferric, ferrous, lithium, magnesium, manganese salts, manganese, potassium, sodium, zinc, and the like. Particularly preferred are the ammonium, calcium, magnesium, potassium, and sodium salts. Salts derived from pharmaceutically acceptable organic non-toxic bases include salts of primary, secondary, and tertiary amines, substituted amines including naturally occurring substituted amines, cyclic amines, and basic ion exchange resins, such as arginine, betaine, caffeine, choline, N,N'-dibenzylethylenediamine, diethylamine, 2-diethylaminoethanol, 2-dimethylaminoethanol, ethanolamine, ethylenediamine, N-ethyl-morpholine, N-ethylpiperidine, glucamine, glucosamine, histidine, hydrabamine, isopropylamine, lysine, methylglucamine, morpholine, piperazine, piperidine, polyamine resins, procaine, purines, theobromine, triethylamine, trimethylamine, tripropylamine, tromethamine, and the like. The term “pharmaceutically acceptable salt” further includes all acceptable salts such as acetate, lactobionate, benzenesulfonate, laurate, benzoate, malate, bicarbonate, maleate, bisulfate, mandelate, bitartrate, mesylate, borate, methylbromide, bromide, methylnitrate, calcium edetate, methylsulfate, camsylate, mucate, carbonate, napsylate, chloride, nitrate, clavulanate, N-methylglucamine, citrate, ammonium salt, dihydrochloride, oleate, edetate, oxalate, edisylate, pamoate (embonate), estolate, palmitate, esylate, pantothenate, fumarate, phosphate/diphosphate, gluceptate, polygalacturonate, gluconate, salicylate, glutamate, stearate, glycolylarsanilate, sulfate, hexylresorcinolate, subacetate, hydrabamine, succinate, hydrobromide, tannate, hydrochloride, tartrate, hydroxynaphthoate, teoclate, iodide, tosylate, isothionate, triethiodide, lactate, pantoate, valerate, and the like which can be used as a dosage form for modifying the solubility or hydrolysis characteristics or can be used in sustained release or pro-

drug formulations. It will be understood that, as used herein, references to specific agents (e.g., neuromedin U receptor agonists or antagonists), also include the pharmaceutically acceptable salts thereof.

**[0206]** The composition of the invention can also advantageously be formulated in order to release 2'-ATP mimetics or derivatives, and/or agonist in the subject in a timely controlled fashion. In a particular embodiment, the composition of the invention is formulated for controlled release.

**[0207]** The agents of the present invention may be modified, such that they acquire advantageous properties for therapeutic use (e.g., stability and specificity), but maintain their biological activity.

**[0208]** In particular embodiments, the agents (e.g., STING signaling antagonists) include a protecting group covalently joined to the N-terminal amino group. In exemplary embodiments, a protecting group covalently joined to the N-terminal amino group of the agonists reduces the reactivity of the amino terminus under in vivo conditions. Amino protecting groups include —C1-10 alkyl, —C1-10 substituted alkyl, —C2-10 alkenyl, —C2-10 substituted alkenyl, aryl, —C1-6 alkyl aryl, —C(O)—(CH<sub>2</sub>)<sub>1-6</sub>-COOH, —C(O)—C1-6 alkyl, —C(O)-aryl, —C(O)—O—C1-6 alkyl, or —C(O)—O-aryl. In particular embodiments, the amino terminus protecting group is selected from the group consisting of acetyl, propyl, succinyl, benzyl, benzyloxycarbonyl, and t-butyloxycarbonyl. In other embodiments, deamination of the N-terminal amino acid is another modification that may be used for reducing the reactivity of the amino terminus under in vivo conditions.

**[0209]** Chemically modified compositions of the agents (wherein the agent is linked to a polymer) are also included within the scope of the present invention. The polymer selected is usually modified to have a single reactive group, such as an active ester for acylation or an aldehyde for alkylation, so that the degree of polymerization may be controlled. Included within the scope of polymers is a mixture of polymers. Preferably, for therapeutic use of the end-product preparation, the polymer will be pharmaceutically acceptable. The polymer or mixture thereof may include but is not limited to polyethylene glycol (PEG), monomethoxy-polyethylene glycol, dextran, cellulose, or other carbohydrate based polymers, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, a polypropylene oxide/ethylene oxide co-polymer, polyoxyethylated polyols (for example, glycerol), and polyvinyl alcohol.

**[0210]** In certain embodiments, the present invention provides for one or more therapeutic agents. In certain embodiments, the one or more agents comprises a small molecule inhibitor, small molecule degrader (e.g., PROTAC), genetic modifying agent, antibody, antibody fragment, antibody-like protein scaffold, aptamer, protein, or any combination thereof.

**[0211]** The terms “therapeutic agent”, “therapeutic capable agent” or “treatment agent” are used interchangeably and refer to a molecule or compound that confers some beneficial effect upon administration to a subject. The beneficial effect includes enablement of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition.

**[0212]** As used herein, “treatment” or “treating,” or “palliating” or “ameliorating” are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested. As used herein “treating” includes ameliorating, curing, preventing it from becoming worse, slowing the rate of progression, or preventing the disorder from re-occurring (i.e., to prevent a relapse). In certain embodiments, the present invention provides for one or more therapeutic agents against combinations of targets identified. Targeting the identified combinations may provide for enhanced or otherwise previously unknown activity in the treatment of disease.

**[0213]** One type of small molecule applicable to the present invention is a degrader molecule. Proteolysis Targeting Chimera (PROTAC) technology is a rapidly emerging alternative therapeutic strategy with the potential to address many of the challenges currently faced in modern drug development programs. PROTAC technology employs small molecules that recruit target proteins for ubiquitination and removal by the proteasome (see, e.g., Zhou et al., Discovery of a Small-Molecule Degradable of Bromodomain and Extra-Terminal (BET) Proteins with Picomolar Cellular Potencies and Capable of Achieving Tumor Regression. *J. Med. Chem.* 2018, 61, 462-481; Bondeson and Crews, Targeted Protein Degradation by Small Molecules, *Annu Rev Pharmacol Toxicol.* 2017 Jan. 6; 57: 107-123; and Lai et al., Modular PROTAC Design for the Degradation of Oncogenic BCR-ABL *Angew Chem Int Ed Engl.* 2016 Jan. 11; 55(2): 807-810).

**[0214]** In certain embodiments, combinations of targets are modulated (e.g., one or more targets related to STING signaling). In certain embodiments, an agent against one of the targets in a combination may already be known or used clinically. In certain embodiments, targeting the combination may require less of the agent as compared to the current standard of care and provide for less toxicity and improved treatment.

**[0215]** Methods of administering the pharmacological compositions, including agonists, antagonists, antibodies or fragments thereof, to an individual include, but are not limited to, intradermal, intrathecal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, by inhalation, and oral routes. The compositions can be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (for example, oral mucosa, rectal and intestinal mucosa, and the like), ocular, and the like and can be administered together with other biologically-active agents. Administration can be systemic or local. In addition, it may be advantageous to administer the composition into the central nervous system by any suitable route, including intraventricular and intrathecal injection. Pulmonary administration may also be employed by use of an inhaler or nebulizer, and formulation with an aerosolizing agent. It may also be desirable to administer the agent locally to the

area in need of treatment; this may be achieved by, for example, and not by way of limitation, local infusion during surgery, topical application, by injection, by means of a catheter, by means of a suppository, or by means of an implant.

**[0216]** Various delivery systems are known and can be used to administer the pharmacological compositions including, but not limited to, encapsulation in liposomes, microparticles, microcapsules; micelles; polymers; capsules; tablets; and the like. In one embodiment, the agent may be delivered in a vesicle, in particular a liposome. In a liposome, the agent is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithin, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Pat. Nos. 4,837,028 and 4,737,323. In yet another embodiment, the pharmacological compositions can be delivered in a controlled release system including, but not limited to: a delivery pump (See, for example, Saudek, et al., *New Engl. J. Med.* 321: 574 (1989) and a semi-permeable polymeric material (See, for example, Howard, et al., *J. Neurosurg.* 71: 105 (1989)). Additionally, the controlled release system can be placed in proximity of the therapeutic target (e.g., a tumor), thus requiring only a fraction of the systemic dose. See, for example, Goodson, In: *Medical Applications of Controlled Release*, 1984. (CRC Press, Boca Raton, Fla.).

#### Nanoparticle

**[0217]** Delivery of therapeutic agents can be accomplished via nanoparticle. Anderson et al. (US 20170079916) provides a modified dendrimer nanoparticle for the delivery of therapeutic, prophylactic and/or diagnostic agents to a subject, comprising: one or more zero to seven generation alkylated dendrimers; one or more amphiphilic polymers; and one or more therapeutic, prophylactic and/or diagnostic agents encapsulated therein. One alkylated dendrimer may be selected from the group consisting of poly(ethyleneimine), poly(polypropyleneimine), diaminobutane amine polypropyleneimine tetramine and poly(amido amine).

**[0218]** Anderson et al. (US20050123596) provides examples of microparticles that are designed to release their payload when exposed to acidic conditions, wherein the microparticles comprise at least one agent to be delivered, a pH triggering agent, and a polymer, wherein the polymer is selected from the group of polymethacrylates and polyacrylates. See also, Anderson et al (US 20020150626) providing lipid-protein-sugar particles for delivery of nucleic acids, wherein the polynucleotide is encapsulated in a lipid-protein-sugar matrix by contacting the polynucleotide with a lipid, a protein, and a sugar; and spray drying mixture of the polynucleotide, the lipid, the protein, and the sugar to make microparticles.

#### Liposomes and Lipids

**[0219]** Semi-solid and soft nanoparticles have been manufactured, and are within the scope of the present invention. A prototype nanoparticle of semi-solid nature is the lipo-

some. Various types of liposome nanoparticles are currently used clinically as delivery systems for anticancer drugs and vaccines. Nanoparticles with one half hydrophilic and the other half hydrophobic are termed Janus particles and are particularly effective for stabilizing emulsions. They can self-assemble at water/oil interfaces and act as solid surfactants.

**[0220]** Berg et al. (US20160174546) a nanolipid delivery system, in particular a nanoparticle concentrate, comprising: a composition comprising a lipid, oil or solvent, the composition having a viscosity of less than 100 cP at 25.degree. C. and a Kauri Butanol solvency of greater than 25 Kb; and at least one amphipathic compound selected from the group consisting of an alkoxyated lipid, an alkoxyated fatty acid, an alkoxyated alcohol, a heteroatomic hydrophilic lipid, a heteroatomic hydrophilic fatty acid, a heteroatomic hydrophilic alcohol, a diluent, and combinations thereof, wherein the compound is derived from a starting compound having a viscosity of less than 1000 cP at 50.degree. C., wherein the concentrate is configured to provide a stable nano emulsion having a D50 and a mean average particle size distribution of less than 100 nm when diluted.

**[0221]** Liu et al. (US 20140301951) provides a protocell nanostructure comprising: a porous particle core comprising a plurality of pores; and at least one lipid bilayer surrounding the porous particle core to form a protocell, wherein the protocell is capable of loading one or more cargo components to the plurality of pores of the porous particle core and releasing the one or more cargo components from the porous particle core across the surrounding lipid bilayer.

**[0222]** Bader et al. (US 20150250725), provides a method for producing a lipid particle comprising the following: i) providing a first solution comprising denatured apolipoprotein, ii) adding the first solution to a second solution comprising at least two lipids and a detergent but no apolipoprotein, and iii) removing the detergent from the solution obtained in ii) and thereby producing a lipid particle.

**[0223]** In another embodiment, the delivery system may be an administration device. As used herein, an administration device can be any pharmaceutically acceptable device adapted to deliver a composition of the invention (e.g., to a subject's nose). A nasal administration device can be a metered administration device (metered volume, metered dose, or metered-weight) or a continuous (or substantially continuous) aerosol-producing device. Suitable nasal administration devices also include devices that can be adapted or modified for nasal administration. In some embodiments, the nasally administered dose can be absorbed into the bloodstream of a subject.

**[0224]** A metered nasal administration device delivers a fixed (metered) volume or amount (dose) of a nasal composition upon each actuation. Exemplary metered dose devices for nasal administration include, by way of example and without limitation, an atomizer, sprayer, dropper, squeeze tube, squeeze-type spray bottle, pipette, ampule, nasal cannula, metered dose device, nasal spray inhaler, breath actuated bi-directional delivery device, pump spray, pre-compression metered dose spray pump, monospray pump, bispray pump, and pressurized metered dose device. The administration device can be a single-dose disposable device, single-dose reusable device, multi-dose disposable device or multi-dose reusable device. The compositions of the invention can be used with any known metered administration device.

**[0225]** A continuous aerosol-producing device delivers a mist or aerosol comprising droplet of a nasal composition dispersed in a continuous gas phase (such as air). A nebulizer, pulsating aerosol nebulizer, and a nasal continuous positive air pressure device are exemplary of such a device. Suitable nebulizers include, by way of example and without limitation, an air driven jet nebulizer, ultrasonic nebulizer, capillary nebulizer, electromagnetic nebulizer, pulsating membrane nebulizer, pulsating plate (disc) nebulizer, pulsating/vibrating mesh nebulizer, vibrating plate nebulizer, a nebulizer comprising a vibration generator and an aqueous chamber, a nebulizer comprising a nozzle array, and nebulizers that extrude a liquid formulation through a self-contained nozzle array.

**[0226]** In certain embodiments, the device can be any commercially available administration devices that are used or can be adapted for nasal administration of a composition of the invention (see, e.g., US patent publication US20090312724A1).

**[0227]** The amount of the agents (e.g., STING signaling agonist) which will be effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and may be determined by standard clinical techniques by those of skill within the art. In addition, in vitro assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the overall seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each patient's circumstances. Ultimately, the attending physician will decide the amount of the agent with which to treat each individual patient. In certain embodiments, the attending physician will administer low doses of the agent and observe the patient's response. Larger doses of the agent may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. In general, the daily dose range lie within the range of from about 0.001 mg to about 100 mg per kg body weight of a mammal, preferably 0.01 mg to about 50 mg per kg, and most preferably 0.1 to 10 mg per kg, in single or divided doses. On the other hand, it may be necessary to use dosages outside these limits in some cases. In certain embodiments, suitable dosage ranges for intravenous administration of the agent are generally about 5-500 micrograms ( $\mu\text{g}$ ) of active compound per kilogram (Kg) body weight. Suitable dosage ranges for intranasal administration are generally about 0.01  $\mu\text{g}/\text{kg}$  body weight to 1 mg/kg body weight. In certain embodiments, a composition containing an agent of the present invention is subcutaneously injected in adult patients with dose ranges of approximately 5 to 5000  $\mu\text{g}/\text{human}$  and preferably approximately 5 to 500  $\mu\text{g}/\text{human}$  as a single dose. It is desirable to administer this dosage 1 to 3 times daily. Effective doses may be extrapolated from dose-response curves derived from in vitro or animal model test systems. Suppositories generally contain active ingredient in the range of 0.5% to 10% by weight; oral formulations preferably contain 10% to 95% active ingredient. Ultimately the attending physician will decide on the appropriate duration of therapy using compositions of the present invention. Dosage will also vary according to the age, weight and response of the individual patient.

**[0228]** Methods for administering antibodies for therapeutic use is well known to one skilled in the art. In certain

embodiments, small particle aerosols of antibodies or fragments thereof may be administered (see e.g., Piazza et al., *J. Infect. Dis.*, Vol. 166, pp. 1422-1424, 1992; and Brown, *Aerosol Science and Technology*, Vol. 24, pp. 45-56, 1996). In certain embodiments, antibodies are used as agonists to depress inflammatory diseases. In certain embodiments, antibodies may be administered in liposomes, i.e., immunoliposomes (see, e.g., Maruyama et al., *Biochim. Biophys. Acta*, Vol. 1234, pp. 74-80, 1995). In certain embodiments, immunoconjugates, immunoliposomes or immunomicrospheres containing an agent of the present invention is administered by inhalation.

**[0229]** In certain embodiments, antibodies may be topically administered to mucosa, such as the oropharynx, nasal cavity, respiratory tract, gastrointestinal tract, eye such as the conjunctival mucosa, vagina, urogenital mucosa, or for dermal application. In certain embodiments, antibodies are administered to the nasal, bronchial or pulmonary mucosa. In order to obtain optimal delivery of the antibodies to the pulmonary cavity in particular, it may be advantageous to add a surfactant such as a phosphoglyceride, e.g. phosphatidylcholine, and/or a hydrophilic or hydrophobic complex of a positively or negatively charged excipient and a charged antibody of the opposite charge.

**[0230]** Other excipients suitable for pharmaceutical compositions intended for delivery of antibodies to the respiratory tract mucosa may be a) carbohydrates, e.g., monosaccharides such as fructose, galactose, glucose, D-mannose, sorbitose, and the like; disaccharides, such as lactose, trehalose, cellobiose, and the like; cyclodextrins, such as 2-hydroxypropyl- $\beta$ -cyclodextrin; and polysaccharides, such as raffinose, maltodextrins, dextrans, and the like; b) amino acids, such as glycine, arginine, aspartic acid, glutamic acid, cysteine, lysine and the like; c) organic salts prepared from organic acids and bases, such as sodium citrate, sodium ascorbate, magnesium gluconate, sodium gluconate, tromethamine hydrochloride, and the like; d) peptides and proteins, such as aspartame, human serum albumin, gelatin, and the like; e) alditols, such mannitol, xylitol, and the like, and f) polycationic polymers, such as chitosan or a chitosan salt or derivative.

**[0231]** For dermal application, the antibodies of the present invention may suitably be formulated with one or more of the following excipients: solvents, buffering agents, preservatives, humectants, chelating agents, antioxidants, stabilizers, emulsifying agents, suspending agents, gel-forming agents, ointment bases, penetration enhancers, and skin protective agents.

**[0232]** Examples of solvents are e.g. water, alcohols, vegetable or marine oils (e.g. edible oils like almond oil, castor oil, cacao butter, coconut oil, corn oil, cottonseed oil, linseed oil, olive oil, palm oil, peanut oil, poppy seed oil, rapeseed oil, sesame oil, soybean oil, sunflower oil, and tea seed oil), mineral oils, fatty oils, liquid paraffin, polyethylene glycols, propylene glycols, glycerol, liquid polyalkylsiloxanes, and mixtures thereof.

**[0233]** Examples of buffering agents are e.g. citric acid, acetic acid, tartaric acid, lactic acid, hydrogenphosphoric acid, diethyl amine etc. Suitable examples of preservatives for use in compositions are parabenes, such as methyl, ethyl, propyl p-hydroxybenzoate, butylparaben, isobutylparaben, isopropylparaben, potassium sorbate, sorbic acid, benzoic acid, methyl benzoate, phenoxyethanol, bronopol, bronidox,

MDM hydantoin, iodopropynyl butylcarbamate, EDTA, benzalconium chloride, and benzylalcohol, or mixtures of preservatives.

**[0234]** Examples of humectants are glycerin, propylene glycol, sorbitol, lactic acid, urea, and mixtures thereof.

**[0235]** Examples of antioxidants are butylated hydroxy anisole (BHA), ascorbic acid and derivatives thereof, tocopherol and derivatives thereof, cysteine, and mixtures thereof.

**[0236]** Examples of emulsifying agents are naturally occurring gums, e.g. gum acacia or gum tragacanth; naturally occurring phosphatides, e.g. soybean lecithin, sorbitan monooleate derivatives; wool fats; wool alcohols; sorbitan esters; monoglycerides; fatty alcohols; fatty acid esters (e.g. triglycerides of fatty acids); and mixtures thereof.

**[0237]** Examples of suspending agents are e.g. celluloses and cellulose derivatives such as, e.g., carboxymethyl cellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropylmethylcellulose, carraghenan, acacia gum, arabic gum, tragacanth, and mixtures thereof.

**[0238]** Examples of gel bases, viscosity-increasing agents or components which are able to take up exudate from a wound are: liquid paraffin, polyethylene, fatty oils, colloidal silica or aluminum, zinc soaps, glycerol, propylene glycol, tragacanth, carboxyvinyl polymers, magnesium-aluminum silicates, Carbopol®, hydrophilic polymers such as, e.g. starch or cellulose derivatives such as, e.g., carboxymethylcellulose, hydroxyethylcellulose and other cellulose derivatives, water-swelling hydrocolloids, carragenans, hyaluronates (e.g. hyaluronate gel optionally containing sodium chloride), and alginates including propylene glycol alginate.

**[0239]** Examples of ointment bases are e.g. beeswax, paraffin, cetanol, cetyl palmitate, vegetable oils, sorbitan esters of fatty acids (Span), polyethylene glycols, and condensation products between sorbitan esters of fatty acids and ethylene oxide, e.g. polyoxyethylene sorbitan monooleate (Tween).

**[0240]** Examples of hydrophobic or water-emulsifying ointment bases are paraffins, vegetable oils, animal fats, synthetic glycerides, waxes, lanolin, and liquid polyalkylsiloxanes. Examples of hydrophilic ointment bases are solid macrogols (polyethylene glycols). Other examples of ointment bases are triethanolamine soaps, sulphated fatty alcohol and polysorbates.

**[0241]** Examples of other excipients are polymers such as carmelose, sodium carmelose, hydroxypropylmethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, pectin, xanthan gum, locust bean gum, acacia gum, gelatin, carbomer, emulsifiers like vitamin E, glyceryl stearates, cetanyl glucoside, collagen, carrageenan, hyaluronates and alginates and chitosans.

**[0242]** The dose of antibody required in humans to be effective in the treatment or prevention of allergic inflammation differs with the type and severity of the allergic condition to be treated, the type of allergen, the age and condition of the patient, etc. Typical doses of antibody to be administered are in the range of 1  $\mu$ g to 1 g, preferably 1-1000  $\mu$ g, more preferably 2-500, even more preferably 5-50, most preferably 10-20  $\mu$ g per unit dosage form. In certain embodiments, infusion of antibodies of the present invention may range from 10-500 mg/m<sup>2</sup>.

**[0243]** There are a variety of techniques available for introducing nucleic acids into viable cells. The techniques

vary depending upon whether the nucleic acid is transferred into cultured cells *in vitro*, or *in vivo* in the cells of the intended host. Techniques suitable for the transfer of nucleic acid into mammalian cells *in vitro* include the use of liposomes, electroporation, microinjection, cell fusion, DEAE-dextran, the calcium phosphate precipitation method, etc. The currently preferred *in vivo* gene transfer techniques include transfection with viral (typically retroviral) vectors and viral coat protein-liposome mediated transfection.

[0244] In another aspect, provided is an administration device, pharmaceutical pack or kit, comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions, such as STING/IFN- $\beta$  signaling agonists or antagonists, and/or additional therapeutic agents.

[0245] Further embodiments are illustrated in the following Examples which are given for illustrative purposes only and are not intended to limit the scope of the invention.

#### EXAMPLES

[0246] The work herein describes the identification and characterization of the compound 2'-ATP.

##### Example 1. Extract Suppressing cGAS-STING Signaling

[0247] Polar Extracts from Cells Expressing Catalytically Active MB21D2 can Suppress cGAS-STING Signaling

[0248] Because Applicants suspected that the cGAS homolog MB21D2 generates a small molecule second messenger, wildtype or inactive MB21D2 was expressed in cells, then harvested the polar fraction of these cells (i.e. small molecules that are water soluble and smaller than 3 kilodaltons). Applicants then transfected these extracts into the assay that was used to measure the effects of MB21D2 expression on interferon signaling (FIG. 1B workflow at top).

[0249] Interestingly, it was found that extracts only from cells that expressed wildtype MB21D2 could suppress the interferon reporter (FIG. 1B chart).

[0250] Purification approaches as detailed in FIG. 2 was used to evaluate the purified fraction chromatographically. Using iterative rounds of purification by HPLC on 3 different columns, the anti-STING activity was localized to a single peak (the major peak in the bottom trace at a retention time of 24.5 minutes). (FIG. 3). The purified fraction contains an unambiguously unique ion in positive and negative ionization modes in the experimental but not control fractions. (FIG. 4). The  $m/z$  505.9891 is chemically similar to 5'-ATP. (FIG. 5).

Snake Venom PDEI Breaks Down the Purified Fraction Down to Adenosine

[0251] Snake venom phosphodiesterase I (SVPDE) is a valuable diagnostic tool that helped us decipher the location of the phosphate groups on the ribose sugar. SVPDE can cleave 2' or 3'-linked phosphates completely, while 5' phosphates are only cleaved to result in a 5'-linked monophosphate. In other words, SVPDE will cleave 5'-ATP into 5'-AMP, not adenosine. The product purified here is cleaved to adenosine (FIG. 6A, negative control in FIG. 6B), which indicates that the triphosphate moiety of the molecule is on the 2' or 3' carbon, NOT on the 5' carbon. SVPDE can only

degrade 5'-ATP to 5'-AMP, and would only produce Adenosine form a 2' or 3' phosphorylated AMP/ADP/ATP.

[0252] Unlike SVPDE, recombinant shrimp alkaline phosphatase (rsAP) is an enzyme that sequentially cleaves terminal phosphate groups from small molecules and proteins. Thus, full digestion of 5'-ATP or 2'-ATP would result in adenosine, while partial digestion of an ATP would yield ATP, ADP, AMP and adenosine.

[0253] Because 2'-, 3'- or 5'-AMP are all available as chemical standards, they were used as benchmarks (FIG. 7, top three panels) to determine which AMP species appeared after partial digestion of purified 2'- or 3'-ATP with rsAP.

[0254] The purified material appears as a major peak when treated with inactive rsAP (FIG. 7, fourth panel), but when treated with active rsAP the breakdown products that appear are adenosine and 2'-AMP (FIG. 7, bottom panel). This indicates that the purified molecule is 2'-ATP.  $C_{10}H_{16}N_5O_{13}P_3$  Neutral MW 506.99575 Da (or g/mol)

NMR Spectroscopy Further Rules Out 5'-ATP as the Identity of the Purified Molecule

[0255] While the biochemical data presented herein confirms that this molecule is 2'-ATP, nuclear magnetic resonance (NMR) spectroscopy will be utilized for additional definitive analysis. Applicants have performed  $^1H$  NMR ("proton NMR") analysis of the purified 2'-ATP at the Broad Institute. The data support the identity of 2'-ATP as being chemically distinct from 5'-ATP. Additional definitive NMR data will come in the form of  $^{13}C$  NMR ("carbon NMR"), forthcoming.

##### Example 2. Synthesis of Adenosine 2'-Triphosphate, Sodium Salt

[0256] Initial two-step proposed synthesis of 2'ATP produced undesired cyclization during the first reaction step. (FIG. 12A) However, a one-pot synthesis as depicted in FIG. 12B produced successful yields Adenosine 2'-monophosphate (100 mg, 288  $\mu$ mol, Merck) was suspended in a mixture of DMF (10 ml, dried over molecular sieve, Acros) and triethylamine (1 ml, Acros) and the solvents were evaporated subsequently. The resulting Adenosine 2'-monophosphate as triethylammonium salt was suspended in DMF, 1,1'-carbonyldiimidazole (280 mg, 1.7 mmol, Merck) was added and the mixture was stirred at room temperature for 25 min. A suspension of pyrophosphate as triethylammonium salt (256 mg, 551  $\mu$ mol, Jena bioscience) in DMF (5 ml) was added, the reaction mixture was further stirred at room temperature overnight and the conversion was monitored by RP-HPLC. The mixture was quenched with water, diluted to a volume of 600 ml and the pH value adjusted to 7.5 using aqueous NaOH. The product was enriched using ion exchange column (Q Sepharose, 300 ml, triethylammonium bicarbonate buffer 50 mM/2 M, 0 $\rightarrow$ 100%) and purified by reversed phase chromatography (C18 silica, 300 ml, triethylammonium bicarbonate 25 mM/70% MeOH in triethylammonium bicarbonate 25 mM, 0 $\rightarrow$ 100%). Solvents were removed *in vacuo* and the residue was dissolved in MeOH (2 ml), precipitated with a solution of sodium perchlorate monohydrate in acetone (2 ml, 1 M), and overlaid with acetone. The precipitate was separated and washed three times with acetone. Drying *in vacuo* led to formation of 28 mg (49  $\mu$ mol, 17% yield) of the title compound as white powder.

- [0257]** RP-HPLC (InertSustain® C18, 50 mM triethylammonium acetate, 1.5% MeCN/MeCN, 0→70%, 1.5 ml/min, 268 nm): Rf 10.3 min.
- [0258]** <sup>1</sup>H NMR (400.22 MHz, d<sub>6</sub>-DMSO): δ8.35 (s, 1H, H-8), 8.12 (s, 1H, H-2), 7.37 (s, 2H, —NH<sub>2</sub>), 5.95 (d, <sup>3</sup>J<sub>H,H</sub>=7.6 Hz, 1H, H-1'), 5.19 (dt, <sup>3</sup>J<sub>H,H</sub>=4.6 Hz, <sup>3</sup>J<sub>P,H</sub>=8.4 Hz, 1H, H-2'), 4.37 (d, <sup>3</sup>J<sub>H,H</sub>=4.8 Hz, 1H, H-3'), 3.98 (t, <sup>3</sup>J<sub>H,H</sub>=3.6 Hz, 1H, H-4'), 3.57 (ddd, <sup>3</sup>J<sub>H,H</sub>=3.6 Hz, <sup>3</sup>J<sub>H,H</sub>=12.0 Hz, H-5'), 3.16 (s, <sup>1</sup>H, —OH, MeOH), 2.97 (q, <sup>3</sup>J<sub>H,H</sub>=7.2 Hz, <sup>3</sup>J<sub>H,H</sub>=14.8 Hz, —CH<sub>2</sub>, TEA), 1.13 (t, <sup>3</sup>J<sub>H,H</sub>=12.8 Hz, —CH<sub>3</sub>, TEA).
- [0259]** <sup>13</sup>C {<sup>1</sup>H} NMR (100.6 MHz, d<sub>6</sub>-DMSO): δ156.2 (s, C-6), 152.4 (s, C-2), 149.3 (s, C-4), 139.9 (s, C-8), 119.3 (s, C-5), 86.5 (s, C-4'), 85.9 (d, <sup>3</sup>J<sub>P,C</sub>=11.0 Hz, C-1'), 76.8 (d, <sup>2</sup>J<sub>P,C</sub>=4.3 Hz, C-2'), 70.1 (s, C-3'), 62.1 (s, C-5'), 48.6 (s, MeOH), 44.9 (s, TEA), 8.3 (s, TEA).
- [0260]** <sup>31</sup>P NMR (162.02 MHz, d<sub>6</sub>-DMSO): δ-7.94 (d, <sup>2</sup>J<sub>P,P</sub>=23.8 Hz, Ψ-P), -9.43 (dd, <sup>3</sup>J<sub>P,H</sub>=8.5 Hz, <sup>2</sup>J<sub>P,P</sub>=26.6, α-P), -20.25 (t, <sup>2</sup>J<sub>P,P</sub>=24.8 Hz, β-P).
- [0261]** MS (ESI neg.): m/z 506 (M-H)<sup>-</sup>.
- [0262]** MS (ESI pos.): m/z 609 (M+H+NEt<sub>3</sub>)<sup>+</sup>, 710 (M+H+2×NEt<sub>3</sub>)<sup>+</sup>, 811 (M+H+3×NEt<sub>3</sub>)<sup>+</sup>.
- [0263]** The following references relate to the examples provided herein and are specifically incorporated by reference in their entirety:
- [0264]** 1. Wu J, Sun L, Chen X, et al. Cyclic-GMP-AMP Is An Endogenous Second Messenger in Innate Immune Signaling by Cytosolic DNA. *Science* (New York, NY). 2013; 339(6121):10.1126/science.1229963. doi:10.1126/science.1229963.
- [0265]** 2. Sun L, Wu J, Du F, Chen X, Chen Z J. Cyclic GMP-AMP Synthase is a Cytosolic DNA Sensor that Activates the Type-I Interferon Pathway. *Science* (New York, NY). 2013; 339(6121):10.1126/science.1232458. doi:10.1126/science.1232458.
- [0266]** 3. Zhang X, Shi H, Wu J, et al. Cyclic GMP-AMP Containing Mixed Phosphodiester Linkages Is An Endogenous High Affinity Ligand for STING. *Molecular cell*. 2013; 51(2):10.1016/j.molcel.2013.05.022. doi:10.1016/j.molcel.2013.05.022.
- [0267]** 4. Cerami E, Gao J, Dogrusoz U, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer discovery*. 2012; 2(5):401-404. doi:10.1158/2159-8290.CD-12-0095.
- [0268]** 5. Campbell J D, Alexandrov A, Kim J, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nature genetics*. 2016; 48(6):607-616. doi:10.1038/ng.3564.
- [0269]** 6. Kuchta K, Knizewski L, Wyrwicz L S, Rychlewski L, Ginalski K. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Research*. 2009; 37(22):7701-7714. doi:10.1093/nar/gkp854.
- [0270]** 7. Hornung, V., Hartmann, R., Ablasser, A. & Hopfner, K.-P. OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat. Rev. Immunol.* 14, 521-528(2014).
- [0271]** 8. Gao D, Li T, Li X-D, et al. Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases. *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112(42): E5699-E5705. doi:10.1073/pnas.1516465112.
- [0272]** 9. Mackenzie K J, Carroll P, Lettice L, et al. Ribonuclease H2 mutations induce a cGAS/STING-dependent innate immune response. *The EMBO Journal*. 2016; 35(8):831-844. doi:10.15252/embj.201593339.
- [0273]** 10. Pawaria S, Sharma S, Baum R, et al. Taking the STING out of TLR-driven autoimmune diseases: good, bad, or indifferent? *Journal of Leukocyte Biology*. 2017; 101(1):121-126. doi:10.1189/jlb.3MR0316-115R.
- [0274]** 11. Vincent J, Adura C, Gao P, et al. Small molecule inhibition of cGAS reduces interferon expression in primary macrophages from autoimmune mice. *Nature Communications*. 2017; 8:750. doi:10.1038/s41467-017-00833-9.
- [0275]** 12. Liu Y, Jesus A A, Marrero B, et al. Activated STING in a Vascular and Pulmonary Syndrome. *The New England journal of medicine*. 2014; 371(6):507-518. doi: 10.1056/NEJMoa1312625.
- [0276]** 13. Jeonghyun Ahn, Sehee Son, Sergio C. Oliveira, Glen N. Barber, STING-Dependent Signaling Underlies IL-10 Controlled Inflammatory Colitis, *Cell Reports*, Volume 21, Issue 13, 2017, Pages 3873-3884.
- [0277]** 14. Bakhom et al., "Chromosomal Instability Drives Metastasis through a Cytosolic DNA Response."
- [0278]** 15. Chen Q, Boire A, Jin X, et al. Carcinoma-astrocyte gap junctions promote brain metastasis by cGAMP transfer. *Nature*. 2016; 533(7604):493-498. doi: 10.1038/nature18268.
- [0279]** Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

## SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 2

<210> SEQ ID NO 1
<211> LENGTH: 288
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence

```

-continued

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Synthetic N/C-terminal capping region

&lt;400&gt; SEQUENCE: 1

```

Met Asp Pro Ile Arg Ser Arg Thr Pro Ser Pro Ala Arg Glu Leu Leu
1          5          10          15

Ser Gly Pro Gln Pro Asp Gly Val Gln Pro Thr Ala Asp Arg Gly Val
20          25          30

Ser Pro Pro Ala Gly Gly Pro Leu Asp Gly Leu Pro Ala Arg Arg Thr
35          40          45

Met Ser Arg Thr Arg Leu Pro Ser Pro Pro Ala Pro Ser Pro Ala Phe
50          55          60

Ser Ala Asp Ser Phe Ser Asp Leu Leu Arg Gln Phe Asp Pro Ser Leu
65          70          75          80

Phe Asn Thr Ser Leu Phe Asp Ser Leu Pro Pro Phe Gly Ala His His
85          90          95

Thr Glu Ala Ala Thr Gly Glu Trp Asp Glu Val Gln Ser Gly Leu Arg
100         105         110

Ala Ala Asp Ala Pro Pro Pro Thr Met Arg Val Ala Val Thr Ala Ala
115         120         125

Arg Pro Pro Arg Ala Lys Pro Ala Pro Arg Arg Arg Ala Ala Gln Pro
130         135         140

Ser Asp Ala Ser Pro Ala Ala Gln Val Asp Leu Arg Thr Leu Gly Tyr
145         150         155         160

Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val
165         170         175

Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His
180         185         190

Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val
195         200         205

Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His Glu Ala
210         215         220

Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala
225         230         235         240

Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp
245         250         255

Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val
260         265         270

Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn
275         280         285

```

&lt;210&gt; SEQ ID NO 2

&lt;211&gt; LENGTH: 183

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Synthetic C-terminal capping region

&lt;400&gt; SEQUENCE: 2

```

Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro
1          5          10          15

Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu
20          25          30

Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala

```



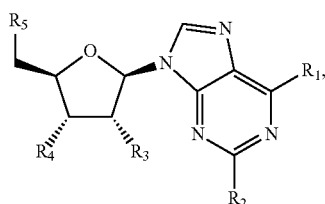
-continued

35	40	45
Pro Ala Leu Ile Lys Arg	Thr Asn Arg Arg Ile	Pro Glu Arg Thr Ser
50	55	60
His Arg Val Ala Asp His	Ala Gln Val Val Arg	Val Leu Gly Phe Phe
65	70	75
80		
Gln Cys His Ser His Pro	Ala Gln Ala Phe Asp	Asp Ala Met Thr Gln
85	90	95
Phe Gly Met Ser Arg His	Gly Leu Leu Gln Leu	Phe Arg Arg Val Gly
100	105	110
Val Thr Glu Leu Glu Ala	Arg Ser Gly Thr Leu	Pro Pro Ala Ser Gln
115	120	125
Arg Trp Asp Arg Ile Leu	Gln Ala Ser Gly Met	Lys Arg Ala Lys Pro
130	135	140
Ser Pro Thr Ser Thr Gln	Thr Pro Asp Gln Ala	Ser Leu His Ala Phe
145	150	155
160		
Ala Asp Ser Leu Glu Arg	Asp Leu Asp Ala Pro	Ser Pro Met His Glu
165	170	175
180		
Gly Asp Gln Thr Arg Ala	Ser	

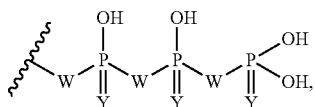
What is claimed is:

1. An agent for inhibiting a Stimulator of Interferon Genes (STING) signaling pathway, wherein the agent is a 2'-ATP derivative or a 3'-ATP derivative.

2. The agent of claim 1, wherein the 2'-ATP derivative is according to the formula:



wherein  $R_1$  and  $R_2$  are independently H or  $\text{NH}_2$ ,  $R_3$  is



wherein Y is O or S, and W is O or  $\text{NH}$ , and  $R_4$  and  $R_5$  are independently OH or F.

3. A method of inhibiting STING/ $\text{IFN}\beta$  signaling comprising;

administering to a subject in need thereof a therapeutically effective amount of a STING signaling antagonist.

4. The method of claim 3, wherein the STING signaling antagonist is 2'-ATP or a derivative thereof.

5. The method of claim 3, comprising administering the STING signaling antagonist via a delivery vehicle comprising liposomes, lipid particles, or nanoparticles.

6. The method of claim 3, wherein the subject suffers from an interferonopathy or auto-inflammatory disease.

7. The method of claim 3, wherein the subject suffers from Aicardi-Goutieres syndrome, Lupus erythematosus, STING-associated vasculopathy with onset in infancy (SAVI), inflammatory bowel disease, colitis, or another disease process where STING signaling drives pathology.

8. The method of any one of claims 3 to 7, wherein the agent is the 2'-ATP or derivative thereof, or a 3'-ATP or derivative thereof.

9. The method of any one of claims 3 to 8, further comprising detecting, prior to the administering step, the presence of innate immunity pathway activation, wherein the one or more agents are administered only if innate immunity pathway activation is detected.

10. The method of claim 9, wherein detecting innate immunity activation comprise detecting mutations in one or more of TREG1, ADAR, STING, or other similar genes involved in innate immune signaling.

11. The method of claim 9, wherein the detecting is by sequencing, amplification, hybridization, or CRISPR-based detection.

12. The method of claim 9, wherein detecting innate immunity pathway activation comprises biochemical detection of STING activation.

13. A kit for detecting levels of 2'-ATP, 3'-ATP or derivatives thereof in a sample comprising a first binding molecule that bind 2'-ATP, 3'-ATP or derivatives thereof and a labeled binding molecule that binds the first binding molecule.

14. The kit of claim 13, further comprising a solid substrate capable of absorbing 2'-ATP, 3'-ATP or derivatives thereof.

\* \* \* \* \*