



(86) Date de dépôt PCT/PCT Filing Date: 2014/09/05
(87) Date publication PCT/PCT Publication Date: 2015/03/12
(45) Date de délivrance/Issue Date: 2023/03/21
(85) Entrée phase nationale/National Entry: 2016/03/04
(86) N° demande PCT/PCT Application No.: US 2014/054252
(87) N° publication PCT/PCT Publication No.: 2015/035139
(30) Priorités/Priorities: 2013/09/06 (US61/874,682);
2014/07/08 (US14/326,329); 2014/07/08 (US14/326,340);
2014/07/08 (US14/326,361)

(51) CI.Int./Int.Cl. C12N 15/113 (2010.01),
A61K 48/00 (2006.01), C12N 15/10 (2006.01),
C12N 15/11 (2006.01), C12N 15/55 (2006.01),
C12N 15/63 (2006.01), C12N 9/22 (2006.01)

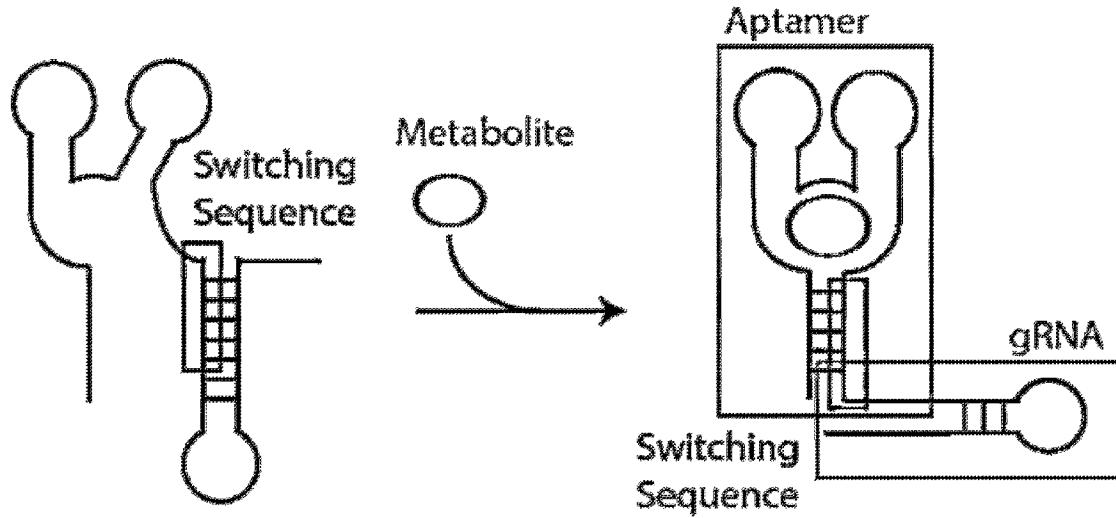
(72) Inventeurs/Inventors:
LIU, DAVID R., US;
HU, JOHNNY HAO, US

(73) Propriétaire/Owner:
PRESIDENT AND FELLOWS OF HARVARD COLLEGE,
US

(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : NUCLEASES CAS9 COMMUTABLES ET LEURS UTILISATIONS

(54) Title: SWITCHABLE CAS9 NUCLEASES AND USES THEREOF



(57) Abrégé/Abstract:

Some aspects of this disclosure provide compositions, methods, systems, and kits for controlling the activity and/or improving the specificity of RNA-programmable endonucleases, such as Cas9. For example, provided are guide RNAs (gRNAs) that are engineered to exist in an "on" or "off" state, which control the binding and hence cleavage activity of RNA-programmable endonucleases. Some aspects of this disclosure provide mRNA-sensing gRNAs that modulate the activity of RNA-programmable endonucleases based on the presence or absence of a target mRNA. Some aspects of this disclosure provide gRNAs that modulate the activity of an RNA-programmable endonuclease based on the presence or absence of an extended DNA (xDNA).

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 March 2015 (12.03.2015)

(10) International Publication Number
WO 2015/035139 A3

(51) International Patent Classification:

C12N 15/113 (2010.01)

(21) International Application Number:

PCT/US2014/054252

(22) International Filing Date:

5 September 2014 (05.09.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/874,682	6 September 2013 (06.09.2013)	US
14/326,329	8 July 2014 (08.07.2014)	US
14/326,340	8 July 2014 (08.07.2014)	US
14/326,361	8 July 2014 (08.07.2014)	US

(71) Applicant: PRISIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: LIU, David, R.; 3 Whitman Circle, Lexington, MA 02420 (US). HU, Johnny, Hao; 553 Cabot Mail Center, 60 Linnaean St., Cambridge, MA 02138 (US).

(74) Agent: BAKER, C., Hunter; Wolf, Greenfield & Sacks. P.C., 600 Atlantic Avenue, Boston, MA 02110-2206 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(88) Date of publication of the international search report:

30 April 2015

(54) Title: SWITCHABLE CAS9 NUCLEASES AND USES THEREOF

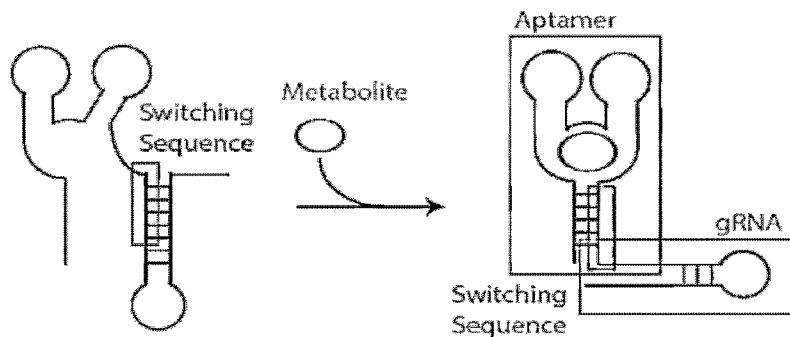


FIG. 1A

(57) Abstract: Some aspects of this disclosure provide compositions, methods, systems, and kits for controlling the activity and/or improving the specificity of RNA-programmable endonucleases, such as Cas9. For example, provided are guide RNAs (gRNAs) that are engineered to exist in an "on" or "off" state, which control the binding and hence cleavage activity of RNA-programmable endonucleases. Some aspects of this disclosure provide mRNA-sensing gRNAs that modulate the activity of RNA-programmable endonucleases based on the presence or absence of a target mRNA. Some aspects of this disclosure provide gRNAs that modulate the activity of an RNA-programmable endonuclease based on the presence or absence of an extended DNA (xDNA).

WO 2015/035139 A3

SWITCHABLE CAS9 NUCLEASES AND USES THEREOF

[0001]

BACKGROUND OF THE INVENTION

[0002] Site-specific endonucleases theoretically allow for the targeted manipulation of a single site within a genome and are useful in the context of gene targeting as well as for therapeutic applications. In a variety of organisms, including mammals, site-specific endonucleases have been used for genome engineering by stimulating either non-homologous end joining or homologous recombination. In addition to providing powerful research tools, site-specific nucleases also have potential as gene therapy agents, and two site-specific endonucleases have recently entered clinical trials: one, CCR5-2246, targeting a human CCR-5 allele as part of an anti-HIV therapeutic approach (NCT00842634, NCT01044654, NCT01252641), and the other one, VF24684, targeting the human VEGF-A promoter as part of an anti-cancer therapeutic approach (NCT01082926).

[0003] Specific cleavage of the intended nuclease target site without or with only minimal off-target activity is a prerequisite for clinical applications of site-specific endonuclease, and also for high-efficiency genomic manipulations in basic research applications. For example, imperfect specificity of engineered site-specific binding domains has been linked to cellular toxicity and undesired alterations of genomic loci other than the intended target. Most nucleases available today, however, exhibit significant off-target activity, and thus may not be suitable for clinical applications. An emerging nuclease platform for use in clinical and research settings are the RNA-guided nucleases, such as Cas9. While these nucleases are able to bind guide RNAs (gRNAs) that direct cleavage of specific target sites, off-target activity is still observed for certain Cas9:gRNA complexes (Pattanayak *et al.*, “High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity.” *Nat Biotechnol.* 2013; doi: 10.1038/nbt.2673). Technology for engineering nucleases with improved specificity is therefore needed.

SUMMARY OF THE INVENTION

[0004] Some aspects of this disclosure are based on the recognition that the reported toxicity of some engineered site-specific endonucleases is based on off-target DNA cleavage. Further, the activity of existing RNA-guided nucleases generally cannot be controlled at the molecular level, for example, to switch a nuclease from an “off” to an “on” state. Controlling the activity of nucleases could decrease the likelihood of incurring off-target effects. Some aspects of this disclosure provide strategies, compositions, systems, and methods to control the binding and/or cleavage activity RNA-programmable endonucleases, such as Cas9 endonuclease.

[0005] Accordingly, one embodiment of the disclosure provides RNA-guided nuclease complexes comprising a “switchable” guide RNA (gRNA). For example, in some embodiments, the invention provides a complex comprising: (i) a gRNA comprising an aptamer, wherein the gRNA does not hybridize to a target nucleic acid in the absence of a specific ligand bound to the aptamer; and (ii) a Cas9 protein. In some embodiments, the aptamer is bound by a ligand. In some aspects, the ligand is any molecule. In some aspects, the ligand is a small molecule, a metabolite, a carbohydrate, a peptide, a protein, or a nucleic acid. In some embodiments, the gRNA:ligand:Cas9 complex binds to and mediates cleavage of a target nucleic acid. See, *e.g.*, Figure 1.

[0006] According to another embodiment, gRNAs comprising an aptamer are provided. In some embodiments, the gRNA does not hybridize to a target nucleic acid in the absence of a ligand bound to the aptamer. Such gRNAs may be referred to as “switchable gRNAs.” For example, in some aspects, the gRNA does not bind Cas9 in the absence of a ligand bound to the aptamer. See, *e.g.*, Figures 1A-B. In some embodiments, the gRNA binds Cas9 when the aptamer is bound by a ligand specific to the aptamer. In some embodiments, the gRNA binds Cas9 in the absence or presence of a ligand bound to the aptamer but binds to a target nucleic acid only in the presence of a ligand bound to the aptamer. In some aspects, the ligand is any molecule. In some aspects, the ligand is a small molecule, a metabolite, a carbohydrate, a peptide, a protein, or a nucleic acid. In some embodiments, the aptamer is an RNA aptamer, for example, an RNA aptamer derived from a riboswitch. In some embodiments, the riboswitch from which the aptamer is derived is selected from a theophylline riboswitch, a thiamine pyrophosphate (TPP) riboswitch, an adenosine cobalamin (AdoCbl) riboswitch, an S-adenosyl methionine (SAM) riboswitch, an SAH riboswitch, a flavin mononucleotide (FMN) riboswitch, a tetrahydrofolate riboswitch, a lysine riboswitch, a glycine riboswitch, a purine riboswitch, a GlnS riboswitch, or a pre-

queosine₁ (PreQ1) riboswitch. In some embodiments, the aptamer is derived from a theophylline riboswitch and comprises SEQ ID NO:3. In other embodiments, the aptamer is non-naturally occurring, and in some aspects, is engineered to bind a specific ligand using a systematic evolution of ligands by exponential enrichment (SELEX) platform. In some embodiments, the non-aptamer portion of the gRNA comprises at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, or at least 150 nucleotides, and the aptamer comprises at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 175, at least 200, at least 250, or at least 300 nucleotides.

[0007] According to another embodiment, methods for site-specific DNA cleavage using the inventive Cas9 variants are provided. For example, in some aspects, the methods comprise contacting a DNA with a complex comprising (i) a gRNA comprising an aptamer, wherein the gRNA comprises a sequence that binds to a portion of the DNA, (ii) a specific ligand bound to the aptamer of the gRNA, and (iii) a Cas9 protein, under conditions in which the Cas9 protein cleaves the DNA.

[0008] According to another embodiment, methods for inducing site-specific DNA cleavage in a cell are provided. For example, in some embodiments, the methods comprise: (a) contacting a cell or expressing within a cell a gRNA comprising an aptamer, wherein the gRNA comprises a sequence capable of binding to a DNA target sequence; (b) contacting a cell or expressing within a cell a Cas9 protein; and (c) contacting the cell with a ligand that binds the aptamer of the gRNA, resulting in the formation of a gRNA:ligand:Cas9 complex that cleaves the DNA target. In some embodiments, the cell produces the ligand intracellularly, for example as a part of physiological or pathophysiological process. In some embodiments, the method comprises (a) contacting the cell with a complex comprising a Cas9 protein and a gRNA comprising an aptamer, wherein the gRNA comprises a sequence capable of binding to a DNA target sequence; and (b) contacting the cell with a ligand that binds the aptamer of the gRNA, resulting in the formation of a gRNA:ligand:Cas9 complex that cleaves the DNA target. In some aspects, steps (a) and (b) are performed simultaneously or sequentially in any order. In some embodiments, the method is performed *in vitro*, whereas in other embodiments the method is performed *in vivo*.

[0009] According to another embodiment, RNA-guided nuclease complexes comprising an mRNA-sensing gRNA are provided. For example, in some embodiments, the complex comprises a Cas9 protein and a gRNA, wherein the gRNA comprises: (i) a region

that hybridizes a region of a target nucleic acid; (ii) another region that partially or completely hybridizes to the sequence of region (i); and (iii) a region that hybridizes to a region of a transcript (mRNA).

[0010] According to another embodiment, mRNA-sensing gRNAs are provided, for example, that comprise: (i) a region that hybridizes a region of a target nucleic acid; (ii) another region that partially or completely hybridizes to the sequence of region (i); and (iii) a region that hybridizes to a region of a transcript (mRNA). See, *e.g.*, Figure 2. In some embodiments, each of the sequences of regions (i), (ii), and (iii) comprise at least 5, at least 10, at least 15, at least 20, or at least 25 nucleotides. In some aspects, the gRNA forms a stem-loop structure in which the stem comprises the sequence of region (i) hybridized to part or all of the sequence of region (ii), and the loop is formed by part or all of the sequence of region (iii). In some embodiments, regions (ii) and (iii) are both either 5' or 3' to region (i). See, *e.g.*, Figure 2A vs. Figure 2C. In some embodiments, the stem-loop structure forms in the absence of the transcript that hybridizes to the sequence of region (iii). In this example, the gRNA is said to be in the “off” state. See, *e.g.*, Figure 2A, 2C. In some embodiments, the binding of the transcript to the sequence of region (iii) results in the unfolding of the stem-loop structure, or prevents the formation of the stem-loop structure, such that the sequence of region (ii) does not hybridize to the sequence of region (i). In this example, the gRNA is said to be in the “on” state. See, *e.g.*, Figure 2B, 2D. In some embodiments, the gRNA binds a Cas9 protein, and the sequence of region (i) hybridizes to the target nucleic acid when the sequence of region (iii) binds (*e.g.*, “senses”) the transcript.

[0011] According to another embodiment, methods for site specific DNA cleavage are provided, for example, that comprise contacting a DNA with the complex comprising a Cas9 protein associated with an mRNA-sensing gRNA, wherein the mRNA-sensing gRNA is bound by an mRNA thereby allowing the complex to bind and cleave the DNA.

[0012] According to another embodiment, extended DNA recognition (xDNA-sensing) gRNAs are provided. See, *e.g.*, Figure 3. In some embodiments, xDNA-sensing gRNAs comprise: (i) a region that hybridizes a region of a target nucleic acid; (ii) another region that partially or completely hybridizes to the sequence of region (i); and (iii) a region that hybridizes to another region of the target nucleic acid. In some embodiments, each of the sequences of regions (i) and (ii) comprise at least 5, at least 10, at least 15, at least 20, or at least 25 nucleotides; and the sequence of region (iii) comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, or at least 100 nucleotides. In some embodiments, the gRNA forms a stem-loop structure in which the stem

comprises the sequence of region (i) hybridized to part or all of the sequence of region (ii), and the loop is formed by part or all of the sequence of region (iii). In some embodiments, regions (ii) and (iii) are both either 5' or 3' to region (i). See *e.g.*, Figure 3A vs. Figure 3C. In some embodiments, the stem-loop structure forms in the absence of the region of the target nucleic acid that complements and binds the sequence in region (iii). See, *e.g.*, Figure 3A, C. In some embodiments, the hybridization of the region of the target nucleic acid to the sequence of region (iii) results in the unfolding of the stem-loop structure, or prevents the formation of the stem-loop structure, such that the sequence of region (ii) does not hybridize to the sequence of region (i). See, *e.g.*, Figure 3B, D. In some embodiments, the gRNA binds a Cas9 protein, and the sequence in (i) binds the target nucleic acid when the sequence in region (iii) binds the target nucleic acid.

[0013] According to another embodiment, complexes comprising an xDNA-sensing gRNA and a Cas9 protein are provided, optionally wherein the complex comprises a target nucleic acid. In some embodiments, the formation of the complex results in the cleavage of the target nucleic acid.

[0014] According to another embodiment, methods for site-specific DNA cleavage are provided comprising contacting a DNA with the complex comprising an xDNA-sensing gRNA and a Cas9 protein.

[0015] Any of the methods provided herein can be performed on DNA in a cell, for example, a cell *in vitro* or *in vivo*. In some embodiments, any of the methods provided herein are performed on DNA in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual, for example, a human.

[0016] According to another embodiment, polynucleotides are provided, for example, that encode any of the gRNAs, complexes, or proteins (*e.g.*, Cas9 proteins) described herein. In some embodiments, vectors that comprise a polynucleotide described herein are provided. In some embodiments, vectors for recombinant expression of any of the gRNAs, complexes, or proteins (*e.g.*, Cas9 proteins) described herein are provided. In some embodiments, cells comprising genetic constructs for expressing any of the gRNAs, complexes, or proteins (*e.g.*, Cas9 proteins) described herein are provided.

[0017] In some embodiments, kits are provided. For example, kits comprising any of the gRNAs, complexes, or proteins (*e.g.*, Cas9 proteins) described herein are provided. In some embodiments, kits comprising any of the polynucleotides described herein are provided. In some embodiments, kits comprising a vector for recombinant expression, wherein the vectors comprise a polynucleotide encoding any of the gRNAs, complexes, or

proteins (*e.g.*, Cas9 proteins) described herein, are provided. In some embodiments, kits comprising a cell comprising genetic constructs for expressing any of the gRNAs, complexes, or proteins (*e.g.*, Cas9 proteins) described herein are provided.

[0018] Other advantages, features, and uses of the invention will be apparent from the Detailed Description of Certain Embodiments of the Invention; the Drawings, which are schematic and not intended to be drawn to scale; and the Claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] **Figures 1A-D** shows certain embodiments of the invention relating to gRNAs linked to aptamers. (A) In this figure, a switchable gRNA comprising an aptamer is schematically depicted. In the absence of a specific ligand (here, a metabolite) that binds the aptamer, the sequence responsible for binding the target nucleic acid is hybridized to aspects of the aptamer (see far left, area depicted as “switching sequence”). Upon binding of the metabolite, the aptamer undergoes conformational changes, such that the sequence responsible for binding the target nucleic acid no longer hybridizes to the aptamer sequence, allowing it to hybridize to the target. (B) Upon switching to an “on” state, the gRNA, when bound to Cas9, directs the nuclease to the target site where it hybridizes to the target site, allowing Cas9 to cleave each strand of the target nucleic acid. (C-D) In this figure, the aptamer linked to a gRNA is derived from the theophylline riboswitch. In the absence of theophylline (C), aspects of the aptamer (depicted as “Theophylline Riboswitch”) bind aspects of the sequence (depicted as “Guide to Cut the Target”) responsible for binding the target nucleic acid (depicted by the double-stranded sequence at the top of the panel), thereby precluding the gRNA from hybridizing to the target nucleic acid. In Figure 1C, the sequences, from top to bottom, correspond to SEQ ID NOs: 4-6. When aptamer is bound by theophylline (depicted as solid small molecule binding the aptamer sequence) (D), it undergoes conformational changes resulting in the “Guide” sequence being free to hybridize to the target nucleic acid. In Figure 1D, the sequences, from top to bottom, correspond to SEQ ID NOs: 4, 5, and 11.

[0020] **Figures 2A-D** shows certain embodiments of the invention relating to mRNA-sensing gRNAs. (A-B) In this figure, a gRNA comprising a 5' transcript sensor/guide block motif is depicted. In the absence of a certain mRNA (A), aspects of the transcript sensor remain unbound, resulting in the formation of a stem-loop structure that blocks certain aspects of the sequence (depicted as “Guide to Cut the Target”) responsible for binding the target nucleic acid (depicted by the double-stranded sequence at the top of the panel), thereby

preventing the gRNA from hybridizing to the target nucleic acid. In Figure 2A, the sequences, from top to bottom, correspond to SEQ ID NOs: 4, 5, and 7. In the presence of the mRNA to which the transcript sensor hybridizes (B), the gRNA undergoes conformational changes resulting in the “Guide” sequence being free to hybridize to the target nucleic acid, sequences, from top to bottom, correspond to SEQ ID NOs: 4, 5, 7, and 12. (C-D) Similarly, the strategy may be applied to gRNAs comprising a 3' transcript sensor/guide block, such that in the absence of the mRNA (C), the gRNA is in the “off” state (sequences, from top to bottom, correspond to SEQ ID NOs: 4, 5, and 8), and in the presence of the mRNA (D), the gRNA is in the “on” state (sequences, from top to bottom, correspond to SEQ ID NOs: 4, 5, and 12).

[0021] **Figures 3A-D** shows certain embodiments of the invention relating to extended DNA (xDNA) recognition strategies. (A-B) In this embodiment, a gRNA comprising a 5' xDNA sensor/guide block motif is depicted. The xDNA sensor motif complements and hybridizes to other aspects of the target nucleic acid (*e.g.*, in addition to the “Guide to Cut Target” sequence). (A) In the absence of the correct target sequence (*e.g.*, comprising both the target of the “guide” sequence as well as the target of the xDNA sensor sequence), aspects of the xDNA sensor remain unbound, resulting in the formation of a stem-loop structure that blocks certain aspects of the sequence (depicted as “Guide to Cut the Target”) responsible for binding the target nucleic acid (depicted by the double-stranded sequence at the top of the panel), thereby preventing the gRNA from hybridizing to the target nucleic acid, sequences, from top to bottom, correspond to SEQ ID NOs: 4, 5, and 9. In the presence of the correct target nucleic acid to which portion(s) of the xDNA sensor hybridize(s) (B), the gRNA undergoes conformational changes resulting in the “Guide” sequence being free to hybridize to the target nucleic acid, sequences, from top to bottom, correspond to 13, 14, and 9. Thus, only in the presence of the correct target nucleic acid does binding of the gRNA (and associated Cas9 protein) occur. This effectively increases (*i.e.*, extends) the number of target nucleotides recognized by a *e.g.*, a Cas9:gRNA complex, which increase specificity. (C-D) Similarly, the strategy may be applied to gRNAs comprising a 3' xDNA sensor/guide block, such that in the absence of the target nucleic acid (C), the gRNA is in the “off” state (sequences, from top to bottom, correspond to SEQ ID NOs: 4, 5, and 10), and in the presence of the target nucleic acid (D), the gRNA is in the “on” state, sequences, from top to bottom and left to right, correspond to SEQ ID NOs: 4, 15, 5, 16, and 10.

DEFINITIONS

[0022] 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 such agents.

[0023] The term “aptamer” refers to nucleic acid or peptide molecules that bind to a specific target molecule, *e.g.*, a specific ligand. In some embodiments, binding of the ligand to the aptamer induces conformational changes in the aptamer, and *e.g.*, other molecules conjugated or linked to the aptamer. In some embodiments, nucleic acid (*e.g.*, DNA or RNA) aptamers are engineered through repeated rounds of *in vitro* selection or equivalently, SELEX (systematic evolution of ligands by exponential enrichment) to bind to various molecular targets, for example, small molecules, macromolecules, metabolites, proteins, carbohydrates, metals, nucleic acids, cells, tissues and organisms. Methods for engineering aptamers to bind small molecules are known in the art and include those described in U.S. Pat Nos. 5,580,737 and 8,492,082; Ellington and Szostak, “In vitro selection of RNA molecules that bind specific ligands.” *Nature*. 1990; 346:818–822; Tuerk and Gold, “Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.” *Science*. 1990; 249:505–510; Burke and Gold, “RNA aptamers to the adenosine moiety of S-adenosyl methionine: structural inferences from variations on a theme and the reproducibility of SELEX.” *Nucleic Acids Res.* 1997; 25(10):2020-4; Ulrich *et al.*, “DNA and RNA aptamers: from tools for basic research towards therapeutic applications.” *Comb Chem High Throughput Screen.* 2006; 9(8):619-32; Svobodová *et al.*, “Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Anal Bioanal Chem.* 2012; 404:835–842 .

Nucleic acid aptamers are also found in nature, for example, those that form part of a riboswitch. A “riboswitch” is a regulatory segment of a mRNA molecule that binds a small molecule, for example, a metabolite, resulting in a change in production of the protein(s) encoded by the mRNA (*e.g.*, proteins involved in the production of the metabolite binding the riboswitch). Riboswitches are often conceptually divided into two parts: an aptamer and an expression platform (*e.g.*, mRNA). The aptamer directly binds the small molecule (*e.g.*, metabolite), and the mRNA undergoes structural changes in response to the changes in the aptamer. Typically, the structural changes in the mRNA result in a decrease or inhibition of protein expression. Aptamers can be cloned from (*e.g.*, separated from) riboswitches and used to control the activity of other molecules (*e.g.*,

RNA, DNA) linked thereto using routine methods in the art. Additionally, aptamers found in nature can be re-engineered to bind to synthetic, non-natural small molecule ligands to control the activities of other molecules linked thereto using known methods. See, e.g., Dixon *et al.*, “Reengineering orthogonally selective riboswitches.” *PNAS* 2010; **107** (7): 2830–2835.

The following

is a non-limiting list of riboswitches that include aptamers:

[0024] *Cobalamin riboswitch* (also *B₁₂-element*), which binds adenosylcobalamin (the coenzyme form of vitamin B₁₂) to regulate cobalamin biosynthesis and transport of cobalamin and similar metabolites, and other genes. See, e.g., Nahvi *et al.*, “Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes.” *Nucleic Acids Res.* 2004; **32**: 143-150; Vitreschak *et al.*, “Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element.” *RNA*. 2003; **9**:1084-1097.

[0025] *cyclic di-GMP riboswitches* bind the signaling molecule cyclic di-GMP in order to regulate a variety of genes controlled by this second messenger. At least two classes of cyclic di-GMP riboswitches are known: cyclic di-GMP-I riboswitches and cyclic di-GMP-II riboswitches. See, e.g., Sudarsan *et al.*, “Riboswitches in eubacteria sense the second messenger cyclic di-GMP.” *Science*. 2008; **321** (5887): 411–3; Lee *et al.*, “An allosteric self-splicing ribozyme triggered by a bacterial second messenger.” *Science*. 2010; **329** (5993): 845–8.

[0026] *FMN riboswitch* (also *RFN-element*) binds flavin mononucleotide (FMN) to regulate riboflavin biosynthesis and transport. See, e.g., Winkler *et al.*, “An mRNA structure that controls gene expression by binding FMN.” *Proc Natl Acad Sci USA*. 2002; **99** (25): 15908–15913; Serganov *et al.*, “Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch.” *Nature*. 2009; **458** (7235): 233–7.

[0027] *GlmS riboswitch* is a ribozyme that cleaves itself when bound by glucosamine-6-phosphate. See, e.g., Winkler *et al.*, “Control of gene expression by a natural metabolite-responsive ribozyme.” *Nature*. 2004; **428**: 281-286; Jansen *et al.*, “Backbone and nucleobase contacts to glucosamine-6-phosphate in the glmS ribozyme.” *Nat Struct Mol Biol*. 2006; **13**: 517-523; Hampel and Tinsley, “Evidence for preorganization of the glmS ribozyme ligand binding pocket.” *Biochemistry*. 2006; **45**: 7861-7871.

[0028] *Glycine riboswitch* binds glycine to regulate glycine metabolism genes, including the use of glycine as an energy source. See, *e.g.*, Mandal *et al.*, “A glycine-dependent riboswitch that uses cooperative binding to control gene expression.” *Science*. 2004; **306** (5694): 275–279; Kwon and Strobel, “Chemical basis of glycine riboswitch cooperativity.” *RNA*. 2008; **14** (1): 25–34.

[0029] *Lysine riboswitch* (also *L-box*) binds lysine to regulate lysine biosynthesis, catabolism and transport. See, *e.g.*, Sudarsan *et al.*, “An mRNA structure in bacteria that controls gene expression by binding lysine.” *Genes Dev.* 2003; **17**:2688-2697; Grundy *et al.*, “The L box regulon: Lysine sensing by leader RNAs of bacterial lysine biosynthesis genes.” *Proc. Natl. Acad. Sci. USA*. 2003; **100**:12057-12062.

[0030] *PreQ1 riboswitches* bind pre-queuosine₁, to regulate genes involved in the synthesis or transport of this precursor to queuosine. At least two distinct classes of PreQ1 riboswitches are known: PreQ1-I riboswitches and PreQ1-II riboswitches. See, *e.g.*, Roth *et al.*, “A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain,” *Nat Struct Mol Biol.* 2007; **14** (4): 308–317; Klein *et al.*, “Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase,” *Nat. Struct. Mol. Biol.* 2009; **16** (3): 343–344; Kang *et al.*, “Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA.” *Mol. Cell* **33** 2009; (6): 784–90; Meyer *et al.*, “Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria.” *RNA* 2008; **14** (4): 685.

[0031] *Purine riboswitches* binds purines to regulate purine metabolism and transport. Different forms of the purine riboswitch bind guanine (a form originally known as the *G-box*) or adenine. The specificity for either guanine or adenine depends completely upon Watson-Crick interactions with a single pyrimidine in the riboswitch at a particular position, *e.g.*, Y74. In the guanine riboswitch this residue is typically a cytosine (*e.g.*, C74), in the adenine riboswitch it is typically a uracil (*e.g.*, U74). Homologous types of purine riboswitches bind deoxyguanosine but have more significant differences than a single nucleotide mutation. See *e.g.*, Serganov *et al.*, “Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs.” *Chem Biol.* 2004; **11** (12): 1729–41; Batey *et al.*, “Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine.” *Nature*. 2004; **432** (7015): 411–415; Mandal and Breaker, “Adenine

riboswitches and gene activation by disruption of a transcription terminator.” *Nat Struct Mol Biol.* 2004; **11** (1): 29–35.

[0032] *SAH riboswitches* bind S-adenosylhomocysteine to regulate genes involved in recycling this metabolite that is produced when S-adenosylmethionine is used in methylation reactions. See, *e.g.*, Wang *et al.*, “Riboswitches that Sense S-adenosylhomocysteine and Activate Genes Involved in Coenzyme Recycling.” *Mol. Cell* 2008; **29** (6): 691–702; Edwards *et al.*, “Structural basis for recognition of S-adenosylhomocysteine by riboswitches.” *RNA* 2010; **16** (11): 2144–2155.

[0033] *SAM riboswitches* bind S-adenosyl methionine (SAM) to regulate methionine and SAM biosynthesis and transport. At least four SAM riboswitches are known: *SAM-I* (originally called *S-box*), *SAM-II*, the *S_{MK} box riboswitch* and *Sam-IV*. *SAM-I* is widespread in bacteria, but *SAM-II* is found only in alpha-, beta- and a few gamma-proteobacteria. The *S_{MK}* box riboswitch is believed to be found only in the order Lactobacillales. *SAM-IV* riboswitches have a similar ligand-binding core to that of *SAM-I* riboswitches, but in the context of a distinct scaffold. See, *e.g.*, Montange *et al.*, “Structure of the S-adenosyl methionine riboswitch regulatory mRNA element.” *Nature*. 2006; 441:1172 -1175; Winkler *et al.*, “An mRNA structure that controls gene expression by binding Sadenosylmethionine.” *Nat Struct Biol.* 2003; 10: 701 -707; Zasha *et al.*, “The aptamer core of *SAM-IV* riboswitches mimics the ligand-binding site of *SAM-I* riboswitches.” *RNA*. 2008; 14(5): 822–828 .

[0034] *Tetrahydrofolate riboswitches* bind tetrahydrofolate to regulate synthesis and transport genes. See, *e.g.*, Ames *et al.*, “A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate.” *Chem. Biol.* 2010; **17** (7): 681–5; Huang *et al.*, “Long-range pseudoknot interactions dictate the regulatory response in the tetrahydrofolate riboswitch.” *Proc. Natl. Acad. Sci. U.S.A.* 2011; **108** (36): 14801–6; Trausch *et al.*, “The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer.” *Structure*. 2011; **19** (10): 1413–23 .

[0035] *Theophylline riboswitch* was identified by SELEX and selectively binds the small molecule theophylline. The aptamer comprises a 15-nucleotide core motif that is required for theophylline binding. See, *e.g.*, Jenison *et al.*, “High-resolution molecular discrimination by RNA.” *Science*. 1994; 263:1425–1429; Zimmerman *et al.*, “Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer.” *RNA*.

2000; 6(5):659-67; Suess *et al.*, “A theophylline responsive riboswitch based on helix slipping controls gene expression *in vivo*.” *Nucleic Acids Res.* 2004; 32(4): 1610–1614.

See also, *e.g.*, Figure 1C-D.

[0036] *TPP riboswitches* (also THI-box) bind thiamin pyrophosphate (TPP) to regulate thiamin biosynthesis and transport, as well as transport of similar metabolites. It is believed to be the only riboswitch found so far in eukaryotes. See, *e.g.*, Edwards *et al.*, “Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition.” *Structure* 2006; **14** (9): 1459–68; Winkler *et al.*, “Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression.” *Nature*. 2002; **419** (6910): 952–956; Serganov *et al.*, “Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch.” *Nature*. 2006; **441** (7097): 1167–1171.

[0037] The term “Cas9” or “Cas9 nuclease” refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof. A Cas9 nuclease is also 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 (*e.g.*, 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 not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNA species. However, single guide RNAs (“sgRNA”, or simply “gRNA”) can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA molecule. See, *e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012).

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)).

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.

In some embodiments, proteins comprising Cas9 or fragments thereof proteins 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), 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, SEQ ID NO:1 (nucleotide); SEQ ID NO:2 (amino acid)).

```
ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGCGGTGATCACTGATG
ATTATAAGGTTCCGTCTAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCT
TATAGGGGCTCTTTATTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGA
AGGTATACACGTCGGAAGAACGTATTGTTATCTACAGGAGATTTTCAAATGAGATGGCGAAAGTAG
ATGATAGTTCTTCATCGACTTGAAGAGTCTTTGGTGGAAAGAACAGCATGAACGTCATCC
TATTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAATCTATCATCTGCGAAAA
AAATTGGCAGATTCTACTGATAAAGCGGATTGCGCTTAATCTATTGGCCTAGCGCATATGATTAAGT
TTCGTGGTCATTTTGATTGAGGGAGATTAAATCCTGATAATAGTGTGGACAAACTATTTATCCA
```

GTTGGTACAAATCTACAATTATTGAAGAAAACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCG
 ATTCTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTATTGCTCAGCTCCCCGGTGAGAAGA
 GAAATGGCTTGGGAATCTCATTGCTTGTATTGGATTGACCCCTAATTTAAATCAAATTTGA
 TTTGGCAGAAGATGCTAAATTACAGCTTCAAAAGATACTTACGATGATGATTAGATAATTATTGGCG
 CAAATTGGAGATCAATATGCTGATTGTTTGGCAGCTAAGAATTTCAGATGCTATTACTTCAG
 ATATCCTAAGAGTAAATAGTGAATACTAAGGCTCCCTATCAGCTTCAATGATTAAGCGTACGATGA
 ACATCATCAAGACTTGCACITTTAAAGCTTAGTTCGACAACAACCTCCAGAAAAGTATAAAGAAATC
 TTTTGATCAATCAAAAAACGGATATGCAGGTTATTGATGGGGAGCTAGCCAAGAAGAATTTCATA
 AATTTCACAAACCAATTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAACACTAAATCGTGAAGA
 TTTGCTGCGCAAGCAACGGACCTTGACAACGGCTTATTCCCCATCAAATTCACTGGGTGAGCTGCAT
 GCTATTTCAGAAGACAAGAAGACTTTATCCATTAAAGACAATCGTGAAGAGATTGAAAAATCT
 TGACTTTCACTTATTATGTTGGCCATTGGCGCGTGGCAATAGTCGTTTGATGGATGACTCG
 GAAGCTGAAGAACAAATTACCCCATGAAATTGAGAAGTTGTCGATAAGGTGCTTCAGCTCAATCA
 TTTATTGAACGCATGACAAACTTGATAAAAATCTCAAATGAAAAAGTACTACCAAAACATAGTTGC
 TTATGAGTATTTCAGGTTATAACGAAATTGACAAAGGTCAAATATGTTACTGAGGGAAATGCGAAAACC
 AGCATTCTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACC
 GTTAAGCAATTAAAGAAGATTATTCAAAAAAATAGAATGTTGATAGTGTGAATTTCAGGAGTTG
 AAGATAGATTAAATGCTTCATTAGCGCCTACCATGATTGCTAAAATTATTAAAGATAAAGATTTTT
 GGATAATGAAGAAAATGAAGATATCTAGAGGATATTGTTAACATTGACCTTATTGAAGATAGGGGG
 ATGATTGAGGAAAGACTTAAACATATGCTCACCTTTGATGATAAGGTGATGAAACAGCTTAAACGTC
 GCCGTTACTGGTGGGACGTTGTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAA
 AACAAATTAGATTGGAAATCAGATGGTTGCCATCGAATTTCAGCTTCTGAGCTGATGATGATGAT
 AGTTGACATTAAAGAAGATATTCAAAGACAGGTGCTGGACAAGGCCATAGTTACATGAACAGA
 TTGCTAACTTAGCTGGCAGTCTGCTATTAAAAAGGTATTACAGACTGTTAAAGTGTGATGAACT
 GGTCAAAGTAATGGGCATAAGCCAGAAAATCGTTATTGAAATGGCACGTGAAAATCAGACAACCTCAA
 AAGGGCCAGAAAATTGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAAATTAGGAAGTCAGA
 TTCTTAAAGAGCATCCTGTTGAAATACTCAATTGCAAATGAAAGCTCTATCTTCTATTACACAAAA
 TGGAAAGAGACATGTATGGACCAAGAATTAGATATTACGTTAAGTGTGATTATGATGTCGATCACATT
 GTTCCACAAAGTTCATTAAGACGATTCAATAGACAATAAGGTACTAACCGCTCTGATAAAAATCGTG
 GTAAATCGGATAACGTTCAAGTGAAGAAGTAGTCAGGAAAGATGAAAAGTATTGGAGACAACCTCTAA
 CGCCAAGTTAATCACTCAACGTAAGTTGATAATTAAACGAAAGCTGAACGTGGAGGTTGAGTGAACCT
 GATAAAAGCTGGTTTATCAAACGCCATTGGTGAACACTGCCAAATCACTAACGATGTGGCACAATT
 TGGATAGTCGATGAATAACTAAACGATGAAAATGATAAACTTATTGAGAGGTTAAAGTGTGATTACCTT
 AAAATCTAAATTAGTTCTGACTTCCGAAAAGATTCCAATTCTATCAAAGTACGTGAGGATTAACAATTAC
 CATCATGCCATGATGCGTATCTAAATGCCGTCGTTGAAACTGCTTGTATTAGAAATATCCAAAACCTG
 AATCGGAGTTGCTATGGTATTAAAGTTATGATGTTGTTAAAGTGTGCTAAGTCTGAGCAAGA
 AATAGGCAAAGCAACCGCAAATTTCTTACTCTAAATATCATGAACTTCTCAGGAAACAGAAATTACA
 CTTGCAAATGGAGAGATTGCAAACGCCCTCTAACGAAACTATGGGAAACTGGAGAAATTGCTGG
 ATAAAGGGCGAGATTGCCCAGTGCAGTAAAGTATTGCTATTGCCAAAGTCAATATGTCAAGAAA
 AGAAGTACAGACAGGGCGGATTCTCCAAGGGAGTCATTTACCAAAAGAAATTCCGACAAGCTTATTGCT
 CGTAAAAAAAGACTGGGATCCAAAAAAATGGGGTTTGTAGTCCAACGGTAGCTTATTGAGCTCCTAG
 TGGTTGCTAAGGTGGAAAAGGGAAATCGAAGAAGTTAAATCCGTTAAAGAGTTACTAGGGATCACAAT
 TATGGAAAGAAGTTCCATTGAAAAAAATCCGATTGACTTTTAAAGCTAAAGGATATAAGGAAGTTAAA
 AAAGACTTAATCATTAAACTACCTAAATATAGTCTTTGAGTTAGAAAACGGTCGAAACGGATGCTGG

CTAGTGC CGGAGAATTACAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTATATT
 AGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAACAATTGTTGTGGAGCAG
 CATAAGCATTATTAGATGAGATTATTGAGCAAATCAGTGAATTTCTAAGCGTGTATTTAGCAGATG
 CCAATTAGATAAAGTCTTAGTGCATATAACAAACATAGAGACAAACATACGTGAACAAGCAGAAAA
 TATTATTCATTTACGGTACGAATCTGGAGCTCCGCTGCTTTAAATATTTGATACAAACAATT
 GATCGTAAACGATATACTACGTACAAAAGAAGTTAGATGCCACTCTTATCCATCAATCCATCACTGGTC
 TTTATGAAACACGCATTGATTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:1)

MDKKYSIGLDIGTNSVGVWAVITDDYKVP SKFKVLGNTRHSIKNLIGALLFGSGETAETRLKRTARR
 RYTRRKNRICYLQEIFSNE MAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
 KLA DSTD KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQIYNQLFEENPINASRVDAKA
 ILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSGLTPNFKSNFDLA EDAK LQLSKDTYDDLDNLLA
 QIGDQYADLFLA AKNLSDAI LLSDI LRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
 AILRROEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAMWTRKSETITPWNEEVVDKGASAQS
 FIERMTNFDKNL PNEKVL PKHSLLY EYFTVYNELTKV KYVTEGMRKPAFLSGE QKKAI VDLLF KTNRKVT
 VKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLKII KDKDFLDNEENEDILEDIVLT LTFEDRG
 MIEERLKIYAHLFDDKVMQKLKRRRYTGWGRLSRKLLINGIRDQSGKTILDFLKSDGFANRNFMQLIHDD
 SLTFKEDIQKAQVSGQGHSLEQIANLAGSPAIIKGILQTVKIVDELVKVGMHKPENIVIEMARENQTTQ
 KGQKNSRERMKRRIE GIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVQELDINRLSDYDVDHI
 VPQSF IKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLTQRKFDNLTKAERGGLSEL
 DKAGFIKRQLVETRQITKHVAQILD SRMNTK YDENDKL I REVK VITL KSKL VSDFRKDFQFYKVREINNY
 HHAHDAYLNAAVG TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
 LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
 RKKDWDPKKYGGFDSPVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
 KDLI I KLPKYS LFELENGRKRMLASAGELQKGNE LALPSKYVNFLY LASHYEKLKGSPEDNEQKQLFVEQ
 HKHYLDEIIEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
 DRKRYTSTKEVLDATL IHQSITGLYETRIDLSQLGGD (SEQ ID NO:2)

[0038] The terms “conjugating,” “conjugated,” and “conjugation” refer to an association of two entities, for example, of two molecules such as two proteins, two domains (*e.g.*, a binding domain and a cleavage domain), or a protein and an agent, *e.g.*, a protein binding domain and a small molecule. In some aspects, the association is between a protein (*e.g.*, RNA-programmable nuclease) and a nucleic acid (*e.g.*, a guide RNA). The association can be, for example, via a direct or indirect (*e.g.*, via a linker) covalent linkage. In some embodiments, the association is covalent. In some embodiments, two molecules are conjugated via a linker connecting both molecules. For example, in some embodiments where two portions of RNA are conjugated to each other, *e.g.*, an aptamer (or nucleic acid sensing domain) and a gRNA, the two RNAs may be conjugated via a polynucleotide linker, *e.g.*, a nucleotide sequence connecting the 3' end of one RNA to the 5' end of the other RNA. In some embodiments, the linker comprises 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, or at least 30 nucleotides.

[0039] The term “consensus sequence,” as used herein in the context of nucleic acid sequences, refers to a calculated sequence representing the most frequent nucleotide residues found at each position in a plurality of similar sequences. Typically, a consensus sequence is determined by sequence alignment in which similar sequences are compared to each other and similar sequence motifs are calculated.

[0040] 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 a nuclease may refer to the amount of the nuclease that is sufficient to induce cleavage of a desired target site specifically bound and cleaved by the nuclease, preferably with minimal or no off-target cleavage. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a nuclease, a hybrid protein, a fusion protein, a protein dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors as, for example, on the desired biological response, the specific allele, genome, target site, cell, or tissue being targeted, and the agent being used.

[0041] The term “engineered,” as used herein refers to a nucleic acid molecule, a protein molecule, complex, substance, or entity that has been designed, produced, prepared, synthesized, and/or manufactured by a human. Accordingly, an engineered product is a product that does not occur in nature.

[0042] The term “linker,” as used herein, refers to a chemical group or a molecule linking two adjacent molecules or moieties, *e.g.*, an aptamer (or nucleic acid sensing domain) and a gRNA. 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 is a nucleotide linker. In some embodiments, the nucleotide linker comprises 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, or at least 30 nucleotides. 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 an organic molecule, group, polymer, or chemical moiety.

[0043] The term “mutation,” as used herein, refers to a substitution of a residue within a sequence, *e.g.*, a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Methods for

making the amino acid substitutions (mutations) provided herein are known in the art and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

[0044] The term “nuclease,” as used herein, refers to an agent, for example, a protein 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 bonds 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 is a RNA-guided (*i.e.*, RNA-programmable) nuclease, which complexes with (*e.g.*, binds with) an RNA (*e.g.*, a guide RNA, “gRNA”) having a sequence that complements a target site, thereby providing the sequence specificity of the nuclease. In some embodiments, a nuclease recognizes a single stranded target site. 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 sites 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 5' 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 also, in some cases, specifically binds to a target site, 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 fused to create nucleases binding specific target sites, are well known to those of skill in the art. For example, the binding domain of RNA-programmable nucleases (*e.g.*, Cas9), or a Cas9 protein having an inactive DNA cleavage domain, can be used as a binding domain (*e.g.*, that binds a gRNA to direct binding to a target site) 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 target site.

[0045] The terms “nucleic acid” and “nucleic acid molecule,” as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, “nucleic acid” refers to individual nucleic acid residues (*e.g.* nucleotides and/or nucleosides). In some embodiments, “nucleic acid” refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms “oligonucleotide” and “polynucleotide” can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, “nucleic acid” encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, an mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a recombinant DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms “nucleic acid,” “DNA,” “RNA,” and/or similar terms include nucleic acid analogs, *i.e.* analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using recombinant expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified

bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5' to 3' direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (*e.g.*, 2'-fluororibose, ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5'-*N*-phosphoramidite linkages).

[0046] The term “pharmaceutical composition,” as used herein, refers to a composition that can be administrated to a subject in the context of treatment of a disease or disorder. In some embodiments, a pharmaceutical composition comprises an active ingredient, *e.g.*, a nuclease or a nucleic acid encoding a nuclease, and a pharmaceutically acceptable excipient.

[0047] The term “proliferative disease,” as used herein, refers to any disease in which cell or tissue homeostasis is disturbed in that a cell or cell population exhibits an abnormally elevated proliferation rate. Proliferative diseases include hyperproliferative diseases, such as pre-neoplastic hyperplastic conditions and neoplastic diseases. Neoplastic diseases are characterized by an abnormal proliferation of cells and include both benign and malignant neoplasias. Malignant neoplasia is also referred to as cancer.

[0048] The terms “protein,” “peptide,” and “polypeptide” are used interchangeably herein and refer 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. The term “fusion protein” as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an “amino-terminal fusion protein” or a “carboxy-terminal fusion protein,” respectively. A protein may comprise different domains, for example, a nucleic acid binding domain (e.g., the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage 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. In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, e.g., RNA. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

[0049] The term “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 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. 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 at least 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. In some embodiments, domain (2) is the “sgRNA Backbone as depicted in any of the Figures 1-4. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA, and comprises a stem-loop structure. For example, in some embodiments, domain (2) is homologous to a tracrRNA as depicted in

Figure 1E of Jinek *et al.*, *Science* 337:816-821(2012).

In some embodiments, domain 2 is at least 90%, at least 95%, at least 98%, or at least 99% identical to the “sgRNA backbone” of any one of Figures 1-4 or the tracrRNA as described by Jinek *et al.*, *Science* 337:816-821(2012). In some embodiments, a gRNA comprises two or more of domains (1) and (2), and may be referred to as an “extended gRNA.” For example, an extended gRNA will bind two or more Cas9 proteins and bind a target nucleic acid at two or more distinct regions. 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. The sequence of a gRNA that binds a target nucleic acid can comprise any sequence that complements a region of the target and is suitable for a nuclease:RNA complex to bind. In some embodiments, the RNA-programmable nuclease is the (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).

[0050] 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)).

[0051] The terms “small molecule” and “organic compound” are used interchangeably herein and refer to molecules, whether naturally-occurring or artificially created (*e.g.*, via chemical synthesis) that have a relatively low molecular weight. Typically, an organic compound contains carbon. An organic compound may contain multiple carbon-carbon bonds, stereocenters, and other functional groups (*e.g.*, amines, hydroxyl, carbonyls, or heterocyclic rings). In some embodiments, organic compounds are monomeric and have a molecular weight of less than about 1500 g/mol. In certain embodiments, the molecular weight of the small molecule is less than about 1000 g/mol or less than about 500 g/mol. In certain embodiments, the small molecule is a drug, for example, a drug that has already been deemed safe and effective for use in humans or animals by the appropriate governmental agency or regulatory body. In certain embodiments, the small molecule is known to bind an aptamer. In some embodiments, the organic compound is an antibiotic drug, for example, an anticancer antibiotic such as dynemicin, neocarzinostatin, calicheamicin, esperamicin, bleomycin, or a derivative thereof.

[0052] The term “subject,” as used herein, refers to an individual organism, for example, an individual mammal. In some embodiments, the subject is a human. 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 sheep, a goat, a cattle, 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 a research animal. In some embodiments, the subject is genetically engineered, *e.g.*, a genetically engineered non-human subject. The subject may be of either sex and at any stage of development.

[0053] The terms “target nucleic acid,” and “target genome,” as used herein in the context of nucleases, refer to a nucleic acid molecule or a genome, respectively, that comprises at least one target site of a given nuclease.

[0054] The term “target site,” used herein interchangeably with the term “nuclease target site,” refers to a sequence within a nucleic acid molecule that is bound and cleaved by a nuclease. A target site may be single-stranded or double-stranded. In the context of RNA-guided (*e.g.*, RNA-programmable) nucleases (*e.g.*, a protein dimer comprising a Cas9 gRNA binding domain and an active Cas9 DNA cleavage domain), a target site typically comprises

a nucleotide sequence that is complementary to a gRNA of the RNA-programmable nuclease, and a protospacer adjacent motif (PAM) at the 3' end adjacent to the gRNA-complementary sequence. For the RNA-guided nuclease Cas9, the target site may be, in some embodiments, 20 base pairs plus a 3 base pair PAM (*e.g.*, NNN, wherein N represents any nucleotide).

Typically, the first nucleotide of a PAM can be any nucleotide, while the two downstream nucleotides are specified depending on the specific RNA-guided nuclease. Exemplary target sites for RNA-guided nucleases, such as Cas9, are known to those of skill in the art and include, without limitation, NNG, NGN, NAG, and NGG, wherein N represents any nucleotide. In addition, Cas9 nucleases from different species (*e.g.*, *S. thermophilus* instead of *S. pyogenes*) recognize a PAM that comprises the sequence: NGGNG. Additional PAM sequences are known, including, but not limited to, NNAGAAW and NAAR (see, *e.g.*, Esvelt and Wang, *Molecular Systems Biology*, 9:641 (2013)).

For example, the target site of an RNA-guided nuclease, such as, *e.g.*, Cas9, may comprise the structure [N_z]-[PAM], where each N is, independently, any nucleotide, and z is an integer between 1 and 50. In some embodiments, z is 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 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, 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, z is 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, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50. In some embodiments, Z is 20. In some embodiments, “target site” may also refer to a sequence within a nucleic acid molecule that is bound but not cleaved by a nuclease.

[0055] 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.

[0056] The term “vector” refers to a polynucleotide comprising one or more recombinant polynucleotides of the present invention, *e.g.*, those encoding a gRNA provided herein and/or a Cas9 protein. Vectors include, but are not limited to, plasmids, viral vectors, cosmids, artificial chromosomes, and phagemids. The vector is one which is able to replicate in a host cell, and which is further characterized by one or more endonuclease restriction sites at which the vector may be cut and into which a desired nucleic acid sequence may be inserted. Vectors may contain one or more marker sequences suitable for use in the identification and/or selection of cells which have or have not been transformed or genetically modified with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics (*e.g.*, kanamycin, ampicillin) or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (*e.g.*, β -galactosidase, alkaline phosphatase or luciferase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies, or plaques. Any vector suitable for the transformation of a host cell, (*e.g.*, *E. coli*, mammalian cells such as CHO cell, insect cells, *etc.*) as embraced by the present invention, for example vectors belonging to the pUC series, pGEM series, pET series, pBAD series, pTET series, or pGEX series. In some embodiments, the vector is suitable for transforming a host cell for recombinant protein production. Methods for selecting and engineering vectors and host cells for expressing gRNAs and/or proteins (*e.g.*, those provided herein), transforming cells, and expressing/purifying recombinant proteins are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

[0057] Site-specific nucleases are powerful tools for targeted genome modification *in vitro* and *in vivo*. Some site-specific nucleases can theoretically achieve a level of specificity for a target cleavage site that would allow one to target a single unique site in a genome for cleavage without affecting any other genomic site. It has been reported that nuclease cleavage in living cells triggers a DNA repair mechanism that frequently results in a modification of the cleaved and repaired genomic sequence, for example, via homologous recombination or non-homologous end-joining. Accordingly, the targeted cleavage of a

specific unique sequence within a genome opens up new avenues for gene targeting and gene modification in living cells, including cells that are hard to manipulate with conventional gene targeting methods, such as many human somatic or embryonic stem cells. Nuclease-mediated modification of disease-related sequences, *e.g.*, the CCR-5 allele in HIV/AIDS patients, or of genes necessary for tumor neovascularization, can be used in the clinical context, and two site specific nucleases are currently in clinical trials (Perez, E.E. *et al.*, "Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases." *Nature Biotechnology*. 26, 808-816 (2008); ClinicalTrials.gov identifiers: NCT00842634, NCT01044654, NCT01252641, NCT01082926). Other diseases that can be treated using site-specific nucleases include, for example, diseases associated with triplet expansion (*e.g.*, Huntington's disease, myotonic dystrophy, spinocerebellar ataxias, *etc.*) cystic fibrosis (by targeting the *CFTR* gene), cancer, autoimmune diseases, and viral infections.

[0058] One important problem with site-specific nuclease-mediated modification is off-target nuclease effects, *e.g.*, the cleavage of genomic sequences that differ from the intended target sequence by one or more nucleotides. Undesired side effects of off-target cleavage range from insertion into unwanted loci during a gene targeting event to severe complications in a clinical scenario. Off-target cleavage of sequences encoding essential gene functions or tumor suppressor genes by an endonuclease administered to a subject may result in disease or even death of the subject. Accordingly, it is desirable to employ new strategies in designing nucleases having the greatest chance of minimizing off-target effects.

[0059] The methods and compositions of the present disclosure represent, in some aspects, an improvement over previous methods and compositions by providing means to control the temporal activity and/or increase the specificity of RNA-guided nucleases. For example, RNA-guided nucleases known in the art, both naturally occurring and those engineered, typically bind to and cleave DNA upon forming a complex with an RNA (*e.g.*, a gRNA) that complements the target. Aspects of the present invention relate to the recognition that having temporal control over the timing of the binding of an RNA-guided nuclease:RNA complex to its target will decrease the likelihood of off-target effects by minimizing or controlling the amount of time a complex is able to bind to and cleave the target. Additionally, engineering gRNAs that only bind the target site to be cleaved, for example, using gRNAs with extended target recognition domains that block binding in the absence of the target, improves the specificity of RNA-guided nucleases and decreases the chances of off-target effects.

[0060] The strategies, methods, compositions, kits, and systems provided herein can be used to control the activity and/or improve the specificity of any RNA-guided nuclease (e.g., Cas9). Suitable nucleases for use with modified gRNA as described herein will be apparent to those of skill in the art based on this disclosure.

[0061] In certain embodiments, the strategies, methods, compositions, kits, and systems provided herein are utilized to control the timing of RNA-guided (e.g., RNA-programmable) nuclease activity. Whereas typical RNA-guided nucleases recognize and cleave a target sequence upon forming a nuclease:RNA complex, the modified gRNAs provided herein allow for control over target binding and cleavage. Other aspects provide gRNAs engineered to bind a target site only when the intended target site is present, thereby improving the specificity of a RNA-guided nuclease. While Cas9:gRNA complexes have been successfully used to modify both cells (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); Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013)) and organisms (Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*. 31, 227-229 (2013)), a study using Cas9:guide RNA complexes to modify zebrafish embryos observed toxicity (e.g., off-target effects) at a rate similar to that of ZFNs and TALENs (Hwang, W.Y. *et al.* *Nature Biotechnology*. 31, 227-229 (2013)). Accordingly, aspects of the present disclosure aim to reduce the chances for Cas9 off-target effects using novel gRNA platforms that control for the timing of target binding and cleavage and/or improve the specificity of RNA-guided nucleases.

[0062] While of particular relevance to DNA and DNA-cleaving nucleases such as Cas9, the inventive concepts, methods, compositions, strategies, kits, and systems provided herein are not limited in this respect, but can be applied to any nucleic acid:nuclease system utilizing nucleic acid templates such as RNA to direct binding to a target nucleic acid.

Modified guide RNAs (gRNAs)

[0063] Some aspects of this disclosure provide gRNAs engineered to have both an “on” and “off” state. In some aspects then, the gRNAs may collectively be referred to as “switchable gRNAs.” For example, a switchable gRNA is said to be in an “off” state when the gRNA is in a structural state that prevents binding of the gRNA to a target nucleic acid. In some aspects, a gRNA in an “off” state can bind to its cognate RNA-guided nuclease (e.g., Cas9), however, the nuclease:gRNA complex (when the gRNA is in an “off” state) is unable

to bind the target nucleic acid to mediate cleavage. In other aspects, a gRNA that is in an “off” state is unable to bind its target sequence or an RNA-guided nuclease, such as Cas9. Conversely, a switchable gRNA is said to be in an “on” state when the gRNA is in a structural state that allows binding of the gRNA to a target nucleic acid (e.g., as a complex with an RNA-guided nuclease such as Cas9). Some embodiments of this disclosure provide complexes comprising an inventive gRNA associated with an RNA-guided nuclease, such as Cas9, and methods of their use. Some embodiments of this disclosure provide nucleic acids encoding such gRNAs and/or RNA-guided nucleases (e.g., Cas9). Some embodiments of this disclosure provide expression constructs comprising such encoding nucleic acids.

Aptamer based gRNAs

[0064] In one embodiment, gRNAs are provided that comprise an aptamer. See, e.g., Figure 1. For example, in some embodiments, a gRNA is linked to an aptamer via a nucleotide linker, as described herein. Aptamers are typically RNA or peptide based molecules that bind a specific ligand with affinities, for example, that rival antibody:antigen interactions. In some embodiments, an aptamer binds its ligand with a K_d between about 1 nM – 10 μ M, between about 1 nM – 1 μ M, between about 1 nM – 500 nM, or between about 1 nM – 100 nM. With RNA-based aptamers, for example, those found in riboswitches of mRNAs, binding of the ligand to the aptamer domain results in conformational changes that control expression (e.g., translation) of the mRNA. RNA aptamers have been successfully cloned and adapted to other molecules, for example, to control gene expression, or have been engineered/selected for particular ligands using SELEX (See, e.g., Dixon *et al.*, “Reengineering orthogonally selective riboswitches.” *PNAS* 2010; **107** (7): 2830–2835; Suess *et al.*, “A theophylline responsive riboswitch based on helix slipping controls gene expression *in vivo*.” *Nucleic Acids Res.* 2004; **32**(4): 1610–1614; Ellington and Szostak, “In vitro selection of RNA molecules that bind specific ligands.” *Nature*. 1990; **346**:818–822; Tuerk and Gold, “Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.” *Science*. 1990; **249**:505–510; Burke and Gold, “RNA aptamers to the adenosine moiety of S-adenosyl methionine: structural inferences from variations on a theme and the reproducibility of SELEX.” *Nucleic Acids Res.* 1997; **25**(10):2020-4; Ulrich *et al.*, “DNA and RNA aptamers: from tools for basic research towards therapeutic applications.” *Comb Chem High Throughput Screen.* 2006; **9**(8):619-32; Svobodová *et al.*, “Comparison of different methods for generation of single-stranded DNA for SELEX processes. *Anal Bioanal Chem.* 2012; **404**:835–842).

Ligands that bind aptamers include, but are not limited to, small molecules, metabolites, carbohydrates, proteins, peptides, or nucleic acids. As shown in Figure 1, gRNAs linked to aptamers exist in an “off” state in the absence of the specific ligand that binds the aptamer. Typically, the “off” state is mediated by a structural feature that prevents all or a part of the sequence of the gRNA that hybridizes to the target nucleic acid from being free to hybridize with the target nucleic acid. For example, in some aspects, the gRNA comprising an aptamer is designed such that part of the aptamer sequence hybridizes to part or all of the gRNA sequence that hybridizes to the target. The sequence of the gRNA that binds a target (e.g., depicted as “Guide to Cut Target” in Figure 1C, D, referred to herein as the “guide” sequence) can be engineered using methods known in the art to include any sequence that targets any desired nucleic acid target, and is therefore not limited to the sequence(s) depicted in the Figures, which are exemplary. Similarly, any suitable aptamer can be linked 5' or 3' to the gRNA sequences, and be modified to include nucleotides that will hybridize to a particular guide sequence in a gRNA using methods routine in the art. In some embodiments, the aptamers linked to any gRNA provided herein are RNA aptamers, as described herein. In some embodiments, the RNA aptamer is derived from (e.g., cloned from) a riboswitch. Any riboswitch may be used in the RNA aptamer. Exemplary riboswitches include, but are not limited to, theophylline riboswitches, thiamine pyrophosphate (TPP) riboswitches, adenosine cobalamin (AdoCbl) riboswitches, S-adenosyl methionine (SAM) riboswitches, SAH riboswitches, flavin mononucleotide (FMN) riboswitches, tetrahydrofolate riboswitches, lysine riboswitches, glycine riboswitches, purine riboswitches, GlnS riboswitches, and pre-queosine₁ (PreQ1) riboswitches. In some embodiments, the RNA aptamer is derived from a theophylline riboswitch. In some embodiments, the aptamer derived from the theophylline riboswitch comprises SEQ ID NO:3. In some embodiments, the underlined, bold portion of SEQ ID NO:3 can be modified such that any nucleotide therein is replaced with any other nucleotide, and/or can be modified by adding or deleting 1 or more nucleotides. For example, the underlined, bold portion can be modified so as to include a sequence that hybridizes to part or all of the sequence of the gRNA that hybridizes to the target nucleic acid. See, e.g., Figure 1C. In some embodiments, the RNA aptamer is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical to SEQ ID NO:3.

5' – GGUGAUACC**AGCAUCGUCUUGAUG**CCCUGGGCAGCACC – 3' (SEQ ID NO:3)
[0065] In some embodiments, the aptamer is non-naturally occurring (e.g., is not found in nature). For example, in some embodiments, the aptamer is engineered or selected

from a library using SELEX. In some embodiments, the aptamer comprises at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 175, at least 200, at least 250, or at least 300 nucleotides. In some embodiments, the aptamer comprises 20-200, 20-150, 20-100, or 20-80 nucleotides. In some embodiments, the gRNA portion of provided RNAs (e.g., RNAs comprising a gRNA linked to an aptamer) comprises at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 175, or at least 200 nucleotides. In some embodiments, the gRNA portion comprises 60-150, 60-100, or 60-80 nucleotides.

mRNA-sensing gRNAs

[0066] According to another embodiment, gRNAs are provided that bind a target nucleic acid under certain conditions (e.g., in the presence of a metabolite, small molecule, nucleic acid, etc.). In some embodiments, the gRNAs are structurally precluded (e.g., are in an “off” state) from binding (e.g., hybridizing to) a target unless another molecule binds to (e.g., hybridizes to) the gRNA, resulting in a structural rearrangement corresponding to an “on” state. In some embodiments, the binding of a particular transcript (e.g., mRNA) to the gRNA turns the gRNA from an “off” state to an “on” state. See, e.g., Figure 2. Such gRNAs are referred to as “mRNA-sensing” gRNAs. For example, in some aspects, gRNAs are provided that comprise: (i) a region that hybridizes a region of a target nucleic acid (e.g., the “guide” sequence); (ii) another region that partially or completely hybridizes to the sequence of region (i) (e.g., the “guide block” sequence); and (iii) a region that hybridizes to a region of a transcript (mRNA) (e.g., the “transcript sensor”). In some embodiments, each region (e.g., i-iii) comprises at least 5, at least 10, at least 15, at least 20, or at least 25 nucleotides. In some embodiments, the gRNA forms a stem-loop structure. In some embodiments the stem comprises the sequence of region (i) hybridized to part or all of the sequence of region (ii), and the loop is formed by part or all of the sequence of region (iii). In some embodiments, regions (ii) and (iii) are both either 5' or 3' to region (i). See, e.g., Figure 2A vs. Figure 2C. The sequence of the gRNA that binds a target (e.g., the “guide” sequence) can be engineered using methods known in the art to include any sequence that targets any desired nucleic acid target, and is therefore not limited to the sequence(s) depicted in the Figures, which are exemplary. Similarly, region (iii) (e.g., the transcript sensor) can be engineered to comprise any sequence that hybridizes an mRNA of interest using methods routine in the art. Likewise, region (ii) can be engineered to comprise a sequence that

hybridizes to part or all of the “guide” sequence using methods routine in the art. For example, in some aspects, the mRNA is one that when expressed in a cell, the genomic modification of a target nucleic acid (*e.g.*, gene) is desired. Thus, in the absence of the mRNA, the gRNA, when delivered to (or expressed in) a cell, remains in the “off” state. When the mRNA is present (*e.g.*, expressed), it binds the transcript sensor of the gRNA, resulting in unfolding of the stem-loop structure that prevented hybridization of the “guide” sequence to the target nucleic acid, thereby turning the gRNA “on”. See, *e.g.*, Figures 2B and 2D. Provided gRNAs in an “on” state are able to associate with and guide RNA-guided nucleases (*e.g.*, Cas9 proteins) to bind a target nucleic acid.

Extended-DNA-sensing (xDNA-sensing) gRNAs

[0067] According to another embodiment, modified gRNAs are provided that remain in an “off” state unless the gRNA hybridizes to a target nucleic acid at least two distinct regions. See, *e.g.*, Figure 3. Such gRNAs provide improved specificity to an RNA-guided nuclease (*e.g.*, Cas9) because they effectively extend the recognition sequence of a particular gRNA/target interaction. Such gRNAs are referred to as “xDNA-sensing” gRNAs (“x” being short for “extended” DNA recognition). For example, gRNAs are provided that comprise: (i) a region that hybridizes a region of a target nucleic acid (*e.g.*, the “guide” sequence); (ii) another region that partially or completely hybridizes to the sequence of region (i) (*e.g.*, the “guide block”); and (iii) a region that hybridizes to another region of the target nucleic acid (*e.g.*, the “xDNA sensor”). In some embodiments, the xDNA sensor must first bind the target nucleic acid before the guide sequence is able to bind the target. In some embodiments, the xDNA sensor binds the same strand of the target that the guide sequence binds. In some embodiments, the xDNA sensor and guide sequence bind different strands of the target nucleic acid. In some embodiments, the sequences of regions (i) and (ii) comprise at least 5, at least 10, at least 15, at least 20, or at least 25 nucleotides. In some embodiments, the sequence of region (iii) comprise at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, or at least 100 nucleotides. In some embodiments, the gRNA forms a stem-loop structure. For example, in some embodiments, the stem comprises the sequence of region (i) hybridized to part or all of the sequence of region (ii), and the loop is formed by part or all of the sequence of region (iii). In some embodiments, regions (i) and (iii) comprise sequences that are adjacent in the gRNA. In some embodiments, regions (ii) and (iii) are both either 5' or 3' to region (i). See, *e.g.*, Figure 3A vs. Figure 3C. In some embodiments, region (ii) is located between regions (i) and (iii). The

sequence of the gRNA that binds a target (*e.g.*, the “guide” sequence) can be engineered using methods known in the art to include any sequence that targets any desired nucleic acid target, and is therefore not limited to the sequence(s) depicted in the Figures, which are exemplary. Similarly, region (iii) (*e.g.*, the xDNA sensor) can be engineered to comprise any sequence that hybridizes another region (*e.g.*, a different region than that targeted by the “guide” sequence) of the target nucleic acid using methods routine in the art. Likewise, region (ii) can be engineered to include a sequence that hybridizes to part or all of the “guide” sequence using methods routine in the art. Thus, in the absence of the correct target nucleic acid (*e.g.*, a target comprising both regions to which the gRNA was designed to hybridize), the gRNA, when delivered to (or expressed in) a cell, remains in the “off” state. Without wishing to be bound by any particular theory, it is expected that when the gRNA (*e.g.*, when associated with Cas9) comes into contact with the target nucleic acid, the xDNA sensor hybridizes to the target, which in turn unravels the stem-loop structure that blocks the “guide” sequence, turning the gRNA “on”. If it is the correct target nucleic acid, the guide sequence will then hybridize to the target, and optionally the complex will cleave the target nucleic acid. See, *e.g.*, Figures 3B and 3D.

Complexes

[0068] In some embodiments, complexes comprising any of the RNAs/gRNAs described herein (*e.g.*, RNAs comprising a gRNA linked to an aptamer, gRNAs that sense mRNAs, or gRNAs comprising xDNA sensors) are provided. In some aspects, a complex comprising a provided RNA/gRNA associated with an RNA-guided nuclease is provided. In some embodiments, the RNA-guided nuclease is Cas9, a variant of Cas9, or a fragment of Cas9, for example as described herein. In some embodiments, the RNA-guided nuclease is any form of the Cas9 protein as provided in U.S. Provisional Patent Application, U.S.S.N. 61/874,609, filed September 6, 2013, entitled “Cas9 Variants And Uses Thereof,” and U.S. Provisional Patent Application, U.S.S.N. 61/874,746, filed September 6, 2013, entitled “Delivery System For Functional Nucleases,” .

[0069] In some embodiments, the complex further comprises a ligand, *e.g.*, a ligand that binds the aptamer of the RNA associated with the RNA-guided nuclease, as described herein. In some embodiments, the complex (*e.g.*, comprising a provided RNA (gRNA):ligand:Cas9 protein) binds to and optionally cleaves a target nucleic acid. In some

aspects, a complex comprising a “sensing” gRNA (*e.g.*, mRNA or xDNA) and Cas9 binds to and optionally cleaves a target nucleic acid.

Pharmaceutical compositions

[0070] In some embodiments, any of the gRNAs described herein are provided as part of a pharmaceutical composition. In some embodiments, the pharmaceutical composition further comprises an RNA-guided nuclease (*e.g.*, Cas9) that forms a complex with an inventive gRNA. For example, some embodiments provide pharmaceutical compositions comprising a gRNA and an RNA-guided nuclease as provided herein, or a nucleic acid encoding such gRNAs and/or nuclease, and a pharmaceutically acceptable excipient. Pharmaceutical compositions may optionally comprise one or more additional therapeutically active substances.

[0071] In some embodiments, compositions provided herein are administered to a subject, for example, to a human subject, in order to effect a targeted genomic modification within the subject. In some embodiments, cells are obtained from the subject and contacted with a provided gRNA associated with an RNA-guided nuclease or nucleic acid(s) encoding such *ex vivo*. In some embodiments, cells removed from a subject and contacted *ex vivo* with an inventive gRNA:nuclease complex are re-introduced into the subject, optionally after the desired genomic modification has been effected or detected in the cells. Methods of delivering pharmaceutical compositions comprising nucleases are known, and are described, for example, in U.S. Pat. Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824.

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 or organisms 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, domesticated animals, pets, and 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.

[0072] 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(s) 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.

[0073] 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)

discloses various excipients used in formulating pharmaceutical compositions and known techniques for the preparation thereof. See also PCT application PCT/US2010/055131 (Publication number WO2011053982 A8, filed Nov. 2, 2010), for additional suitable methods, reagents, excipients and solvents for producing pharmaceutical compositions comprising a nuclease. 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 disclosure.

[0074] 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.

Methods for site-specific nucleic acid cleavage

[0075] In another embodiment of this disclosure, methods for site-specific nucleic acid (e.g., DNA) cleavage are provided. In some embodiments, the methods comprise contacting a DNA with any of the Cas9:RNA complexes described herein. For example, in some embodiments, the method comprises contacting a DNA with a complex comprising: (i) gRNA linked to an aptamer as described herein, wherein the gRNA comprises a sequence that binds to a portion of the DNA; (ii) a ligand bound to the aptamer of the gRNA; and (iii) an RNA-guided nuclease (e.g., a Cas9 protein), under suitable conditions for the Cas9 nuclease to cleave DNA.

[0076] In some embodiments, methods for inducing site-specific DNA cleavage in a cell are provided. In some embodiments, the method comprises: (a) contacting a cell or expressing within a cell a gRNA comprising an aptamer as described herein, wherein the gRNA comprises a sequence capable of binding to a DNA target sequence; (b) contacting a cell or expressing within a cell an RNA-guided nuclease (e.g., a Cas9 protein); and (c) contacting the cell with a specific ligand that binds the aptamer of the gRNA, resulting in the formation of a gRNA:ligand:Cas9 complex that cleaves the DNA target. In some embodiments, the method comprises: (a) contacting the cell with a complex comprising a Cas9 protein and a gRNA comprising an aptamer as described herein, wherein the gRNA comprises a sequence capable of binding to a DNA target sequence; and (b) contacting the cell with a specific ligand that binds the aptamer of the gRNA, resulting in the formation of a gRNA:ligand:Cas9 complex that cleaves the DNA target. In some embodiments, steps (a) and (b) are performed simultaneously. In some embodiments, steps (a) and (b) are performed sequentially. Thus in some embodiments, wherein the cell is contacted with the ligand subsequent to the cell being contacted with the complex, control of cleavage is achieved because cleavage only occurs once the ligand has been delivered to the cell. In some embodiments of these methods, the ligand is not delivered to the cell, but is produced internally by the cell, for example as part of a physiological or pathophysiological process.

[0077] In some embodiments, methods for site-specific DNA cleavage are provided that utilize mRNA-sensing gRNAs as described herein. For example, in some embodiments,

the method comprises contacting a DNA with a complex comprising an RNA-guided nuclease (e.g., a Cas9 protein) and an mRNA-sensing gRNA, wherein the gRNA comprises: (i) a region that hybridizes a region of a target nucleic acid; (ii) another region that partially or completely hybridizes to the sequence of region (i); and (iii) a region that hybridizes to a region of a transcript (mRNA). In some embodiments, cleavage occurs after the sequence in region (iii) hybridizes to the mRNA.

[0078] In other embodiments, methods for site-specific DNA cleavage are provided that utilize xDNA-sensing gRNAs as described herein. For example, in some embodiments, the method comprises contacting a DNA with a complex comprising an RNA-guided nuclease (e.g., a Cas9 protein) and an xDNA-sensing gRNA, wherein the gRNA comprises: (i) a region that hybridizes a region of a target nucleic acid; (ii) another region that partially or completely hybridizes to the sequence of region (i); and (iii) a region that hybridizes to another region of the target nucleic acid. In some embodiments, cleavage occurs after the sequence in region (iii) hybridizes to the region of the target nucleic acid that is not targeted by the “guide” sequence.

[0079] In some embodiments, any of the methods provided herein can be performed on DNA in a cell. For example, in some embodiments the DNA contacted by any RNA/gRNA-comprising complex provided herein is in a eukaryotic cell. In some embodiments, the eukaryotic cell is in an individual. In some embodiments, the individual is a human. In some embodiments, any of the methods provided herein are performed *in vitro*. In some embodiments, any of the methods provided herein are performed *in vivo*.

Polynucleotides, Vectors, Cells, Kits

[0080] In another embodiment of this disclosure, polynucleotides are provided that encode any of the gRNAs (and optionally any Cas9 protein) described herein. For example, polynucleotides encoding any of the gRNAs and/or Cas9 proteins described herein are provided, *e.g.*, for recombinant expression and purification of inventive gRNAs, or complexes comprising such, *e.g.*, complexes comprising inventive gRNAs and an RNA-guided nuclease (e.g., a Cas9 protein). In some embodiments, provided polynucleotides comprises one or more sequences encoding a gRNA, alone or in combination with a sequence encoding any of the Cas9 proteins described herein.

[0081] In some embodiments, vectors encoding any of the gRNAs (and optionally any Cas9 protein) described herein are provided, *e.g.*, for recombinant expression and purification of inventive gRNAs, or complexes comprising inventive gRNAs and an RNA-

guided nuclease (*e.g.*, a Cas9 protein). In some embodiments, the vector comprises or is engineered to include a polynucleotide, *e.g.*, those described herein. In some embodiments, the vector comprises one or more sequences encoding a gRNA and/or any Cas9 protein (*e.g.*, as described herein). Typically, the vector comprises a sequence encoding an inventive gRNA operably linked to a promoter, such that the gRNA is expressed in a host cell.

[0082] In some embodiments, cells are provided for recombinant expression and purification of any of the gRNAs (and optionally any Cas9 protein) described herein. The cells include any cell suitable for recombinant RNA expression and optionally protein expression, for example, cells comprising a genetic construct expressing or capable of expressing an inventive gRNA (*e.g.*, cells that have been transformed with one or more vectors described herein, or cells having genomic modifications that express an inventive gRNA and optionally any Cas9 protein provided herein from an allele that has been incorporated in the cell's genome). Methods for transforming cells, genetically modifying cells, and expressing genes and proteins in such cells are well known in the art, and include those provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)) and Friedman and Rossi, *Gene Transfer: Delivery and Expression of DNA and RNA, A Laboratory Manual* (1st ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2006)).

[0083] Some aspects of this disclosure provide kits comprising any of the inventive gRNAs or complexes provided herein and optionally any Cas9 protein described herein. In some embodiments, the kit comprises any of the polynucleotides encoding a provided gRNA, and optionally any Cas9 protein. In some embodiments, the kit comprises a vector for recombinant expression of any inventive gRNA and optionally any Cas9 protein. In some embodiments, the kit comprises a cell that comprises a genetic construct for expressing any of the inventive gRNAs, complexes, and optionally any Cas9 protein provided herein. In some embodiments, the kit comprises an excipient and instructions for contacting any of the inventive compositions with the excipient to generate a composition suitable for contacting a nucleic acid with *e.g.*, a complex of an inventive gRNA and a RNA-guided nuclease, such as Cas9. In some embodiments, the composition is suitable for contacting a nucleic acid within a genome. In some embodiments, the composition is suitable for delivering an inventive composition (*e.g.*, a gRNA, complexes thereof with Cas9) to a cell. In some embodiments, the composition is suitable for delivering an inventive composition (*e.g.*, a gRNA, complexes

thereof with Cas9) to a subject. In some embodiments, the excipient is a pharmaceutically acceptable excipient.

EQUIVALENTS AND SCOPE

[0084] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above description, but rather is as set forth in the appended claims.

[0085] 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 also 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.

[0086] Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, *etc.*, from one or more of the claims or from relevant portions of the description 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. Furthermore, where the claims recite a composition, it is to be understood that methods of using the composition for any of the purposes disclosed herein are included, and methods of making the composition according to any of the methods of making disclosed herein or other methods known in the art are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[0087] Where elements are presented as lists, *e.g.*, in Markush group format, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It is also noted that the term “comprising” is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, steps, *etc.*, certain embodiments of the invention or aspects of

the invention consist, or consist essentially of, such elements, features, steps, *etc.* For purposes of simplicity those embodiments have not been specifically set forth *in haec verba* herein. Thus for each embodiment of the invention that comprises one or more elements, features, steps, *etc.*, the invention also provides embodiments that consist or consist essentially of those elements, features, steps, *etc.*

[0088] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value 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. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

[0089] In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

[0090]

CLAIMS

What is claimed is:

1. A complex comprising: (i) a single-guide RNA (sgRNA) comprising an aptamer, wherein the sgRNA does not hybridize to a target nucleic acid in the absence of a specific ligand bound to the aptamer; and (ii) a Cas9 protein.
2. The complex of claim 1, wherein the aptamer is bound by the ligand.
3. The complex of claim 2, wherein the ligand is a small molecule, a metabolite, a peptide, or a nucleic acid.
4. The complex of claim 3, wherein the sgRNA:ligand:Cas9 complex binds the target nucleic acid.
5. An sgRNA comprising an aptamer, wherein the sgRNA does not hybridize to a target nucleic acid in the absence of a specific ligand bound to the aptamer.
6. The sgRNA of claim 5, wherein the sgRNA does not bind Cas9 in the absence of the ligand bound to the aptamer.
7. The sgRNA of claim 5 or 6, wherein the sgRNA binds Cas9 when the aptamer is bound by the specific ligand.
8. The sgRNA of claim 5, wherein the sgRNA binds Cas9 in the absence or presence of the specific ligand, but binds to the target nucleic acid only in the presence of the specific ligand.
9. The sgRNA of any one of claims 5-8, wherein the ligand is a small molecule, a metabolite, a peptide, or a nucleic acid.
10. The sgRNA of any one of claims 5-9, wherein the aptamer is an RNA aptamer.

11. The sgRNA of claim 10, wherein the RNA aptamer is derived from a riboswitch.
12. The sgRNA of claim 11, wherein the riboswitch from which the aptamer is derived is a theophylline riboswitch, a thiamine pyrophosphate (TPP) riboswitch, an adenosine cobalamin (AdoCbl) riboswitch, an S-adenosyl methionine (SAM) riboswitch, an S-adenosyl homocysteine (SAH) riboswitch, a flavin mononucleotide (FMN) riboswitch, a tetrahydrofolate riboswitch, a lysine riboswitch, a glycine riboswitch, a purine riboswitch, a glucosamine-6-phosphate synthase (GlmS) riboswitch, or a pre-queosine₁ (PreQ1) riboswitch.
13. The sgRNA of claim 12, wherein the aptamer is derived from a theophylline riboswitch and comprises SEQ ID NO:3.
14. The sgRNA of any one of claims 5-13, wherein the aptamer is non-naturally occurring.
15. The sgRNA of any one of claims 5-14, wherein the aptamer is engineered to bind the specific ligand using a systematic evolution of ligands by exponential enrichment (SELEX) platform.
16. The sgRNA of any one of claims 5-15, wherein the non-aptamer portion of the sgRNA comprises at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, or at least 150 nucleotides, and the aptamer comprises at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 175, at least 200, at least 250, or at least 300 nucleotides.
17. A polynucleotide encoding the sgRNA of any one of claims 5-16.
18. A vector comprising the polynucleotide of claim 17.

19. A vector for recombinant expression comprising a polynucleotide encoding the sgRNA of any one of claims 5-16.
20. A cell comprising a genetic construct for expressing the sgRNA of any one of claims 5-16.
21. The cell of claim 20, wherein the cell expresses a Cas9 protein.
22. An *in vitro* method for site-specific DNA cleavage comprising contacting a DNA with a complex comprising (i) the sgRNA of any one of claims 5-16, wherein the sgRNA comprises a sequence that binds to a portion of the DNA, (ii) the specific ligand bound to the aptamer of the sgRNA, and (iii) a Cas9 protein, under conditions in which the Cas9 protein cleaves the DNA.
23. The method of claim 22, wherein the DNA is in a cell.
24. The method of claim 23, wherein the cell is a eukaryotic cell.
25. An *in vitro* method for inducing site-specific DNA cleavage in a cell comprising:
 - (a) contacting the cell or expressing within the cell the sgRNA of any one of claims 5-16, wherein the sgRNA comprises a sequence that binds to a portion of the DNA;
 - (b) contacting the cell or expressing within the cell a Cas9 protein; and
 - (c) contacting the cell with the specific ligand that binds the aptamer of the sgRNA, resulting in the formation of an sgRNA:ligand:Cas9 complex that cleaves the DNA.
26. An *in vitro* method for inducing site-specific DNA cleavage in a cell comprising (a) contacting the cell with a complex comprising a Cas9 protein and the sgRNA of any one of claims 5-16, wherein the sgRNA comprises a sequence that binds to a portion of the DNA, and (b) contacting the cell with the specific ligand that binds the aptamer of the sgRNA, resulting in the formation of a sgRNA:ligand:Cas9 complex that cleaves the DNA.

27. The method of claim 25 or 26, wherein steps (a) and (b) are performed simultaneously or sequentially in any order.

28. The method of any one of claims 25-27, wherein the cell is a eukaryotic cell.

29. A kit comprising the sgRNA of any one of claims 5-16 and either an excipient, instructions, or both.

30. A kit comprising a polynucleotide encoding the sgRNA of any one of claims 5-16 and either an excipient, instructions, or both.

31. A kit comprising a vector for recombinant expression, wherein the vector comprises a polynucleotide encoding the sgRNA of any one of claims 5-16 and either an excipient, instructions, or both.

32. A kit comprising a cell that comprises a genetic construct for expressing the sgRNA of any one of claims 5-16 and either an excipient, instructions, or both, and optionally a Cas9 protein.

33. The kit of any one of claims 29-32, further comprising one or more Cas9 proteins and/or constructs encoding a Cas9 protein.

34. Use of a complex for site-specific DNA cleavage, the complex comprising:

(i) the sgRNA of any one of claims 5-16, wherein the sgRNA comprises a sequence that binds to a portion of DNA;
(ii) the specific ligand bound to the aptamer of the sgRNA; and
(iii) a Cas9 protein,

wherein the complex is for contact with the DNA under conditions in which the Cas9 protein cleaves the DNA.

35. The use of claim 34, wherein the DNA is in a cell.

36. The use of claim 35, wherein the cell is a eukaryotic cell.
37. The use of claim 36, wherein the eukaryotic cell is in an individual.
38. The use of claim 37, wherein the individual is a human.
39. Use of the sgRNA of any one of claims 5-16, a Cas9 protein, and the specific ligand that binds the aptamer of the sgRNA for inducing site-specific DNA cleavage in a cell, wherein:
 - the sgRNA is for contact with the cell or for expression within the cell, and comprises a sequence that binds to a portion of the DNA;
 - the Cas9 protein is for contact with the cell or for expression within the cell; and
 - the specific ligand that binds the aptamer of the sgRNA is for contact with the cell, and results in the formation of an sgRNA:ligand:Cas9 complex that cleaves the DNA.
40. The use as defined in claim 39, wherein the sgRNA for contact with the cell or expression within the cell and the Cas9 protein for contact with the cell or for expression within the cell may be used simultaneously or sequentially in any order prior to contact with the specific ligand that binds the aptamer of the sgRNA.
41. Use of a complex comprising the sgRNA of any one of claims 5-16 and a Cas9 protein, and the specific ligand that binds the aptamer of the sgRNA for inducing site-specific DNA cleavage in a cell, wherein:
 - the sgRNA comprises a sequence that binds to a portion of the DNA;
 - the complex is for contact with the cell; and
 - the specific ligand that binds the aptamer of the sgRNA is for contact with the cell, and results in the formation of an sgRNA:ligand:Cas9 complex that cleaves the DNA.
42. The use as defined in claim 41, wherein the complex for contact with the cell and the specific ligand that binds the aptamer of the sgRNA for contact with the cell may be used simultaneously or sequentially in any order.

43. The use of any one of claims 39-42, wherein the cell is a eukaryotic cell.
44. The use of claim 43, wherein the eukaryotic cell is *in vitro*.
45. The use of claim 43, wherein the eukaryotic cell is *in vivo*.
46. The use of claim 45, wherein the eukaryotic cell is in an individual.
47. The use of claim 46, wherein the individual is a human.
48. The complex of any one of claims 1-4, wherein the aptamer is an RNA aptamer.
49. The complex of claim 48, wherein the RNA aptamer is derived from a riboswitch.
50. The complex of claim 49, wherein the riboswitch from which the aptamer is derived is a theophylline riboswitch, a thiamine pyrophosphate (TPP) riboswitch, an adenosine cobalamin (AdoCbl) riboswitch, an S-adenosyl methionine (SAM) riboswitch, an S-adenosyl homocysteine (SAH riboswitch), a flavin mononucleotide (FMN) riboswitch, a tetrahydrofolate riboswitch, a lysine riboswitch, a glycine riboswitch, a purine riboswitch, a glucosamine-6-phosphate synthase (GlmS) riboswitch, or a pre-queosine1 (PreQ1) riboswitch.
51. The complex of claim 50, wherein the aptamer is derived from a theophylline riboswitch and comprises SEQ ID NO:3.
52. The complex of any one of claims 1-4 and 48-51, wherein the aptamer is non-naturally occurring.
53. The complex of any one of claims 1-4 and 48-52, wherein the aptamer is engineered to bind the specific ligand using a systematic evolution of ligands by exponential enrichment (SELEX) platform.

54. The complex of any one of claims 1-4 and 48-53, wherein the non-aptamer portion of the sgRNA comprises at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, or at least 150 nucleotides, and the aptamer comprises at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 175, at least 200, at least 250, or at least 300 nucleotides.

55. A cell comprising the complex of any one of claims 1-4 and 48-54.

56. The cell of claim 55, wherein the cell expresses a Cas9 protein.

57. A cell comprising the polynucleotide of claim 17.

58. A cell comprising the vector of claim 18 or 19.

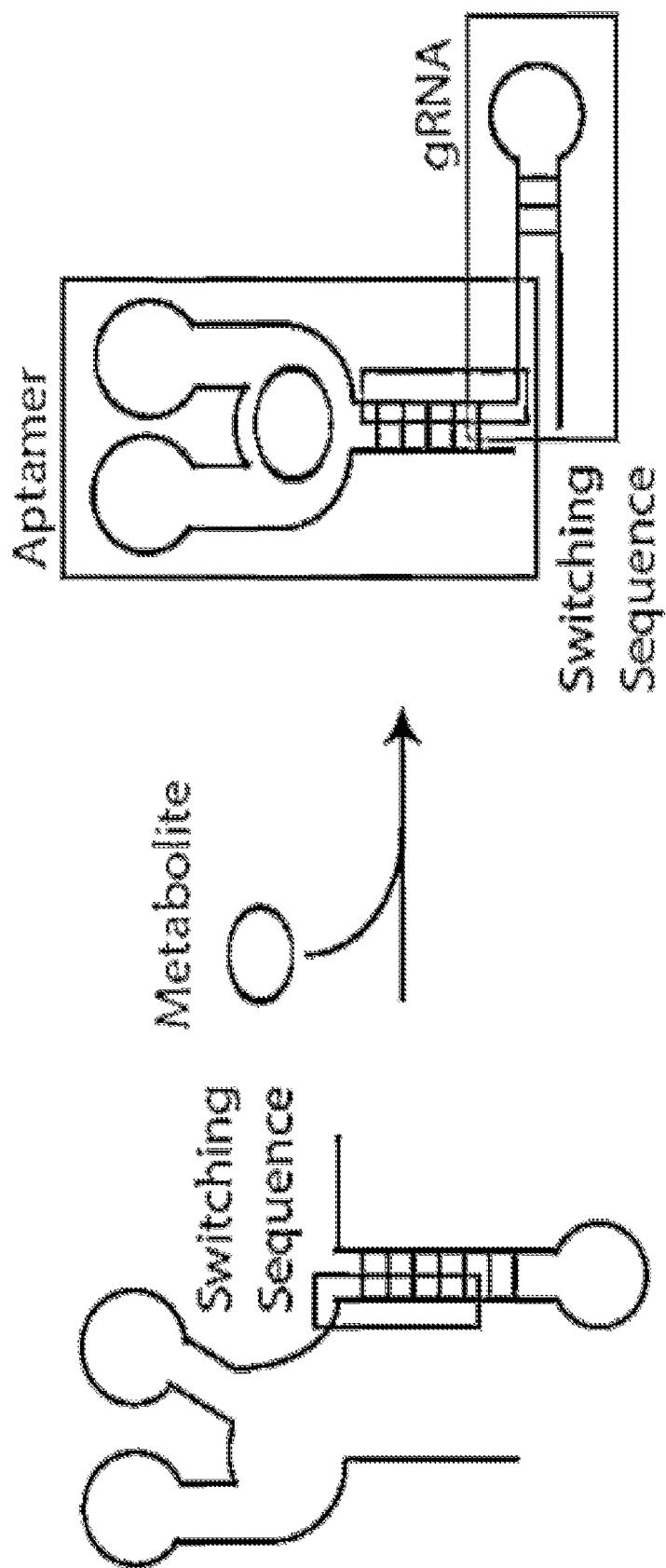


FIG. 1A

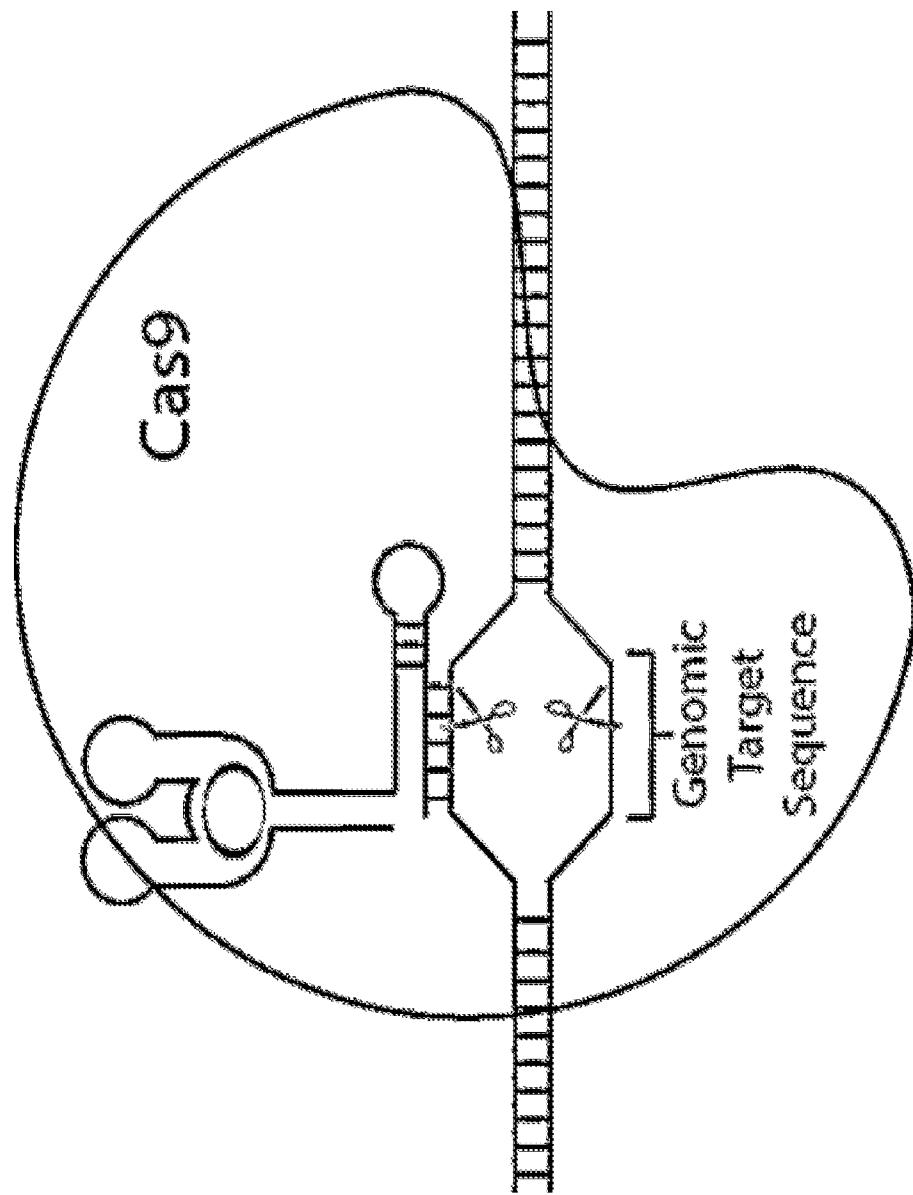
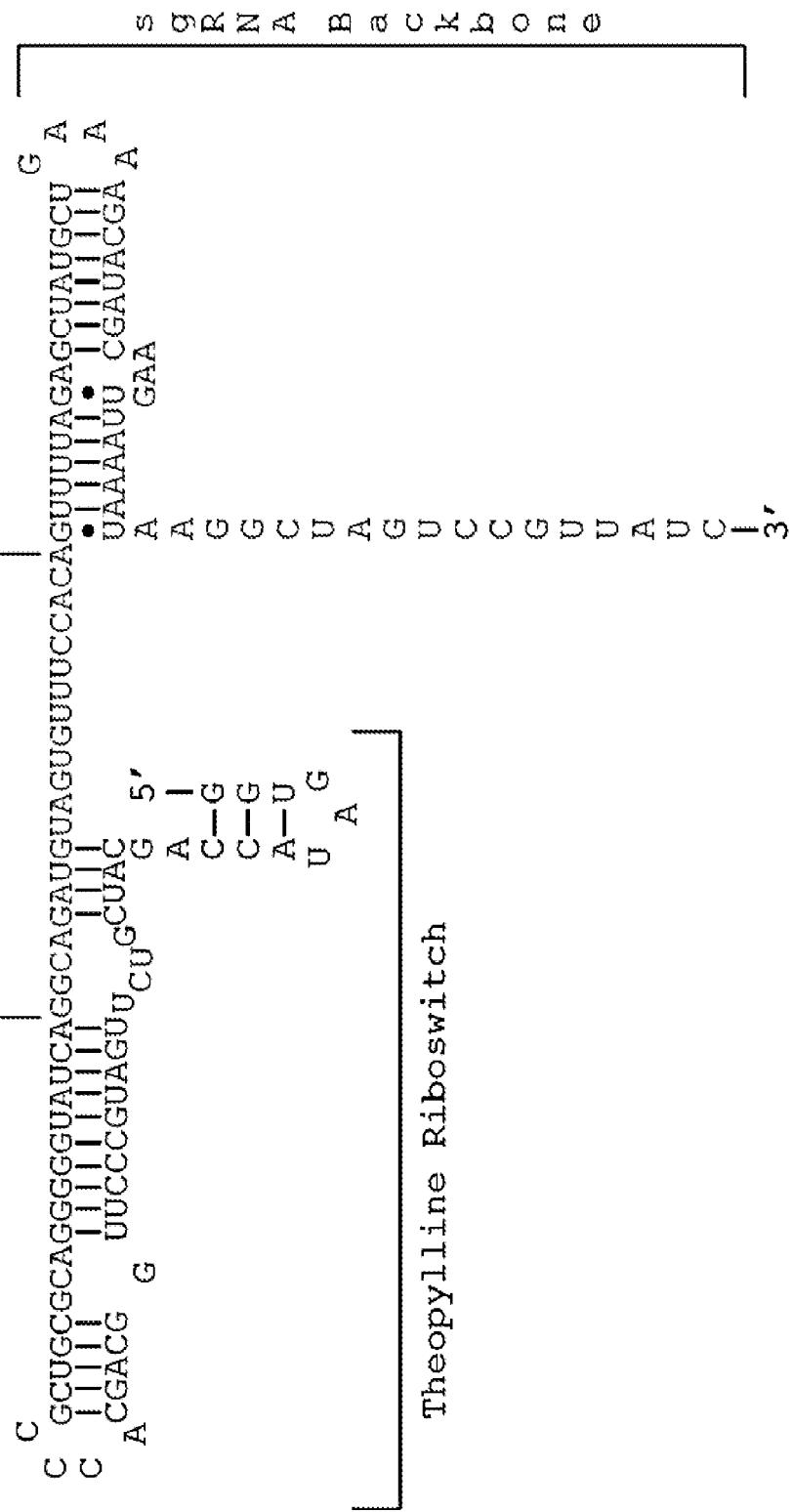


FIG. 1B

Theophylline sensitive gRNA, Off State

5' - — — GGCAGATGTAGTGTTCACAGGG — — — 3'
3' - — — CCGTCTACATCACAAAGGTGTCCC — — — 5'

Guide to Cut Target



Theophylline Riboswitch

FIG. 1C

Theophylline sensitive gRNA, On State

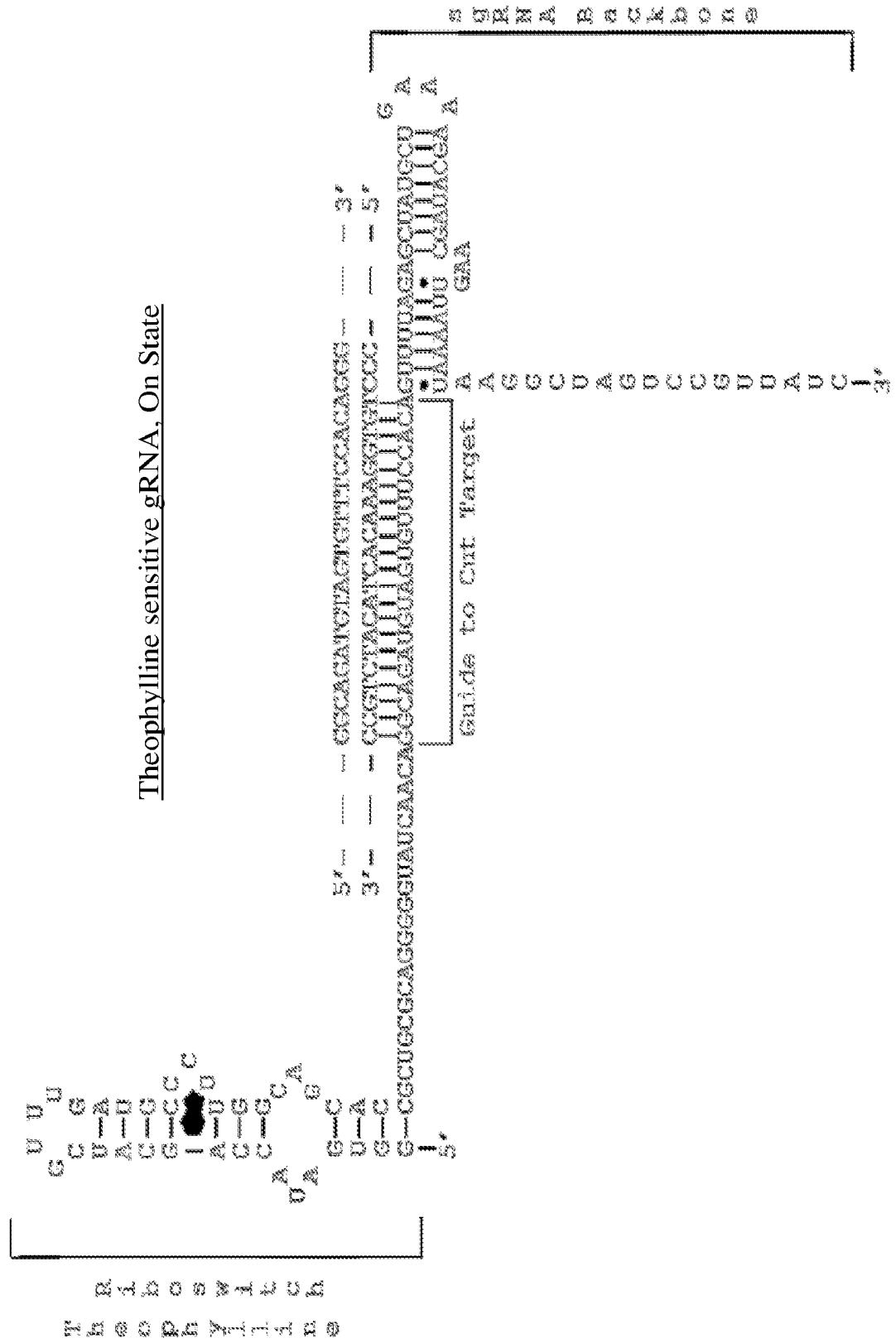


FIG. 1D

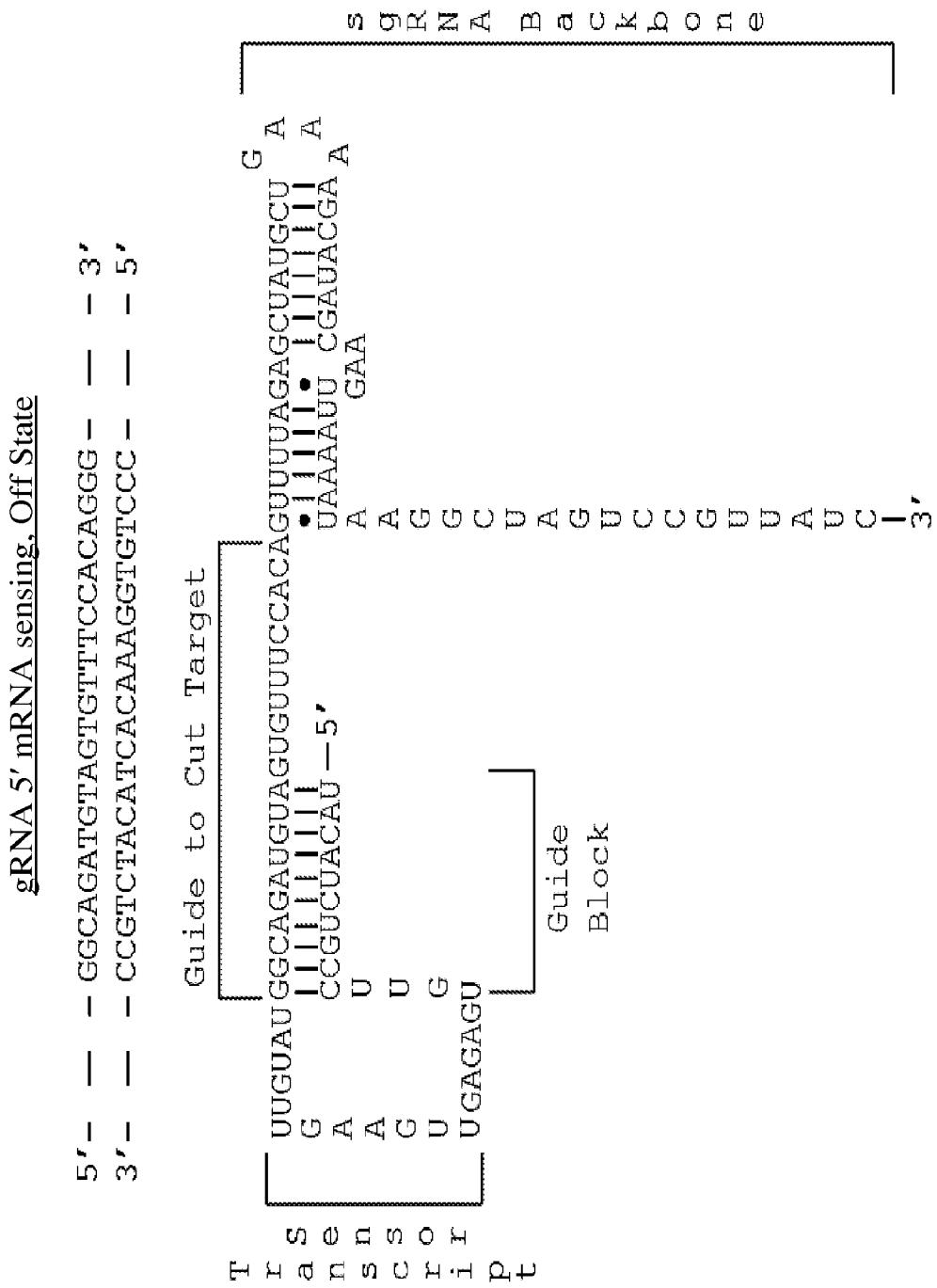


FIG. 2A

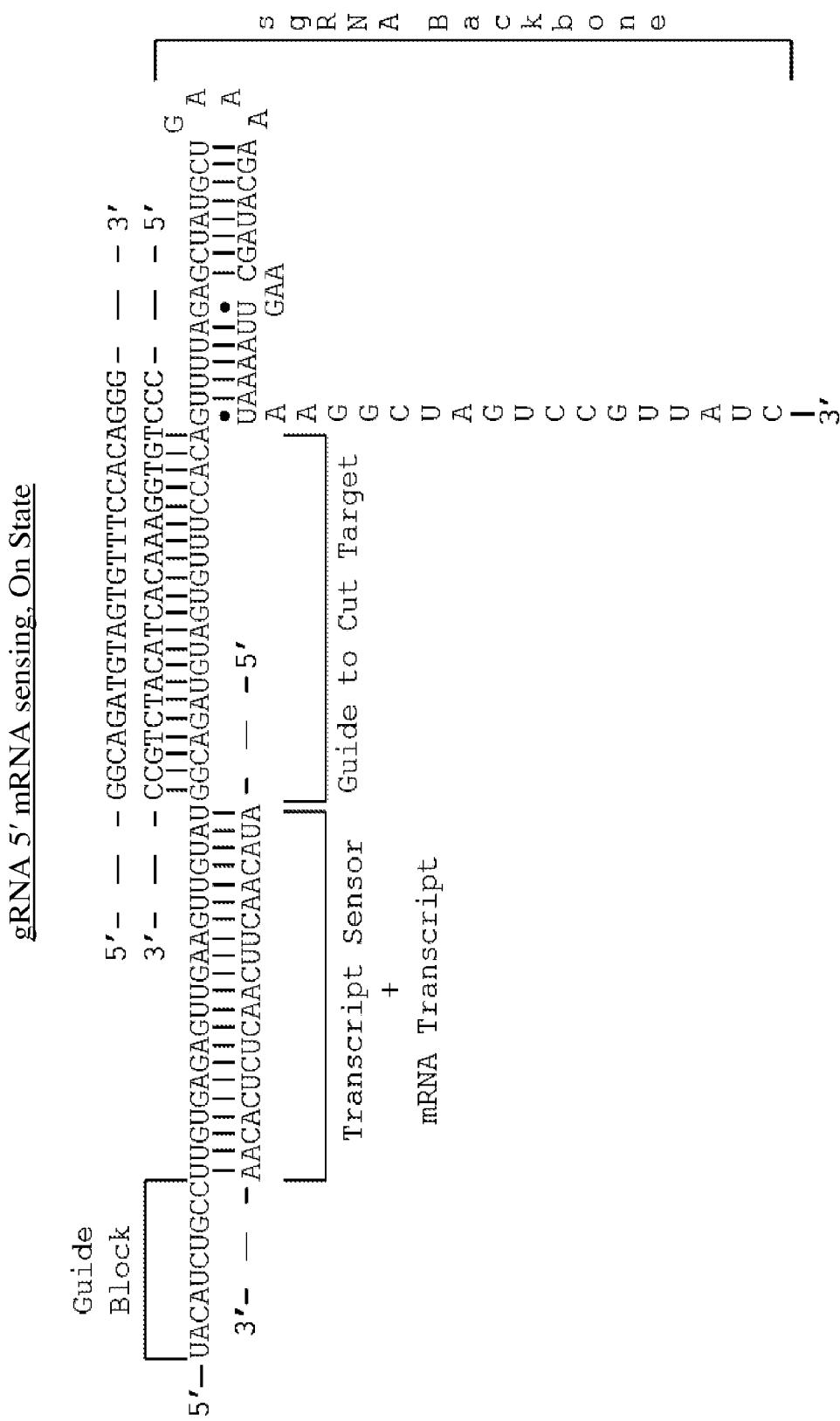


FIG. 2B

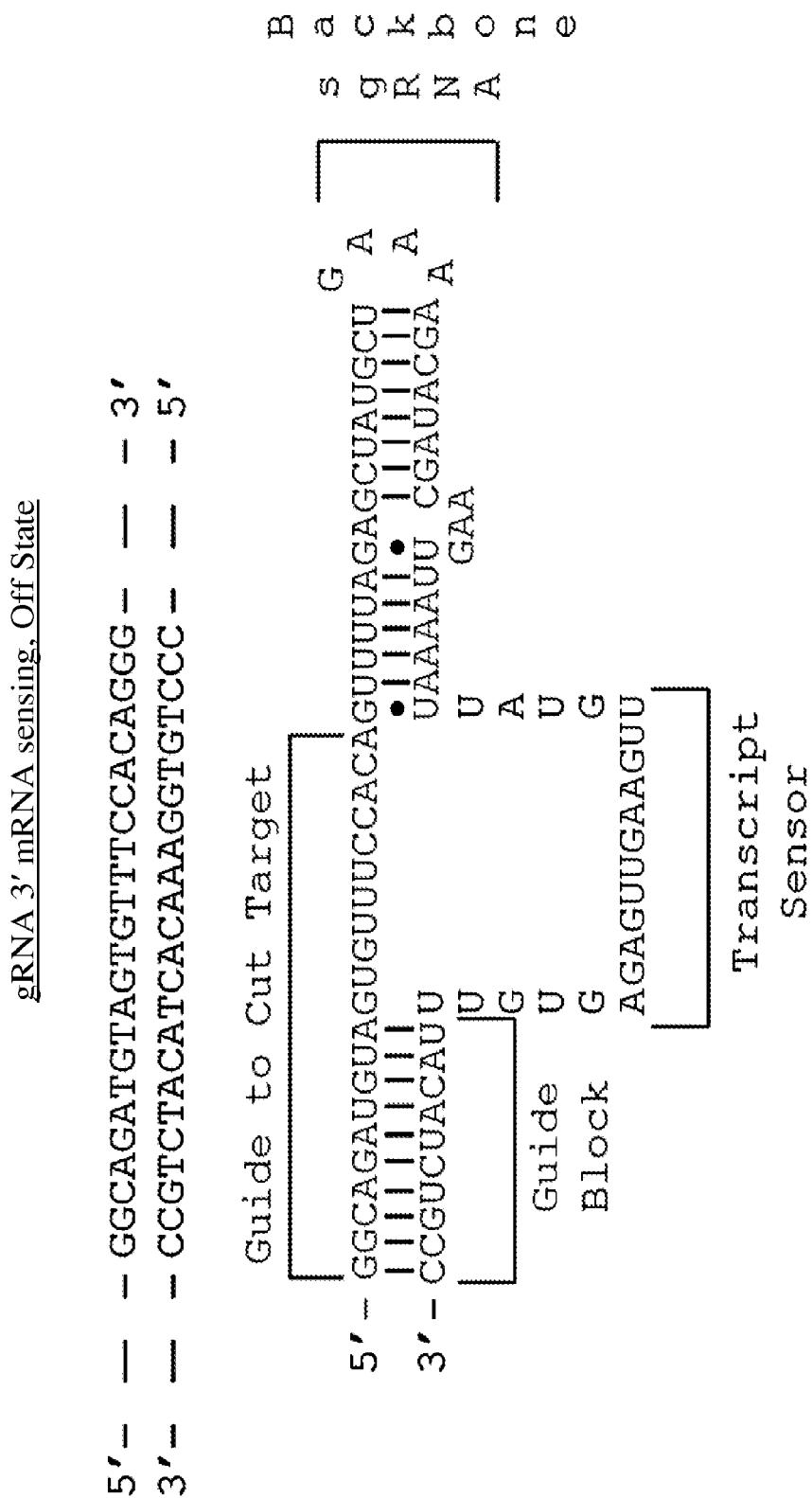


FIG. 2C

gRNA 3' mRNA sensing, On State

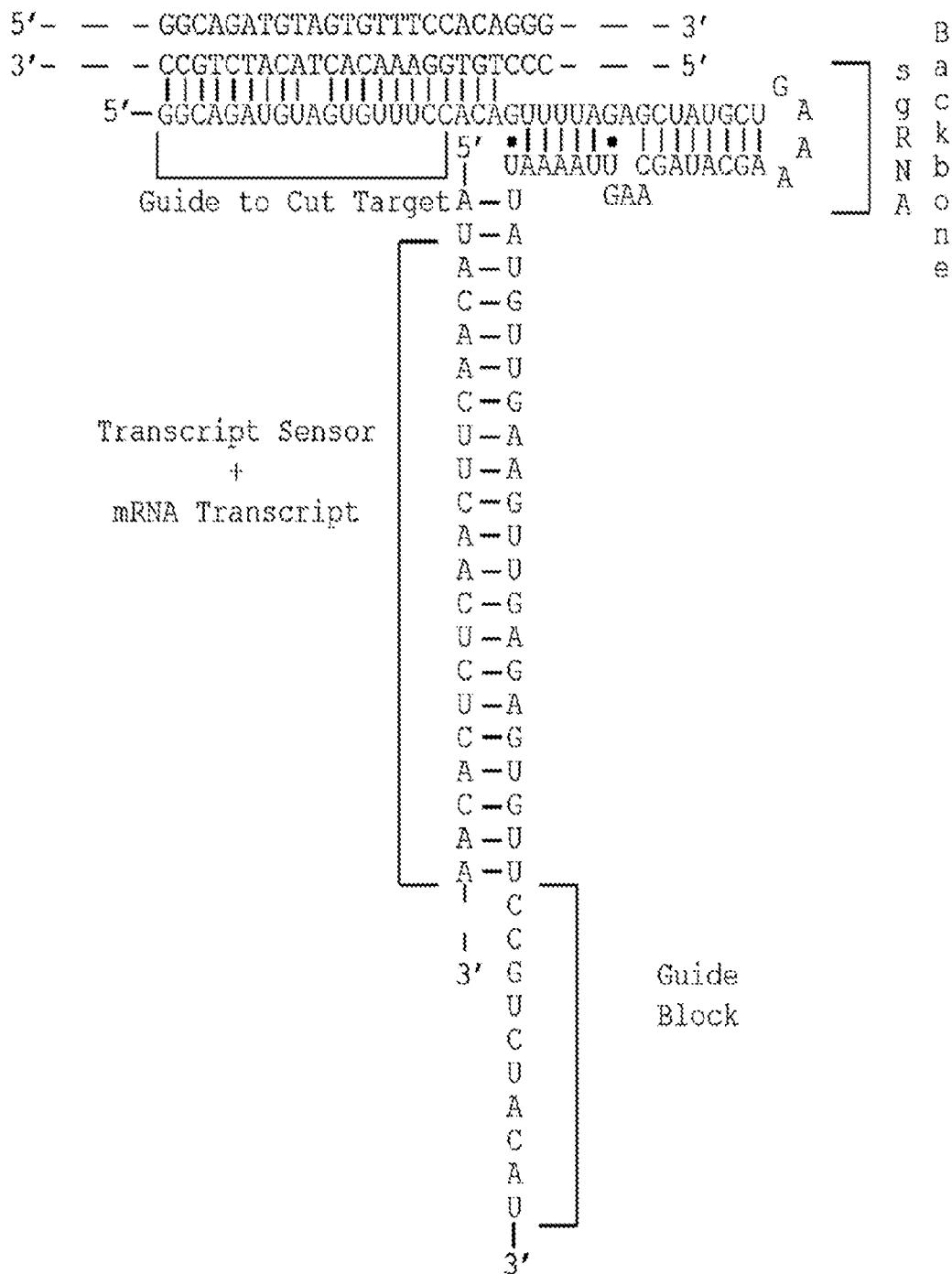


FIG. 2D

gRNA 5'xDNA sensing, Off State

5' - — — GGCAGATGTAGTGTTCACAGGG — — — 3'
3' - — — CCGTCTACATCACAAAGGTGCCC — — — 5'

Guide to Cut Target

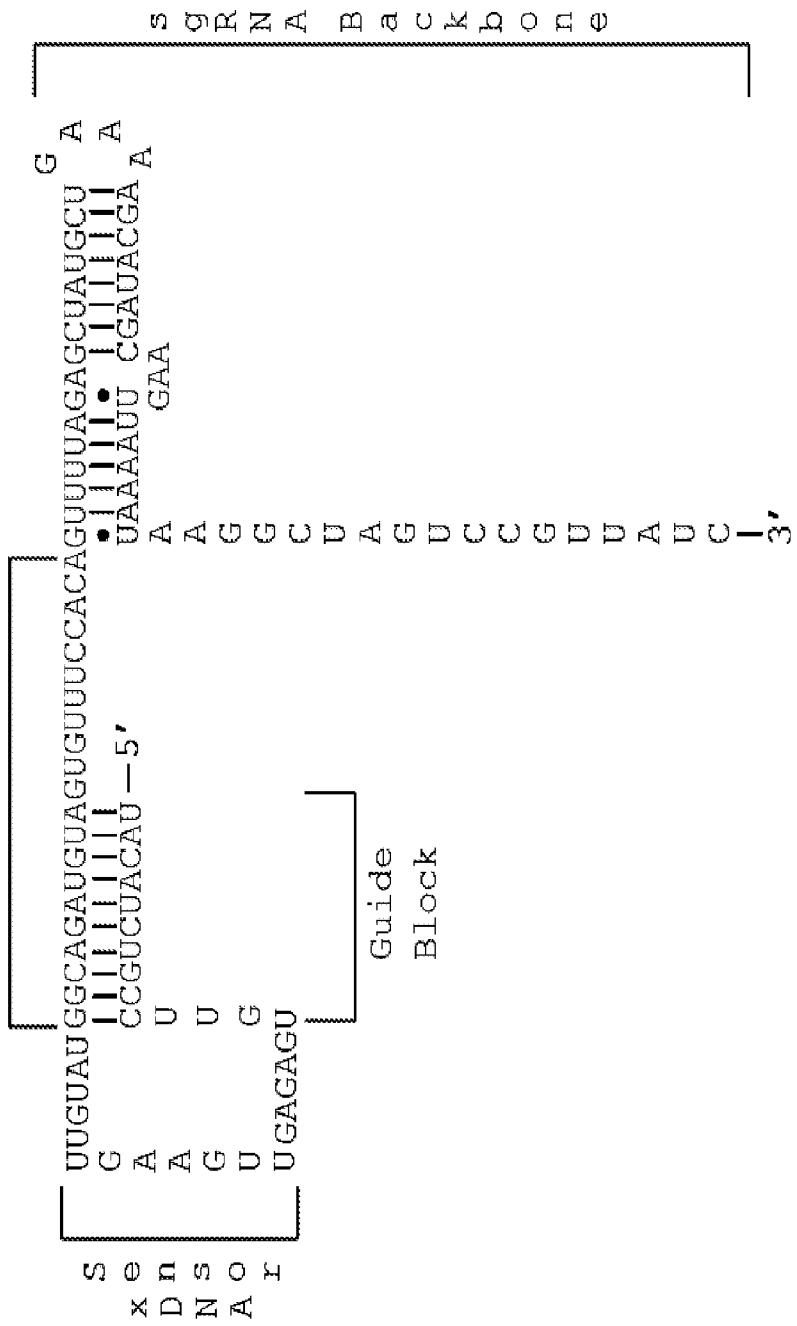


FIG. 3A

gRNA 5' xDNA sensing, On State

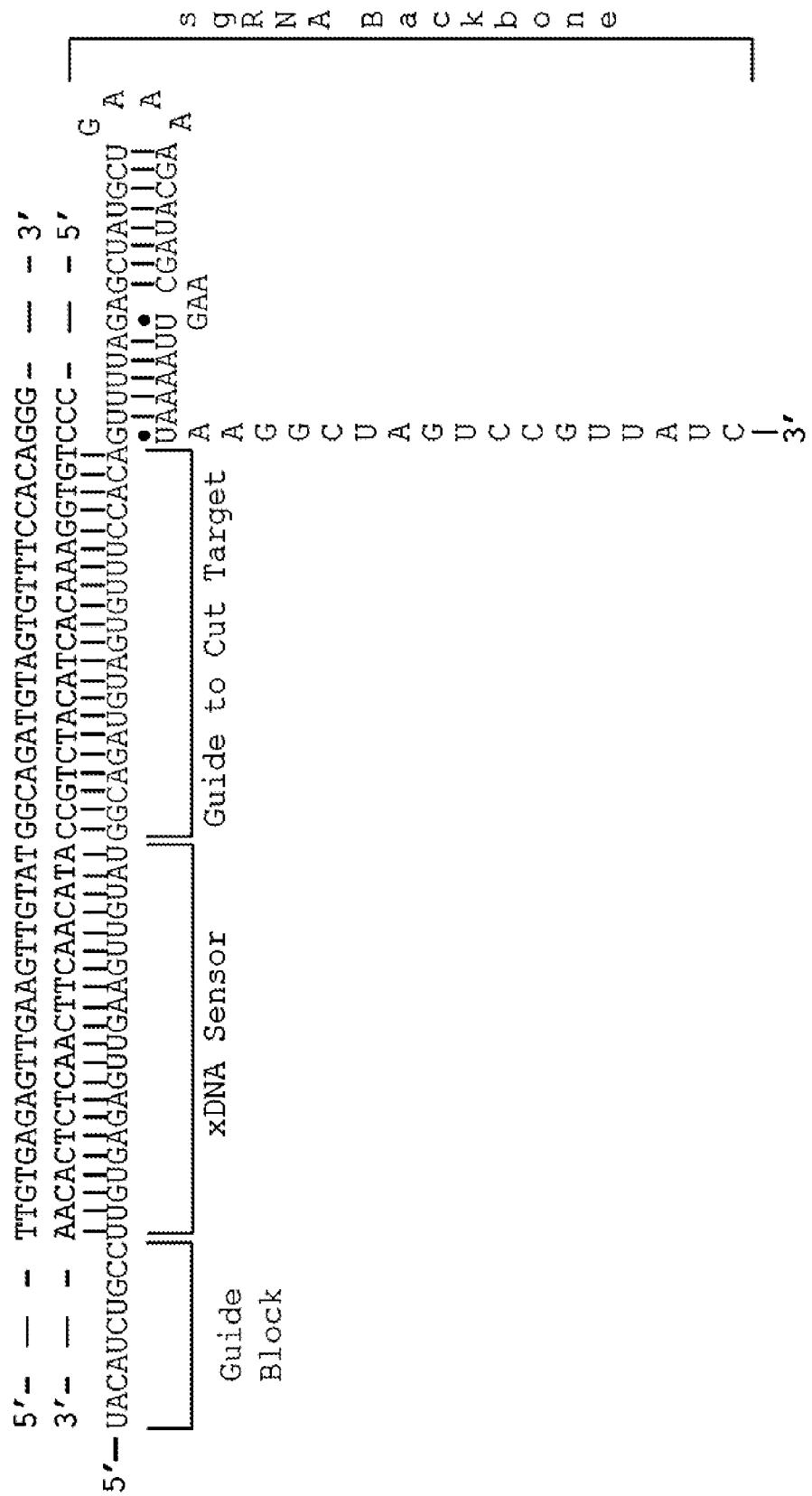


FIG. 3B

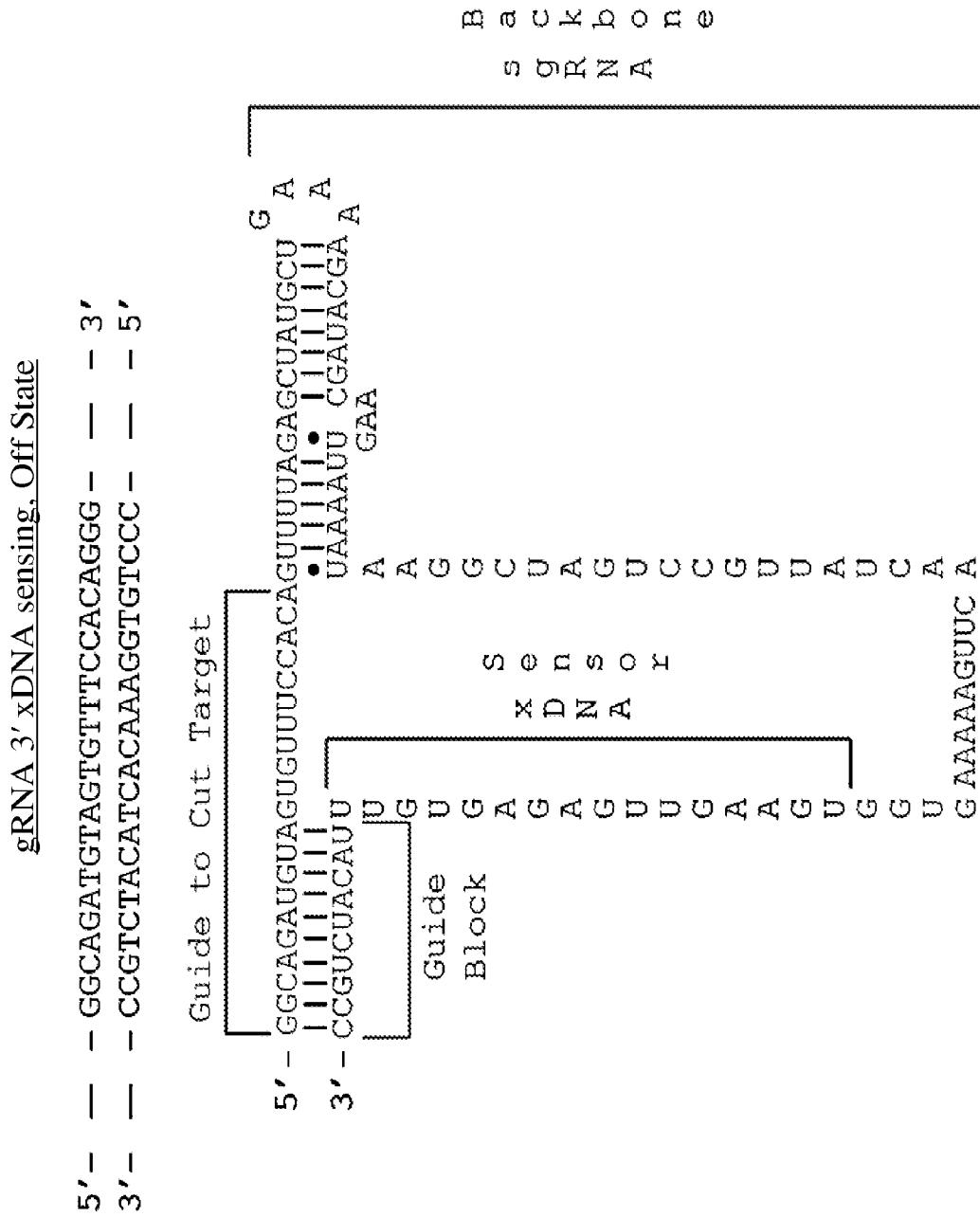


FIG. 3C

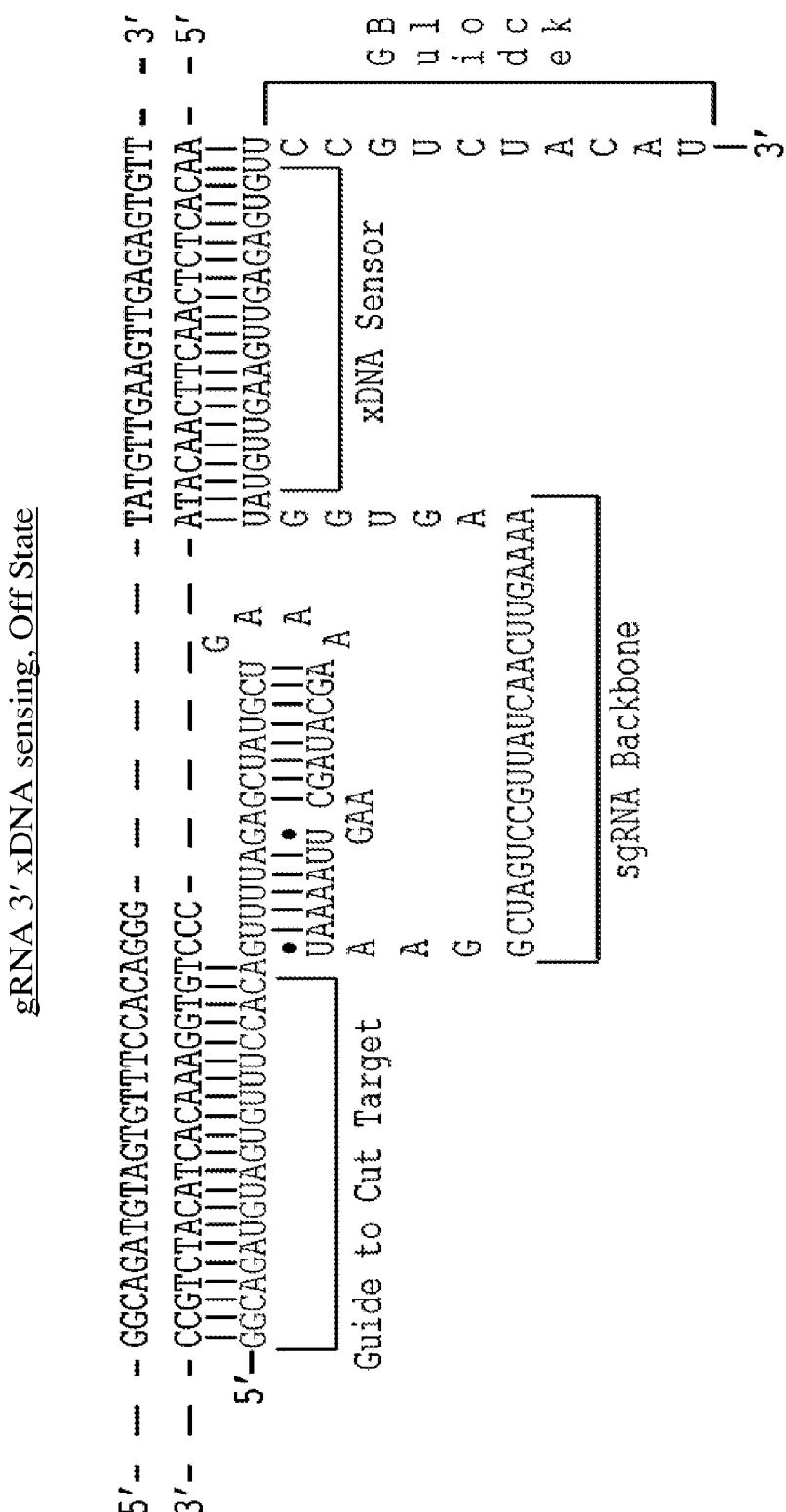


FIG. 3D

