



## (51) International Patent Classification:

C07K 14/47 (2006.01) C12N 9/22 (2006.01)

## (21) International Application Number:

PCT/US2020/014789

## (22) International Filing Date:

23 January 2020 (23.01.2020)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

62/795,912 23 January 2019 (23.01.2019) US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, MA 02142 (US). **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **KIM, Yongjoo**; 605 W Chapel Hill Street, Apartment 704, Durham, NC 27701 (US). **LIU, David, R.**; 3 Whitman Circle, Lexington, MA 02420 (US). **ZHAO,**

**Kevin, Tianmeng**; 3510 S. Seeley Avenue, Unit A, Chicago, IL 60609 (US).

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

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

## (54) Title: SUPERNEGATIVELY CHARGED PROTEINS AND USES THEREOF

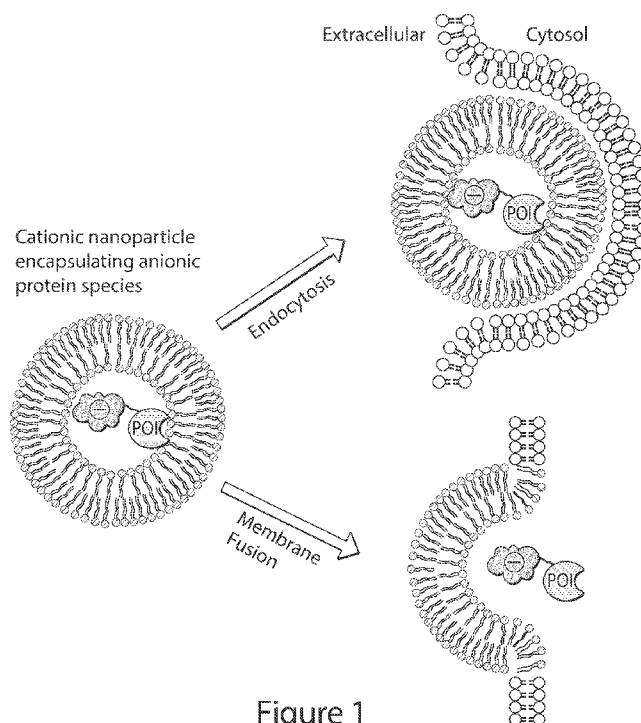


Figure 1

(57) Abstract: Provided herein are compositions, systems, and methods for delivering an effector protein into a cell. The present disclosure, in some aspects, provide novel proteins delivering an effector protein into a cell. The novel proteins are supernegatively charged proteins derived from highly anionic proteins identified from the proteome (e.g., human proteome). The novel protein tags can be associated (e.g., covalently or noncovalently) with the protein to be delivered to facilitate delivery of the effector protein into a cell.



EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,  
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,  
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,  
KM, ML, MR, NE, SN, TD, TG).

**Published:**

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

## **SUPERNEGATIVELY CHARGED PROTEINS AND USES THEREOF**

### **RELATED APPLICATIONS**

**[0001]** This application claims the benefit under 35 U.S.C. § 119(e) to U.S. Provisional Application No. 62/795912, filed January 23, 2019, and entitled “SUPERNEGATIVELY CHARGED PROTEINS AND USES THEREOF,” the entire contents of which are incorporated herein by reference.

### **GOVERNMENT SUPPORT**

**[0002]** This invention was made with government support under grant no. HR0011-17-2-0049, awarded by the Defense Advanced Research Projects Agency (DARPA), and grant nos. RM1 HG009490, R01 EB022376, U01 AI142756, and R35 GM118062, awarded by the National Institutes of Health. The government has certain rights in this invention.

### **BACKGROUND OF THE INVENTION**

**[0003]** Proteins including genome editing agents represent an increasing proportion of biomedical research tools and new human therapeutics. Due to their inability to spontaneously cross the lipid bilayer, however, current uses of protein tools and therapeutic agents are largely limited to targeting extracellular components. Technologies to facilitate cytosolic access by proteins are critical to expanding the potential targets that can be accessed by exogenous proteins.

### **SUMMARY OF THE INVENTION**

**[0004]** The present disclosure, in some aspects, provide novel proteins delivering an effector protein into a cell. The novel proteins are supernegatively charged proteins derived from highly anionic proteins identified from the proteome (e.g., human proteome). The novel protein tags can be associated (e.g., covalently or noncovalently) with the protein to be delivered to facilitate delivery of the effector protein into a cell. Compositions and methods of delivering proteins into a cell (e.g., a cultured cell or a cell *in vivo*) using the novel protein tags and also provided. The supernegatively charged proteins described herein are selected from the group consisting of prothymosin alpha (ProTα), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and variants thereof. In some embodiments, the supernegatively charged proteins are human proteins.

**[0005]** Accordingly, some aspects of the present disclosure provide compositions comprising a cationic lipid or a cationic polymer and a supernegatively charged protein associated with an effector protein, wherein the supernegatively charged protein is selected from the group consisting of prothymosin alpha (ProT $\alpha$ ), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and variants thereof.

**[0006]** In some embodiments, the supernegatively charged protein is a human protein. In some embodiments, the supernegatively charged protein is human prothymosin alpha (ProT $\alpha$ ) or a variant thereof. In some embodiments, the supernegatively charged protein comprises the amino acid sequence of any one of SEQ ID NOs: 1-3. In some embodiments, the supernegatively charged protein is human DPH3 homolog or a variant thereof. In some embodiments, the supernegatively charged protein comprises the amino acid sequence of SEQ ID NO: 4. In some embodiments, the supernegatively charged protein is human ADP ribosylation factor-like protein 2-binding protein or a variant thereof. In some embodiments, the supernegatively charged protein comprises the amino acid sequence of SEQ ID NO: 5. In some embodiments, the supernegatively charged protein is human protein S100-B or a variant thereof. In some embodiments, the supernegatively charged protein comprises the amino acid sequence of SEQ ID NO: 6. In some embodiments, the supernegatively charged protein is human Sirtuin-1 or a variant thereof. In some embodiments, the supernegatively charged protein comprises the amino acid sequence of SEQ ID NO: 7.

**[0007]** In some embodiments, the effector protein is an enzyme, a transcriptional regulator, a therapeutic protein, or a diagnostic protein. In some embodiments, the enzyme is a recombinase, a nuclease, or an epigenetic modifier. In some embodiments, the effector protein is a recombinase. In some embodiments, the recombinase is a Cre recombinase, a Gin recombinase, or a Tn3 recombinase. In some embodiments, the effector protein is a nuclease. In some embodiments, the nuclease is a zinc-finger nuclease (ZFN), a transcription activator-like effector nucleases (TALEN), or a RNA-guided nuclease. In some embodiments, the RNA-guided nuclease is a Cas9 nuclease, a variant, or a fusion protein thereof. In some embodiments, the effector protein is a transcriptional regulator. In some embodiments, the transcriptional regulator is a transcriptional activator or a transcriptional repressor. In some embodiments, the transcriptional activator is selected from the group consisting of VP16, VP64, and p65. In some embodiments, the transcriptional repressor is a KRAB protein or a SID protein. In some embodiments, the effector protein is a therapeutic protein or a diagnostic protein.

[0008] In some embodiments, the supernegatively charged protein is covalently associated with the effector protein. In some embodiments, the supernegatively charged protein is associated with the effector protein via a linker. In some embodiments, the linker is a cleavable linker. In some embodiments, the linker is a UV-cleavable linker or a linker that is cleaved by a lysosomal enzyme. In some embodiments, the supernegatively charged protein is non-covalently associated with the effector protein.

[0009] In some embodiments, the composition is a pharmaceutical composition. In some embodiments, the pharmaceutical composition comprises a pharmaceutically acceptable excipient. In some embodiments, the composition is formulated for administration to a subject to deliver the effector protein to the subject. In some embodiments, the composition is administered to the subject for diagnosing or treating a disease.

[0010] In some embodiments, the cationic lipid is selected from Lipofectamine® 2000, Lipofectamine® 3000, Lipofectamine® RNAiMAX, Lipofectamine® LTX, Lipofectamine® CRISPRMAX, Lipofectamine® MessengerMax, and FuGENE HD.

[0011] Further provided here are fusion proteins comprising a supernegatively charged protein fused to an effector protein, wherein the supernegatively charged protein is selected from the group consisting of prothymosin alpha (ProTα), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and functional variants thereof.

[0012] Further provided herein are nucleic acids encoding the fusion protein described herein, vectors comprising such nucleic acid, and cells comprising such fusion protein, such nucleic acid, or such vector. Compositions comprising the fusion protein described herein and a cationic lipid or cationic polymer are also provided. Complexes comprising a supernegatively charged protein associated with an effector protein are provided, wherein the supernegatively charged protein is selected from the group consisting of prothymosin alpha (ProTα), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and variants thereof.

[0013] Other aspects of the present disclosure provide methods comprising contacting a cell with the composition described herein, whereby the contacting results in the delivery of the effector protein into the cell. In some embodiments, the method further comprises detecting the effector protein in the cell. In some embodiments, the contacting is *in vitro*. In some embodiments, the contacting is *in vivo*. In some embodiments, the contacting is *ex vivo*. In some embodiments, the cell is a mammalian cell. In some embodiments, the mammalian cell is a human cell.

[0014] In some embodiments, the composition exhibits low toxicity when contacted with a population of cells. In some embodiments, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the cells are viable following contacting the composition.

[0015] Also provided herein are method sof treating a disease comprising administering to a subject in need thereof an effective amount of the composition described herein, wherein the effector protein is a therapeutic protein.

[0016] Also provided herein are methods of diagnosing a disease comprising administering to a subject in need thereof an effective amount of the composition of or the fusion protein described herein, wherein the effector protein is a diagnostic protein. In some embodiments, the step of administering comprises a route of administration selected from the group consisting of oral, intravenous, intramuscular, intra-arterial, subcutaneous, intraventricular, topical, inhalational, and mucosal delivery. In some embodiments, the effector protein enters at least one cell in the subject.

[0017] The details of one or more embodiments of the invention are set forth in the Detailed Description of Certain Embodiments, as described below. Other features, objects, and advantages of the invention will be apparent from the Definitions, Examples, and Claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The accompanying drawings, which constitute a part of this specification, illustrate several embodiments of the invention and together with the description, serve to explain the principles of the invention. For purposes of clarity, not every component may be labeled in every drawing. The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee. In the drawings:

[0019] **Figure 1** depicts cationic lipid-mediated protein delivery. A protein of interest (POI) fused to an anionic protein is encapsulated by cationic lipids. The resulting nanoparticles are localized to the cell membrane, followed by endocytosis. Membrane fusion may contribute to functional delivery.

[0020] **Figure 2** shows that ProT $\alpha$  enables efficient cationic lipid-mediated protein delivery into HeLa-DsRed, BSR-tdTomato, and HEK293-RFP cell lines. EC<sub>50</sub> is defined as the protein concentration at which 50% of analyzed cells contain red fluorescence. In all cell lines tested, ProT $\alpha$  delivers Cre at low nM concentrations, ~10-fold more potent than (–30)GFP and ~100-fold more efficient than unmodified Cre.

[0021] **Figures 3A-3B** show that ProTα delivery potency cannot be recapitulated by simple oligomers of acidic residues. *Figure 3A* shows sequences of ProTα (SEQ ID NO: 88) and charge analogs that contain –30 (SEQ ID NO: 89) or –44 (SEQ ID NO: 90) charge for comparison. *Figure 3B* shows that despite containing the same amount of charge as that of ProTα, (–44)PolyD/E-Cre does not promote efficient protein delivery. EC<sub>50</sub> is defined as the protein concentration at which 50% of analyzed cells contain red fluorescence.

[0022] **Figures 4A-4B** depict the truncation of ProTα. *Figure 4A* illustrates schematics of several truncated variants of ProTα-Cre. Regions of deletion are denoted in black. *Figure 4B* shows delivery efficiencies of truncated variants. Removal of the central region containing stretches of acidic residues (B3, B5) results in a catastrophic reduction in delivery efficiency. Deletion of either the N-terminus or the C-terminus of ProTα (B4, B2) slightly reduces delivery efficiency, suggesting that the termini also play a role in the delivery process. EC<sub>50</sub> is defined as the protein concentration at which 50% of analyzed cells contain red fluorescence.

[0023] **Figure 5** shows the effect of various endocytosis inhibitors on ProTα- and (–30)GFP-mediated lipid nanoparticle delivery. Cationic nanoparticles containing either ProTα-Cre fusion protein or (–30)GFP-Cre fusion protein was delivered to HEK293-RFP cells that had been pre-treated with inhibitors that block various endocytosis pathways. After 2 days, delivery efficiency was measured via flow cytometry, and the percent of cells containing red fluorescence was quantified and normalized against a no-inhibitor control. Chlorpromazine (CPZ) and wortmannin (WTM) modestly decreased delivery efficiency, whereas MBCD significantly abrogated Cre recombination.

[0024] **Figures 6A-6B** show the delivery of zinc finger nucleases (ZFNs) using ProTα. *Figure 6A* illustrates a schematic of ProTα-fused ZFN. *Figure 6B* shows ZFNs targeting the AAVS1 site in HEK293T cells were delivered using Lipofectamine RNAiMAX. ProTα enables efficient delivery of ZFNs and induce indels at a mid-nanomolar concentration range. Both left and right ZFN components are required for efficient indel generation.

[0025] **Figure 7.** Identification of proteins from the human proteome with high negative charge-to-MW ratio. Blue dots represent various surface-modified GFPs. The different lines in the graph represent hypothetical cutoffs for proteins with charge-to-MW (kDa) ratio of 0.75, 0.9, 1.1, respectively. Proteins with charge-to-MW ratio greater than 0.5 were manually inspected.

[0026] **Figures 8A-8F** show examples of flow cytometry analysis of HEK293-RFP Cre reporter cells treated with (–30)GFP-Cre, ProTα-Cre, or Cre at 0.25 nM (*Figures 8A, 8B, and 8C, respectively*) and 5 nM (*Figures 8D, 8E, and 8F, respectively*) concentrations.

[0027] **Figure 9** depicts microscopy of HEK293-RFP Cre reporter cells treated with ProTα-Cre, (-30)GFP-Cre, or Cre at 0.25 nM and 5 nM concentrations. While both ProTα-Cre and (-30)GFP-Cre facilitate efficient delivery of Cre at 5 nM, only ProTα-Cre enables efficient delivery at 0.25 nM protein concentration.

[0028] **Figure 10** shows the cytotoxicity of ProTα- and Lipofectamine RNAiMAX-mediated Cre delivery to HEK293-RFP cells. Trypan blue was used to quantify the percentage of live cells at the end of the experiment for three different protein constructs. Minimal toxicity was observed for all cases.

[0029] **Figure 11** shows that Cre does not benefit from an additional NLS. Simple attachment of an SV40 NLS to Cre (NLSCre) does not significantly enhance the percentage of red fluorescent cells when delivered to HEK293-RFP cells using Lipofectamine RNAiMAX.

[0030] **Figure 12** shows the cytotoxicity of HEK293-RFP cells with various endocytosis inhibitors after delivery of ProTα-Cre or (-30)GFP-Cre. Various endocytosis inhibitors were used as described in the Methods section of Example 1, below. After the experiment, % live cells were determined by a Trypan blue assay. At the doses used in the experiment, most inhibitors did not result in significant cell death.

[0031] **Figures 13A-13B** show that fusion of ProTα does not affect ZFN activity. **Figure 13A** depicts an *in vitro* DNA cleavage assay of ZFN and ProTα-ZFN on purified DNA fragment containing a human AAVS1 sequence. **Figure 13B** depicts plasmid transfection of various constructs under a constitutive CMV promoter in HEK293T cells. ZFN and ProTα-ZFN both induce equally significant indels when delivered in plasmid form.

[0032] **Figure 14** depicts the optimization of protein and lipid concentrations used for ZFN delivery. Various concentrations of left and right ProTα-ZFNs were complexed with a range of Lipofectamine RNAiMAX reagents and delivered into HEK293T cells. After 2 days, indel percentage was quantified by HTS. Maximal indel was achieved with protein concentration of 200 nM and 3-3.5 μL of lipid per well.

[0033] **Figure 15** shows the cytotoxicity of ZFN protein delivery. HEK293T cells were treated with ZFN variants with or without lipid complexation. 2 days after treatment, Cell Titer Glo was used to quantify the amount of cells in each well, normalized to a well without treatment. No significant cytotoxicity was observed from any particular treatment conditions.

## DEFINITIONS

**[0034]** As used herein and in the claims, the singular forms “a,” “an,” and “the” include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to “an agent” includes a single agent and a plurality of agents.

**[0035]** The term “associated with” as used herein in the context of two or more moieties (*e.g.*, proteins or protein domains) refers to the fact that the moieties are physically associated with or connected to one another, either directly or via one or more additional moieties that serve as a linking agent, to form a structure that is sufficiently stable so that the moieties remain physically associated under the conditions in which the structure is used, *e.g.*, under physiological conditions. A supernegatively charged protein may be associated with an effector protein through non-covalent interactions (*e.g.*, electrostatic interactions). In some embodiments, a supernegatively charged protein may be associated with an effector protein through electrostatic interactions to form a complex. In some embodiments, a sufficient number of weaker interactions can provide sufficient stability for moieties to remain physically associated under a variety of different conditions. In some embodiments, a supernegatively charged protein is associated with an effector protein via a covalent bond (*e.g.*, an amide bond). In some embodiments, an effector protein is associated with a supernegatively charged protein directly by a peptide bond, or indirectly via a linker.

**[0036]** The term “fusion protein” refers to a protein comprising a plurality of heterologous proteins, protein domains, or peptides, *e.g.*, a supernegatively charged protein and an effector protein, associated with each other via a peptide linkage, thus forming a single amino acid sequence. In some embodiments, a fusion protein is encoded by a gene.

**[0037]** The term “supernegatively charged protein” refers to a protein that has a net negative charge of at least -1, at least -2, at least -3, at least -4, at least -5, at least -6, at least 7, at least 8, at least -9, at least -10, at least -15, at least -20, at least -25, at least -30, at least -35, at least -40, at least -45, at least -50, or more. Supernegatively charged proteins provided in the present disclosure include: prothymosin alpha (ProTα), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and variants thereof.

**[0038]** The term “effector protein” refers to a protein that modulates a biological function of a cell when introduced into the cell, *e.g.*, a modification of a nucleic acid molecule in the cell (such as a cleavage, deamination, recombination, *etc.*), or a modulation (*e.g.*, increases or decreases) the expression or the expression level of a gene in the cell.

**[0039]** The term “nuclease,” as used herein, refers to an agent, for example, a protein, nucleic acid, or a small molecule, capable of cleaving a phosphodiester bond connecting nucleotide residues in a nucleic acid molecule. In some embodiments, a nuclease is a protein, *e.g.*, an

enzyme that can bind a nucleic acid molecule and cleave a phosphodiester bond connecting nucleotide residues within the nucleic acid molecule. A nuclease may be an endonuclease, cleaving a phosphodiester bond within a polynucleotide chain, or an exonuclease, cleaving a phosphodiester bond at the end of the polynucleotide chain. In some embodiments, a nuclease is a site-specific nuclease, binding and/or cleaving a specific phosphodiester bond within a specific nucleotide sequence, which is also referred to herein as the “recognition sequence,” the “nuclease target site,” or the “target site.” In some embodiments, a nuclease recognizes a single stranded target site, while in other embodiments, a nuclease recognizes a double-stranded target site, for example, a double-stranded DNA target site. The target sites of many naturally occurring nucleases, for example, many naturally occurring DNA restriction nucleases, are well known to those of skill in the art. In many cases, a DNA nuclease, such as EcoRI, HindIII, or BamHI, recognize a palindromic, double-stranded DNA target site of 4 to 10 base pairs in length, and cut each of the two DNA strands at a specific position within the target site. Some endonucleases cut a double-stranded nucleic acid target site symmetrically, *i.e.*, cutting both strands at the same position so that the ends comprise base-paired nucleotides, also referred to herein as blunt ends. Other endonucleases cut a double-stranded nucleic acid target site asymmetrically, *i.e.*, cutting each strand at a different position so that the ends comprise unpaired nucleotides. Unpaired nucleotides at the end of a double-stranded DNA molecule are also referred to as “overhangs,” *e.g.*, as “5'-overhang” or as “3'-overhang,” depending on whether the unpaired nucleotide(s) form(s) the 5' or the 3' end of the respective DNA strand. Double-stranded DNA molecule ends ending with unpaired nucleotide(s) are also referred to as sticky ends, as they can “stick to” other double-stranded DNA molecule ends comprising complementary unpaired nucleotide(s). A nuclease protein typically comprises a “binding domain” that mediates the interaction of the protein with the nucleic acid substrate, and a “cleavage domain” that catalyzes the cleavage of the phosphodiester bond within the nucleic acid backbone. In some embodiments, a nuclease protein can bind and cleave a nucleic acid molecule in a monomeric form, while, in other embodiments, a nuclease protein has to dimerize or multimerize in order to cleave a target nucleic acid molecule. Binding domains and cleavage domains of naturally occurring nucleases, as well as modular binding domains and cleavage domains that can be combined to create nucleases that bind specific target sites, are well known to those of skill in the art. For example, transcriptional activator like elements can be used as binding domains to specifically bind a desired target site, and fused or conjugated to a cleavage domain, for example, the cleavage domain of FokI, to create an engineered nuclease cleaving the desired target site.

**[0040]** The term “cationic lipid” refers to a lipid which has a cationic, or positive, charge at physiologic pH. Cationic lipids can take a variety of forms including, but not limited to, liposomes or micelles. Cationic lipids useful for certain aspects of the present disclosure are known in the art, and, generally comprise both polar and non-polar domains, bind to polyanions, such as nucleic acid molecules or supernegatively charged proteins, and are typically known to facilitate the delivery of nucleic acids into cells. Examples of useful cationic lipids include polyethylenimine, polyamidoamine (PAMAM) starburst dendrimers, Lipofectin (a combination of DOTMA and DOPE), Lipofectase, LIPOFECTAMINE® (*e.g.*, LIPOFECTAMINE® 2000, LIPOFECTAMINE® 3000, LIPOFECTAMINE® RNAiMAX, LIPOFECTAMINE® LTX Lipofectamine® MessengerMax), SAINT-RED (Synvolux Therapeutics, Groningen Netherlands), FuGENE HD DOPE (Promega), Cytofectin (Gilead Sciences, Foster City, Calif.), and Eufectins (JBL, San Luis Obispo, Calif.). Exemplary cationic liposomes can be made from N-[1-(2,3-dioleoloxo)-propyl]-N,N,N-trimethylammonium chloride (DOTMA), N-[1-(2,3-dioleoloxo)-propyl]-N,N,N-trimethylammonium methylsulfate (DOTAP), 3β-[N-(N',N'-dimethylaminoethane)carbamoyl]cholesterol (DC-Chol), 2,3,-dioleoyloxy-N-[2(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate (DOSPA), 1,2-dimyristyloxypropyl-3-dimethyl-hydroxyethyl ammonium bromide; and dimethyldioctadecylammonium bromide (DDAB). Cationic lipids have been used in the art to deliver nucleic acid molecules to cells (see, *e.g.*, U.S. Pat. Nos. 5,855,910; 5,851,548; 5,830,430; 5,780,053; 5,767,099; 8,569,256; 8,691,750; 8,748,667; 8,758,810; 8,759,104; 8,771,728; Lewis *et al.* 1996. *Proc. Natl. Acad. Sci. USA* 93:3176; Hope *et al.* 1998. *Molecular Membrane Biology* 15:1). In addition, other lipid compositions are also known in the art and include, *e.g.*, those taught in U.S. Pat. No. 4,235,871; U.S. Pat. Nos. 4,501,728; 4,837,028; 4,737,323.

**[0041]** The term “cationic polymer,” as used herein, refers to a polymer having a net positive charge. Cationic polymers are well known in the art, and include those described in Samal *et al.*, Cationic polymers and their therapeutic potential. *Chem Soc Rev.* 2012 Nov 7;41(21):7147-94; in published U.S. patent applications U.S. 2014/0141487 A1, U.S. 2014/0141094 A1, U.S. 2014/0044793 A1, U.S. 2014/0018404 A1, U.S. 2014/0005269 A1, and U.S. 2013/0344117 A1; and in U.S. Pat. Nos. 8,709,466; 8,728,526; 8,759,103; and 8,790,664; the entire contents of each are incorporated herein by reference. Exemplary cationic polymers include, but are not limited to, polyallylamine (PAH); polyethyleneimine (PEI); poly(L-lysine) (PLL); poly(L-arginine) (PLA); polyvinylamine homo- or copolymer; a poly(vinylbenzyl-tri-C<sub>1</sub>-C<sub>4</sub>-

alkylammonium salt); a polymer of an aliphatic or araliphatic dihalide and an aliphatic N,N,N',N'-tetra-C<sub>1</sub>-C<sub>4</sub>-alkyl-alkylenediamine; a poly(vinylpyridin) or poly(vinylpyridinium salt); a poly(N,N-diallyl-N,N-di-C<sub>1</sub>-C<sub>4</sub>-alkyl-ammoniumhalide); a homo- or copolymer of a quaternized di-C<sub>1</sub>-C<sub>4</sub>-alkyl-aminoethyl acrylate or methacrylate; POLYQUAD<sup>TM</sup>; a polyaminoamide; and the like.

**[0042]** The terms “transcriptional activator” and “transcriptional repressor” refer to an agent such as a protein (*e.g.*, a transcription factor or fragment thereof), that binds a target nucleic acid sequence and causes an increase or decrease of the level of expression of a gene product associated with the target nucleic acid sequence, respectively. For example, if the target nucleic acid sequence is located within a regulatory region of a gene, a transcriptional activator causes an increase of the level of expression of a gene product encoded by the gene (conversely, a transcriptional repressor causes a decrease of the level of expression of a gene product encoded by the gene). The gene product can be an RNA transcribed from the gene (*e.g.*, an mRNA) or a polypeptide translated from an mRNA transcribed from the gene. Typically an increase or decrease in the level of an mRNA results in an increase or decrease in the level of a polypeptide translated therefrom. The level of expression may be determined using standard techniques for measuring mRNA or protein.

**[0043]** The term “Transcriptional Activator-Like Effector,” (TALE) as used herein, refers to effector proteins comprising a DNA binding domain, which contains a highly conserved 33-34 amino acid sequence comprising a highly variable two-amino acid motif (Repeat Variable Diresidue, RVD). The RVD motif determines binding specificity to a nucleic acid sequence, and can be engineered according to methods well known to those of skill in the art to specifically bind a desired DNA sequence (see, *e.g.*, Miller, Jeffrey; *et.al.* (February 2011). “A TALE nuclease architecture for efficient genome editing”. *Nature Biotechnology* **29** (2): 143–8; Zhang, Feng; *et.al.* (February 2011). “Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription”. *Nature Biotechnology* **29** (2): 149–53; Geißler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. “Transcriptional Activators of Human Genes with Programmable DNA-Specificity”. *PLoS ONE* **6** (5): e19509; Boch, Jens (February 2011). “TALEs of genome targeting”. *Nature Biotechnology* **29** (2): 135–6; Boch, Jens; *et.al.* (December 2009). “Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors”. *Science* **326** (5959): 1509–12; and Moscou, Matthew J.; Adam J. Bogdanove (December 2009). “A Simple Cipher Governs DNA Recognition by TAL Effectors”. *Science* **326** (5959): 1501; the entire contents of each of which are incorporated

herein by reference). The simple relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DNA binding domains by selecting a combination of repeat segments containing the appropriate RVDs. TALE effector proteins include, without limitation, TALE nucleases (TALENs) and TALE transcriptional activators and repressors.

**[0044]** The term “transcriptional repressor” refers to a transcription factor, *e.g.*, a protein, that binds a target nucleic acid sequence and causes a reduction of the level of expression of a gene product associated with the target nucleic acid sequence. For example, if the target nucleic acid sequence is located within a regulatory region of a gene, a transcriptional repressor causes a reduction of the level of expression of a gene product encoded by the gene. The gene product can be an RNA transcribed from the gene (*e.g.*, an mRNA) or a polypeptide translated from an mRNA transcribed from the gene. Typically a reduction in the level of an mRNA results in a reduction in the level of a polypeptide translated therefrom. The level of expression may be determined using standard techniques for measuring mRNA or protein.

**[0045]** The term “Transcriptional Activator-Like Element Nuclease,” (TALEN) as used herein, refers to an artificial nuclease comprising a transcriptional activator like effector DNA binding domain to a DNA cleavage domain, for example, a FokI domain. A number of modular assembly schemes for generating engineered TALE constructs have been reported (Zhang, Feng; *et.al.* (February 2011). “Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription”. *Nature Biotechnology* **29** (2): 149–53; Geißler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. “Transcriptional Activators of Human Genes with Programmable DNA-Specificity”. *PLoS ONE* **6** (5): e19509; Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V. *et al.* (2011). “Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting”. *Nucleic Acids Research*; Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). “Assembly of custom TALE-type DNA binding domains by modular cloning”. *Nucleic Acids Research*; Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). “Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes”. *Nucleic Acids Research*; Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). Bendahmane, Mohammed. ed. “Assembly of Designer TAL Effectors by Golden Gate Cloning”. *PLoS ONE* **6** (5): e19722; each of which is incorporated herein by reference).

[0046] The term “zinc finger nuclease,” as used herein, refers to a nuclease comprising a nucleic acid cleavage domain conjugated to a binding domain that comprises a zinc finger array. In some embodiments, the cleavage domain is the cleavage domain of the type II restriction endonuclease FokI. Zinc finger nucleases can be designed to target virtually any desired sequence in a given nucleic acid molecule for cleavage, and the possibility to design zinc finger binding domains to bind unique sites in the context of complex genomes allows for targeted cleavage of a single genomic site in living cells, for example, to achieve a targeted genomic alteration of therapeutic value. Targeting a double-strand break to a desired genomic locus can be used to introduce frame-shift mutations into the coding sequence of a gene due to the error-prone nature of the non-homologous DNA repair pathway. Zinc finger nucleases can be generated to target a site of interest by methods well known to those of skill in the art. For example, zinc finger binding domains with a desired specificity can be designed by combining individual zinc finger motifs of known specificity. The structure of the zinc finger protein Zif268 bound to DNA has informed much of the work in this field and the concept of obtaining zinc fingers for each of the 64 possible base pair triplets and then mixing and matching these modular zinc fingers to design proteins with any desired sequence specificity has been described (Pavletich NP, Pabo CO (May 1991). “Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å”. *Science* **252** (5007): 809–17, the entire contents of which are incorporated herein). In some embodiments, separate zinc fingers that each recognizes a 3 base pair DNA sequence are combined to generate 3-, 4-, 5-, or 6-finger arrays that recognize target sites ranging from 9 base pairs to 18 base pairs in length. In some embodiments, longer arrays are contemplated. In other embodiments, 2-finger modules recognizing 6-8 nucleotides are combined to generate 4-, 6-, or 8- zinc finger arrays. In some embodiments, bacterial or phage display is employed to develop a zinc finger domain that recognizes a desired nucleic acid sequence, for example, a desired nuclease target site of 3-30 bp in length. Zinc finger nucleases, in some embodiments, comprise a zinc finger binding domain and a cleavage domain fused or otherwise conjugated to each other via a linker, for example, a polypeptide linker. The length of the linker determines the distance of the cut from the nucleic acid sequence bound by the zinc finger domain. If a shorter linker is used, the cleavage domain will cut the nucleic acid closer to the bound nucleic acid sequence, while a longer linker will result in a greater distance between the cut and the bound nucleic acid sequence. In some embodiments, the cleavage domain of a zinc finger nuclease has to dimerize in order to cut a bound nucleic acid. In some such embodiments, the dimer is a heterodimer of two monomers, each of which comprise a different zinc finger binding domain.

For example, in some embodiments, the dimer may comprise one monomer comprising zinc finger domain A conjugated to a FokI cleavage domain, and one monomer comprising zinc finger domain B conjugated to a FokI cleavage domain. In this non-limiting example, zinc finger domain A binds a nucleic acid sequence on one side of the target site, zinc finger domain B binds a nucleic acid sequence on the other side of the target site, and the dimerize FokI domain cuts the nucleic acid in between the zinc finger domain binding sites.

**[0047]** The term “zinc finger,” as used herein, refers to a small nucleic acid-binding protein structural motif characterized by a fold and the coordination of one or more zinc ions that stabilize the fold. Zinc fingers encompass a wide variety of differing protein structures (see, *e.g.*, Klug A, Rhodes D (1987). “Zinc fingers: a novel protein fold for nucleic acid recognition”. *Cold Spring Harb. Symp. Quant. Biol.* 52: 473–82, the entire contents of which are incorporated herein by reference). Zinc fingers can be designed to bind a specific sequence of nucleotides, and zinc finger arrays comprising fusions of a series of zinc fingers, can be designed to bind virtually any desired target sequence. Such zinc finger arrays can form a binding domain of a protein, for example, of a nuclease, *e.g.*, if conjugated to a nucleic acid cleavage domain. Different types of zinc finger motifs are known to those of skill in the art, including, but not limited to, Cys<sub>2</sub>His<sub>2</sub>, Gag knuckle, Treble clef, Zinc ribbon, Zn<sub>2</sub>/Cys<sub>6</sub>, and TAZ2 domain-like motifs (see, *e.g.*, Krishna SS, Majumdar I, Grishin NV (January 2003). “Structural classification of zinc fingers: survey and summary”. *Nucleic Acids Res.* 31 (2): 532–50). Typically, a single zinc finger motif binds 3 or 4 nucleotides of a nucleic acid molecule. Accordingly, a zinc finger domain comprising 2 zinc finger motifs may bind 6-8 nucleotides, a zinc finger domain comprising 3 zinc finger motifs may bind 9-12 nucleotides, a zinc finger domain comprising 4 zinc finger motifs may bind 12-16 nucleotides, and so forth. Any suitable protein engineering technique can be employed to alter the DNA-binding specificity of zinc fingers and/or design novel zinc finger fusions to bind virtually any desired target sequence from 3 – 30 nucleotides in length (see, *e.g.*, Pabo CO, Peisach E, Grant RA (2001). “Design and selection of novel cys<sub>2</sub>His<sub>2</sub> Zinc finger proteins”. *Annual Review of Biochemistry* 70: 313–340; Jamieson AC, Miller JC, Pabo CO (2003). “Drug discovery with engineered zinc-finger proteins”. *Nature Reviews Drug Discovery* 2 (5): 361–368; and Liu Q, Segal DJ, Ghiara JB, Barbas CF (May 1997). “Design of polydactyl zinc-finger proteins for unique addressing within complex genomes”. *Proc. Natl. Acad. Sci. U.S.A.* 94 (11); the entire contents of each of which are incorporated herein by reference). Fusions between engineered zinc finger arrays and protein domains that cleave a nucleic acid can be used to generate a “zinc finger nuclease.” A zinc finger nuclease typically comprises a zinc finger domain that

binds a specific target site within a nucleic acid molecule, and a nucleic acid cleavage domain that cuts the nucleic acid molecule within or in proximity to the target site bound by the binding domain. Typical engineered zinc finger nucleases comprise a binding domain having between 3 and 6 individual zinc finger motifs and binding target sites ranging from 9 base pairs to 18 base pairs in length. Longer target sites are particularly attractive in situations where it is desired to bind and cleave a target site that is unique in a given genome.

**[0048]** The terms “RNA-programmable nuclease” and “RNA-guided nuclease” are used interchangeably herein and refer to a nuclease that forms a complex with (*e.g.*, binds or associates with) one or more RNA molecule that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. RNA-programmable nucleases include Cas9. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). gRNAs can exist as a complex of two or more RNAs, or as a single RNA molecule. gRNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though “gRNA” is used interchangeably to refer to guide RNAs that exist as either single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (*e.g.*, and directs binding of a Cas9 complex to the target); and (2) a domain that binds a Cas9 protein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to said target site and providing the sequence specificity of the nuclease:RNA complex.

**[0049]** The term “Cas9” or “Cas9 nuclease” refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active or inactive DNA cleavage domain of Cas9 or a partially inactive DNA cleavage domain (*e.g.*, a Cas9 “nickase”), and/or the gRNA binding domain of Cas9). In some embodiments, the term “Cas9” refers to a fusion protein comprising Cas9 or a fragment thereof.

**[0050]** In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC\_015683.1, NC\_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC\_016782.1, NC\_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC\_021284.1); *Prevotella intermedia* (NCBI Ref: NC\_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC\_021846.1); *Streptococcus iniae* (NCBI Ref: NC\_021314.1); *Belliella baltica* (NCBI Ref: NC\_018010.1); *Psychroflexus torquis* (NCBI Ref: NC\_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP\_820832.1); *Listeria innocua* (NCBI Ref: NP\_472073.1); *Campylobacter jejuni* (NCBI Ref: YP\_002344900.1); or *Neisseria meningitidis* (NCBI Ref: YP\_002342100.1). In some

embodiments, the Cas9 encompasses to any one of the Cas9 homologues and orthologues known in the art, e.g., as described in Klompe *et al.*, The CRISPR Journal Vol. 1, No. 2, 2018, incorporated herein by reference. In some embodiments, the term “Cas9” also refers to any functional variants, or fusion proteins known to those skilled in the art.

**[0051]** The term “recombinase,” as used herein, refers to a site-specific enzyme that mediates the recombination of DNA between recombinase recognition sequences, which results in the excision, integration, inversion, or exchange (e.g., translocation) of DNA fragments between the recombinase recognition sequences. Recombinases can be classified into two distinct families: serine recombinases (e.g., resolvases and invertases) and tyrosine recombinases (e.g., integrases). Examples of serine recombinases include, without limitation, Hin, Gin, Tn3,  $\beta$ -six, CinH, ParA,  $\gamma\delta$ , Bxb1,  $\phi$ C31, TP901, TG1,  $\phi$ BT1, R4,  $\phi$ RV1,  $\phi$ FC1, MR11, A118, U153, and gp29. Examples of tyrosine recombinases include, without limitation, Cre, FLP, R, Lambda, HK101, HK022, and pSAM2. The serine and tyrosine recombinase names stem from the conserved nucleophilic amino acid residue that the recombinase uses to attack the DNA and which becomes covalently linked to the DNA during strand exchange. Recombinases have numerous applications, including the creation of gene knockouts/knock-ins and gene therapy applications. See, e.g., Brown *et al.*, “Serine recombinases as tools for genome engineering.” *Methods*. 2011;53(4):372-9; Hirano *et al.*, “Site-specific recombinases as tools for heterologous gene integration.” *Appl. Microbiol. Biotechnol.* 2011; 92(2):227-39; Chavez and Calos, “Therapeutic applications of the  $\Phi$ C31 integrase system.” *Curr. Gene Ther.* 2011;11(5):375-81; Turan and Bode, “Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications.” *FASEB J.* 2011; 25(12):4088-107; Venken and Bellen, “Genome-wide manipulations of *Drosophila melanogaster* with transposons, FLP recombinase, and  $\Phi$ C31 integrase.” *Methods Mol. Biol.* 2012; 859:203-28; Murphy, “Phage recombinases and their applications.” *Adv. Virus Res.* 2012; 83:367-414; Zhang *et al.*, “Conditional gene manipulation: Creating a new biological era.” *J. Zhejiang Univ. Sci. B.* 2012; 13(7):511-24; Karpenshif and Bernstein, “From yeast to mammals: recent advances in genetic control of homologous recombination.” *DNA Repair (Amst)*. 2012; 1;11(10):781-8; the entire contents of each are hereby incorporated by reference in their entirety. The recombinases provided herein are not meant to be exclusive examples of recombinases that can be used in embodiments of the invention. The methods and compositions of the invention can be expanded by mining databases for new orthogonal recombinases or designing synthetic recombinases with defined DNA specificities (See, e.g., Groth *et al.*, “Phage integrases: biology and applications.” *J. Mol. Biol.* 2004; 335, 667-678; Gordley *et al.*, “Synthesis of

programmable integrases.” *Proc. Natl. Acad. Sci. U S A.* 2009; 106, 5053-5058; the entire contents of each are hereby incorporated by reference in their entirety). Other examples of recombinases that are useful in the methods and compositions described herein are known to those of skill in the art, and any new recombinase that is discovered or generated is expected to be able to be used in the different embodiments of the invention. In some embodiments, a recombinase (or catalytic domain thereof) is fused to a Cas9 protein (*e.g.*, dCas9).

**[0052]** The term “epigenetic modifier,” as used herein, refers to a protein or catalytic domain thereof having enzymatic activity that results in the epigenetic modification of DNA, for example chromosomal DNA. Epigenetic modifications include, but are not limited to DNA methylation and demethylation; histone modifications including methylation and demethylation (*e.g.*, mono-, di- and tri-methylation), histone acetylation and deacetylation, as well as histone ubiquitylation, phosphorylation, and sumoylation.

**[0053]** The term “linker,” as used herein, refers to a chemical group or a molecule linking two molecules or moieties, *e.g.*, a supernegatively charged protein and a nuclease. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker comprises an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is a cleavable linker, *e.g.*, the linker comprises a bond that can be cleaved upon exposure to a cleaving activity, such as UV light or a hydrolytic enzyme, such as a lysosomal protease. In some embodiments, the linker is any stretch of amino acids having at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, or more amino acids. In some embodiments, the peptide linker comprises repeats of the tri-peptide Gly-Gly-Ser, *e.g.*, comprising the sequence (GGG)<sub>n</sub>, wherein n represents at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more repeats. In some embodiments, the linker comprises the sequence (GGG)<sub>6</sub> (SEQ ID NO: 8). In some embodiments, the peptide linker is the 16 residue “XTEN” linker, or a variant thereof (See, *e.g.*, Schellenberger *et al.* A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat. Biotechnol.* **27**, 1186–1190 (2009)). In some embodiments, the XTEN linker comprises the sequence SGSETPGTSESATPES (SEQ ID NO: 9), SGSETPGTSESA (SEQ ID NO:10), or SGSETPGTSESATPEGGSGGS (SEQ ID NO: 11). In some embodiments, the peptide linker is one or more selected from VPFLLEPDNINGKTC (SEQ ID NO: 12), GSAGSAAGSGEF (SEQ ID NO: 13), SIVAQLSRPDPA (SEQ ID NO: 14), MKIIEQLPSA (SEQ ID NO: 15), VRHKLKRVGS

(SEQ ID NO: 16), GHGTGSTGSGSS (SEQ ID NO: 17), MSRPDPA (SEQ ID NO: 18); or GGSM (SEQ ID NO: 19).

**[0054]** The term “pharmaceutical composition,” as used herein, refers to a composition that can be administrated to a subject, for example, in the context of treatment of a disease or disorder. In some embodiments, a pharmaceutical composition comprises an active ingredient, *e.g.*, a supernegatively charged protein associated with an effector protein, such as a nuclease, or a nucleic acid encoding a supernegatively charged protein and an effector protein, *e.g.*, in the form of a fusion protein, and a pharmaceutically acceptable excipient.

**[0055]** The term “physiological pH” as used herein refers to a pH value that is found in a normal, non-pathologic cell or subject. In some embodiments, physiological pH is between pH 5 – 8. In some embodiments, physiological pH is pH 7-7.5, for example, pH 7.0, pH 7.1, pH 7.2, pH 7.3, pH 7.4, or pH 7.5. In some embodiments, physiological pH is pH 6.5-7.5. In some embodiments, physiological pH is pH 5, pH 5.5, pH 6, pH 6.5, pH 7, pH 7.5, or pH 8.

**[0056]** The term “protein” is interchangeably used herein with the terms “peptide” and “polypeptide” and refers to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof. A protein may comprise different domains, for example, a TALE effector protein may comprise a nucleic acid binding domain and an effector domain, *e.g.*, a nucleic acid cleavage domain or a transcriptional activator or repressor domain. In some embodiments, a protein comprises a proteinaceous part, *e.g.*, an amino acid sequence constituting a nucleic acid binding domain, and an organic compound, *e.g.*, a compound that can act as a nucleic acid cleavage agent.

**[0057]** The term “subject,” as used herein, refers to an individual organism. In some embodiments, the subject is a human of either sex at any stage of development. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a

non-human primate. In some embodiments, the subject is a rodent. In some embodiments, the subject is a laboratory animal, for example, a mouse, a rat, a gerbil, a guinea pig, a fish, a frog, or a fly. In some embodiments, the subject is a farm animal, for example, a sheep, a goat, a pig, or a cattle. In some embodiments, the subject is a companion animal, for example, a cat or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is genetically engineered, *e.g.*, a genetically engineered non-human subject.

**[0058]** The term “effective amount,” as used herein, refers to an amount of a biologically active agent that is sufficient to elicit a desired biological response. For example, in some embodiments, an effective amount of an effector protein (*e.g.*, nucleases, transcriptional activators/repressors, recombinases, Cas9 proteins including variants and fusions thereof, *etc.*) may refer to the amount of the protein that is sufficient to induce a detectable effect (*e.g.*, cleavage of a target site, modification of a target site, modulation of gene expression, *etc.*). Such an effect may be detected in a suitable assay, *e.g.*, in a cell-free assay, or in a target cell, tissue, or subject organism. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, an effector protein, may vary depending on various factors as, for example, on the desired biological response, the specific allele to be targeted, the genome, target site, cell, or tissue being targeted, and the supernegatively charged protein being used.

**[0059]** The terms “treatment,” “treat,” and “treating,” refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms “treatment,” “treat,” and “treating” refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In other embodiments, treatment may be administered in the absence of symptoms, *e.g.*, to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (*e.g.*, in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

#### **DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION**

The present disclosure, in some aspects, provides conjugates, complexes, compositions, preparations, kits, systems, and related methods and uses for the delivery of effector proteins,

*e.g.*, enzymes, transcriptional regulators, therapeutic proteins, or diagnostic proteins, to a cell by associating the effector protein with one or more of the supernegatively charged proteins described herein (*e.g.*, ProT $\alpha$ , DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and variants thereof) and a cationic polymer or cationic lipid. Typically, the effector protein is delivered to the interior of a cell, *e.g.*, to cause a biological effect in the cell. In some embodiments, the biological effect exerts a therapeutic benefit to a subject in which the cell is found. The complexes, conjugates, compositions, preparations, systems, kits, and related methods and uses for delivery of effector proteins are useful for introducing an effector protein into a cell, *e.g.*, in the context of manipulating the cell for research or therapeutic purposes.

**[0060]** Accordingly, some aspects of the present disclosure provide supernegatively charged proteins. Supernegatively charged proteins may be derived from any species of plant, animal, and/or microorganism. In some embodiments, the supernegatively charged protein is a mammalian protein. In some embodiments, the supernegatively charged protein is a human protein. In some embodiments, the protein is derived from an organism typically used in research. For example, the protein to be modified may be from a primate (*e.g.*, ape, monkey), rodent (*e.g.*, rabbit, hamster, gerbil), pig, dog, cat, fish (*e.g.*, *Danio rerio*), nematode (*e.g.*, *C. elegans*), yeast (*e.g.*, *Saccharomyces cerevisiae*), or bacteria (*e.g.*, *E. coli*). In some embodiments, the protein is non-immunogenic. In some embodiments, the protein is non-antigenic. In some embodiments, the protein does not have inherent biological activity or has been modified to have no biological activity.

**[0061]** In some embodiments, the theoretical net charge of the supernegatively protein is least -1, at least -2, at least -3, at least -4, at least -5, at least -10, at least -15, at least -20, at least -25, at least -30, at least -35, at least -40, at least -45, or at least -50. In some embodiments, the supernegatively charged proteins described herein are human proteins. For example, the present disclosure identified several supernegatively charged proteins from the human proteome, including prothymosin alpha (ProT $\alpha$ ), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, and Sirtuin-1. In some embodiments, the supernegatively charged protein comprises the amino acid sequence of any one of SEQ ID NOs: 1-7. In some embodiments, the supernegatively charged protein consists essentially of the amino acid sequence of any one of SEQ ID NOs: 1-7. In some embodiments, the supernegatively charged protein consists of the amino acid sequence of any one of SEQ ID NOs: 1-7.

**[0062]** Variants of the supernegatively charged proteins described herein are also provided. In some embodiments, a variant of a supernegatively charged protein described herein has modifications that changes or does not change its amino acid sequence. For example, in some embodiments, one or more amino acids may be added, deleted, or changed from the primary sequence. For example, a poly-histidine tag or other tag may be added to the supercharged protein to aid in the purification of the protein. Other peptides or proteins may be added onto the supernegatively charged protein to alter the biological, biochemical, and/or biophysical properties of the protein but without altering its functions in delivering the effector protein to a cell. For example, an endosomolytic peptide may be added to the primary sequence of the supernegatively charged protein, or a targeting peptide, may be added to the primary sequence of the supernegatively charged protein. Other modifications of the supernegatively charged protein include, but are not limited to, post-translational modifications (*e.g.*, glycosylation, phosphorylation, acylation, lipidation, farnesylation, acetylation, proteolysis, *etc.*). In some embodiments, the supernegatively charged protein is modified to reduce its immunogenicity. In some embodiments, the supernegatively charged protein is modified to enhance its ability to deliver an effector protein to a cell. In some embodiments, the supernegatively charged protein is modified to enhance its stability. In some embodiments, the supernegatively charged protein is conjugated to a polymer. For example, the protein may be PEGylated by conjugating the protein to a polyethylene glycol (PEG) polymer.

**[0063]** In some embodiments, a functional variant of the supernegatively charged protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or more identical to the amino acid sequence of any one of SEQ ID NOs: 1-7. In some embodiments, a functional variant of the supernegatively charged protein comprises an amino acid sequence that is 70%, 75%, 80%, 85%, 90%, 95%, 99%, or more identical to the amino acid sequence of any one of SEQ ID NOs: 1-7. It is to be understood that, when a “supernegatively charged protein” is referred to herein, it encompasses the wild-type protein and any variants thereof. In some embodiments, the variant has a deletion, addition, or amino acid substitution as compared to a wild type sequence. The amino acid sequences of the supernegatively charged proteins and non-limiting exemplary variants are provided in Table 1.

*Table 1 Exemplary Supernegatively Charged Proteins*

| Supernegatively charged protein | Amino acid sequence | NCBI accession No./Uniprot Accession No. | SEQ ID NO: |
|---------------------------------|---------------------|------------------------------------------|------------|
|---------------------------------|---------------------|------------------------------------------|------------|

|                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |               |   |
|---------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|---|
| ProTα                                                               | MGASDAAVDTSSSEITTKDLKEKKEVVVEEAENGRDA<br>PANGNAENEENGEQEADNEVDEEEEEEGGEEEEEEEE<br>EGDGEEEDGDEDEEAESATGKRAAEDDEDDDDVDT<br>KKQKTDEDDT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 5757/P06454   | 1 |
| ProTα- B2                                                           | MGASDAAVDTSSSEITTKDLKEKKEVVVEEAENGRDA<br>PANGNAENEENGEQEADNEVDEEEEEEGGEEEEEEEE<br>EGDGEEEDGDEDEEAESATGKRAAEDDEDDDDVDT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | N/A           | 2 |
| ProTα-B4                                                            | MGAENEENGEQEADNEVDEEEEEEGGEEEEEEEEEGD<br>GEEEDGDEDEEAESATGKRAAEDDEDDDDVDTKKQ<br>KTDEDD                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | N/A           | 3 |
| DPH3 homolog                                                        | MAVFHDEVEIEDFQYDEDSETYFYPCPCGDNFSITK<br>EDLENGEDVATCPSCSLIKVIYDKDQFVCGETVPAP<br>SANKELVKC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 285381/Q96FX2 | 4 |
| ADP<br>ribosylation<br>factor-like<br>protein 2-<br>binding protein | MDALEGESFALSFSSASDAEFDAVVGYLEDIIMDDE<br>FQLLQRNFMKYYLEFEDTEENKLIYTPIFNEYISLV<br>EKYIEEQLLQRIPEFNMAAFTTTTLQHHKDEVAGDIF<br>DMLLTFTDFLAFKEMFLDYRAEKEGRGLDLSSGLV<br>VTSLCKSSSLPASQNNLRH                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 23568/Q9Y2Y0  | 5 |
| S100B                                                               | MSELEKAMVALIDVFHQYSGREGDKHKLKSELKE<br>LINNELSHFLEEIKEQEVVDKVMETLDNDGDGECDF<br>QEFMAFVAMVTTACHEFFEHE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 6285/P04271   | 6 |
| Sirtuin-1                                                           | MADEAALALQPGGSPSAAGADREAASSPAGEPLRK<br>RPRRDGGLERSPGEPGGAAPEREPVAAARGCPGA<br>AAAALWREAEAEAAAAGGEQEAQATAAAGEGDN<br>GPGQLQGPSREPPLADNLYDEDDDDGEEEEEEAAAA<br>AIGYRDNLLFGDEIITNGFHSCESDEEDRASHASSD<br>WTPRPRIGPYTFVQQHLMIGTDPRTILKDLLPETIPPP<br>ELDDMTLWQIVINILSEPPKRRKKRDINTIEDAVKLL<br>QECKKIIVLTGAGVSVSCGIPDFRSRDGIYARLAVDF<br>PDLDPQAMFDIEYFRKDPRPFFKFAKEIYPGQFQPS<br>LCHKFIASDKEGKLLRNQYTNIDTLEQVAGIQRIIQ<br>CHGSFATASCLICKYKVDCEAVRGDIFNQVVRCP<br>CPADEPLAIMKPEIVFFGENLPEQFHRAMKYDKDEV<br>DLLIVIGSSLKVRPVALIPSSIPHEVPQILINREPLPHL<br>HFDVELLGDCDVIINELCHRLGGEYAKLCCNPVKLS<br>EITEKPPRTQKELAYLSELPTPLHVSEDSSSPERTSP<br>PDSSVIVTLLDQAAKSNDLDVSESKGCMEEKPQE<br>VQTSRNVESIAEQMENPDL<br>KNVGSSTGEKNERTSVAGTVRKCWPNRVAKEQISR<br>RLDGNQYLFLPPNRYIFHGAEVYSDSEDDVLSSSSC<br>GSNSDSGTCQSPSLEEPMEDESEIEEFYNGLEDEPDV<br>PERAGGAGFGTDGDDQEAINAISVKQEVTDNMNYP<br>SNKS | 23411/Q96EB6  | 7 |

[0063] The present disclosure provides systems and methods for the delivery of effector proteins to cells *in vivo*, *ex vivo*, or *in vitro*. Such systems and methods typically involve association of the effector protein with a supernegatively charged protein to form a complex or a fusion protein, and delivery of the complex or fusion protein to a cell. In some embodiments, the effector protein to be delivered by the supernegatively charged protein has therapeutic activity. In some embodiments, delivery of the complex or fusion protein to a cell involves administering the complex or fusion protein comprising a supernegatively charged protein associated with a effector protein to a subject in need thereof.

**[0064]** Effector proteins suitable for delivery to a target cell *in vivo*, *ex vivo*, or *in vitro*, by a system or method provided herein will be apparent to those of skill in the art and include, for example, enzymes (e.g., without limitation, nucleases, recombinases, epigenetic modifiers), transcriptional regulators (e.g., without limitation, transcriptional activators and transcriptional repressors), therapeutic proteins (e.g., without limitation, antibodies, antigens, hormones, cytokines), and diagnostic proteins (e.g., without limitation, fluorescent proteins).

**[0065]** In some embodiments, the effector protein is an enzyme. An enzyme is a substance produced by a living organism that acts as a catalyst to bring about a specific biochemical reaction. Non-limiting examples of enzymes that may be used in accordance with the present disclosure include nucleases, recombinases, and epigenetic modifiers. The delivery of other types of enzymes using the compositions and methods described herein are also contemplated.

**[0066]** In some embodiments, the effector protein is a nuclease. In some embodiments, the effector protein is a transcription activator-like effector nucleases (TALEN). TALE nucleases, or TALENs, are artificial nucleases comprising a transcriptional activator-like effector DNA binding domain associated with a DNA cleavage domain, for example, a FokI domain. A number of modular assembly schemes for generating engineered TALE constructs have been reported (Zhang, Feng; et.al. (February 2011). “Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription”. *Nature Biotechnology* 29 (2): 149–53; Geißler, R.; Scholze, H.; Hahn, S.; Streubel, J.; Bonas, U.; Behrens, S. E.; Boch, J. (2011), Shiu, Shin-Han. ed. “Transcriptional Activators of Human Genes with Programmable DNA-Specificity”. *PLoS ONE* 6 (5): e19509; Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V. et al. (2011). “Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting”. *Nucleic Acids Research*; Morbitzer, R.; Elsaesser, J.; Hausner, J.; Lahaye, T. (2011). “Assembly of custom TALE-type DNA binding domains by modular cloning”. *Nucleic Acids Research*; Li, T.; Huang, S.; Zhao, X.; Wright, D. A.; Carpenter, S.; Spalding, M. H.; Weeks, D. P.; Yang, B. (2011). “Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes”. *Nucleic Acids Research*; Weber, E.; Gruetzner, R.; Werner, S.; Engler, C.; Marillonnet, S. (2011). Bendahmane, Mohammed. ed. “Assembly of Designer TAL Effectors by Golden Gate Cloning”. *PLoS ONE* 6 (5): e19722; the entire contents of each of which are incorporated herein by reference). Those of skill in the art will understand that TALE nucleases can be engineered to target virtually any genomic sequence with high specificity, and that such engineered nucleases can be used in embodiments of the present technology to manipulate the genome of a cell, e.g., by delivering the respective

TALEN via a method or strategy disclosed herein under circumstances suitable for the TALEN to bind and cleave its target sequence within the genome of the cell. In some embodiments, the delivered TALEN targets a gene or allele associated with a disease or disorder. In some embodiments, delivery of the TALEN to a subject confers a therapeutic benefit to the subject. **[0067]** In some embodiments, the effector protein is a zinc finger nuclease (ZFN). Zinc finger nucleases are a class of artificial nucleases that comprise a DNA cleavage domain and a zinc finger DNA binding domain. In some embodiments, the DNA cleavage domain is a non-specific DNA cleavage domain of a restriction endonuclease, for example, of FokI. In some embodiments, the DNA cleavage domain is a domain that only cleaves double-stranded DNA when dimerized with a second DNA cleavage domain of the same type. In some embodiments, the DNA cleavage domain is fused to the C-terminus of the zinc finger domain via a linker, for example, a peptide linker. In some embodiments, the zinc finger domain comprises between about 3 and about 6 zinc fingers and specifically recognizes and binds a target sequence of about 9-20 nucleotides in length. In some embodiments, a plurality of zinc finger nuclease molecules is delivered to a target cell by a system or method provided by this invention, with the zinc finger domain of one zinc finger nuclease molecule binding a target sequence in close proximity of the target sequence of a second zinc finger nuclease molecule. In some embodiments, the zinc finger domains of the zinc finger nuclease molecules binding target sequences in close proximity to each other are different. In some embodiments, a zinc finger nuclease molecule delivered to a cell by a system or method provided herein binds a target nucleic acid sequence in close proximity to the target sequence of another zinc finger nuclease molecule, so that the DNA cleavage domains of the molecules dimerize and cleave a DNA molecule at a site between the two target sequences.

**[0068]** Methods for engineering, generation, and isolation of nucleases targeting specific sequences, e.g., TALE, or zinc finger nucleases, and editing cellular genomes at specific target sequences, are well known in the art (see, e.g., Mani et al., *Biochemical and Biophysical Research Communications* 335:447-457, 2005; Perez et al., *Nature Biotechnology* 26:808-16, 2008; Kim et al., *Genome Research*, 19:1279-88, 2009; Urnov et al., *Nature* 435:646-51, 2005; Carroll et al., *Gene Therapy* 15:1463-68, 2005; Lombardo et al., *Nature Biotechnology* 25:1298-306, 2007; Kandavelou et al., *Biochemical and Biophysical Research Communications* 388:56-61, 2009; and Hockemeyer et al., *Nature Biotechnology* 27(9):851-59, 2009, as well as the reference recited in the respective section for each nuclease). The skilled artisan will be able to ascertain suitable methods for use in the context of the present disclosure based on the guidance provided herein.

**[0069]** In some embodiments, the effector protein is a RNA-guided nuclease. In some embodiments, the RNA guided protein is associated with a guide RNA (gRNA) in the composition described herein. In some embodiments, the RNA-programmable protein is a Cas9 nuclease, a Cas9 variant, or a fusion of a Cas9 protein, which is delivered to a target cell by a system or method provided herein. In some embodiments, the RNA-programmable nuclease is a (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (Csn1) from *Streptococcus pyogenes* (see, e.g., “Complete genome sequence of an M1 strain of *Streptococcus pyogenes*.” Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); “CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.” Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821(2012), the entire contents of each of which are incorporated herein by reference. Because RNA-programmable nucleases (e.g., Cas9) use RNA:DNA hybridization to determine target DNA cleavage sites, these proteins are able to cleave, in principle, any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (e.g., to modify a genome) are known in the art (see e.g., Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013); Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013); Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nature biotechnology 31, 227-229 (2013); Jinek, M. et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013); Dicarlo, J.E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic acids research (2013); Jiang, W. et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology 31, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

**[0070]** A Cas9 nuclease may also be referred to sometimes as a *casn1* nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR

clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand that is not complementary to the crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNA. However, single guide RNAs ("sgRNA", or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See e.g., Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, e.g., "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L. expand/collapse author list McLaughlin R.E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference).

**[0071]** Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, proteins comprising Cas9 proteins or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to wild type Cas9.

In some embodiments, the Cas9 variant comprises a fragment of Cas9 (e.g., a gRNA binding domain or a DNA-cleavage domain, an N-terminal domain or a C-terminal domain, etc.), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to the corresponding fragment of wild type Cas9. In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC\_017053.1). In some embodiments, a Cas9 protein has an inactive (e.g., an inactivated) DNA cleavage domain. A nuclease-inactivated Cas9 protein may interchangeably be referred to as a “dCas9” protein (for nuclease “dead” Cas9). In some embodiments, dCas9 corresponds to, or comprises in part or in whole. In some embodiments, variants of dCas9 are provided. For example, in some embodiments, variants having mutations other than D10A and H840A are provided, which result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H840, or other substitutions within the nuclease domain of Cas9 (e.g., substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain). In some embodiments, variants or homologues of dCas9 are provided which are at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to SEQ ID NO: 22. In some embodiments, variants of dCas9 are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 22, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more. In some embodiments, Cas9 “nickases” are provided which comprise a mutation which inactivates a single nuclease domain in Cas9. Such nickases induce a single strand break in a target nucleic acid as opposed to a double strand break.

#### Cas9

```
ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGG
CGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAAT
ACAGACCGCCACAGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGG
AGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGT
CGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGT
AGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGGAAGAAGACAAGA
AGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAG
AAATATCCAACATATCTATCATCTGCGAAAAAATTGGCAGATTCTACTGATAAAGC
GGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTT
```

TTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAACTATTTATCC  
AGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAG  
AGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAA  
AATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCAT  
TGCTTTGTGCTATTGGGATTGACCCCTAATTTTAAATCAAATTTTGATTGAGCAGAAG  
ATGCTAAATTACAGCTTTCAAAGATACTTACGATGATGATTTAGATAATTTATTG  
GCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGA  
TGCTATTTTACTTTTACAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCC  
TATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTA  
AAAGCTTTAGTTTCGACAACAACCTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCA  
ATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTT  
TATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGT  
GAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCT  
ATTCCCCATCAAATTCCTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGA  
CTTTTATCCATTTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTC  
GAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATG  
ACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTGCGATA  
AAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAACTTTGATAAAAAATCTT  
CCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTA  
TAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATT  
CTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAA  
AAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAAATAGAATGTTTTGA  
TAGTGTTGAAATTTTCAAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACC  
ATGATTTGCTAAAAATTATTAAGATAAAGATTTTTTGGATAATGAAGAAAATGA  
AGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGATGA  
TTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACA  
GCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATG  
GTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATTTTTTGAATCAGATGGT  
TTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGA  
AGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATT  
GCTAACTTAGCTGGCAGTCCTGCTATTA AAAAAGGTATTTTACAGACTGTAAAAAT  
TGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAA  
ATGGCACGTGAAAATCAGACAACCTCAAAGGGCCAGAAAATTCGCGAGAGCGT  
ATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGC  
ATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAA  
AATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAAGTGATT  
ATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAGACGATTCAATAGACAAT  
AAGGTACTAACGCGTTCTGATAAAAAATCGTGGTAAATCGGATAACGTTCCAAGTG  
AAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACCTTCTAAACGCCAAGTT  
AATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGT  
GAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCAC  
TAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAAT  
GATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGA  
CTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATG  
CCCATGATGCGTATCTAAATGCCGTCGTTGGAAGTCTTTGATTAAGAAATATCCA  
AACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTTCGTAAAT  
GATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTAC  
TCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCTG  
CAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAA  
GGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGT  
CAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAA

AGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATG  
 GTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAA  
 AAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTA  
 TGGAAGAAGTTCCTTTGAAAAAATCCGATTGACTTTTTAGAAAGCTAAAGGATA  
 TAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGT  
 TAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAA  
 ATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTAT  
 GAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGC  
 AGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGT  
 GTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAG  
 AGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACG  
 AATCTTGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACG  
 ATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTG  
 GTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:  
 20)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET  
 AEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLLESFLVEEDKKHERH  
 PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN  
 PDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKK  
 NGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLA  
 AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF  
 DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP  
 HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSE  
 ETITPWNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVK  
 YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVDR  
 FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLFEDREMIEERLKTYAHLFDDK  
 VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTF  
 KEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIE  
 MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG  
 RDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV  
 KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ  
 ILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNA  
 VVGTA LIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTE  
 ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKE  
 SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGI  
 TIMERSSEKPNIDFLEAKGYKEVKKDLIILPKYSLFELENGRKRMLASAGELQKGNE  
 LALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLAD  
 ANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL  
 DATLIHQ SITGLYETRIDLSQLGGD (SEQ ID NO: 21)

dCas9 (D10A and H840A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET  
 AEATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLLESFLVEEDKKHERH  
 PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLN  
 PDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKK  
 NGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLA  
 AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF  
 DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP  
 HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSE

ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVK  
 YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR  
 FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK  
 VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTF  
 KEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIE  
 MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG  
 RDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV  
 KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ  
 ILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA  
 VVGTAIIKKYPKLESEFVYGDYKVYDVVRKMIKSEQEIGKATAKYFFYSNIMNFFKTE  
 ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVIVKKTEVQTGGFSKE  
 SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGI  
 TIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENRKRMLASAGELQKGNE  
 LALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLAD  
 ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL  
 DATLIHQSIITGLYETRIDLSQLGGD (SEQ ID NO: 22)

Cas9 nickase (D10A):

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET  
 AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERH  
 PIFGNIVDEVAYHEKYPTIYHLRKKLVDSSTDKADRLIYLALAHMIKFRGHFLIEGDLN  
 PDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK  
 NGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDITYDDDLNLLAQIGDQYADFLA  
 AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF  
 DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP  
 HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVVGPLARGNSRFAWMTRKSE  
 ETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVK  
 YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR  
 FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDK  
 VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTF  
 KEDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIE  
 MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNG  
 RDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV  
 KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ  
 ILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA  
 VVGTAIIKKYPKLESEFVYGDYKVYDVVRKMIKSEQEIGKATAKYFFYSNIMNFFKTE  
 ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVIVKKTEVQTGGFSKE  
 SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGI  
 TIMERSSEFKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENRKRMLASAGELQKGNE  
 LALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLAD  
 ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL  
 DATLIHQSIITGLYETRIDLSQLGGD (SEQ ID NO: 23)

**[0072]** In some embodiments, the effector protein is a Cas9 fusion protein comprising a catalytically inactive Cas9 (dCas9) or Cas9 nickase (nCas9) fused to an effector domain (e.g., an enzymatic domain such as a base modifying enzyme, a transcription regulator, an epigenetic modifier, or a nuclease). Any of the nucleases, transcriptional regulators, and epigenetic modifiers described herein may be fused to a dCas9 or a nCas9 to form a Cas9 fusion protein.

**[0073]** In some embodiments, the Cas9 fusion protein comprises a dCas9 or nCas9 fused to a deaminase (e.g., a cytosine deaminase or adenosine deaminase). A “cytosine deaminase” refers to an enzyme that catalyzes the chemical reaction “cytosine + H<sub>2</sub>O → uracil + NH<sub>3</sub>” or “5-methyl-cytosine + H<sub>2</sub>O → thymine + NH<sub>3</sub>.” As it may be apparent from the reaction formula, such chemical reactions result in a C to U/T nucleobase change. In the context of a gene, such a nucleotide change, or mutation, may in turn lead to an amino acid change in the protein, which may affect the protein’s function, e.g., loss-of-function or gain-of-function. Cas9 fusion proteins comprising a dCas9 or nCas9 fused to a cytosine deaminase have been described in the art, e.g., in US Patent 9,068,179, US Patent Application Publications US 2015/0166980, published June 18, 2015; US 2015/0166981, published June 18, 2015; US 2015/0166982, published June 18, 2015; US 2015/0166984, published June 18, 2015; and US 2015/0165054, published June 18, 2015; PCT Application, PCT/US2016/058344, filed October 22, 2016, US Patent Application, U.S.S.N. 15/311,852, filed October 22, 2016; and in Komor *et al.*, *Nature*, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, 533, 420–424 (2016), the entire contents of each of which is incorporated herein by reference.

**[0074]** An “adenosine deaminase” is an enzyme involved in purine metabolism. It is needed for the breakdown of adenosine from food and for the turnover of nucleic acids in tissues. Its primary function in humans is the development and maintenance of the immune system. An adenosine deaminase catalyzes hydrolytic deamination of adenosine (forming inosine, which base pairs as G) in the context of DNA. There are no known adenosine deaminases that act on DNA. Instead, known adenosine deaminase enzymes only act on RNA (tRNA or mRNA). Evolved deoxyadenosine deaminase enzymes that accept DNA substrates and deaminate dA to deoxyinosine, which were used in Cas9 fusion proteins comprising dCas9 or nCas9 fused to an adenosine deaminase have been described, e.g., in PCT Application, PCT/US2017/045381, filed August 3, 2017; incorporated herein by reference.

**[0075]** When the functional effector protein is a RNA guided nuclease (e.g., Cas9, a Cas9 homolog, a Cas9 variant, or a Cas9 fusion protein), the effector protein may be associated with a guide RNA (gRNA). A “gRNA” (guide ribonucleic acid) herein refers to a fusion of a CRISPR-targeting RNA (crRNA) and a trans-activation crRNA (tracrRNA), providing both targeting specificity and scaffolding/binding ability for Cas9 nuclease. A “crRNA” is a bacterial RNA that confers target specificity and requires tracrRNA to bind to Cas9. A “tracrRNA” is a bacterial RNA that links the crRNA to the Cas9 nuclease and typically can bind any crRNA. The sequence specificity of a Cas DNA-binding protein is determined by

gRNAs, which have nucleotide base-pairing complementarity to target DNA sequences. The native gRNA comprises a 20 nucleotide (nt) Specificity Determining Sequence (SDS), which specifies the DNA sequence to be targeted, and is immediately followed by a 80 nt scaffold sequence, which associates the gRNA with Cas9. In some embodiments, an SDS of the present disclosure has a length of 15 to 100 nucleotides, or more. For example, an SDS may have a length of 15 to 90, 15 to 85, 15 to 80, 15 to 75, 15 to 70, 15 to 65, 15 to 60, 15 to 55, 15 to 50, 15 to 45, 15 to 40, 15 to 35, 15 to 30, or 15 to 20 nucleotides. In some embodiments, the SDS is 20 nucleotides long. For example, the SDS may be 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides long. At least a portion of the target DNA sequence is complementary to the SDS of the gRNA. For Cas9 to successfully bind to the DNA target sequence, a region of the target sequence is complementary to the SDS of the gRNA sequence and is immediately followed by the correct protospacer adjacent motif (PAM) sequence (e.g., NGG for Cas9 and TTN, TTTN, or YTN for Cpf1). In some embodiments, an SDS is 100% complementary to its target sequence. In some embodiments, the SDS sequence is less than 100% complementary to its target sequence and is, thus, considered to be partially complementary to its target sequence. For example, a targeting sequence may be 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, or 90% complementary to its target sequence. In some embodiments, the SDS of template DNA or target DNA may differ from a complementary region of a gRNA by 1, 2, 3, 4 or 5 nucleotides.

**[0076]** In addition to the SDS, the gRNA comprises a scaffold sequence (corresponding to the tracrRNA in the native CRISPR/Cas system) that is required for its association with Cas9 (referred to herein as the “gRNA handle”). In some embodiments, the gRNA comprises a structure 5'-[SDS] -[gRNA handle]-3'. In some embodiments, the scaffold sequence comprises the nucleotide sequence of 5'-guuuuagagcuagaaauagcaaguuaaaauaaaggcuagucgguaucaacuugaaaaaguggcaccgagucggugcuuuuu-3' (SEQ ID NO: 1). Other non-limiting, suitable gRNA handle sequences that may be used in accordance with the present disclosure are listed in Table 2.

**[0077]** In some embodiments, the guide RNA is about 15-120 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 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, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, or 120

nucleotides long. In some embodiments, the guide RNA comprises a sequence of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more contiguous nucleotides that is complementary to a target sequence. Sequence complementarity refers to distinct interactions between adenine and thymine (DNA) or uracil (RNA), and between guanine and cytosine.

**[0078]** A “protospacer adjacent motif” (PAM) is typically a sequence of nucleotides located adjacent to (e.g., within 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 nucleotide(s) of a target sequence). A PAM sequence is “immediately adjacent to” a target sequence if the PAM sequence is contiguous with the target sequence (that is, if there are no nucleotides located between the PAM sequence and the target sequence). In some embodiments, a PAM sequence is a wild-type PAM sequence. Examples of PAM sequences include, without limitation, NGG, NGR, NNGRR(T/N), NNNNGATT, NNAGAAW, NGGAG, and NAAAAC, AWG, CC. In some embodiments, a PAM sequence is obtained from *Streptococcus pyogenes* (e.g., NGG or NGR). In some embodiments, a PAM sequence is obtained from *Staphylococcus aureus* (e.g., NNGRR(T/N)). In some embodiments, a PAM sequence is obtained from *Neisseria meningitidis* (e.g., NNNNGATT). In some embodiments, a PAM sequence is obtained from *Streptococcus thermophilus* (e.g., NNAGAAW or NGGAG). In some embodiments, a PAM sequence is obtained from *Treponema denticola* NGGAG (e.g., NAAAAC). In some embodiments, a PAM sequence is obtained from *Escherichia coli* (e.g., AWG). In some embodiments, a PAM sequence is obtained from *Pseudomonas aeruginosa* (e.g., CC). Other PAM sequences are contemplated. A PAM sequence is typically located downstream (i.e., 3') from the target sequence, although in some embodiments a PAM sequence may be located upstream (i.e., 5') from the target sequence.

**[0079]** In some embodiments, the effector protein is a recombinase. A “recombinase” is an enzyme that catalyzes directionally sensitive DNA exchange reactions between short (30–40 nucleotides) target site sequences that are specific to each recombinase. These reactions enable four basic functional modules: excision/insertion, inversion, translocation and cassette exchange, which have been used individually or combined in a wide range of configurations to control gene expression. Recombinases are known to those of skill in the art, and include, for example, those described herein. In some embodiments, the recombinase is selected from a Cre recombinase, a Tn3 resolvase, a Hin recombinase, or a Gin recombinase.

**[0080]** Non-limiting, exemplary nucleotide and amino acid sequences for recombinases are provided below. Functional variants for these recombinases can also be used.

**[0081]** Stark Tn3 recombinase (nucleotide: SEQ ID NO: 24; amino acid: SEQ ID NO: 25):

ATGGCCCTGTTTGGCTACGCACGCGTGTCTACCAGTCAACAGTCACTCGATTTGCA  
 AGTGAGGGCTCTTAAAGATGCCGGAGTGAAGGCAAACAGAATTTTTACTGATAAG  
 GCCAGCGGAAGCAGCACAGACAGAGAGGGGCTGGATCTCCTGAGAATGAAGGTA  
 AAGGAGGGTGATGTGATCTTGGTCAAAAAATTGGATCGACTGGGGAGAGACACAG  
 CTGATATGCTTCAGCTTATTAAGAGTTTGACGCTCAGGGTGTTGCCGTGAGGTTT  
 ATCGATGACGGCATCTCAACCGACTCCTACATTGGTCTTATGTTTGTGACAATTTT  
 GTCCGCTGTGGCTCAGGCTGAGCGGAGAAGGATTCTCGAAAGGACGAATGAGGGA  
 CGGCAAGCAGCTAAGTTGAAAGGTATCAAATTTGGCAGACGAAGG (SEQ ID NO:  
 24)

MALFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVKE  
 GDVILVKKLDRLGRDADMLQLIKEFDAQGVAVRFIDDGISTDSYIGLMFVTILSAVA  
 QAERRRILERTNEGRQAARKLKGIFGRRR (SEQ ID NO: 25)

**[0082]** Hin Recombinase (nucleotide: SEQ ID NO: 26; amino acid: SEQ ID NO: 27):

ATGGCAACCATTGGCTACATAAGGGTGTCTACCATCGACCAAAATATCGACCTGC  
 AGCGCAACGCTCTGACATCCGCCAACTGCGATCGGATCTTCGAGGATAGGATCAG  
 TGGCAAGATCGCCAACCGGCCCGGTCTGAAGCGGGCTCTGAAGTACGTGAATAAG  
 GGCGATACTCTGGTTGTGTGGAAGTTGGATCGCTTGGGTAGATCAGTGAAGAATCT  
 CGTAGCCCTGATAAGCGAGCTGCACGAGAGGGGTGCACATTTCCATTCTCTGACC  
 GATTCCATCGATACGTCTAGCGCCATGGGCCGATTCTTCTTTTACGTCATGTCCGCC  
 CTCGCTGAAATGGAGCGCGAACTTATTGTTGAACGGACTTTGGCTGGACTGGCAG  
 CGGCTAGAGCACAGGGCCGACTTGGA (SEQ ID NO: 26)

MATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPGLKRALKYVNBKDTL  
 VVWKLDRLGRSVKNLVALISELHERGAHFHSLTDSIDTSSAMGRFFFYVMSALAEME  
 RELIVERTLAGLAAARAQGR LG (SEQ ID NO: 27)

**[0083]** Gin beta recombinase (nucleotide: SEQ ID NO: 28; amino acid: SEQ ID NO: 29):

ATGCTCATTGGCTATGTAAGGGTCAGCACCAATGACCAAAACACAGACTTGCAAC  
 GCAATGCTTTGGTTTGCGCCGGATGTGAACAGATATTTGAAGATAAACTGAGCGG  
 CACTCGGACAGACAGACCTGGGCTTAAGAGAGCACTGAAAAGACTGCAGAAGGG  
 GGACACCCTGGTTCGTCTGGAACTGGATCGCCTCGGACGCAGCATGAAACATCTG  
 ATTAGCCTGGTTGGTGAGCTTAGGGAGAGAGGAATCAACTTCAGAAGCCTGACCG  
 ACTCCATCGACACCAGTAGCCCCATGGGACGATTCTTCTTCTATGTGATGGGAGCA  
 CTTGCTGAGATGGAAAGAGAGCTTATTATCGAAAGAACTATGGCTGGTATCGCTG  
 CTGCCCGGAACAAAGGCAGACGGTTCGGCAGACCGCCGAAGAGCGGC (SEQ ID  
 NO: 28)

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD  
 LVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEME  
 RELIERTMAGIAAARNKGRRFGRPPKSG (SEQ ID NO: 29)

**[0084]** In some embodiments, the effector protein is an epigenetic modifier. An “epigenetic modifier” refers to an enzyme that modifies a DNA base without changing the DNA sequence.

For example, an epigenetic modifier catalyzes DNA methylation (and demethylation), acetylation (and deacetylation), or histone modifications (e.g., histone methylation/demethylation, acetylation/deacetylation, ubiquitylation, phosphorylation, sumoylation, etc.). The presence of one more epigenetic modifications can affect the transcriptional activity of one or more genes, for example turning genes from an “on” state to an “off” state, and vice versa. Epigenetic modifiers include, but are not limited to, histone demethylase, histone methyltransferase, hydroxylase, histone deacetylase, and histone acetyltransferase. Exemplary epigenetic modifiers can be found in Konermann et al., *Nature*. 2013; 500, 472-476; Mendenhall et al., *Nat. Biotechnol.* 2013; 31, 1133-1136; and Maeder et al., *Nat. Biotechnol.* 2013; 31, 1137-1142; the entire contents of each are incorporated herein by reference. Non-limiting, exemplary amino acid sequences of epigenetic modifiers are provided below.

**[0085]** LSD1, isoform a (human):

MLSGKKAAAAAAAAAAAAATGTEAGPGTAGGSENGSEVAAQPAGLSGPAEVGPGAV  
 GERTPRKKEPPRASPPGGLAEPPGSAGPQAGPTVVPGSATPMETGIAETPEGRRTSRRK  
 RAKVEYREMDLANLSEDEYYSEEERNAKAEKEKKLPPPPPQAPPEEENESEPEEPSG  
 QAGGLQDDSSGGYGDGQASGVEGAFAFQSRPLPHDRMTSQEAACFPDIISGPQQTQKVF  
 LFIRNRTLQLWLDNPKIQLTFEATLQQLEAPYNSDTVLVHRVHSYLERHGLINFGIYKR  
 IKPLPTKKTGKVIIIGSGVSGLAARQLQSFQMDVTLLLEARDRVGGRVATFRKGNVVA  
 DLGAMVVTGLGGNPMMAVVSQVNMELAKIKQKCPLYEANGQADTVKVPKEKDEMV  
 EQEFNRLLEATSYLSHQLDFNVLNNKPVSLGQALEVVIQLQEKHVKDEQIEHWKKIVK  
 TQEELKELLNKMVNLEKIKELHQYKEASEVKPPRDITAEFLVKS KHRDLTALCKEY  
 DELAETQGKLEEKLELEANPPSDVYLSSRDRQILDWHFANLEFANATPLSTLSLKHWD  
 QDDDDFEFTGSHLTVRNGYSCVPVALAEGLDIKLNTAVRQVRYTASGCEVIAVNTRST  
 SQTFIYKCDAVLCTPLGLVKQPPAVQFVPLPEWKTSVAVQRMGFGNLNKVVLCD  
 RVFWDPSVNLFGHVGSTTASRGELFLFWNLYKAPILLALVAGEAAGIMENISDDVIVG  
 RCLAILKGIFGSSAVPQPKETVVSRRADPWARGSYSYVAAGSSGNDYDLMAQPITP  
 GPSIPGAPQPIRLFFAGEHTIRNYPATVHGALLSGLREAGRIADQFLGAMYTLPRQATP  
 GVPAQQSPSM (SEQ ID NO: 30)

**[0086]** LSD1, isoform b (human):

MLSGKKAAAAAAAAAAAAATGTEAGPGTAGGSENGSEVAAQPAGLSGPAEVGPGAV  
 GERTPRKKEPPRASPPGGLAEPPGSAGPQAGPTVVPGSATPMETGIAETPEGRRTSRRK  
 RAKVEYREMDLANLSEDEYYSEEERNAKAEKEKKLPPPPPQAPPEEENESEPEEPSG  
 VEGAFAFQSRPLPHDRMTSQEAACFPDIISGPQQTQKVF LFIRNRTLQLWLDNPKIQLTFE  
 ATLQQLEAPYNSDTVLVHRVHSYLERHGLINFGIYKR IKPLPTKKTGKVIIIGSGVSGLA  
 AARQLQSFQMDVTLLLEARDRVGGRVATFRKGNVVA DLGAMVVTGLGGNPMMAVVSQ  
 QVNMELAKIKQKCPLYEANGQAVPKEKDEMVEQEFNRLLEATSYLSHQLDFNVLNN  
 KPVSLGQALEVVIQLQEKHVKDEQIEHWKKIVKTQEELKELLNKMVNLEKIKELHQ  
 QYKEASEVKPPRDITAEFLVKS KHRDLTALCKEYDELAETQGKLEEKLELEANPPSD  
 VYLSSRDRQILDWHFANLEFANATPLSTLSLKHWDQDDDDFEFTGSHLTVRNGYSCVP  
 VALAEGLDIKLNTAVRQVRYTASGCEVIAVNTRSTSQTFIYKCDAVLCTPLGLVKQQ

PPAVQFVPPLPEWKTS AVQRMGFGNLNKVVLCFDRVFWDP SVNLF GHV GSTTASRGE  
LFLFWNL YKAPILLALVAGEAAGIMENISDDVIVGRCLA ILKGIFGSSAVPQPKETVVS  
RWRADPWARGSY SYVAAGSSGNDYDLMAQPITPGPSIPGAPQPIPR LFFAGEHTIRNYP  
ATVHGALLSGLREAGRIADQFLGAMYTLPRQATPGVPAQQSPSM (SEQ ID NO: 31)

**[0087]** TET1:

SIVAQLSRPDPALAAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTS  
HRVADHAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSGGGS LPTCSC LDRVIQKDKG  
PYYTHLGAGPSVAAVREIMENRYGQKGN AIRIEIVVYTGKEGKSSHGCPIAKWVLRRS  
SDEEKVLC LVRQRTGHHCP TAVMVVLIMVWDGIPLMADR LYTELTENLKS YNGHPT  
DRRCTLNENRTCTCQGIDPETCGASFSGCSWSMYFNGCKFGRSPSPRRFRIDPSSPLH  
EKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLGSKEGRPFSGVTACL  
DFCAHPHRDIHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVLPLYKLSDTDEFGSKE  
GMEAKIKSGAIEVLAPRRKKRTCFTQPVPRSGKKRAAMMTEVLAHKIRAVEKKPIPRI  
KRKNNSTTTNNSKPSLPTLGSNTETVQPEVKSETEPHFILKSSDNTKTYSLMPSAPHPV  
KEASPGFSWSPKTASATPAPLKNDATASCGFSERSSTPHCTMPSGRLSGANAAAADGP  
GISQLGEVAPLPTLSAPVMEPLINSEPSTGVTEPLTPHQPNHQPSFLTSPQDLASSPMEE  
DEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWS DSEHIFLDANIGGVAIAPAHGSVL  
IECARRELHATTPVEHPNRNHPTRL SLV FYQHKNLNKPQHGFELN KIKFEAKEAKNKK  
MKASEQKDQAANEGPEQSSEVNELNQIPSHKALT LTHDNVVTVSPYAL THVAGPYNH  
WV (SEQ ID NO: 32)

**[0088]** HDAC:

**[0089]** HDAC8 (*X. laevis*):

ASSPKKKRKVEASMSRVVKPKVASMEEMAA FH TDAYLQHLHKVSEEGDNDDPETLE  
YGLGYDCPITEGIYDYAAAVGGATLTAAEQLIEGKTRIAVNWPGGWHHAKKDEASGF  
CYLND AVLGLKLREKFDRVLYVDMDLHHGDGVEDAFSFTSKVMTVSLHKFSPGFFP  
GTGDVSDIGLGKGRYY SINVPLQDGIQDDKYYQICEGV LKEVFTTFNPEAVVLQLGAD  
TIAGDPMCSFNMTPEGIGKCLKYVLQWQLPTLILGGGGYHLPNTARCW TYLTALIVGR  
TLSSEIPDHEFFTEYGP DYVLEITPSCRPD RNDTQKVQEILQSIKGNLKR VVEF (SEQ ID  
NO: 33)

**[0090]** RPD3 (*S. cerevisiae*):

ASSPKKKRKVEASRRVAYFYDADVGN YAYGAGHPMKPHRIRMAHSLIMNYGLYKK  
MEIYRAKPA TKQEMCQFHTDEYIDFLSRVTPDNLEMFKRESVKFNVGDDCPVFDGLY  
EYCSISGGGSMEGAARLN RGKCDVAVNYAGGLHHAKKSEASGFCYLNDIVLGIIELLR  
YHPRVLYIDIDVHHGDGV EEA FYTTDRVMTCSFH KYGEFFPGTGELRDIGVGAGKNY  
AVNVPLRDGIDDATYRSVFEPVIKKIMEWYQPSAVVLQCGGDSLSGDRLGCFNLSME  
GHANCVNYVKSFGIPMMVVG GGYTMRNVARTWCFETG LLNNVVLDKDLPYEF  
(SEQ ID NO: 34)

**[0091]** MesoLo4 (*M. loti*):

ASSPKKKRKVEASMPLQIVHHPDYDAGFATNHRFPMSKYPLLMEALRARGLASPDAL  
NTTEPAPASWLKLAHAADYVDQVISCSVPEKIEREIGFPVGPRVSLRAQLATGGTILAA  
RLALRHGIACNTAGGSHHARRAQGAGFCTFNDVAVASLVLLDEGAAQNILVVDLDV  
HQGDGTADILSDEPGVFTFSMHGERNYPVRKIASDL DIALPDGTGDAA YLRRLATILPE

LSARARWDIVFYNAGVDVHAEDRLGRLALSNGGLRARDEMVIGHFRALGIPVCGVIG  
GGYSTDPALASRHAILFEVASTYAEF (SEQ ID NO: 35)

**[0092]** HDAC11 (human):

ASSPKKKRKVEASMLHTTQLYQHVPETRWPIVYSPRYNITFMGLEKLHPFDAGKWGK  
VINFLKEEKLLSDSMLVEAREASEEDLLVVHTRRYLNELKWSFAVATITEIPPVIFLPNF  
LVQRKVLRLRTQTGGTIMAGKLAVERGWAINVGGGFHHCSSDRGGGFCAYADITLA  
IKFLFERVEGISRATIIDLDAHQGNNGHERDFMDDKRVIYIMDVYNRHIYPGDRFAKQAIR  
RKVELEWGTEDDEYLDKVERNIKKSLQEHLPDVVVYNAGTDILEGDRLGGLSISPAGI  
VKRDELVFRMVRGRRVPILMVTSGGYQKRTARIADSILNLFGLGLIGPESPSVSAQNS  
DTPLLPPAVPEF (SEQ ID NO: 36)

**[0093]** HDT1 (*A. thaliana*):

ASSPKKKRKVEASMEFWGIEVKSGKPVTVTPEEGILIHVSQASLGECKNKKGEFVPLH  
VKVGNQNLVLGTLSTENIPQLFCDLVFDKEFELSHTWKGKGSVYFVGKYKTPNIEPQGYS  
EEEEEEEEVPAGNAKAVAKPKAKPAEVKPAVDDEEDESDDGMDDEDDSDGEDSEE  
EEPTPKKPASSKKRANETTPKAPVSAKKAKVAVTPQKTDEKKKGGKAANQSEF (SEQ  
ID NO: 37)

**[0094]** SIRT3 (human):

ASSPKKKRKVEASMGAGISTPSGIPDFRSPGSGLYSNLQQYDLPYPEAIFELPFFFHNP  
KPFFTLAKELYPGNYKPNVTHYFLRLLHDKGLLLRLYTQNIDGLERVSGIPASKLVEA  
HGTFASATCTVCQRPFGEDIRADVMADRVPRCPVCTGVVKPDIVFFGEPLPQRFLH  
VVDFFMADLLLLILGTSLEVEPFASLTEAVRSSVPRLINRDLVGPLAWHPRS RDVAQLG  
DVVHGVESLVELLGWTEEMRDLVQRETGKLDGPDKEF (SEQ ID NO: 38)

**[0095]** HST2 (*S. cerevisiae*):

ASSPKKKRKVEASTEMSVRKIAAHMKSNNPAKVIFMVGAGISTSCGIPDFRSPGTGLY  
HNLARLKLPEYAEVFDVDFQSDPLPFYTLAKELYPGNFRPSKFHYLLKLFQDKDVLK  
RVYTQNIDTLERQAGVKDDLIEAHGSFAHCHCIGCGKVYPPQVFKSKLAEHPIKDFVK  
CDVCGELVKPAIVFFGEDLPDSFSETWLNDSEWLREKITTSGKHPQQPLVIVVGTSLAV  
YPFASLPEEIPRKVKRVLNLETVGDFKANKRPTDLIVHQYSDEFAEQLVEELGWQED  
FEKILTAQGGMGEF (SEQ ID NO: 39)

**[0096]** CobB (*E. coli* (K12)):

ASSPKKKRKVEASMEKPRVLVLTGAGISAESGIRTFRAADGLWEEHRVEDVATPEGFD  
RDPELVQAFYNARRRQLQQPEIQPNAHLALAKLQDALGDRFLLVTQNIDNLHERAG  
NTNVIHMHGELLKVRCSQSGQVLDWTGDVTPEDKCHCCQFPAPLRPHVWVWFGEMPL  
GMDEIYMALSMADIFIAIGTSGHVYPAAAGFVHEAKLHGAHTVELNLEPSQVGNEFAEK  
YYGPASQVVPEFVEKLLKGLKAGSIAEF (SEQ ID NO: 40)

**[0097]** HST2 (*C. albicans*):

ASSPKKKRKVEASMPSLDDILKPVAEAVKNGKKVTFENGAGISTGAGIPDFRSPDTGL  
YANLAKLNLPEAEVFDIDFFKEDPKPFYTLAEELYPGNFAPTKFHHFIKLLQDQGSLLK  
RVYTQNIDTLERLAGVEDKYIVEAHGSFASNHCVDCHKEMTTETLKTMYMKDKKIPSC

QHCEGYVKPDIVFFGEGLPVKFFDLWEDDCEDVEVAIVAGTSLTVFPFASLPGEVNKK  
CLRVLVNKEKVGTFKHEPRKSDIIALHDCDIVAERLCTLLGLDDKLNEVYEKEKIKYS  
KAETKEIKMHEIEDKLKEEAHLKEDKHTTKVDKKEKQNDANDKELEQLIDKAKAEF  
(SEQ ID NO: 41)

[0098] SIRT5 (human):

ASSPKKKRKVEASSSSMADFRKFFAKAKHIVIISGAGVSAESGVPTFRGAGGYWRKW  
QAQDLATPLAFAHNPSRVWEFYHYRREVMGSKEPNAGHRAIAECETRLGKQGRRVV  
VITQNIDELHRKAGTKNLLEIHGSLFKTRCTSCGVVAENYKSPICPALSGKGAPEPGTQ  
DASIPVEKLPRCEEAGCGLLRPHVWVWFGENLDPAILVEVDRELAHCDLCLVVGTSV  
VYPAAMFAPQVAARGVPVAEFNTETTPATNRFRHFQGPCGTTLPEALACHENETVSE  
F (SEQ ID NO: 42)

[0099] Sir2A (*P. falciparum*):

ASSPKKKRKVEASMGNLMISFLKKDTQSITLEELAKIIKKCKHVVALTGSGTSAESNIPS  
FRGSSNSIWSKYDPRIYGTIWGFWKYPEKIWEVIRDISSDYEIEINNGHVALSTLES LGY  
LKS VVTQNV DGLHEASGNTKVISLHGNVFEAVCCTCNKIVKLNKIMLQKTSHFMHQL  
PPECPCGGIFKPNILFGEVVSSDLLKEAEEIEIAKCDLLVIGTSSTVSTATNLCHFACKK  
KKKIVEINISKTYITNKMSDYHVC AKFSELT KVANILKGSSEKNKKIMEF (SEQ ID NO:  
43)

[00100] SIRT6 (human):

ASSPKKKRKVEASMSVNYAAGLSPYADKGKCGLPEIFDPPEELERK V WELARLVWQS  
SSVVFHTGAGISTASGIPDFRGPHGVWTMEERGLAPKFDTTFESARPTQTHMALVQLE  
RVGLLRFLVSQNV DGLHVRSGFPRDKLAELHGNMFVEECAKCKTQYVRDTVVGTMG  
LKATGRLCTVAKARGLRACRGELRDTILDWEDSLPDRDLALADEASRNADLSITLGTS  
LQIRPSGNLPLATKRRGGRLVIVNLQPTKHDRHADLRHGYVDEVMTRLMKHLGLEIP  
AWDGPRVLERALPPLEF (SEQ ID NO: 44)

[00101] HMT effector domains:

[00102] NUE (*C. trachomatis*):

ASSPKKKRKVEASMTTNSTQDTLYLSLHGGIDSAIPYPVRRVEQLLQFSFLPELQFQNA  
AVKQRIQRLCYREEKRLAVSSSLAKWLGQLHKQRLRAPKNPPVAICWINSYVGYGVFA  
RESIPAWSYIGEYTGILRRRQALWLDENDYCFRYPVPRYSFRYFTIDSGMQGNVTRFIN  
HSDNP NLEAIGAFENGIFHIIIRA IKDILPGEELCYHYGPLYWKHRKKREEFVPQEEEF  
(SEQ ID NO: 45)

[00103] vSET (*P. bursaria* chlorella virus):

ASSPKKKRKVEASMFNDRVIVKKSPLGGYG V FARKSFEKGELVEECLCIVRHND DWG  
TALEDYLF SRKNMSAMALGFGAIFNH SKDPNARHELTAGLKRMRIFTIKPIAIGEEITIS  
YGDDYWLSRPRLTQNEF (SEQ ID NO: 46)

[00104] SUV39H1 (human):

ASSPKKKRKVEASNLKCV RILKQFHKDLERELLRRHRSKTPRHLDPSLANYL VQKA  
KQRRALRRWEQELNAKRSHLGRITVENEVDLDGPPRAFVYINEYRVGEGITLNQVAV

GCECQDCLWAPTGGCCPGASLHKFAYNDQGQVRLRAGLPIYECNSRRCRCGYDCPNR  
 VVQKGIRYDLCIFRTDDGRGWGVRTLEKIRKNSFVMEYVGEIITSEEAEERRGQIYDRQ  
 GATYLFDLDYVEDVYTVDAAYYGNISHFVNHSCDPNLQVYNVFIDNLDERLPRIAFFA  
 TRTIRAGEELTFDYNMQVDPVDMESTRMDSNFGLAGLPGSPKKRVRIECKCGTESCRK  
 YLFEF (SEQ ID NO: 47)

**[00105]** DIM5 (*N. crassa*):

ASSPKKKRKVEASMEKAFRPHFFNHGKPDANPKEKKNCHWCQIRSFATHAQLPISIVN  
 REDDAFLNPNFRFIDHSIIGKNVPVADQSFVGCSCASDEECMYSTCQCLDEMAPDS  
 EEADPYTRKKRFAYYSQGAKKGLLRDRVLQSQEPIYECHQGCACSKDCPNRVVERGR  
 TVPLQIFRTKDRGWGVKCPVNIKRQGFVDRYLGEIITSEEADRRRAESTIARRKDVYLF  
 ALDKFSDPDSLDPPLLAGQPLEVDGEYMSGPTRFINHSCDPNMAIFARVGDHADKHIHD  
 LALFAIKDIPKGTELTFDYVNGLTGLES DAHDPSKISEMTKCLCGTAKCRGYLWEF  
 (SEQ ID NO: 48)

**[00106]** KYP (*A. thaliana*):

ASSPKKKRKVEASDISGGLEFKGIPATNRVDDSPVSPTS GFTYIKSLIIEPNVIIPKSSTGC  
 NCRGSCDTSKKCACAKLNGGNFPYVDLNDGRLIESRDVVFECGPHCGCGPKCVNRTS  
 QKRLRFNLEVFRSAKKGWAVRSWEYIPAGSPVCEYIGVVRRTADVDTISDNEYIFEIDC  
 QQTMQGLGGRQRRRLRDVA VPMNNGVSQSSEDENAPEFCIDAGSTGNFARFINHSCEP  
 NLFVQCVLSSHQDRLARVVLFAADNISPMQELTYDYGALDSVHEF (SEQ ID NO: 49)

**[00107]** SUVR4 (*A. thaliana*):

ASSPKKKRKVEASQSAYLHVSLARISDEDCANCKGNCLSA DFPCTCARETSGEYAYT  
 KEGLLKEKFLDTCLMKKEPDSFPKVYCKDCPLERDHDKGTYGKCDGHLIRKFIKEC  
 WRKCGCDMQCGNRVVQRGIRCQLQVYFTQEGKGWGLRTLQDLPKGT FICEYIGEILT  
 NTELYDRNVRSSESRHTYPVTLDADWGSEKDLKDEEALCLDATICGNVARFINHRCE  
 DANMIDIPIEJETPDRHYHIAFFTLDVKAMDELTDWYDYMIDFNDKSHPVKA FRCCCG  
 SESCRDRKIKGSQGKSIERRKIVSAKKQQGSKEVSKKRKEF (SEQ ID NO: 50)

**[00108]** Set4 (*C. elegans*):

ASSPKKKRKVEASMLHEQIANISVTFNDIPRSDHSMTPTELCYFDDFATTLVVDSVLN  
 FTTHKMSKKRRYLYQDEYRTARTVMKTFREQRDWTNAIYGLLT LRSVSHFLSKLPPN  
 KLFEFRDHIVRFLNMFILDSGYTIQECKRYSQEGHQGAKLVSTGVWSRGDKIERLSGV  
 VCLLSSEDEDSILAQEGSDFSVMYSTRKRCSTLWLGPAYINHDCRPTCE FVSHGSTA  
 HIRVLRDMVPGDEITCFYGSEFFGPNNIDCECCTCEKNMNGAFSYLRGNENA EPIISEK  
 KTKYELRSRSEF (SEQ ID NO: 51)

**[00109]** Set1 (*C. elegans*):

ASSPKKKRKVEASMKVAAKKLATSRMRKDRAAAA SPSSDIENSENPSSLASHSSSSGR  
 MTPSKNTRSRKGVSVKDVSNHKITEFFQVRRSNRKT SKQISDEAKHALRDTV LKGTNE  
 RLLEVYKDVVKGRGIRTKVNFKEGDFVVEYRGVMMEYSEAKVIEEQYSNDEEIGSYM  
 YFFEHNKKWCIDATKESPWKGRLINHSVLRPNLKT KVV EIDGSHHLILVARRQIAQG  
 EELLYDYGDRSAETIAKNPWLVNTEF (SEQ ID NO: 52)

**[00110]** SETD8 (human)

ASSPKKKRKVEASSCDSTNAIAIAKQALKKKPIKGGKQAPRKKAAQGKTQQNRKLTDFFPV  
 RRSSRKSKAELQSEERKRIDELIESGKEEGMKIDLIDGKGRGVIA TKQFSRGDFVVEYH  
 GDLIEITDAKKREALYAQDPSTGCYMYFQYLSKTYCVDATRETNRLGRLINHSKCG  
 NCQTKLHDIDGVPHLILIASRDIAAGEELLYDYGDRSKASIEAFPWLKHEF (SEQ ID  
 NO: 53)

**[00111]** TgSET8 (*T. gondii*):

ASSPKKKRKVEASASRRTGEFLRDAQAPSRWLKRSKTGQDDGAFCLWTWLAGAGDD  
 AAGGERGRDREGAADKAKQREERRQKELEERFEEMKVEFEEKAQRMIAARRAALTGEI  
 YSDGKGSKKPRVPSLPENDDALIEIIIDPEQGILKWPLSVMSIRQRTVIYQECLRRDLT  
 ACIHLTKVPGKGRAVFAADTILKDDFVVEYKGELCSEREAREQRYNRSKVPMGSF  
 MFYFKNGSRMMAIDATDEKQDFGPRLINHSRRNPNMTPRAITLGD FNSEPR LIFVAR  
 RNIEKGEELLVDYGERDPDVIKEHPWLNSEF (SEQ ID NO: 54)

**[00112]** In some embodiments, the effector protein is a transcriptional regulator. A transcriptional regulator refers to a protein that regulates gene expression by activating or repressing the transcription of a gene. A transcriptional regulator may be a transcriptional activator or a transcriptional repressor. A transcriptional activator is a protein that increases gene transcription of a gene or set of genes. Most activators function by binding sequence-specifically to a DNA site located in or near a promoter and making protein–protein interactions with the general transcription machinery (RNA polymerase and general transcription factors), thereby facilitating the binding of the general transcription machinery to the promoter. A transcriptional repressor is a DNA- or RNA-binding protein that inhibits the expression of one or more genes by binding to the operator or associated silencers. A DNA-binding repressor blocks the attachment of RNA polymerase to the promoter, thus preventing transcription of the genes into messenger RNA. An RNA-binding repressor binds to the mRNA and prevents translation of the mRNA into protein. Non-limiting, exemplary amino acid sequences of transcriptional regulators are provided below.

**[00113]** VP64

GSGRADALDDFDLDM LGSDALDDFDLDM LGSDALDDFDLDM LGSDALDDFDLDM LI  
 N (SEQ ID NO: 55)

**[00114]** VP16

APPTDVSLGDELHLDGEDVAMAHADALDDFDLDM LGDGDSPGPGFTPHDSAPYGAL  
 DMADFEFEQMFTDALGIDEYGGFEFPGIRR (SEQ ID NO: 56)

**[00115]** p65:

PSGQISNQALALAPSSAPVLAQTMVPSSAMVPLAQPPAPAPVLTPGPPQSL SAPVPKST  
QAGEGTLSEALLHLQFDADEDLGALLGNSTDPGVFTDLASVDNSEFQQLNQGVSMS  
HSTAEPMLMEYPEAITRLVTGSQRPPDPAPTPLGTSGLPNGLSGDEDFSSIADMDFSAL  
LSQISSSGQ (SEQ ID NO: 57)

[00116] dCas9-VP64-6xHis:

MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET  
AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERH  
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLN  
PDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKK  
NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA  
AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF  
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP  
HQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYVVGPLARGNSRFAWMTRKSE  
ETITPWNFEVVDKGAQAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVK  
YVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVDR  
FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDK  
VMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTF  
KEDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIE  
MARENQTTQKGQKNSRERMKRIEIEGKELGSQILKEHPVENTQLQNEKLYLYYLQNG  
RDMYVDQELDINRLSDYDVAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVV  
KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVETRQITKHVAQ  
ILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNA  
VVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTE  
ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTETVQTGGFSKE  
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGI  
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNE  
LALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFSKRVLAD  
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL  
DATLIHQSIITGLYETRIDLSQLGGDGSPKKRKRKVSDDYKDHDGDYKDHDIDYKDDDD  
KAAGGGGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDD  
FDLDMLHHHHHH (SEQ ID NO: 58)

[00117] KRAB (human; GenBank: AAD20972.1)

MNMFKEAVTFKDVAVAFTEEEELGGLGPAQRKLYRDVMVENFRNLLSVGHPPFKQDV  
SPIERNEQLWIMTTATRRQGNLDTLPVKALLLYDLAQT (SEQ ID NO: 59)

[00118] KRAB protein domain, partial (human; GenBank: CAB52478.1):

EQVSFKDVCVDFTQEEWYLLDPAQKILYRDVILENYSNLVSVGYCITKPEVIFKIEQGE  
EPWILEKGFPSQCHP (SEQ ID NO: 60)

[00119] KRAB A domain, partial (human; GenBank: AAB03530.1):

EAVTFKDVAVVFTEEEELGGLDPAQRKLYRDVMLENFRNLLSV (SEQ ID NO: 61)

[00120] KRAB (mouse; C2H2 type domain containing protein; GenBank: CAM27971.1):

MDLVITYDDVHVNFTQDEWALLDPSQKSPLYKGVMLETYKNLTAIGYIWEEHTIEDHFQ  
TSRSHGSNKKTH (SEQ ID NO: 62)

**[00121]** SID repressor domain:

GSGMNIQMLLEAADYLERREREAHEGYASMLP (SEQ ID NO: 63)

**[00122]** SID4x repressor domain:

GSGMNIQMLLEAADYLERREREAHEGYASMLPGSGMNIQMLLEAADYLERREREA  
HGYASMLPGSGMNIQMLLEAADYLERREREAHEGYASMLPGSGMNIQMLLEAADYL  
ERREREAHEGYASMLPSR (SEQ ID NO: 64)

**[00123]** In some embodiments, the effector protein is a therapeutic protein. Non-limiting examples of therapeutic proteins include enzymes, regulatory proteins (e.g., immuno-regulatory proteins), antigens, antibodies or antibody fragments, and structural proteins. Suitable enzymes include, without limitation, oxidoreductases, transferases, polymerases, hydrolases, lyases, synthases, isomerases, and ligases, digestive enzymes (e.g., proteases, lipases, carbohydrases, and nucleases). In some embodiments, the enzyme is selected from the group consisting of lactase, beta-galactosidase, a pancreatic enzyme, an oil-degrading enzyme, mucinase, cellulase, isomaltase, alginase, digestive lipases (e.g., lingual lipase, pancreatic lipase, phospholipase), amylases, cellulases, lysozyme, proteases (e.g., pepsin, trypsin, chymotrypsin, carboxypeptidase, elastase,), esterases (e.g. sterol esterase), disaccharidases (e.g., sucrase, lactase, beta-galactosidase, maltase, isomaltase), DNases, and RNases.

**[00124]** Non-limiting examples of antibodies and fragments thereof include: bevacizumab (AVASTIN®), trastuzumab (HERCEPTIN®), alemtuzumab (CAMPATH®, indicated for B cell chronic lymphocytic leukemia,), gemtuzumab (MYLOTARG®, hP67.6, anti-CD33, indicated for leukemia such as acute myeloid leukemia), rituximab (RITUXAN®), tositumomab (BEXXAR®, anti-CD20, indicated for B cell malignancy), MDX-210 (bispecific antibody that binds simultaneously to HER-2/neu oncogene protein product and type I Fc receptors for immunoglobulin G (IgG) (Fc gamma RI)), oregovomab (OVAREX®, indicated for ovarian cancer), edrecolomab (PANOREX®), daclizumab (ZENAPAX®), palivizumab (SYNAGIS®, indicated for respiratory conditions such as RSV infection), ibritumomab tiuxetan (ZEVALIN®, indicated for Non-Hodgkin's lymphoma), cetuximab (ERBITUX®), MDX-447, MDX-22, MDX-220 (anti-TAG-72), IOR-C5, IOR-T6 (anti-CD1), IOR EGF/R3, celogovab (ONCOSCINT® OV103), epratuzumab (LYMPHOCIDE®), pemtumomab (THERAGYN®), Gliomab-H (indicated for brain cancer, melanoma). In some embodiments, the antibody is an antibody that inhibits an immune check point protein, e.g., an anti-PD-1

antibody such as pembrolizumab (Keytruda®) or nivolumab (Opdivo®), or an anti-CTLA-4 antibody such as ipilimumab (Yervoy®).

**[00125]** A regulatory protein that is a therapeutic protein may be, in some embodiments, a transcription factor or an immunoregulatory protein. Non-limiting, exemplary transcriptional factors include: those of the NFkB family, such as Rel-A, c-Rel, Rel-B, p50 and p52; those of the AP-1 family, such as Fos, FosB, Fra-1, Fra-2, Jun, JunB and JunD; ATF; CREB; STAT-1, -2, -3, -4, -5 and -6; NFAT-1, -2 and -4; MAF; Thyroid Factor; IRF; Oct-1 and -2; NF-Y; Egr-1; and USF-43, EGR1, Sp1, and E2F1.

**[00126]** In some embodiments, the therapeutic protein is an immunoregulatory protein. An immunoregulatory protein is a protein that regulates an immune response. Non-limiting examples of immunoregulatory proteins include: antigens, adjuvants (e.g., flagellin, muramyl dipeptide), cytokines including interleukins (e.g., IL-2, IL-7, IL-15 or superagonist/mutant forms of these cytokines), IL-12, IFN-gamma, IFN-alpha, GM-CSF, FLT3-ligand), and immunostimulatory antibodies (e.g., anti-CTLA-4, anti-CD28, anti-CD3, or single chain/antibody fragments of these molecules).

**[00127]** In some embodiments, the therapeutic protein is an antigen. An antigen is a molecule or part of a molecule that is bound by the antigen-binding site of an antibody. In some embodiments, an antigen is a molecule or moiety that, when administered to or expression in the cells of a subject, activates or increases the production of antibodies that specifically bind the antigen. Antigens of pathogens are well known to those of skill in the art and include, but are not limited to parts (coats, capsules, cell walls, flagella, fimbriae, and toxins) of bacteria, viruses, and other microorganisms. Examples of antigens that may be used in accordance with the disclosure include, without limitation, cancer antigens, self-antigens, microbial antigens, allergens and environmental antigens. In some embodiments, the antigen is a cancer antigen. A cancer antigen is an antigen that is expressed preferentially by cancer cells (i.e., it is expressed at higher levels in cancer cells than on non-cancer cells) and, in some instances, it is expressed solely by cancer cells. Cancer antigens may be expressed within a cancer cell or on the surface of the cancer cell. Cancer antigens that may be used in accordance with the disclosure include, without limitation, MART-1/Melan-A, gp100, adenosine deaminase-binding protein (ADAbp), FAP, cyclophilin b, colorectal associated antigen (CRC)--C017-1A/GA733, carcinoembryonic antigen (CEA), CAP-1, CAP-2, etv6, AML1, prostate specific antigen (PSA), PSA-1, PSA-2, PSA-3, prostate-specific membrane antigen (PSMA), T cell receptor/CD3-zeta chain and CD20. The cancer antigen may be selected from the group consisting of MAGE-A1, MAGE-A2, MAGE-A3, MAGE-A4, MAGE-A5, MAGE-A6,

MAGE-A7, MAGE-A8, MAGE-A9, MAGE-A10, MAGE-A11, MAGE-A12, MAGE-Xp2 (MAGE-B2), MAGE-Xp3 (MAGE-B3), MAGE-Xp4 (MAGE-B4), MAGE-C1, MAGE-C2, MAGE-C3, MAGE-C4 and MAGE-C5. The cancer antigen may be selected from the group consisting of GAGE-1, GAGE-2, GAGE-3, GAGE-4, GAGE-5, GAGE-6, GAGE-7, GAGE-8 and GAGE-9. The cancer antigen may be selected from the group consisting of BAGE, RAGE, LAGE-1, NAG, GnT-V, MUM-1, CDK4, tyrosinase, p53, MUC family, HER2/neu, p21ras, RCAS1,  $\alpha$ -fetoprotein, E-cadherin,  $\alpha$ -catenin,  $\beta$ -catenin,  $\gamma$ -catenin, p120ctn, gp100Pmel117, PRAME, NY-ESO-1, cdc27, adenomatous polyposis coli protein (APC), fodrin, Connexin 37, Ig-idiotypic, p15, gp75, GM2 ganglioside, GD2 ganglioside, human papilloma virus proteins, Smad family of tumor antigens, Imp-1, P1A, EBV-encoded nuclear antigen (EBNA)-1, brain glycogen phosphorylase, SSX-1, SSX-2 (HOM-MEL-40), SSX-3, SSX-4, SSX-5, SCP-1 and CT-7, CD20 and c-erbB-2.

**[00128]** In some embodiments, the effector protein is a diagnostic protein. In some embodiments, the diagnostic protein is a detectable protein. In some embodiments, a detectable protein is a fluorescent protein. A fluorescent protein is a protein that emits a fluorescent light when exposed to a light source at an appropriate wavelength (e.g., light in the blue or ultraviolet range). Suitable fluorescent proteins that may be used in accordance with the present disclosure include, without limitation, eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, and mHoneydew. In some embodiments, a detectable protein is an enzyme that hydrolyzes an substrate to produce a detectable signal (e.g., a chemiluminescent signal). Such enzymes include, without limitation, beta-galactosidase (encoded by LacZ), horseradish peroxidase, or luciferase.

**[00129]** The examples of therapeutic proteins or diagnostic proteins provided herein are not meant to be limiting. Any therapeutic proteins known to one skilled in the art may be delivered into a cell using the conjugates, compositions, and methods described herein. In some embodiments, an effector protein by itself may not be able to enter a cell, but is able to enter a cell when associated with a supernegatively charged protein described herein (e.g., prothymosin alpha (ProT $\alpha$ ), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and variants thereof), for example, via a covalent bond or a non-covalent interaction. In some embodiments, the supernegatively charged protein is non-covalently associated with the effector protein. Non-covalent association results in a complex comprising the effector protein and supernegatively charged protein.

**[00130]** In some embodiments, the effector protein to be delivered is contacted with the supernegatively charged protein to form a complex. In some embodiments, formation of

complexes is carried out at or around pH 7. In some embodiments, formation of complexes is carried out at about pH 5, about pH 6, about pH 7, about pH 8, or about pH 9. Formation of complexes is typically carried out at a pH that does not negatively affect the function of the supernegatively charged protein and/or the effector protein. In some embodiments, formation of complexes is carried out at room temperature. In some embodiments, formation of complexes is carried out at or around 37 °C. In some embodiments, formation of complexes is carried out below 4 °C, at about 4 °C, at about 10 °C, at about 15 °C, at about 20 °C, at about 25 °C, at about 30 °C, at about 35 °C, at about 37 °C, at about 40 °C, or higher than 40 °C.

Formation of complexes is typically carried out at a temperature that does not negatively affect the function of the supernegatively charged protein and/or effector protein. In some embodiments, formation of complexes is carried out in serum-free medium.

**[00131]** In some embodiments, formation of complexes is carried out using concentrations of effector protein of about 100 nM. In some embodiments, formation of complexes is carried out using concentrations of effector protein of about 25 nM, about 50 nM, about 75 nM, about 90 nM, about 100 nM, about 110 nM, about 125 nM, about 150 nM, about 175 nM, or about 200 nM. In some embodiments, formation of complexes is carried out using concentrations of supernegatively charged protein of about 40 nM. In some embodiments, formation of complexes is carried out using concentrations of supernegatively charged protein of about 10 nM, about 20 nM, about 30 nM, about 40 nM, about 50 nM, about 60 nM, about 70 nM, about 80 nM, about 90 nM, or about 100 nM.

**[00132]** In some embodiments, formation of complexes is carried out under conditions of excess effector protein. In some embodiments, formation of complexes is carried out with ratios of effector protein:supernegatively charged protein of about 20:1, about 10:1, about 9:1, about 8:1, about 7:1, about 6:1, about 5:1, about 4:1, about 3:1, about 2:1, or about 1:1. In some embodiments, formation of complexes is carried out with ratios of effector protein:supernegatively charged protein of about 3:1. In some embodiments, formation of complexes is carried out with ratios of supernegatively charged protein: effector protein of about 20:1, about 10:1, about 9:1, about 8:1, about 7:1, about 6:1, about 5:1, about 4:1, about 3:1, about 2:1, or about 1:1.

**[00133]** In some embodiments, formation of complexes is carried out by mixing supernegatively charged protein with effector protein, and agitating the mixture (*e.g.*, by inversion). In some embodiments, formation of complexes is carried out by mixing supernegatively charged protein with effector protein, and allowing the mixture to sit still. In some embodiments, the formation of the complex is carried out in the presence of a

pharmaceutically acceptable carrier or excipient. In some embodiments, the complex is further combined with a pharmaceutically acceptable carrier or excipient. Exemplary excipients or carriers include water, solvents, lipids, proteins, peptides, endosomolytic agents (*e.g.*, chloroquine, pyrene butyric acid), small molecules, carbohydrates, buffers, natural polymers, synthetic polymers (*e.g.*, PLGA, polyurethane, polyesters, polycaprolactone, polyphosphazenes), pharmaceutical agents, *etc.*

**[00134]** In some embodiments, complexes comprising supernegatively charged protein and effector protein may migrate more slowly in gel electrophoresis assays than either the supernegatively charged protein alone or the effector protein alone.

**[00135]** In some embodiments, the supernegatively charged protein is covalently associated with the effector protein. For example, a supernegatively charged protein may be fused to an effector protein to be delivered. Covalent attachment may be direct or indirect (*e.g.*, through a linker). In some embodiments, a covalent attachment is mediated through one or more linkers.

**[00136]** A “linker” refers to a chemical group or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein. Typically, the linker is positioned between, or flanked by, two groups, molecules, domains, or other moieties and connected to each one via a covalent bond, thus connecting the two. The linker may be as simple as a covalent bond, or it may be a polymeric linker many atoms in length.

**[00137]** In some embodiments, the linker is a polypeptide or based on amino acids. In some embodiments, the linker is not peptide-like. In some embodiments, the linker is a covalent bond (*e.g.*, a carbon-carbon bond, disulfide bond, carbon-heteroatom bond, *etc.*). In some embodiments, the linker is a carbon-nitrogen bond of an amide linkage. In some embodiments, the linker is a cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic or heteroaliphatic linker. In some embodiments, the linker is polymeric (*e.g.*, polyethylene, polyethylene glycol, polyamide, polyester, *etc.*). In some embodiments, the linker comprises a monomer, dimer, or polymer of aminoalkanoic acid. In some embodiments, the linker comprises an aminoalkanoic acid (*e.g.*, glycine, ethanoic acid, alanine, beta-alanine, 3-aminopropanoic acid, 4-aminobutanoic acid, 5-pentanoic acid, *etc.*). In some embodiments, the linker comprises a monomer, dimer, or polymer of aminohexanoic acid (Ahx). In some embodiments, the linker is based on a carbocyclic moiety (*e.g.*, cyclopentane, cyclohexane). In other embodiments, the linker comprises a polyethylene glycol moiety (PEG). In other embodiments, the linker comprises amino acids. In some embodiments, the linker comprises a peptide. In some embodiments, the linker comprises an aryl or heteroaryl moiety. In some embodiments, the linker is based on a phenyl ring. The linker may include functionalized

moieties to facilitate attachment of a nucleophile (e.g., thiol, amino) from the peptide to the linker. Any electrophile may be used as part of the linker. Exemplary electrophiles include, but are not limited to, activated esters, activated amides, Michael acceptors, alkyl halides, aryl halides, acyl halides, and isothiocyanates.

**[00138]** In some embodiments, the linker is an amino acid or a plurality of amino acids (e.g., a peptide or protein). In some embodiments, the linker is a bond (e.g., a covalent bond), an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 1-100 amino acids in length, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-110, 110-120, 120-130, 130-140, 140-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated.

**[00139]** In some embodiments, a linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 9), which may also be referred to as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence SGGS (SEQ ID NO: 65). In some embodiments, a linker comprises (SGGS)<sub>n</sub> (SEQ ID NO: 66), (GGGS)<sub>n</sub> (SEQ ID NO: 67), (GGGGGS)<sub>n</sub> (SEQ ID NO: 68), (G)<sub>n</sub> (SEQ ID NO: 69), (EAAAK)<sub>n</sub> (SEQ ID NO: 70), (GGS)<sub>n</sub> (SEQ ID NO: 71), SGSETPGTSESATPES (SEQ ID NO: 9), or (XP)<sub>n</sub> motif (SEQ ID NO: 72), or a combination of any of these, wherein n is independently an integer between 1 and 30, and wherein X is any amino acid. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, a linker comprises SGSETPGTSESATPES (SEQ ID NO: 9), and SGGS (SEQ ID NO: 65). In some embodiments, a linker comprises SGGSSGSETPGTSESATPES (SEQ ID NO: 73). In some embodiments, a linker comprises SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 74). In some embodiments, a linker comprises

GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS (SEQ ID NO: 75).

**[00140]** In some embodiments, the linker is a cleavable linker. In some embodiments, the cleavable linker comprises an amide, ester, or disulfide bond. For example, the linker may be an amino acid sequence that is cleavable by a cellular enzyme. In some embodiments, the enzyme is a protease. In other embodiments, the enzyme is an esterase. In some embodiments, the enzyme is one that is more highly expressed in certain cell types than in other cell types. For example, the enzyme may be one that is more highly expressed in tumor cells than in non-tumor cells. Exemplary linkers and enzymes that cleave those linkers are presented below.

**[00141]** In some embodiments, the linker comprises a bond that can be cleaved upon exposure to a cleaving activity, such as UV light or a hydrolytic enzyme, such as a lysosomal protease. For example, the linker may be a UV-cleavable linker or a linker that is cleaved by a lysosomal enzyme (e.g., a protease or esterase). In some embodiments, the cleavable linker comprises an amide, ester, or disulfide bond. Non-limiting, exemplary cleavable linkers are provided in Table 2.

Table 2. Cleavable linker

| Linker Sequence                           | Enzyme(s) Targeting Linker                                                                                                                                                                                     |
|-------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| X <sup>1</sup> -AGVF-X<br>(SEQ ID NO: 76) | lysosomal thiol proteinases (see, e.g., Duncan <i>et al.</i> , 1982, <i>Biosci. Rep.</i> , 2:1041-46; incorporated herein by reference)                                                                        |
| X-GFLG-X<br>(SEQ ID NO: 77)               | lysosomal cysteine proteinases (see, e.g., Vasey <i>et al.</i> , <i>Clin. Canc. Res.</i> , 1999, 5:83-94; incorporated herein by reference)                                                                    |
| X-FK-X                                    | Cathepsin B – ubiquitous, overexpressed in many solid tumors, such as breast cancer (see, e.g., Dubowchik <i>et al.</i> , 2002, <i>Bioconjugate Chem.</i> , 13:855-69; incorporated herein by reference)       |
| X-A*L-X                                   | Cathepsin B – ubiquitous, overexpressed in many solid tumors, such as breast cancer (see, e.g., Trouet <i>et al.</i> , 1982, <i>Proc. Natl. Acad. Sci., USA</i> , 79:626-29; incorporated herein by reference) |
| X-A*LA*L-X<br>(SEQ ID NO: 78)             | Cathepsin B – ubiquitous, overexpressed in many solid tumors (see, e.g., Schmid <i>et al.</i> , 2007, <i>Bioconjugate Chem.</i> , 18:702-16; incorporated herein by reference)                                 |
| X-AL*AL*A-X<br>(SEQ ID NO: 79)            | Cathepsin D – ubiquitous (see, e.g., Czerwinski <i>et al.</i> , 1998, <i>Proc. Natl. Acad. Sci., USA</i> , 95:11520-25; incorporated herein by reference)                                                      |

\* indicates cleavage site

**[00142]** In some embodiments, other proteins or peptides are fused to the supernegatively charged protein or to a fusion protein comprising a supernegatively charged protein and an effector protein. For example, a targeting peptide may be fused to the supernegatively charged protein in order to selectively deliver an effector protein to a particular cell type. Peptides or proteins that enhance cellular uptake of the effector protein may also be used. In some embodiments, the peptide fused to the supernegatively charged protein is a peptide hormone. In some embodiments, the peptide fused to the supernegatively charged protein is a peptide ligand.

**[00143]** Further provided herein are nucleic acids encoding the fusion protein comprising the supernegatively charged protein and the effector protein. The fusion protein can generally be produced by expression from recombinant nucleic acids in appropriate cells (e.g., bacterial cell or eukaryotic cells) and isolated. To produce the fusion protein, nucleic acids encoding the fusion protein may be introduced to a cell (e.g., a bacterial cell or a eukaryotic cell such as a yeast cell or an insect cell). The cells may be cultured under conditions that allow the fusion protein to express from the nucleic acids encoding the fusion protein. Fusion proteins comprising a signal peptide can be secreted, e.g., into the culturing media and can subsequently be recovered. The fusion protein may be isolated using any methods of purifying a protein known in the art.

**[00144]** The nucleic acids encoding the fusion protein described herein may be obtained, and the nucleotide sequence of the nucleic acids determined, by any method known in the art. One skilled in the art is able to identify the nucleotide sequence encoding the fusion protein from the amino acid sequence of the fusion protein. The nucleic acids encoding the fusion protein of the present disclosure, may be DNA or RNA, double-stranded or single stranded. In some embodiments, the nucleotide sequence encoding the fusion protein may be codon optimized to adapt to different expression systems (e.g., for mammalian expression).

**[00145]** In some embodiments, the nucleic acid is comprised within a vector, such as an expression vector. In some embodiments, the vector comprises a promoter operably linked to the nucleic acid.

**[00146]** A variety of promoters can be used for expression of the fusion proteins described herein, including, but not limited to, cytomegalovirus (CMV) intermediate early promoter, a viral LTR such as the Rous sarcoma virus LTR, HIV-LTR, HTLV-1 LTR, the simian virus 40 (SV40) early promoter, E. coli lac UV5 promoter, and the herpes simplex tk virus promoter.

**[00147]** Regulatable promoters can also be used. Such regulatable promoters include those using the lac repressor from E. coli as a transcription modulator to regulate transcription from lac operator-bearing mammalian cell promoters (Brown, M. et al., Cell, 49:603-612 (1987)), those using the tetracycline repressor (tetR) (Gossen, M., and Bujard, H., Proc. Natl. Acad. Sci. USA 89:5547-5551 (1992); Yao, F. et al., Human Gene Therapy, 9:1939-1950 (1998); Shockelt, P., et al., Proc. Natl. Acad. Sci. USA, 92:6522-6526 (1995)). Other systems include FK506 dimer, VP16 or p65 using estradiol, RU486, diphenol murislerone, or rapamycin. Inducible systems are available from Invitrogen, Clontech, and Ariad.

**[00148]** Regulatable promoters that include a repressor with the operon can be used. In one embodiment, the lac repressor from Escherichia coli can function as a transcriptional

modulator to regulate transcription from lac operator-bearing mammalian cell promoters (M. Brown et al., *Cell*, 49:603-612 (1987)); Gossen and Bujard (1992); (M. Gossen et al., *Natl. Acad. Sci. USA*, 89:5547-5551 (1992)) combined the tetracycline repressor (tetR) with the transcription activator (VP 16) to create a tetR-mammalian cell transcription activator fusion protein, tTa (tetR-VP 16), with the tetO-bearing minimal promoter derived from the human cytomegalovirus (hCMV) major immediate-early promoter to create a tetR-tet operator system to control gene expression in mammalian cells. In one embodiment, a tetracycline inducible switch is used (Yao et al., *Human Gene Therapy*; Gossen et al., *Natl. Acad. Sci. USA*, 89:5547-5551 (1992); Shockett et al., *Proc. Natl. Acad. Sci. USA*, 92:6522-6526 (1995)).

**[00149]** Additionally, the vector can contain, for example, some or all of the following: a selectable marker gene, such as the neomycin gene for selection of stable or transient transfectants in mammalian cells; enhancer/promoter sequences from the immediate early gene of human CMV for high levels of transcription; transcription termination and RNA processing signals from SV40 for mRNA stability; SV40 polyoma origins of replication and Cole1 for proper episomal replication; internal ribosome binding sites (IRESes), versatile multiple cloning sites; and T7 and SP6 RNA promoters for in vitro transcription of sense and antisense RNA. Suitable vectors and methods for producing vectors containing transgenes are well known and available in the art.

**[00150]** An expression vector comprising the nucleic acid can be transferred to a host cell by conventional techniques (e.g., electroporation, liposomal transfection, and calcium phosphate precipitation) and the transfected cells are then cultured by conventional techniques to produce the fusion proteins described herein. In some embodiments, the expression of the fusion proteins described herein is regulated by a constitutive, an inducible, or a tissue-specific promoter.

**[00151]** The host cells used to express the fusion proteins described herein may be either bacterial cells such as *Escherichia coli*, or, preferably, eukaryotic cells. In particular, mammalian cells, such as Chinese hamster ovary cells (CHO), in conjunction with a vector such as the major intermediate early gene promoter element from human cytomegalovirus is an effective expression system for immunoglobulins (Foecking et al. (1986) "Powerful And Versatile Enhancer-Promoter Unit For Mammalian Expression Vectors," *Gene* 45:101-106; Cockett et al. (1990) "High Level Expression Of Tissue Inhibitor Of Metalloproteinases In Chinese Hamster Ovary Cells Using Glutamine Synthetase Gene Amplification," *Biotechnology* 8:662-667).

**[00152]** A variety of host-expression vector systems may be utilized to express the fusion proteins described herein. Such host-expression systems represent vehicles by which the coding sequences of the isolated fusion proteins described herein may be produced and subsequently purified, but also represent cells which may, when transformed or transfected with the appropriate nucleotide coding sequences, express the fusion proteins described herein in situ. These include, but are not limited to, microorganisms such as bacteria (e.g., *E. coli* and *B. subtilis*) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing coding sequences for the fusion proteins described herein; yeast (e.g., *Saccharomyces pichia*) transformed with recombinant yeast expression vectors containing sequences encoding the fusion proteins described herein; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing the sequences encoding the fusion proteins described herein; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus (CaMV) and tobacco mosaic virus (TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing sequences encoding the fusion proteins described herein; or mammalian cell systems (e.g., COS, CHO, BHK, 293, 293T, 3T3 cells, lymphotic cells (see U.S. Pat. No. 5,807,715), Per C.6 cells (human retinal cells developed by Crucell) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter).

**[00153]** In bacterial systems, a number of expression vectors may be advantageously selected depending upon the use intended for the fusion proteins being expressed. For example, when a large quantity of such a protein is to be produced, for the generation of pharmaceutical compositions of fusion proteins described herein, vectors which direct the expression of high levels of fusion protein products that are readily purified may be desirable. Such vectors include, but are not limited, to the *E. coli* expression vector pUR278 (Rüther et al. (1983) "Easy Identification Of cDNA Clones," EMBO J. 2:1791-1794), in which the coding sequence may be ligated individually into the vector in frame with the lac Z coding region so that a fusion protein is produced; pIN vectors (Inouye et al. (1985) "Up-Promoter Mutations In The lpp Gene Of Escherichia Coli," Nucleic Acids Res. 13:3101-3110; Van Heeke et al. (1989) "Expression Of Human Asparagine Synthetase In Escherichia Coli," J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption and binding to a matrix glutathione-

agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

**[00154]** In an insect system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes. The virus grows in *Spodoptera frugiperda* cells. The coding sequence may be cloned individually into non-essential regions (e.g., the polyhedrin gene) of the virus and placed under control of an AcNPV promoter (e.g., the polyhedrin promoter).

**[00155]** In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, the coding sequence of interest may be ligated to an adenovirus transcription/translation control complex, e.g., the late promoter and tripartite leader sequence. This chimeric gene may then be inserted in the adenovirus genome by in vitro or in vivo recombination. Insertion in a non-essential region of the viral genome (e.g., region E1 or E3) will result in a recombinant virus that is viable and capable of expressing the immunoglobulin molecule in infected hosts (e.g., see Logan et al. (1984) "Adenovirus Tripartite Leader Sequence Enhances Translation Of mRNAs Late After Infection," Proc. Natl. Acad. Sci. USA 81:3655-3659). Specific initiation signals may also be required for efficient translation of inserted antibody coding sequences. These signals include the ATG initiation codon and adjacent sequences. Furthermore, the initiation codon must be in phase with the reading frame of the desired coding sequence to ensure translation of the entire insert. These exogenous translational control signals and initiation codons can be of a variety of origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements, transcription terminators, etc. (see Bitter et al. (1987) "Expression And Secretion Vectors For Yeast," Methods in Enzymol. 153:516-544).

**[00156]** In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may be important for the function of the protein. Purification and modification of recombinant proteins is well known in the art such that the design of the polyprotein precursor could include a number of embodiments readily appreciated by a skilled worker. Any known proteases or peptidases known in the art can be used for the described modification of the precursor molecule, e.g., thrombin or factor Xa (Nagai et al. (1985) "Oxygen Binding Properties Of Human Mutant Hemoglobins Synthesized In Escherichia Coli," Proc. Nat. Acad. Sci. USA

82:7252-7255, and reviewed in Jenny et al. (2003) "A Critical Review Of The Methods For Cleavage Of Fusion Proteins With Thrombin And Factor Xa," Protein Expr. Purif. 31:1-11, each of which is incorporated by reference herein in its entirety)), enterokinase (Collins-Racie et al. (1995) "Production Of Recombinant Bovine Enterokinase Catalytic Subunit In Escherichia Coli Using The Novel Secretory Fusion Partner DsbA," Biotechnology 13:982-987 hereby incorporated by reference herein in its entirety)), furin, and AcTEV (Parks et al. (1994) "Release Of Proteins And Peptides From Fusion Proteins Using A Recombinant Plant Virus Proteinase," Anal. Biochem. 216:413-417; hereby incorporated by reference herein in its entirety)) and the Foot and Mouth Disease Virus Protease C3.

**[00157]** Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells which possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product may be used. Such mammalian host cells include but are not limited to CHO, VERY, BHK, HeLa, COS, MDCK, 293, 293T, 3T3, WI38, BT483, Hs578T, HTB2, BT20 and T47D, CRL7030 and Hs578Bst.

**[00158]** A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler et al. (1977) "Transfer Of Purified Herpes Virus Thymidine Kinase Gene To Cultured Mouse Cells," Cell 11: 223-232), hypoxanthine-guanine phosphoribosyltransferase (Szybalska et al. (1992) "Use Of The HPRT Gene And The HAT Selection Technique In DNA-Mediated Transformation Of Mammalian Cells First Steps Toward Developing Hybridoma Techniques And Gene Therapy," Bioessays 14: 495-500), and adenine phosphoribosyltransferase (Lowy et al. (1980) "Isolation Of Transforming DNA: Cloning The Hamster aprt Gene," Cell 22: 817-823) genes can be employed in tk-, hgppt- or aprt- cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for the following genes: dhfr, which confers resistance to methotrexate (Wigler et al. (1980) "Transformation Of Mammalian Cells With An Amplifiable Dominant-Acting Gene," Proc. Natl. Acad. Sci. USA 77:3567-3570; O'Hare et al. (1981) "Transformation Of Mouse Fibroblasts To Methotrexate Resistance By A Recombinant Plasmid Expressing A Prokaryotic Dihydrofolate Reductase," Proc. Natl. Acad. Sci. USA 78: 1527-1531); gpt, which confers resistance to mycophenolic acid (Mulligan et al. (1981) "Selection For Animal Cells That Express The Escherichia coli Gene Coding For Xanthine-Guanine Phosphoribosyltransferase," Proc. Natl. Acad. Sci. USA 78: 2072-2076); neo, which confers resistance to the

aminoglycoside G-418 (Tolstoshev (1993) "Gene Therapy, Concepts, Current Trials And Future Directions," *Ann. Rev. Pharmacol. Toxicol.* 32:573-596; Mulligan (1993) "The Basic Science Of Gene Therapy," *Science* 260:926-932; and Morgan et al. (1993) "Human Gene Therapy," *Ann. Rev. Biochem.* 62:191-217) and hygromycin, which confers resistance to hygromycin (Santerre et al. (1984) "Expression Of Prokaryotic Genes For Hygromycin B And G418 Resistance As Dominant-Selection Markers In Mouse L Cells," *Gene* 30:147-156).

Methods commonly known in the art of recombinant DNA technology which can be used are described in Ausubel et al. (eds.), 1993, *Current Protocols in Molecular Biology*, John Wiley & Sons, NY; Kriegler, 1990, *Gene Transfer and Expression, A Laboratory Manual*, Stockton Press, NY; and in Chapters 12 and 13, Dracopoli et al. (eds), 1994, *Current Protocols in Human Genetics*, John Wiley & Sons, NY.; Colberre-Garapin et al. (1981) "A New Dominant Hybrid Selective Marker For Higher Eukaryotic Cells," *J. Mol. Biol.* 150:1-14.

**[00159]** The expression levels of the fusion protein described herein can be increased by vector amplification (for a review, see Bebbington and Hentschel, *The use of vectors based on gene amplification for the expression of cloned genes in mammalian cells in DNA cloning*, Vol. 3 (Academic Press, New York, 1987). When a marker in the vector system expressing a fusion protein described herein is amplifiable, increase in the level of inhibitor present in culture of host cell will increase the number of copies of the marker gene. Since the amplified region is associated with the nucleotide sequence of a fusion protein described herein or a fusion protein described herein, production of the fusion protein will also increase (Crouse et al. (1983) "Expression And Amplification Of Engineered Mouse Dihydrofolate Reductase Minigenes," *Mol. Cell. Biol.* 3:257-266).

**[00160]** Once a fusion protein described herein has been recombinantly expressed, it may be purified by any method known in the art for purification of polypeptides, polypeptides or antibodies (e.g., analogous to antibody purification schemes based on antigen selectivity) for example, by chromatography (e.g., ion exchange, affinity, particularly by affinity for the specific antigen (optionally after Protein A selection where the polypeptide comprises an Fc domain (or portion thereof)), and sizing column chromatography), centrifugation, differential solubility, or by any other standard technique for the purification of polypeptides or antibodies.

**[00161]** In some embodiments, to facilitate purification, e.g., by affinity chromatography, the fusion protein described herein further contains a fusion domain. Well known examples of such fusion domains include, without limitation, polyhistidine, Glu-Glu, glutathione S transferase (GST), thioredoxin, protein A, protein G, an immunoglobulin heavy chain constant region (Fc), maltose binding protein (MBP), or human serum albumin. A fusion domain may

be selected so as to confer a desired property. For example, some fusion domains are particularly useful for isolation of the fusion proteins by affinity chromatography. For the purpose of affinity purification, relevant matrices for affinity chromatography, such as glutathione-, amylase-, and nickel- or cobalt- conjugated resins are used. Many of such matrices are available in “kit” form, such as the Pharmacia GST purification system and the QIAexpress™ system (Qiagen) useful with (HIS6) fusion partners.

**[00162]** In some embodiments, the complex or the fusion protein comprising the supernegatively charged protein and the effector protein has an overall net negative charge. In some embodiments, the complex further comprises a cationic lipid and/or cationic polymer.

**[00163]** The present disclosure further provides compositions comprising a supernegatively charged protein associated with an effector protein to be delivered. In some embodiments, the composition further comprises a cationic polymer or cationic lipids. A “cationic lipid” refers to a lipid which has a cationic, or positive, charge at physiologic pH. Cationic lipids can take a variety of forms including, but not limited to, liposomes or micelles. Cationic lipids useful for certain aspects of the present disclosure are known in the art, and, generally comprise both polar and non-polar domains, bind to polyanions, such as nucleic acid molecules or supernegatively charged proteins, and are typically known to facilitate the delivery of nucleic acids into cells. Any cationic lipid formulations known in the art for delivery a biological molecule to a cell may be used in accordance with the present disclosure. For example, Wang et al. (PNAS, 113 (11) 2868-2873; 2016, incorporated herein by reference) describes cationic lipid formulations that may be used to deliver gene-editing agents to the cell and such cationic lipid formulations can be used in accordance with the present disclosure. Other non-limiting examples of useful cationic lipids include polyethylenimine, polyamidoamine (PAMAM) starburst dendrimers, Lipofectin (a combination of DOTMA and DOPE), Lipofectase, LIPOFECTAMINE® (e.g., LIPOFECTAMINE® 2000, LIPOFECTAMINE® 3000, LIPOFECTAMINE® RNAiMAX, LIPOFECTAMINE® LTX, Lipofectamine® MessengerMax, Lipofectamine® CRISPRMAX, from ThermoFisher ), SAINT-RED (Synvolux Therapeutics, Groningen Netherlands), DOPE, Cytfectin (Gilead Sciences, Foster City, Calif.), Eufectins (JBL, San Luis Obispo, Calif.), and FuGENE HD (Promega). Exemplary cationic liposomes can be made from N-[1-(2,3-dioleoloxo)-propyl]-N,N,N-trimethylammonium chloride (DOTMA), N-[1-(2,3-dioleoloxo)-propyl]-N,N,N-trimethylammonium methylsulfate (DOTAP), 3β-[N-(N',N'-dimethylaminoethane)carbamoyl]cholesterol (DC-Chol), 2,3,-dioleyloxy-N-[2(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate (DOSPA),

1,2-dimyristyloxypropyl-3-dimethyl-hydroxyethyl ammonium bromide; and dimethyldioctadecylammonium bromide (DDAB). Cationic lipids have been used in the art to deliver nucleic acid molecules to cells (see, e.g., U.S. Pat. Nos. 5,855,910; 5,851,548; 5,830,430; 5,780,053; 5,767,099; 8,569,256; 8,691,750; 8,748,667; 8,758,810; 8,759,104; 8,771,728; Lewis et al. 1996. Proc. Natl. Acad. Sci. USA 93:3176; Hope et al. 1998. Molecular Membrane Biology 15:1). In addition, other lipid compositions are also known in the art and include, e.g., those taught in U.S. Pat. No. 4,235,871; U.S. Pat. Nos. 4,501,728; 4,837,028; 4,737,323.

**[00164]** A “cationic polymer,” as used herein, refers to a polymer having a net positive charge. Cationic polymers are well known in the art, and include those described in Samal et al., Cationic polymers and their therapeutic potential. Chem Soc Rev. 2012 Nov 7;41(21):7147-94; in published U.S. patent applications U.S. 2014/0141487 A1, U.S. 2014/0141094 A1, U.S. 2014/0044793 A1, U.S. 2014/0018404 A1, U.S. 2014/0005269 A1, and U.S. 2013/0344117 A1; and in U.S. Pat. Nos. 8,709,466; 8,728,526; 8,759,103; and 8,790,664; the entire contents of each are incorporated herein by reference. Exemplary cationic polymers include, but are not limited to, polyallylamine (PAH); polyethyleneimine (PEI); poly(L-lysine) (PLL); poly(L-arginine) (PLA); polyvinylamine homo- or copolymer; a poly(vinylbenzyl-tri-C1-C4-alkylammonium salt); a polymer of an aliphatic or araliphatic dihalide and an aliphatic N,N,N',N'-tetra-C1-C4-alkyl-alkylenediamine; a poly(vinylpyridin) or poly(vinylpyridinium salt); a poly(N,N-diallyl-N,N-di-C1-C4-alkyl-ammoniumhalide); a homo- or copolymer of a quaternized di-C1-C4-alkyl-aminoethyl acrylate or methacrylate; POLYQUADTM; a polyaminoamide; and the like.

**[00165]** In some embodiments, the composition is the composition is a pharmaceutical composition formulated for administration to a subject to deliver the effector protein to the subject (e.g., deliver to a cell in the subject). For example, the composition may be formulated for administration to a subject for diagnosing or treating a disease. Pharmaceutical compositions may optionally comprise one or more additional therapeutically active substances. In accordance with some embodiments, a method of administering pharmaceutical compositions comprising one or more supernegatively charged proteins associated with an effector protein to be delivered to a subject in need thereof is provided. In some embodiments, compositions are administered to humans.

**[00166]** Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally

suitable for administration to animals of all sorts. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, including commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as chickens, ducks, geese, and/or turkeys.

**[00167]** Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, shaping and/or packaging the product into a desired single- or multi-dose unit.

**[00168]** A pharmaceutical composition in accordance with the invention may be prepared, packaged, and/or sold in bulk, as a single unit dose, and/or as a plurality of single unit doses. As used herein, a “unit dose” is discrete amount of the pharmaceutical composition comprising a predetermined amount of the active ingredient. The amount of the active ingredient is generally equal to the dosage of the active ingredient which would be administered to a subject and/or a convenient fraction of such a dosage such as, for example, one-half or one-third of such a dosage.

**[00169]** Relative amounts of the active ingredient, the pharmaceutically acceptable excipient, and/or any additional ingredients in a pharmaceutical composition in accordance with the invention will vary, depending upon the identity, size, and/or condition of the subject treated and further depending upon the route by which the composition is to be administered. By way of example, the composition may comprise between 0.1% and 100% (w/w) active ingredient.

**[00170]** Pharmaceutical formulations may additionally comprise a pharmaceutically acceptable excipient, which, as used herein, includes any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, lubricants and the like, as suited to the particular dosage form desired. Remington’s The Science and Practice of Pharmacy, 21st Edition, A. R. Gennaro (Lippincott, Williams & Wilkins, Baltimore, MD, 2006; incorporated herein by reference) discloses various excipients used in formulating pharmaceutical compositions and known techniques for the preparation thereof. Except insofar

as any conventional excipient medium is incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition, its use is contemplated to be within the scope of this invention.

**[00171]** In some embodiments, a pharmaceutically acceptable excipient is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% pure. In some embodiments, an excipient is approved for use in humans and for veterinary use. In some embodiments, an excipient is approved by United States Food and Drug Administration. In some embodiments, an excipient is pharmaceutical grade. In some embodiments, an excipient meets the standards of the United States Pharmacopoeia (USP), the European Pharmacopoeia (EP), the British Pharmacopoeia, and/or the International Pharmacopoeia.

**[00172]** Pharmaceutically acceptable excipients used in the manufacture of pharmaceutical compositions include, but are not limited to, inert diluents, dispersing and/or granulating agents, surface active agents and/or emulsifiers, disintegrating agents, binding agents, preservatives, buffering agents, lubricating agents, and/or oils. Such excipients may optionally be included in pharmaceutical formulations. Excipients such as cocoa butter and suppository waxes, coloring agents, coating agents, sweetening, flavoring, and/or perfuming agents can be present in the composition, according to the judgment of the formulator. The “active ingredient” of the pharmaceutical composition described herein is, generally, the effector protein. In the context when the effector protein is a RNA-guided nuclease (e.g., Cas9) or a variant or fusion protein thereof, the active ingredient further includes the RNA that guides the RNA-guided nuclease to its target site.

**[00173]** Exemplary diluents include, but are not limited to, calcium carbonate, sodium carbonate, calcium phosphate, dicalcium phosphate, calcium sulfate, calcium hydrogen phosphate, sodium phosphate lactose, sucrose, cellulose, microcrystalline cellulose, kaolin, mannitol, sorbitol, inositol, sodium chloride, dry starch, cornstarch, powdered sugar, etc., and/or combinations thereof.

**[00174]** Exemplary granulating and/or dispersing agents include, but are not limited to, potato starch, corn starch, tapioca starch, sodium starch glycolate, clays, alginic acid, guar gum, citrus pulp, agar, bentonite, cellulose and wood products, natural sponge, cation-exchange resins, calcium carbonate, silicates, sodium carbonate, cross-linked poly(vinyl-pyrrolidone) (crospovidone), sodium carboxymethyl starch (sodium starch glycolate), carboxymethyl cellulose, cross-linked sodium carboxymethyl cellulose (croscarmellose), methylcellulose, pregelatinized starch (starch 1500), microcrystalline starch, water insoluble starch, calcium

carboxymethyl cellulose, magnesium aluminum silicate (Veegum), sodium lauryl sulfate, quaternary ammonium compounds, etc., and/or combinations thereof.

**[00175]** Exemplary surface active agents and/or emulsifiers include, but are not limited to, natural emulsifiers (e.g. acacia, agar, alginic acid, sodium alginate, tragacanth, chondrux, cholesterol, xanthan, pectin, gelatin, egg yolk, casein, wool fat, cholesterol, wax, and lecithin), colloidal clays (e.g. bentonite [aluminum silicate] and Veegum® [magnesium aluminum silicate]), long chain amino acid derivatives, high molecular weight alcohols (e.g. stearyl alcohol, cetyl alcohol, oleyl alcohol, triacetin monostearate, ethylene glycol distearate, glyceryl monostearate, and propylene glycol monostearate, polyvinyl alcohol), carbomers (e.g. carboxy polymethylene, polyacrylic acid, acrylic acid polymer, and carboxyvinyl polymer), carrageenan, cellulosic derivatives (e.g. carboxymethylcellulose sodium, powdered cellulose, hydroxymethyl cellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, methylcellulose), sorbitan fatty acid esters (e.g. polyoxyethylene sorbitan monolaurate (Tween®20), polyoxyethylene sorbitan (Tween®60), polyoxyethylene sorbitan monooleate (Tween®80), sorbitan monopalmitate (Span®40), sorbitan monostearate (Span®60), sorbitan tristearate (Span®65), glyceryl monooleate, sorbitan monooleate (Span®80), polyoxyethylene esters (e.g. polyoxyethylene monostearate (Myrj®45), polyoxyethylene hydrogenated castor oil, polyethoxylated castor oil, polyoxymethylene stearate, and Solutol®), sucrose fatty acid esters, polyethylene glycol fatty acid esters (e.g. Cremophor®), polyoxyethylene ethers, (e.g. polyoxyethylene lauryl ether (Brij®30)), poly(vinyl-pyrrolidone), diethylene glycol monolaurate, triethanolamine oleate, sodium oleate, potassium oleate, ethyl oleate, oleic acid, ethyl laurate, sodium lauryl sulfate, Pluronic®F 68, Poloxamer®188, cetrimonium bromide, cetylpyridinium chloride, benzalkonium chloride, docusate sodium, etc. and/or combinations thereof.

**[00176]** Exemplary binding agents include, but are not limited to, starch (e.g. cornstarch and starch paste); gelatin; sugars (e.g. sucrose, glucose, dextrose, dextrin, molasses, lactose, lactitol, mannitol,); natural and synthetic gums (e.g. acacia, sodium alginate, extract of Irish moss, panwar gum, ghatti gum, mucilage of isapol husks, carboxymethylcellulose, methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropyl cellulose, hydroxypropyl methylcellulose, microcrystalline cellulose, cellulose acetate, poly(vinyl-pyrrolidone), magnesium aluminum silicate (Veegum®), and larch arabogalactan); alginates; polyethylene oxide; polyethylene glycol; inorganic calcium salts; silicic acid; polymethacrylates; waxes; water; alcohol; etc.; and combinations thereof.

[00177] Exemplary preservatives may include, but are not limited to, antioxidants, chelating agents, antimicrobial preservatives, antifungal preservatives, alcohol preservatives, acidic preservatives, and/or other preservatives. Exemplary antioxidants include, but are not limited to, alpha tocopherol, ascorbic acid, acorbyl palmitate, butylated hydroxyanisole, butylated hydroxytoluene, monothioglycerol, potassium metabisulfite, propionic acid, propyl gallate, sodium ascorbate, sodium bisulfite, sodium metabisulfite, and/or sodium sulfite. Exemplary chelating agents include ethylenediaminetetraacetic acid (EDTA), citric acid monohydrate, disodium edetate, dipotassium edetate, edetic acid, fumaric acid, malic acid, phosphoric acid, sodium edetate, tartaric acid, and/or trisodium edetate. Exemplary antimicrobial preservatives include, but are not limited to, benzalkonium chloride, benzethonium chloride, benzyl alcohol, bronopol, cetrimide, cetylpyridinium chloride, chlorhexidine, chlorobutanol, chlorocresol, chloroxylonol, cresol, ethyl alcohol, glycerin, hexetidine, imidurea, phenol, phenoxyethanol, phenylethyl alcohol, phenylmercuric nitrate, propylene glycol, and/or thimerosal. Exemplary antifungal preservatives include, but are not limited to, butyl paraben, methyl paraben, ethyl paraben, propyl paraben, benzoic acid, hydroxybenzoic acid, potassium benzoate, potassium sorbate, sodium benzoate, sodium propionate, and/or sorbic acid. Exemplary alcohol preservatives include, but are not limited to, ethanol, polyethylene glycol, phenol, phenolic compounds, bisphenol, chlorobutanol, hydroxybenzoate, and/or phenylethyl alcohol. Exemplary acidic preservatives include, but are not limited to, vitamin A, vitamin C, vitamin E, beta-carotene, citric acid, acetic acid, dehydroacetic acid, ascorbic acid, sorbic acid, and/or phytic acid. Other preservatives include, but are not limited to, tocopherol, tocopherol acetate, deteroxime mesylate, cetrimide, butylated hydroxyanisole (BHA), butylated hydroxytoluene (BHT), ethylenediamine, sodium lauryl sulfate (SLS), sodium lauryl ether sulfate (SLES), sodium bisulfite, sodium metabisulfite, potassium sulfite, potassium metabisulfite, Glydant Plus®, Phenonip®, methylparaben, Germall®115, Germaben®II, Neolone™, Kathon™, and/or Euxyl®.

[00178] Exemplary buffering agents include, but are not limited to, citrate buffer solutions, acetate buffer solutions, phosphate buffer solutions, ammonium chloride, calcium carbonate, calcium chloride, calcium citrate, calcium gluconate, calcium gluceptate, calcium gluconate, D-gluconic acid, calcium glycerophosphate, calcium lactate, propanoic acid, calcium levulinate, pentanoic acid, dibasic calcium phosphate, phosphoric acid, tribasic calcium phosphate, calcium hydroxide phosphate, potassium acetate, potassium chloride, potassium gluconate, potassium mixtures, dibasic potassium phosphate, monobasic potassium phosphate, potassium phosphate mixtures, sodium acetate, sodium bicarbonate, sodium chloride, sodium

citrate, sodium lactate, dibasic sodium phosphate, monobasic sodium phosphate, sodium phosphate mixtures, tromethamine, magnesium hydroxide, aluminum hydroxide, alginic acid, pyrogen-free water, isotonic saline, Ringer's solution, ethyl alcohol, etc., and/or combinations thereof.

**[00179]** Exemplary lubricating agents include, but are not limited to, magnesium stearate, calcium stearate, stearic acid, silica, talc, malt, glyceryl behenate, hydrogenated vegetable oils, polyethylene glycol, sodium benzoate, sodium acetate, sodium chloride, leucine, magnesium lauryl sulfate, sodium lauryl sulfate, etc., and combinations thereof.

**[00180]** Exemplary oils include, but are not limited to, almond, apricot kernel, avocado, babassu, bergamot, black current seed, borage, cade, camomile, canola, caraway, carnauba, castor, cinnamon, cocoa butter, coconut, cod liver, coffee, corn, cotton seed, emu, eucalyptus, evening primrose, fish, flaxseed, geraniol, gourd, grape seed, hazel nut, hyssop, isopropyl myristate, jojoba, kukui nut, lavandin, lavender, lemon, litsea cubeba, macademia nut, mallow, mango seed, meadowfoam seed, mink, nutmeg, olive, orange, orange roughy, palm, palm kernel, peach kernel, peanut, poppy seed, pumpkin seed, rapeseed, rice bran, rosemary, safflower, sandalwood, sasquana, savoury, sea buckthorn, sesame, shea butter, silicone, soybean, sunflower, tea tree, thistle, tsubaki, vetiver, walnut, and wheat germ oils. Exemplary oils include, but are not limited to, butyl stearate, caprylic triglyceride, capric triglyceride, cyclomethicone, diethyl sebacate, dimethicone 360, isopropyl myristate, mineral oil, octyldodecanol, oleyl alcohol, silicone oil, and/or combinations thereof.

**[00181]** Liquid dosage forms for oral and parenteral administration include, but are not limited to, pharmaceutically acceptable emulsions, microemulsions, solutions, suspensions, syrups, and/or elixirs. In addition to active ingredients, liquid dosage forms may comprise inert diluents commonly used in the art such as, for example, water or other solvents, solubilizing agents and emulsifiers such as ethyl alcohol, isopropyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, dimethylformamide, oils (in particular, cottonseed, groundnut, corn, germ, olive, castor, and sesame oils), glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols and fatty acid esters of sorbitan, and mixtures thereof. Besides inert diluents, oral compositions can include adjuvants such as wetting agents, emulsifying and suspending agents, sweetening, flavoring, and/or perfuming agents. In some embodiments for parenteral administration, compositions are mixed with solubilizing agents such as Cremophor®, alcohols, oils, modified oils, glycols, polysorbates, cyclodextrins, polymers, and/or combinations thereof.

**[00182]** Injectable preparations, for example, sterile injectable aqueous or oleaginous suspensions may be formulated according to the known art using suitable dispersing agents, wetting agents, and/or suspending agents. Sterile injectable preparations may be sterile injectable solutions, suspensions, and/or emulsions in nontoxic parenterally acceptable diluents and/or solvents, for example, as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, U.S.P., and isotonic sodium chloride solution. Sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil can be employed including synthetic mono- or diglycerides. Fatty acids such as oleic acid can be used in the preparation of injectables.

**[00183]** Injectable formulations can be sterilized, for example, by filtration through a bacterial-retaining filter, and/or by incorporating sterilizing agents in the form of sterile solid compositions which can be dissolved or dispersed in sterile water or other sterile injectable medium prior to use.

**[00184]** In order to prolong the effect of an active ingredient, it is often desirable to slow the absorption of the active ingredient from subcutaneous or intramuscular injection. This may be accomplished by the use of a liquid suspension of crystalline or amorphous material with poor water solubility. The rate of absorption of the drug then depends upon its rate of dissolution which, in turn, may depend upon crystal size and crystalline form. Alternatively, delayed absorption of a parenterally administered drug form is accomplished by dissolving or suspending the drug in an oil vehicle. Injectable depot forms are made by forming microencapsule matrices of the drug in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are prepared by entrapping the drug in liposomes or microemulsions which are compatible with body tissues.

**[00185]** Compositions for rectal or vaginal administration are typically suppositories which can be prepared by mixing compositions with suitable non-irritating excipients such as cocoa butter, polyethylene glycol or a suppository wax which are solid at ambient temperature but liquid at body temperature and therefore melt in the rectum or vaginal cavity and release the active ingredient.

**[00186]** Solid dosage forms for oral administration include capsules, tablets, pills, powders, and granules. In such solid dosage forms, an active ingredient is mixed with at least one inert, pharmaceutically acceptable excipient such as sodium citrate or dicalcium phosphate and/or

fillers or extenders (e.g. starches, lactose, sucrose, glucose, mannitol, and silicic acid), binders (e.g. carboxymethylcellulose, alginates, gelatin, polyvinylpyrrolidinone, sucrose, and acacia), humectants (e.g. glycerol), disintegrating agents (e.g. agar, calcium carbonate, potato or tapioca starch, alginic acid, certain silicates, and sodium carbonate), solution retarding agents (e.g. paraffin), absorption accelerators (e.g. quaternary ammonium compounds), wetting agents (e.g. cetyl alcohol and glycerol monostearate), absorbents (e.g. kaolin and bentonite clay), and lubricants (e.g. talc, calcium stearate, magnesium stearate, solid polyethylene glycols, sodium lauryl sulfate), and mixtures thereof. In the case of capsules, tablets and pills, the dosage form may comprise buffering agents.

**[00187]** Solid compositions of a similar type may be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular weight polyethylene glycols and the like. Solid dosage forms of tablets, dragees, capsules, pills, and granules can be prepared with coatings and shells such as enteric coatings and other coatings well known in the pharmaceutical formulating art. They may optionally comprise opacifying agents and can be of a composition that they release the active ingredient(s) only, or preferentially, in a certain part of the intestinal tract, optionally, in a delayed manner.

Examples of embedding compositions which can be used include polymeric substances and waxes. Solid compositions of a similar type may be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular weight polyethylene glycols and the like.

**[00188]** Dosage forms for topical and/or transdermal administration of a composition may include ointments, pastes, creams, lotions, gels, powders, solutions, sprays, inhalants and/or patches. Generally, an active ingredient is admixed under sterile conditions with a pharmaceutically acceptable excipient and/or any needed preservatives and/or buffers as may be required. Additionally, the present invention contemplates the use of transdermal patches, which often have the added advantage of providing controlled delivery of a compound to the body. Such dosage forms may be prepared, for example, by dissolving and/or dispensing the compound in the proper medium. Alternatively or additionally, rate may be controlled by either providing a rate controlling membrane and/or by dispersing the compound in a polymer matrix and/or gel.

**[00189]** Suitable devices for use in delivering intradermal pharmaceutical compositions described herein include short needle devices. Intradermal compositions may be administered by devices which limit the effective penetration length of a needle into the skin and functional equivalents thereof. Jet injection devices which deliver liquid compositions to the dermis via a

liquid jet injector and/or via a needle which pierces the stratum corneum and produces a jet which reaches the dermis are suitable. Ballistic powder/particle delivery devices which use compressed gas to accelerate vaccine in powder form through the outer layers of the skin to the dermis are suitable. Alternatively or additionally, conventional syringes may be used in the classical mantoux method of intradermal administration.

**[00190]** Formulations suitable for topical administration include, but are not limited to, liquid and/or semi liquid preparations such as liniments, lotions, oil in water and/or water in oil emulsions such as creams, ointments and/or pastes, and/or solutions and/or suspensions. Topically-administrable formulations may, for example, comprise from about 1% to about 10% (w/w) active ingredient, although the concentration of active ingredient may be as high as the solubility limit of the active ingredient in the solvent. Formulations for topical administration may further comprise one or more of the additional ingredients described herein.

**[00191]** A pharmaceutical composition may be prepared, packaged, and/or sold in a formulation suitable for pulmonary administration via the buccal cavity. Such a formulation may comprise dry particles which comprise the active ingredient and which have a diameter in the range from about 0.5 nm to about 7 nm or from about 1 nm to about 6 nm. Such compositions are conveniently in the form of dry powders for administration using a device comprising a dry powder reservoir to which a stream of propellant may be directed to disperse the powder and/or using a self-propelling solvent/powder dispensing container such as a device comprising the active ingredient dissolved and/or suspended in a low-boiling propellant in a sealed container. Such powders comprise particles wherein at least 98% of the particles by weight have a diameter greater than 0.5 nm and at least 95% of the particles by number have a diameter less than 7 nm. Alternatively, at least 95% of the particles by weight have a diameter greater than 1 nm and at least 90% of the particles by number have a diameter less than 6 nm. Dry powder compositions may include a solid fine powder diluent such as sugar and are conveniently provided in a unit dose form.

**[00192]** Low boiling propellants generally include liquid propellants having a boiling point of below 65 °F at atmospheric pressure. Generally the propellant may constitute 50% to 99.9% (w/w) of the composition, and active ingredient may constitute 0.1% to 20% (w/w) of the composition. A propellant may further comprise additional ingredients such as a liquid non-ionic and/or solid anionic surfactant and/or a solid diluent (which may have a particle size of the same order as particles comprising the active ingredient).

**[00193]** Pharmaceutical compositions formulated for pulmonary delivery may provide an active ingredient in the form of droplets of a solution and/or suspension. Such formulations may be prepared, packaged, and/or sold as aqueous and/or dilute alcoholic solutions and/or suspensions, optionally sterile, comprising active ingredient, and may conveniently be administered using any nebulization and/or atomization device. Such formulations may further comprise one or more additional ingredients including, but not limited to, a flavoring agent such as saccharin sodium, a volatile oil, a buffering agent, a surface active agent, and/or a preservative such as methylhydroxybenzoate. Droplets provided by this route of administration may have an average diameter in the range from about 0.1 nm to about 200 nm.

**[00194]** Formulations described herein as being useful for pulmonary delivery are useful for intranasal delivery of a pharmaceutical composition. Another formulation suitable for intranasal administration is a coarse powder comprising the active ingredient and having an average particle from about 0.2  $\mu\text{m}$  to 500  $\mu\text{m}$ . Such a formulation is administered in the manner in which snuff is taken, i.e. by rapid inhalation through the nasal passage from a container of the powder held close to the nose.

**[00195]** Formulations suitable for nasal administration may, for example, comprise from about as little as 0.1% (w/w) and as much as 100% (w/w) of active ingredient, and may comprise one or more of the additional ingredients described herein. A pharmaceutical composition may be prepared, packaged, and/or sold in a formulation suitable for buccal administration. Such formulations may, for example, be in the form of tablets and/or lozenges made using conventional methods, and may, for example, 0.1% to 20% (w/w) active ingredient, the balance comprising an orally dissolvable and/or degradable composition and, optionally, one or more of the additional ingredients described herein. Alternately, formulations suitable for buccal administration may comprise a powder and/or an aerosolized and/or atomized solution and/or suspension comprising active ingredient. Such powdered, aerosolized, and/or aerosolized formulations, when dispersed, may have an average particle and/or droplet size in the range from about 0.1 nm to about 200 nm, and may further comprise one or more of any additional ingredients described herein.

**[00196]** A pharmaceutical composition may be prepared, packaged, and/or sold in a formulation suitable for ophthalmic administration. Such formulations may, for example, be in the form of eye drops including, for example, a 0.1/1.0% (w/w) solution and/or suspension of the active ingredient in an aqueous or oily liquid excipient. Such drops may further comprise buffering agents, salts, and/or one or more other of any additional ingredients described herein. Other ophthalmically-administrable formulations which are useful include those which comprise

the active ingredient in microcrystalline form and/or in a liposomal preparation. Ear drops and/or eye drops are contemplated as being within the scope of this invention.

**[00197]** General considerations in the formulation and/or manufacture of pharmaceutical agents may be found, for example, in Remington: The Science and Practice of Pharmacy 21st ed., Lippincott Williams & Wilkins, 2005 (incorporated herein by reference).

**[00198]** Other aspects of the present disclosure provide methods comprising contacting a cell with the any of the compositions described herein, wherein the contacting results in the delivery of the effector protein into the cell. The contacting may be in vitro (e.g., using cells cultured in vitro), ex vivo (e.g., using cells isolated from a subject), or in vivo (e.g., contacting a cell in a subject). In some embodiments, the method further comprises detecting the effector protein in the cell. Those skilled in the art are familiar with methods of detecting the delivery of an effector protein into a cell, e.g., by western blot, functional assays, or immunostaining. In some embodiments, the cell may be a mammalian cell, such as a human cell. Other mammalian cells are also contemplated. Suitable cells and cell types for delivery of effector proteins according to some aspects of this disclosure include, but are not limited to, human cells, mammalian cells, T-cells, neurons, stem cells, progenitor cells, blood cells, fibroblasts, epithelial cells, neoplastic cells, and tumor cells.

**[00199]** While liposomal delivery of cargo such as DNA and RNA has been known to induce toxicity in targeted cells, the compositions described herein deliver the effector protein both in vitro and in vivo surprisingly with no or low toxicity. For example, in some embodiments, the compositions comprising an effector protein described herein exhibit low toxicity when administered to a population of cells (e.g., in vitro or in vivo). In some embodiments, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the cells in a population are viable following administration of an the composition described herein. One skilled in the art is familiar with methods for assessing the toxicity of a composition when administered to a population of cells.

**[00200]** In some aspects, the present disclosure provides methods comprising administering the fusion protein or complex comprising the supernegatively charged protein and the effector protein, or compositions comprising such to a subject in need thereof. In some embodiments, the effector protein in the composition is a therapeutic protein and the composition is administered for treating a disease. In some embodiments, the effector protein in the composition is a diagnostic protein and the composition is administered for diagnosing a disease.

**[00201]** Such compositions may be administered to a subject using any amount and any route of administration effective for preventing, treating, diagnosing, or imaging a disease, disorder, and/or condition. In some embodiments, the step of administering comprises a route of administration selected from the group consisting of oral, intravenous, intramuscular, intra-arterial, subcutaneous, intraventricular, topical, inhalational, and mucosal delivery.

**[00202]** The exact amount required will vary from subject to subject, depending on the species, age, and general condition of the subject, the severity of the disease, the particular composition, its mode of administration, its mode of activity, and the like. Compositions in accordance with the invention are typically formulated in dosage unit form for ease of administration and uniformity of dosage. It will be understood, however, that the total daily usage of the compositions of the present invention will be decided by the attending physician within the scope of sound medical judgment. The specific therapeutically effective, prophylactically effective, or appropriate imaging dose level for any particular patient will depend upon a variety of factors including the disorder being treated and the severity of the disorder; the activity of the specific compound employed; the specific composition employed; the age, body weight, general health, sex and diet of the patient; the time of administration, route of administration, and rate of excretion of the specific compound employed; the duration of the treatment; drugs used in combination or coincidental with the specific compound employed; and like factors well known in the medical arts.

**[00203]** Non-limiting, exemplary routes include oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, subcutaneous, intraventricular, transdermal, interdermal, rectal, intravaginal, intraperitoneal, topical (e.g., by powders, ointments, creams, gels, lotions, and/or drops), mucosal, nasal, buccal, enteral, vitreal, intratumoral, sublingual; by intratracheal instillation, bronchial instillation, and/or inhalation; as an oral spray, nasal spray, and/or aerosol, and/or through a portal vein catheter. In some embodiments, the administration is via systemic intravenous injection. In some embodiments, the administration is oral. In some embodiments, the fusion protein or complex comprising the supernegatively charged protein or compositions comprising such may be administered in a way which allows the effector protein to cross the blood-brain barrier, vascular barrier, or other epithelial barrier.

**[00204]** In some embodiments, the fusion protein or complex comprising the supernegatively charged protein or compositions comprising such may be administered at dosage levels sufficient to deliver an amount of effector protein of from about 0.0001 mg/kg to about 100 mg/kg, from about 0.01 mg/kg to about 50 mg/kg, from about 0.1 mg/kg to about 40 mg/kg, from about 0.5 mg/kg to about 30 mg/kg, from about 0.01 mg/kg to about 10 mg/kg, from

about 0.1 mg/kg to about 10 mg/kg, or from about 1 mg/kg to about 25 mg/kg, of subject body weight per day, one or more times a day, to obtain the desired therapeutic, diagnostic, prophylactic, or imaging effect. The desired dosage may be delivered three times a day, two times a day, once a day, every other day, every third day, every week, every two weeks, every three weeks, or every four weeks. In some embodiments, the desired dosage may be delivered using multiple administrations (e.g., two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, or more administrations).

**[00205]** The fusion protein or complex comprising the supernegatively charged protein or compositions comprising such may be administered in combination with one or more other therapeutic, prophylactic, diagnostic, or imaging agents. By “in combination with,” it is not intended to imply that the agents must be administered at the same time and/or formulated for delivery together, although these methods of delivery are within the scope of the invention. Compositions can be administered concurrently with, prior to, or subsequent to, one or more other desired therapeutics or medical procedures. In general, each agent will be administered at a dose and/or on a time schedule determined for that agent. In some embodiments, the invention encompasses the delivery of pharmaceutical, prophylactic, diagnostic, or imaging compositions in combination with agents that may improve their bioavailability, reduce and/or modify their metabolism, inhibit their excretion, and/or modify their distribution within the body.

**[00206]** The compositions and methods described herein may be used to treat or prevent any disease that can benefit, *e.g.*, from the delivery of an agent to a cell. The compositions and methods described herein may also be used to transfect or treat cells for research purposes.

**[00207]** In some embodiments, compositions in accordance with the present disclosure may be used for research purposes, *e.g.*, to efficiently deliver effector proteins to cells in a research context. In some embodiments, compositions in accordance with the present invention may be used for therapeutic purposes. In some embodiments, compositions in accordance with the present invention may be used for treatment of any of a variety of diseases, disorders, and/or conditions, including, but not limited to, one or more of the following: autoimmune disorders (*e.g.*, diabetes, lupus, multiple sclerosis, psoriasis, rheumatoid arthritis); inflammatory disorders (*e.g.*, arthritis, pelvic inflammatory disease); infectious diseases (*e.g.*, viral infections (*e.g.*, HIV, HCV, RSV), bacterial infections, fungal infections, sepsis); neurological disorders (*e.g.* Alzheimer’s disease, Huntington’s disease; autism; Duchenne muscular dystrophy); cardiovascular disorders (*e.g.* atherosclerosis, hypercholesterolemia, thrombosis, clotting disorders, angiogenic disorders such as macular degeneration); proliferative disorders (*e.g.*

cancer, benign neoplasms); respiratory disorders (*e.g.* chronic obstructive pulmonary disease); digestive disorders (*e.g.* inflammatory bowel disease, ulcers); musculoskeletal disorders (*e.g.* fibromyalgia, arthritis); endocrine, metabolic, and nutritional disorders (*e.g.* diabetes, osteoporosis); urological disorders (*e.g.* renal disease); psychological disorders (*e.g.* depression, schizophrenia); skin disorders (*e.g.* wounds, eczema); blood and lymphatic disorders (*e.g.* anemia, hemophilia); *etc.*

**[00208]** Compositions of the invention may be used in a clinical setting. For example, a supernegatively charged protein may be associated with an effector protein that can be used for therapeutic applications. Such effector protein may be, for example, nucleases or transcriptional activators. Other compositions comprising a genome editing agent (*e.g.*, TALEN, ZFN, or Cas9 proteins, variants, and fusion proteins thereof) and a cationic lipid may also be used for therapeutic applications.

**[00209]** In some embodiments, the supernegatively charged protein or effector protein associated with a supernegatively charged protein includes a detectable label. These molecules can be used in detection, imaging, disease staging, diagnosis, or patient selection. Suitable labels include fluorescent, chemiluminescent, enzymatic labels, colorimetric, phosphorescent, density-based labels, *e.g.*, labels based on electron density, and in general contrast agents, and/or radioactive labels.

**[00210]** In some embodiments, the effector protein, *e.g.*, a transcription factor, is delivered to a cell in an amount sufficient to activate or inhibit transcription of a target gene of the transcription factor within the cell. In some embodiments, a transcription factor is delivered in an amount and over a time period sufficient to effect a change in the phenotype of a target cell, for example, a change in cellular function, or a change in developmental potential. Exemplary transcription factors are described herein, and the skilled artisan will be able to identify additional suitable transcription factors based on the guidance provided herein and the knowledge of such transcription factors in the art.

**[00211]** In some embodiments, a target cell is contacted repeatedly with an effector protein associated with a supernegatively charged protein and a cationic lipid and/or cationic polymer as provided herein until the formation of a desired cellular effect is detected. Methods for detecting cellular effects and gene expression are well known to those in the art and include, for example, morphological analysis, and detection of marker gene expression by well-established methods such as immunohistochemistry, fluorescence activated cell sorting (FACS), or fluorescent microscopy. In some embodiments, a target cell is contacted with an effector protein associated with a supernegatively charged protein as provided herein for a

period of at least 3 hours, at least 6 hours, at least 12 hours, at least 1 day, at least 2 days, at least 3 days, at least 4 days, at least 5 days, at least 6 days, at least 7 days, at least 10-12 days, at least 12-15 days, at least 15-20 days, at least 20-25 days, at least 25-30 days, at least 30-40 days, at least 40-50 days, at least 50-60 days, at least 60-70, or at least 70-100 days.

**[00212]** A useful concentration of an effector protein associated with a supernegatively charged protein and a cationic lipid and/or cationic polymer for delivery to a specific cell type can be established by those of skill in the art by routine experimentation. In some embodiments a target cell is contacted in vitro or ex vivo with an effector protein associated with a supernegatively charged protein and a cationic lipid and/or cationic polymer at a concentration of about 1 pM to about 1  $\mu$ M. In some embodiments, a target cell is contacted in vitro or ex vivo with the effector protein associated to a supernegatively charged protein at a concentration of about 1 pM, about 2.5 pM, about 5 pM, about 7.5 pM, about 10 pM, about 20 pM, about 25 pM, about 30 pM, about 40 pM, about 50 pM, about 60 pM, about 70 pM, about 75 pM, about 80 pM, about 90 pM, about 100 pM, about 200 pM, about 250 pM, about 300 pM, about 400 pM, about 500 pM, about 600 pM, about 700 pM, about 750 pM, about 800 pM, about 900 pM, about 1 nM, about 2nM, about 3nM, about 4nM, about 5nM, about 6nM, about 7 nM, about 8 nM, about 9nM, about 10nM, about 20nM, about 25 nM, about 30 nM, about 40 nM, about 50 nM, about 60 nm, about 70 nM, about 75 nM, about 80 nM, about 90 nM, about 100 nM, about 200 nM, about 250 nM, about 300 nM, about 400 nM, about 500 nM, about 600 nM, about 700 nM, about 750 nM, about 800 nM, about 900 nM, or about 1  $\mu$ M. A useful time of exposure of the target cell to the effector protein, and, if necessary, incubation after administration in the absence of the effector protein, as well as a number of administration/incubation cycles useful to achieve a desired biological effect (e.g., change in gene transcription, cleavage of a target site by a delivered nuclease, etc.), or a desired cellular phenotype can also be established by those of skill in the art by routine experimentation.

**[00213]** In some embodiments, the target cell for delivery of an effector protein by a system or method provided herein, is a primary cell obtained by a biopsy from a subject. In some embodiments, the subject is diagnosed as having a disease. In some embodiments the disease is a degenerative disease characterized by diminished function of a specific cell type, for example, a neural cell. In some embodiments, a cell treated with an effector protein according to the strategies or methods disclosed herein, or the progeny of such a cell, is used in a cell-replacement therapeutic approach. In some embodiments, the treated cells are administered to the subject from which the somatic cell was obtained in an autologous cell replacement therapeutic approach.

## EXAMPLES

**[00214]** In order that the invention described herein may be more fully understood, the following examples are set forth. The synthetic examples described in this application are offered to illustrate the compounds and methods provided herein and are not to be construed in any way as limiting their scope.

**[00215]** Delivery into mammalian cells remains a significant barrier to many applications of proteins as research tools and therapeutics. The cationic lipid-mediated delivery of cargo proteins fused to a surface-modified supernegatively charged (–30)GFP was recently reported, which results in the encapsulation and delivery of the cargo protein into mammalian cells *in vitro* and *in vivo*<sup>1</sup>. To discover polyanionic proteins with optimal properties for lipid-mediated protein delivery, highly negatively charged proteins from the human proteome were evaluated for their ability to deliver cargo proteins into cultured mammalian cells. It was discovered that a small, widely expressed, intrinsically disordered human protein, ProTα, enables ~10-fold more efficient cationic lipid-mediated protein delivery compared to (–30)GFP. ProTα enables efficient delivery of two unrelated genome editing proteins, Cre recombinase and zinc-finger nucleases, under conditions in which (–30)GFP, lipid alone, or the cargo proteins alone do not result in substantial delivered genome editing activity. High negative charge is necessary, but not sufficient, for ProTα's protein delivery activity, which is blocked by small molecules that impede liposome fusion, but not by drugs that inhibit clathrin-mediated endocytosis. ProTα can enable protein delivery applications, including genome editing, when delivery potency is limiting.

**[00216]** Proteins including genome editing agents represent an increasing proportion of biomedical research tools and new human therapeutics<sup>2</sup>. Due to their inability to spontaneously cross the lipid bilayer, however, current uses of protein tools and therapeutic agents are largely limited to targeting extracellular components<sup>3</sup>. Technologies to facilitate cytosolic access by proteins are critical to expanding the potential targets that can be accessed by exogenous proteins.

**[00217]** Researchers have developed several approaches to facilitate protein translocation across the cytosol. Positively charged cell-penetrating peptides, such as the HIV-1 transactivator of transcription (Tat) peptide<sup>4,5</sup>, poly-arginine<sup>6</sup>, or superpositively charged proteins<sup>3,7,8</sup>, allow a fused cargo protein to reach the cytosol via interaction with the highly anionic proteoglycans on the cell surface, followed by endocytosis<sup>9</sup>. However, escape from endosomes is generally very inefficient, and moreover any unencapsulated escaped proteins are

susceptible to proteolytic degradation and neutralization by serum proteins and the extracellular matrix<sup>10</sup>.

**[00218]** Lipid nanoparticles can encapsulate proteins to prevent degradation and neutralization by antibodies<sup>11</sup>. While cationic lipids are routinely used to transfect polyanionic nucleic acids into mammalian cells, they have not been widely used for protein delivery, as proteins vary in charge, and hence, vary in their ability to be encapsulated by cationic lipids<sup>12</sup>. To impart nucleic-acid-like properties onto proteins, it was previously demonstrated that fusing an anionic protein, such as an engineered (−30)GFP with many surface-exposed negative charges, to a protein of interest enables its efficient encapsulation and delivery into mammalian cells (Figure 1)<sup>1</sup>.

**[00219]** While (−30)GFP enables cationic lipids to deliver a variety of fused cargo proteins<sup>1</sup>, its discovery from protein engineering efforts unrelated to delivery<sup>13</sup> suggests that more potent anionic proteins that mediate lipid-based protein delivery may exist. Moreover, the bacterial origin of (−30)GFP likely will result in immunogenicity, potentially compromising the safety and efficacy of using this protein *in vivo*<sup>14</sup>.

**[00220]** To address both limitations, it was sought to discover an improved protein tag for lipid-mediated protein delivery from the human proteome. A possible outcome was that the human proteome, which contains many proteins with high theoretical net charge, likely includes some highly anionic proteins that would mediate efficient lipid-based delivery of fused cargo proteins into mammalian cells.

**[00221]** A candidate list of human proteins with > 0.75 net theoretical negative charges per kD was generated from the UniProt protein database (Figure 7)<sup>15</sup>. To focus on proteins that are more likely to be easily expressed and purified as fusion partners with a wide range of potential cargo proteins, the list was narrowed down to proteins with published structures, known bacterial expression, and no extensive disulfide bonds or post-translational modifications. Based on these criteria, 12 candidate proteins were identified for further analysis (Table 3).

**Table 3. Protein delivery using Lipofectamine RNAiMAX in HeLa-DsRed cell line.** 12

human proteins were chosen as candidates for lipid nanoparticle delivery. EC<sub>50</sub> is defined as the concentration at which 50 % of the cells contain red fluorescence.

| Protein fused to Cre                                   | Concentration at which 50% of cells are recombined (nM) |
|--------------------------------------------------------|---------------------------------------------------------|
| ProTα                                                  | 1.0 ± 0.2                                               |
| Polyadenylate-binding protein-interacting protein 2    | 290 ± 60                                                |
| Troponin C                                             | 57 ± 19                                                 |
| DPH3 homolog                                           | 36 ± 5                                                  |
| RNA Polymerase II Subunit F                            | 110 ± 58                                                |
| Multiple coagulation factor deficiency protein 2       | 140 ± 170                                               |
| ADP-ribosylation factor-like protein 2-binding protein | 20 ± 9                                                  |
| NF-kappa-B inhibitor alpha                             | 63 ± 7                                                  |
| DNA damage-inducible transcript 3 protein              | 89 ± 5                                                  |
| Carbonic anhydrase VIII                                | 84 ± 22                                                 |
| Protein S100-B                                         | ~ 47                                                    |
| Sirtuin-1                                              | 49 ± 18                                                 |
| (-30)GFP                                               | 10.1 ± 2.6                                              |
| No fusion                                              | 73 ± 12                                                 |

[00222] The genes encoding the candidate human proteins were cloned as an N-terminal fusion to Cre recombinase and overexpressed in *E. coli*. Each purified fusion protein was encapsulated with Lipofectamine RNAiMAX, a commercially available lipid formulation previously used to mediate robust protein delivery of anionic proteins and protein complexes<sup>1</sup>. HeLa cells containing a genomically integrated dsRed gene preceded by a floxed transcriptional terminator<sup>7</sup> were used to screen the fusion proteins for Cre delivery activity. Functional Cre delivery catalyzes recombination to remove the terminator, resulting in dsRed expression and red fluorescence. Delivery potency was defined as the protein concentration at which 50% of analyzed cells contain red fluorescence (EC<sub>50</sub>). While unfused Cre with RNAiMAX lipid induced red fluorescence in the assay at an EC<sub>50</sub> of 73 ± 12 nM, fusion of (-30)GFP to Cre lowered the EC<sub>50</sub> with RNAiMAX to 10.1 ± 2.6 nM (Table 3)<sup>1</sup>.

[00223] Seven of the 12 anionic human proteins tested from the set of 12 candidates did not substantially improve the efficacy of protein delivery when fused to the N-terminus of Cre recombinase compared to that of Cre alone (73 nM EC<sub>50</sub>) (Table 3). Their inability to do so may be due to an inability to support lipid encapsulation, or to interference with recombinase activity when fused to Cre. Four human proteins showed compatibility with lipid complexation and enhanced Cre delivery compared to that of unfused Cre, but their delivery potency did not exceed that of (-30)GFP (10 nM EC<sub>50</sub>) (Table 3). One protein, human prothymosin alpha

(ProT $\alpha$ ), a small, intrinsically disordered protein<sup>16</sup> with a net theoretical charge of  $-44$ , greatly improved the potency of Cre protein delivery 10-fold compared to that of  $(-30)$ GFP, to an EC<sub>50</sub> of  $1.0 \pm 0.2$  nM (Table 3 and Figure 2).

**[00224]** To verify the highly potent lipid encapsulation and delivery of ProT $\alpha$ , ProT $\alpha$ -Cre was delivered using lipid nanoparticles to two additional human cell lines containing genomically integrated RFP reporters based on HEK293 and BSR<sup>3</sup> cells. Consistent with the delivery improvement observed in HeLa cells, ProT $\alpha$ -Cre also resulted in a large (approximately 10-fold) improvement in delivery potency compared to  $(-30)$ GFP-Cre in both cell lines (Figure 2 and Figures 8A-8F). Cre-mediated fluorescent activation was visually confirmed via microscopy to be consistent with these quantified changes in delivery potency (Figure 9). There was not any significant cellular toxicity observed during ProT $\alpha$ -mediated delivery (Figure 10).

**[00225]** ProT $\alpha$  contains a stretch of aspartate and glutamate residues that provide a concentrated region of anionic charge<sup>16</sup>. To investigate if a simple stretch of acidic residues of similar negative charge magnitude is sufficient to confer compatibility with cationic liposome delivery, two constructs were generated that have the same theoretical net charge with those of  $(-30)$ GFP and ProT $\alpha$ ,  $(-30)$  and  $-44$ , respectively) but with sequences containing scrambled Glu and Asp residues:  $(-30)$ polyD/E and  $(-44)$ polyD/E (Figure 3A). Importantly, Cre fusions to  $(-30)$ polyD/E or  $(-44)$ polyD/E resulted in much poorer protein delivery potencies (EC<sub>50</sub> =  $62 \pm 7.6$  nM and  $21 \pm 3.2$  nM, respectively) than fusions with  $(-30)$ GFP or ProT $\alpha$  (Figure 3B). These results show that additional features beyond net theoretical charge within ProT $\alpha$  strongly contribute to its ability to enable highly potent lipid-mediated protein delivery.

**[00226]** ProT $\alpha$  is universally expressed in human tissues and contains multiple domains that are hypothesized to be modular in function<sup>16</sup>. Researchers have associated multiple biological functions with ProT $\alpha$ , including histone chaperone activity implicated in nucleosome exchange<sup>17</sup>, extracellular immune modulation<sup>16</sup>, and cell proliferation<sup>18</sup>. To identify the domains within ProT $\alpha$  responsible for cationic lipid-mediated protein delivery, a series of truncated ProT $\alpha$  variants were generated and the Cre fusion protein delivery titrations were repeated in the presence of RNAiMAX lipid (Figure 4A).

**[00227]** Deletion of the central anionic region almost completely abolished delivery by ProT $\alpha$  (EC<sub>50</sub> =  $36 \pm 4.7$  nM), consistent with the hypothesis that negative charges are critical for complexation with cationic lipids (Figure 4B). Deletion of the N-terminal region, hypothesized to have receptor-binding properties<sup>16</sup>, also slightly increased the EC<sub>50</sub> to  $2.9 \pm 0.4$  nM, suggesting that ProT $\alpha$  might mediate liposome-cell membrane interaction via surface-exposed

receptors, in addition to simple electrostatic charge attraction (Figure 4B). The deletion of its C-terminal region resulted in a 7.7-fold impairment of delivery potency ( $EC_{50} = 7.7 \pm 0.8$  nM, Figure 4B). The C-terminus of ProT $\alpha$  contains a putative NLS<sup>16</sup>. To test the possibility that the NLS is responsible for delivery enhancements during ProT $\alpha$ -mediated delivery, Cre containing an additional N-terminal SV40 NLS was cloned, expressed, and delivered. Cre delivery did not benefit from an additional NLS, consistent with the known ability of Cre to localize to the nucleus spontaneously, and suggesting that the C-terminus domain assists ProT $\alpha$  protein delivery potency through mechanisms beyond facilitating nuclear localization (Figure 11). Taken together, these results indicate that multiple domains within ProT $\alpha$  contribute to its unusual cationic lipid-mediated protein delivery potency.

**[00228]** Next, potential mechanisms of ProT $\alpha$ -mediated protein delivery were probed compared to those of canonical cationic liposomes, which are thought to enter cells through clathrin-dependent endocytosis<sup>19,20</sup>. Several inhibitors of various pathways involved in endocytosis and micropinocytosis were applied to observe possible effects on delivery, including sodium azide, which depletes cellular ATP<sup>21</sup>; ethylisopropyl amiloride (EIPA), a micropinocytosis inhibitor<sup>22</sup>; Dynasore (Dyna), a dynamin inhibitor<sup>23</sup>; methyl-B-cyclodextrin (MBCD), a cholesterol-depleting molecule that prevents lipid raft formation<sup>24</sup>; chloropramazine (CPZ), a clathrin-dependent endocytosis inhibitor<sup>25</sup>, and wortmannin (WTM), a PI3-kinase inhibitor that prevents clathrin-dependent endocytosis<sup>26</sup>. After pre-incubating various inhibitors with HEK293 cells containing the RFP Cre reporter, either (–30)GFP-Cre or ProT $\alpha$ -Cre was combined with RNAiMAX lipid and cells with the resulting complex were treated for 4 hours. Excess liposomes and inhibitors were washed off through several washes with buffer containing heparin, and the cells were further incubated for 2 days.

**[00229]** Flow cytometry analysis showed that MBCD almost completely negated cytosolic access, while CPZ and WTM caused mild inhibition of delivery (Figure 5). These data are consistent with a previous study<sup>27</sup> that examined the mechanism of siRNA transfection using cationic lipids and implicated direct liposome fusion, which is acutely inhibited by MBCD, rather than clathrin-mediated endocytosis followed by endosomal escape, as the major source of functional cargo in the cytoplasm. Moreover, (–30)GFP-Cre and ProT $\alpha$ -Cre were inhibited by the same molecules to nearly identical degrees, suggesting that the large delivery potency increase enabled by ProT $\alpha$  is predominantly due to its apparent ability to engage cationic lipids at ~10-fold lower concentrations than (–30)GFP (Figure 5). Minimal cell toxicity was observed at the doses chosen for study (Figure 12).

[00230] Next, ProTα was applied as an unusually potent mediator of protein delivery to deliver zinc-finger nucleases (ZFNs), chimeric genome editing proteins composed of a modular DNA-binding zinc-finger domains and a heterodimeric FokI nuclease<sup>28</sup>. Two ZFNs (left and right) when targeted to adjacent half-sites of a genomic locus will bind together, enabling their fused FokI nuclease domains to dimerize and induce a double stranded DNA cut, initiating end-joining processes that result in indels at the target locus<sup>28</sup>. ZFNs are promising research tools and therapeutics, and are in multiple clinical trials for the treatment of diseases including HIV<sup>29</sup>.

[00231] ZFNs have been shown to enter cells spontaneously at high concentrations under some conditions<sup>30</sup>. However, self-delivery of ZFNs requires serum-free media and μM protein concentrations, which are not relevant for some cell culture and for most *in vivo* applications. It was sought to test the ability of ProTα to mediate potent delivery of ZFNs with cationic lipids. Both left and right ZFNs were expressed targeting the AAVS1 safe harbor site in the human genome fused with ProTα at the N-terminus (Figure 6A). Before conducting protein delivery, it was confirmed that ProTα did not affect the activity of ZFNs in DNA cleavage assays *in vitro* conducted using purified substrates (Figure 13A) and in plasmid transfections of ZFN variants (Figure 13B). The concentration of protein and lipid was also optimized for ZFN delivery into HEK293T cells (Figure 14).

[00232] Finally, ProTα–ZFN fusions or unmodified ZFNs complexed with RNAiMAX lipid were delivered into HEK293T cells in the presence of 10% serum. After 2 days, high-throughput sequencing (HTS) showed robust ZFN-mediated indel formation only in cells treated with both pairs of ProTα–ZFN fusions and the lipid in the mid-nM concentration regime (Figure 6B). In contrast, cells treated with unmodified ZFNs complexed with RNAiMAX resulted in background level indels (Figure 6B). Neither ZFNs only nor the ProTα–ZFN fusions resulted in substantial cytotoxicity at the concentrations tested when complexed with RNAiMAX (Figure 15). Taken together, the findings show that ProTα enables potent delivery of both Cre recombinase and ZFNs at nM concentrations into human cells when combined with a simple commercially available cationic lipid.

[00233] It is believed that ProTα represents the most potent reported protein that enables potent delivery of fused proteins via cationic liposomes. As ProTα expression is known in all human tissues tested<sup>16</sup>, it may serve as a less immunogenic domain for protein delivery than other non-human alternatives, such as (–30)GFP. Based on previous studies on the use of anionic proteins to mediate cationic lipid-based protein delivery<sup>1,31</sup>, it is anticipated that ProTα will be compatible with a variety of commercial and non-commercial cationic lipid reagents. It

is also envisioned that the ProTα may be particularly enabling when delivering proteins with adverse properties that preclude naked protein delivery via conventional cell-penetrating peptides, or that are not tolerated by the cell at higher concentrations.

## **Methods**

### **Cloning**

[00234] PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (New England BioLabs). Candidate human protein DNAs were purchased from IDT as gBlock Gene Fragments. Bacterial expression plasmids encoding human protein fused to Cre were made using USER-cloning (New England BioLabs). Truncation of ProTα was done using blunt-end ligation to delete regions of ProTα. Following PCR, KLD enzyme mix (New England BioLabs) was used to phosphorylate and circularize the PCR product before transformation into NEB10beta cells (New England BioLabs).

### **Protein Expression**

[00235] BL21 Star (DE3) chemically competent *E. coli* cells (ThermoFisher Scientific) were transformed with plasmids encoding the human proteins fused with Cre with a His<sub>6</sub> C-terminal purification tag. A single colony was grown overnight in 2xYT broth containing 50 µg/ml Carbenicillin at 37 °C. The cells were diluted 1:20 into 1 L of the same media and grown until OD<sub>600</sub> ~ 0.5. The cultures were incubated on ice for 60 min and protein expression was induced with 0.5 mM isopropyl-b-D-1-thiogalactopyranoside (IPTG, GoldBio Sciences). Protein was expressed for 14–16 h with shaking at 16 °C. Cells were centrifuged at 10,000 rpm for 20 min, and then resuspended in a high salt buffer (100 mM tris(hydroxymethyl)-aminomethane (Tris)-HCl, pH 8.0, 1 M NaCl, 20% glycerol, 5 mM tris(2-carboxyethyl)phosphine (TCEP; GoldBio) with a protease inhibitor pellet (Roche). The cells were lysed using sonication and the supernatant was incubated with His-Pur nickel nitriloacetic acid (nickel-NTA) resin (ThermoFisher) with rotation at 4 °C for 30 min. The resin was washed with the high salt buffer before the protein was eluted with an elution buffer (high salt buffer supplemented with 200 mM imidazole). The eluent was purified on a 5 ml Hi-Trap Q (GE Healthcare) anion exchange column with an FPLC (AKTA Pure). The purified protein was quantified by a Pierce microplate BCA protein assay kit (Pierce Biotechnology) and snap-frozen in liquid nitrogen and stored at -80 °C until before use. ZFNs were purified according to previous literature<sup>1</sup>.

***In vitro* DNA cleavage assay**

[00236] DNA containing the AAVS1 locus was amplified from purified HEK293T genomic DNA using PCR. ~350 bp PCR product was purified using Minelute columns (Qiagen). 100 ng of DNA substrate was incubated with 300 nM of ZFN or ProTα-ZFN pairs in Cutsmart buffer with 1 mM Arginine and 100 μM ZnCl<sub>2</sub> (New England Biolabs) at room temperature for 16 hours. The cleavage product was detected by running the mixture on an agarose gel without further purification.

**Cell culture**

[00237] HeLa-DsRed, BSR-TdTomato, HEK293-loxP-GFP-RFP (HEK293-RFP) (GenTarget), and HEK293T cells were cultured in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher Scientific) supplemented with 10% (v/v) FBS, at 37 °C with 5% CO<sub>2</sub>.

**Protein delivery assays**

[00238] Protein was diluted to 12.5 μL in OptiMem and was complexed with 1.5 μL of Lipofectamine RNAiMAX (ThermoFisher Scientific) in another 12.5 μL of OptiMem. The resulting complex was delivered to cells that had been seeded on a 48-well collagen-coated BioCoat plate (Corning) at ~70% confluency (250 μL final volume). After 3 days, the cells were trypsinized using TrypLE reagent (ThermoFisher Scientific), and resuspended in culture media before being analyzed on the CytoFlex flow cytometer (Beckman Coulter).

[00239] For ZFN experiments, equimolar amounts of 'left' and 'right' ZFNs was diluted to 12.5 μL in OptiMem and was complexed with 12.5 μL of OptiMem containing 3.5 μL of Lipofectamine RNAiMAX (ThermoFisher Scientific). The resulting complex was delivered to cells that had been seeded on a 48-well collagen-coated BioCoat plate (Corning) at ~70% confluency at the final volume of 100 μL per well. After 4 hours, the cells were incubated with fresh media for further 48 hours. Then, the cells were lysed and DNA was purified DNA was isolated using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) according to the manufacturer's instructions. AAVS1 site was amplified by PCR with flanking high-throughput sequencing primer pairs. Then, DNA was further amplified by PCR with primers containing Illumina sequencing adaptors. The products were gel-purified and quantified using KAPA Library Quantification Kit-Illumina (KAPA Biosystems). Samples were sequenced on an Illumina MiSeq as previously described. Indel was quantified within the 30-base window surrounding the cleavage site among the high-quality reads (Q > 30) using a custom Matlab script<sup>2</sup>.

### **Inhibitor Assays**

**[00240]** HEK293-RFP cells at ~70% confluency were incubated with one of the following endocytosis inhibitors for 1 hr: NaN<sub>3</sub> (Sigma), 5-(N-ethyl-N-isopropyl)amiloride (EIPA, Santa Cruz Biotechnology), Dynasore (Abcam), Chlorpromazine (CPZ, Sigma), Methyl- $\beta$ -cyclodextrin (MBCD, Sigma), Wortmannin (Sigma). Then, liposomes containing either (-30)GFP-Cre or ProT $\alpha$ -Cre were delivered and incubated at 37 °C with 5% CO<sub>2</sub> for 4 hours. The cells were washed 3 times with PBS containing heparin (Stem Cell Technologies) at 20  $\mu$ g/ml. The cells were further recovered for 2 days, and the cells were analyzed by the CytoFlex flow cytometer (Beckman Coulter).

### **ZFN Transfections**

**[00241]** HEK293T cells were plated on a 48-well collagen-coated BioCoat plate (Corning) 1 day prior to experiment. At ~70% confluency, 500 ng of 'left' and 'right' CMV-ZFNs and CMV-ProT $\alpha$ -ZFNs (1  $\mu$ g total DNA content) were transfected using 1.5  $\mu$ L Lipofectamine 2000 (ThermoFisher Scientific) according to the manufacturer's protocol.

### **Sequences**

#### **[00242] ProT $\alpha$ -Cre:**

MGASDAAVDTSSEITTKDLKEKKEVVVEEAENGRDAPANGNAENEENGEQEADNEVD  
 EEEEEGGEEEEEEEEEGDGEEDGDEDEEAESATGKRAAEDDEDDVDTKKQKTDDEDD  
 TGGSGGSGGSGGSGGSGGSGGSGGSGGTASNLLTVHQNLPALPVDATSDEVKRNLM  
 MFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFPAEPEDVRDYLLYLQARGLA  
 VKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERTD  
 FDQVRSLMENS DRCQDIRNLAFLGIAYNTLLRIAELIRVKDISRTDGGRMLIHIGRTK  
 TLVSTAGVEKALSLGVTKLVERWISVSGVADDPNNYLCRVRKNGVAAPSATSQSLST  
 RALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGG  
 WTNVNIVMNYIRNLDSETGAMVRLLEDGDGGSHHHHHH (SEQ ID NO: 80)

#### **[00243] (-30)GFP-Cre:**

MGASKGEELFDGVVPILVELDGDVNGHEFSVRGEGEGDATEGELTLKFICTTGELPVP  
 WPTLVTTLTLYGVQCFS DYPDHMDQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVK  
 FEGDTLVNRIELKGIDFKEDGNILGHKLEYNFN SHDVYITADKQENG IKA EFEIRHNVE  
 DGSVQLADHYQQNTPIGDGPVLLPDDHYLSTESALSKDPNEDRDHMLLEFVTAAGI

DHGMDELYKTGGSGGSGGSGGSGGSGGSGGSGGSGGSGGTASNLLTVHQNLPALPVDAT  
SDEVKRNLMDFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFPAEPEDVRDY  
LLYLQARGLA VKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERA  
KQALAFERTDFDQVRSLMENS DRCQDIRNLAFLGIA YNTLLRIA EIRIRVKDISRTDG  
GRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERWISVSGVADDPNNYLFCRVRKNGV  
AAPSATSQ LSTRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMARAGV  
SIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGDGGSHHHHHH (SEQ ID NO:  
81)

### Charge variants

#### [00244] (-30)PolyD/E-Cre:

MGADDEESDEEELDEEEDELEEDTDEEGGDEELEDDELDETGGSGGSGGSGGSGGSGG  
SGGSGGSGGSGGTASNLLTVHQNLPALPVDATSDDEVKRNLMDFRDRQAFSEHTWKML  
LSVCRSWAAWCKLNNRKWFPAEPEDVRDYLLYLQARGLA VKTIQQHLGQLNMLHR  
RSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENS DRCQD  
IRNLAFLGIA YNTLLRIA EIRIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGV  
TKLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQ LSTRALEGIFEATHRLIYG  
KDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSE  
TGAMVRLLEDGDGGSHHHHHH (SEQ ID NO: 82)

#### [00245] (-44)PolyD/E-Cre:

MGADDEESDEEELDEEEDELEEDTDEEGGDEELEDDEDEDEETDDEESDEDEDEET  
GGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGTASNLLTVHQNLPALPVDATSDDEVKRNLM  
DFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFPAEPEDVRDYLLYLQARGLA  
VKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERTD  
FDQVRSLMENS DRCQDIRNLAFLGIA YNTLLRIA EIRIRVKDISRTDGGRMLIHIGRTK  
TLVSTAGVEKALSLGVTKLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQ LST  
RALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGG  
WTNVNIVMNYIRNLDSETGAMVRLLEDGDGGSHHHHHH (SEQ ID NO: 83)

**Truncated ProTα sequences****[00246] in B2:**

MGASDAAVDTSSEITTKDLKEKKEVVVEEAENGRDAPANGNAENEENGQEADNEVD  
 EEEEEGGEEEEEEEEEGDGEEEDGDEDEEAESATGKRAAEDDEDDVDVT (SEQ ID NO:  
 84)

**[00247] in B3:**

MGAESATGKRAAEDDEDDVDTKKQKTDEDD (SEQ ID NO: 85)

**[00248] in B4:**

MGAENEENGQEADNEVDDEEEEEGGEEEEEEEEEGDGEEEDGDEDEEAESATGKRAAE  
 DDEDDVDTKKQKTDEDD (SEQ ID NO: 86)

**[00249] in B5:**

MGASDAAVDTSSEITTKDLKEKKEVVVEEAENGRDAPANGNAESATGKRAAEDDEDD  
 DVDTKKQKTDEDD (SEQ ID NO: 87)

**EQUIVALENTS AND SCOPE**

**[00250]** In the claims articles such as “a,” “an,” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include “or” between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

**[00251]** Furthermore, the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, and descriptive terms from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Where elements are presented as lists, *e.g.*, in Markush group format, each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should be understood that, in general, where

the invention, or aspects of the invention, is/are referred to as comprising particular elements and/or features, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements and/or features. For purposes of simplicity, those embodiments have not been specifically set forth in haec verba herein. It is also noted that the terms “comprising” and “containing” are intended to be open and permits the inclusion of additional elements or steps. Where ranges are given, endpoints are included. Furthermore, unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub-range within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

**[00252]** This application refers to various issued patents, published patent applications, journal articles, and other publications, all of which are incorporated herein by reference. If there is a conflict between any of the incorporated references and the instant specification, the specification shall control. In addition, any particular embodiment of the present invention that falls within the prior art may be explicitly excluded from any one or more of the claims. Because such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular embodiment of the invention can be excluded from any claim, for any reason, whether or not related to the existence of prior art.

**[00253]** Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation many equivalents to the specific embodiments described herein. The scope of the present embodiments described herein is not intended to be limited to the above Description, but rather is as set forth in the appended claims. Those of ordinary skill in the art will appreciate that various changes and modifications to this description may be made without departing from the spirit or scope of the present invention, as defined in the following claims.

#### REFERENCES

1. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73–80 (2015).
2. Pisal, D. S., Kosloski, M. P. & Balu-Iyer, S. V. Delivery of Therapeutic Proteins. *J. Pharm. Sci.* **99**, 2557–2575 (2010).
3. Cronican, J. J. *et al.* Potent Delivery of Functional Proteins into Mammalian Cells *in Vitro* and *in Vivo* Using a Supercharged Protein. *ACS Chem. Biol.* **5**, 747–752 (2010).

4. Alan D. Frankel & Carl O. Pabo. Cellular Uptake of the Tat Protein from Human Immunodeficiency Virus.
5. Fawell, S. *et al.* Tat-mediated delivery of heterologous proteins into cells. *Proc. Natl. Acad. Sci.* **91**, 664–668 (1994).
6. Fuchs, S. M. & Raines, R. T. Polyarginine as a multifunctional fusion tag. *Protein Sci.* **14**, 1538–1544 (2009).
7. Cronican, J. J. *et al.* A Class of Human Proteins that Deliver Functional Proteins into Mammalian Cells In Vitro and In Vivo. *Chem. Biol.* **18**, 833–838 (2011).
8. Thompson, D. B., Villaseñor, R., Dorr, B. M., Zerial, M. & Liu, D. R. Cellular Uptake Mechanisms and Endosomal Trafficking of Supercharged Proteins. *Chem. Biol.* **19**, 831–843 (2012).
9. Heitz, F., Morris, M. C. & Divita, G. Twenty years of cell-penetrating peptides: from molecular mechanisms to therapeutics: Peptide-based drug delivery technology. *Br. J. Pharmacol.* **157**, 195–206 (2009).
10. Allen, T. M. & Cullis, P. R. Liposomal drug delivery systems: From concept to clinical applications. *Adv. Drug Deliv. Rev.* **65**, 36–48 (2013).
11. Judge, A. D., Bola, G., Lee, A. C. H. & MacLachlan, I. Design of Noninflammatory Synthetic siRNA Mediating Potent Gene Silencing in Vivo. *Mol. Ther.* **13**, 494–505 (2006).
12. Colletier, J.-P., Chaize, B., Winterhalter, M. & Fournier, D. Protein encapsulation in liposomes: efficiency depends on interactions between protein and phospholipid bilayer. *BMC Biotechnol.* **8** (2002).
13. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging Proteins Can Impart Unusual Resilience. *J. Am. Chem. Soc.* **129**, 10110–10112 (2007).
14. Ansari, A. M. *et al.* Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments. *Stem Cell Rev. Rep.* **12**, 553–559 (2016).
15. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
16. Piñeiro, A., Cordero, O. J. & Nogueira, M. Fifteen years of prothymosin alpha: contradictory past and new horizons. *Peptides* **21**, 1433–1446 (2000).
17. Borgia, A. *et al.* Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).
18. Magdalena, C., Dominguez, F., Loidi, L. & Puente, J. L. Tumour prothymosin alpha content, a potential prognostic marker for primary breast cancer. *Br. J. Cancer* **82**, 584–590 (2000).

19. J. H. Felgner et al. Enhanced Gene Delivery and Mechanism Studies with a Novel Series of Cationic Lipid Formulations. (1994).
20. Godbey, W. T., Wu, K. K. & Mikos, A. G. Tracking the intracellular path of poly(ethylenimine)/DNA complexes for gene delivery. *Proc. Natl. Acad. Sci.* **96**, 5177–5181 (1999).
21. Hoffmann, J. & Mendgen, K. Endocytosis and Membrane Turnover in the Germ Tube of *Uromyces fabae*. *Fungal Genet. Biol.* **24**, 77–85 (1998).
22. Koivusalo, M. et al. Amiloride inhibits macropinocytosis by lowering submembranous pH and preventing Rac1 and Cdc42 signaling. *J. Cell Biol.* **188**, 547–563 (2010).
23. Preta, G., Cronin, J. G. & Sheldon, I. M. Dynasore - not just a dynamin inhibitor. *Cell Commun. Signal.* **13**, (2015).
24. Rodal, S. K., Skretting, G., Garred, Ø. & Vilhardt, F. Extraction of Cholesterol with Methyl- $\alpha$ -Cyclodextrin Perturbs Formation of Clathrin-coated Endocytic Vesicles. *Mol. Biol. Cell* **10**, 14 (1999).
25. Vercauteren, D. et al. The Use of Inhibitors to Study Endocytic Pathways of Gene Carriers: Optimization and Pitfalls. *Mol. Ther.* **18**, 561–569 (2010).
26. Posor, Y. et al. Spatiotemporal control of endocytosis by phosphatidylinositol-3,4-bisphosphate. *Nature* **499**, 233–237 (2013).
27. Lu, J. J., Langer, R. & Chen, J. A Novel Mechanism Is Involved in Cationic Lipid-Mediated Functional siRNA Delivery. *Mol. Pharm.* **6**, 763–771 (2009).
28. Hockemeyer, D. et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851–857 (2009).
29. Tebas, P. et al. Gene Editing of *CCR5* in Autologous CD4 T Cells of Persons Infected with HIV. *N. Engl. J. Med.* **370**, 901–910 (2014).
30. Liu, J., Gaj, T., Wallen, M. C. & Barbas, C. F. Improved Cell-Penetrating Zinc-Finger Nuclease Proteins for Precision Genome Engineering. *Mol. Ther. - Nucleic Acids* **4**, e232 (2015).
31. Wang, M. et al. Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proc. Natl. Acad. Sci.* **113**, 2868–2873 (2016).

## CLAIMS

What is claimed is:

1. A composition comprising a cationic lipid or a cationic polymer and a supernegatively charged protein associated with an effector protein, wherein the supernegatively charged protein is selected from the group consisting of prothymosin alpha (ProT $\alpha$ ), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and variants thereof.
2. The composition of claim 1, wherein the supernegatively charged protein is a human protein.
3. The composition of claim 2, wherein the supernegatively charged protein is human prothymosin alpha (ProT $\alpha$ ) or a variant thereof.
4. The composition of claim 3, wherein the supernegatively charged protein comprises the amino acid sequence of any one of SEQ ID NOs: 1-3.
5. The composition of claim 2, wherein the supernegatively charged protein is human DPH3 homolog or a variant thereof.
6. The composition of 1, wherein the supernegatively charged protein comprises the amino acid sequence of SEQ ID NO: 4.
7. The composition of claim 2, wherein the supernegatively charged protein is human ADP ribosylation factor-like protein 2-binding protein or a variant thereof.
8. The composition of 7, wherein the supernegatively charged protein comprises the amino acid sequence of SEQ ID NO: 5.
9. The composition of claim 2, wherein the supernegatively charged protein is human protein S100-B or a variant thereof.

10. The composition of 9, wherein the supernegatively charged protein comprises the amino acid sequence of SEQ ID NO: 6.
11. The composition of claim 2, wherein the supernegatively charged protein is human Sirtuin-1 or a variant thereof.
12. The composition of 11, wherein the supernegatively charged protein comprises the amino acid sequence of SEQ ID NO: 7.
13. The composition of any one of claims 1-12, wherein the effector protein is an enzyme, a transcriptional regulator, a therapeutic protein, or a diagnostic protein.
14. The composition of any one of claims 1-13, wherein the enzyme is a recombinase, a nuclease, or an epigenetic modifier.
15. The composition of claim 13, wherein the effector protein is a recombinase.
16. The composition of claim 15, wherein the recombinase is a Cre recombinase, a Gin recombinase, or a Tn3 recombinase.
17. The composition of claim 13, wherein the effector protein is a nuclease.
18. The composition of claim 15, wherein the nuclease is a zinc-finger nuclease (ZFN), a transcription activator-like effector nucleases (TALEN), or a RNA-guided nuclease.
19. The composition of claim 18, wherein the RNA-guided nuclease is a Cas9 nuclease, a variant, or a fusion protein thereof.
20. The composition of claim 13, wherein the effector protein is a transcriptional regulator.
21. The composition of claim 20, wherein the transcriptional regulator is a transcriptional activator or a transcriptional repressor.

22. The composition of claim 21, wherein the transcriptional activator is selected from the group consisting of VP16, VP64, and p65.

23. The composition of claim 21, wherein the transcriptional repressor is a KRAB protein or a SID protein.

24. The composition of claim 13, wherein the effector protein is a therapeutic protein or a diagnostic protein.

25. The composition of any one of claims 1-24, wherein the supernegatively charged protein is covalently associated with the effector protein.

26. The composition of 1-25, wherein the supernegatively charged protein is associated with the effector protein via a linker.

27. The composition of claim 26, wherein the linker is a cleavable linker.

28. The composition of claim 27, wherein the linker is a UV-cleavable linker or a linker that is cleaved by a lysosomal enzyme.

29. The composition of any one of claims 1-28, wherein the supernegatively charged protein is non-covalently associated with the effector protein.

30. The compositions of any one of claims 1-29, wherein the composition is a pharmaceutical composition.

31. The composition of claim 30, wherein the pharmaceutical composition comprises a pharmaceutically acceptable excipient.

32. The composition any one of claims 1-31, wherein the composition is formulated for administration to a subject to deliver the effector protein to the subject.

33. The composition of any one of claim 32, wherein the composition is administered to the subject for diagnosing or treating a disease.

34. The composition of any one of claims 1-33, wherein the cationic lipid is selected from Lipofectamine® 2000, Lipofectamine® 3000, Lipofectamine® RNAiMAX, Lipofectamine® LTX, Lipofectamine® CRISPRMAX, Lipofectamine® MessengerMax, and FuGENE HD.

35. A fusion protein comprising a supernegatively charged protein fused to an effector protein, wherein the supernegatively charged protein is selected from the group consisting of prothymosin alpha (ProT $\alpha$ ), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and functional variants thereof.

36. A nucleic acid encoding the fusion protein of claim 35.

37. A vector comprising the nucleic acid of claim 36.

38. A cell comprising the fusion protein of claim 35, the nucleic acid of claim 36, or the vector of claim 37.

39. A composition comprising the fusion protein of claim 35 and a cationic lipid or cationic polymer.

40. A complex comprising a supernegatively charged protein associated with an effector protein, wherein the supernegatively charged protein is selected from the group consisting of prothymosin alpha (ProT $\alpha$ ), DPH3 homolog, ADP ribosylation factor-like protein 2-binding protein, protein S100-B, Sirtuin-1, and variants thereof.

41. A method comprising contacting a cell with the composition of any one of claims 1-31 and 39, whereby the contacting results in the delivery of the effector protein into the cell.

42. The method of claim 41 further comprising detecting the effector protein in the cell.

43. The method of claim 41 or claim 42, wherein the contacting is *in vitro*.

44. The method of claim 41 or claim 42, wherein the contacting is *in vivo*.

45. The method of claim 41 or claim 42, wherein the contacting is *ex vivo*.

46. The method of any one of claims 41-45, wherein the cell is a mammalian cell.

47. The method of claim 46, wherein the mammalian cell is a human cell.

48. The method of any one of claims 41-47, wherein the composition exhibits low toxicity when contacted with a population of cells.

49. The composition of claim 48, wherein at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the cells are viable following contacting the composition.

50. A method of treating a disease comprising administering to a subject in need thereof an effective amount of the composition of any one of claims 1-34 and 39, wherein the effector protein is a therapeutic protein.

51. A method of diagnosing a disease comprising administering to a subject in need thereof an effective amount of the composition of any one of claims 1-34 and 39 or the fusion protein of claim 35, wherein the effector protein is a diagnostic protein.

52. The method of claim 50 or claim 51, wherein the step of administering comprises a route of administration selected from the group consisting of oral, intravenous, intramuscular, intra-arterial, subcutaneous, intraventricular, topical, inhalational, and mucosal delivery.

53. The method of any one of claims 50-52, wherein the effector protein enters at least one cell in the subject.

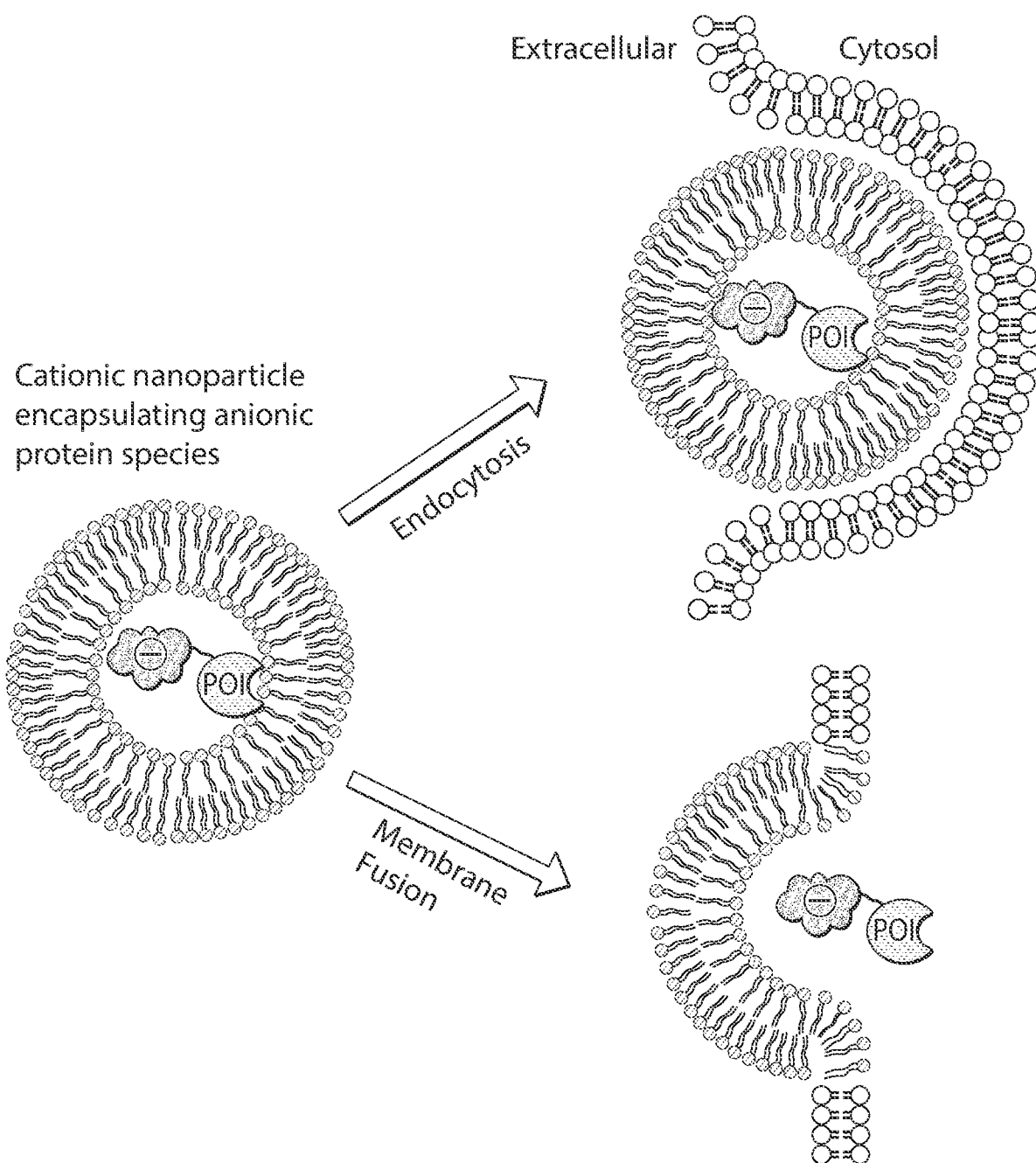


Figure 1

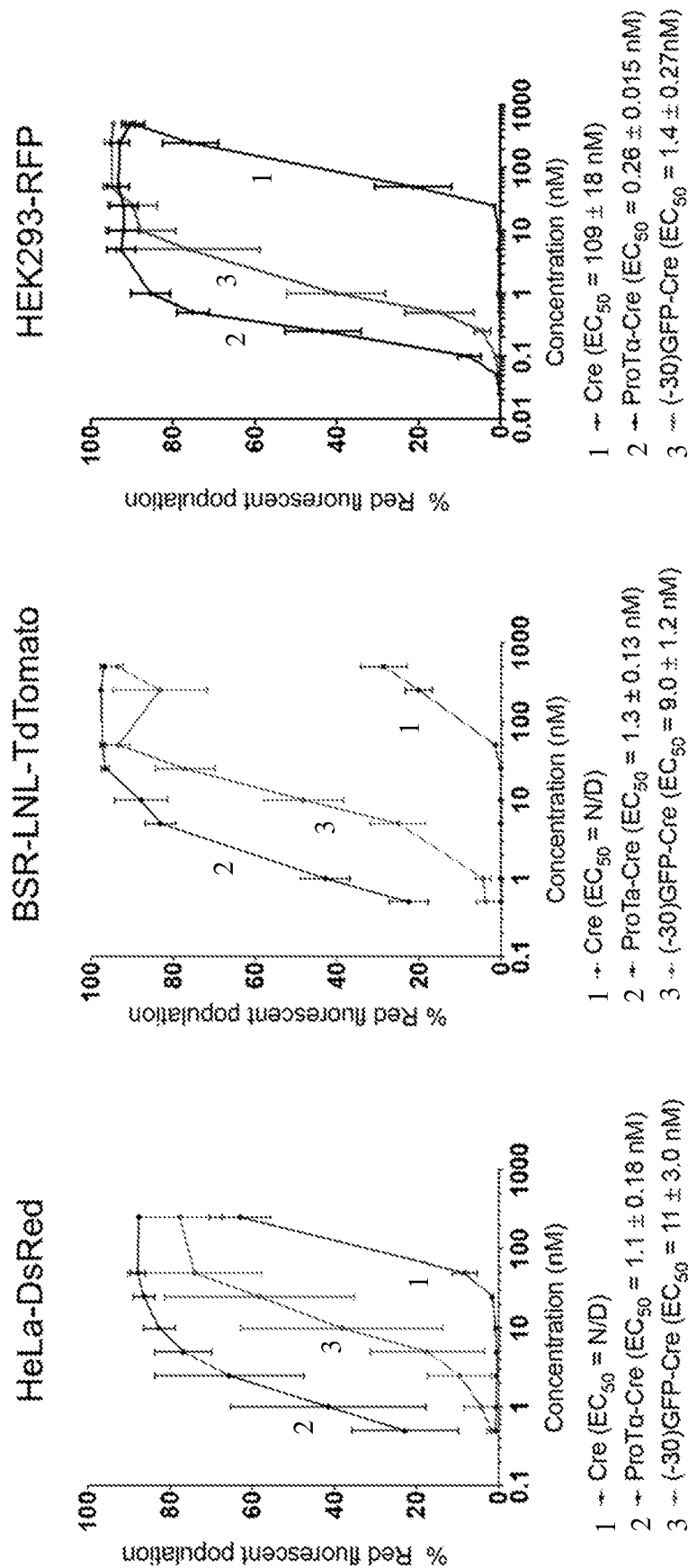
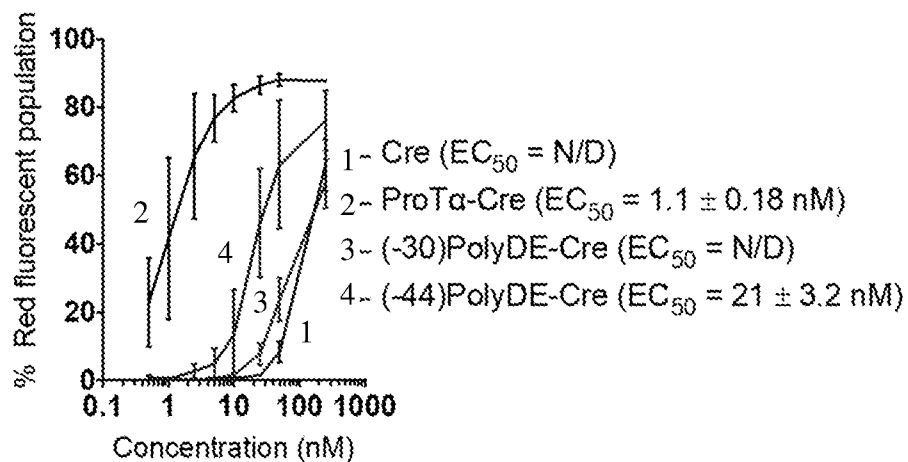
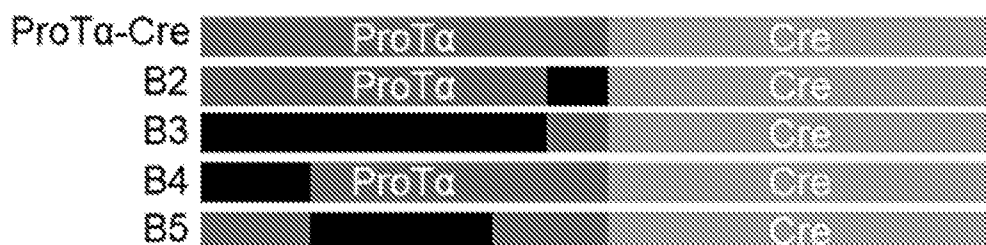
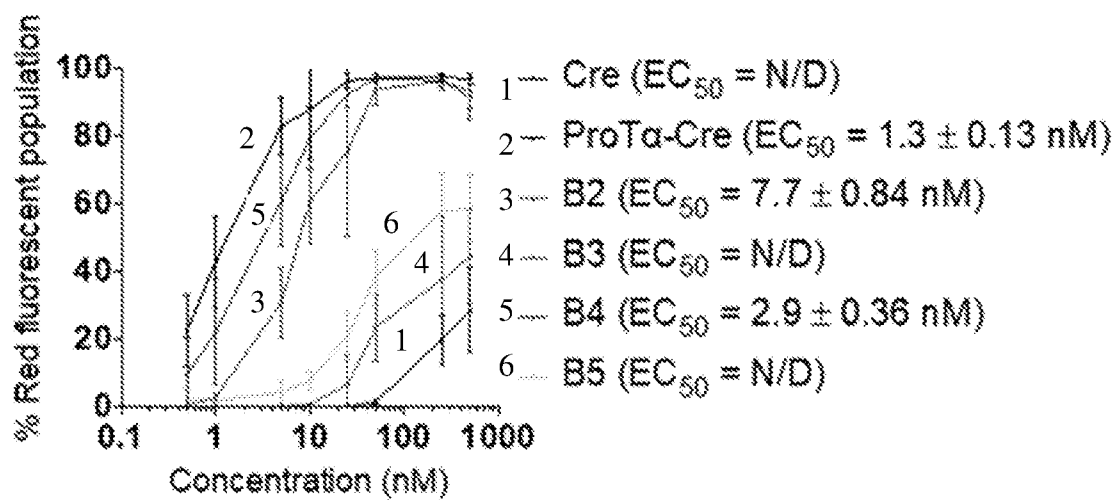


Figure 2

| Name         | Sequence                                                                                                                          |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------|
| ProTα-Cre    | MGASDAAVDTSSSEITTKDLKEKKEVVEEAEN<br>GRDAPANGNAENEENGEGEQEADNEVDEEEE<br>EGGEEEEEEEEEGDGEEEDGDEDEEEAESAT<br>GKRAAEDDEDDVDTKKQKTDEDD |
| (-44)polyD/E | MGADEEESDEEELDEEEDELEEDTDEEG<br>GDEELEDELDEDEETDDEESDEDEDEE                                                                       |
| (-30)polyD/E | MGADEEESDEEELDEEEDELEEDTDEEG<br>GDEELEDELDE                                                                                       |

**Figure 3A****Figure 3B**

**Figure 4A****Figure 4B**

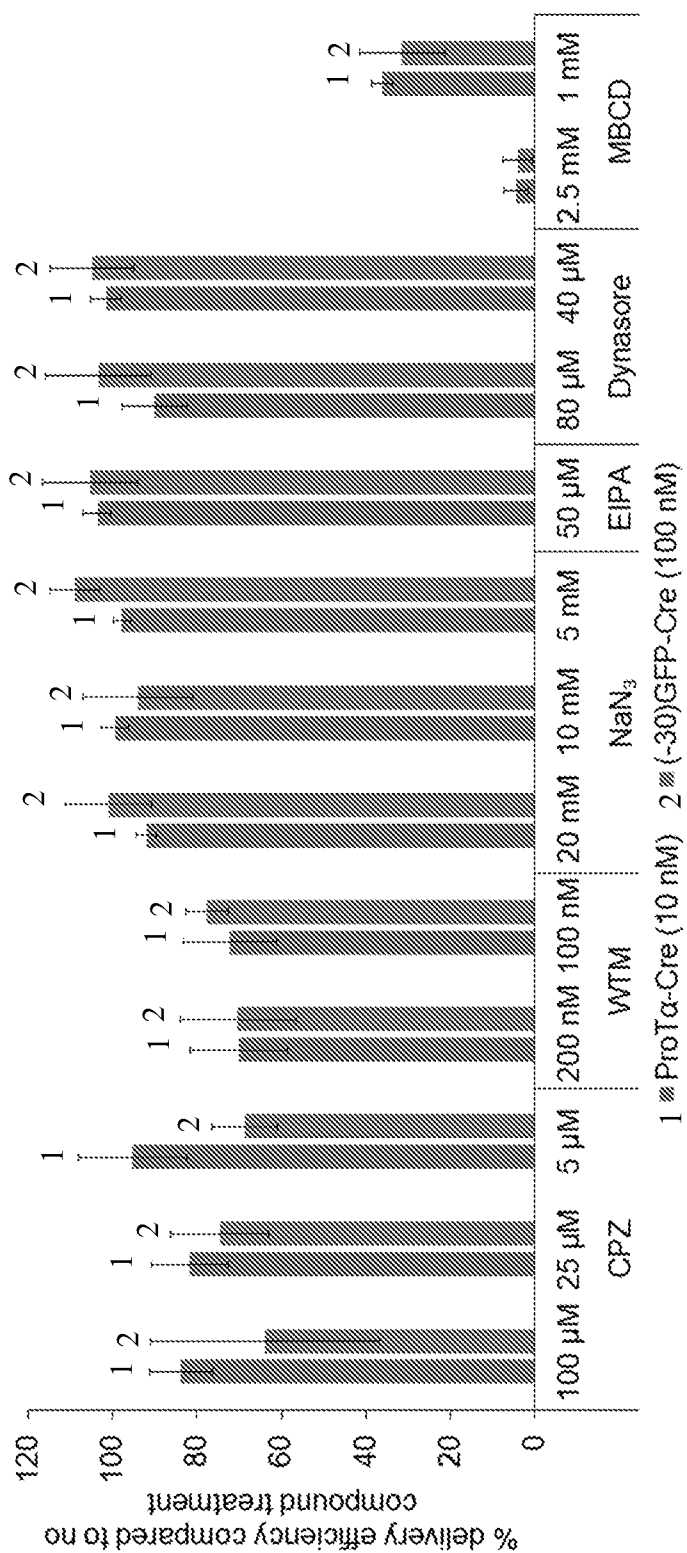
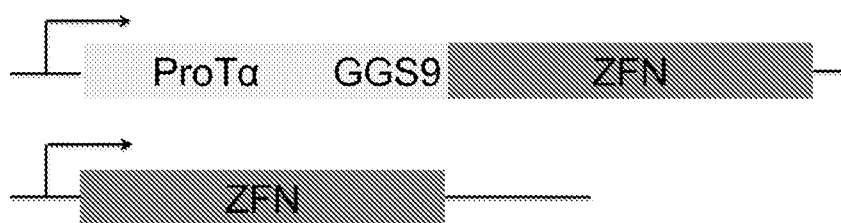
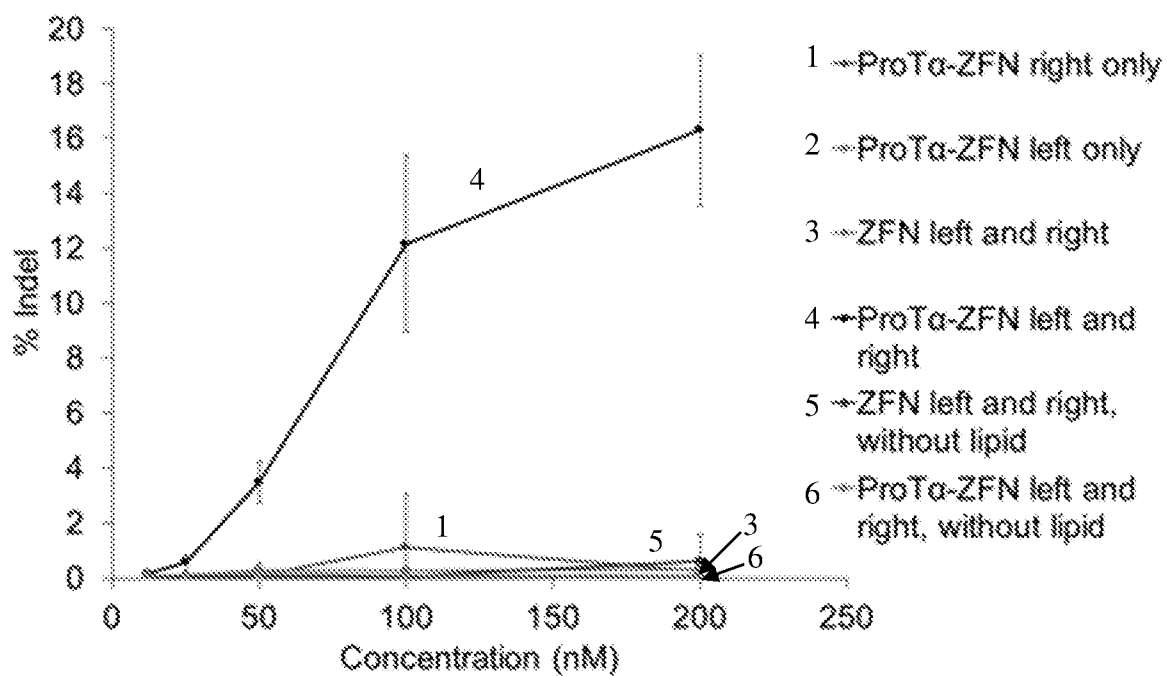


Figure 5

**Figure 6A****Figure 6B**

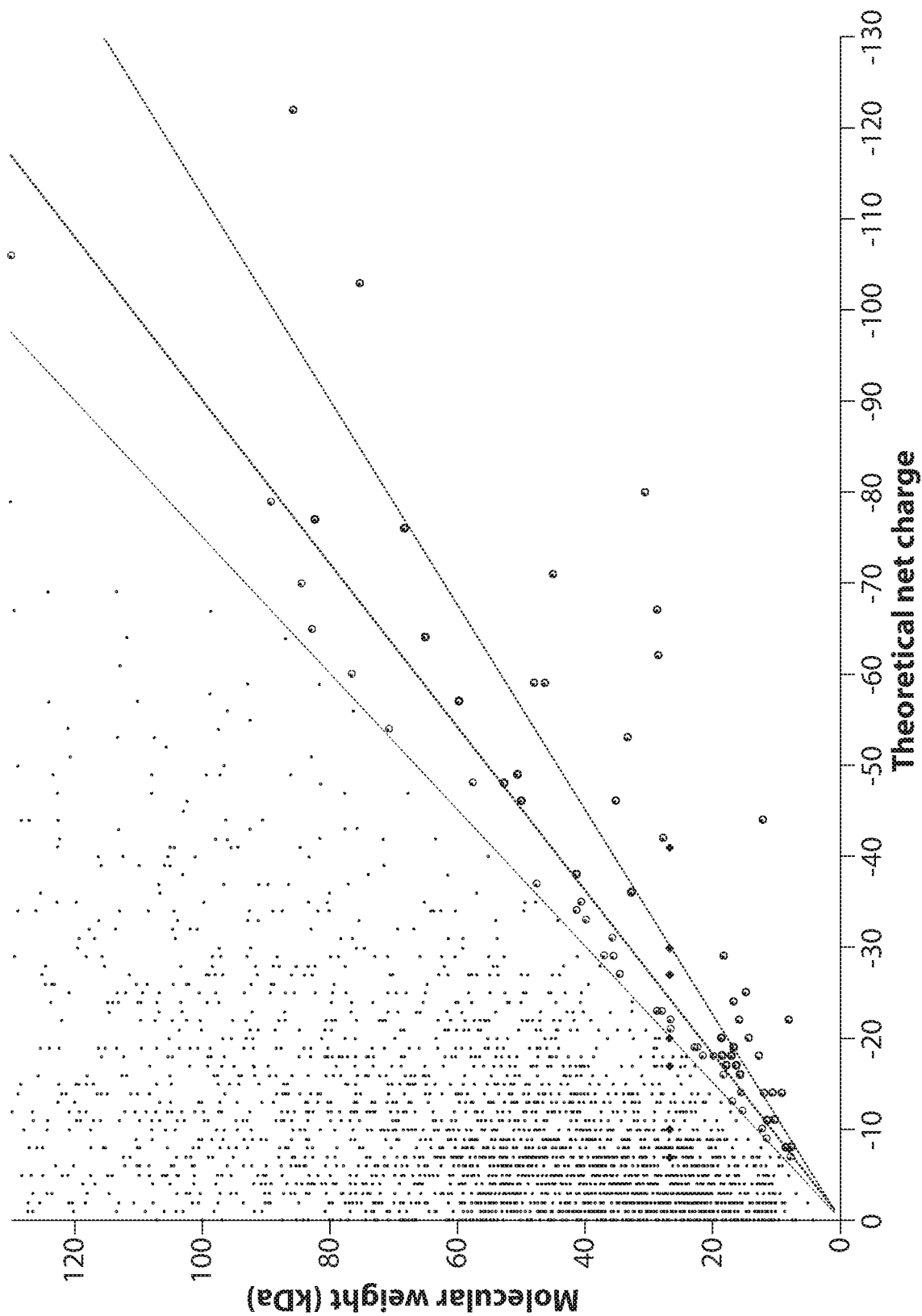


Figure 7

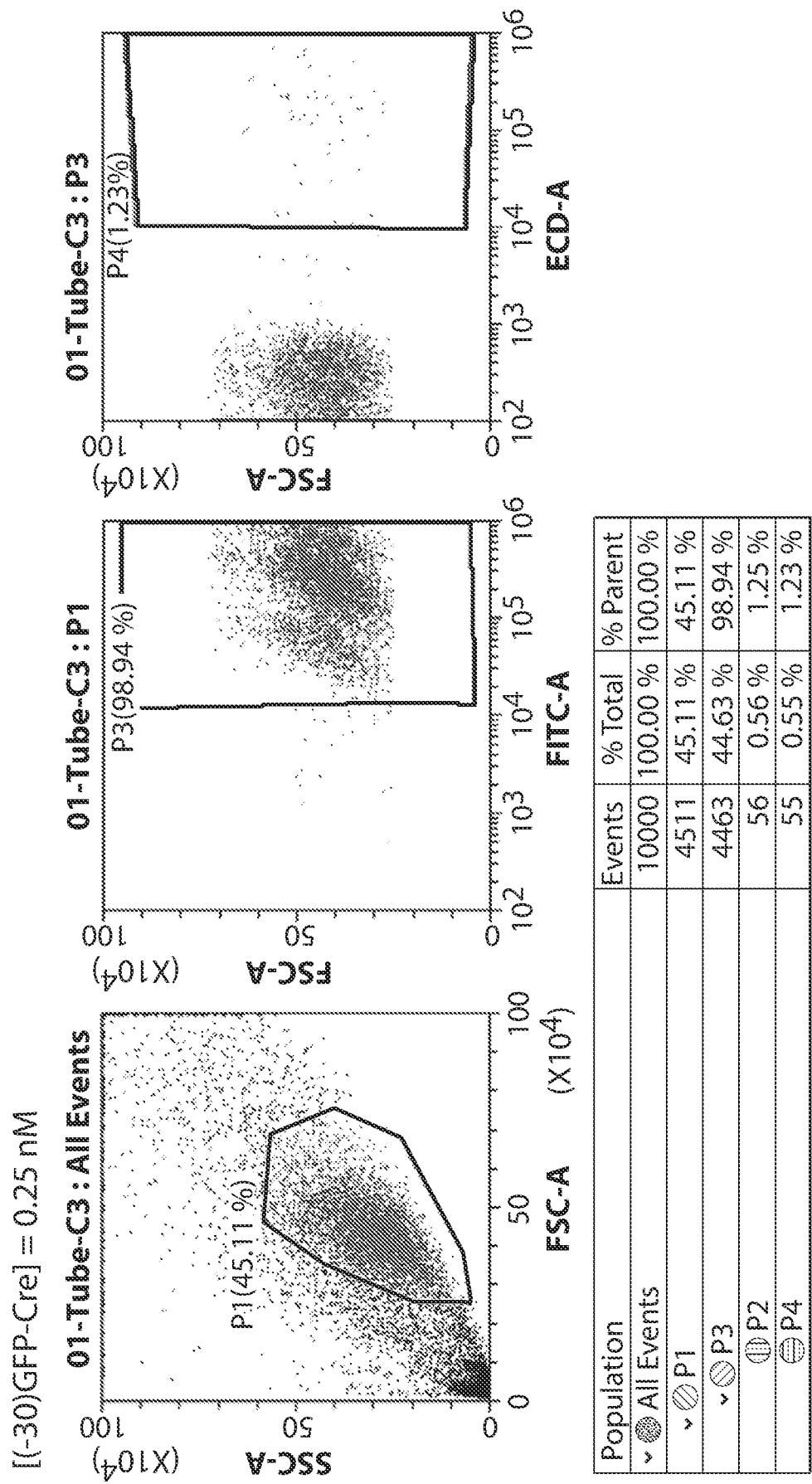


Figure 8A

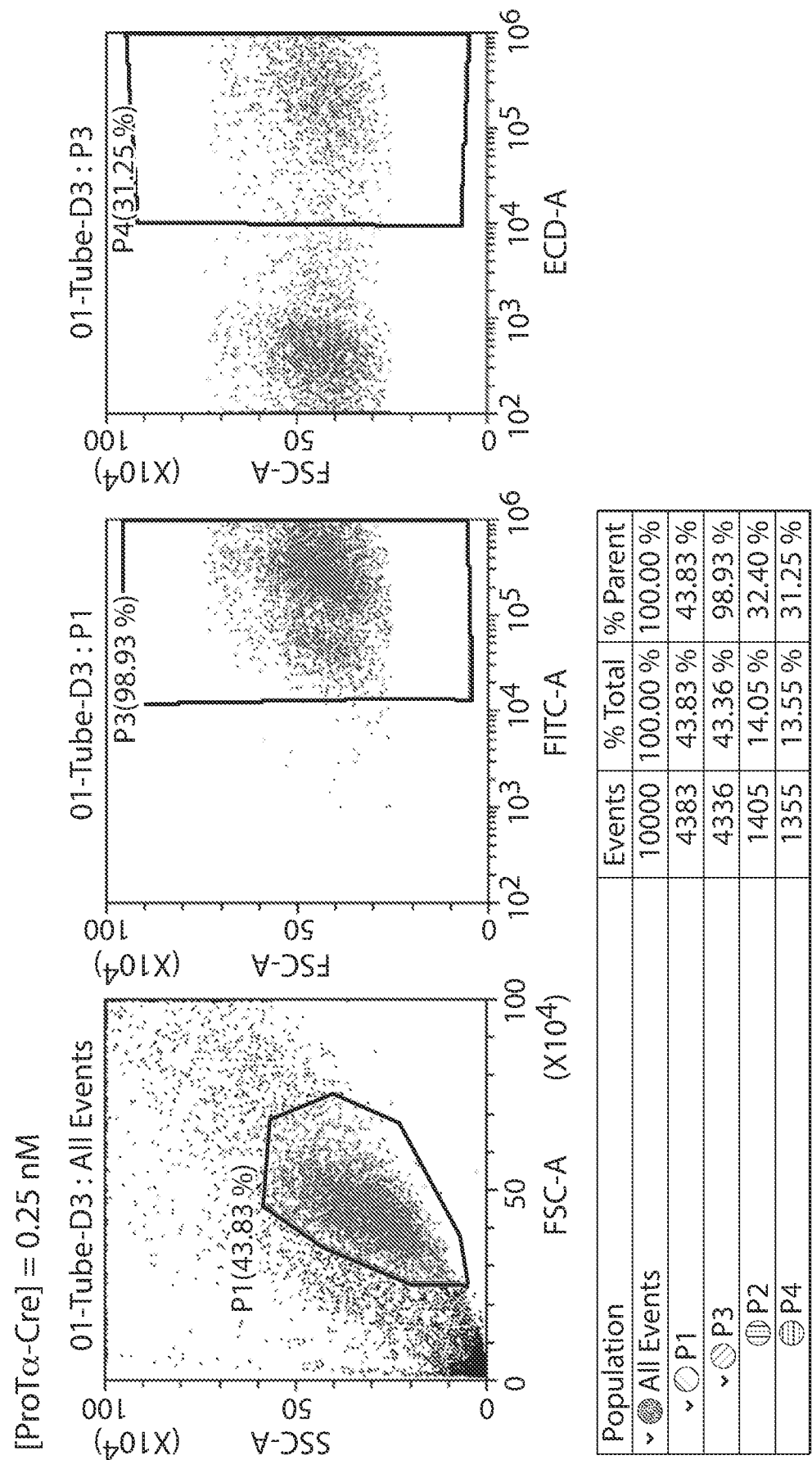
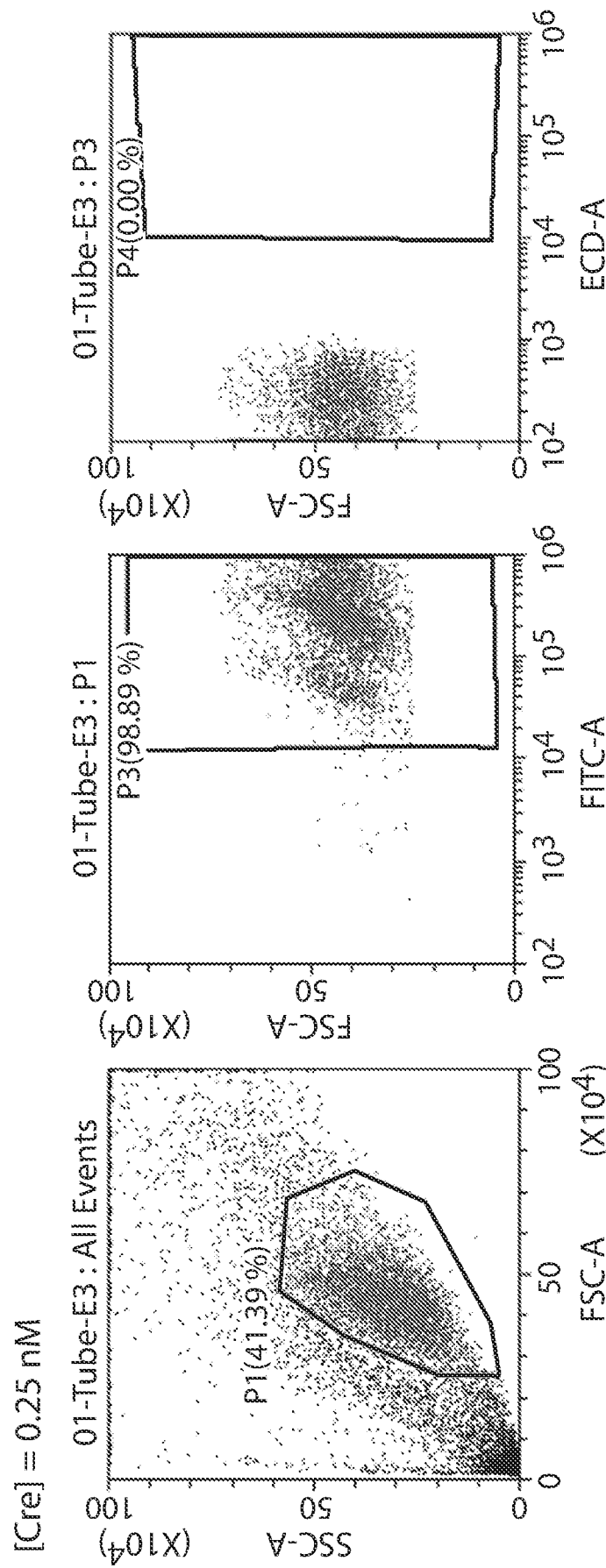


Figure 8B



| Population     | Events | % Total  | % Parent |
|----------------|--------|----------|----------|
| ▼ ● All Events | 10000  | 100.00 % | 100.00 % |
| ▼ ⊘ P1         | 4139   | 41.39 %  | 41.39 %  |
| ▼ ⊙ P3         | 4093   | 40.93 %  | 98.89 %  |
| ⊖ P2           | 0      | 0.00 %   | 0.00 %   |
| ⊖ P4           | 0      | 0.00 %   | 0.00 %   |

Figure 8C

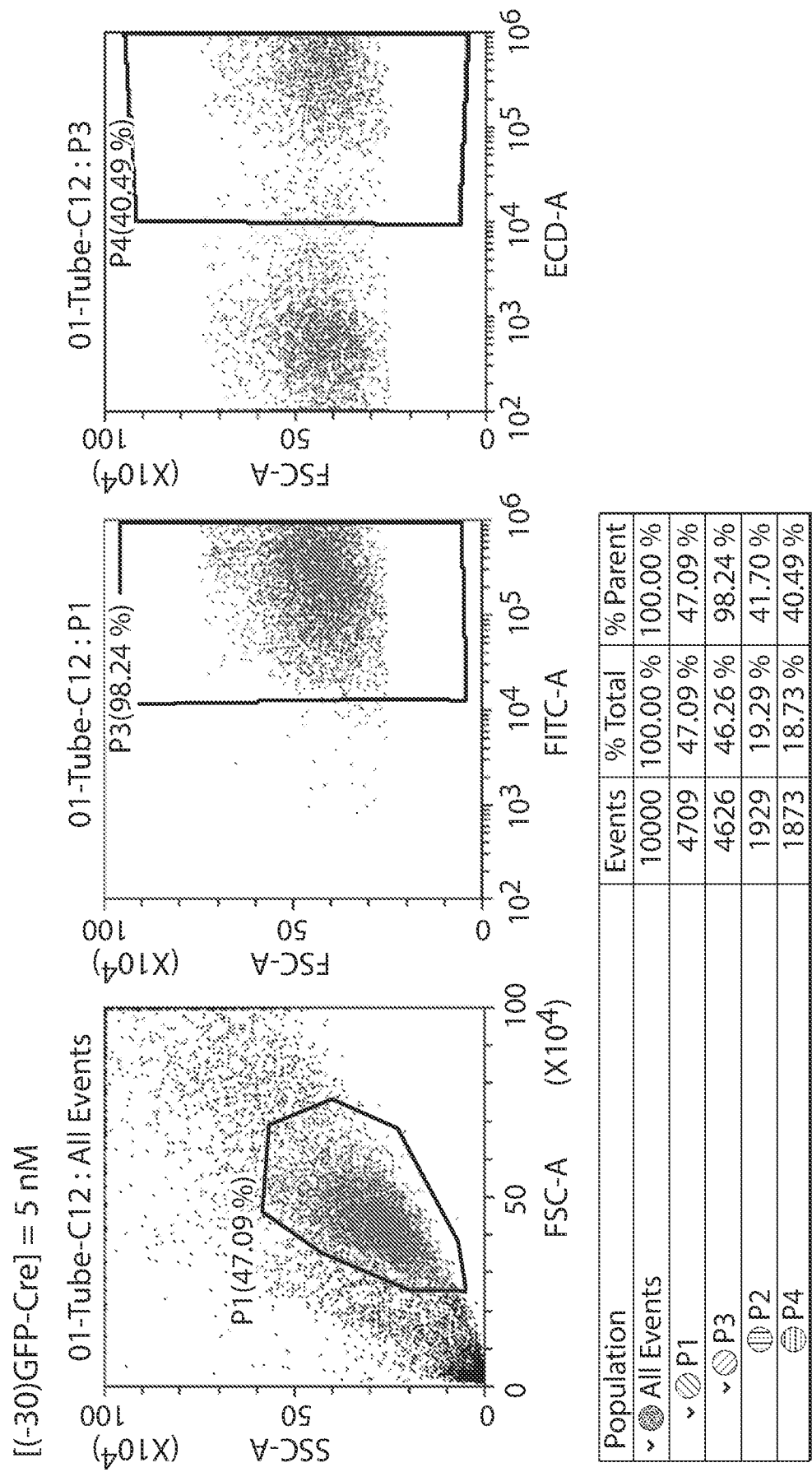


Figure 8D

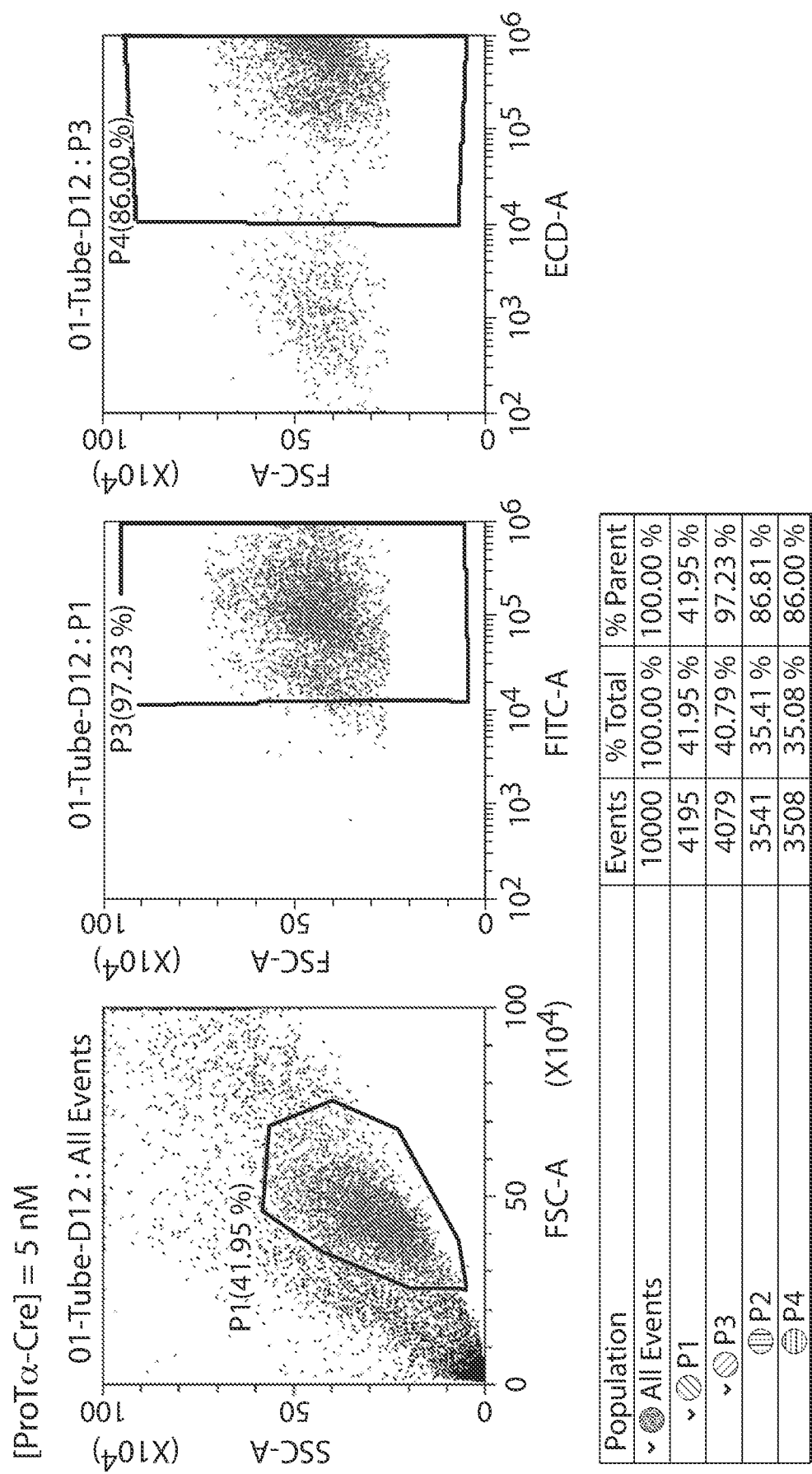


Figure 8E

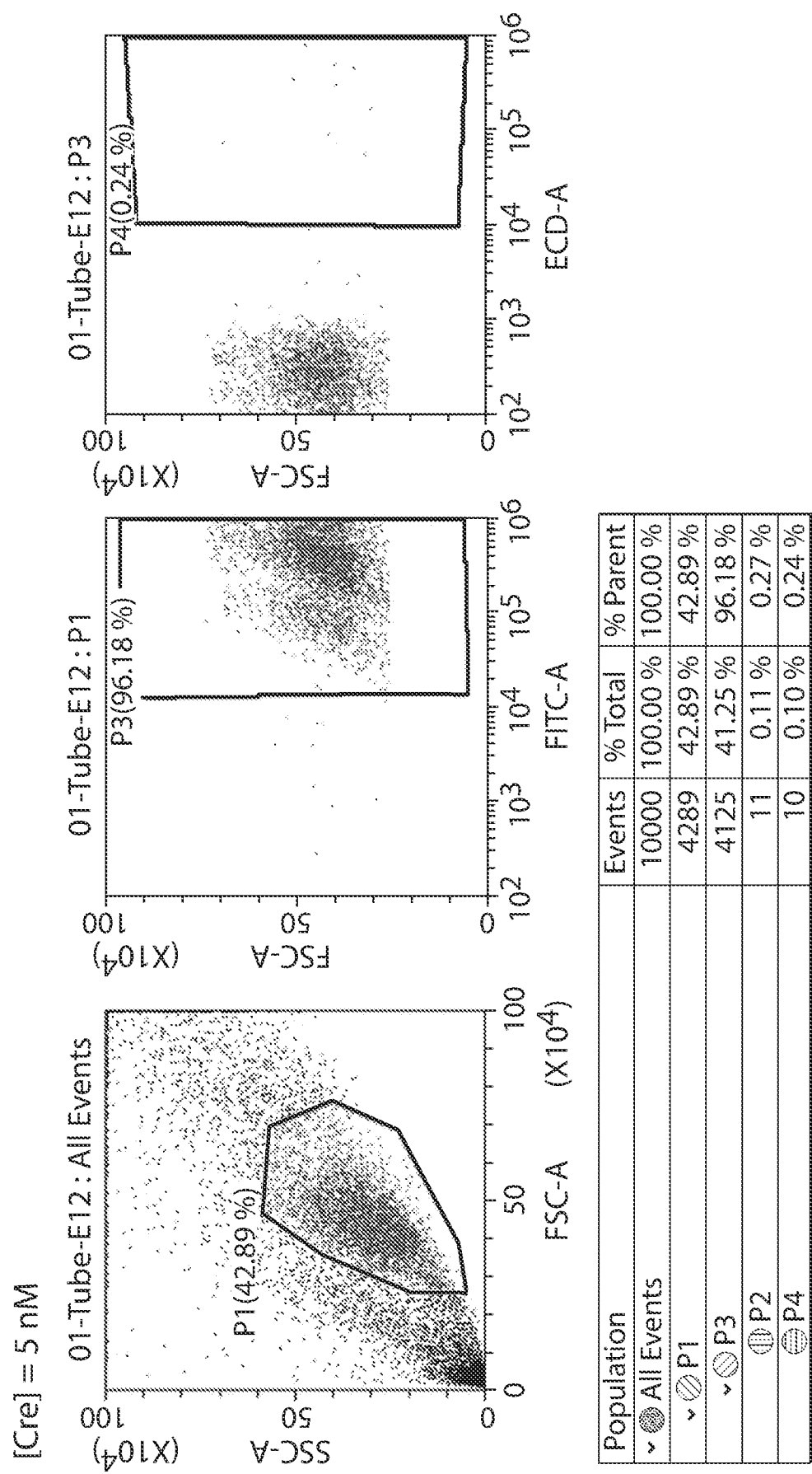
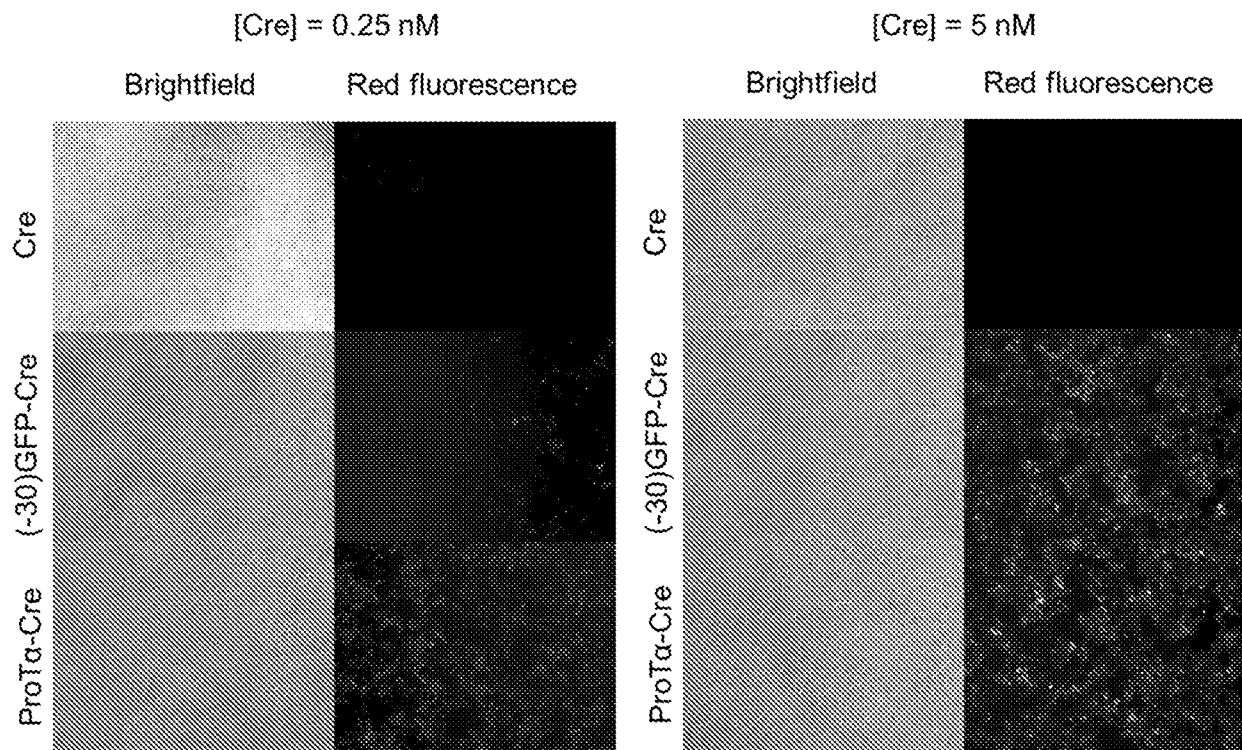
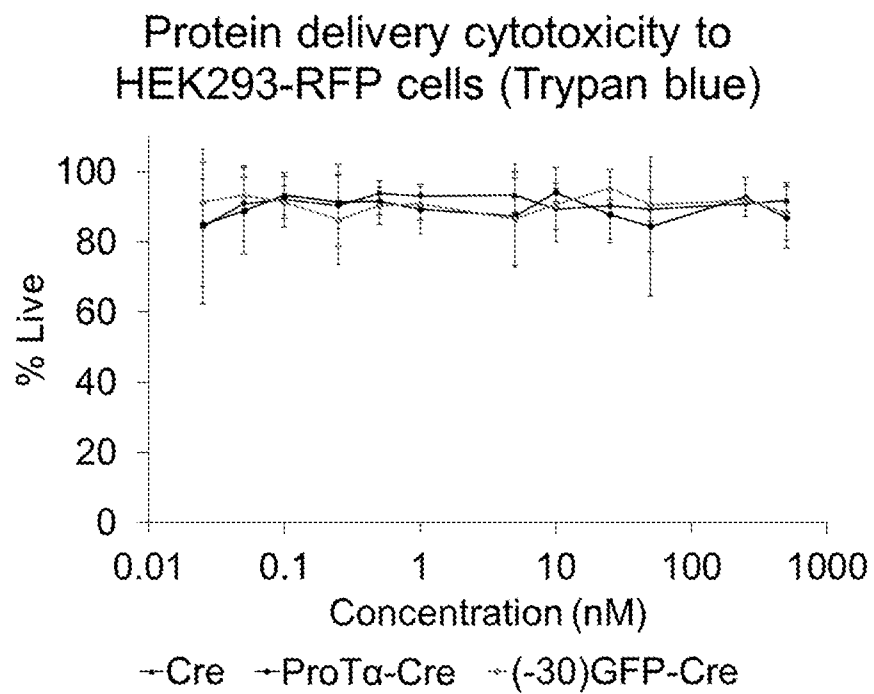
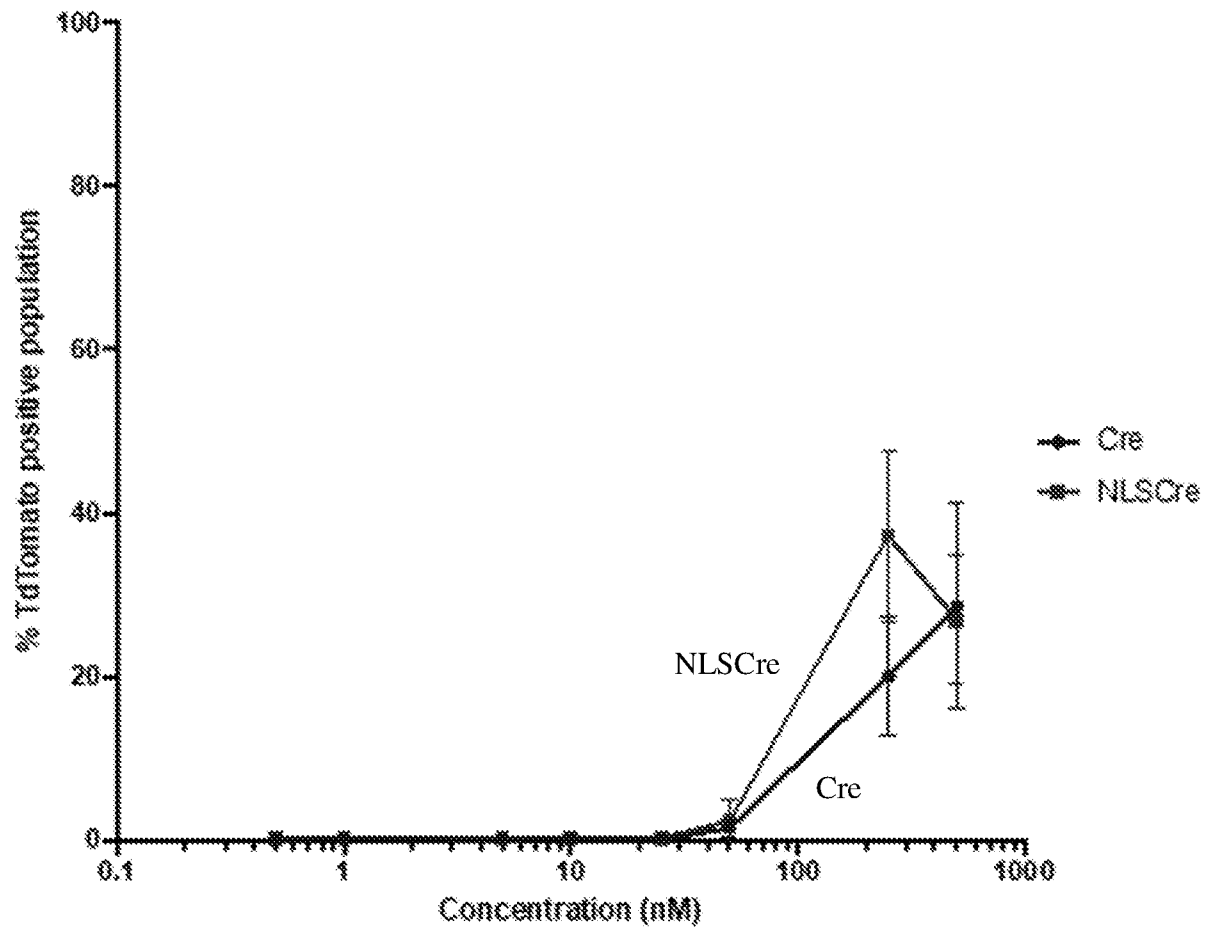


Figure 8F

**Figure 9**

**Figure 10**

**Figure 11**

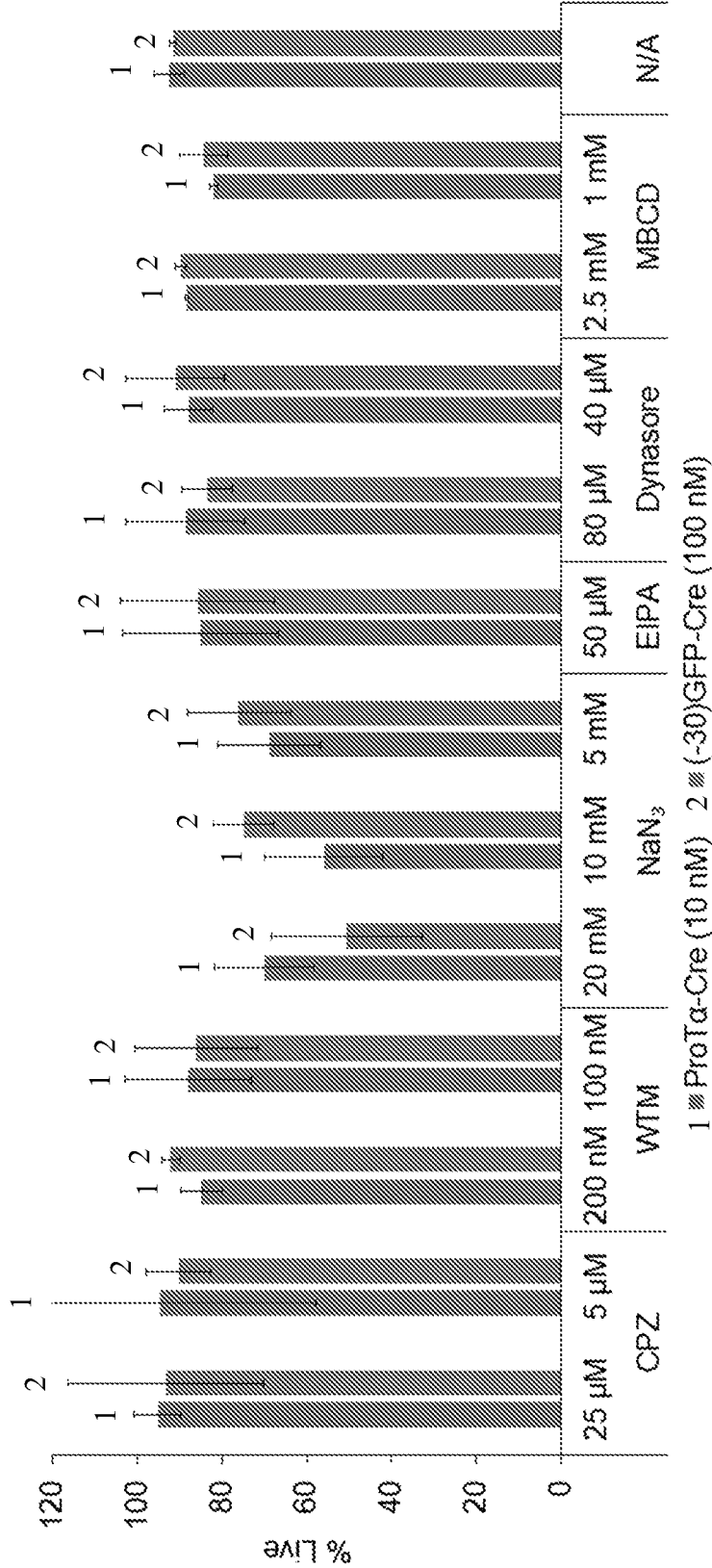


Figure 12

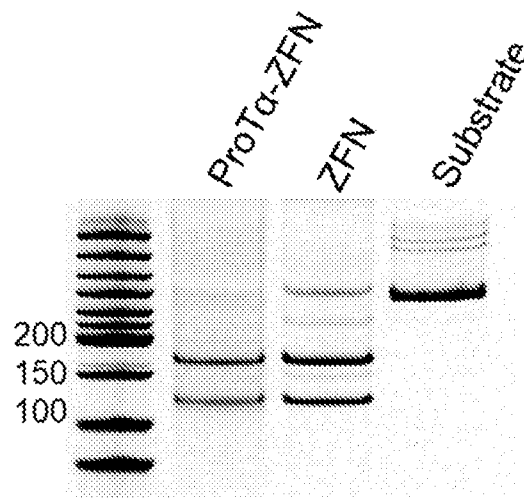


Figure 13A

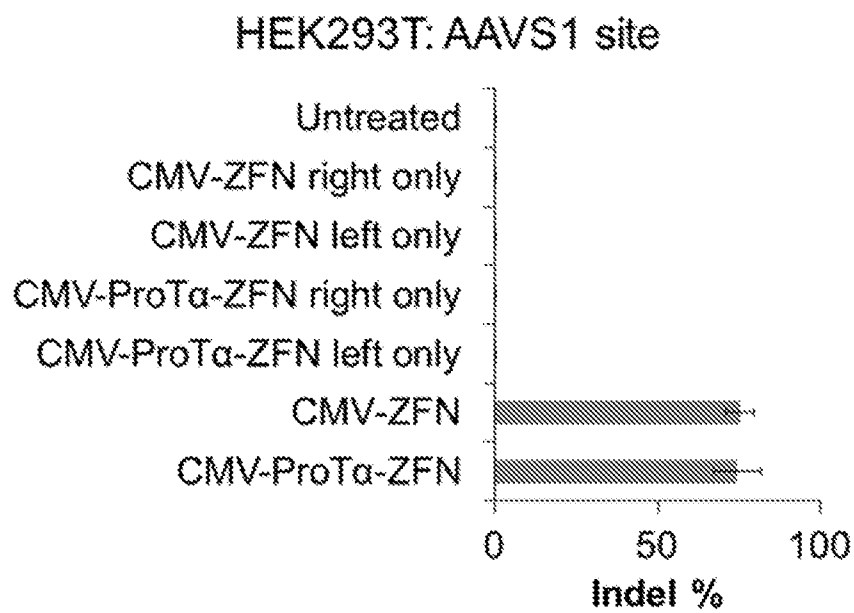


Figure 13B

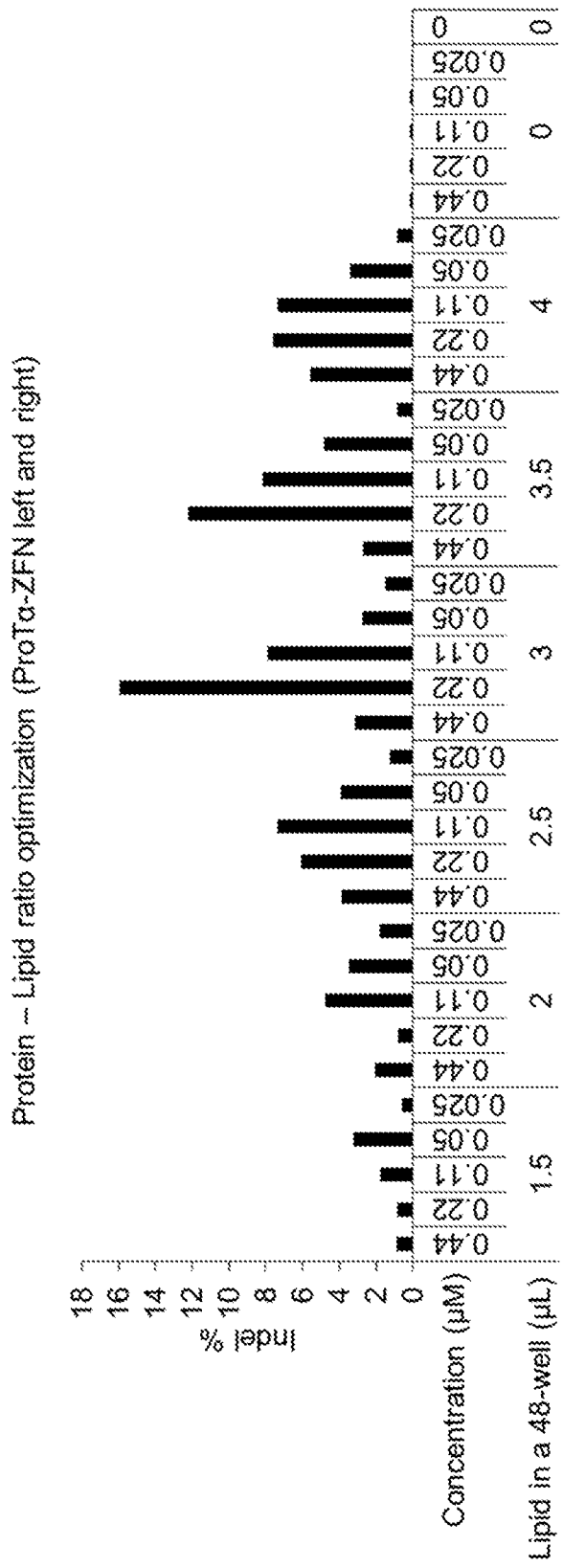


Figure 14

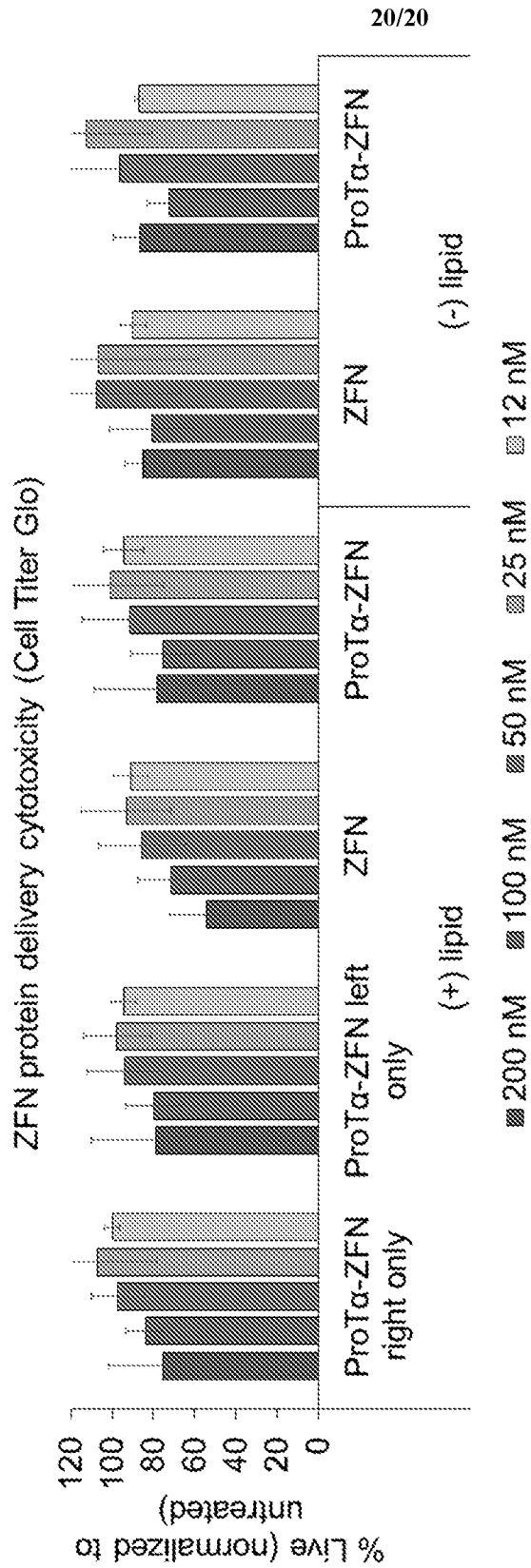


Figure 15

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2020/014789

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C07K 14/47; C12N 9/22 (2020.01)

CPC - A61K 2800/5426; C07K 14/4702; C07K 2319/033; C07K 2319/10 (2020.02)

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

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

USPC - 424/94.6; 435/69.7 (keyword delimited)

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

See Search History document

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                     | Relevant to claim No.              |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| X         | US 2018/0236081 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 23 August 2018 (23.08.2018) entire document                                                                                                                              | 1-3, 5, 7, 9, 11, 13, 15-24, 35-40 |
| --        |                                                                                                                                                                                                                                        |                                    |
| Y         |                                                                                                                                                                                                                                        | 6, 8, 10, 12                       |
| Y         | US 2004/0203109 A1 (LAL et al) 14 October 2004 (14.10.2004) entire document                                                                                                                                                            | 6                                  |
| Y         | SHARER et al. "The ARF-like 2 (ARL2)-binding protein, BART. Purification, cloning, and initial characterization," Journal of Biological Chemistry, 24 September 1999 (24.09.1999), Vol. 274, No. 39, Pgs. 27553-27561. entire document | 8                                  |
| Y         | US 2003/0096337 A1 (HILLMAN et al) 22 May 2003 (22.05.2003) entire document                                                                                                                                                            | 10                                 |
| Y         | US 2005/0136429 A1 (GUARENTE et al) 23 June 2005 (23.06.2005) entire document                                                                                                                                                          | 12                                 |
| A         | WO 2017/191274 A2 (CUREVAC AG) 09 November 2017 (09.11.2017) entire document                                                                                                                                                           | 1-13, 15-24, 35-40                 |
| A         | WO 2018/049073 A1 (FLAGSHIP PIONEERING, INC.) 15 March 2018 (15.03.2018) entire document                                                                                                                                               | 1-13, 15-24, 35-40                 |

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

\* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

30 March 2020

Date of mailing of the international search report

22 APR 2020

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2020/014789

## Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☒ Claims Nos.: 14, 25-34, 41-53  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.