



(51) International Patent Classification:

C07D 471/04 (2006.01) C12N 9/22 (2006.01)

(21) International Application Number:

PCT/US2017/059365

(22) International Filing Date:

31 October 2017 (31.10.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/416,017 01 November 2016 (01.11.2016) US

(71) Applicants: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 QUINCY STREET, CAMBRIDGE, Massachusetts 02138 (US). **THE BRIGHAM AND WOMEN'S HOSPITAL, INC.** [US/US]; 75 FRANCIS STREET, BOSTON, Massachusetts 02115 (US).

(72) Inventors; and

(71) Applicants: **CHOUDHARY, Amit** [US/US]; c/o President and Fellows of Harvard College, 17 Quincy Street, Cambridge, Massachusetts 02138 (US). **WU, Peng** [US/US]; c/o The Brigham and Women's Hospital, Inc., 75 Francis Street, Boston, Massachusetts 02115 (US). **MAJI, Basudeb** [US/US]; c/o The Brigham and Women's Hospital, Inc., 75 Francis Street, Boston, Massachusetts 02115 (US).

(74) Agent: **HUNTER-ENSOR, Melissa**; C/O GREENBERG TRAUIG, LLP, ONE INTERNATIONAL PLACE, BOSTON, Massachusetts 02110 (US).

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

(54) Title: INHIBITORS OF RNA GUIDED NUCLEASES AND USES THEREOF

(57) Abstract: Compositions and methods are provided for the inhibition of the function of RNA guided endonucleases, including the identification and use of such inhibitors.



INHIBITORS OF RNA GUIDED NUCLEASES AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of and priority to U.S. Provisional Patent Application serial number 62/416,017, filed November 1, 2016, which is incorporated herein by reference in its entirety.

STATEMENT OF RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH

This invention was made with government support under Grant No. All 26239 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

The CRISPR (clustered regularly interspaced short palindromic repeat) system is an adaptive immune system used by bacteria and archaea to defend against invading phages or mobile genetic elements. The most studied CRISPR system employs an RNA-guided endonuclease Cas9, which can cleave double-stranded target DNA in multiple cell types. Cas9 identifies the target sequence by two recognition mechanisms: (i) Watson-Crick base-pairing between the target DNA sequence and guide RNA and (ii) Protospacer Adjacent Motif (PAM) sequence on the target DNA. Upon target recognition, Cas9 induces double-strand breaks in the target gene, which when repaired by non-homologous end joining (NHEJ) can result in frameshift mutations and gene knockdown. Alternatively, homology-directed repair (HDR) at the double-strand break site can allow insertion of the desired sequence.

Two common variants of Cas9 are SpCas9 and SaCas9, which naturally occur in *S. pyogenes* and *S. aureus*, respectively, and recently another endonuclease called Cpf1 has been reported. The relative ease of targeting Cas9/Cpf1 to specific genomic loci has enabled the development of revolutionary biomedical technologies. For example, catalytically inactive Cas9 (called dCas9), when fused to transcriptional activators, has enabled genome-wide screening of gene targets. Further, by targeting dCas9 to the promoter or exonic sequences, transcriptional repression has been accomplished. In yet another example, a fusion of dCas9 to acetyltransferases has enabled epigenome editing. Imaging of specific genomic loci has been accomplished by fusing dCas9 to GFP.

There are multiple reasons to establish controls on Cas9 activity. First, as described by Paracelsus' "The dose makes the poison", dosable control of the therapeutic activity is important for effective therapeutic strategies. Indeed, Cas9 exhibits undesirable off-target editing and chromosomal translocations when present at high concentrations. Second, most gene delivery systems have constitutively active Cas9, which is important to be terminated rapidly following on-target gene-editing. Third, Cas9-based technologies (e.g., transcriptional regulation) would benefit from dosable and temporal control of Cas9 activity.

The rapid ascension of CRISPR-based genome editing technologies has raised serious biosafety and bioterrorism concerns, leading to calls for a moratorium and responsible conduct. In particular, much concern has surrounded CRISPR-based gene drives. In sexual reproduction, the progenies receive two versions of a gene, one from each parent. Gene drives enable replacement of one version of the gene with the other "selfish" version of the gene, thereby converting a heterozygous individual to homozygous individual. In laboratory settings, CRISPR-based gene drives have successfully enabled self-propagation of engineered genes in multiple organisms (e.g., mosquitoes) and complete annihilation of wild-type genes. For example, using gene drives engineered mosquitoes have been generated that can wipe out the entire species by ensuring that every female progeny is infertile. Gene drives can be used to propagate a particular trait in the entire ecosystem, which may find use in the elimination of diseases (e.g., malaria, dengue fever) or invasive species, and reversing pesticide resistance in plants. On the other hand, there exists the malevolent use of gene drives in entomological and agricultural settings.

Reports of small-molecule controlled Cas9 activity are present in literature and involve fusing Cas9 to small-molecule controlled protein domains. Genetic-fusions of Cas9 to small-molecule controlled degrons (e.g., Wandless' destabilized domains) may allow aforementioned controls, but such fusions to have unacceptably high background activity presumably owing to the large size of Cas9. These systems also do not ensure dosage control-the small molecules act merely as an inducer of Cas9 activity. Further, these "inducer" small molecules cannot control gene drives containing wild-type Cas9/Cpf1. A general approach would be desirable to control all variants of Cas9/Cpf1, including the wild type and engineered versions. The use of "inducible" systems to control gene drives is also questionable given that the "inducer" small molecules are toxic at the organismal level (albeit not at the cellular level, where these systems were developed).

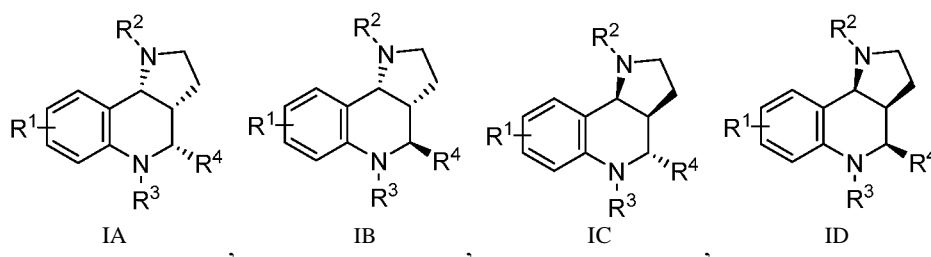
More importantly, large-sized genetic-fusion constructs are incompatible with the most common Cas9 gene delivery systems under investigation for therapeutic gene therapy. The application of these "inducible systems" in a therapeutic setting will be challenging as they involve fusion of large genes to Cas9 gene. Since Cas9 is a large protein, fitting even Cas9 gene into virus delivery systems (e.g., AAV) has been an enormous challenge. Even the smallest of the small-molecule controlled systems will aggrandize the delivery problems. Finally, many small-molecule "inducible" Cas9 constructs exist, but none allow dosable control. The reported "inducible" systems are not reversible upon removal of the small molecule, and therefore, do not allow complete temporal control.

Currently, no method exists for rapid, reversible dosage and temporal control of CRISPR-based technologies or to thwart the malevolent use of gene drives. Accordingly, a need exists for compositions and methods for inhibiting one or more activities of RNA guided endonuclease (e.g., Cas9, Cpf1). Such compositions and methods are useful for regulating the activity of RNA guided endonucleases (e.g., in genome editing).

SUMMARY OF THE INVENTION

The invention provides compositions and methods for inhibiting the activity of RNA guided endonucleases (e.g., Cas9, Cpf1), and methods of use therefor, including rapid, reversible, dosage, and/or temporal control of RNA guided endonuclease technologies. Also provided are high-throughput biochemical and cellular assays for detecting one or more activities of RNA guided endonucleases, and methods of using them to identify or screen agents that inhibit RNA guided nucleases.

In one aspect, the invention provides a compound having the structure of Formula IA, IB, IC, or ID:



where R^1 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R^2 - R^4 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

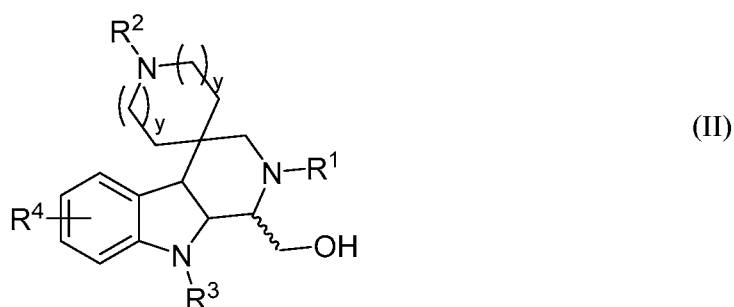
X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, NR₂, or halogen (*e.g.*, -Cl, -F, -Br, etc.);

Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)-O-(CH₂)_n-, -(CH₂)_n-C(=O)-NH-(CH₂)_n-, -(CH₂)_n-NH-C(=O)-(CH₂)_n-, -(CH₂)_n-NH-SO₂-(CH₂)_n-, -(CH₂)_n-SO₂-NH-(CH₂)_n-, -(CH₂)_n-SO₂-(CH₂)_n-, -(CH₂)_n-SO₂-NH-C(=O)-(CH₂)_n-, -(CH₂)_n-R^L-, -R^L-C(=O)-O-, -S-, -S(O)-, -SO₂-NH-(CH₂)_n-, -R^L-NH-C(=O)-(CH₂)_n-, -R^L-NH-S(O)₂-(CH₂)_n-, -S(O)₂-; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C₁-C₁₂ linear and/or branched and/or cyclic and/or aromatic bivalent radicals (*e.g.*, alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (*e.g.*, 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C₁-C₁₂ hydrocarbons (*e.g.*, alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In another aspect, the invention provides a compound having the structure of Formula (II):



where "y" is 0 or 1,

R⁴ is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R¹-R³ are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, NR₂, or halogen (*e.g.*, -Cl, -F, -Br, etc.);

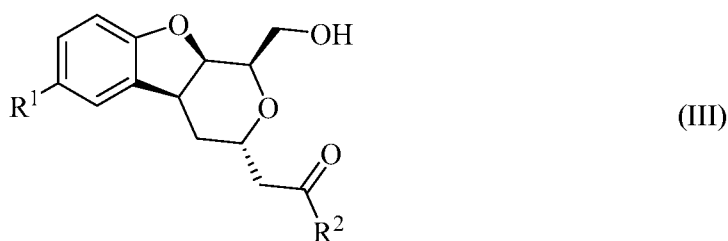
Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)-O-, -(CH₂)_n-C(=O)-NH-(CH₂)_n-, -(CH₂)_n-NH-C(=O)-(CH₂)_n-, -(CH₂)_n-NH-SO₂-(CH₂)_n-, -(CH₂)_n-SO₂-NH-(CH₂)_n-, -(CH₂)_n-SO₂-(CH₂)_n-, -(CH₂)_n-SO₂-NH-C(=O)-(CH₂)_n-, -(CH₂)_n-R^L-, -R^L-C(=O)-O-, -SO₂-NH-

$(\text{CH}_2)_n$ -, $-\text{R}^{\text{L}}\text{-NH-C(0)-}(\text{CH}_2)_n$ -, $-\text{R}^{\text{L}}\text{-NH-S(0)}_2(\text{CH}_2)_n$ -, $-\text{S-}$, $-\text{S(O)-}$, $-\text{S(O)}_2$ -; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^{L} is independently selected at each occurrence from Ci-Ci_2 linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from Ci-i_2 hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In another aspect, the invention provides a compound having the structure of Formula (III):



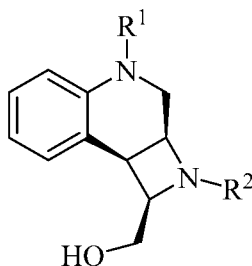
where $\text{R}^1\text{-R}^2$ are independently selected from hydrogen, $-\text{X}$, $-\text{R}$, $-\text{Li-X}$, or $-\text{Li-R}$; where X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, NR_2 , or halogen (e.g., $-\text{Cl}$, $-\text{F}$, $-\text{Br}$, etc.);

Li is selected from $-(\text{CH}_2)_n$ -, $-(\text{CH}_2)_n\text{-C(0)-O-}$, $-(\text{CH}_2)_n\text{-C(0)-NH-}(\text{CH}_2)_n$ -, $-(\text{CH}_2)_n\text{-NH-C(0)-}(\text{CH}_2)_n$ -, $-(\text{CH}_2)_n\text{-NH-SO}_2(\text{CH}_2)_n$ -, $-(\text{CH}_2)_n\text{-SO}_2\text{-NH-}(\text{CH}_2)_n$ -, $-(\text{CH}_2)_n\text{-SO}_2(\text{CH}_2)_n$ -, $-(\text{CH}_2)_n\text{-SO}_2\text{-NH-C(0)-}(\text{CH}_2)_n$ -, $-(\text{CH}_2)_n\text{-R}^{\text{L}}\text{-}$, $-\text{R}^{\text{L}}\text{-C(0)-O-}$, $-\text{S-}$, $-\text{S(O)-}$, $-\text{S(O)}_2$ -; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^{L} is independently selected at each occurrence from Ci-Ci_2 linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from Ci-i_2 hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In another aspect, the invention provides a compound having the structure:



where each wavy bond may be in the R or the S configuration,

R^1 - R^2 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

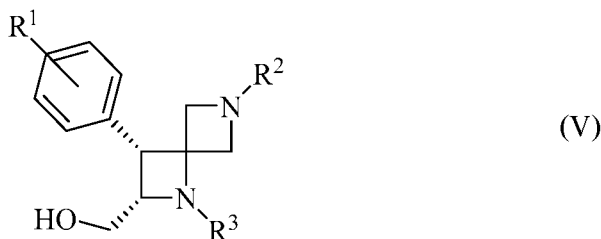
X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen (e.g., -Cl, -F, -Br, etc.);

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)-O-(CH_2)_n-$, $-(CH_2)_n-C(=O)-NH-(CH_2)_n-$, $-(CH_2)_n-NH-C(=O)- (CH_2)_n-$, $-(CH_2)_n-NH-SO_2-(CH_2)_n-$, $-(CH_2)_n-SO_2-NH-(CH_2)_n-$, $-(CH_2)_n-SO_2-(CH_2)_n-$, $-(CH_2)_n-SO_2-NH-C(=O)- (CH_2)_n-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)-O-$, $-S-$, $-S(O)-$, $-S(O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1 - C_{12} hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In another aspect, the invention provides a compound having the structure of formula (V):



where R^1 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R^2 - R^3 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, NR_2 , or halogen (e.g., -Cl, -F, -Br, etc.);

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-(CH_2)_n-$, $-(CH_2)_n-C(=O)-NH-(CH_2)_n-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-$, $-(CH_2)_n-NH-SO_2-(CH_2)_n-$, $-(CH_2)_n-SO_2-NH-(CH_2)_n-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-(CH_2)_n-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)O-$, $-S-$, $-S(O)-$, $-S(O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C_1-C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1-i_2 hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In another aspect, the invention provides a method of inhibiting the activity of an RNA guided endonuclease-guide RNA complex, involving contacting the RNA guided endonuclease-guide RNA complex with a small molecule.

In another aspect, the invention provides a method of detecting nucleic acid binding of an RNA guided endonuclease-guide RNA complex involving contacting the RNA guided endonuclease-guide RNA complex with a probe including an oligonucleotide having at least 4-12 PAM sites and a fluorescent reporter at the 5' terminus; and detecting an increase in fluorescence polarization relative to a reference, thereby detecting nucleic acid binding of the RNA guided endonuclease-guide RNA complex.

In another aspect, the invention provides a method of detecting nuclease activity of an RNA guided endonuclease-guide RNA complex involving providing an RNA guided endonuclease-guide RNA complex bound to a probe including an oligonucleotide having at least one PAM site and a fluorescent reporter at the 5' terminus; and detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nucleic acid binding of the RNA guided endonuclease-guide RNA complex.

In another aspect, the invention provides a method of identifying an agent that inhibits an RNA guided endonuclease-guide RNA complex, involving contacting an RNA guided endonuclease-guide RNA complex with an agent in the presence of a probe including an oligonucleotide having at least 4-12 PAM sites and a fluorescent reporter at the 5' terminus; and failing to detect an increase in fluorescence polarization to a reference, thereby identifying the agent as an agent that inhibits an RNA guided endonuclease-guide RNA complex.

In another aspect, the invention provides a method of identifying an agent that inhibits nuclease activity of an RNA guided endonuclease-guide RNA complex, involving contacting an RNA guided endonuclease-guide RNA complex with an agent, where the RNA guided endonuclease-guide RNA complex is bound to a probe including an oligonucleotide having at least one PAM site and a fluorescent reporter at the 5' terminus; and detecting an increase in fluorescence polarization relative to a reference, thereby identifying the agent as an agent that inhibits nuclease activity of the RNA guided endonuclease-guide RNA complex.

In another aspect, the invention provides a method of identifying an agent that inhibits an RNA guided endonuclease-guide RNA complex, involving contacting a nucleic acid expressing a nucleic acid aptamer with the RNA guided endonuclease-guide RNA complex in the presence of a fluorophore molecule and an agent, where binding of the aptamer to the fluorophore molecule generates a fluorescent signal; and failing to detect a decrease in fluorescent signal relative to a reference, thereby identifying the agent as an agent that inhibits an RNA guided endonuclease-guide RNA complex.

In another aspect, the invention provides a method of detecting nucleic acid binding and/or nuclease activity of an RNA guided endonuclease-guide RNA complex involving contacting a nucleic acid expressing a nucleic acid aptamer with the RNA guided endonuclease-guide RNA complex in the presence of a fluorophore molecule, where binding of the aptamer to the fluorophore molecule generates a fluorescent signal, and detecting a decrease in fluorescent signal relative to a reference, thereby detecting nucleic acid binding or nuclease activity of the RNA guided endonuclease-guide RNA complex.

In another aspect, the invention provides a method of identifying an agent that inhibits an RNA guided endonuclease-guide RNA complex, involving one or more methods according to any aspect delineated herein.

In another aspect, the invention provides an oligonucleotide having at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or more PAM sites and a fluorescent reporter at the 5' or 3' terminus.

In another aspect, the invention provides a kit containing the oligonucleotide according to any aspect delineated herein.

In various embodiments of any aspect delineated herein, the nucleic acid binding and/or nuclease activity is inhibited or prevented by an agent (e.g., a small molecule). In various embodiments of any aspect delineated herein, the agent or small molecule inhibits one or more of nucleic acid binding or nuclease activity.

In various embodiments of any aspect delineated herein, the method is *in vitro* or *in vivo*. In various embodiments of any aspect delineated herein, the method is performed in a cell (e.g., a prokaryotic or eukaryotic cell). In some embodiments, the prokaryotic cell is a bacterium. In other embodiments, the eukaryotic cell is a human, mammalian, insect, plant, or yeast cell. In particular embodiments, the cell is in an organism (e.g., a human, mammal, vertebrate, invertebrate, insect, or plant).

In various embodiments of any aspect delineated herein, the agent or small molecule inhibits binding of the RNA guided endonuclease-guide RNA complex to a PAM site.

In various embodiments, the RNA guided endonuclease is Cas9 or Cpf1. In various embodiments, the one or more PAM sites are 3' or 5' of a target nucleic acid (with reference to the non-target strand). In various embodiments of any aspect delineated herein, the one or more PAM sites have the nucleic acid sequence 5'-NGG-3' (SpCas9), 5'-TTN-3' (FnCpf1), 5'-NNGGGT-3' (SaCas9), or 5'-TTTN-3' (As/LbCpf1). In various embodiments of any aspect delineated herein, the fluorescent reporter is Fluorescein isothiocyanate (FITC).

In various embodiments, the probe includes an oligonucleotide having one or more of the nucleic acid sequences:

5'-

GGCTiXiACCACGCGGGAAAATCCACCT^A AGGTGGTTCCTCTTCGGATGTTCCATCC
TTT/36-FAM-3",

3'-

CCGACCTCKJTCKXK:CCTTTTAG<JTG<JATCCACCAAGGAGAA<K:CTACAAGGTAG

G

AAA~5',

5'-6-FAM/TAA1'ACGACTC ACTATAGGACGCGACCGAAATGGTGAAGGACGGGT-

3',

5'-6-

FAM/ACTCACTATAGGGACGCGACCGAAATGGTGAAGGACGGGTCCAGTGCTTCG

G-3\

5'-CGTCCTTCACCATTTTCGGTCGCGTCCCTATAGTGAGTCGTATTAG-TTCCAT/6-

FAM~3\ and

5'-6-

FAM'ATG<JA\CTA\TACGACTCACTATACGGACX)(X)ACCGA A\TCGTGAAGGAC

G-3'.

In various embodiments, the probe or oligonucleotide specifically binds one or more of the nucleic acid sequences:

5'-ATAGTGAGTCGTATTA/3IABkFQ-3' ,
 5'-CGTCCCTATAGTGAGT/3IABkFQ-3' ,
 5'-5IABkFQ/ATGGAACTAATACGAC-3' , and
 5'-GTCGTATTAGTTCCAT/3IABkFQ-3' .

In various embodiments, the guide RNA includes the nucleic acid sequence:

5'-
 GCUAUAGGACGCGACCGAAAGUUUAGAGCUAGAAAUAGCAAGUUAAAAUAA
 GGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' .

In various embodiments of any aspect delineated herein, the reference is the fluorescence polarization of the probe in the presence of the agent or small compound and in the absence of the RNA guided endonuclease-guide RNA complex.

In various embodiments of any aspect delineated herein, the nucleic acid expresses a nucleic acid aptamer including a PAM site. In various embodiments, the nucleic acid aptamer has the sequence

5'-
 GCUAUAGGACGCGACCGAAAGUUUAGAGCUAGAAAUAGCAAGUUAAAAUAA
 GGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU -3',
 5'GGUGACGCGACCGAAAUGGUGAAGGACGGGUCCAGUGCUUCGGCACUGUUG
 AGUAGAGUGUGAGCUCCGUAACUGGUCGCGUC-3' , or
 5'GGNNNGACGCGACCGAAAUGGUGAAGGACGGGUCCAGUGCUUCGGCACUG
 UUGAGUAGAGUGUGAGCUCCGUAACUGGUCGCGUC-3' .

In various embodiments, the DNA template of the nucleic acid has the sequence:

5'-
 GCGCGCTTTCTAATACGACTCACTATAGGGTGACGCGACCGAAATGGTGAAGGA
 CGGGTCCAGTGCTTCGGCACTGTTGAGTAGAGTGTGAGCTCCGTAACCTGGTCGCG
 TC-3'

or

5'-
 GCGCGCNNNNTAATACGACTCACTATAGGGNNNNGACGCGACCGAAATGGTGAA

GGACGGGTCCAGTGCTTCGGCACTGTTGAGTAGAGTGTGAGCTC CGTAACTGGTC GCGTC-3'.

In various embodiments, the guide RNA includes the nucleic acid sequence:

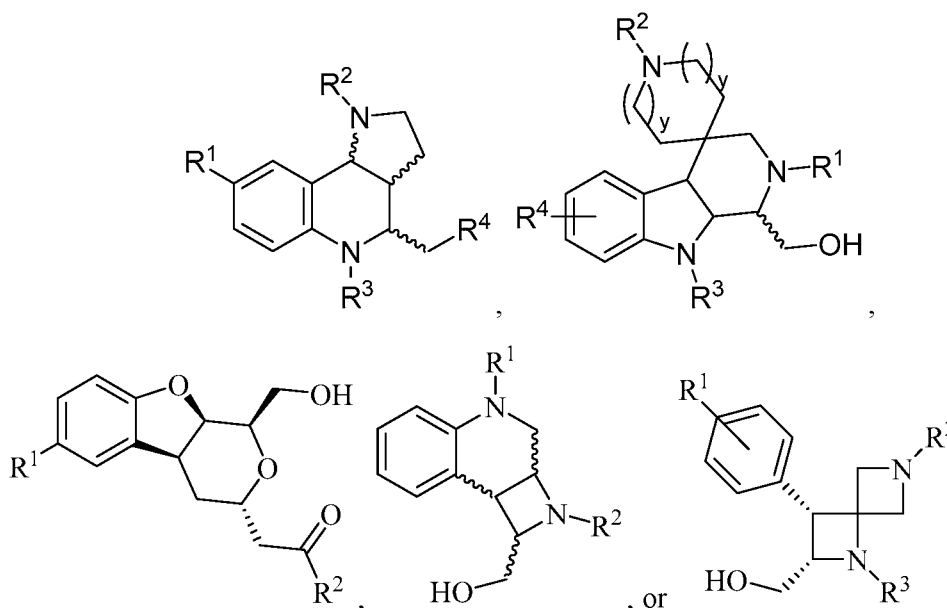
5'-

GCUAUAGGACGCGACCGAAAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAA GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'

In various embodiments of any aspect delineated herein, the fluorophore molecule is one or more of 4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1 *H*-imidazol-5-one (DFHBI). In various embodiments, the reference is the fluorescence in the absence of the RNA guided endonuclease-guide RNA complex.

In various embodiments of any aspect delineated herein, the method further involves assaying the agent or small compound in one or more of an EGFP disruption assay, DNA cleavage assay, or Surveyor assay.

In various embodiments of any aspect delineated herein, the agent or small molecule is one or more of:



where each wavy bond may be in either the R or the S configuration, with the proviso that two wavy bonds of membered of a five or six membered fused ring have the same stereochemistry,

R¹-R⁴ are independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; where

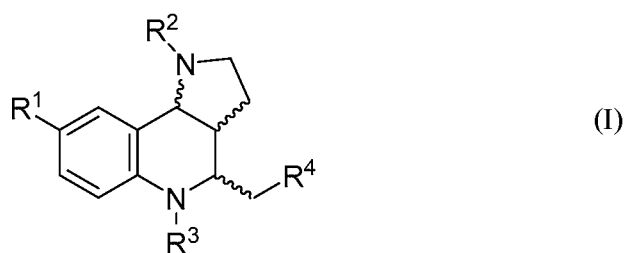
X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen (e.g., -Cl, -F, -Br, etc.);

Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)O-, -(CH₂)_n-C(=O)-NH-, -(CH₂)_n-NH-C(=O)-, -(CH₂)_n-NH-SO₂-, -NH-C(=O)-(CH₂)_n-, -(CH₂)_n-SO₂-NH-, -SO₂-NH-(CH₂)_n-, -(CH₂)_n-SO₂-NH-C(=O)-, -R^L-NH-C(=O)-(CH₂)_n-, -R^L-NH-S(=O)₂-(CH₂)_n-, -(CH₂)_n-R^L-, -R^L-C(=O)-O-, -S-, -S(O)-, -S(O)₂-; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C₁-C₆ linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C₁-C₁₀ hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In various embodiments of any aspect delineated herein, the small molecule or compound has the formula of Formula I:



where R¹ is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R₂-R₄ are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen (e.g., -Cl, -F, -Br, etc.);

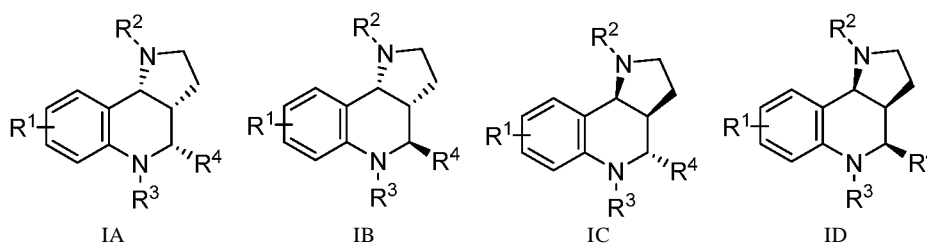
Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)O-, -(CH₂)_n-C(=O)-NH-, -(CH₂)_n-NH-C(=O)-, -(CH₂)_n-NH-SO₂-, -(CH₂)_n-SO₂-NH-, -(CH₂)_n-SO₂-, -(CH₂)_n-SO₂-NH-C(=O)-, -(CH₂)_n-R^L-, -R^L-C(=O)-O-, -R^L-NH-C(=O)-(CH₂)_n-, -R^L-NH-S(=O)₂-(CH₂)_n-, -S-, -S(O)-, -S(O)₂-; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C₁-C₆ linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-

alkyl, and combinations thereof); optionally substituted with one or more (*e.g.*, 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C₁₋₁₂ hydrocarbons (*e.g.*, alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In various embodiments, the small molecule or compound has the formula of Formula IA, IB, IC, or ID:



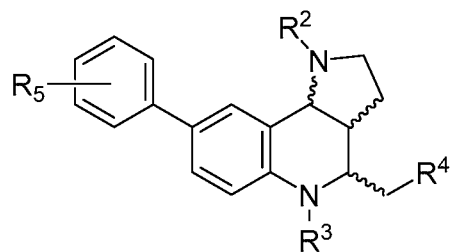
where R¹ is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R²-R⁴ are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen (*e.g.*, -Cl, -F, -Br, etc.);

Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)-, -(CH₂)_n-C(=O)-NH-, -(CH₂)_n-NH-C(=O)-, -(CH₂)_n-NH-SO₂-, -(CH₂)_n-SO₂-NH-, -(CH₂)_n-SO₂-, -(CH₂)_n-SO₂-NH-C(=O)-, -(CH₂)_n-R^L-, -R^L-C(=O)-, -R^L-NH-C(=O)-(CH₂)_n-, -R^L-NH-SO₂-(CH₂)_n-, -S-, -S(O)-, -S(O)₂-;

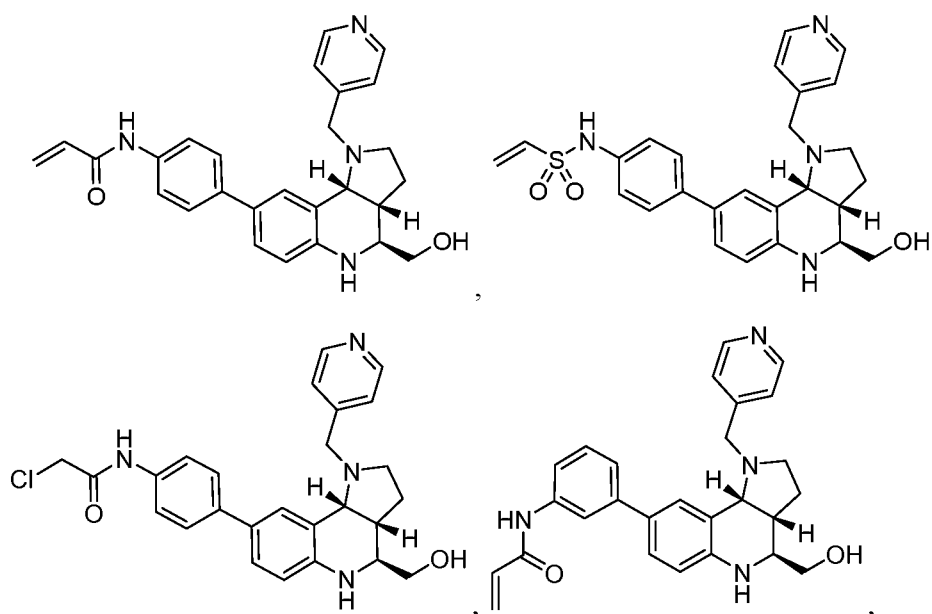
R^L is independently selected at each occurrence from C₁-C₁₂ linear and/or branched and/or cyclic and/or aromatic bivalent radicals (*e.g.*, alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (*e.g.*, 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

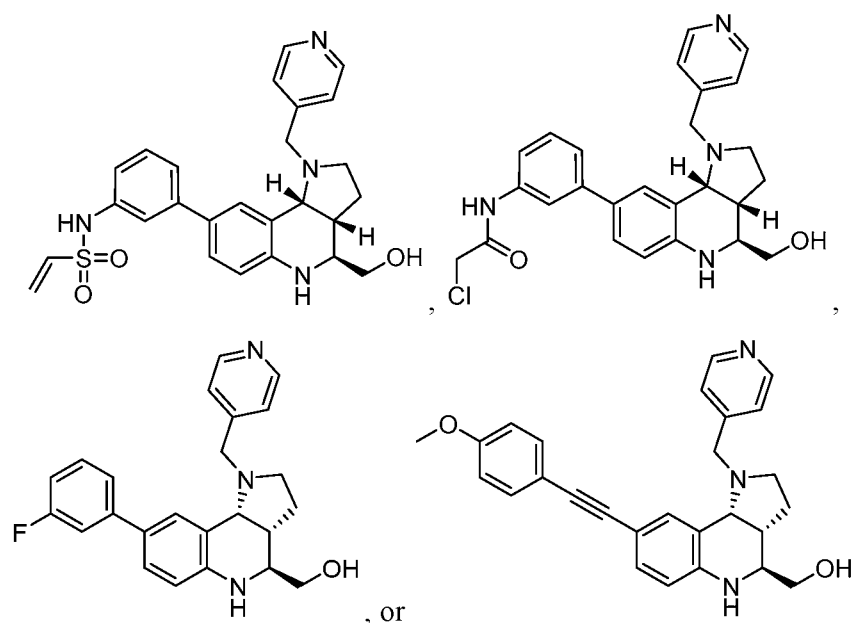
R is selected from C₁₋₁₂ hydrocarbons (*e.g.*, alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof. In some embodiments, the small molecule or compound has the structure:



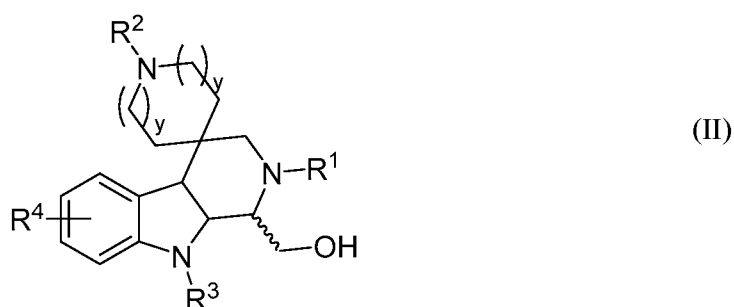
wherein R_5 is selected from hydrogen, -R, -X, -Li-X, or -Li-R. In some embodiments, R_5 is selected from halogen, alkoxy, -NH-CO-alkene, -NH-S(0)₂-alkene, or -NH-C(O)-(CH₂)_n-halogen.

In certain embodiments, the small molecule or compound is one or more of





In various embodiments of any aspect delineated herein, the small molecule or compound has the formula of Formula II:



where "y" is 0 or 1,

R^4 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R;
and

R^1 - R^3 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

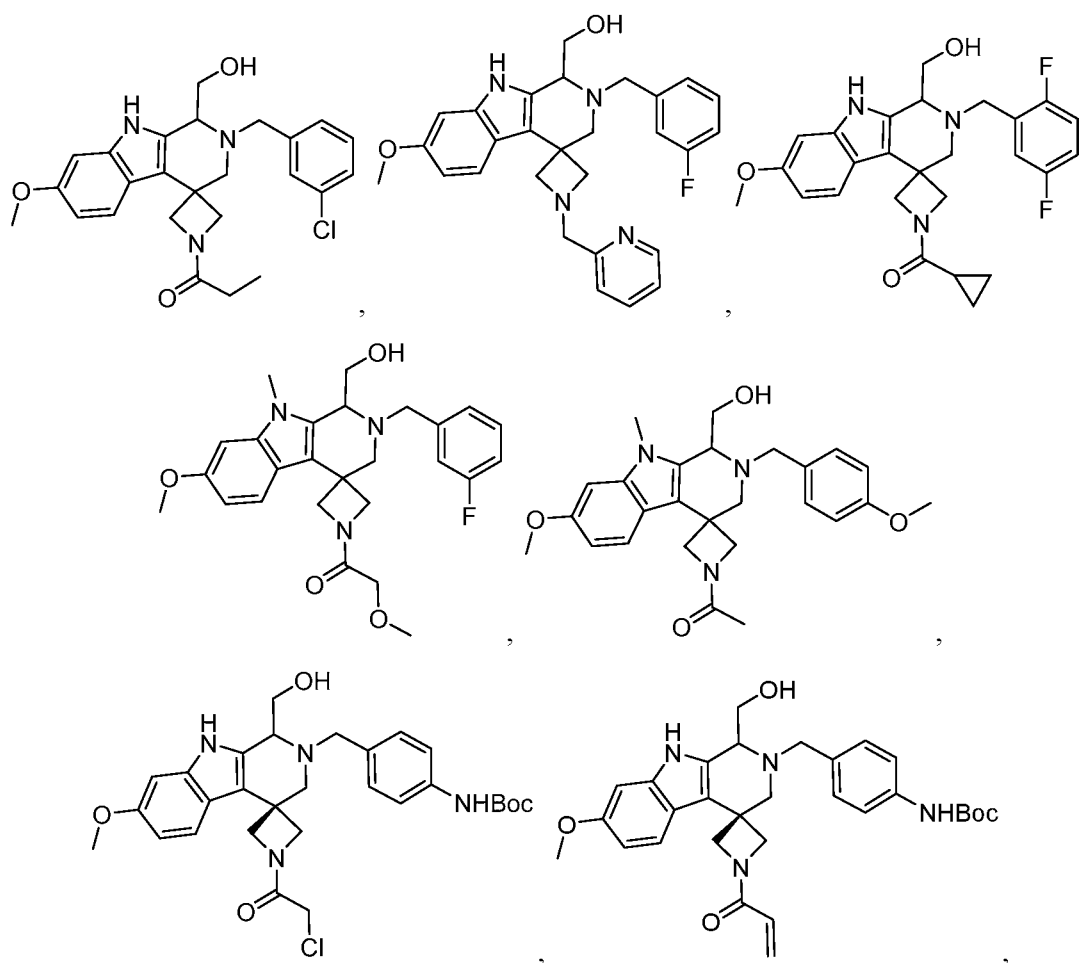
X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen (*e.g.*, -Cl, -F, -Br, etc.);

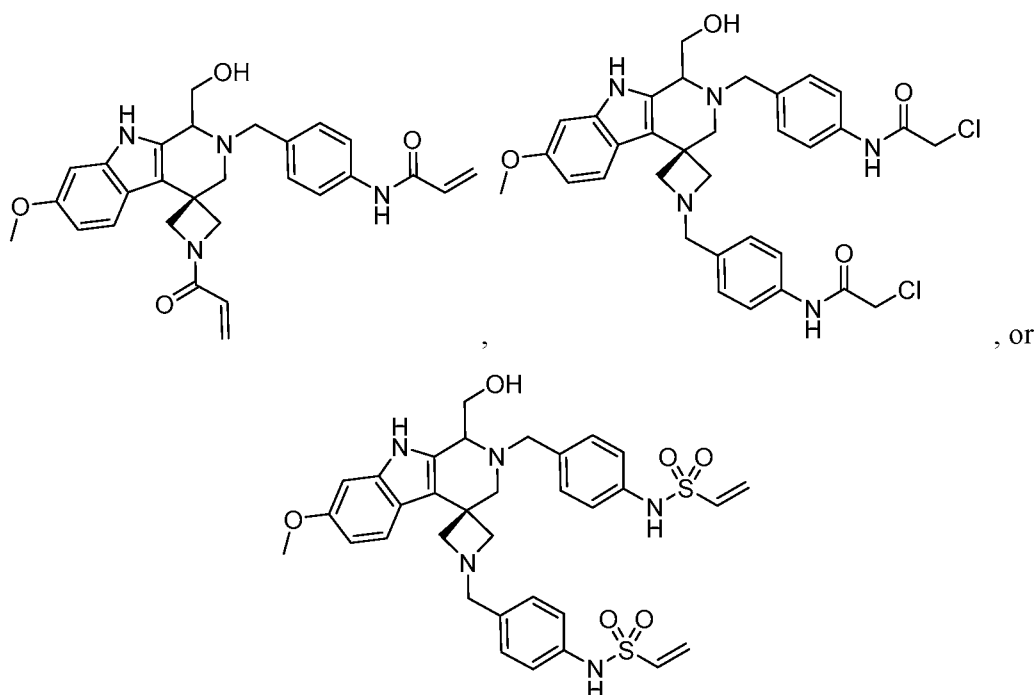
Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-C(=O)-NH-(CH_2)_n-$, $-(CH_2)_n-SO_2-NH-$, $-SO_2-NH-(CH_2)_n-$, $-R^L-NH-C(=O)-(CH_2)_n-$, $-R^L-NH-SO_2-(CH_2)_n-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L$, $-R^L-C(=O)-$, $-S-$, $-S(O)-$, $-S(O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C1-C12 linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

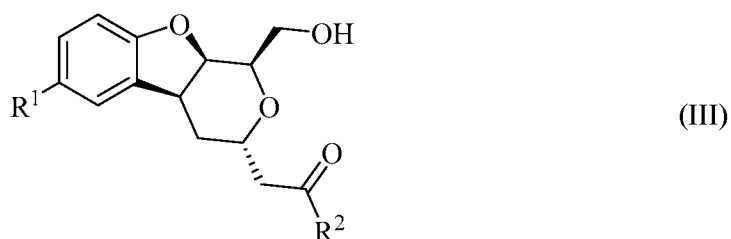
R is selected from C1-12 hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In various embodiments, the small molecule or compound is one or more of:





In various embodiments of any aspect delineated herein, the small molecule or compound has the formula of Formula III:



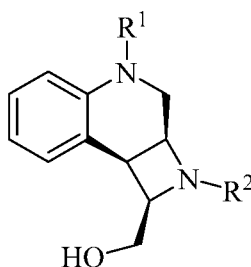
where R¹-R² are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen (*e.g.*, -Cl, -F, -Br, etc.);

Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)-O-, -(CH₂)_n-C(=O)-NH-, -(CH₂)_n-NH-C(=O)-, -(CH₂)_n-NH-SO₂-, -(CH₂)_n-SO₂-NH-, -(CH₂)_n-SO₂-, -(CH₂)_n-SO₂-NH-C(=O)-, -(CH₂)_n-R^L-, -R^L-C(=O)-O-, -S-, -S(O)-, -S(O)₂-;

R^L is independently selected at each occurrence from C₁-C₁₂ linear and/or branched and/or cyclic and/or aromatic bivalent radicals (*e.g.*, alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (*e.g.*, 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C₁₋₁₂ hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In various embodiments of any aspect delineated herein, the small molecule or compound has the formula:



where each wavy bond may be in the R or the S configuration,

R¹-R² are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

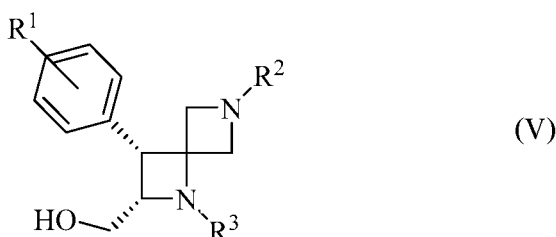
X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen (e.g., -Cl, -F, -Br, etc.);

Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)-, -(CH₂)_n-C(=O)-NH-, -(CH₂)_n-NH-C(=O)-, -(CH₂)_n-NH-SO₂-, -(CH₂)_n-SO₂-NH-, -(CH₂)_n-SO₂-, -(CH₂)_n-SO₂-NH-C(=O)-, -(CH₂)_n-R^L-, -R^L-C(=O)-, -S-, -S(O)-, -S(O)₂-;

R^L is independently selected at each occurrence from C₁-C₁₂ linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C₁₋₁₂ hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In various embodiments of any aspect delineated herein, the small molecule or compound has the formula of Formula V:



where R^1 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R^2 - R^3 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

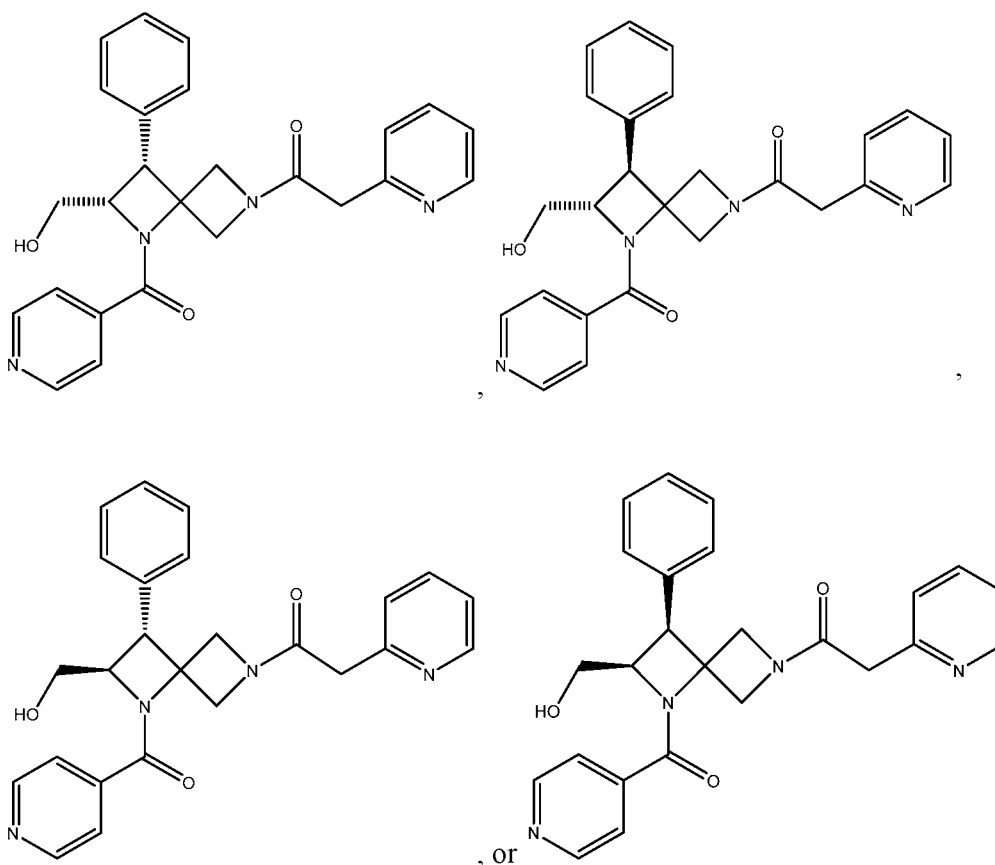
X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen (e.g., -Cl, -F, -Br, etc.);

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)-$, $-S-$, $-S(O)-$, $-S(O)_2-$;

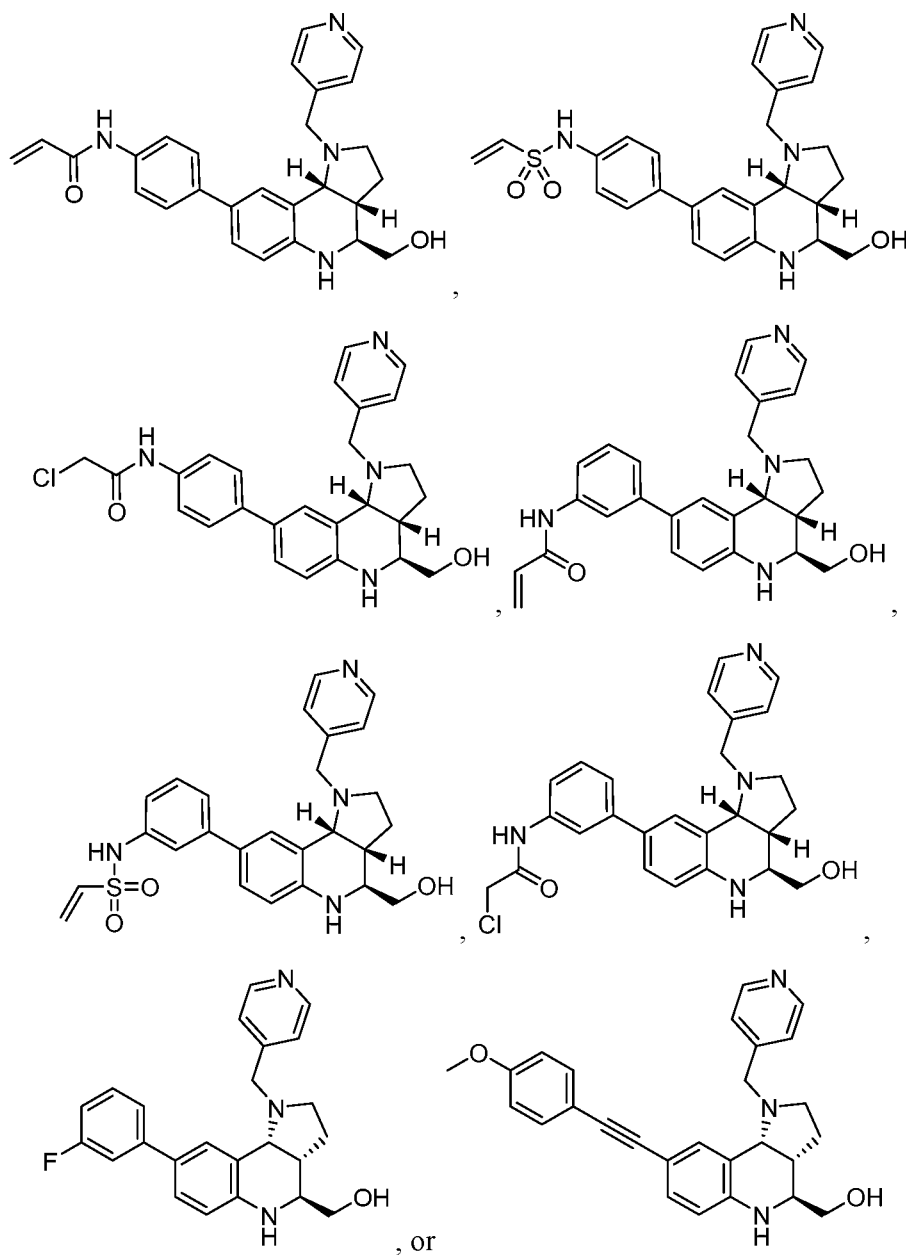
R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1 - C_{12} hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

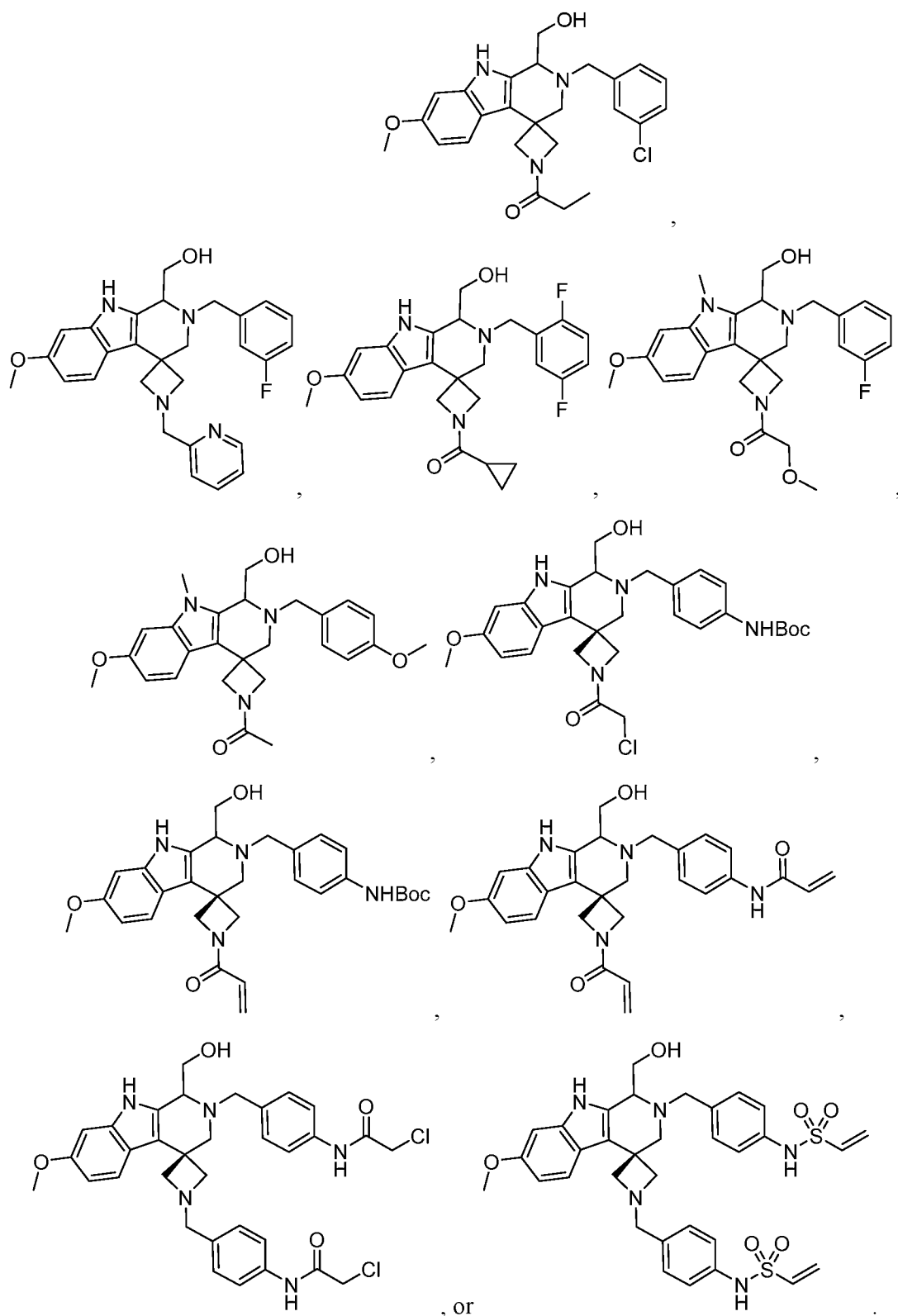
In various embodiments, the small molecule or compound is one or more of



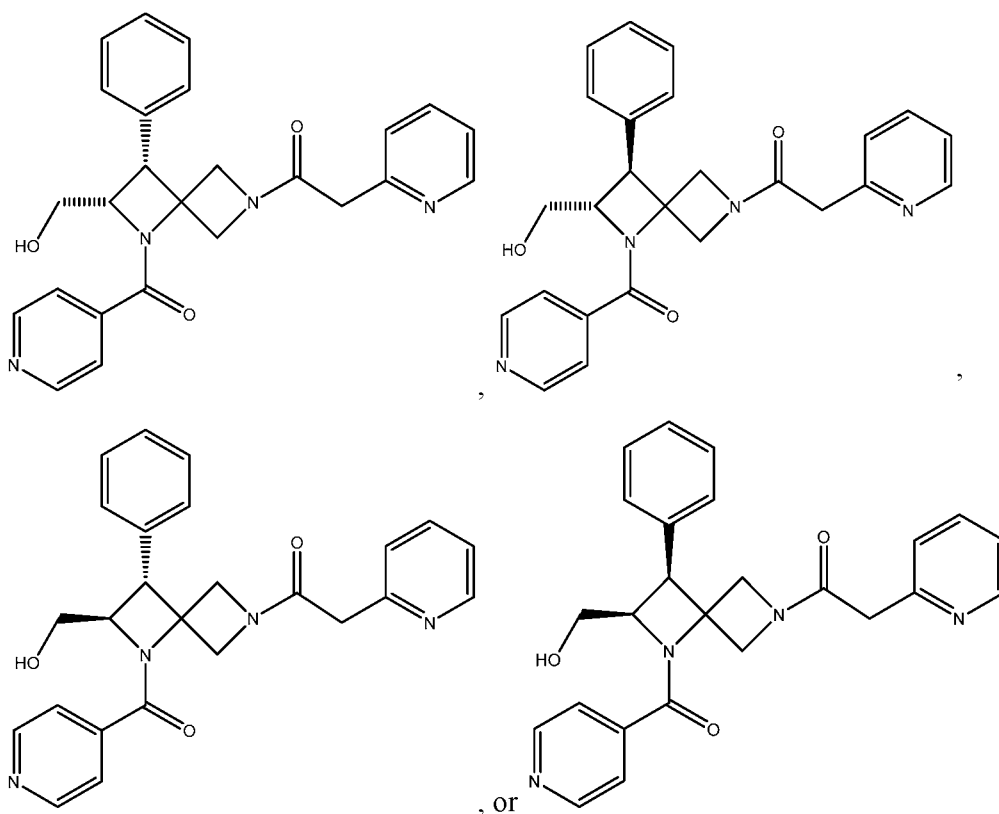
In various embodiments of any aspect delineated herein, the small molecule or compound is one or more of:



In various embodiments of any aspect delineated herein, the small molecule or compound is one or more of:



In various embodiments of any aspect delineated herein, the small molecule or compound is one or more of:



In various embodiments of any aspect delineated herein, compositions comprising any of the compounds delineated may include stereocenters which each may independently be in the (R) configuration, the (S) configuration, or racemic mixtures.

Other features and advantages of the invention will be apparent from the detailed description, and from the claims.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton *et al*, Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger *et al*. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them below, unless specified otherwise.

By "RNA guided endonuclease" is meant a polypeptide having RNA binding activity, DNA binding activity, and/or DNA cleavage activity. RNA guided endonucleases form a complex with a guide RNA, which contains a sequence that is able to bind a target sequence on double stranded DNA. In some embodiments, the RNA guided endonuclease cleaves the double stranded target DNA. Exemplary, RNA guided endonucleases include, without limitation, Cpf1, Cas9, and active fragments, derivatives, and variants thereof.

By "Protospacer adjacent motif (PAM)" is meant a nucleic acid sequence immediately adjacent the nucleic acid sequence targeted by an RNA guided endonuclease (e.g., Cas9, Cpf1). In certain embodiments, the PAM sequence is the Cas9 PAM sequence: 5'-NGG-3'.

By "Cas9 (CRISPR associated protein 9)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_269215 and having RNA binding activity, DNA binding activity, and/or DNA cleavage activity (e.g., endonuclease or nickase activity). An exemplary Cas9 polypeptide sequence is provided below.

```

1  mdkkysigld igtntsvgwav itdeykvpsk kfkvlgntrd hsikknliga llfdsgetae
61  atrlkrtrarr rytrrknrnc ylqeifsnem akvddsffhr leesflveed kkherhpifg
121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlndpnsd
181 vdklfiqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknlgfng
241 lialslgltp nfksnfdlae daklqlskdt yddldnlla qigdqyadlf laaknlsdai
301 llsdilrvnt eitkaplsas mikrydehmq dltllkalvr qqlpekykei ffdqskngya
361 gyidggasqe efykfikpil ekmdgteell vklnredllr kqrtfdngsi phqihlgelg
421 ailrrqedfy pflkdnreki ekiltfrip yvgplargns rfawmtrkse etitpwnfee
481 vvdkgasags fiermtnfdk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
541 sgeqkkaivd llfktnrkvt vkqlkedyfk kiecfdsvei sgvedrfnas lgtyhdllki
601 ikdkdfldne enedilediv ltltlfedre mieerlktya hlfdkvmkq lkrrrytgwg
661 rlsrklingi rdkqsgktil dflksdgfan rnfmqlihdd sltfkediql aqvsgqgds
721 hehianlags paikkgilqt kvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
841 ivpqsflkdd sidnkvlttr dknrgksdnv pseevvkkmk nywrqlnak litqrkfdnl
901 tkaergglse ldagfikrq lvetrqitkh vaqildsrnm tkydendkli revkvitlks
961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvdyvrk
1021 miaksegeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
1081 atvrkvlsmg qvniavktev qtggfskesi lpknsdkli arkkdwdpkk yggfidsptva
1141 ysvllvakve kgkskklksv kelligitime rssfeknpid fleakgykev kkdliiklpk
1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkspe dneqkqlfve

```


1261 qhkhyldi eiisefskrv iladanldkv lsaynkhrdk pireqaenii hlftltlnlga
 1321 paafkyfdtt idrkrytstk evldatlihq sitglyetri dlsqllggd

"Cas9 function" can be defined by any of a number of assays including, but not limited to, fluorescence polarization-based nucleic acid bind assays, fluorescence polarization-based strand invasion assays, transcription assays, EGFP disruption assays, DNA cleavage assays, and/or Surveyor assays, for example, as described herein.

By "Cas 9 nucleic acid molecule" is meant a polynucleotide encoding a Cas9 polypeptide or fragment thereof. An exemplary Cas9 nucleic acid molecule sequence is provided at NCBI Accession No. NC_002737 and is shown below.

```

1 atggataaga aataactcaat aggcttagat atcggcacaa atagcgtcgg atgggacggtg
61 atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc
121 cacagtatca aaaaaaatct tataggggct cttttatattg acagtggaga gacagcggaa
181 gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatattgt
241 tatctacagg agatTTTTTt aaatgagatg gcgaaagtag atgatagttt ctttcatcga
301 cttgaagagt ctttttttgt ggaagaagac aagaagcatg aacgtcatcc tatttttggg
361 aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa
421 aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcgt
481 atgattaagt ttctgtgtca ttttttgatt gagggagatt taaatcctga taatagtgtat
541 gtggacaaac tatTTtatcca gttggtacaa acctacaatc aattatttga agaaaaccct
601 attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga
661 cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttggaagt
721 ctcatgtgct tgctattggg tttgacctct aattttaaat caaattttga tttggcagaa
781 gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg
841 caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt
901 ttactttcag atatcctaag agtaaatact gaaataacta aggtctccct atcagcttca
961 atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga
1021 caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca
1081 gggttatattg atggggggagc tagccaagaa gaattttata aatttatcaa accaatttta
1141 gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc
1201 aagcaacgga ctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat
1261 gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt
1321 gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt
1381 cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa
1441 gttgtcgata aagggtgctt agctcaatca tttattgaac gcatgacaaa ctttgataaa
1501 aatcttccaa atgaaaaagt actacaaaaa catagtttgc tttatgagta ttttacggtt
1561 tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt
1621 tcagggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc

```

```

1681 gttaagcaat taaaagaaga ttattttcaaa aaaatagaat gttttgatag tgttgaaatt
1741 tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt
1801 attaaagata aagatTTTTT ggataatgaa gaaaatgaag atatcttaga ggatattggt
1861 ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct
1921 cacctctttg atgataaggt gatgaaacag cttaaactgc gccgttatac tggttgggga
1981 cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta
2041 gattttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat
2101 agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta
2161 catgaacata ttgcaaattt agctggtagc cctgctatta aaaaagggtat tttacagact
2221 gtaaaagtgt ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcggt
2281 attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt
2341 atgaaacgaa tcgaagaagg tatcaaaaga ttaggaagtc agattcttaa agagcatcct
2401 gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga
2461 gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac
2521 attgttccac aaagtttcct taaagacgat tcaatagaca ataaggctct aacgcgttct
2581 gataaaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa
2641 aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta
2701 acgaaagctg aacgtggagg tttgagtga cttgataaag ctggttttat caaacgcaa
2761 ttggttgaaa ctgcgcaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat
2821 actaaatacg atgaaaatga taaacttatt cgagagggtta aagtgattac cttaaaatct
2881 aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat
2941 taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa
3001 tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa
3061 atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct
3121 aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc
3181 cctctaactc aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt
3241 gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta
3301 cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt
3361 gctcgtaaaa aagactggga tccaaaaaaa tatgggtggt ttgatagtcc aacggtagct
3421 tattcagtc tagtggttgc taagggtgga aaagggaaat cgaagaagtt aaaatccgtt
3481 aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac
3541 tttttagaag ctaaaggata taaggaggtt aaaaaagact taatcattaa actacctaaa
3601 tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta
3661 caaaaaggaa atgagctggc tctgccaaagc aaatatgtga attttttata tttagctagt
3721 cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag
3781 cagcataagc attattttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt
3841 attttagcag atgccaatTT agataaagtt cttagtgc atatacaaca tagagacaaa
3901 ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttgagact
3961 cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa
4021 gaagtttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt

```

4081 gatttgagtc agctaggagg tgactga

By "Cpfl (CRISPR associated protein Cpfl)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to GenBank Accession No. AJ161 006.1 and having RNA binding activity, DNA binding activity, and/or DNA cleavage activity (e.g., endonuclease or nickase activity). An exemplary Cas9 polypeptide sequence is provided below.

```

1 msyqefvnk yslsktlrfe lipqgktlen ikarglildd ekrakdykka kqiidkyhqf
61 fieeilssvc isedllqnys dvyfklkksd ddnlqkdfks akdtikkqis eyikdsekfk
121 nlfngnlida kkggesdlil wlqgskdngi elfkansdit didealeiik sfkgwttyfk
181 gfhenrknavy ssndiptsii yrivddnlpk flenkakyes lkdkaeain yeqikkdlae
241 eltfdidykt sevnqrvfsl devfeianfn nylngsgitk fntiiggkfv ngentkrkgi
301 neyinlysqq indktlkkyk msvlfkqils dtesksfvid kleddsdrvvt tmqsfyeqia
361 afktveeksi ketlsllfdd lkaqkldlsk iyfkndkslt dlsqqvfddy svigtavley
421 itqqiapknl dnpskkeqel iakktekaky lsletiklal eefnkhrdid kqcrfeeila
481 nfaaipmifd eiaqnkdnl qisikyqnqg kkdllqasae ddvkaikdll dqtnnllhkl
541 kifhisqsed kanildkdeh fylvfeecyf elanivplyn kirnyitqkp ysdekfklnf
601 enstlangwd knkepdntai lfikddkyyl gvmnkknki fddkaikenk gegykkivyk
661 llpgankmlp kvffsaksik fynpsedilr irnhsthtkn gspqkgyekf efniedcrkf
721 idfykqsisk hpewkdfgfr fsdtqrynsl defyrevenq gykltfenis esyidsvvnq
781 gklylfqiyn kdfsayskgr pnlhtlywka lfdernlqdv vyklngael fyrkqsipkk
841 ithpakeaia nknkdnpkke svfeydlikd krftedkfff hcpitinfs sgankfnde
901 nlllkekand vhilsidrge rhllytldv gkgniikqdt fniigndrmk tnyhdklaai
961 ekdrdsarkd wkkinnikem kegylsqvvh eiaklvieyn aivvfedlnf gfkgrgfkve
1021 kqvyqklekm lieklnylvf kdnefdktgg vlayqltap fetfkkmqkq tgiyyvpag
1081 ftskicpvtg fvnqlypkye svsksqeffs kfdkicynld kgyfefsfdy knfgdkaakg
1141 kwtiasfgsr linfrnsdkn hnwdtrevyp tkelekllkd ysieyghgec ikaaicgesd
1201 kkffakltsv lntilqmrns ktgteldyli spvadvngnf fdsrqapknm pqdadangay
1261 higlkgmlml griknqegk klnlviknee yfefvqnrnn

```

"Cpfl function" can be defined by any of a number of assays including, but not limited to, fluorescence polarization-based nucleic acid bind assays, fluorescence polarization-based strand invasion assays, transcription assays, EGFP disruption assays, DNA cleavage assays, and/or Surveyor assays, for example, as described herein.

By "Cpfl nucleic acid molecule" is meant a polynucleotide encoding a Cpfl polypeptide or fragment thereof. An exemplary Cpfl nucleic acid molecule sequence is

provided at GenBank Accession No. CP009633, nucleotides 652838 - 656740 and is shown below.

```
1  atgtcaatztatcaagaatttggttaataaatatagtttaagtaaaactctaagatttgag
61  ttaatcccacagggtaaaacacttgaaaacataaaagcaagaggtttgattttagatgat
121 gagaaaagagctaaagactacaaaaaggctaaacaaataattgataaatatcatcagttt
181 tttatagaggagatattaagttcgggtttgtatttagcgaagatttattacaaaactattct
241 gatgtttattttaaacttaaaaagagtgatgatgataatctacaaaaagattttaaaagt
301 gcaaaagatacgataaagaacaaatatctgaatatataaaggactcagagaaatttaag
361 aatttgtttaatacaaaccttatcgatgctaaaaaagggaagagtcagatttaattcta
421 tggctaagcaatctaaggataatggtatagaactatttaaagccaatagtgatatcaca
481 gatatagatgaggcgtagaaataatcaaactcttttaaagggttgacaacttattttaag
541 gggtttcatgaaaatagaaaaatgtttatagtagcaatgatattcctacatctattatt
601 tataggatagtagatgataatttgcctaatttctagaaaaataaagctaagtatgagagt
661 ttaaaagacaaagctccagaagctataaactatgaacaaattaaaaaagatttggcagaa
721 gagctaacctttgatattgactacaaaacatctgaagttaatcaaagagtttttctactt
781 gatgaagtttttgagatagcaaacctttaataattatctaaatcaaagtggtattactaaa
841 ttttaactattatttggtggttaaatttgtaaatggtgaaaatacaaagagaaaaggata
901 aatgaatatataaatctatactcacagcaataaatgataaaacactcaaaaaatataaaa
961 atgagtgttttatttaagcaaattttaagtgtacagaatctaaatcttttgtaattgat
1021 aagttagaagatgatagtgatgtagttacaacgatgcaaagtttttatgagcaaatagca
1081 gctttttaaacagtagaagaaaaatctattaaagaaacactatctttattatttgatgat
1141 ttaaaagctcaaaaacttgatttgagtaaaatttattttaaaaatgataaatctcttact
1201 gatctatcacacaagtttttgatgattatagtggtatttggtacagcggtactagaatat
1261 ataactcaacaaatagcacctaaaaatcttgataaccctagtaagaaagagcaagaatta
1321 atagccaaaaaaactgaaaaagcaaaatacttatctctagaaactataaagcttgacctta
1381 gaagaatttaataagcatagagatatagataaacagtgtaggtttgaagaaatacttgca
1441 aactttgcggtctattccgatgatatttgatgaaatagctcaaaacaaagacaatttgga
1501 cagatatctatcaaatatcaaatcaaggtaaaaaagacctacttcaagctagtgcggaa
1561 gatgatgttaaagctatcaaggatcttttagatcaaactaataatctcttacataaacta
1621 aaaatatttcatatttagtcagtcagaagataaggcaaatattttagacaaggatgagcat
1681 ttttatctagtatttgaggagtgtactttgagctagcgaatatagtgacctctttataac
1741 aaaattagaaactatataactcaaaagccatatagtgatgagaaatttaagctcaatttt
1801 gagaactcgactttggctaatagggtgggataaaaaataaagagcctgacaatacggcaatt
1861 ttatttatcaaagatgataaatattatctgggtgtgatgaataagaaaaataacaaaata
1921 tttgatgataaagctatcaaaagaaaataaaggcgaggggtataaaaaaattgtttataaa
1981 cttttacctggcgcaataaaaatgttacctaagggttttcttttctgctaaatctataaaa
2041 ttttataatcctagtgaagatatacttagaataagaaatcattccacacatacaaaaaat
2101 ggtagtcctcaaaaaggatataaaaaatttgagtttaatatgaagattgccgaaaattt
```

```

2161 atagatTTTTtataaacagtctataagtaagcatccggagtggaaagattttggatttaga
2221 ttttctgatactcaaagatataattctatagatgaattttatagagaagttgaaaatcaa
2281 ggctacaaactaacttttgaaaatatatcagagagctatatattgatagcgtagttaatcag
2341 ggtaaattgtacctattccaaatctataataaagatttttcagcttatagcaaagggcga
2401 ccaaactacatacttttatattggaaagcgtgtttgatgagagaaatcttcaagatgtg
2461 gtttataagctaaatggtgaggcagagcttttttatcgtaaacaatcaatacctaaaaaa
2521 atcactcaccagctaaagaggcaatagctaataaaaaacaaagataatcctaaaaaagag
2581 agtggttttgaatatgatttaatacaagataaacgctttactgaagataagtttttcttt
2641 cactgtcctattacaatcaattttaaatctagtggagctaataagtttaatgatgaaatc
2701 aatttattgctaaaaagaaaaagcaaatgatgttcatatattaagtatagatagaggtgaa
2761 agacatttagcttactatactttggtagatggtaaaggcaatatcatcaacaagatact
2821 ttcaacatcattggtaatgatagaatgaaaacaaactacatgataagcttgctgcaata
2881 gagaaagatagggattcagctaggaaagactggaaaaagataaataacatcaaagagatg
2941 aaagagggctatctatctcaggtagttcatgaaatagctaagctagttatagagtataat
3001 gctattgtggTTTTtgaggatttaaattttggatttaaaagagggcgtttcaaggtagag
3061 aagcaggtctatcaaaagttagaaaaaatgctaattgagaaactaaactatctagttttc
3121 aaagataatgagtttgataaaactgggggagtgcttagagcttatcagctaacagcacct
3181 tttgagacttttataaaagatgggtaaacaaacaggtattatctactatgtaccagctggg
3241 tttacttcaaaaatttgcctgtaactggttttgtaaatcagttatatcctaagtatgaa
3301 agtgtcagcaaatctcaagagttcttttagtaagtttgacaagatttgttataaccttgat
3361 aagggctattttgagtttagttttgattataaaaacttttggtgacaaggctgccaaaggc
3421 aagtggactatagctagctttgggagtagattgattaacttttagaaattcagataaaaaat
3481 cataattgggatactcgagaagtttatccaactaaagagttggagaaattgctaaaagat
3541 tattctatcgaatatgggcatggcgaatgtatcaaagcagctatttgcggtgagagcgac
3601 aaaaagtttttgcctaagctaactagtgtcctaataactatcttacaatgcgtaactca
3661 aaaacaggtactgagtttagattatctaatttcaccagtagcagatgtaaatggcaatttc
3721 tttgattcgcgacaggcgccaaaaaatatgcctcaagatgctgatgccaatggtgcttat
3781 catattgggctaaaaggtctgatgctactaggtaggatcaaaaataatcaagagggcaaa
3841 aaactcaatttggttatcaaaaatgaagagtattttgagttcgtgcagaataggaataac
3901 taa

```

It will be understood that "substitution" or "substituted with" includes the implicit proviso that such substitution is in accordance with permitted valence of the substituted atom and the substituent, and that the substitution results in a stable compound, e.g., which does not spontaneously undergo transformation such as by rearrangement, cyclization, elimination, etc. In a broad aspect, the permissible substituents include acyclic and cyclic, branched and unbranched, carbocyclic and heterocyclic, aromatic and nonaromatic substituents of organic compounds. The permissible substituents can be one or more and the same or different for

appropriate organic compounds. For purposes of this disclosure, the heteroatoms such as nitrogen may have hydrogen substituents and/or any permissible substituents of organic compounds described herein which satisfy the valences of the heteroatoms. Non-limiting examples of optional substituents as referred to herein include halogen, alkyl, aralkyl, heteroalkyl, alkenyl, alkynyl, cycloalkyl, hydroxyl, alkoxy, amino, amido, nitro, cyano, amido, carbonyl, carboxyl, silyl, ether, alkylthio, sulfonyl, sulfonamido, ketone, aldehyde, ester, heterocyclyl, aryl, and heteroaryl.

A "substituted" hydrocarbon may have as a substituent one or more hydrocarbon radicals, substituted hydrocarbon radicals, or may comprise one or more heteroatoms. Examples of substituted hydrocarbon radicals include, without limitation, heterocycles, such as heteroaryls. Unless otherwise specified, a hydrocarbon substituted with one or more heteroatoms will comprise from 1-20 heteroatoms. In other embodiments, a hydrocarbon substituted with one or more heteroatoms will comprise from 1-12 or from 1-8 or from 1-6 or from 1-4 or from 1-3 or from 1-2 heteroatoms. Examples of heteroatoms include, but are not limited to, oxygen, nitrogen, sulfur, phosphorous, halogen (F, Cl, Br, I, etc.), boron, silicon, etc. In some embodiments, heteroatoms will be selected from the group consisting of oxygen, nitrogen, sulfur, phosphorous, and halogen (F, Cl, Br, I, etc.). In some embodiments, a heteroatom or group may substitute a carbon. In some embodiments, a heteroatom or group may substitute a hydrogen. In some embodiments, a substituted hydrocarbon may comprise one or more heteroatoms in the backbone or chain of the molecule (e.g., interposed between two carbon atoms, as in "oxa"). In some embodiments, a substituted hydrocarbon may comprise one or more heteroatoms pendant from the backbone or chain of the molecule (e.g., covalently bound to a carbon atom in the chain or backbone, as in "oxo").

In some embodiments, any hydrocarbon or substituted hydrocarbon disclosed herein may be substituted with one or more substituents X, where X is independently selected at each occurrence from one or more (e.g., 1-20) heteroatoms or one or more (e.g., 1-10) heteroatom-containing groups, where, for example, X may be selected from -F; -Cl; -Br; -I; -OH; -OR^{*}; -NH₂; -NHR^{*}; -N(R^{*})₂; -N(R^{*})₃⁺; -N(R^{*})-OH; -N(→O)(R^{*})₂; -O-N(R^{*})₂; -N(R^{*})-O-R^{*}; -N(R^{*})-N(R^{*})₂; -C=N-R^{*}; -N=C(R^{*})₂; -C=N-N(R^{*})₂; -C(=NR^{*})(-N(R^{*})₂); -C(H)(=N-OH); -SH; -SR^{*}; -CN; -NC; -C(=O)-R^{*}; -CHO; -C(=O)₂H; -C(=O)₂⁻; -C(=O)₂R^{*}; -C(=O)-S-R^{*}; -O-(C=O)-H; -O-(C=O)-R^{*}; -S-C(=O)-R^{*}; -(C=O)-NH₂; -C(=O)-N(R^{*})₂; -NH-(C=O)-R^{*}; -NH-(C=O)-R^{*}; -N(R^{*})-C(=O)-R^{*}; -C(=O)-NHNH₂; -O-C(=O)-NHNH₂; -C(=S)-NH₂; -(C=S)-N(R^{*})₂; -N(R^{*})-CHO; -N(R^{*})-

$C(=O)-R^*$; $-C(=NR^*)-O-R^*$; $-O-C(=NR^*)-R^*$; $-SCN$; $-NCS$; $-NSO$; $-SSR^*$; $-N(R^*)-C(=O)-N(R^*)_2$; $-N(R^*)-C(=S)-N(R^*)_2$; $-S(=O)_2-R^*$; $-O-S(=O)_2-R^*$; $-S(=O)_2-OR^*$; $-N(R^*)-S(=O)_2-R^*$; $-S(=O)_2-N(R^*)_2$; $-O-SO_3$; $-O-S(=O)_2-OR^*$; $-O-S(=O)-OR^*$; $-O-S(=O)-R^*$; $-S(=O)-OR^*$; $-S(=O)-R^*$; $-NO$; $-NO_2$; $-NO_3$; $-O-NO$; $-O-NO_2$; $-N_3$; $-N_2-R^*$; $-N(C_2H_4)$; $-Si(R^*)_3$; $-CF_3$; $-O-CF_3$; $-O-CH_3$; $-O-(CH_2)_{1-6}-CH_3$; $-PR^*_2$; $-O-P(=O)(OR^*)_2$; and $-P(=O)(OR^*)_2$; where, independently at each occurrence, R^* may be H or a **Ci**₁₋₁₀ or **Ci**_g or **Ci**₆ or **C**₁₋₄ hydrocarbon, including without limitation alkyl, alkenyl, alkynyl, aryl (e.g., phenyl), alkyl-aryl (e.g., benzyl), aryl-alkyl (e.g., toluyl), etc. In other embodiments, X may comprise a **Ci**-**Cg** or **C**₁₋₆ or **C**₂₋₄ perfluoroalkyl. In other embodiments, X may be a **Ci**-**Cg** or **C**₂₋₆ or **C**₃₋₅ heterocycle (e.g., heteroaryl radical). The term "halo" or "halogen" refers to any radical of fluorine, chlorine, bromine or iodine. In some embodiments, X is independently selected at each occurrence from $-OH$, $-SH$, $-NH_2$, $-N(R^*)_2$, $-F$, and $-Cl$.

In addition, the phrase "substituted with a[n]," as used herein, means the specified group may be substituted with one or more of any or all of the named substituents. For example, where a group, such as an alkyl or heteroaryl group, is "substituted with an unsubstituted **Ci**-**C**₂₀ alkyl, or unsubstituted 2 to 20 membered heteroalkyl," the group may contain one or more unsubstituted **Ci**-**C**₂₀ alkyls, and/or one or more unsubstituted 2 to 20 membered heteroalkyls. Moreover, where a moiety is substituted with an R substituent, the group may be referred to as "R-substituted." Where a moiety is R-substituted, the moiety is substituted with at least one R substituent and each R substituent is optionally different.

The term "alkyl" refers to a saturated hydrocarbon chain that may be a straight chain or branched chain, containing the indicated number of carbon atoms. For example, **C**₁₋₆ alkyl indicates that the group may have from 1 to 6 (inclusive) carbon atoms in it. Any atom can be optionally substituted, e.g., by one or more substituents. Examples of alkyl groups include without limitation methyl, ethyl, *n*-propyl, isopropyl, and *tert*-butyl.

As used herein, the term "straight chain **C**_{n-m} alkylene," employed alone or in combination with other terms, refers to a non-branched divalent alkyl linking group having n to m carbon atoms. Any atom can be optionally substituted, e.g., by one or more substituents. Examples include methylene (i.e., $-CH_2-$).

The term "haloalkyl" refers to an alkyl group, in which at least one hydrogen atom is replaced by halo. In some embodiments, more than one hydrogen atom (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14) are replaced by halo. In these embodiments, the hydrogen atoms can each be replaced by the same halogen (e.g., fluoro) or the hydrogen atoms can be

replaced by a combination of different halogens (e.g., fluoro and chloro). "Haloalkyl" also includes alkyl moieties in which all hydrogens have been replaced by halo (sometimes referred to herein as perhaloalkyl, e.g., perfluoroalkyl, such as trifluoromethyl). Any atom can be optionally substituted, e.g., by one or more substituents.

As referred to herein, the term "alkoxy" refers to a group of formula -O(alkyl). Alkoxy can be, for example, methoxy (-OCH₃), ethoxy, propoxy, isopropoxy, butoxy, isobutoxy, sec-butoxy, pentoxy, 2-pentoxy, 3-pentoxy, or hexyloxy. Likewise, the term "thioalkoxy" refers to a group of formula -S(alkyl). Finally, the terms "haloalkoxy" and "halothioalkoxy" refer to -O(haloalkyl) and -S(haloalkyl), respectively. The term "sulfhydryl" refers to -SH. As used herein, the term "hydroxyl," employed alone or in combination with other terms, refers to a group of formula -OH.

The term "aralkyl" refers to an alkyl moiety in which an alkyl hydrogen atom is replaced by an aryl group. One of the carbons of the alkyl moiety serves as the point of attachment of the aralkyl group to another moiety. Any ring or chain atom can be optionally substituted e.g., by one or more substituents. Non-limiting examples of "aralkyl" include benzyl, 2-phenylethyl, and 3-phenylpropyl groups.

The term "alkenyl" refers to a straight or branched hydrocarbon chain containing the indicated number of carbon atoms and having one or more carbon-carbon double bonds. Any atom can be optionally substituted, e.g., by one or more substituents. Alkenyl groups can include, e.g., vinyl, allyl, 1-butenyl, and 2-hexenyl. One of the double bond carbons can optionally be the point of attachment of the alkenyl substituent.

The term "alkynyl" refers to a straight or branched hydrocarbon chain containing the indicated number of carbon atoms and having one or more carbon-carbon triple bonds. Alkynyl groups can be optionally substituted, e.g., by one or more substituents. Alkynyl groups can include, e.g., ethynyl, propargyl, and 3-hexynyl. One of the triple bond carbons can optionally be the point of attachment of the alkynyl substituent.

The term "heterocyclyl" refers to a fully saturated monocyclic, bicyclic, tricyclic or other polycyclic ring system having one or more constituent heteroatom ring atoms independently selected from O, N (it is understood that one or two additional groups may be present to complete the nitrogen valence and/or form a salt), or S. The heteroatom or ring carbon can be the point of attachment of the heterocyclyl substituent to another moiety. Any atom can be optionally substituted, e.g., by one or more substituents. Heterocyclyl groups can include, e.g., tetrahydrofuryl, tetrahydropyranyl, piperidyl (piperidino), piperazinyl,

morpholinyl (morpholino), pyrrolinyl, and pyrrolidinyl. By way of example, the phrase "heterocyclic ring containing from 5-6 ring atoms, wherein from 1-2 of the ring atoms is independently selected from N, NH, N(Ci-C₆ alkyl), NC(=O)(Ci-C₆ alkyl), O, and S; and wherein said heterocyclic ring is optionally substituted with from 1-3 independently selected R^a" would include (but not be limited to) tetrahydrofuryl, tetrahydropyranyl, piperidyl (piperidino), piperazinyl, morpholinyl (morpholino), pyrrolinyl, and pyrrolidinyl.

The term "heterocycloalkenyl" refers to partially unsaturated monocyclic, bicyclic, tricyclic, or other polycyclic hydrocarbon groups having one or more (e.g., 1-4) heteroatom ring atoms independently selected from O, N (it is understood that one or two additional groups may be present to complete the nitrogen valence and/or form a salt), or S. A ring carbon (e.g., saturated or unsaturated) or heteroatom can be the point of attachment of the heterocycloalkenyl substituent. Any atom can be optionally substituted, e.g., by one or more substituents. Heterocycloalkenyl groups can include, e.g., dihydropyridyl, tetrahydropyridyl, dihydropyranyl, 4,5-dihydrooxazolyl, 4,5-dihydro-1H-imidazolyl, 1,2,5,6-tetrahydro-pyrimidinyl, and 5,6-dihydro-2H-[1,3]oxazinyl.

The term "cycloalkyl" refers to a fully saturated monocyclic, bicyclic, tricyclic, or other polycyclic hydrocarbon groups. Any atom can be optionally substituted, e.g., by one or more substituents. A ring carbon serves as the point of attachment of a cycloalkyl group to another moiety. Cycloalkyl moieties can include, e.g., cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl, adamantyl, and norbornyl (bicycle[2.2.1]heptyl).

The term "cycloalkenyl" refers to partially unsaturated monocyclic, bicyclic, tricyclic, or other polycyclic hydrocarbon groups. A ring carbon (e.g., saturated or unsaturated) is the point of attachment of the cycloalkenyl substituent. Any atom can be optionally substituted e.g., by one or more substituents. Cycloalkenyl moieties can include, e.g., cyclohexenyl, cyclohexadienyl, or norbornenyl.

As used herein, the term "cycloalkylene" refers to a divalent monocyclic cycloalkyl group having the indicated number of ring atoms.

As used herein, the term "heterocycloalkylene" refers to a divalent monocyclic heterocyclyl group having the indicated number of ring atoms.

The term "aryl" refers to an aromatic monocyclic, bicyclic (2 fused rings), or tricyclic (3 fused rings), or polycyclic (> 3 fused rings) hydrocarbon ring system. One or more ring atoms can be optionally substituted, e.g., by one or more substituents. Aryl moieties include, e.g., phenyl and naphthyl.

The term "heteroaryl" refers to an aromatic monocyclic, bicyclic (2 fused rings), tricyclic (3 fused rings), or polycyclic (> 3 fused rings) hydrocarbon groups having one or more heteroatom ring atoms independently selected from O, N (it is understood that one or two additional groups may be present to complete the nitrogen valence and/or form a salt), or S. One or more ring atoms can be optionally substituted, e.g., by one or more substituents. Examples of heteroaryl groups include, but are not limited to, 2H-pyrrolyl, 3H-indolyl, 4H-quinoliziny, acridinyl, benzo[b]thienyl, benzothiazolyl, β -carbolinyl, carbazolyl, coumarinyl, chromenyl, cinnolinyl, dibenzo[b,d]furanyl, furazanyl, furyl, imidazolyl, imidizolyl, indazolyl, indolyl, isobenzofuranyl, isoindolyl, isoquinolyl, isothiazolyl, isoxazolyl, naphthyridinyl, oxazolyl, perimidinyl, phenanthridinyl, phenanthrolinyl, phenarsazinyl, phenazinyl, phenothiazinyl, phenoxathiinyl, phenoxazinyl, phthalazinyl, pteridinyl, purinyl, pyranyl, pyrazinyl, pyrazolyl, pyridazinyl, pyridyl, pyrimidinyl, pyrrolyl, quinazoliny, quinolyl, quinoxalinyl, thiadiazolyl, thianthrenyl, thiazolyl, thienyl, triazolyl, and xanthenyl.

The terms "arylcycloalkyl" and "arylheterocyclyl" refer to bicyclic, tricyclic, or other polycyclic ring systems that include an aryl ring fused to a cycloalkyl and heterocyclyl, respectively. Similarly, the terms "heteroarylheterocyclyl," and "heteroarylcycloalkyl" refer to bicyclic, tricyclic, or other polycyclic ring systems that include a heteroaryl ring fused to a heterocyclyl and cycloalkyl, respectively. Any atom can be substituted, e.g., by one or more substituents. For example, arylcycloalkyl can include indanyl; arylheterocyclyl can include 2,3-dihydrobenzofuryl, 1,2,3,4-tetrahydroisoquinolyl, and 2,2-dimethylchromanyl.

The term "vicinal" refers to the configuration in which any two atoms or groups are, respectively, bonded to two adjacent atoms (*i.e.*, the two atoms are directly bonded to one another). The term "geminal" describes a configuration in which any atoms or two functional groups are bonded to the same atom. As used herein, when any two groups are said to together form a ring, unless otherwise indicated, it is meant that a bond is formed between each of said two groups, with the valences of the atoms appropriately adjusted to accommodate at least a bond (*e.g.*, a hydrogen atom may be removed from each group).

The descriptors "C=O" or "C(O)" or "carbonyl" refers to a carbon atom that is doubly bonded to an oxygen atom. "Alkyl carbonyl" has a common formula of R-C(O)- wherein R may be C₁₋₁₂ alkyl, C₂₋₁₂ alkenyl, C₂₋₁₂ alkynyl, C₃₋₁₂ cycloalkyl, C₆₋₁₂ aryl, C₃₋₁₂ heteroaryl, or C₃₋₁₂ heterocyclyl.

The term "oxo" refers to double bonded oxygen which can be a substituent on carbon or other atoms. When oxo is a substituent on nitrogen or sulfur, it is understood that the resultant groups has the structures $N \rightarrow O^-$ and $S(O)$ and SO_2 , respectively.

As used herein, the term "cyano," employed alone or in combination with other terms, refers to a group of formula $-CN$, wherein the carbon and nitrogen atoms are bound together by a triple bond. The term "azide" refers to a group of formula $-N_3$. The term "nitro" refers to a group of formula $-NO_2$. The term "amine" includes primary ($-NH_2$), secondary ($-NHR$), tertiary ($-NRR'$), and quaternary ($-N^+RR'R''$) amine having one, two or three independently selected substituents such as straight chain or branched chain alkyl, alkenyl, alkynyl, cycloalkyl, aryl, heteroaryl, heterocycle, and the like.

When any variable (*e.g.*, R^1) occurs more than one time in any constituent or formula for a compound, its definition at each occurrence is independent of its definition at every other occurrence. Thus, for example, if a group is shown to be substituted with one or more R^1 moieties, then R^1 at each occurrence is selected independently from the Markush group recited for R^1 . Also, combinations of substituents and/or variables are permissible, but only if such combinations result in stable compounds within a designated atom's normal valency.

As used herein, "unsaturated" refers to compounds or structures having at least one degree of unsaturation (*e.g.*, at least one double or triple bond).

The term "acid addition salt" refers to a salt of a compound prepared by reaction of a compound with a mineral or organic acid. For exemplification of pharmaceutically acceptable acid addition salts, see, *e.g.*, Berge, S.M., Bighley, L.D., and Monkhouse, D.C., *J. Pharm. Sci.*, 66: 1, 1977. For example, amine compounds are basic in nature and accordingly react with any of a number of inorganic and organic acids to form pharmaceutically acceptable acid addition salts.

Pharmaceutically acceptable acid addition salts can be formed by the reaction of a disclosed compound with an equimolar or excess amount of acid. Alternatively, hemi-salts can be formed by the reaction of a compound with the desired acid in a 2:1 ratio, compound to acid. The reactants are generally combined in a mutual solvent such as diethyl ether, tetrahydrofuran, methanol, ethanol, *iso*-propanol, benzene, or the like. The salts normally precipitate out of solution within, *e.g.*, about one hour to about ten days and can be isolated by filtration or other conventional methods.

Inorganic acids commonly employed to form such salts include hydrochloric acid, hydrobromic acid, hydroiodic acid, sulfuric acid, phosphoric acid, and the like. Organic acids

commonly employed to form such salts include p-toluenesulfonic acid, methanesulfonic acid, oxalic acid, p-bromophenylsulfonic acid, carbonic acid, succinic acid, citric acid, benzoic acid, acetic acid and the like. Examples of such pharmaceutically acceptable salts thus are the sulfate, pyrosulfate, bisulfate, sulfite, bisulfite, phosphate, monohydrogenphosphate, dihydrogenphosphate, metaphosphate, pyrophosphate, chloride, bromide, iodide, acetate, propionate, decanoate, caprylate, acrylate, formate, zso-butyrates, caproate, heptanoate, propiolate, oxalate, malonate, succinate, hemisuccinate, suberate, sebacate, fumarate, maleate, butyne-1,4-dioate, hexyne-1,6-dioate, benzoate, chlorobenzoate, methylbenzoate, dinitrobenzoate, hydroxybenzoate, methoxybenzoate, phthalate, sulfonate, xylenesulfonate, phenylacetate, phenylpropionate, phenylbutyrate, citrate, lactate, β -hydroxybutyrate, glycolate, tartrate, methanesulfonate, propanesulfonate, naphthalene-1-sulfonate, naphthalene-2-sulfonate, mandalate and the like.

By "agent" is meant any small molecule chemical compound, antibody, nucleic acid molecule, or polypeptide, or fragments thereof.

By "decreases" is meant a negative alteration. Such alterations are by 5%, 10%, 25%, 50%, 75%, 85%, 90% or even by 100% of a reference value.

By "effective amount" is meant the amount of an agent required to ameliorate the symptoms of a condition, disease or disorder relative to an untreated subject or organism. In particular embodiments, the subject or organism expresses an active RNA guided endonuclease polypeptide. The effective amount of active agent(s) used to practice the present invention varies depending upon the manner of administration, the age, body weight, and general health of the subject.

By "fragment" is meant a portion of a polypeptide or nucleic acid molecule. This portion contains, preferably, at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the entire length of the reference nucleic acid molecule or polypeptide. A fragment may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or amino acids.

By "identity" is meant the amino acid or nucleic acid sequence identity between a sequence of interest and a reference sequence. Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of

homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

By "hybridize" is meant pair to form a double-stranded molecule between complementary polynucleotide sequences, or portions thereof, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507).

For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C. in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C. in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C. in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably

less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C, more preferably of at least about 42° C, and most preferably of at least about 68° C. In a preferred embodiment, wash steps will occur at 25° C. in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C. in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68° C. in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art.

Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel et al. (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

By "increases" is meant a positive alteration. Such alterations are by 5%, 10%, 25%, 50%, 75%, 85%, 90% or even by 100% of a reference value.

In some aspects, the compound is an isomer. "Isomers" are different compounds that have the same molecular formula. "Stereoisomers" are isomers that differ only in the way the atoms are arranged in space. As used herein, the term "isomer" includes any and all geometric isomers and stereoisomers. For example, "isomers" include geometric double bond cis- and trans-isomers, also termed E- and Z-isomers; R- and S-enantiomers; diastereomers, (d)-isomers and (l)-isomers, racemic mixtures thereof; and other mixtures thereof, as falling within the scope of this disclosure.

Geometric isomers can be represented by the symbol ---- which denotes a bond that can be a single, double or triple bond as described herein. Provided herein are various geometric isomers and mixtures thereof resulting from the arrangement of substituents around a carbon-carbon double bond or arrangement of substituents around a carbocyclic ring. Substituents around a carbon-carbon double bond are designated as being in the "Z" or "E" configuration wherein the terms "Z" and "E" are used in accordance with IUPAC standards. Unless otherwise specified, structures depicting double bonds encompass both the "E" and "Z" isomers.

Substituents around a carbon-carbon double bond alternatively can be referred to as "cis" or "trans," where "cis" represents substituents on the same side of the double bond and

"trans" represents substituents on opposite sides of the double bond. The arrangement of substituents around a carbocyclic ring can also be designated as "cis" or "trans." The term "cis" represents substituents on the same side of the plane of the ring, and the term "trans" represents substituents on opposite sides of the plane of the ring. Mixtures of compounds wherein the substituents are disposed on both the same and opposite sides of plane of the ring are designated "cis/trans."

The term "enantiomers" refers to a pair of stereoisomers that are non-superimposable mirror images of each other. An atom having an asymmetric set of substituents can give rise to an enantiomer. A mixture of a pair of enantiomers in any proportion can be known as a "racemic" mixture. The term "(±)" is used to designate a racemic mixture where appropriate. "Diastereoisomers" are stereoisomers that have at least two asymmetric atoms, but which are not mirror-images of each other. The absolute stereochemistry is specified according to the Cahn-Ingold-Prelog R-S system. When a compound is an enantiomer, the stereochemistry at each chiral carbon can be specified by either R or S. Resolved compounds whose absolute configuration is unknown can be designated (+) or (-) depending on the direction (dextro- or levorotatory) which they rotate plane polarized light at the wavelength of the sodium D line. Certain of the compounds described herein contain one or more asymmetric centers and can thus give rise to enantiomers, diastereomers, and other stereoisomeric forms that can be defined, in terms of absolute stereochemistry at each asymmetric atom, as (R)- or (S)-. The present chemical entities, pharmaceutical compositions and methods are meant to include all such possible isomers, including racemic mixtures, optically substantially pure forms and intermediate mixtures. In some chemical structures, stereocenters may be identified with "wavy" bonds indicating that the stereocenter may be in the R or S configuration, unless otherwise specified. However, stereocenters without a wavy bond (*i.e.*, a "straight" bond) may also be in the (R) or (S) configuration, unless otherwise specified. Compositions comprising compounds may comprise stereocenters which each may independently be in the (R) configuration, the (S) configuration, or racemic mixtures.

Optically active (R)- and (S)-isomers can be prepared, for example, using chiral synthons or chiral reagents, or resolved using conventional techniques. Enantiomers can be isolated from racemic mixtures by any method known to those skilled in the art, including chiral high pressure liquid chromatography (HPLC), the formation and crystallization of chiral salts, or prepared by asymmetric syntheses.

Optical isomers can be obtained by resolution of the racemic mixtures according to conventional processes, e.g., by formation of diastereoisomeric salts, by treatment with an optically active acid or base. Examples of appropriate acids are tartaric, diacetyltartaric, dibenzoyltartaric, ditoluoyltartaric, and camphorsulfonic acid. The separation of the mixture of diastereoisomers by crystallization followed by liberation of the optically active bases from these salts affords separation of the isomers. Another method involves synthesis of covalent diastereoisomeric molecules by reacting disclosed compounds with an optically pure acid in an activated form or an optically pure isocyanate. The synthesized diastereoisomers can be separated by conventional means such as chromatography, distillation, crystallization or sublimation, and then hydrolyzed to deliver the enantiomerically enriched compound. Optically active compounds can also be obtained by using active starting materials. In some embodiments, these isomers can be in the form of a free acid, a free base, an ester or a salt.

In certain embodiments, a disclosed compound can be a tautomer. As used herein, the term "tautomer" is a type of isomer that includes two or more interconvertible compounds resulting from at least one formal migration of a hydrogen atom and at least one change in valency (e.g., a single bond to a double bond, a triple bond to a single bond, or vice versa). "Tautomerization" includes prototropic or proton-shift tautomerization, which is considered a subset of acid-base chemistry. "Prototropic tautomerization" or "proton-shift tautomerization" involves the migration of a proton accompanied by changes in bond order. The exact ratio of the tautomers depends on several factors, including temperature, solvent, and pH. Where tautomerization is possible (e.g., in solution), a chemical equilibrium of tautomers can be reached. Tautomerizations (i.e., the reaction providing a tautomeric pair) can be catalyzed by acid or base, or can occur without the action or presence of an external agent. Exemplary tautomerizations include, but are not limited to, keto-to-enol; amide-to-imide; lactam-to-lactim; enamine-to-imine; and enamine-to-(a different) enamine tautomerizations. A specific example of keto-enol tautomerization is the interconversion of pentane-2,4-dione and 4-hydroxypent-3-en-2-one tautomers. Another example of tautomerization is phenol-keto tautomerization. A specific example of phenol-keto tautomerization is the interconversion of pyridin-4-ol and pyridin-4(1H)-one tautomers.

All chiral, diastereomeric, racemic, and geometric isomeric forms of a structure are intended, unless specific stereochemistry or isomeric form is specifically indicated. All processes used to prepare compounds and intermediates made therein are encompassed by the

present disclosure. All tautomers of shown or described compounds are also encompassed by the present disclosure.

The terms "pharmaceutical" or "pharmaceutically acceptable", when used herein as an adjective, mean substantially non-toxic and substantially non-deleterious to the subject.

As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The pharmaceutically acceptable carrier or excipient does not destroy the pharmacological activity of the disclosed compound and is nontoxic when administered in doses sufficient to deliver a therapeutic amount of the compound. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions as disclosed herein is contemplated. Non-limiting examples of pharmaceutically acceptable carriers and excipients include sugars such as lactose, glucose and sucrose; starches such as corn starch and potato starch; cellulose and its derivatives such as sodium carboxymethyl cellulose, ethyl cellulose and cellulose acetate; powdered tragacanth; malt; gelatin; talc; cocoa butter and suppository waxes; oils such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; glycols, such as polyethylene glycol and propylene glycol; esters such as ethyl oleate and ethyl laurate; agar; buffering agents such as magnesium hydroxide and aluminum hydroxide; alginic acid; isotonic saline; Ringer's solution; ethyl alcohol; phosphate buffer solutions; non-toxic compatible lubricants such as sodium lauryl sulfate and magnesium stearate; coloring agents; releasing agents; coating agents; sweetening, flavoring and perfuming agents; preservatives; antioxidants; ion exchangers; alumina; aluminum stearate; lecithin; self-emulsifying drug delivery systems (SEDDS) such as d- α -tocopherol polyethyleneglycol 1000 succinate; surfactants used in pharmaceutical dosage forms such as Tweens or other similar polymeric delivery matrices; serum proteins such as human serum albumin; glycine; sorbic acid; potassium sorbate; partial glyceride mixtures of saturated vegetable fatty acids; water, salts or electrolytes such as protamine sulfate, disodium hydrogen phosphate, potassium hydrogen phosphate, sodium chloride, and zinc salts; colloidal silica; magnesium trisilicate; polyvinyl pyrrolidone; cellulose-based substances; polyacrylates; waxes; and polyethylene-polyoxypropylene-block polymers. Cyclodextrins such as α -, β -, and γ -cyclodextrin, or chemically modified derivatives such as hydroxyalkylcyclodextrins, including 2- and 3-

hydroxypropyl-cyclodextrins, or other solubilized derivatives can also be used to enhance delivery of compounds described herein.

By "pharmaceutical formulation" it is further meant that the carrier, solvent, excipient(s) and/or salt must be compatible with the active ingredient of the formulation (e.g. a disclosed compound). It is understood by those of ordinary skill in this art that the terms "pharmaceutical formulation" and "pharmaceutical composition" are generally interchangeable, and they are so used for the purposes of this application.

By "prodrug" is meant any compound that must undergo bioactivation before exhibiting its intended pharmacological effects. Since prodrugs are known to enhance numerous desirable qualities of pharmaceuticals (e.g., solubility, bioavailability, manufacturing, etc.), the disclosed compounds can be delivered in prodrug form. Thus, both prodrugs of the compounds, methods of delivering the same and compositions containing the same are disclosed herein. Prodrugs are intended to include any covalently bonded carriers that release an active parent drug (compound) *in vivo* when such prodrug is administered to a subject. Prodrugs are prepared, for example, by modifying functional groups present in the compound in such a way that the modifications are cleaved, either in routine manipulation or *in vivo*, to the parent compound. Prodrugs include compounds wherein a hydroxyl or amino group is bonded to any group that, when the prodrug is administered to a subject, it cleaves to form a free hydroxyl or free amino group, respectively. Examples of prodrugs include, but are not limited to, acetate, formate, and benzoate derivatives of alcohol and amine functional groups in the disclosed compounds. Examples of prodrugs include, but are not limited to, benzamide derivatives of an amine functional group in the active compound and the like. Other examples of prodrugs include compounds that comprise -NO, -NO₂, -ONO, or -ONO₂ moieties.

For example, if a disclosed compound or a pharmaceutically acceptable form of the compound contains a carboxylic acid functional group, a prodrug can comprise a pharmaceutically acceptable ester formed by the replacement of the hydrogen atom of the acid group with a group such as (Ci-g)alkyl, (Ci-i2)alkanoyloxymethyl, 1- (alkanoyloxy) ethyl having from 4 to 9 carbon atoms, 1-methyl- 1-(alkanoyloxy)-ethyl having from 5 to 10 carbon atoms, alkoxycarbonyloxymethyl having from 3 to 6 carbon atoms, 1- (alkoxycarbonyloxy)ethyl having from 4 to 7 carbon atoms, 1-methyl- 1- (alkoxycarbonyloxy)ethyl having from 5 to 10 carbon atoms, N- (alkoxycarbonyl)aminomethyl having from 3 to 9 carbon atoms, 1-(N-

(alkoxycarbonyl)amino)ethyl having from 4 to 10 carbon atoms, 3-phthalidyl, 4-crotonolactonyl, gamma-butyrolacton-4-yl, di-N,N-(Ci-2)alkylamino(C2-3)alkyl (such as [3-dimethylaminoethyl), carbamoyl-(Ci-2)alkyl, N,N-di(Ci-2)alkylcarbamoyl-(Ci-2)alkyl and piperidino-, pyrrolidino- or morpholino(C2-3)alkyl.

Similarly, if a disclosed compound or a pharmaceutically acceptable form of the compound contains an alcohol functional group, a prodrug can be formed by the replacement of the hydrogen atom of the alcohol group with a group such as (Ci-6)alkanoyloxymethyl, 1-((Ci-6)alkanoyloxy)ethyl, 1-methyl-1-((Ci-6)alkanoyloxy)ethyl, (Ci-6)alkoxycarbonyloxymethyl, N-(Ci-6)alkoxycarbonylaminomethyl, succinoyl, (Ci-6)alkanoyl, a-amino(Ci-4)alkanoyl, arylacyl, and a-aminoacyl, or a-aminoacyl-a-aminoacyl, where each a-aminoacyl group is independently selected from the naturally occurring L-amino acids, -P(0)(OH)₂, -P(0)(0(Ci-6)alkyl)₂ or glycosyl (the radical resulting from the removal of a hydroxyl group of the hemiacetal form of a carbohydrate).

If a disclosed compound or a pharmaceutically acceptable form of the compound incorporates an amine functional group, a prodrug can be formed by the replacement of a hydrogen atom in the amine group with a group such as R-carbonyl, RO-carbonyl, NRR'-carbonyl where R and R' are each independently (Ci-10)alkyl, (C₃₋₇)cycloalkyl, benzyl, a natural α-aminoacyl or natural a-aminoacyl-natural-a-aminoacyl, -C(OH)C(0)OY¹ wherein Y¹ is H, (Ci-6)alkyl or benzyl, C(OY²)Y³ wherein Y² is (Ci-4)alkyl and Y³ is (Ci-6)alkyl, carboxy(Ci-6)alkyl, amino(Ci-4)alkyl or mono-N- or di-N,N-(Ci-6)alkylaminoalkyl, -C(Y⁴)Y⁵ wherein Y⁴ is H or methyl and Y⁵ is mono-N- or di-N,-(Ci-6)alkylamino, morpholino, piperidin-1-yl or pyrrolidin-1-yl.

In certain cases, a prodrug has improved physical and/or delivery properties over the parent compound. Prodrugs can increase the bioavailability of the compound when administered to a subject (e.g., by permitting enhanced absorption into the blood following oral administration) or which enhance delivery to a biological compartment of interest (e.g., the brain or lymphatic system) relative to the parent compound. Exemplary prodrugs include derivatives of a disclosed compound with enhanced aqueous solubility or active transport through the gut membrane, relative to the parent compound.

By "reference" is meant a standard or control condition.

A "reference sequence" is a defined sequence used as a basis for sequence comparison. In one embodiment, the reference sequence is Cas9.

By "small molecule" is meant any chemical compound.

The terms "selecting a subject" or "identifying a subject" are understood as choosing one or more members of a mixed population of individuals based on specific characteristics including, but not limited to, physical symptoms, and/or clinical characteristics as determined by diagnostic methods.

The term "solvate" means a solvent addition form that contains either a stoichiometric or non- stoichiometric amount of solvent. Some compounds have a tendency to trap a fixed molar ratio of solvent molecules in a solid state, thus forming a solvate. If the solvent is water, the solvate formed is a hydrate; when the solvent is alcohol, the solvate formed is an alcoholate. Hydrates are formed by the combination of one or more molecules of water with one of the substances in which the water retains its molecular state as H_2O , such combination being able to form one or more hydrates.

The term "suitable solvent" refers to any solvent, or mixture of solvents, that may be inert to the ongoing reaction that sufficiently solubilizes the reactants to afford a medium within which to effect the desired reaction.

By "specifically binds" is meant recognizes and binds a polynucleotide or polypeptide of the invention, but which does not substantially recognize and bind other molecules in a sample.

By "subject" is meant an organism, including, but not limited to, a human or non-human mammal, such as a bovine, equine, canine, ovine, or feline; vertebrate; invertebrate, such as an insect; or plant; or any commercially relevant organism.

As used herein, the terms "treat," "treating," "treatment," and the like refer to reducing or ameliorating a disorder and/or symptoms associated therewith. It will be appreciated that, although not precluded, treating a disorder or condition does not require that the disorder, condition or symptoms associated therewith be completely eliminated.

As used herein, the terms "prevent," "preventing," "prevention," "prophylactic treatment" and the like refer to reducing the probability of developing a disorder or condition in a subject, who does not have, but is at risk of or susceptible to developing a disorder or condition.

Unless specifically stated or obvious from context, as used herein, the term "about" is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01 % of the stated value. Unless otherwise clear from context, all numerical values provided herein are modified by the term about.

Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

Any compounds, compositions, or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein.

Unless specifically stated or obvious from context, as used herein, the term "or" is understood to be inclusive. Unless specifically stated or obvious from context, as used herein, the terms "a", "an", and "the" are understood to be singular or plural. Thus, for example, reference to "an amino acid substitution" includes reference to more than one amino acid substitution.

The term "including" is used herein to mean, and is used interchangeably with, the phrase "including but not limited to."

As used herein, the terms "comprises," "comprising," "containing," "having" and the like can have the meaning ascribed to them in U.S. Patent law and can mean "includes," "including," and the like; "consisting essentially of" or "consists essentially" likewise has the meaning ascribed in U.S. Patent law and the term is open-ended, allowing for the presence of more than that which is recited so long as basic or novel characteristics of that which is recited is not changed by the presence of more than that which is recited, but excludes prior art embodiments.

Other features and advantages of the invention will be apparent from the following description of the desirable embodiments thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A is a schematic depicting a binding assay that measures Cas9 complex binding to PAM-rich DNA (FP-assay). Fluorescence polarization (FP) increases when [Cas9:guideRNA] complex binds to PAM-rich DNA.

FIG. 1B is a graph showing a dose-dependent increase in FP signal with increasing equivalents of [Cas9:guideRNA] complex relative to those of 12 PAM-DNA.

FIG. 1C is a graph showing results of competition experiments using DNA sequences with varying numbers of PAM-sites. A decrease in FP signal was observed as the number of PAM-sites increased on the competitor.

FIG. 1D is a graph showing that thermal stability of [Cas9:guideRNA] complex increased as PAM-sites increased on the interacting DNA.

FIG. 1E is a schematic of a fluorescence-based strand displacement assay for monitoring Cas9 nuclease activity. Following Cas9 cleavage, a fluorophore bearing double stranded oligo (DS-oligo) is displaced by a quencher (Q)-bearing displacer strand (Q-oligo), resulting in a decrease in fluorescent signal.

FIG. 1F depicts gel-monitored cleavage of fluorophore labeled oligos. Fluorophore labeled oligos (100 nM) were cleaved by SpCas9 (500 nM) in a PAM-dependent manner. Gel is representative of 2 biological replicates.

FIG. 1G is a graph showing that DS-oligo fluorescence is not quenched in the presence of Q-oligo unless the duplex is disrupted by cleavage via an active Cas9:gRNA complex. A single DNA strand with fluorophore (SS-Oligo) can be completely quenched by the Q-oligo in the absence of a duplex. Error bars represent standard deviation from 3 technical replicates (n=3), and is representative of 2 biological replicates.

FIG. 1H is a graph showing that quenching via strand displacement is dependent on the presence of a NGG PAM in the DS-oligo when using SpCas9, indicating the specificity of the interaction. Error bars represent standard deviation from 3 technical replicates (n=3), and is representative of 2 biological replicates.

FIG. 1I is a graph showing that strand displacement is generalizable to SaCas9. SaCas9 strand displacement had comparable efficiency to SpCas9, and was dependent on an NNGGGT PAM sequence. Error bars represent standard deviation from 3 technical replicates (n=3), and is representative of 2 biological replicates.

FIG. 1J is a graph depicting optimization of the relative ratio of the SpCas9:gRNA complex (1 - 200 nM) to DS-oligo (fixed at 1 nM) while holding the Q-oligo concentration fixed (5 nM). Using a 5-fold excess of SpCas9:gRNA maximizes activity while minimizing background quenching from SpCas9 simply binding to DNA. Data is presented as the average background-subtracted fluorescence from 3 technical replicates. Error bars represent standard deviation (n=3).

FIG. 1K is a graph depicting optimization of the relative amounts of Q-oligo (1 - 200 nM) and DS-oligo (fixed at 1 nM) while holding the SpCas9:gRNA concentration fixed (5 nM). A 2-fold excess of Q-oligo is sufficient to displace the cut strand. Data is presented as the average background-subtracted fluorescence from 3 technical replicates. Error bars represent standard deviation (n=3).

FIG. 1L is a graph depicting a determination of the DS-oligo limit of detection, fixing [SpCas9] and [Q-oligo] at 5-fold relative amount of DS-oligo and conducting the reaction for 120 min. Data is presented as the average background-subtracted fluorescence from 3 technical replicates, and is representative of 2 biological replicates. Error bars represent standard deviation (n=3). Inset is enlarged view of the 1, 0.3, and 0.1 nM points.

FIG. 1M is a graph depicting a time course of strand displacement, fixing [SpCas9] and [Q-oligo] at 5-fold relative amount of DS-oligo (1 nM). Reactions were incubated at either 25°C or 37°C. Data is presented as fraction with 3 technical replicates, and is representative of 2 biological replicates. Error bars represent standard deviation (n=3).

FIG. 2A is a plot of normalized FP values for two screening replicates. Hit compounds identified by the FP-assay are circled.

FIG. 2B depicts that stereoisomers of spirocyclicazetidine hit compounds have different IC₅₀ in the FP-assay.

FIG. 2C is a graph depicting dose-dependent reduction in FP-signal by the compound BRD3033.

FIG. 2D is a graph depicting dose-dependent reduction in FP-signal by the compound BRD4172.

FIG. 2E is DNA cleavage assay shows inhibition of SpCas9 activity by BRD3033 and BRD4172.

FIG. 2F show dose-dependence inhibition of SpCas9 in DNA cleavage assay by BRD 3033 and BRD4172.

FIG. 3A is a schematic of a spinach-based *in vitro* transcription assay for monitoring Cas9 nuclease activity. In absence of Cas9, T7 RNA polymerase is recruited to a T7 promoter-containing DNA template to transcribe the spinach RNA aptamer, which can bind to the fluorogenic molecule DFHBI. Cleavage of the DNA by Cas9 results in complete termination of transcription or production of unproductive RNA, resulting in loss of fluorescence. Cas9 can recognize PAM sites native to the T7 and spinach sequences, or variable PAMs proximal and distal to the T7 promoter.

FIG. 3B is a schematic of the DNA template, detailing gRNA sites, both engineered and native.

FIG. 3C is a graph depicting validation of the Spinach-assay.

FIG. 3D is a graph depicting that SpCas9-mediated knockdown of spinach gene led to loss of fluorescence.

FIG. 3E is a graph showing that SpCas9:gRNA targeting site Sp g-2 caused loss of spinach fluorescence in a dose-dependent manner. ApoCas9 at 5 nM did not result in cleavage, indicating that this loss is due to cleavage of the spinach DNA template. Error bars represent the standard deviation from n=3 technical replicates.

FIG. 3F is a graph showing that SpCas9:gRNA-mediated fluorescence loss was dependent on the position of the gRNA, with PAM sites closer to the T7 promoter (in order: Sp g-2, g-3, g-4, and g-5) being more efficient. ApoCas9 at 2 nM did not result in cleavage. Error bars represent the standard deviation from n=3 technical replicates.

FIG. 3G is a graph showing Cas nuclease-mediated inhibition of IVT in SaCas9. Active SaCas9:gRNA (5 nM) can be used at both an endogenous PAM site (Sa g-1) and an installed GGGT proximal PAM site (Sa g-2). ApoSaCas9 (5 nM) did not result in cleavage. Error bars represent the standard deviation from n=3 technical replicates.

FIG. 3H is a set of graphs showing Cas nuclease-mediated inhibition of IVT in AsCpfI. Active AsCpfI: gRNA can cleave an installed distal TTTC PAM site (CpfI gRNA-1; upper) or native TTTC site (CpfI gRNA-2; lower) in a dose dependent manner, albeit with lower efficiency compared to other tested Cas nucleases. Error bars represent the standard deviation from n=3 technical replicates.

FIG. 3I is a set of graphs showing Cas nuclease-mediated inhibition of IVT in LbCpfI. Active LbCpfI:gRNA can cleave an installed distal TTTC PAM site (CpfI gRNA-1; upper) or native TTTC site (CpfI gRNA-2; lower) in a dose dependent manner, albeit with lower efficiency compared to other tested Cas nucleases. Error bars represent the standard deviation from n=3 technical replicates.

FIG. 3J is a set of graphs showing Cas nuclease-mediated inhibition of IVT in FnCpfI. Active FnCpfI:gRNA can cleave an installed distal TTTC PAM site (CpfI gRNA-1; upper) or native TTTC site (CpfI gRNA-2; lower) in a dose dependent manner, albeit with lower efficiency compared to other tested Cas nucleases. Error bars represent the standard deviation from n=3 technical replicates.

FIG. 3K is a schematic depicting an EGFP-disruption assay in U2OS.EGFP cells.

FIG. 3L depicts the results of an EGFP knockdown experiment, showing images corresponding to Control EGFP-disruption assay (left panel); EGFP-disruption assay - Cas9

(center panel); and EGFP-disruption assay - Cas9 + BRD2940. EGFP knockdown using SpCas9 led to loss of fluorescence, which was recovered by the compound BRD2940.

FIG. 3M is a graph showing testing of compounds identified by the FP-assay in the EGFP-assay

FIG. 3N depicts the results of a Surveyor assay in the absence (left lanes) or presence (right lanes) of compounds identified by the FP-assay. The Surveyor fragment was absent in compound treated cells but present in control assays. The results of the Surveyor assay confirmed Cas9 inhibition by the FP-assay hit compounds. Exemplary data are shown of assay performed in replicate.

FIG. 3O depicts selected compound libraries identified as having Cas9 inhibitory activity. The selected compound libraries are used to optimize the potency of Cas9 inhibitors. Number of compounds available in each library is shown in parenthesis.

FIG. 3P is a workflow for optimization of the potency of the Cas9 inhibitors. Workflows comprise one or more of the Fluorescence polarization (FP) assay; a counterscreen to the FP assay to identify false positive results; Spinach assay; EGFP disruption assay; DNA cleavage assay and/or Surveyor assay.

FIG. 3Q are images depicting surface-show of the binding pose for two selected Cas9 inhibitors with a Pictet-Spengler scaffold determined by Glide docking.

FIG. 3R are images depicting ribbon-show of the binding pose for two selected Cas9 inhibitors with a Pictet-Spengler scaffold determined by Glide docking.

FIG. 4A is an image of an SDS-PAGE gel showing purified SaCas9 and SpCas9.

FIG. 4B is a schematic depicting a fluorescence polarization-based strand invasion assay to monitor nuclease activity of Cas9.

FIG. 4C is a schematic depicting measurement of Cas9 nuclease activity by strand invasion.

FIG. 4D is a graph depicting optimization of Cas9: Substrate ratio in the fluorescence polarization-based strand invasion assay. Data columns corresponding to the 12PAM DNA template were included as positive control.

FIG. 4E is a graph depicting optimization of Cas9: Substrate ratio in the fluorescence polarization-based strand invasion assay.

FIG. 4F is a graph depicting fluorescence polarization (FP) in the Cas9 strand invasion assay. The results validate the strand invasion experiment.

FIG. 4G is a graph depicting kinetics of strand invasion in the Cas9 strand invasion assay. The results validate the strand invasion experiment.

FIG. 5A is a schematic depicting a fluorescence quencher based Cas9 strand invasion assay.

FIG. 5B is a graph showing Cas9-DNA binding in the fluorescence quencher based Cas9 strand invasion assay.

FIG. 5C is a graph showing that the 3'-quencher-DNA strand efficiently diminished FAM fluorescence only in the ss-non-target DNA in a ratiometric fashion while there was no effect on the c³/4-target DNA in the fluorescence quencher based Cas9 strand invasion assay.

FIG. 5D is a graph showing systematically enhanced readout was observed with increasing quencher concentration in different Cas9 compositions in the fluorescence quencher based Cas9 strand invasion assay.

FIG. 5E depicts gel electrophoresis results of the fluorescence quencher based Cas9 strand invasion assay showing that DNA fragments were generated in the presence of Cas9.

FIG. 6A depicts the chemical structure of hit compounds BRD-K33182911-001-01-8 (BRD2911); BRD-K36151368-001-01-8 (BRD1368); BRD-K86027682-001-01-8 (BRD7682); and BRD-K96745813-001-01-5 (BRD5813).

FIG. 6B is a graph depicting compound mediated inhibition of Cas9 activity in egfp-U2OS cells (48h treatment) for the compounds shown in FIG. 5A.

FIG. 6C are cell images for compound mediated inhibition of Cas9 activity in egfp-U2OS cells (48h treatment) for the compounds shown in FIG. 5A.

FIG. 6D are images showing compound mediated inhibition of Cas9 activity in HEK293T cells (48h treatment) as measured by Surveyor assay for the compounds shown in FIG. 5A.

FIG. 7A is a graphical listing of Cas9 residues (PDB structure 4008) found to interact with gRNA and DNA via salt bridges, hydrogen bonds, or hydrophobic/stacking interactions.

FIG. 7B is a graph representing the number of Cas9: DNA (top) and Cas9:gRNA (bottom) interactions as a function of base location.

FIG. 7C is a graph depicting experimental cleavage efficiencies measured for single base pair mismatches with DNA at each gRNA base location (averaged over the 3 mismatches at each location) relative to perfectly matched DNA. The tolerance for mismatches at each base location was inversely correlated to the number of Cas9:gRNA

interactions at each base location, determined from the StCas9 and two SpCas9 crystal structures.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides compositions and methods for inhibiting the activity of RNA guided endonucleases (e.g., Cas9, Cpf1), and methods of use therefore, as well as to inhibit or prevent Cas9 genome editing. The invention is based, at least in part, on the discovery of small molecule inhibitors of RNA guided endonucleases. As described herein, high-throughput biochemical and cellular assays, and workflows comprising combinations of such assays, were developed for screening and identifying small molecules with the ability to inhibit one or more activities of RNA guided endonucleases. Methods involving small molecule inhibitors of RNA guided endonucleases are useful for the modulation of RNA guided endonuclease activity, including rapid, reversible, dosage, and/or temporal control of RNA guided endonuclease technologies.

CRISPR (clustered regularly interspaced short palindromic repeat)-Cas system

The CRISPR (clustered regularly interspaced short palindromic repeat)-Cas system is an adaptive immune system used by bacteria and archaea to defend against invading phages or mobile genetic elements. Three types of CRISPR-Cas system occur naturally and the type II system was first to be employed for genome editing in mammalian cells. This system employs an RNA-guided endonuclease Cas9, which cleaves double stranded DNA using RuvC and HNH nuclease domains. Cas9 identifies the target sequence by two recognition mechanisms: (i) Watson-Crick base-pairing between the target sequence and gRNA, and (ii) Protospacer Adjacent Motif (PAM) sequence on the target sequence.

Upon target recognition, Cas9 induces double strand breaks in the target gene, which when repaired by non-homologous endjoining (NHEJ) can result in frameshift mutations and gene knockdown. Alternatively, homology-directed repair (HDR) at the double-strand break site can allow insertion of the desired sequence.

Cas9 is a spectacular molecular machine with several fascinating attributes. First, the Cas9 employs molecular interactions and recognition between all the three elements of the central dogma—DNA, RNA, and protein. Second, Cas9 unwinds target DNA and facilitates strand invasion without utilizing ATP. Third, Cas9 efficiently induces DNA strand breaks in both prokaryotic and eukaryotic genome despite enormous differences in their genome size, structure, and organization. Finally, unlike transcription factors that often employ ID

diffusion and hopping for target search, Cas9:gRNA complex accomplishes target search by 3D diffusion only.

While Cas9 is a highly efficient molecular machine, its specificity is poor at best. Using genomewide, unbiased identification of double strand breaks enabled by sequencing (GUIDE-seq), Joung and co-workers showed that Cas9's off-targets were present on nearly every chromosome for the on-target gene EMX1. In another specificity study, Zhang and co-workers systematically studied the effect of single-base mismatch between gRNA and target sequence on Cas9 cleavage efficiency for the EMX1 gene. They found that Cas9 tolerated mismatches at PAM-distal sites on gRNA for multiple locations on EMX1 gene. Similar trends in mismatch tolerance were also reported by Doudna and Joung laboratories. Finally, Alt and co-workers have shown that in addition to off-target editing, Cas9 induces chromosomal translocations leading to dicentric chromosomes, which will generate genomic instability. They and others also demonstrated the inverse correlation between Cas9 activity and specificity.

Compounds

The compounds of the invention can be prepared from commercially available starting materials, compounds known in the literature, or readily prepared intermediates, by employing standard synthetic methods and procedures known to those skilled in the art. Standard synthetic methods and procedures for the preparation of organic molecules and functional group transformations and manipulations can be readily obtained from the relevant scientific literature or from standard textbooks in the field. It will be appreciated that where typical or preferred process conditions (i.e., reaction temperatures, times, mole ratios of reactants, solvents, pressures, etc.) are given, other process conditions can also be used unless otherwise stated. Optimum reaction conditions may vary with the particular reactants or solvent used, but such conditions can be determined by one skilled in the art by routine optimization procedures. Those skilled in the art of organic synthesis will recognize that the nature and order of the synthetic steps presented may be varied for the purpose of optimizing the formation of the compounds described herein.

Synthetic chemistry transformations (including protecting group methodologies) useful in synthesizing the compounds described herein are known in the art and include, for example, those such as described in R.C. Larock, *Comprehensive Organic Transformations*, 2d.ed., Wiley-VCH Publishers (1999); P.G.M. Wuts and T.W. Greene, *Protective Groups in Organic Synthesis*, 4th Ed., John Wiley and Sons (2007); L. Fieser and M. Fieser, *Fieser and*

Fieser's Reagents for Organic Synthesis, John Wiley and Sons (1994); and L. Paquette, ed., *Encyclopedia of Reagents for Organic Synthesis*, John Wiley and Sons (1995), and subsequent editions thereof.

The processes described herein can be monitored according to any suitable method known in the art. For example, product formation can be monitored by spectroscopic means, such as nuclear magnetic resonance spectroscopy (e.g., ^1H or ^{13}C), infrared spectroscopy (FT-IR), spectrophotometry (e.g., UV-visible), or mass spectrometry (MS), or by chromatography such as high performance liquid chromatography (HPLC) or thin layer chromatography (TLC).

Preparation of compounds can involve the protection and deprotection of various chemical groups. The need for protection and deprotection, and the selection of appropriate protecting groups can be readily determined by one skilled in the art. The chemistry of protecting groups can be found, for example, in Greene, et al, *Protective Groups in Organic Synthesis*, 2d. Ed., Wiley & Sons, 1991, which is incorporated herein by reference in its entirety.

The reactions of the processes described herein can be carried out in suitable solvents which can be readily selected by one of skill in the art of organic synthesis. Suitable solvents can be substantially nonreactive with the starting materials (reactants), the intermediates, or products at the temperatures at which the reactions are carried out, i.e., temperatures which can range from the solvent's freezing temperature to the solvent's boiling temperature. A given reaction can be carried out in one solvent or a mixture of more than one solvent. Depending on the particular reaction step, suitable solvents for a particular reaction step can be selected.

Resolution of racemic mixtures of compounds can be carried out by any of numerous methods known in the art. An example method includes preparation of the Mosher's ester or amide derivative of the corresponding alcohol or amine, respectively. The absolute configuration of the ester or amide is then determined by proton and/or ^{19}F NMR spectroscopy. An example method includes fractional recrystallization using a "chiral resolving acid" which is an optically active, salt-forming organic acid. Suitable resolving agents for fractional recrystallization methods are, for example, optically active acids, such as the D and L forms of tartaric acid, diacetyltartaric acid, dibenzoyltartaric acid, mandelic acid, malic acid, lactic acid or the various optically active camphorsulfonic acids. Resolution of racemic mixtures can also be carried out by elution on a column packed with an optically

active resolving agent (e.g., dinitrobenzoylphenylglycine). Suitable elution solvent compositions can be determined by one skilled in the art.

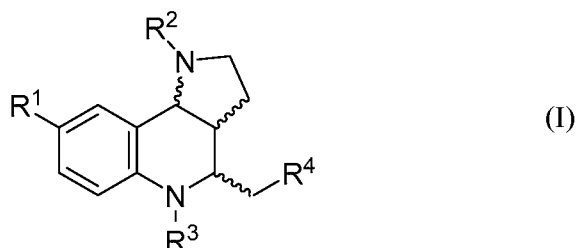
Disclosed herein are inhibitors of Cas9, an RNA-guided DNA endonuclease that naturally occurs in *S. pyogenes* (SpCas9) and *S. aureus* (SaCas9), respectively. Cas9 recognizes foreign DNA using Protospacer Adjacent Motif (PAM) sequence and the base-pairing of the target DNA by the guide RNA (gRNA). The relative ease of inducing targeted strand breaks at any genomic loci by Cas9 has enabled efficient genome editing in multiple cell types and organisms. Cas9 derivatives can also be used as transcriptional activators/repressors.

A challenge posed by Cas9 is that its cleavage selectivity is low. Off-target editing activity can result in undesired undesirable chromosomal translocation. This activity limits the use of Cas9 in a therapeutic setting due to unreliable gene manipulation and lack of ability to control the action of Cas9. The Cas9 inhibitors disclosed herein provide rapid, dosable, and/or temporal control of Cas9 that increases Cas9 specificity and enables external control and manipulation of gene targeting.

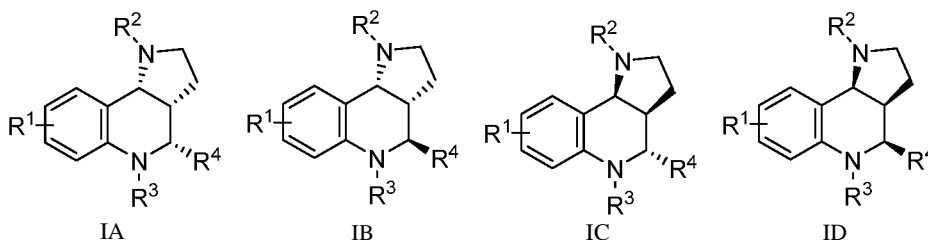
Provided herein are compounds having Formula I, II, III, IV, or V which may be used as Cas9 inhibitors in any embodiment of the invention described. Although each of the disclosed inventive compounds having formulas I-V may be disclosed with wedged bonds to indicate the specific stereochemistries, each stereocenter may independently have the R or S configuration, unless reference is given to specific compounds. For any compound having the structure of Formula I-V described herein, R¹-R⁴ may be defined as R¹ is selected from Hydrogen, aryl, heteroaryl, -C₁₋₆alkyl, -C₂₋₆alkenyl, -C₂₋₆alkynyl, -(CH₂)_nCOOH, -(CH₂)_nCOOC₁₋₆alkyl, -(CH₂)_nCOC₁₋₆alkyl, -(CH₂)_nOH, -(CH₂)_nCONHR, -(CH₂)_nNHCOR, -(CH₂)_nNHCOC₁₋₆alkyl, -(CH₂)_nNHSO₂R, -(CH₂)_nSO₂NHR, -(CH₂)_nSO₂R, -(CH₂)_nSO₂NHCOR, -(CH₂)_nSO₂NHCOOR, -(CH₂)_nSO₂NHCONRR, -(CH₂)_nCONHSO₂R, -(CH₂)_nNHCONRR, -(CH₂)_nC₃₋₁₀cycloalkyl-COOR, -SC₁₋₆alkyl, SOC₁₋₆alkyl, SO₂C₁₋₆alkyl, -C₃₋₁₀-iocyloheteroalkenyl, -C₃₋₁₀-iocyloheteroalkyl, substituted or unsubstituted phenyl, -(CH₂)_n-phenyl, -(CH₂)_n-aryl and -(CH₂)_n-heteroaryl, -(CH₂)_nC₃₋₁₀-iocyloalkyl, -(CH₂)_nC₃₋₁₀-iocyloalkyl-aryl, -(CH₂)_nC₃₋₁₀-iocyloalkyl-heteroaryl, -(CH₂)_nC₄₋₁₀-iocyloalkenyl, -(CH₂)_nC₄₋₁₀-iocyloalkenyl-aryl, -(CH₂)_nC₄₋₁₀-iocyloalkenyl-heteroaryl, -(CH₂)_nC₂₋₁₀-iocyloheteroalkyl, -(CH₂)_nC₁₋₁₀-iocyloheteroalkenyl, -C₂₋₆alkenyl-alkyl, -C₂₋₆alkenyl-aryl, -C₂₋₆alkenyl-heteroaryl, -C₂₋₆alkenyl-C₃₋₇cycloalkyl, -C₂₋₆alkenyl-C₃₋₇cycloalkenyl, -C₂₋₆alkenyl-C₂₋₇cycloheteroalkyl, -C₂₋₆alkynyl-(CH₂)_n-aryl, -C₂₋₆alkynyl-alkyl, -C₂₋₆alkynyl-aryl, -C₂₋₆alkynyl-heteroaryl, -C₂₋

$\text{C}_6\text{alkynyl-C3-7cycloalkyl}$, $\text{-C2-}_6\text{alkynyl-C3-7cycloalkenyl}$, $\text{-C2-}_6\text{alkynyl-C3-7cycloheteroalkyl}$, $\text{-C2-}_6\text{alkynyl-C3-7cycloheteroalkenyl}$, $\text{-CONH-(CH}_2\text{)}_n\text{phenyl}$, wherein n equals to 0 to 6 (*i.e.*, 0, 1, 2, 3, 4, 5, or 6), and each CH_2 is unsubstituted or substituted with one or two substituents selected from $\text{Ci-C}_6\text{alkyl}$, -OH , -CN , -CF_3 , halogen, COOH , $\text{COCi-C}_6\text{alkyl}$, $\text{COOCi-C}_6\text{alkyl}$, and -NH_2 , wherein each NH is unsubstituted or substituted with Ci-Cealkyl , -OH , halogen, COOH , COCi-Cealkyl , $\text{COOCi-C}_6\text{alkyl}$.

In some embodiments, the Cas9 inhibitors are compounds having the structure of Formula (I):



wherein each wavy bond indicates the presence of a stereocenter, and each stereocenter may be independently at each occurrence in the R or S configuration. In some embodiments, both stereocenters of the five membered fused ring are in the same configuration (*i.e.*, both (R) or both (S)). The groups $\text{R}^1\text{-R}^4$ may be defined as shown below. In some embodiments, the Cas9 inhibiting compounds may have the structure of Formula IA, IB, IC, or ID:



wherein R_i is independently selected at each occurrence from hydrogen, -X , -R , -Li-X , or -Li-R ; and

$\text{R}_2\text{-R}_4$ are independently selected from hydrogen, -X , -R , -Li-X , or -Li-R ; where

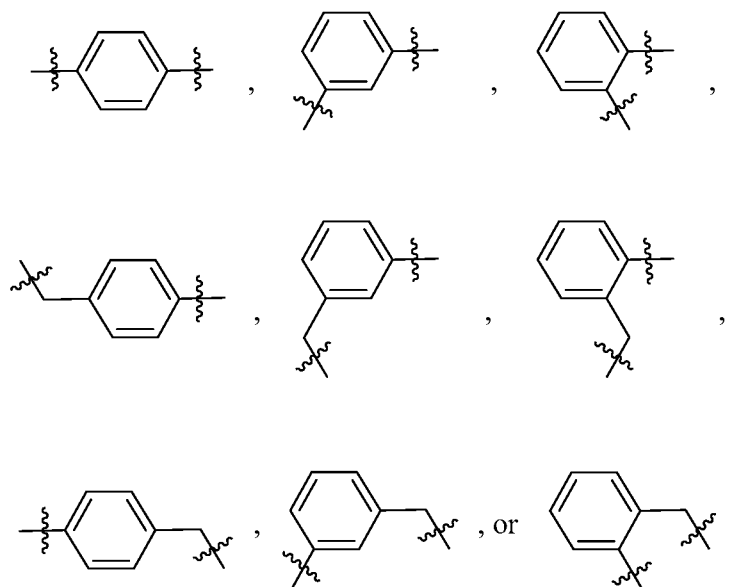
X is independently selected at each occurrence from CN , OH , CF_3 , COOH , OR , OR , NR_2 , or halogen (*e.g.*, -Cl , -F , -Br , etc.);

Li is selected from $\text{-(CH}_2\text{)}_n\text{-}$, $\text{-(CH}_2\text{)}_n\text{-C(0)O-}$, $\text{-(CH}_2\text{)}_n\text{-C(0)-NH-}$, $\text{-(CH}_2\text{)}_n\text{-NH-C(0)-}$, $\text{-(CH}_2\text{)}_n\text{-NH-SO}_2\text{-}$, $\text{-(CH}_2\text{)}_n\text{-SO}_2\text{-NH-}$, $\text{-(CH}_2\text{)}_n\text{-SO}_2\text{-}$, $\text{-(CH}_2\text{)}_n\text{-SO}_2\text{-NH-C(0)-}$, $\text{-(CH}_2\text{)}_n\text{-R}^L\text{-}$, $\text{-R}^L\text{-C(0)-O-}$, $\text{-R}^L\text{-NH-C(0)-(CH}_2\text{)}_n\text{-}$, $\text{-R}^L\text{-NH-S(0)}_2\text{-(CH}_2\text{)}_n\text{-}$, -S- , -S(O)- , $\text{-S(O)}_2\text{-}$; wherein n is independently 0, 1, 2, 3, 4, 5, or 6;

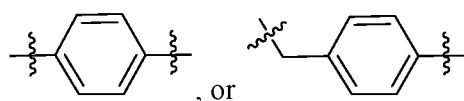
R^L is independently selected at each occurrence from C1-C12 linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C1-12 hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

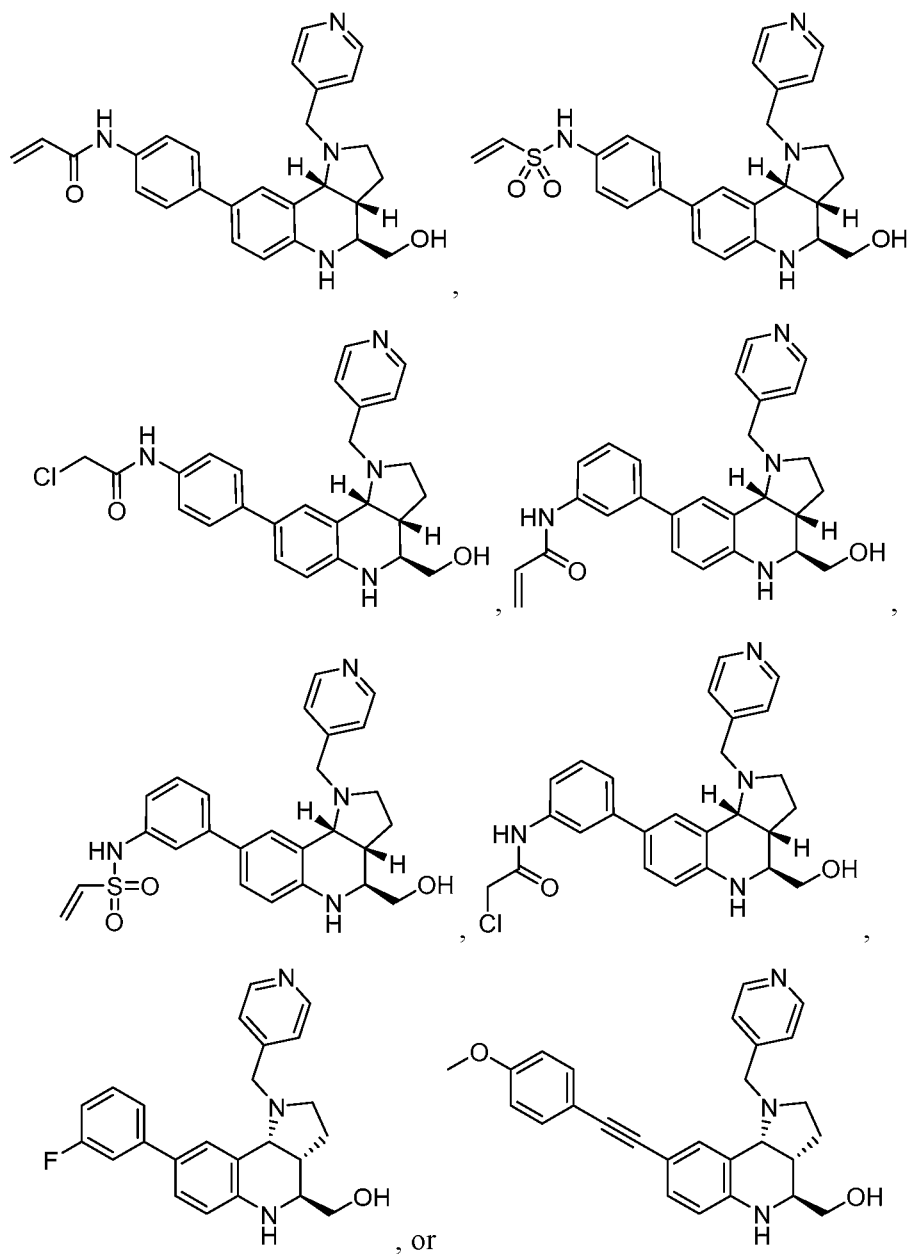
In some embodiments, R^2 is a group $-CH_2-R$. In some embodiments, R^1-R^4 are independently selected at each occurrence from hydrogen, -R, or -Li -R, -Li-X where Li independently selected at each occurrence from $-(CH_2)_i-$ or $-C\equiv C-$. In some embodiments, R is a five or six membered optionally aromatic and optionally substituted ring. In some embodiments, R is selected from phenyl, benzyl, or pyridinyl. In some embodiments, -X is selected from -F, -Cl, -OH, or -OR. In some embodiments, Li is selected from $-(CH_2)_i-$ or $-C\equiv C-$. In some embodiments, R is a C_{3-6} optionally aromatic cyclic hydrocarbon optionally substituted with one or more (e.g., two, three, etc.) functional groups selected from -F, -Cl, -OR* (e.g., -OCH₃), or -NH-C(=O)-C(CH₃)₃ (i.e. NHBoc). In some embodiments, R is methyl, ethyl, propyl, butyl, ethenyl (i.e., $-CH=CH_2$), or ethynyl (e.g., $-C\equiv CH$). In some embodiments, R^2 is -Li-X, or -Li -R, where L_1 is $-R^L-NH-C(=O)-(CH_2)_n-$, $-R^L-NH-S(=O)_2-(CH_2)_n-$, or $-C(=O)-(CH_2)_n-$. In some embodiments, R^L has the structure:



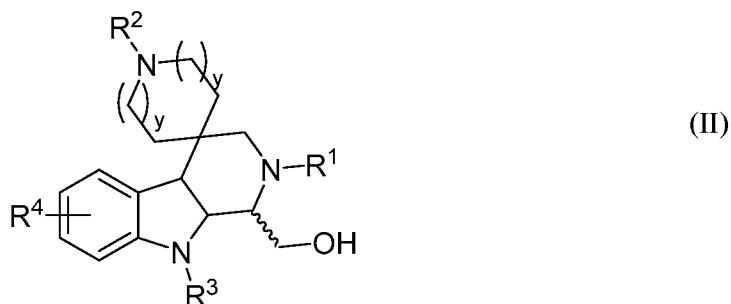
In some embodiments, R^L has the structure:



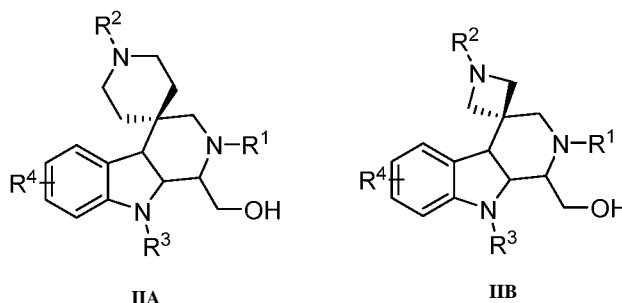
. In some embodiments, each R^1 is hydrogen. In some embodiments, one of R^1 is a group $-R$, or $-R^L-R$ and each other R^1 is hydrogen. The Cas9 inhibiting compounds may have the structure:



In some embodiments, the Cas9 inhibiting compounds may have the structure of Formula (II):



wherein "y" is independently selected at each occurrence from 0 or 1, and the wavy bond indicates the presence of a stereocenter which may be in the (R) or (S) configuration. In some embodiments, "y" is selected from 0 or 1. In some embodiments, the Cas9 inhibiting compounds have the structure of Formula IIA or IIB:



wherein R_4 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R^1 - R^3 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen (e.g., -Cl, -F, -Br, etc.);

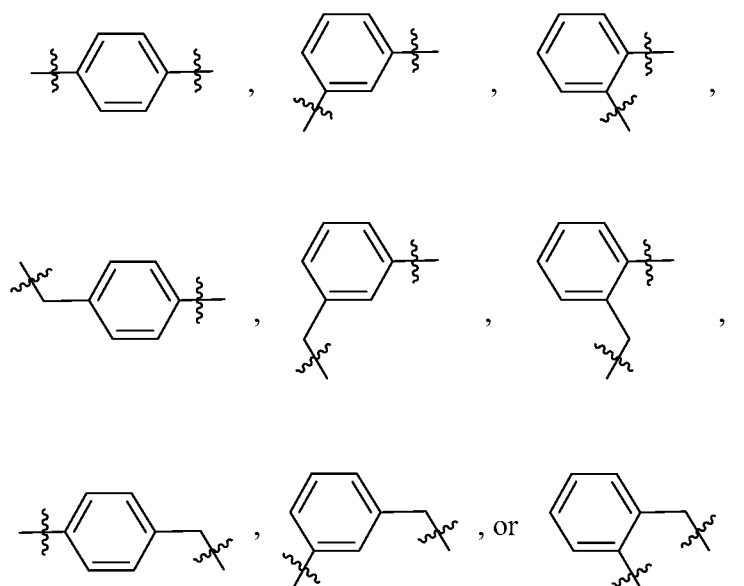
Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-NH-C(=O)-(CH_2)_n-$, $-R^L-NH-S(=O)_2-(CH_2)_n-$, $-R^L-C(=O)O-$, -S-, -S(O)-, $-S(O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

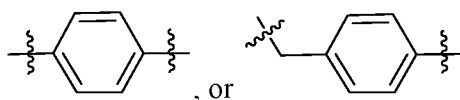
R is selected from C_1 - C_{12} hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups

X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

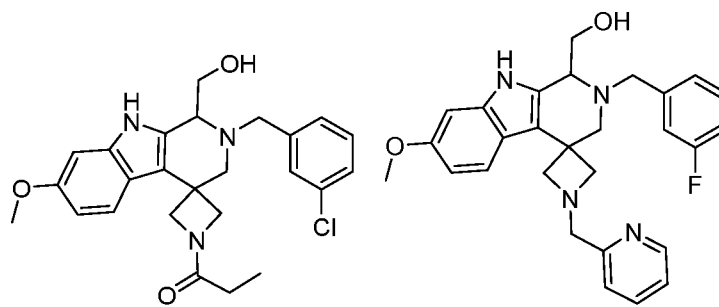
In some embodiments, R^4 is selected at each occurrence from hydrogen or -OR. In some embodiments, one of R^4 is -OR and each other of R^4 is hydrogen. In some embodiments, R^2 is a group -Li-R, -Li-X, where Li is selected from -C(O), $-(CH_2)_n-$, $-C(O)-(CH_2)_n-$, $-CH_2)_n-C(O)-$. In some embodiments, R is a C_{3-6} optionally aromatic cyclic hydrocarbon optionally substituted with one or more (e.g., two, three, etc.) functional groups selected from -F, -Cl, -OR* (e.g., -OCH₃), or -NH-C(O)O-C(CH₃)₃ (i.e. NHBoc). In some embodiments, R is methyl, ethyl, propyl, butyl, ethenyl (i.e., $-CH=CH_2$), or ethynyl (e.g., $-C\equiv CH$). In some embodiments, R^2 is -Li-X, or -Li-R, where Li is $-R^L-NH-C(O)-(CH_2)_n-$, $-R^L-NH-S(O)_2-(CH_2)_n-$, or $-C(O)-(CH_2)_n-$. In some embodiments, R^L has the structure:

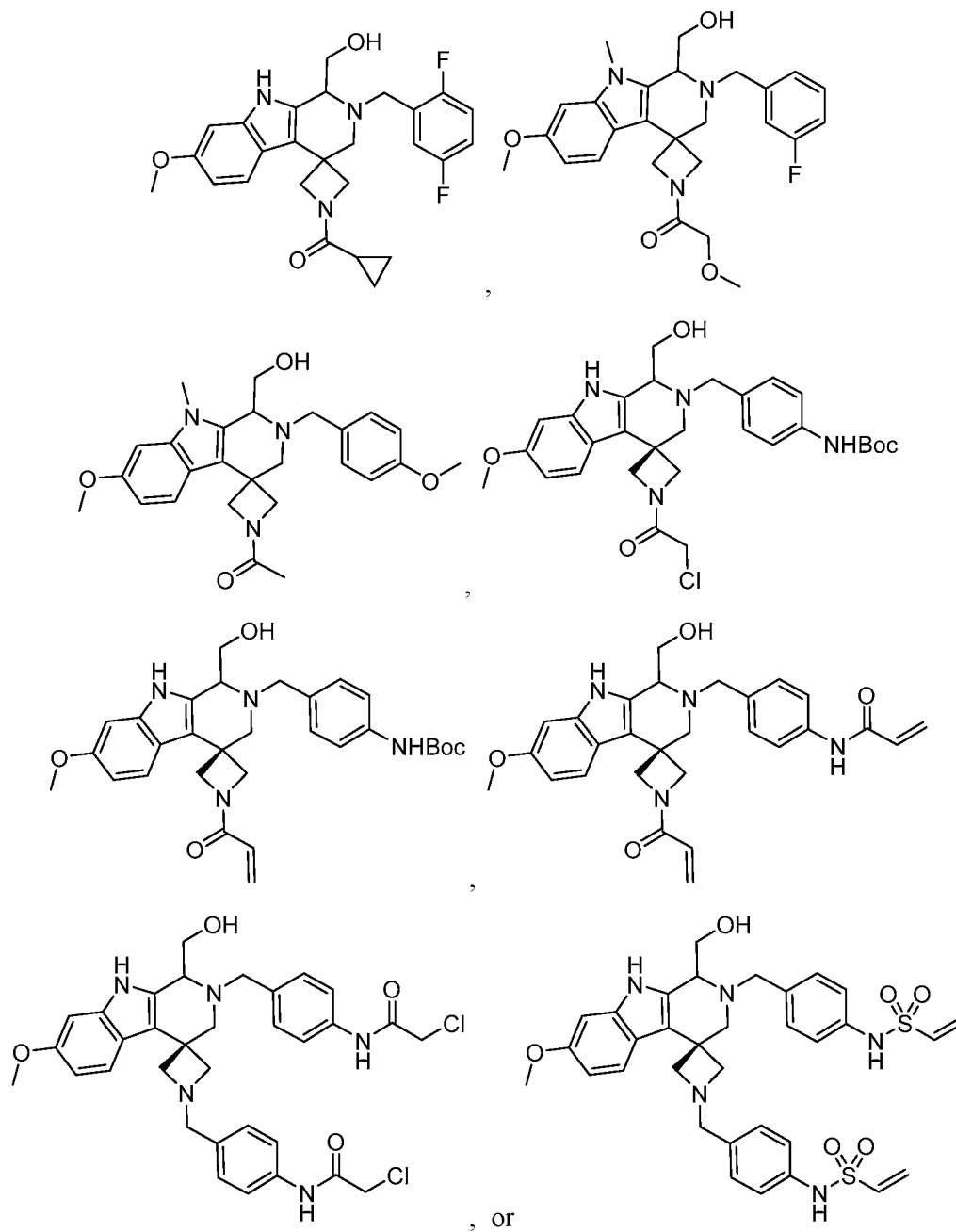


In some embodiments, R^L has the structure:

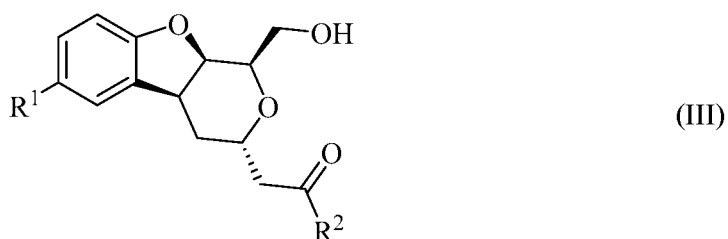


In some embodiments, the Cas9 inhibiting compounds have the structure:





In some embodiments, the Cas9 inhibiting compounds have the structure of Formula (III):



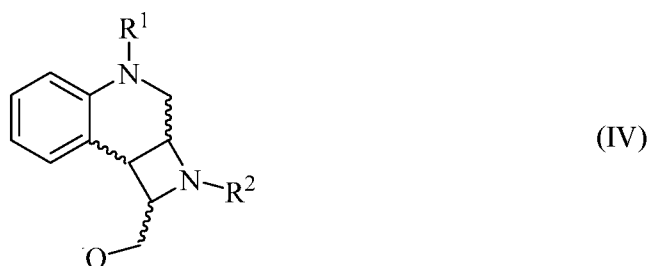
wherein R^1 - R^2 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen (*e.g.*, -Cl, -F, -Br, etc.);

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)-$, -S-, -S(O)-, $-S(O)_2-$; wherein n is independently 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (*e.g.*, alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (*e.g.*, 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1 - C_{12} hydrocarbons (*e.g.*, alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (*e.g.*, 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

In some embodiments, the Cas9 inhibiting compounds have the structure of Formula (IV):



wherein each wavy bond may be in the R or the S configuration,

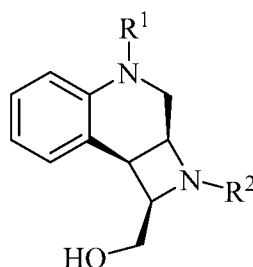
R^1 - R^2 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen (*e.g.*, -Cl, -F, -Br, etc.);

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, -

$(\text{CH}_2)_n$ - R^L -, $-\text{R}^L\text{-C(=O)-O-}$ -, $-\text{S-}$ -, $-\text{S(O)-}$ -, $-\text{S(O)}_2$ -; wherein n is independently 0, 1, 2, 3, 4, 5, or 6;

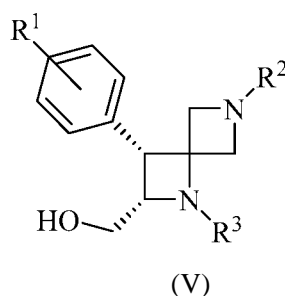
R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_{1-12} hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof. In some embodiments, the Cas9 inhibiting compounds have the structure:



In some embodiments, the Cas9 inhibiting compounds have the structure of Formula

V:



wherein R^1 is independently selected at each occurrence from hydrogen, $-\text{X}$ -, $-\text{R}$ -, $-\text{Li-X}$ -, or $-\text{Li-R}$;

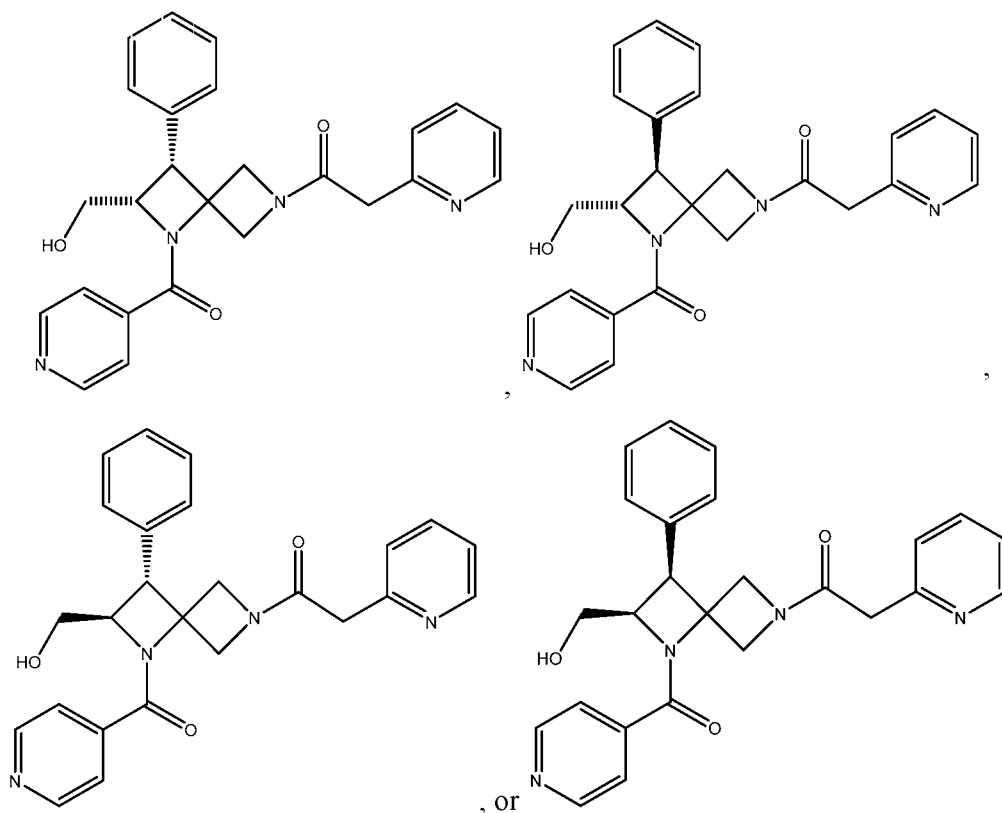
R^2 - R^3 are independently selected from hydrogen, $-\text{X}$ -, $-\text{R}$ -, $-\text{Li-X}$ -, or $-\text{Li-R}$; where

X is independently selected at each occurrence from CN , OH , CF_3 , COOH , OR , NR_2 , or halogen (e.g., $-\text{Cl}$ -, $-\text{F}$ -, $-\text{Br}$ -, etc.);

Li is selected from $-(\text{CH}_2)_n$ -, $-(\text{CH}_2)_n\text{-C(=O)-O-}$ -, $-(\text{CH}_2)_n\text{-C(=O)-NH-}$ -, $-(\text{CH}_2)_n\text{-NH-C(=O)-}$ -, $-(\text{CH}_2)_n\text{-NH-SO}_2$ -, $-(\text{CH}_2)_n\text{-SO}_2\text{-NH-}$ -, $-(\text{CH}_2)_n\text{-SO}_2$ -, $-(\text{CH}_2)_n\text{-SO}_2\text{-NH-C(=O)-}$ -, $-(\text{CH}_2)_n$ - R^L -, $-\text{R}^L\text{-C(=O)-O-}$ -, $-\text{S-}$ -, $-\text{S(O)-}$ -, $-\text{S(O)}_2$ -; wherein n is independently 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C1-C12 linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C1-12 hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof. In some embodiments, each of R^1 is hydrogen. In some embodiments, R^2 or R^3 is independently $-Li-R$, where Li is selected from $-C(O)-$, or $-C(O)-(CH_2)_n-$. In some embodiments R is a six membered aromatic ring optionally substituted with N. In some embodiments, the Cas9 inhibitors have the structure:



The disclosed compounds can be in free base form unassociated with other ions or molecules, or they can be a pharmaceutically acceptable salt, solvate, or prodrug thereof. One aspect provides a disclosed compound or a pharmaceutically acceptable salt. One aspect provides a disclosed compound or a pharmaceutically acceptable salt or solvate thereof. One aspect provides a pharmaceutically acceptable salt of a disclosed compound. One aspect

provides a solvate of a disclosed compound. One aspect provides a hydrate of a disclosed compound. One aspect provides a prodrug of a disclosed compound.

The disclosed compounds can be in free base form unassociated with other ions or molecules, or they can be a pharmaceutically acceptable salt, solvate, or prodrug thereof. One aspect provides a disclosed compound or a pharmaceutically acceptable salt. One aspect provides a disclosed compound or a pharmaceutically acceptable salt or solvate thereof. One aspect provides a pharmaceutically acceptable salt of a disclosed compound. One aspect provides a solvate of a disclosed compound. One aspect provides a hydrate of a disclosed compound. One aspect provides a prodrug of a disclosed compound.

Methods of Use

Small molecule inhibitors of RNA guided endonucleases (e.g., Cas9) were developed that have the potential to allow rapid, dosable, and/or temporal control of Cas9 activities. Reports of small-molecule controlled Cas9 activity are present in literature (Senis et al, *Biotechnol J* 2014, 9, 1402-12; Wright et al., *Proc Natl Acad Sci USA*. 2015 Mar 10;112(10):2984-9; Gonzalez et al, *Cell Stem Cell* 2014, 15, 215-26; Davis et al., *Nat Chem Biol* 2015, 11, 316-8). However, none of them ensure dosability-the small molecules act merely as inducers of Cas9 activity. Further, most of these small molecule systems are not reversible upon removal of the small molecule (Zetsche et al., *Nat Biotech* 2015, 33, 139-142; Davis et al, *Nat Chem Biol* 2015, 11, 316-8), and therefore, do not allow precise temporal control in transcriptional regulatory technologies.

Small molecule inhibitors of RNA guided endonucleases (e.g., Cas9) have potential therapeutic uses for regulating genome editing technologies involving RNA guided endonucleases. Dosable control of the therapeutic activity of RNA guided endonucleases introduced into a subject or cell of a subject is important for effective genome editing therapeutic strategies. Small molecule inhibitors of RNA guided endonucleases can be administered to a subject undergoing RNA guided endonuclease based gene therapy or any other RNA guided endonuclease based therapy. In certain embodiments, the subject is a human or mammal. Small molecule inhibitors of RNA guided endonucleases eliminate or reduce undesirable off-target editing and chromosomal translocations when present at high concentrations. Furthermore, small molecule inhibitors of RNA guided endonucleases can be used to rapidly terminate constitutively active Cas9, following on-target gene-editing.

Small molecule inhibitors of RNA guided endonucleases can also be used to regulate genome editing technologies in other organisms, including invertebrates, plants, and

unicellular organisms (e.g., bacteria). Potential uses include regulating gene drives for entomological and agricultural uses. In addition, it is anticipated that Cas9 inhibitors will be valuable probes to understand the role of Cas9 in CRISPR-mediated bacterial immunity (e.g., spacer acquisition) (Nunez et al, *Nature*. 2015 Mar 12;519(7542): 193-8; Heler et al., *Nature* 2015, 519, 199-202). Along similar lines, Cas9 inhibitors can be deployed for directed evolution of Cas9. It is hypothesized that Cas9 inhibitors will disrupt bacterial immunity against bacteriophages (or toxic DNA) by interfering with the CRISPR-Cas9-based immune surveillance system in bacteria. Akin to the development of antibiotic resistance, bacteria will be forced to evolve Cas9 protein. Accordingly, the inhibitors may also be used as an anti-infective agent.

Formulations

Agents described herein, including analogs thereof, and/or agents discovered to have medicinal value using the methods described herein are useful as a drug for inhibiting RNA guided nucleases (e.g., Cas9, Cpf1). For therapeutic uses, the compositions or agents identified using the methods disclosed herein may be administered systemically, for example, formulated in a pharmaceutically-acceptable buffer such as physiological saline. Preferable routes of administration include, for example, subcutaneous, intravenous, interperitoneally, intramuscular, or intradermal injections that provide continuous, sustained levels of the drug in the patient. Treatment of human patients or other animals will be carried out using a therapeutically effective amount of a therapeutic identified herein in a physiologically-acceptable carrier. Suitable carriers and their formulation are described, for example, in Remington's Pharmaceutical Sciences by E. W. Martin. The amount of the therapeutic agent to be administered varies depending upon the manner of administration, the age and body weight of the patient, and with the clinical symptoms. Generally, amounts will be in the range of those used for other agents used in the treatment of disease.

The disclosed compounds may be administered alone (e.g., in saline or buffer) or using any delivery vehicles known in the art. For instance the following delivery vehicles have been described: Cochleates; Emulsomes, ISCOMs; Liposomes; Live bacterial vectors (e.g., Salmonella, Escherichia coli, Bacillus calmatte-guerin, Shigella, Lactobacillus); Live viral vectors (e.g., Vaccinia, adenovirus, Herpes Simplex); Microspheres; Nucleic acid vaccines; Polymers; Polymer rings; Proteosomes; Sodium Fluoride; Transgenic plants; Virosomes; Virus-like particles. Other delivery vehicles are known in the art and some additional examples are provided below.

The disclosed compounds may be administered by any route known, such as, for example, orally, transdermally, intravenously, cutaneously, subcutaneously, nasally, intramuscularly, intraperitoneally, intracranially, and intracerebroventricularly.

In certain embodiments, disclosed compounds are administered at dosage levels greater than about 0.001 mg/kg, such as greater than about 0.01 mg/kg or greater than about 0.1 mg/kg. For example, the dosage level may be from about 0.001 mg/kg to about 50 mg/kg such as from about 0.01 mg/kg to about 25 mg/kg, from about 0.1 mg/kg to about 10 mg/kg, or from about 1 mg/kg to about 5 mg/kg of subject body weight per day, one or more times a day, to obtain the desired therapeutic effect. It will also be appreciated that dosages smaller than about 0.001 mg/kg or greater than about 50 mg/kg (for example about 50-100 mg/kg) can also be administered to a subject.

In one embodiment, the compound is administered once-daily, twice-daily, or three-times daily. In one embodiment, the compound is administered continuously (i.e., every day) or intermittently (e.g., 3-5 days a week). In another embodiment, administration could be on an intermittent schedule.

Further, administration less frequently than daily, such as, for example, every other day may be chosen. In additional embodiments, administration with at least 2 days between doses may be chosen. By way of example only, dosing may be every third day, bi-weekly or weekly. As another example, a single, acute dose may be administered. Alternatively, compounds can be administered on a non-regular basis e.g., whenever symptoms begin. For any compound described herein the effective amount can be initially determined from animal models.

Toxicity and efficacy of the compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds that exhibit large therapeutic indices may have a greater effect when practicing the methods as disclosed herein. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

Data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage of the compounds disclosed herein for use in humans. The

dosage of such agents lies within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the disclosed methods, the effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography. In certain embodiments, pharmaceutical compositions may comprise, for example, at least about 0.1% of an active compound. In other embodiments, the active compound may comprise between about 2% to about 75% of the weight of the unit, or between about 25% to about 60%, for example, and any range derivable therein. Multiple doses of the compounds are also contemplated.

The formulations disclosed herein are administered in pharmaceutically acceptable solutions, which may routinely contain pharmaceutically acceptable concentrations of salt, buffering agents, preservatives, compatible carriers, and optionally other therapeutic ingredients.

For use in therapy, an effective amount of one or more disclosed compounds can be administered to a subject by any mode that delivers the compound(s) to the desired surface, e.g., mucosal, systemic. Administering the pharmaceutical composition of the present disclosure may be accomplished by any means known to the skilled artisan. Disclosed compounds may be administered orally, transdermally, intravenously, cutaneously, subcutaneously, nasally, intramuscularly, intraperitoneally, intracranially, or intracerebroventricularly .

For oral administration, one or more compounds can be formulated readily by combining the active compound(s) with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a subject to be treated.

Pharmaceutical preparations for oral use can be obtained as solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations

such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Optionally the oral formulations may also be formulated in saline or buffers, i.e. EDTA for neutralizing internal acid conditions or may be administered without any carriers.

Also specifically contemplated are oral dosage forms of one or more disclosed compounds. The compound(s) may be chemically modified so that oral delivery of the derivative is efficacious. Generally, the chemical modification contemplated is the attachment of at least one moiety to the compound itself, where said moiety permits (a) inhibition of proteolysis; and (b) uptake into the blood stream from the stomach or intestine. Also desired is the increase in overall stability of the compound(s) and increase in circulation time in the body. Examples of such moieties include: polyethylene glycol, copolymers of ethylene glycol and propylene glycol, carboxymethyl cellulose, dextran, polyvinyl alcohol, polyvinyl pyrrolidone and polyproline. Other polymers that could be used are poly-1,3-dioxolane and poly-1,3,6-tioxocane. In some aspects for pharmaceutical usage, as indicated above, are polyethylene glycol moieties.

The location of release may be the stomach, the small intestine (the duodenum, the jejunum, or the ileum), or the large intestine. One skilled in the art has available formulations which will not dissolve in the stomach, yet will release the material in the duodenum or elsewhere in the intestine. In some aspects, the release will avoid the deleterious effects of the stomach environment, either by protection of the compound or by release of the biologically active material beyond the stomach environment, such as in the intestine.

To ensure full gastric resistance a coating impermeable to at least pH 5.0 is important. Examples of the more common inert ingredients that are used as enteric coatings are cellulose acetate trimellitate (CAT), hydroxypropylmethylcellulose phthalate (HPMCP), HPMCP 50, HPMCP 55, polyvinyl acetate phthalate (PVAP), Eudragit L30D, Aquateric, cellulose acetate phthalate (CAP), Eudragit L, Eudragit S, and Shellac. These coatings may be used as mixed films.

A coating or mixture of coatings can also be used on tablets, which are not intended for protection against the stomach. This can include sugar coatings, or coatings which make the tablet easier to swallow. Capsules may consist of a hard shell (such as gelatin) for

delivery of dry therapeutic i.e. powder; for liquid forms, a soft gelatin shell may be used. The shell material of cachets could be thick starch or other edible paper. For pills, lozenges, molded tablets or tablet triturates, moist massing techniques can be used.

The disclosed compounds can be included in the formulation as fine multiparticulates in the form of granules or pellets of particle size about 1 mm. The formulation of the material for capsule administration could also be as a powder, lightly compressed plugs or even as tablets. The compound could be prepared by compression.

Colorants and flavoring agents may all be included. For example, the compound may be formulated (such as by liposome or microsphere encapsulation) and then further contained within an edible product, such as a refrigerated beverage containing colorants and flavoring agents.

One may dilute or increase the volume of compound delivered with an inert material. These diluents could include carbohydrates, especially mannitol, α -lactose, anhydrous lactose, cellulose, sucrose, modified dextrans and starch. Certain inorganic salts may be also be used as fillers including calcium triphosphate, magnesium carbonate and sodium chloride. Some commercially available diluents are Fast-Flo, Emdex, STA-Rx 1500, Emcompress and Avicell. Disintegrants may be included in the formulation of the therapeutic into a solid dosage form. Materials used as disintegrates include but are not limited to starch, including the commercial disintegrant based on starch, Explotab. Sodium starch glycolate, Amberlite, sodium carboxymethylcellulose, ultramylopectin, sodium alginate, gelatin, orange peel, acid carboxymethyl cellulose, natural sponge and bentonite may all be used. Another form of the disintegrants is the insoluble cationic exchange resins. Powdered gums may be used as disintegrants and as binders and these can include powdered gums such as agar, Karaya or tragacanth. Alginic acid and its sodium salt are also useful as disintegrants.

Binders may be used to hold the therapeutic together to form a hard tablet and include materials from natural products such as acacia, tragacanth, starch and gelatin. Others include methyl cellulose (MC), ethyl cellulose (EC) and carboxymethyl cellulose (CMC). Polyvinyl pyrrolidone (PVP) and hydroxypropylmethyl cellulose (HPMC) could both be used in alcoholic solutions to granulate the therapeutic.

An anti-frictional agent may be included in the formulation of the compound to prevent sticking during the formulation process. Lubricants may be used as a layer between the compound and the die wall, and these can include but are not limited to; stearic acid including its magnesium and calcium salts, polytetrafluoroethylene (PTFE), liquid paraffin,

vegetable oils and waxes. Soluble lubricants may also be used such as sodium lauryl sulfate, magnesium lauryl sulfate, polyethylene glycol of various molecular weights, Carbowax 4000 and 6000. Glidants that might improve the flow properties of the drug during formulation and to aid rearrangement during compression might be added. The glidants may include starch, talc, pyrogenic silica and hydrated silicoaluminate.

To aid dissolution of the compound into the aqueous environment a surfactant might be added as a wetting agent. Surfactants may include anionic detergents such as sodium lauryl sulfate, dioctyl sodium sulfosuccinate and dioctyl sodium sulfonate. Cationic detergents might be used and could include benzalkonium chloride or benzethonium chloride. The list of potential non-ionic detergents that could be included in the formulation as surfactants are lauromacrogol 400, polyoxyl 40 stearate, polyoxyethylene hydrogenated castor oil 10, 50 and 60, glycerol monostearate, polysorbate 40, 60, 65 and 80, sucrose fatty acid ester, methyl cellulose and carboxymethyl cellulose. These surfactants could be present in the formulation of the compound either alone or as a mixture in different ratios.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. Microspheres formulated for oral administration may also be used. Such microspheres have been well defined in the art. All formulations for oral administration should be in dosages suitable for such administration.

For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present disclosure may be conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of e.g. gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

Also contemplated herein is pulmonary delivery of the compounds of the disclosure. The compound is delivered to the lungs of a mammal while inhaling and traverses across the lung epithelial lining to the blood stream using methods well known in the art.

Contemplated for use in the practice of methods disclosed herein are a wide range of mechanical devices designed for pulmonary delivery of therapeutic products, including but not limited to nebulizers, metered dose inhalers, and powder inhalers, all of which are familiar to those skilled in the art. Some specific examples of commercially available devices suitable for the practice of these methods are the Ultravent nebulizer, manufactured by Mallinckrodt, Inc., St. Louis, Missouri; the Acorn II nebulizer, manufactured by Marquest Medical Products, Englewood, Colorado; the Ventolin metered dose inhaler, manufactured by Glaxo Inc., Research Triangle Park, North Carolina; and the Spinhaler powder inhaler, manufactured by Fisons Corp., Bedford, Massachusetts.

All such devices require the use of formulations suitable for the dispensing of compound. Typically, each formulation is specific to the type of device employed and may involve the use of an appropriate propellant material, in addition to the usual diluents, and/or carriers useful in therapy. Also, the use of liposomes, microcapsules or microspheres, inclusion complexes, or other types of carriers is contemplated. Chemically modified compound may also be prepared in different formulations depending on the type of chemical modification or the type of device employed. Formulations suitable for use with a nebulizer, either jet or ultrasonic, will typically comprise compound dissolved in water at a concentration of about 0.1 to about 25 mg of biologically active compound per mL of solution. The formulation may also include a buffer and a simple sugar (e.g., for stabilization and regulation of osmotic pressure). The nebulizer formulation may also contain a surfactant, to reduce or prevent surface induced aggregation of the compound caused by atomization of the solution in forming the aerosol.

Formulations for use with a metered-dose inhaler device will generally comprise a finely divided powder containing the compound suspended in a propellant with the aid of a surfactant. The propellant may be any conventional material employed for this purpose, such as a chlorofluorocarbon, a hydrochlorofluorocarbon, a hydrofluorocarbon, or a hydrocarbon, including trichlorofluoromethane, dichlorodifluoromethane, dichlorotetrafluoroethanol, and 1,1,1,2-tetrafluoroethane, or combinations thereof. Suitable surfactants include sorbitan trioleate and soya lecithin. Oleic acid may also be useful as a surfactant.

Formulations for dispensing from a powder inhaler device will comprise a finely divided dry powder containing compound and may also include a bulking agent, such as lactose, sorbitol, sucrose, or mannitol in amounts which facilitate dispersal of the powder from the device, e.g., about 50 to about 90% by weight of the formulation. The compound should most advantageously be prepared in particulate form with an average particle size of less than 10 mm (or microns), such as about 0.5 to about 5 mm, for an effective delivery to the distal lung.

Nasal delivery of a disclosed compound is also contemplated. Nasal delivery allows the passage of a compound to the blood stream directly after administering the therapeutic product to the nose, without the necessity for deposition of the product in the lung. Formulations for nasal delivery include those with dextran or cyclodextran.

For nasal administration, a useful device is a small, hard bottle to which a metered dose sprayer is attached. In one embodiment, the metered dose is delivered by drawing the pharmaceutical composition solution into a chamber of defined volume, which chamber has an aperture dimensioned to aerosolize and aerosol formulation by forming a spray when a liquid in the chamber is compressed. The chamber is compressed to administer the pharmaceutical composition. In a specific embodiment, the chamber is a piston arrangement. Such devices are commercially available.

Alternatively, a plastic squeeze bottle with an aperture or opening dimensioned to aerosolize an aerosol formulation by forming a spray when squeezed is used. The opening is usually found in the top of the bottle, and the top is generally tapered to partially fit in the nasal passages for efficient administration of the aerosol formulation. In some aspects, the nasal inhaler will provide a metered amount of the aerosol formulation, for administration of a measured dose of the drug.

The compound, when it is desirable to deliver them systemically, may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions.

Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Alternatively, the active compounds may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal or vaginal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols.

Suitable liquid or solid pharmaceutical preparation forms are, for example, aqueous or saline solutions for inhalation, microencapsulated, encochleated, coated onto microscopic gold particles, contained in liposomes, nebulized, aerosols, pellets for implantation into the skin, or dried onto a sharp object to be scratched into the skin. The pharmaceutical compositions also include granules, powders, tablets, coated tablets, (micro)capsules, suppositories, syrups, emulsions, suspensions, creams, drops or preparations with protracted release of active compounds, in whose preparation excipients and additives and/or auxiliaries such as disintegrants, binders, coating agents, swelling agents, lubricants, flavorings, sweeteners or solubilizers are customarily used as described above. The pharmaceutical compositions are suitable for use in a variety of drug delivery systems.

The compounds may be administered per se (neat) or in the form of a pharmaceutically acceptable salt. When used in medicine the salts should be pharmaceutically acceptable, but non-pharmaceutically acceptable salts may conveniently be

used to prepare pharmaceutically acceptable salts thereof. Such salts include, but are not limited to, those prepared from the following acids: hydrochloric, hydrobromic, sulphuric, nitric, phosphoric, maleic, acetic, salicylic, p-toluene sulphonic, tartaric, citric, methane sulphonic, formic, malonic, succinic, naphthalene-2-sulphonic, and benzene sulphonic. Also, such salts can be prepared as alkaline metal or alkaline earth salts, such as sodium, potassium or calcium salts of the carboxylic acid group.

Suitable buffering agents include: acetic acid and a salt (about 1-2% w/v); citric acid and a salt (about 1-3% w/v); boric acid and a salt (about 0.5-2.5% w/v); and phosphoric acid and a salt (about 0.8-2% w/v). Suitable preservatives include benzalkonium chloride (about 0.003-0.03% w/v); chlorobutanol (about 0.3-0.9% w/v); parabens (about 0.01-0.25% w/v) and thimerosal (about 0.004- 0.02% w/v).

The pharmaceutical compositions contain an effective amount of a disclosed compound optionally included in a pharmaceutically acceptable carrier. The term pharmaceutically acceptable carrier means one or more compatible solid or liquid filler, diluents or encapsulating substances which are suitable for administration to a human or other vertebrate animal. The term carrier denotes an organic or inorganic ingredient, natural or synthetic, with which the active ingredient is combined to facilitate the application. The components of the pharmaceutical compositions also are capable of being commingled with the compounds, and with each other, in a manner such that there is no interaction which would substantially impair the desired pharmaceutical efficiency.

Provided herein are methods of synthesizing disclosed compounds. A compound provided herein can be synthesized using a variety of methods known in the art. The schemes and description below depict general routes for the preparation of disclosed compounds.

Fluorescence Polarization-based Assays

In one aspect, the invention provides an assay that monitors the change in the fluorescence polarization of the fluorophore-labelled PAM-rich target DNA (henceforth called 12PAM-DNA) upon binding to [Cas9:guideRNA] complex. In this assay, the complexation of [Cas9:guideRNA] to 12PAM-DNA shows a dose-dependent increase in fluorophore polarization.

Fluorescence polarization is a useful technique to monitor the interaction between two molecules, including for example, Cas9-gRNA (ribonucleoprotein) complex and target DNA (12PAM). Exemplary fluorophore-labelled PAM-rich target DNA is shown below:

5'-GGCTGGACCACGCGGGAAAAATCCACCTAGGTGGTT CCTCTTCGGATGTTCCATCCTTT/36-FAM-3'

3'-CCGACCTGGTGCGCCCTTTTAGGTGGATCCACCAAGGAGAAGCCTACAAGGTAGGAAA-5'

The technique is based on the change in the tumbling rate or mass after complexation. Following the FP principle, smaller fragment was fluorescently labelled and polarizations were compared before and after complexation in the presence and absence of compounds. While complexation of Cas9:gRNA-DNA showed an enhancement in the FP value, the inhibitors should revert back the enhanced signal intensity. To verify this assumption, Cas9:gRNA-DNA complexation was formed in the presence of excess unlabeled DNA template (with no fluorophore tagging) and measured the FP value. As expected, a sharp decrease in the FP value was observed. Without being bound to theory, this was due at least in part to the displacement of the fluorophore labeled DNA by the cold one. Furthermore, it was investigate whether the displacement by the cold DNA was random or specific. To address this, a competition assay was performed where Cas9-gRNA was incubated with 12PAM DNA template either in the absence or presence of unlabeled DNA with increasing number of PAM density. Interestingly, zero (0) PAM DNA template did not show any considerable inhibition of fluorophore labeled 12PAM binding. In contrast, both 4PAM and 12PAM showed efficient inhibition though the extent was considerably more for 12PAM than the 4PAM DNA. These findings confirmed that the interaction between Cas9:gRNA and DNA template was specific and can be precisely monitored by reading out the FP value. In conclusion, the FP based assay can be used for the Cas9 inhibitor screening assay.

At an initial stage, FP-based screening of 10,000 compounds was performed consisting of structurally diverse scaffolds with vast functional variability. DMSO was used as a negative control while 12PAM cold DNA was used as the competitive positive control. Upon transformation of the large set of data into scatter plot, nearly 0.5% (>3 standard deviation) of the compounds were identified as the potential hits. Interestingly, a large portion of the hit compounds found to have similarity in their molecular scaffold with variation in the stereochemistry and functionality. Numerous structural and functional diversities in the compound library offered a wide scope of medicinal chemistry. Some of the hit compounds like spirocyclic library showed exciting stereo-centric dependence of the compound on their activity. To further understand the nature of these hit compounds, their

dose dependence on activity was examined. The potential hit compounds showed excellent dose dependent Cas9 inhibition activity with an IC_{50} value as low as 0.6 μ M. These finding further re-confirms the validation of the primary assay.

Spinach Transcription Assay

In one aspect, the invention provides a transcription assay to detect the activity of an RNA guided endonuclease. In one embodiment, the level of transcription is suppressed by Cas9 nuclease activity in an *in vitro* assay. In various embodiments, the transcription assay involves expression of a nucleic acid aptamer that binds a molecular fluorophore to generate a fluorescent signal. Such aptamer-fluorophore combinations are known in the art, including for example, the Spinach aptamer having the sequence

5'- GGGAGACGCAACUGAAUGAAAUGGUGAAGGACGGGUCCAGGUGU
GGCUGCUUCGGCAGUGCAGCUUGUUGAGUAGAGUGUGAGCUCCGCG
UAACUAGUCGCGUC AC-3'

and the fluorophore 4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1 *H*-imidazol-5-one (DFHBI) (see, e.g., US20120252699 and US20140220560, each of which is incorporated herein in their entirety). In the Spinach assay, Cas9 can cleave the DNA template and thus inhibit *in vitro* transcription of the nucleic acid aptamer. In certain embodiments, the guide RNA targeting the Spinach aptamer has the sequence

5'- GCUAUAGGACGCGACCGAAAGUUUUAGAGCUAGAAAUAGC AAG
UUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA
GUCGGUGCUUUU-3'

In the presence of fluorophore, suppression in transcription results in the reduction of RNA aptamer-fluorophore concentration and hence in the fluorescence signal. *In vitro* transcription reactions may comprise a purified linear DNA template containing a promoter operatively linked to a nucleic acid sequence encoding an RNA aptamer, ribonucleotide triphosphates, a buffer system (e.g., including DTT and magnesium ions, and an appropriate phage RNA polymerase (e.g., T7 polymerase).

EGFP Disruption Assay

In some embodiments, a quantitative human cell-based reporter assay that enables rapid quantitation of targeted nuclease activities is used to characterize off-target cleavage of Cas9-based RNA guided endonucleases. In this assay, the activities of nucleases targeted to a single integrated EGFP reporter gene can be quantified by assessing loss of fluorescence signal in human U20S.EGFP cells caused by inactivating frameshift insertion/deletion (indel) mutations introduced by error prone non-homologous end-joining (NHEJ) repair of nuclease-induced double-stranded breaks (DSBs).

In one protocol, U20S.EGFP cells harboring a single integrated copy of an EGFP-PEST fusion gene are cultured (see e.g., Reyon et al, Nat Biotech 30, 460-465 (2012), which is herein incorporated by reference in its entirety). For transfections, 200,000 cells are Nucleofected with gRNA expression plasmid and pJDS246 together with 30 ng of a Td-tomato-encoding plasmid using the SE Cell Line 4D-Nucleofector™ X Kit (Lonza) according to the manufacturer's protocol. Cells are analyzed 2 days post-transfection using a BD LSRII flow cytometer. Transfections for optimizing gRNA/Cas9 plasmid concentration are performed in triplicate and all other transfections are performed in duplicate.

PCR amplification is used for sequence verification of endogenous human genomic sites. PCR reactions are performed using Phusion Hot Start II high-fidelity DNA polymerase (NEB). Loci are amplified using touchdown PCR (98° C , 10 s; 72-62° C , -1° C./cycle, 15 s; 72° C , 30 s] 10 cycles, [98° C , 10 s; 62° C , 15 s; 72° C , 30 s] 25 cycles). Alternatively, PCR for other targets are performed with 35 cycles at a constant annealing temperature of 68° C. or 72° C. and 3% DMSO or 1M betaine, if necessary. PCR products are analyzed on a QIAXCEL capillary electrophoresis system to verify both size and purity. Validated products are treated with ExoSap-IT (Affymetrix) and sequenced by the Sanger method (MGH DNA Sequencing Core) to verify each target site.

SURVEYOR Nuclease Assay

In various embodiments, SURVEYOR nuclease assay is used to assess genome modification (see e.g., US20150356239, which is herein incorporated by reference in its entirety). In one protocol, 293FT cells are transfected with plasmid DNA. Cells were incubated at 37°C for 72 hours post-transfection prior to genomic DNA extraction. Genomic DNA is extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer's protocol. Briefly, pelleted cells are resuspended in QuickExtract solution and incubated at 65°C for 15 minutes and 98°C for 10 minutes.

The genomic region flanking the CR1 SPR target site for each gene is PCR amplified, and products are purified using QiaQuick Spin Column (Qiagen) following the manufacturer's protocol. 400ng total of the purified PCR products are mixed with 2 μ l 10X Taq DNA Polymerase PCR buffer (Enzytsaties) and ultrapure water to a final volume of 20 μ l, and subjected to a re-annealing process to enable heteroduplex formation: 95°C for 10 min, 95°C to 85°C ramping at - 2°C/s, 85°C to 25°C at - 0.25°C/s, and 25°C hold for 1 minute. After re-annealing, products are treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer's recommended protocol, and analyzed on 4-20% Novex TBE poly-acrylamide gels (Life Technologies). Gels re stained with SYBR Gold DNA stain (Life Technologies) for 30 minutes and imaged with a Gel Doe gel imaging system (Bio-rad). Quantification is based on relative band intensities.

Test Compounds and Extracts

In general, small molecule compounds are known in the art or are identified from large libraries of both natural product or synthetic (or semi-synthetic) extracts or chemical libraries or from polypeptide or nucleic acid libraries, according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Compounds used in screens may include known compounds (for example, known therapeutics used for other diseases or disorders). Alternatively, virtually any number of unknown chemical extracts or compounds can be screened using the methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds.

Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from Brandon Associates (Merrimack, N.H.) and Aldrich Chemical (Milwaukee, Wis.). Alternatively, chemical compounds to be used as candidate compounds can be synthesized from readily available starting materials using standard synthetic techniques and methodologies known to those of ordinary skill in the art. For example, a library of 8,000 novel small molecules is available, which was created using combinatorial methods of Diversity-Oriented Synthesis (DOS) (Comer et al, Proc Natl Acad

Sci U S A 108, 6751 (Apr 26, 2011 ; Lowe et al, J Org Chem 77, 7187 (Sep 7, 2012); Marcaurelle et al, J Am Chem Soc 132, 16962 (Dec 1, 2010))— to investigate chemical compounds not represented in traditional pharmaceutical libraries (Schreiber, S. L. (2000). Science 287, 1964-1969; Schreiber et al, Nat Biotechnol 28, 904 (Sep, 2010), each of which is herein incorporated by reference in their entirety). Synthetic chemistry transformations and protecting group methodologies (protection and deprotection) useful in synthesizing the compounds identified by the methods described herein are known in the art and include, for example, those such as described in R. Larock, Comprehensive Organic Transformations, VCH Publishers (1989); T. W. Greene and P. G. M. Wuts, Protective Groups in Organic Synthesis, 2nd ed., John Wiley and Sons (1991); L. Fieser and M. Fieser, Fieser and Fieser's Reagents for Organic Synthesis, John Wiley and Sons (1994); and L. Paquette, ed., Encyclopedia of Reagents for Organic Synthesis, John Wiley and Sons (1995), and subsequent editions thereof.

Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceanographics Institute (Ft. Pierce, Fla.), and PharmaMar, U.S.A. (Cambridge, Mass.). In addition, natural and synthetically produced libraries are produced, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al*, *Proc. Natl. Acad. Sci. U.S.A.* 90:6909, 1993; Erb *et al*, *Proc. Natl. Acad. Sci. USA* 91:11422, 1994; Zuckermann *et al*, *J. Med. Chem.* 37:2678, 1994; Cho *et al*, *Science* 261:1303, 1993; Carrell *et al*, *Angew. Chem. Int. Ed. Engl.* 33:2059, 1994; Carell *et al*, *Angew. Chem. Int. Ed. Engl.* 33:2061, 1994; and Gallop *et al*, *J. Med. Chem.* 37:1233, 1994. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

Libraries of compounds may be presented in solution (e.g., Houghten, *Biotechniques* 13:412-421, 1992), or on beads (Lam, *Nature* 354:82-84, 1991), chips (Fodor, *Nature* 364:555-556, 1993), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner U.S. Patent No. 5,223,409), plasmids (Cull *et al*, *Proc Natl Acad Sci USA* 89:1865-1869, 1992) or on phage (Scott and Smith, *Science* 249:386-390, 1990; Devlin, *Science* 249:404-406, 1990; Cwir\&et al. *Proc. Natl. Acad. Sci.* 87:63786382, 1990; Felici, *J. Mol. Biol.* 222:301-310, 1991 ; Ladner *supra*.).

In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their activity should be employed whenever possible.

When a crude extract is identified as containing a compound of interest, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect. Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity within the crude extract that achieves a desired biological effect. Methods of fractionation and purification of such heterogenous extracts are known in the art.

Small molecules of the invention preferably have a molecular weight below 2,000 daltons, more preferably between 300 and 1,000 daltons, and most preferably between 400 and 700 daltons. It is preferred that these small molecules are organic molecules.

Kits

The present compositions may be assembled into kits or pharmaceutical systems. The kits can include instructions for the treatment regime, reagents, equipment (test tubes, reaction vessels, needles, syringes, etc.) and standards for calibrating or conducting the treatment. The instructions provided in a kit according to the invention may be directed to suitable operational parameters in the form of a label or a separate insert. Optionally, the kit may further comprise a standard or control information so that the test sample can be compared with the control information standard to determine if whether a consistent result is achieved.

The container means of the kits will generally include at least one vial, test tube, flask, bottle, or other container means, into which a component may be placed, and preferably, suitably aliquoted. Where there is more than one component in the kit, the kit also will generally contain additional containers into which the additional components may be separately placed. However, various combinations of components may be comprised in a container. The kits of the present invention also will typically include a means for packaging the component containers in close confinement for commercial sale. Such packaging may include injection or blow-molded plastic containers into which the desired component containers are retained.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook, 1989); "Oligonucleotide Synthesis" (Gait, 1984); "Animal Cell Culture" (Freshney, 1987); "Methods in Enzymology" "Handbook of Experimental Immunology" (Weir, 1996); "Gene Transfer Vectors for Mammalian Cells" (Miller and Calos, 1987); "Current Protocols in Molecular Biology" (Ausubel, 1987); "PCR: The Polymerase Chain Reaction", (Mullis, 1994); "Current Protocols in Immunology" (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides of the invention, and, as such, may be considered in making and practicing the invention. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the assay, screening, and therapeutic methods of the invention, and are not intended to limit the scope of what the inventors regard as their invention.

EXAMPLES

Example 1. Synthesis and Characterization of Compounds

The following Examples illustrate the synthesis of a representative number of compounds and the use of these compounds in the treatment of malaria. Accordingly, the Examples are intended to illustrate but not to limit the disclosure. Additional compounds not specifically exemplified may be synthesized using conventional methods in combination with the methods described herein.

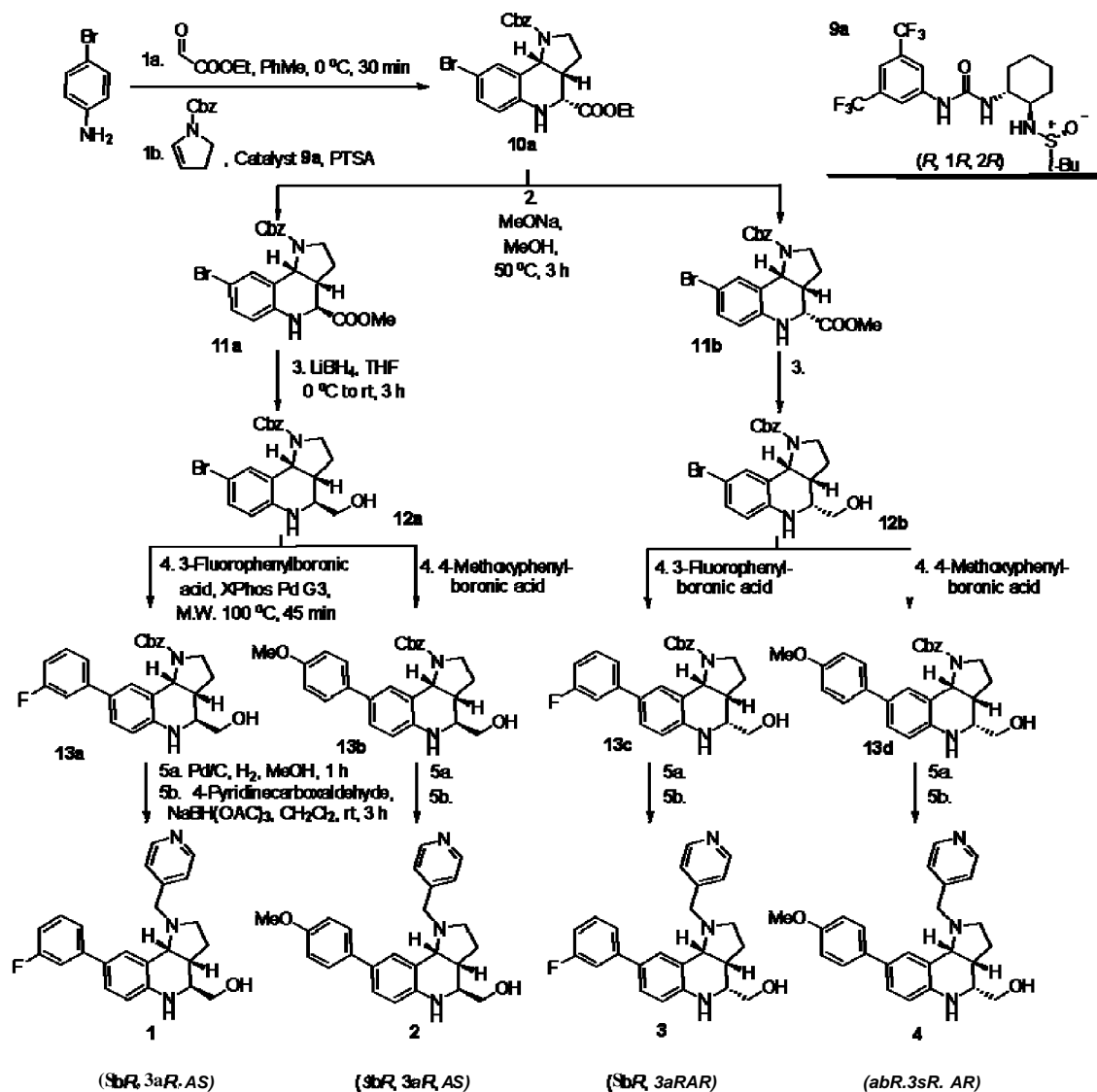
Unless otherwise noted, reactions were performed under an argon atmosphere using freshly dried HPLC grade solvents in flame-dried glassware. All reagents were purchased and used as received from commercial sources or synthesized based on cited procedures. XPhos-Pd-G3, 4-pyridinecarboxaldehyde, palladium on carbon, and sodium triacetoxyborohydride were purchased from Sigma Aldrich; 3-fluorophenylboronic acid was purchased from Oakwood Chemical, and 4-methoxyphenylboronic acid was purchased from Combi-Blocks. All reactions were monitored by thin-layer chromatography (TLC) using Merck Silica gel 60 F254 pre-coated plates (0.25 mm) visualized by UV light at 254 nm. Yields refer to pure

compounds after purification by flash column chromatography, unless otherwise noted. Flash column chromatography was performed using silica gel (60 Å mesh, 20-40 μm) on a Teledyne Isco CombiFlash Rf system.

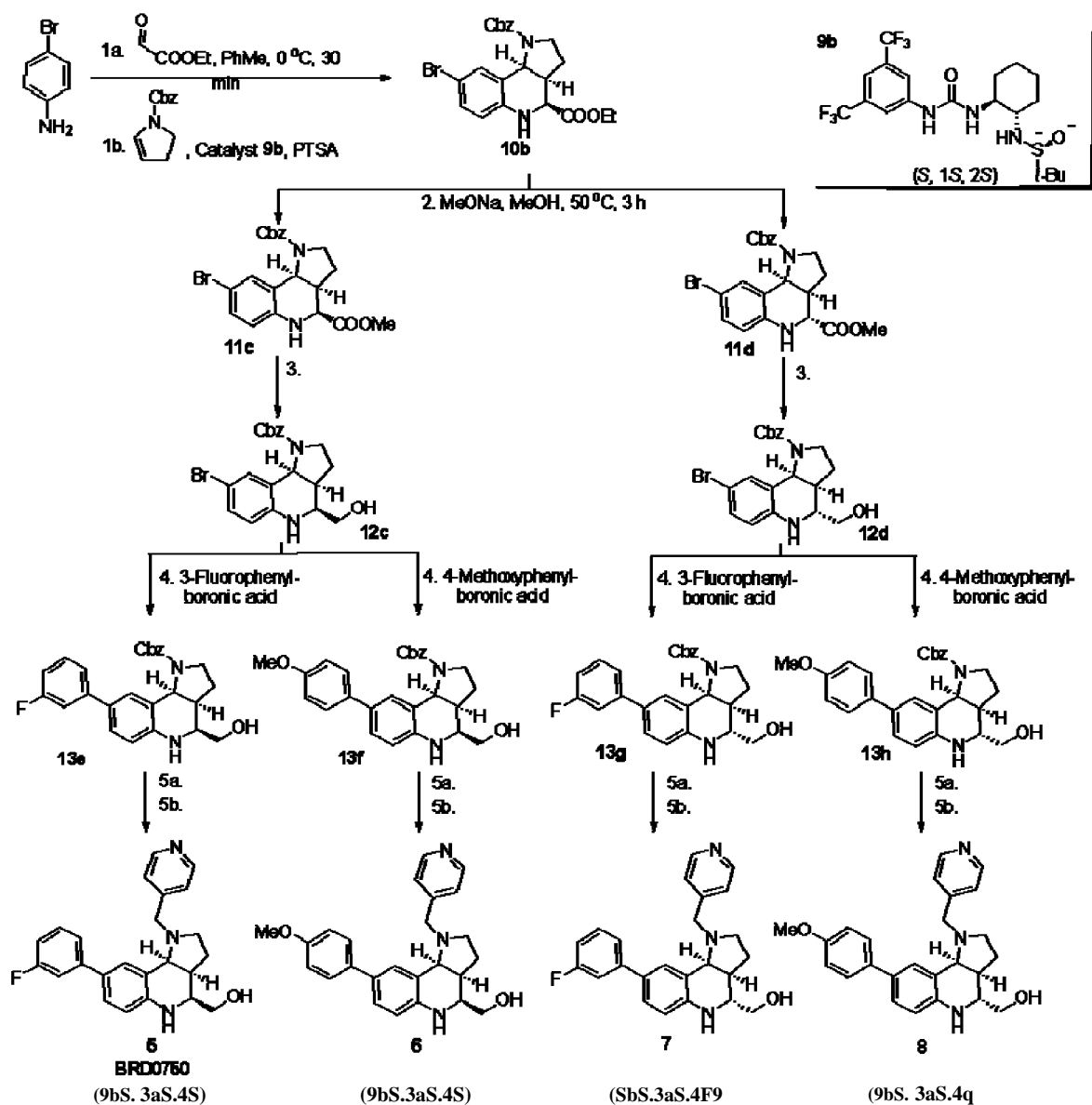
Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker 400 Spectrometer (^1H NMR, 400 MHz; ^{13}C , 100 MHz; Dept-135 Carbon, 100 MHz; ^{19}F NMR, 376 MHz). Chemical shifts are reported in parts per million (ppm) relative to the solvent used. NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. NMR data were obtained in CDCl_3 or $\text{OMSO}-d_6$. Data for ^1H NMR are reported as follows: chemical shift value in ppm, multiplicity (s = singlet, d = doublet, t = triplet, dd = double doublet, and m = multiplet), integration value, and coupling constant value in Hz. Optical rotations were recorded on an Autopol IV automatic Rudolph Research Analytical polarimeter. For each test, 4 mg of the appropriate compound was dissolved in 1 mL chloroform. The reported optical rotations values are averages of five independent measurements at 22 °C (set temperature at 20 °C). Enantiopurity of compounds was determined by analytic supercritical fluid chromatography (SFC) on a Waters UPC2 convergence chromatography system connected to a QDa single quadrupole mass spectrometer with Chiralcel AD-H, AS-H, IC, and OD-H columns using chiral stationary phase with mobile phase A consisting of supercritical carbon dioxide and mobile phase B consisting of isopropanol (IPA) at 45 °C. Infrared spectra were recorded on a Nicolet IR 100 FTIR from Thermo Scientific and are reported in frequency of absorption (cm^{-1}). Tandem liquid chromatography mass spectrometry (LCMS) was performed on a Waters 2795 separations module with a 3100 mass detector. High-resolution mass-spectra (HRMS) were acquired on an Agilent 1290 Infinity separations module coupled to a 6230 time-of-flight (TOF) mass detector operating in ESI^+ or ESI^- mode. "Find-by-Formula" feature in the MassHunter Qualitative Analysis Vb.06.00 was used to confirm mass values, which are averages of three independent measurements.

Compounds may be synthesized using the Synthetic Schemes S1 and S2. References to a substrate number include all compound permutations of that substrate number with alphabetic demarcations. For example, Substrate **12** includes compounds **12a**, **12b**, **12c**, and **12d** and Substrate **13** includes compounds **13a-h**.

Scheme SI: Synthetic scheme for the ((3a*R*, 9b*R*)-1-(pyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-1*H*-pyrrolo[3,2-*c*]quinolin-4-yl)methanols 1-4



Scheme S2, Synthetic scheme for the ((3a*S*, 9b*S*)-1-(pyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-1*H*-pyrrolo[3,2-*c*]quinolin-4-yl)methanols 5-8.



General Procedure A: Microwave- Assisted Suzuki Coupling to Give Hexahydropyrroloquinoline substrates 13a-h

The four isomers of benzyl 8-bromo-4-(hydroxymethyl)-2,3,3a,4,5,9b-hexahydro-1*H*-pyrrolo[3,2-*c*]quinoline-1-carboxylate **12a-d** were synthesized as described by Jacobsen et al. and Marcaurelle et al. using the chiral urea catalyst **9a** or **9b** (H. Xu, H. Zhang, E. N. Jacobsen, *Nat. Protoc.* **2014**, 9, 1860-1866; B. Gerald, M. W. O'Shea, E. Donckele, S. Kesavan, L. B. Akella, H. Xu, E. N. Jacobsen, L. S. Marcaurelle, *ACS Comb. Sci.* **2012**, 74,

621-630). Urea catalysts **9a** and **9b** were synthesized as described by Jacobsen et al. The obtained NMR spectral data were consistent with those reported in the literature (K. L. Tan, E. N. Jacobsen, *Angew. Chem. Int. Ed.* **2007**, *46*, 1315-1317.).

The microwave reactions were performed in a Biotage single-mode microwave reactor with a power of 0 to 400 W. A 10-20 mL Biotage microwave reaction vial was charged with the hexahydropyrroloquinoline substrate **12** (1.0 equiv., >90% ee), 3-fluorophenylboronic acid or 4-methoxyphenylboronic acid (1.2 equiv.), potassium carbonate (2.0 equiv.), XPhos Palladium third generation catalyst (5% mol), and a mixture solvent of THF-H₂O (v/v, 2/1). The vial was sealed with a septum cap, degassed under high vacuum, and backfilled with an argon atmosphere. The degassing step was repeated three times, and the resulting reaction mixture was microwave irradiated for 45 min at 100 °C. The reaction mixture was then cooled to room temperature and filtered through a short pad of Celite. The filtrate was evaporated under vacuum to give crude substrate, usually as off-yellow oily substance, which was purified by flash column chromatography on silica gel eluting with hexane and ethyl acetate (or dichloromethane and methanol).

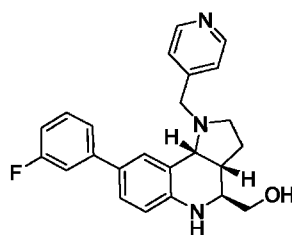
General Procedure B: Reductive Amination to Give Pyridinylmethylhexahydropyrroloquinoline Compounds 1 to 8

A round-bottom flask was charged with hexahydropyrroloquinoline substrate **13** (1.0 equiv.), palladium on carbon (10% weight), and methanol (0.05 M). The flask was sealed with a rubber septum, degassed under high vacuum, and backfilled with a hydrogen atmosphere. The degassing and hydrogen refilling step was repeated three times, and the resulting reaction mixture was stirred at room temperature for one hour or until the full conversion of the starting material monitored by TLC (methanol in CH₂Cl₂). The reaction mixture was filtered through a Celite pad and the filtrate was evaporated under vacuum to give the corresponding Cbz-deprotected hexahydropyrroloquinoline substrate.

A flame-dried round-bottom flask was charged with the Cbz-deprotected hexahydropyrroloquinoline substrate (1.0 equiv.) dissolved in dry CH₂Cl₂ (0.05 M), 4-pyridinecarboxyaldehyde (1.5 equiv.), and acetic acid (2.0 equiv.). The reaction mixture was stirred at room temperature for one hour before the adding of NaBH(OAc)₃ (3.0 equiv.). The reaction mixture was stirred at room temperature for another three hours or until the full conversion of the starting material monitored by TLC (methanol in CH₂Cl₂). The reaction mixture was then diluted with CH₂Cl₂, quenched with a saturated NaHCO₃ aqueous solution,

and extracted with CH_2Cl_2 (three times). Organic layers were combined, washed with brine, dried over anhydrous Na_2SO_4 , filtered, and concentrated in vacuo to give a crude residue, usually as off-white or light yellow oily substance, which was purified by flash column chromatography on silica gel eluting with hexane and ethyl acetate (or dichloromethane and methanol).

((3aR,4S,9bR)-8-(3-Fluorophenyl)-1-(pyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2-c]quinolin-4-yl)methanol (1/BRD7087)



Prepared from benzyl (3aR,4S,9bR)-8-(3-fluorophenyl)-4-(hydroxymethyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2-c]quinoline-1-carboxylate **13a** (667 mg, 1.54 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 5% methanol in dichloromethane gave the desired product **3** as a white solid (348 mg, yield 58%).

R_f = 0.38 (silica gel, 10% methanol in dichloromethane, UV).

^1H NMR (400 MHz, CDCl_3): δ 8.45 (d, 2H, J = 4.8 Hz, aromatic H), 7.35 (d, 2H, J = 8.2 Hz, aromatic H), 7.28-7.25 (m, 4H, aromatic H), 7.21 (d, 1H, J = 10.5 Hz, aromatic H), 6.97 (t, 1H, J = 8.6 Hz, aromatic H), 6.73 (d, 1H, J = 8.4 Hz, aromatic H), 4.38 (d, 1H, J = 13.8 Hz, CH_2OH), 4.01 (d, 1H, J = 8.6 Hz, CH_2NCH), 3.56-3.53 (m, 2H, NHCHCH and CH_2NCH), 3.33 (s, 1H, CH_2NCH), 3.27 (d, 1H, J = 13.8 Hz, CH_2OH), 2.97-2.93 (m, 1H, $\text{NCH}_2\text{C}^3/4$), 2.21-2.19 (m, 1H, NCH_2CH_2), 2.10-2.03 (m, 2H, NCH_2CH_2 and NHCHCH), 1.65-1.62 (m, 1H, NCH_2CH_2).

^{13}C NMR (100 MHz, CDCl_3): δ 164.5 and 162.1 (d, $^1J_{\text{C},\text{F}}$ = 243.4 Hz, aromatic C), 150.0 (aromatic C), 149.1 (2) (pyridinyl C), 144.7 (pyridinyl C), 143.6 and 143.5 (d, $^3J_{\text{C},\text{F}}$ = 7.8 Hz, aromatic C), 130.3 (aromatic C), 130.1 and 130.0 (d, $^3J_{\text{C},\text{F}}$ = 8.8 Hz, aromatic C), 127.8 (aromatic C), 127.6 (aromatic C), 123.6 (2) (pyridinyl C), 121.7 and 121.7 (d, $^4J_{\text{C},\text{F}}$ = 2.8 Hz, aromatic C), 118.3 (aromatic C), 114.6 (aromatic C), 113.0 and 112.8 (d, $^2J_{\text{C},\text{F}}$ = 21.8 Hz, aromatic C), 112.8 and 112.6 (d, $^2J_{\text{C},\text{F}}$ = 21.0 Hz, aromatic C), 64.7 (CH_2NCH), 64.1

(CH₂NCH), 56.0 (CH₂OH), 54.5 (NHCHCH), 51.4 (NCH₂CH₂), 35.8 (NHCHCH), 25.7 (NCH₂CH₂).

¹⁹F NMR (376 MHz, CDCl₃): δ -113.5.

[α]_D²² = +34.4° (c = 0.4, CHCl₃).

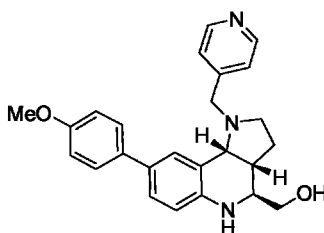
Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et₃N in CO₂, λ = 210 nm): ¾(minor) = 6.4 min, ¾(major) = 7.0 min.

IR (thin film, cm⁻¹): ν_{max} 3413, 2925, 1608, 1522, 1484, 1325, 1261, 1198, 1159, 1077, 869, 819, 782, 752, 693.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 95% by UV, rt = 0.92 min, m/z 390.1 (M+H)⁺, m/z 434.5 (M+FA-H)⁻.

HRMS (ESI, m/z): calcd for C₂₄H₂₄FN₃O (M+H)⁺: 390.1982, found: 390.1976.

((3a[^],4,S',9bi[?])-8-(4-Methoxyphenyl)-l-(pyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2 -c]quinolin-4-yl)methanol (2/BRD5779)



Prepared from benzyl (3ai[?],45',9bi[?])-4-(hydroxymethyl)-8 -(4-methoxyphenyl)-2, 3, 3a, 4, 5, 9b-hexahydro-1 H-pyrrolo [3,2-c]quinoline-1-carboxylate **13b** (186 mg, 0.42 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 70% ethyl acetate in hexane gave the desired product **2** as an off-white solid (72 mg, yield **43%**).

R_f = 0.69 (silica gel, 10% methanol in dichloromethane, UV).

¹H NMR (400 MHz, CDCl₃): δ 8.46 (d, 2H, J = 4.9 Hz, aromatic H), **7.44** (d, 2H, J = 8.2 Hz, aromatic H), **7.33** (d, 1H, J = 8.2 Hz, aromatic H), **7.25-7.24** (m, **3H**, aromatic H), **6.97** (d, 2H, J = 8.2 Hz, aromatic H), **6.74** (d, 1H, J = 8.2 Hz, aromatic H), **4.41** (d, 1H, J = 13.8 Hz, CH₂OH), **3.99** (d, 1H, J = 9.6 Hz, CH₂NCH), **3.85** (s, 3H, OCH₃), **3.59** (d, 1H, J = 9.6 Hz, CH₂NCH), **3.53-3.51** (m, 1H, NHCHCH), **3.32** (s, 1H, CH₂NCH), **3.25** (d, 1H, J = 13.8 Hz, CH₂OH), **2.97** (t, 1H, J = 9.2 Hz, NCH₂CH₂), 2.20-2.18 (m, 1H, NCH₂CH₂), 2.10-2.05 (m, 2H, NCH₂CH₂ and NHCHCH), 1.66-1.62 (m, 1H, NCH₂CH₂).

^{13}C NMR (100 MHz, CDCl_3): δ 158.3 (aromatic C), 149.9 (aromatic C), 149.2 (2) (aromatic C), 143.7 (aromatic C), 134.0 (aromatic C), 130.0 (aromatic C), 129.3 (aromatic C), 127.3 (3) (aromatic C), 123.6 (2) (aromatic C), 118.5 (aromatic C), 114.8 (aromatic C), 114.2 (2) (aromatic C), 64.7 (CH_2NCH), 64.1 (CH_2NCH), 56.1 (CH_2OH), 55.4 (OCH_3), 54.5 (NHCHCH), 51.4 (NCH_2CH_2), 35.9 (NHCHCH), 25.7 (NCH_2CH_2).

$[\alpha]_{\text{D}}^{22} = +33.1^\circ$ ($c = 0.4$, CHCl_3).

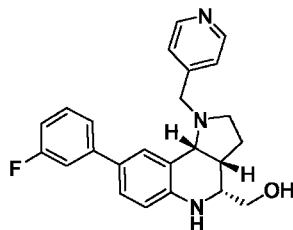
Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et_3N in CO_2 , $\lambda = 210$ nm): $\frac{3}{4}$ (minor) = 6.1 min, $\frac{3}{4}$ (major) = 7.3 min.

IR (thin film, cm^{-1}): ν_{max} 3402, 2929, 1614, 1499, 1480, 1246, 1180, 1028, 817, 753.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 95% by UV, $\text{rt} = 0.84$ min, m/z 402.2 ($\text{M}+\text{H}$) $^+$, m/z 446.5 ($\text{M}+\text{FA}-\text{H}$) $^-$.

HRMS (ESI, m/z): calcd for $\text{C}_{25}\text{H}_{27}\text{N}_3\text{O}_2$ ($\text{M}+\text{H}$) $^+$: 402.2182, found: 402.2172.

((3a * ,4i * ,9bi *)-8-(3-Fluorophenyl)-1-(pyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-L/7-pyrrolo[3,2-c]quinolin-4-yl)methanol (3/BRD2161)



Prepared from benzyl (3aS',45',9b5)-8-(3-fluorophenyl)-4-(hydroxymethyl)-2,3,3a,4,5,9b-hexahydro-1*H*-pyrrolo[3,2-c]quinoline-1-carboxylate **13c** (276 mg, 0.64 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 5% methanol in dichloromethane gave the desired product **3** as a white solid (66 mg, yield 27%).

$R_f = 0.23$ (silica gel, 5% methanol in dichloromethane, UV).

^1H NMR (400 MHz, CDCl_3): δ 8.55-8.53 (m, 2H, aromatic H), 7.33-7.26 (m, 6H, aromatic H), 7.21 (d, 1H, $J = 10.8$ Hz, aromatic H), 6.97 (t, 1H, $J = 8.4$ Hz, aromatic H), 6.73 (d, 1H, $J = 8.4$ Hz, aromatic H), 4.40 (d, 1H, $J = 13.5$ Hz, CH_2OH), 3.90-3.86 (m, 1H, CH_2NCH), 3.71-3.65 (m, 2H, NHCHCH and CH_2NCH), 3.51-3.47 (m, 2H, CH_2NCH and CH_2OH), 2.95-2.92 (m, 1H, NCH_2CH_2), 2.83-2.79 (m, 1H, NCH_2CH_2), 2.39-2.35 (m, 1H, NCH_2CH_2), 2.02-1.94 (m, 2H, NHCHCH and NCH_2CH_2).

^{13}C NMR (100 MHz, CDCl_3): δ 164.5 and 162.1 (d, $^1J_{\text{C},\text{F}} = 244$ Hz, aromatic C), 149.5 (2) (pyridinyl C), 148.2 (aromatic C), 145.8 (pyridinyl C), 143.5 and 143.4 (d, $^3J_{\text{C},\text{F}} = 7.9$ Hz, aromatic C), 130.1 and 130.0 (d, $^3J_{\text{C},\text{F}} = 8.2$ Hz, aromatic C), 129.4 (aromatic C), 128.8 (aromatic C), 127.5 (aromatic C), 123.9 (2) (pyridinyl C), 121.7 and 121.7 (d, $^1J_{\text{C},\text{F}} = 2.3$ Hz, aromatic C), 119.4 (aromatic C), 115.2 (aromatic C), 113.0 and 112.8 (d, $^2J_{\text{C},\text{F}} = 21.8$ Hz, aromatic C), 112.9 and 112.7 (d, $^2J_{\text{C},\text{F}} = 21.0$ Hz, aromatic C), 64.3 (CH_2NCH), 63.3 (CH_2NCH), 58.1 (CH_2OH), 54.4 (NHCHCH), 51.6 (NCH_2CH_2), 38.1 (NHCHCH), 23.6 (NCH_2CH_2).

^{19}F NMR (376 MHz, CDCl_3): δ -113.4.

$[\alpha]_{\text{D}}^{22} = -29.8^\circ$ ($c = 0.4$, CHCl_3).

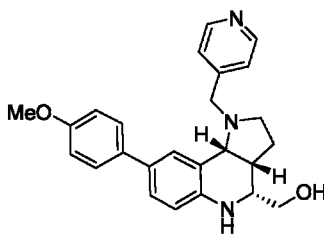
Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et_3N in CO_2 , $\lambda = 210$ nm): $\frac{3}{4}$ (minor) = 6.2 min, $\frac{3}{4}$ (major) = 6.9 min.

IR (thin film, cm^{-1}): ν_{max} 3364, 2917, 1608, 1516, 1480, 1314, 1262, 1193, 1164, 1076, 869, 822, 788, 752, 691.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 95% by UV, $\text{rt} = 0.82$ min, m/z 390 ($\text{M}+\text{H}$) $^+$, m/z 434 ($\text{M}+\text{FA}-\text{H}$) $^-$.

HRMS (ESI, m/z): calcd for $\text{C}_{24}\text{H}_{24}\text{FN}_3\text{O}$ ($\text{M}+\text{H}$) $^+$: 390.1982, found: 390.1972.

((3a * ,4i * ,9bi *)-8-(4-Methoxyphenyl)-1-(pyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo [3,2-c]quinolin-4-yl)methanol (4/BRD 1490)



Prepared from benzyl (3ai * ,4i * ,9bi *)-4-(hydroxymethyl)-8-(4-methoxy-phenyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2-c]quinoline-1-carboxylate **13d** (142 mg, 0.32 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 70% ethyl acetate in hexane gave the desired product **4** as an off-white solid (52 mg, yield 41%).

$R_f = 0.40$ (silica gel, 5% methanol in dichloromethane, UV).

^1H NMR (400 MHz, CDCl_3): δ 8.54 (d, 2H, $J = 4.9$ Hz, aromatic H), 7.44 (d, 2H, $J = 8.4$ Hz, aromatic H), 7.32-7.25 (m, 4H, aromatic H), 6.96 (d, 2H, $J = 8.4$ Hz, aromatic H), 6.73 (d, 1H, $J = 8.0$ Hz, aromatic H), 4.43 (d, 1H, $J = 13.2$ Hz, CH_2OH), 3.90-3.89 (m, 1H, CH_2NCH), 3.85 (s, 3H, OCH_3), 3.71-3.65 (m, 2H, CH_2NCH and NHCHCH), 3.49-3.45 (m, 2H, CH_2NCH and CH_2OH), 2.94-2.80 (m, 2H, NCH_2CH_2 and NHCHCH), 2.36-2.34 (m, 1H, $\text{NCH}_2\text{C}^3/4$), 2.00-1.97 (m, 2H, NCH_2CH_2).

^{13}C NMR (100 MHz, CDCl_3): δ 158.4 (aromatic C), 150.3 (aromatic C), 149.9 (2) (aromatic C), 144.9 (aromatic C), 134.0 (aromatic C), 130.1 (aromatic C), 129.1 (aromatic C), 127.3 (2) (aromatic C), 127.2 (aromatic C), 123.8 (2) (aromatic C), 119.4 (aromatic C), 115.2 (aromatic C), 114.2 (2) (aromatic C), 64.4 (CH_2NCH), 63.3 (CH_2NCH), 58.1 (CH_2OH), 55.4 (OCH_3), 54.5 (NHCHCH), 51.6 (NCH_2CH_2), 38.2 (NHCHCH), 23.6 (NCH_2CH_2).

$[\alpha]_{\text{D}}^{22} = -34.1^\circ$ ($c = 0.4$, CHCl_3).

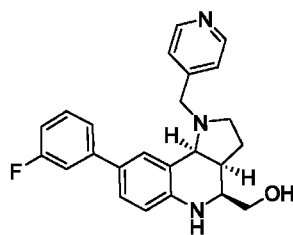
IR (thin film, cm^{-1}): ν_{max} 3364, 2911, 1609, 1495, 1246, 1180, 1045, 1027, 819, 754.

Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et_3N in CO_2 , $\lambda = 210$ nm): $3/4$ (minor) = 6.8 min, $3/4$ (major) = 7.1 min.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 95% by UV, $\text{rt} = 0.79$ min, m/z 402.5 ($\text{M}+\text{H}$) $^+$, m/z 446.6 ($\text{M}+\text{FA}-\text{H}$) $^-$.

HRMS (ESI, m/z): calcd for $\text{C}_{25}\text{H}_{27}\text{N}_3\text{O}_2$ ($\text{M}+\text{H}$) $^+$: 402.2182, found: 402.2171.

((3aS,4S,9bS)-8-(3-Fluorophenyl)-1-ipyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2-c]quinolin-4-yl)methanol (5/BRD0750)



Prepared from benzyl (3aS',4S',9bS)-8-(3-fluorophenyl)-4-(hydroxymethyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2-c]quinoline-1-carboxylate **13e** (168 mg, 0.39 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 80% ethyl acetate in hexane gave the desired product **5** as an off-white solid (71 mg, yield 47%).

$R_f = 0.25$ (silica gel, 5% methanol in dichloromethane, UV)

^1H NMR (400 MHz, CDCl_3): δ 8.53 (d, 2H, $J = 5.0$ Hz, aromatic H), 7.34-7.33 (m, 3H, aromatic H), 7.28-7.26 (m, 3H, aromatic H), 7.21 (d, 1H, $J = 10.6$ Hz, aromatic H), 6.96 (t, 1H, $J = 8.4$ Hz, aromatic H), 6.73 (d, 1H, $J = 8.4$ Hz, aromatic H), 4.39 (d, 1H, $J = 13.2$ Hz, CH_2OH), 3.87 (dd, 1H, $J = 11.6$ Hz and 5.0 Hz, CH_2NCH), 3.71-3.67 (m, 2H, NHCHCH and CH_2NCH), 3.51 (d, 1H, $J = 13.2$ Hz, CH_2OH), 3.47-3.45 (m, 1H, CH_2NCH), 2.96-2.90 (m, 1H, NCH_2CH_2), 2.82-2.79 (m, 1H, NCH_2CH_2), 2.39 (dd, 1H, $J = 17.8$ Hz and 8.8 Hz, NCH_2CH_2), 1.99-1.96 (m, 2H, NHCHCH and NCH_2CH_2).

^{13}C NMR (100 MHz, CDCl_3): δ 164.5 and 162.1 (d, $^1J_{\text{C,F}} = 243$ Hz, aromatic C), 149.7 (2) (pyridinyl C), 148.0 (aromatic C), 145.9 (pyridinyl C), 143.5 and 143.4 (d, $V_{\text{C,F}} = 7.5$ Hz, aromatic C), 130.1 and 130.0 (d, $^3J_{\text{C,F}} = 8.6$ Hz, aromatic C), 129.4 (aromatic C), 128.8 (aromatic C), 127.5 (aromatic C), 123.9 (2) (pyridinyl C), 121.8 and 121.7 (d, $^4J_{\text{C,F}} = 2.8$ Hz, aromatic C), 119.5 (aromatic C), 115.2 (aromatic C), 113.0 and 112.8 (d, $^2J_{\text{C,F}} = 21.4$ Hz, aromatic C), 112.9 and 112.7 (d, $^2J_{\text{C,F}} = 20.9$ Hz, aromatic C), 64.2 (CH_2NCH), 63.3 (CH_2NCH), 58.1 (CH_2OH), 54.4 (NHCHCH), 51.6 (NCH_2CH_2), 38.1 (NHCHCH), 23.6 (NCH_2CH_2).

^{19}F NMR (376 MHz, CDCl_3): δ -113.4.

$[\alpha]_{\text{D}}^{22} = +23.6^\circ$ ($c = 0.4$, CHCl_3).

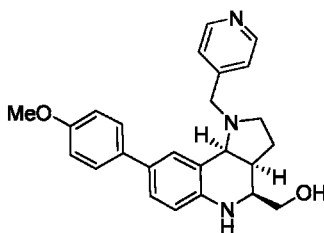
Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et_3N in CO_2 , $\lambda = 210$ nm): $\frac{3}{4}$ (major) = 6.2 min, $\frac{3}{4}$ (minor) = 6.9 min.

IR (thin film, cm^{-1}): ν_{max} 3332, 2916, 1608, 1517, 1480, 1300, 1262, 1193, 1167, 1077, 867, 821, 784, 753, 692.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 95% by UV, $\text{rt} = 0.86$ min, m/z 390.5 ($\text{M}+\text{H}$) $^+$, m/z 434.5 ($\text{M}+\text{FA}-\text{H}$) $^-$.

HRMS (ESI, m/z): calcd for $\text{C}_{24}\text{H}_{24}\text{FN}_3\text{O}$ ($\text{M}+\text{H}$) $^+$: 390.1982, found: 390.1973.

((3a5,4,S',9b^)-8-(4-Methoxyphenyl)-1-(pyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2-c]quinolin-4-yl)methanol (6/BRD6201)



Prepared from benzyl (3aS',45',9b5)-4-(hydroxymethyl)-8-(4-methoxy-phenyl)-2,3,3a,4,5,9b-hexahydro-1*H*-pyrrolo[3,2-*c*]quinoline-1-carboxylate **13f** (178 mg, 0.45 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 65% ethyl acetate in hexane gave the desired product **6** as an off-white solid (65 mg, yield 41%).

R_f = 0.43 (silica gel, 5% methanol in dichloromethane, UV).

¹H NMR (400 MHz, CDCl₃): δ 8.54 (d, 2H, *J* = 5.0 Hz, aromatic H), 7.44 (d, 2H, *J* = 8.2 Hz, aromatic H), 7.33-7.25 (m, 4H, aromatic H), 6.96 (d, 2H, *J* = 8.2 Hz, aromatic H), 6.73 (d, 1H, *J* = 8.2 Hz, aromatic H), 4.44 (d, 1H, *J* = 13.2 Hz, CH₂OH), 3.92-3.86 (m, 1H, CH₂NCH), 3.85 (s, 3H, OCH₃), 3.71 (dd, 1H, *J* = 11.3 Hz and 3.6 Hz, CH₂NCH), 3.65-3.64 (m, 1H, NHCHCH), 3.48-3.44 (m, 2H, CH₂NCH and CH₂OH), 2.95-2.91 (m, 1H, NCH₂CH₂), 2.84-2.81 (m, 1H, NCH₂CH₂), 2.36-2.34 (m, 1H, NCH₂CH₂), 2.03-1.95 (m, 2H, NCH₂CH₂).

¹³C NMR (100 MHz, CDCl₃): δ 158.4 (aromatic C), 149.8 (2) (aromatic C), 148.0 (aromatic C), 144.9 (aromatic C), 134.0 (aromatic C), 130.1 (aromatic C), 129.1 (aromatic C), 127.3 (2) (aromatic C), 127.3 (aromatic C), 123.8 (2) (aromatic C), 119.3 (aromatic C), 115.2 (aromatic C), 114.2 (2) (aromatic C), 64.4 (CH₂NCH), 63.3 (CH₂NCH), 58.1 (CH₂OH), 55.4 (OCH₃), 54.4 (NHCHCH), 51.6 (NCH₂CH₂), 38.1 (NHCHCH), 23.7 (NCH₂CH₂).

[α]_D²² = -24.6° (c = 0.4, CHCl₃).

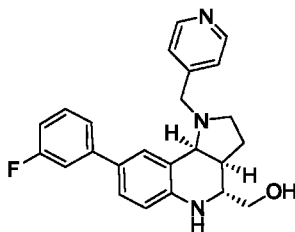
Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et₃N in CO₂, λ = 210 nm): ¾(major) = 6.9 min, ¾(minor) = 7.2 min.

IR (thin film, cm⁻¹): ν_{max} 3365, 2925, 1610, 1496, 1246, 1180, 1045, 1027, 818, 756.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 95% by UV, rt = 0.80 min, m/z 402.1 (M+H)⁺, m/z 446.6 (M+FA-H)⁻.

HRMS (ESI, m/z): calcd for C₂₅H₂₇N₃O₂ (M+H)⁺: 402.2182, found: 402.2180.

((3a5,4i?,9b^)-8^3-Fluorophenyl)-1^pyridin-4-ylmethyl)-2^^a,4,5,9b-hexahydro-L&-pyrrolo[3,2-c]quinolin-4-yl)methanol (7/BRD5039)



Prepared from benzyl (3aS,4i?,9b5)-8-(3-fluorophenyl)-4-(hydroxymethyl)-2,3,3a,4,5,9b-hexahydro-1*H*-pyrrolo[3,2-*c*]quinoline-1-carboxylate **13g** (168 mg, 0.39 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 80% ethyl acetate in hexane gave the desired product **7** as an off-white solid (83 mg, yield 55%).

R_f = 0.39 (silica gel, 10% methanol in dichloromethane, UV)

¹H NMR (400 MHz, CDCl₃): δ 8.46 (d, 2H, *J* = 5.0 Hz, aromatic H), 7.35-7.33 (m, 2H, aromatic H), 7.28-7.25 (m, 4H, aromatic H), 7.21-7.19 (m, 1H, aromatic H), 6.97 (t, 1H, *J* = 8.6 Hz, aromatic H), 6.73 (d, 1H, *J* = 8.4 Hz, aromatic H), 4.37 (d, 1H, *J* = 13.6 Hz, CH₂OH), 4.01 (d, 1H, *J* = 9.4 Hz, CH₂NCH), 3.57-3.53 (m, 2H, NHCHCH and CH₂NCH), 3.37 (s, 1H, CH₂NCH), 3.30 (d, 1H, *J* = 13.8 Hz, CH₂OH), 2.97-2.93 (m, 1H, NCH₂CH₂), 2.24-2.21 (m, 1H, NCH₂CH₂), 2.10-2.04 (m, 2H, NCH₂CH₂ and NHCHCH), 1.67-1.64 (m, 1H, NCH₂CH₂).

¹³C NMR (100 MHz, CDCl₃): δ 164.5 and 162.1 (d, ¹*J*_{C,F} = 243.4 Hz, aromatic C), 149.8 (aromatic C), 149.0 (2) (pyridinyl C), 144.6 (pyridinyl C), 143.6 and 143.5 (d, ³*J*_{C,F} = 8.4 Hz, aromatic C), 130.3 (aromatic C), 130.1 and 130.0 (d, ¹*J*_{C,F} = 8.9 Hz, aromatic C), 127.9 (aromatic C), 127.6 (aromatic C), 123.6 (2) (pyridinyl C), 121.7 and 121.7 (d, ⁴*J*_{C,F} = 1.7 Hz, aromatic C), 118.2 (aromatic C), 114.7 (aromatic C), 113.0 and 112.8 (d, ²*J*_{C,F} = 21.5 Hz, aromatic C), 112.8 and 112.6 (d, ²*J*_{C,F} = 21.0 Hz, aromatic C), 64.7 (CH₂NCH), 64.2 (CH₂NCH), 56.0 (CH₂OH), 54.5 (NHCHCH), 51.4 (NCH₂CH₂), 35.8 (NHCHCH), 25.7 (NCH₂CH₂).

¹⁹F NMR (376 MHz, CDCl₃): δ -113.4.

[α]_D²² = -30.4° (c = 0.4, CHCl₃).

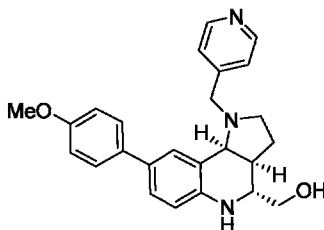
Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et₃N in CO₂, λ = 210 nm): ³/₄(major) = 6.4 min, ³/₄(minor) = 7.0 min.

IR (thin film, cm⁻¹): ν_{max} 3334, 2927, 1608, 1522, 1484, 1326, 1261, 1198, 1160, 1076, 868, 819, 782, 752, 694.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 95% by UV, rt = 0.87 min, m/z 390.2 (M+H)⁺, m/z 434.6 (M+FA-H)⁻.

HRMS (ESI, *m/z*): calcd for C₂₄H₂₄FN₃O (M+H)⁺: 390.1982, found: 390.1976.

((3aS',4i?,9b^)-8-(4-Methoxyphenyl)-1-(pyridin-4-ylmethyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2-c]quinolin-4-yl)methanol (8/BRD0739)



Prepared from benzyl (3aS',4i?,9b5)-4-(hydroxymethyl)-8-(4-methoxy-phenyl)-2,3,3a,4,5,9b-hexahydro-1H-pyrrolo[3,2-c]quinoline-1-carboxylate **13h** (143 mg, 0.32 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 70% ethyl acetate in hexane gave the desired product **8** as an off-white solid (62 mg, yield 48%).

R_f = 0.67 (silica gel, 10% methanol in dichloromethane, UV)

^1H NMR (400 MHz, CDCl_3): δ 8.46 (d, 2H, J = 4.8 Hz, aromatic H), 7.44-7.43 (d, 2H, J = 8.2 Hz, aromatic H), 7.32-7.27 (m, 4H, aromatic H), 6.96 (d, 2H, J = 8.2 Hz, aromatic H), 6.73 (d, 1H, J = 8.2 Hz, aromatic H), 4.39 (d, 1H, J = 13.8 Hz, CH_2OH), 3.99 (d, 1H, J = 10.0 Hz, CH_2NCH), 3.85 (s, 3H, OCH_3), 3.58-3.52 (m, 2H, CH_2NCH and NHCHCH), 3.37 (s, 1H, CH_2NCH), 3.29 (d, 1H, J = 13.8 Hz, CH_2OH), 2.96 (t, 1H, J = 9.0 Hz, NCH_2CH_2), 2.24-2.18 (m, 1H, NCH_2CH_2), 2.10-2.07 (m, 2H, NCH_2CH_2 and NHCHCH), 1.68-1.64 (m, 1H, NCH_2CH_2).

^{13}C NMR (100 MHz, CDCl_3): δ 158.3 (aromatic C), 149.8 (aromatic C), 149.1 (2) (aromatic C), 143.7 (aromatic C), 134.0 (aromatic C), 129.9 (aromatic C), 129.3 (aromatic C), 127.4 (aromatic C), 127.3 (2) (aromatic C), 123.7 (2) (aromatic C), 118.2 (aromatic C), 114.8 (aromatic C), 114.2 (2) (aromatic C), 64.6 (CH_2NCH), 64.2 (CH_2NCH), 56.1 (CH_2OH), 55.4 (OCH_3), 54.6 (NHCHCH), 51.4 (NCH_2CH_2), 35.9 (NHCHCH), 25.7 (NCH_2CH_2).

$[\alpha]_D^{22} = -23.4^\circ$ (c = 0.4, CHCl_3).

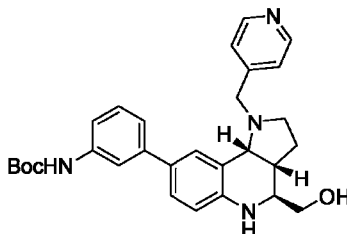
Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et_3N in CO_2 , λ = 210 nm): $\frac{3}{4}$ (major) = 6.1 min, $\frac{3}{4}$ (minor) = 7.3 min.

IR (thin film, cm^{-1}): ν_{max} 3402, 2929, 1614, 1499, 1246, 1180, 1029, 817, 753.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 95% by UV, rt = 0.80 min, m/z 402.2 ($\text{M}+\text{H}$) $^+$, m/z 446.6 ($\text{M}+\text{FA}-\text{H}$) $^-$.

HRMS (ESI, m/z): calcd for $\text{C}_{25}\text{H}_{27}\text{N}_3\text{O}_2$ ($\text{M}+\text{H}$) $^+$: 402.2182, found: 402.2172.

***tert*-Butyl (3-((3*a*?,4*S'*,9*bi*?)*-*4-(hydroxymethyl)-1-(pyridin-4-ylmethyl)-2,3,3*a*,4,5,9*Z*-hexahydro-1*H*-pyrrolo[3,2-*c*]quinolin-8-yl)phenyl)carbamate (14)**



Prepared from benzyl (3*a*?,4*S'*,9*bi*?)*-*8-(3-(((*tert*-butoxycarbonyl)amino)-phenyl)-4-(hydroxymethyl)-2, 3, 3*a*, 4, 5, 9*b*-hexahydro-1*H*-pyrrolo[3,2-*c*]quinoline-1-carboxylate **13i** (210 mg, 0.53 mmol) according to **General Procedure B**. Purification by flash column chromatography eluting with 60% to 90% ethyl acetate in hexane gave the desired product **14** as a white solid (205 mg, yield 79%).

R_f = 0.28 (silica gel, 10% methanol in dichloromethane, UV).

^1H NMR (400 MHz, CDCl_3): δ 8.45 (d, 2H, J = 5.0 Hz, aromatic H), 7.50 (s, 1H, aromatic H), 7.35-7.29 (m, 5H, aromatic H), 7.18 (d, 1H, J = 7.3 Hz, aromatic H), 6.71 (d, 1H, J = 8.2 Hz, aromatic H), 4.34 (d, 1H, J = 13.8 Hz, CH_2OH), 3.98 (d, 1H, J = 10.2 Hz, CH_2NCH), 3.58 (d, 1H, J = 10.2 Hz, CH_2NCH), 3.53-3.51 (m, 1H, NHCHCH), 3.36-3.32 (m, 2H, CH_2NCH and CH_2OH), 2.99-2.97 (m, 1H, NCH_2CH_2), 2.27-2.25 (m, 1H, NCH_2CH_2), 2.10-2.03 (m, 2H, NCH_2CH_2 and NHCHCH), 1.69-1.67 (m, 1H, NCH_2CH_2), 1.54 (s, 9H, *tert*-butyl CH_3).

^{13}C NMR (100 MHz, CDCl_3): δ 171.2 (CONH), 152.9 (aromatic C), 148.8 (aromatic C), 144.3 (aromatic C), 142.0 (aromatic C), 138.8 (aromatic C), 130.3 (aromatic C), 129.3 (aromatic C), 127.9 (aromatic C), 123.8 (aromatic C), 121.0 (aromatic C), 116.5 (aromatic C), 116.4 (aromatic C), 114.8 (aromatic C), 80.5 (*tert*-butyl C), 64.6 (CH_2NCH), 64.3 (CH_2NCH), 56.0 (CH_2OH), 54.4 (NHCHCH), 51.4 (NCH_2CH_2), 35.8 (NHCHCH), 28.4 (*tert*-butyl CH_3), 25.8 (NCH_2CH_2).

$[\alpha]_D^{22} = +57.6^\circ$ (c = 0.5, CHCl_3).

Chiral SFC (AS-H, 1.5 mL/min, MeOH with 0.05% Et_3N in CO_2 , λ = 210 nm): $\frac{3}{4}$ (minor) = 6.5 min, $\frac{3}{4}$ (major) = 6.9 min.

IR (thin film, cm^{-1}): ν_{max} 3425, 2930, 1706, 1606, 1514, 1366, 1241, 1162, 1065, 788, 754, 699.

LCMS (UV Chromatogram, 210 nm, 2.5 min run): Purity > 85% by UV, rt = 0.99 min, m/z 487.6 ($\text{M}+\text{H}$) $^+$, m/z 531.7 ($\text{M}+\text{FA}-\text{H}$) $^-$.

HRMS (ESI, m/z): calcd for $C_{29}H_{35}N_4O_3$ ($M+H$)⁺: 487.2709, found: 487.2720.

Example 2. Assays for detection of RNA guided endonuclease activities.

Fluorescence polarization-based assay

Binding to PAM-site (NGG for SpCas9) is an important first step in target recognition by SpCas9. Since disruption of PAM-site binding (e.g., by mutating SpCas9 residues involved in PAM-site recognition or mutating PAM-sequence) disrupts SpCas9 activity, it was hypothesized that inhibitors disrupting SpCas9's PAM-binding also render SpCas9 inactive. The affinity of SpCas9 for a single PAM site is weak, which portends well for identifying potential inhibitors, but the low affinity creates a challenge in developing a robust assay for SpCas9-PAM binding activity.

SpCas9 affinity for DNA sequences increases monotonically with the increase in the number of PAM sites. Fluorescence polarization (FP) is a useful technique to monitor protein:DNA interaction, including for example, Cas9-gRNA (ribonucleoprotein) complex and target DNA (12PAM). It was discovered that the fluorescence polarization signal of the fluorophore appended to the DNA increased when target DNA bound to the [Cas9:guideRNA] complex (FIG. 1A). The following fluorophore-labelled PAM-rich target DNA was used:

5'-GGCTGG.ACCACGCGGGAAAATCCACCTAGGTGGTTCCTCTTCGGA.TGTTCCATCCTTT/36-FAM-3'
3'-CCGACCTGGTGCGCCCTTTTAGGTGGATCCACCAAGGAG,AAGCCTACAAGGTAGGAA.A-5'

As the target DNA is much smaller than the [Cas9:guideRNA] complex, the target DNA's tumbling rate is significantly reduced upon binding to [Cas9:guideRNA] complex. Exploiting the above results, an assay was developed that monitors the change in the fluorescence polarization of the fluorophore-labelled PAM-rich target DNA (henceforth called 12PAM-DNA) upon binding to [Cas9:guideRNA] complex. In this assay, the complexation of [Cas9:guideRNA] to 12PAM-DNA showed a dose-dependent increase in fluorophore polarization (FIG. 1B).

The FP-assay was validated using competition and differential scanning fluorimetry experiments. In the competition experiment, 12PAM-DNA was competed with DNA sequences containing a varying number of PAM-sites. A drop in FP-signal of 12PAM-DNA was observed that was proportional to the number of PAM-sites on the competitor DNA

(FIG. 1C). Next, differential scanning fluorimetry was used, which can detect the perturbation in thermal stability of a protein upon ligand binding. The melting temperature of the [SpCas9:guideRNA] complex was determined in the presence of DNA sequences with varying number of PAM-sites (FIG. 1D). The melting temperature increase was proportional to the number of PAM-sites on the target DNA. It was found that Z-prime of the FP-assay (in a 384-well format) is between 0.6-0.8, which further confirms the robustness of the assay.

Motivated by these findings, a pilot screening of ~15,000 compounds was performed in two replicates (FIG. 2A). These compounds belonged to Broad Institute's "informer set" and the "performance diverse set" (Wawer et al, *Proc Natl Acad Sci USA* 2014, 111, 10911-6) which primarily includes compounds from commercial sources. Broad Institute has ~100,000 diversity-oriented synthesis compounds in various libraries. However, screening all of these compounds has the potential to be inefficient, as compounds within a single library are relatively similar to each other, and may perform similarly in assays. The Computational Chemical Biology group at the Broad Institute has established a list of ~10,000 compounds, called the "Informer set," that maximally represent the diversity across all diversity-oriented synthesis compounds. Finally, 12PAM-DNA without fluorophore was used as a positive control. Compounds were classified as "hits" (circled in red; FIG. 2A) if they lowered the FP-signal by values greater than 3 standard deviations (3σ) of DMSO.

All the "hits" came from Broad Institute's in-house libraries and none from the commercial libraries. Interestingly, compounds from specific libraries were observed to be highly enriched in hits pointing to the strong structure-activity relationship. Most excitingly, dose curves of "hits" from spirocyclic-azetidine library demonstrated stereochemical-dependence on the activity (FIG. 2B); compounds, BRD4172 and BRD1656, are diastereomers that have > 15-fold difference in IC_{50} . Without intending to be bound by theory, the stereochemical-activity relationship points to specific nature of the interaction between these compounds and SpCas9. The differential potency of these compounds was able to be confirmed in a DNA cleavage assay.

In a typical experimental protocol, a plasmid DNA substrate containing target gene corresponding to the guide RNA (gRNA) was incubated with Cas9/gRNA complex at 37 °C for 10 min before quenching by addition of EDTA. For compound testing, Cas9/gRNA RNP complex was incubated with compounds at the indicated concentration for 30min before introducing cleavable plasmid DNA into it. Reaction mixtures were then analyzed by running on an agarose gel with ethidium bromide as the DNA staining agent.

Briefly, SpCas9 activity was followed by quantifying the amount of cleaved product of a target gene using agarose DNA gel (FIGS. 2E and 2F). Dose-dependence was observed in a reduction in the amount of cleaved product only for the two most potent compounds from the screen (FIGS. 2E and 2F).

Spinach transcription assay

A mechanism-independent assay was developed to assess any Cas9 nuclease activity *in vitro*. Thus, an *in vitro* transcription based assay was developed wherein the transcribed mRNA is the "spinach aptamer" that fluoresces in the presence of a small molecule (Paige et al., *Science* 201 1, 333, 642-6) (FIG. 3A; Table 1). Coupling of an *in vitro* transcription (IVT) reaction produces the RNA aptamer Spinach which, upon binding to the small molecule DFHBI, produces a fluorescent complex. A synthetic gene-like construct was designed to use the bacteriophage T7 RNA Polymerase (T7 RNAP) to drive the production of the Spinach RNA aptamer. This dsDNA construct, termed a 'genelet,' consists of a T7 RNAP promoter upstream of the region that codes for the spinach RNA. Cas9 gRNAs designed to bind to and/or cleave PAM-containing sites within the Spinach DNA template (FIG. 3B), would be able to interfere with T7 RNAP transcription and inhibit production of a functional DFHBI-binding RNA, and hence decrease fluorescence (FIG. 3A).

Table 1. Spinach Transcription Assay sequences

Spinach aptamer	GGGAGACGCAACUGAAUGAAAUGGUGAAGGACGGGUCCAGGUGUG GCUGCUUCGGCAGUGCAGCUUGUUGAGUAGAGUGUGAGCUCCGCG UAACUAGUCGCGUCAC
Guide RNA	GCUAUAGGACGCGACCGAAAGUUUUAGAGCUAGAAAUAGCAAGUUA AAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCG GUGCUUUU
Spinach DNA NonTemplate Strand	TAATACGACTCACTATAGGACGCGACCGAAATGGTGAAGGACGGGT
Spinach DNA Template Strand	ACCCGTCCTTCACCATTTTCGGTCGCGTCCTATAGTGAGTCGTATTA
Spinach DNA Template	GCGCGCTTTCTAATACGACTCACTATAGGGTGACGCGACCGAAATGG TGAAGGACGGGTCCAGTGCTTCGGCACTGTTGAGTAGAGTGTGAGCT CCGTAACCTGGTCGCGTC
Spinach RNA aptamer	GGUGACGCGACCGAAAUGGUGAAGGACGGGUCCAGUGCUUCGGCA CUGUUGAGUAGAGUGUGAGCUCCGUAACUGGUCGCGUC
Spinach DNA Template (general)	GCGCGC>NNNNTAATACGACTCACTATAGGG>NNNNGACGCGACCGAA ATGGTGAAGGACGGGTCCAGTGCTTCGGCACTGTTGAGTAGAGTGTG AGCTCCGTAACCTGGTCGCGTC
Spinach RNA aptamer (general)	GGNNNNGACGCGACCGAAAUGGUGAAGGACGGGUCCAGUGCUUCG GCACUGUUGAGUAGAGUGUGAGCUCCGUAACUGGUCGCGUC

T7 Promoter = TAATACGACTCACTA

Cpf1 PAM = TTTN

SpCas9 PAM = NGG

SaCas9 PAM = NNGGGT

NNNN represents any nucleotide A, G, T, or C. The length of the string of NNNN is arbitrary, and can be expanded to accommodate the PAM consensus motif of any RNA-programmable DNA nuclease. The first "NNNN" site accommodates distal-PAM binding nucleases such as those of the Cpf1 family, while the second "NNNN" site accommodates both distal and proximal PAM binding nucleases (such as those of the Cpf1 family and Cas9 family, respectively).

Using a T7 polymerase for transcription, as low as 1 nM of the spinach gene was able to be detected (FIG. 3B). When [SpCas9: guideRNA] complex was added to the transcription mix, transcription was blocked as evident by the lack of increase in fluorescence

over time (FIG. 3C). Note that addition of SpCas9 or guideRNA alone did not block transcription. Finally, it was found that one of the potent spirocyclic compounds from the primary screening successfully blocked the transcription inhibition of [SpCas9: guideRNA] complex.

Without being bound by theory, as long as the correct PAM sequence is present in the DNA template, it is possible to use any Cas nuclease with the appropriate gRNA (Table 1). Analysis of the Spinach sequence revealed a number of NGG sites evenly distributed throughout the sequence, allowing for preliminary optimization of the assay with SpCas9. Indeed, titrating the amount of DNA template used (0.1 nM) was able to detect nanomolar levels of SpCas9 activity using a guide RNA that targeted site Sp-g1 (FIG. 3C). This activity was dependent on the SpCas9 concentration and was highly dependent on the site of cleavage - scanning the length of the spinach sequence with 4 different SpCas9 guides (Sp-g1 through g4) revealed that binding events 5' to the DFHBI-binding L_{12} loop resulted in fluorescence loss, while binding after this loop produced fluorescence (FIG. 3D).

While this indicated that no modifications would be needed to assess this assay in the context of SpCas9, it hindered the generalizability of the assay to include Cas9 nucleases with more complex PAM recognitions. Indeed, the spinach gene only contained only one NNGGGT and TTTN site each, which are the PAM recognition sequences for SaCas9 and AsCpf1/LbCpf1, respectively. To overcome this limitation, additional nucleotides were inserted that could accommodate arbitrary PAM sites - one between the T7 promoter and the spinach gene (proximal site, intended for 3'-PAM binding Cas enzymes), and one upstream of the T7 promoter (distal site, intended for 5'-PAM binding Cas enzymes like Cpf1. The proximal site contained a TAGGGT SaCas9 PAM, and the distal site contained a TTTC Cpf1 PAM (FIG. 3B). Early termination of spinach transcription resulted in optimal fluorescence loss. Without being bound by theory, it was reasoned that these sites would allow direct targeting of the T7 promoter to completely abolish transcription. When comparing the activity of SaCas9 with a guide RNA targeting an internal spinach site (Sa-g1) and the proximal variable site (Sa-g2), comparable loss of DFHBI fluorescence was observed with nanomolar levels of SaCas9 (FIG. 3E).

To assess the generalizability of our assay, it was assessed whether the activities 3 different Cpf1 orthologs - *Acidaminococcus* sp. Cpf1 (AsCpf1), *Lachnospiraceae bacterium ND2006* Cpf1 (LbCpf1), or *Francisella tularensis subsp. novicida* Cpf1 (FnCpf1) could be sensitively detected. In general, the Cpf1 orthologs had lower cleavage efficiency compared

to the Cas9 nucleases, as was previously reported (Kim, 2016), although nanomolar detection was observed (FIGS. 4A-4C). Of the orthologs tested, FnCpfI exhibited the widest dose response range with similar activity for the two target sites tested (FIG. 4C). While successful inhibition of fluorescence was observed for both AsCpfI and LbCpfI, their efficiencies were much lower than FnCpfI, SpCas9, or SaCas9, and required >200 fold excess protein to detect cleavage of 0.1 nM DNA (FIGS. 4A-4B). Interestingly, each CpfI ortholog yielded different activity depending on the gRNA site, with the installed distal PAM site generally being active toward all CpfIs tested. LbCpfI was not able to cleave the endogenous TTTC PAM site, but was very active toward the distal site PAM (FIG. 4B). This trend of activity was reversed for AsCpfI, although it was capable of cleaving both targets (FIG. 4A). In all cases, denaturing gels confirmed the expected sizes of gRNAs and crRNAs.

EGFP disruption assay

Joung and co-workers (Fu et al, *Nat Biotechnol* 2013, 31, 822-6; Kleinstiver et al, *Nature* 2015, 523, 481-5) have reported a U2OS.EGFP cell line wherein knockdown of EGFP gene by leads to loss of EGFP fluorescence (FIGS. 3K and 3L). By quantifying the percentage of EGFP positive cells using flow cytometry, one can estimate SpCas9 activity. The low throughput, flow cytometry-readout was replaced with a high-throughput readout using a high-content, automated microscope imaging. This assay can be run in a 96-well format and the data can be analyzed by a high-throughput image analysis platform, called ImageExpress. A 70-80% reduction in EGFP positive cells was observed upon SpCas9 mediated knockdown. Further, it was found that most of FP-assay "hits" inhibited SpCas9 in cells, as the compound-treated cells did not lose EGFP fluorescence (FIG. 3M). It has been confirmed that these "hits" are active in the Surveyor assay, which is considered to be the gold standard for Cas9 activity in cells, but is highly tedious and low-throughput. These compounds successfully inhibited indel formation in the EMX1 gene in HEK293T cells (FIG. 3N) at 10-20[^]M of compound concentration.

Strand invasion assay

To measure the Cas9 nuclease activity, a technique was designed based on DNA strand invasion. It was hypothesized that, after a DSB by Cas9 on the substrate DNA, the fluorophore labeled (FAM) 5'-end of the non-target strand can be replaced by a corresponding single-stranded cold DNA (FIG. 4B). Without being bound by theory, the

displacement of the cleaved fluorophore labeled 5'-end by cold DNA leads to the loss of fluorophore from the Cas9:gRNA-DNA ternary complex which results in a decrease in the fluorescence polarization signal (FIG. 4C). Thus, the extent of loss in the fluorescence polarization signal provides a readout of the Cas9 nuclease activity. It is anticipated that this assay would be useful to evaluate the potency of a Cas9 inhibitor.

Initially, performed a fluorescence polarization assay was performed with increasing concentration of Cas9:gRNA keeping the substrate DNA concentration constant and a dose curve was generated (FIGS. 4D and 4E). From the dose curve two RNP:DNA ratios were picked below the saturation point and the displacement assay was performed. While binding of Cas9:gRNA with DNA template enhanced the fluorescence polarization value, addition of excess of unlabeled DNA corresponding to the 5'-end revert back the signal in a time dependent manner (FIGS. 4F and 4G). Moreover, addition of the excess cold DNA to the substrate DNA did not affect its FP value indicating a lack of interference with the assay. These results indicated that, Cas9 nuclease activity can be followed up using this fluorescence polarization based strand-invasion assay. It is anticipated that, the strand-invasion assay can be used to screen Cas9 inhibitors in a high throughput manner.

Strand displacement assay.

In another approach, the previously described strand invasion assay was modified to make it more sensitive and effective with an orthogonal readout of fluorescence instead of fluorescence polarization. In this assay, the substrate DNA remained the same as the strand invasion assay, though the sequence of the invading cold DNA was changed in such a way that it can hybridize with the 5'-end free DNA available only after the Cas9 mediated cleavage. Moreover, the DNA strand was conjugated with a fluorescence quencher at the 3'-end which can readily quench FAM fluorescence only when it hybridizes with the labile non-target strand (FIG. 5A). As a proof of principle, the DNA strand bearing a quencher on the 3'-end was incubated with both c³/4-target DNA and ss-non-target DNA and the fluorescence signal measured. It was found that the 3'-quencher-DNA strand efficiently diminished the FAM fluorescence only in the ss-non-target DNA in a ratiometric fashion while there was no effect on the c³/4-target DNA (FIGS. 5B and 5C). Based on these result, fluorescence quencher based strand displacement technique was applied to measure Cas9 nuclease activity in a biochemical setup. Upon addition of the quencher strand into the Cas9 reaction, a significant drop in the FAM fluorescence signal was observed. Without being bound by

theory, this indicated formation of FAM-quencher DNA duplex with the nicked non-target strand. The phenomenon was validated in different Cas9 compositions with increasing quencher concentration and a systematically enhanced readout was observed (FIG. 5D). To further confirm that the fluorescence signal drop was due to the Cas9 cleavage and FAM-quencher DNA duplex formation, the samples were analyzed by gel electrophoresis (FIG. 5E). The gel electrophoresis results showed the appearance of two cleaved DNA fragments in presence of Cas9. Moreover, upon introduction of the quencher strand, a clear disappearance of the larger DNA fragments corresponding to the 5'-end cleaved product could be observed while a new band appeared. Without being bound by theory, the appearance of the new band was due to the FAM-quencher DNA duplex formation. These results confirmed this design strategy can quantitatively detect Cas9 cleaved DNA product and hence provide a way to measure Cas9 activity.

Although SpCas9 binds to its DNA substrate with nanomolar affinity, even following double stranded cleavage, it was discovered that of the 4 resulting DNA fragments, the distal non-target strand is weakly held, and can be displaced upon addition of excess complementary single stranded DNA (Richardson, 2016) (FIG. 1E). In this system, the 5' end of the non-target strand is fluorescently labeled, and the fluorescence is quenched in a cleavage-dependent manner by adding in excess a complementary DNA strand labeled with a 3' quencher. Upon displacement and annealing of the two strands, fluorescence is quenched by a FRET mechanism, thereby providing a proxy measurement for Cas9 activity at the RuvC domain based on the extent of fluorescence loss. Double stranded oligos (DS-oligo) labeled with 6-carboxyfluorescein were generated containing either a TGG PAM motif for recognition by SpCas9, or ACC and TGC PAMs that should not be recognized by SpCas9 (Table 2). Cleavage of DS-oligo substrate oligos by SpCas9:gRNA was verified by monitoring the FAM fluorescence in a denaturing gel (FIG. 1F), validating the PAM dependence on activity.

An oligo complementary to the 5'-end of the non-target strand was generated, containing an Iowa-Black FQ quencher on the 3' terminus (Q-oligo) (Table 2). Excess Q-oligo (5 nM) could not disrupt the fluorescence of duplex DS-oligo (1 nM), but was capable of quenching the FAM-labeled strand outside of a duplex (SS-oligo, 1 nM). When SpCas9:gRNA complex (5 nM) was added, a significant loss of fluorescence was observed. This activity was dependent on gRNA-mediated cleavage of DNA and not local DNA melting caused by Cas9 binding to the PAM motif, as addition of ApoCas9 to DS-oligo and

Q-oligo did not result in fluorescence loss (FIG. 1G). In agreement with the substrate cleavage observed via denaturing gel, strand displacement was dependent on the correct TGG PAM motif, as no quenching was observed with the ACC or TGC PAM oligos (FIG. 1H). Strand displacement is generalizable to SaCas9 with comparable efficiency to SpCas9, and is dependent on an NNGGT PAM sequence (FIG. II).

Table 2. Strand Displacement Assay sequences

SpCas9 DNA Substrate	5'-6-FAM/ <u>TAATACGACTCACTATAGG</u> ACGCGACCGAAAT TGGT GAAGGACGGGT-3'
SaCas9 DNA Substrate	5'-6-FAM/ <u>ACTCACTATAGGGACGCGACCGAAATGGTGAAGGACGGGT</u> CCAGTGCTTCGG-3'
Cpf1 (all species) DNA Substrate	5'-CGTCCTTCACCAT TTTCGGT CGCGTCCCTATAGTGAGTCGTATTAGT <u>TCCAT</u> /6-FAM-3'
DNA Substrate	5'-6-FAM/ <u>ATGGA</u> ACTAATACGACTCACTATAGGGACGCGACCG AAATG GTGAAGGACG-3'
SpCas9 Quencher Strand	5'-ATAGTGAGTCGTATTA/3IABkFQ-3'
SaCas9 Quencher Strand	5'-CGTCCCTATAGTGAGT/3IABkFQ-3'
Cpf1 (all species) Quencher Strand	5'-5IABkFQ/ATGGAACTAATACGAC-3'
Quencher Strand	5'-GTCGTATTAGTTCCAT/3IABkFQ-3'

DNA substrates for the strand displacement assay shown are double stranded, and only include a fluorophore (6-FAM) on the strand shown.

Quencher strand sequences for the strand displacement assay shown are single stranded, and include a quencher (Iowa Black® FQ).

The ratio of Q-oligo and Cas9:gRNA to the DS-oligo substrate was characterized by testing relative ratios of 1:1, 1:2, 1:5, 1:10, and 1:20, and found that a 5-fold excess of each

reagent relative to DS-oligo is sufficient to yield maximum quenching (FIGS. 1J and IK). Using these optimized conditions, the assay was capable of detecting low (<5 nM) nanomolar quantities of SpCas9 (FIG. 1L). Optimizing the kinetics of this assay showed that 2.5 hours was sufficient to see > 80% quenching at 37°C (FIG. 1M). In agreement with previous reports, SpCas9 activity at room temperature was very weak although observable using our strand displacement assay (FIG. 1M). Thus, the assay provides a sensitive, specific, and potentially high-throughput readout of SpCas9 nuclease activity, at least as it pertains to the RuvC nuclease domain.

Encouraged by the success at assessing SpCas9 activity, the assay scheme was applied to other CRISPR nucleases. It was possible that such generalizability might be hindered by lack of detailed studies on the catalytic mechanisms of Cas9-nucleases from other classes and bacterial species. However, given the similarities between *Staphylococcus aureus* Cas9 (SaCas9) and SpCas9 protein fold and modes of DNA substrate binding, SaCas9 strand displacement was tested to see if it would proceed in the same manner. Using FAM-labeled oligos containing an SaCas9-recognizable ACGGGT PAM and a ACGGTT non-target PAM (ref Friedland 2015) with the appropriate Q-oligo, PAM- and Cas9:gRNA-dependent loss of fluorescence was observed (FIG. II) with similar efficiencies and detection limits as SpCas9. Similar results were observed with Cas nucleases of the Cpf1 family, particularly FnCpf1.

Example 3. Identification of small molecule inhibitors using assays for detection of RNA guided endonuclease activities.

Using the primary screening assay (FP-assay) (FIG. 1A) and multiple secondary assays (e.g., DNA cleavage, Spinach aptamer, EGFP disruption, and Surveyor assays), several SpCas9 inhibitors were identified with moderate potencies in biochemical ($IC_{50} = 600\text{nM}$) and cellular assays ($\sim 10\text{ }\mu\text{M}$). It was found that "hits" belonged to specific diversity-oriented synthesis libraries (FIG. 30). Schrodinger software was used to perform docking of selected active Cas9 inhibitors. The calculations show that active compounds bind at a pocket proximal to the Arginine-guanine interactions which are important PAM-recognition interactions and that electrostatic and hydrogen-bonding interactions play an important role in the binding (FIGS. 3Q and 3R). Interestingly, the inactive analogs had poorer docking scores in comparison with that of the active analogs, indicating that a computational approach was able to reproduce experimental data. Importantly, the calculations were performed with

relatively small computational effort (~1 CPU hr) as a result of the small number of rotatable bonds in the identified active compounds, raising the possibility of using this software for high-throughput screening of compound libraries.

Structure-activity optimizations involve synthesis and potency evaluation of the structural analogs of the "hits" in an iterative fashion. However, this iterative approach is tedious, labor intensive, expensive, and time-consuming. Conveniently, multiple analogs of the "hits" (FIG. 30) already exist at the Broad Institute as a part of their compound library. Structure-activity optimization entails testing compounds present in the selected libraries (FIG. 30) following the workflow described in FIG. 3P. Briefly, FP-based primary screening assay was used followed by a counterscreen assay to eliminate false positives. In the counterscreen assay, fluorophore labeled DNA template was incubated with either DMSO or compounds and the FP values measured. Thus, the counterscreen assay identified compounds that affect the fluorescence polarization of the 12PAM-DNA in the absence of [SpCas9:guideRNA] complex. Compounds that altered the fluorescence polarization signal of the fluorophore labeled DNA template was identified as false positive and subsequently removed from the data.

Following counter-screening, the "surviving hits" are tested using EGFP disruption assay and Spinach assay (FIGS. 6A-6C). Finally, DNA cleavage and Surveyor assay are used to test the "hits" of the Spinach and EGFP disruption assays (FIG. 6D). The goal is to improve the current potency from 600nM to <100nM in the biochemical assays.

Example 4. Structure-guided enhancement of DNA recognition by Cas9:gRNA complex.

Cas9:gRNA complex is considerably more tolerant to base pair mismatches between the gRNA and the target DNA at PAM-distal sites compared to PAM-proximal sites. To investigate the molecular basis for such a gradient in sequence specificity, the interactions of Cas9 with gRNA and with target DNA were analyzed in three different crystal structures of the ternary complex (FIG. 7A). The number of salt bridges, hydrogen bonds, water mediated hydrogen-bonds, and hydrophobic (stacking mostly) interactions were calculated as a function of base location from PAM site. A clear gradient in the total number of gRNA-Cas9 interactions with the base location was observed, with a broad stretch of strong interactions occurring at locations 1 to 8 bases from the PAM site and a smaller stretch of weak interactions at locations 14 to 20 bases away (FIG. 7B). In contrast, the DNA-Cas9

interaction profile was uniformly weak with no clear gradient across the 20-base stretch (FIG. 7B). Without intending to be bound by theory, these results indicate that Cas9 tightly contacts the gRNA portion of the RNA-DNA duplex near the PAM site while the DNA portion is contacted weakly throughout the duplex. Interestingly, the total number of interactions correlated well with the inverse of the measured tolerance for mismatches at each base location (FIG. 7C). Without intending to be bound by theory, it was hypothesized that the low sequence specificity of Cas9 at locations distant from the PAM site arises from the paucity of strong interactions holding gRNA at this distal region.

The results described herein were obtained using the following materials and methods.

Materials and instruments

All oligos were purchased from IDT, and were either purified by HPLC for use in strand displacement assays or by desalting for use in *in vitro* transcription experiments.. Single time point fluorescence measurements were taken using an Envision plate reader with a FITC top mirror (403), FITC 485 excitation filter (102), and BODIPY TMP FP 531 emission filter (105). Gel images were acquired with an Azure Biosystems C400 or C600.

Oligonucleotide and plasmid cleavage assays

Oligonucleotides were annealed by heating to 95°C for 5 minutes, followed by slow cooling to 25°C at a rate of 0.1°C/sec to produce a double stranded oligo (DS-oligo). Oligo-annealing solutions were prepared by mixing 10 µM of each complementary strand together in the presence of 1x Cas9 assay buffer (20 mM Tris-HCl, pH = 7.5, 150 mM KCl, 1 mM EDTA, 50 mM MgC⁺). A T7-promoter spinach (ref) sequence cloned into pUC57-Kan and linearized with AsiSI was used as the plasmid substrate for Cas9 cleavage.

A Cas9:gRNA complex was first preformed by mixing each component at a ratio of 1:1.2 (Cas9:gRNA) and incubating at room temperature for 15 minutes. Cas9:gRNA complexes (500 nM) were mixed with either 100 nM of oligonucleotide or 5 nM (100 ng) of linearized plasmid in 1x assay buffer, and incubated at 37°C for 1 hour. For oligonucleotide cleavage assays, Proteinase K (Qiagen) and RNase (Qiagen) were added to final concentrations of 200 µg/µL and 100 µg/µL, respectively, and incubated at 37°C for at least 30 minutes. Samples were boiled in loading buffer and 50 mM EDTA for 10 min, and run on

a 15% TBE-Urea gel (ThermoFisher EC68855) for 70 minutes at 200 V. FAM fluorescence was measured prior to staining with SYBR gold (ThermoFisher) to visualize total nucleotide content. For plasmid cleavage assays, loading buffer was directly added to reactions and run on 1.6-2% agarose gels with 0.01% ethidium bromide.

Fluorescence strand displacement assays (SDA)

Assay components and solutions were prepared at 10x working stocks prior to mixing. Concentrations are given as the final concentrations. In a typical assay, a Cas9:gRNA complex was first formed as described above. Cas9 without gRNA (ApoCas9) was treated similarly. DS-oligo (1 nM) was mixed with quencher oligo (Q-oligo, 5 nM) in 1x Cas9 assay buffer. Reactions were initiated by addition of Cas9:gRNA (5 nM), distributed among a 384-well plate (Corning 3575) (3 technical replicates per experiment), and incubated at 37°C for 2-3 hours. Fluorescence was read on an Envision plate reader, using 485 nm emission and 535 nm excitation wavelengths. Typical controls included replacing Cas9:gRNA with ApoCas9 (maximum possible fluorescence), replacing DS-oligo with the single stranded FAM-labeled oligo (SS-oligo, maximum possible quenching), and omitting FAM labeled oligos altogether (background fluorescence from ApoCas9 and Q-oligo). Fraction cleaved was calculated by subtracting SS-oligo controls from matched Apo-Cas9 + DS-oligo and Cas9:gRNA + DS-oligo samples and normalizing to ApoCas9 + DS-oligo samples.

Cas nuclease binding in vitro transcription Spinach assay

HiScribe T7 High Yield RNA in vitro transcription kits were purchased from NEB (E2040S). The Spinach aptamer template and non-template oligonucleotides were annealed as described above. In a typical assay, a Cas9/CpfI:gRNA complex was first formed as described above. Cas9/CpfI without gRNA (ApoCas9/CpfI) was treated similarly. A typical assay was performed by mixing the following components together from the 10x stocks to get the indicated final concentrations: NTPs (6.7 mM), 10x T7 reaction buffer (0.67x), murine RNase inhibitor (M0314L) (1.3 U), DFHBI (1 mM), DNA template (0.1 nM), and water to a final volume of 25 mL. ApoCas9 or Cas9:gRNA complexes (10x) were added to initiate cleavage and incubated at 37°C for 30 minutes. Transcription was initiated by adding 2 mL of T7 RNA polymerase, or was omitted to assess background fluorescence.

Reactions (27 μ L) were transferred to a 384-well plate and the fluorescence was monitored at 37°C.

Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

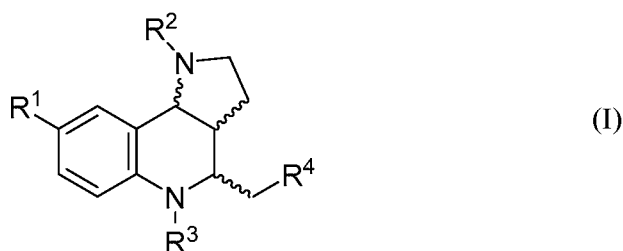
All patents and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent and publication was specifically and individually indicated to be incorporated by reference. In the case of conflict, the present specification, including definitions, will control.

CLAIMS

What is claimed is:

1. A method of inhibiting the activity of an RNA guided endonuclease-guide RNA complex, the method comprising contacting the RNA guided endonuclease-guide RNA complex with a small molecule.

2. The method of claim 1, wherein the small molecule is a compound having the formula of Formula I:



wherein R^1 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

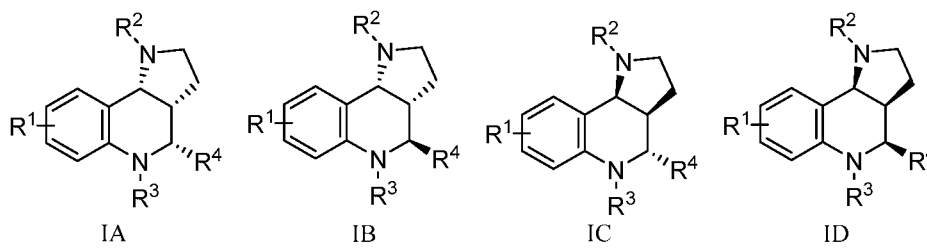
R^2 - R^4 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen;

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-$, $-(CH_2)_n-C(=O)-NH-$, $-C(=O)-NH-(CH_2)_n-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-NH-SO_2-(CH_2)_n-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)-O-$, $-R^L-NH-C(=O)-(CH_2)_n-$, $-R^L-NH-S(=O)_2-(CH_2)_n-$, -S-, -S(O)-, $-S(O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from $Ci-Ci_2$ linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from $Ci-i_2$ hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

3. The method of claim 1, wherein the small molecule is a compound of Formula IA, IB, IC, or ID:



wherein R¹ is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R²-R⁴ are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

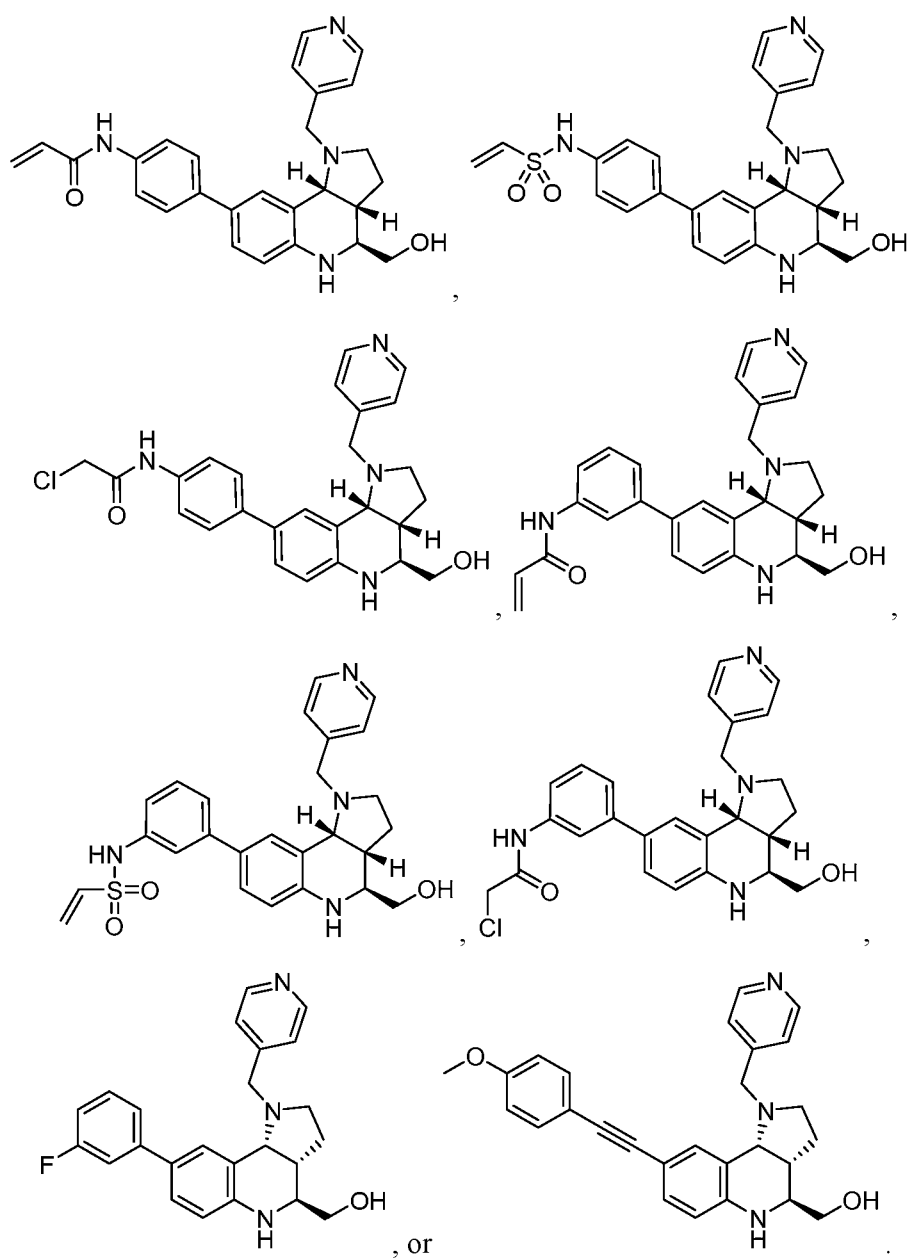
X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen;

Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)-O-, -(CH₂)_n-C(=O)-NH-, -(CH₂)_n-NH-C(=O)-, -(CH₂)_n-NH-SO₂-, -(CH₂)_n-SO₂-NH-, -(CH₂)_n-SO₂-, -(CH₂)_n-SO₂-NH-C(=O)-, -(CH₂)_n-R^L-, R^L-C(=O)-O-, -R^L-NH-C(=O)-(CH₂)_n-, -R^L-NH-SO₂-(CH₂)_n-, -S-, -S(O)-, -S(O)₂-; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

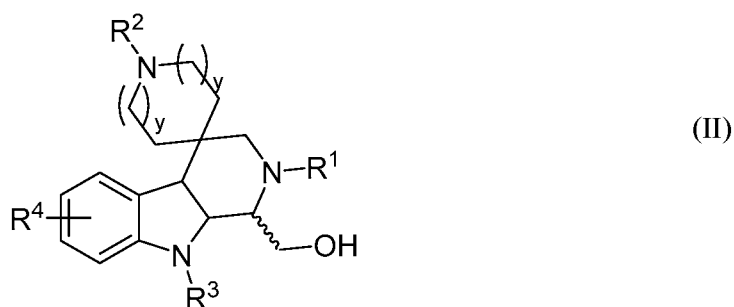
R^L is independently selected at each occurrence from C₁-C₁₂ linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C₁-C₁₂ hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

4. The method of claim 3, wherein the small molecule is selected from a compound having the structure



5. The method of claim 1, wherein the small molecule is a compound of Formula II:



wherein "y" is 0 or 1,

R^4 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R;

and

R^1 - R^3 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

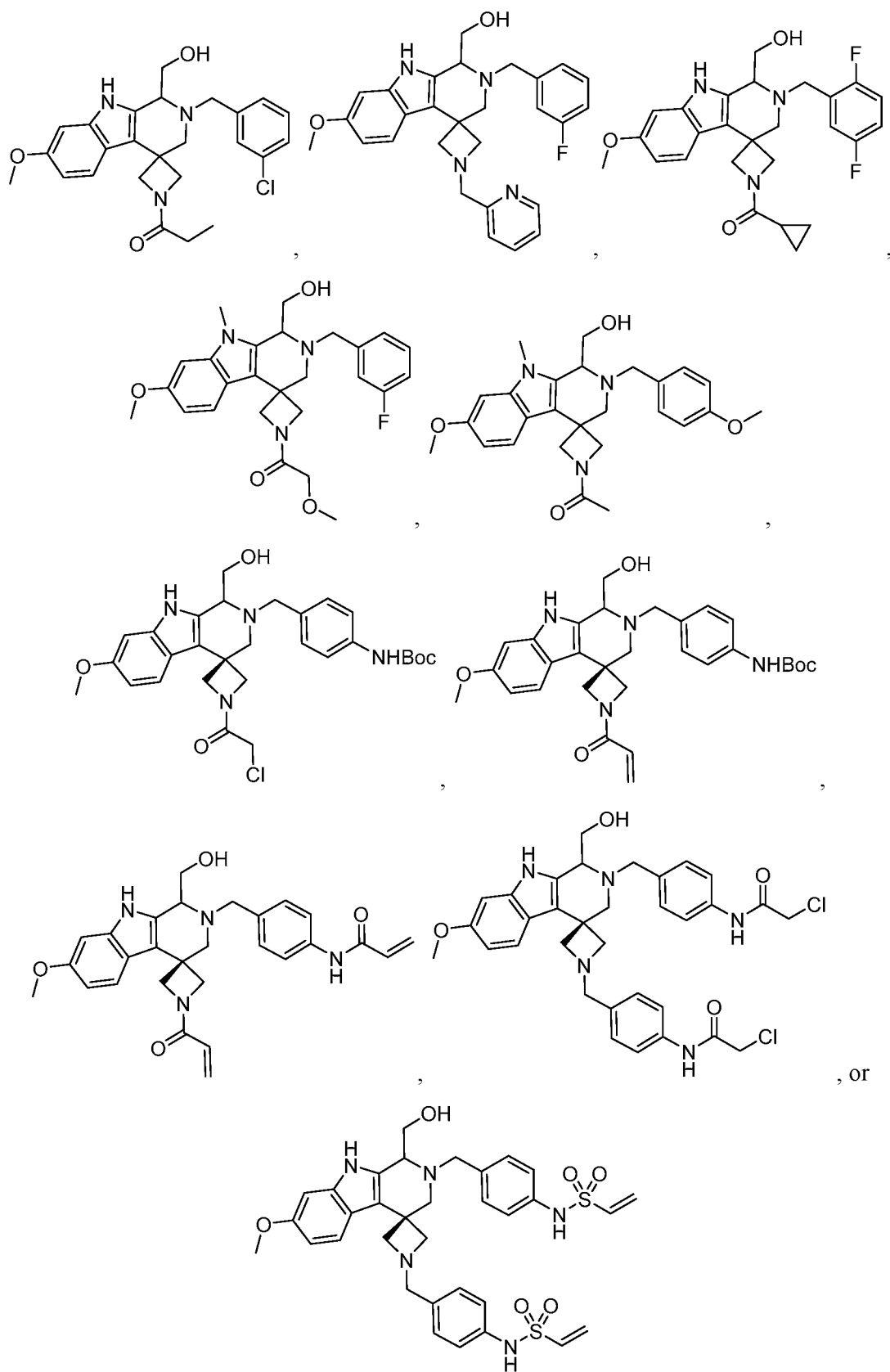
X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen ;

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-R^L-NH-C(=O)-(CH_2)_n-$, $-R^L-NH-S(=O)_2-(CH_2)_n-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)-$, $-S-$, $-S(O)-$, $-S(=O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

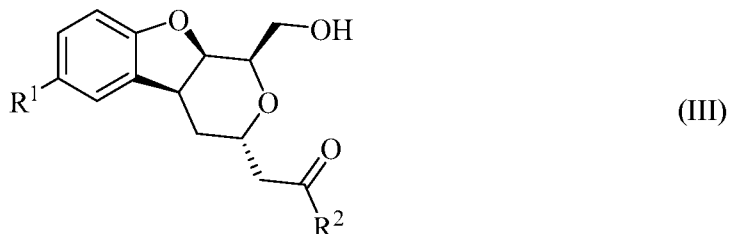
R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1 - C_{12} hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

6. The method of claim 5, wherein the small molecule has a structure:



7. The method of claim 1, wherein the small molecule is a compound of Formula III:



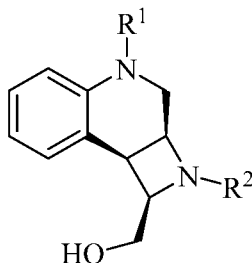
wherein R^1 - R^2 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen;

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)O-$, $-S-$, $-S(O)-$, $-S(O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1 - C_{12} hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

8. The method of claim 1, wherein the small molecule is a compound having the structure:



wherein each wavy bond may be in the R or the S configuration,

R^1 - R^2 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

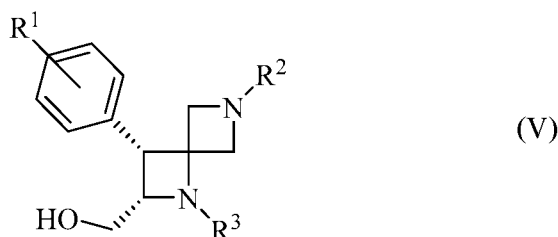
X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen;

Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)O-, -(CH₂)_n-C(=O)-NH-, -(CH₂)_n-NH-C(=O)-, -(CH₂)_n-NH-SO₂-, -(CH₂)_n-SO₂-NH-, -(CH₂)_n-SO₂-, -(CH₂)_n-SO₂-NH-C(=O)-, -(CH₂)_n-R^L-, -R^L-C(=O)O-, -S-, -S(O)-, -S(=O)₂-; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C₁-C₂ linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C₁-C₁₀ hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

9. The method of claim 1, wherein the small molecule is a compound having the structure of Formula V:



wherein R¹ is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R²-R³ are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

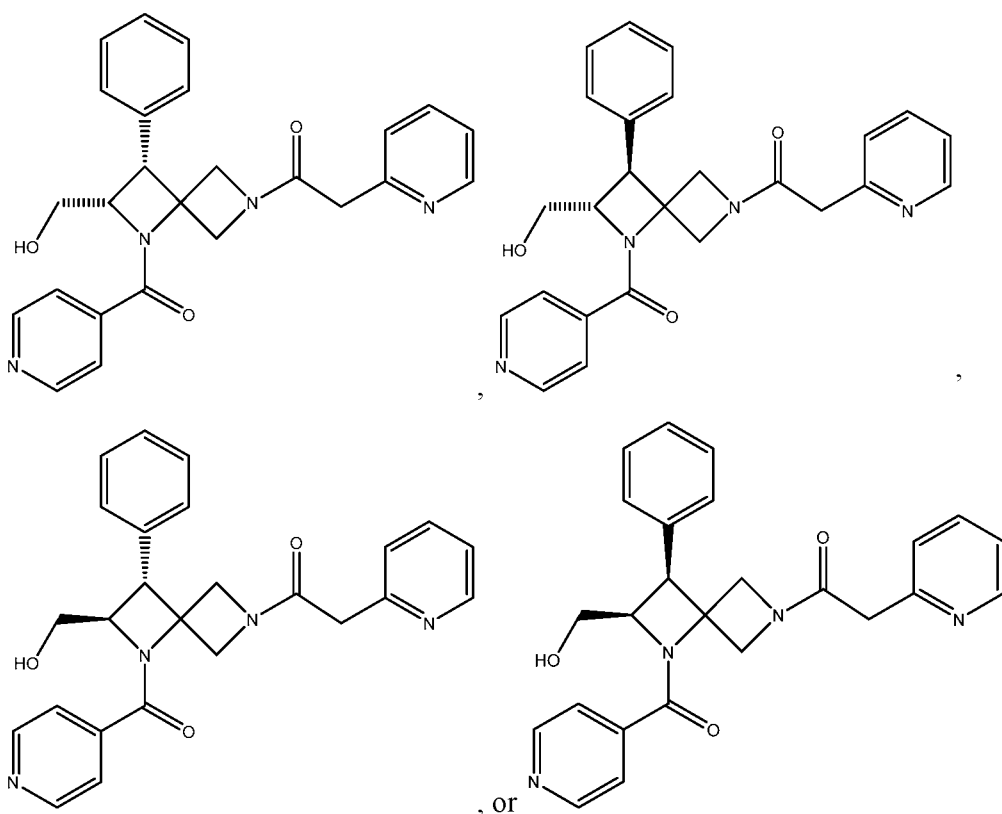
X is independently selected at each occurrence from CN, OH, CF₃, COOH, OR, OR, NR₂, or halogen;

Li is selected from -(CH₂)_n-, -(CH₂)_n-C(=O)O-, -(CH₂)_n-C(=O)-NH-, -(CH₂)_n-NH-C(=O)-, -(CH₂)_n-NH-SO₂-, -(CH₂)_n-SO₂-NH-, -(CH₂)_n-SO₂-, -(CH₂)_n-SO₂-NH-C(=O)-, -(CH₂)_n-R^L-, -R^L-C(=O)O-, -S-, -S(O)-, -S(=O)₂-; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C₁-C₂ linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C₁₋₁₂ hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

10. The method of claim 9, wherein the small molecule is selected from



11. The method of any one of claims 1-10, wherein the method is *in vitro* or *in vivo*.
12. The method of any one of claims 1-11, wherein the method is performed in a cell.
13. The method of claim 12, wherein the cell is a prokaryotic or eukaryotic cell.
14. The method of claim 13, wherein the prokaryotic cell is a bacterium.
15. The method of claim 13, wherein the eukaryotic cell is a human, mammalian, insect, plant, or yeast cell.

16. The method of any one of claims 12-15, wherein the cell is in an organism.
17. The method of claim 16, wherein the organism is a human, mammal, vertebrate, invertebrate, insect, or plant.
18. A method of detecting nucleic acid binding of an RNA guided endonuclease-guide RNA complex comprising contacting the RNA guided endonuclease-guide RNA complex with a probe comprising an oligonucleotide comprising at least 4-12 PAM sites and a fluorescent reporter at the 5' terminus; and detecting an increase in fluorescence polarization relative to a reference, thereby detecting nucleic acid binding of the RNA guided endonuclease-guide RNA complex.
19. A method of detecting nuclease activity of an RNA guided endonuclease-guide RNA complex comprising providing an RNA guided endonuclease-guide RNA complex bound to a probe comprising an oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus; and detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nucleic acid binding of the RNA guided endonuclease-guide RNA complex.
20. The method of claim 18 or 19, wherein the nucleic acid binding and/or nuclease activity is inhibited or prevented by an agent.
21. A method of identifying an agent that inhibits an RNA guided endonuclease-guide RNA complex, the method comprising contacting an RNA guided endonuclease-guide RNA complex with an agent in the presence of a probe comprising an oligonucleotide comprising at least 4-12 PAM sites and a fluorescent reporter at the 5' terminus; and failing to detect an increase in fluorescence polarization to a reference, thereby identifying the agent as an agent that inhibits an RNA guided endonuclease-guide RNA complex.
22. A method of identifying an agent that inhibits nuclease activity of an RNA guided endonuclease-guide RNA complex, the method comprising contacting an RNA guided endonuclease-guide RNA complex with an agent, wherein the RNA guided endonuclease-

guide RNA complex is bound to a probe comprising an oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus; and detecting an increase in fluorescence polarization relative to a reference, thereby identifying the agent as an agent that inhibits nuclease activity of the RNA guided endonuclease-guide RNA complex.

23. The method of any one of claims 18-22, wherein the agent is a small molecule.
24. The method of any one of claims 18-23, wherein the agent or small molecule inhibits one or more of nucleic acid binding or nuclease activity.
25. The method of any one of claims 18-24, wherein the agent or small molecule inhibits binding of the RNA guided endonuclease-guide RNA complex to a PAM site.
26. The method of any one of claims 18-25, wherein the RNA guided endonuclease is Cas9 or Cpf1.
27. The method of any one of claims 18-26, wherein the one or more PAM sites are 3' or 5' of a target nucleic acid with reference to the non-target strand.
28. The method of any one of claims 18-27, wherein the one or more PAM sites have the nucleic acid sequence 5'-NGG-3', 5'-NNGGGT-3', 5'-TTN-3', or 5'-TTTN-3'.
29. The method of any one of claims 18-28, wherein the fluorescent reporter is Fluorescein isothiocyanate (FITC).
30. The method of any one of claims 18-29, wherein the probe comprises an oligonucleotide having one or more of the nucleic acid sequences:
5'-
GGCTGGACCACGCGGGAAAATCCACCTAGGTGGTTCCTCTTCGGATGTTCCATCC
TTT/36-FAM-3',
3'-
CCGAGGCTGGTCCGCCCTTTTAGGTGGATCCACV AAGGAGAAGCCTATTAAAGGTAG
G

AAA~5',

5'-6-FAM/TAATACGACTC ACTATAGGACGCG ACCG AAATGGTGA AGGACGGGT-3',

5'-6-

FAM/ACTCACTATAGGGACGCGACCGAAATGGTGAAGGACGGGTCCAGTGC 'TTCG G~3',

5'-CGTCCTTCACCATTTCGGTCGCGTCCCTATAGTGAGTCGTATTAGTTCCAT/6-FAM-3', and

5'-6~

FAM/ATGGAACIAATACGACTCACIATAGGGACGCGACCGAAATGfflGAAGGAC G-3'.

31. The method of any one of claims 18-30 wherein the guide RNA comprises the nucleic acid sequence:

5'-

GCUAUAGGACGCGACCGAAAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAA GGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'.

32. The method of any one of claims 18-31, wherein the reference is the fluorescence polarization of the probe in the presence of the agent or small compound and in the absence of the RNA guided endonuclease-guide RNA complex.

33. A method of detecting nucleic acid binding and/or nuclease activity of an RNA guided endonuclease-guide RNA complex comprising contacting a nucleic acid expressing a nucleic acid aptamer with the RNA guided endonuclease-guide RNA complex in the presence of a fluorophore molecule, wherein binding of the aptamer to the fluorophore molecule generates a fluorescent signal, and detecting a decrease in fluorescent signal relative to a reference, thereby detecting nucleic acid binding or nuclease activity of the RNA guided endonuclease-guide RNA complex.

34. The method of claim 33, wherein the nucleic acid binding and/or nuclease activity is inhibited or prevented by an agent.

35. A method of identifying an agent that inhibits an RNA guided endonuclease-guide RNA complex, the method comprising contacting a nucleic acid expressing a nucleic acid aptamer with the RNA guided endonuclease-guide RNA complex in the presence of a fluorophore molecule and an agent, wherein binding of the aptamer to the fluorophore molecule generates a fluorescent signal; and failing to detect a decrease in fluorescent signal relative to a reference, thereby identifying the agent as an agent that inhibits an RNA guided endonuclease-guide RNA complex.

36. The method of claim 35, wherein the agent is a small molecule.

37. The method of any one of claims 35-36, wherein the agent or small molecule inhibits one or more of nucleic acid binding or nuclease activity.

38. The method of any one of claims 35-37, wherein the nucleic acid expressing a nucleic acid aptamer comprises a PAM site.

39. The method of any one of claims 35-38, wherein the RNA guided endonuclease is Cas9 or Cpf1.

40. The method of claim 38 or 39, wherein the PAM sites are 3' or 5' of a target nucleic acid with reference to the non-target strand.

41. The method of any one of claims 38-40, wherein the PAM sites have the nucleic acid sequence 5'-NGG-3', 5'-NNGGGT-3', 5'-TTN-3', or 5'-TTTN-3'.

42. The method of any one of claims 35-41, wherein the nucleic acid aptamer has the sequence 5'-

GCUAUAGGACGCGACCGAAAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAA
GGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU -3'.

43. The method of any one of claims 35-42, wherein the guide RNA comprises the nucleic acid sequence:

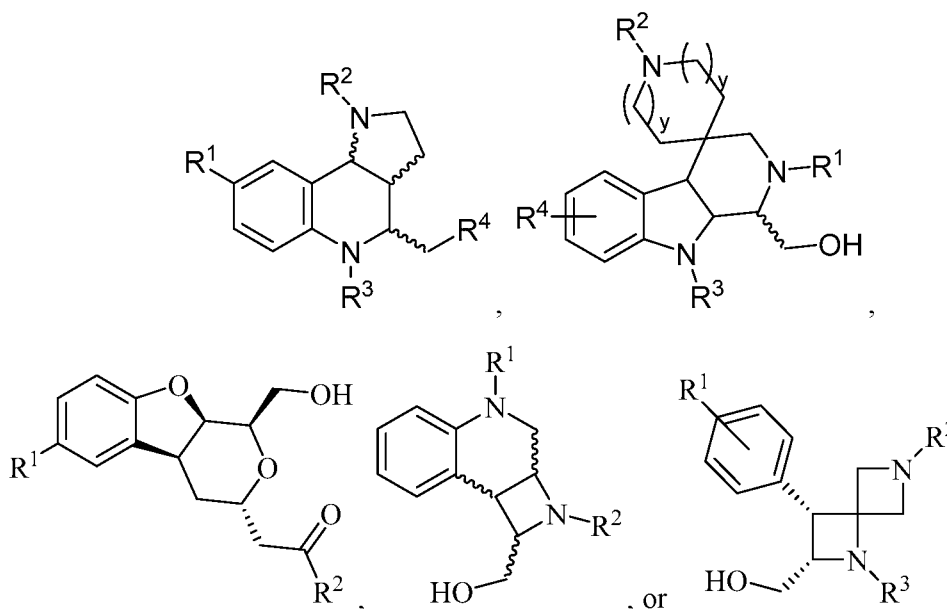
5'-

GCUAUAGGACGCGACCGAAAGUUUUAGAGCUAGAAAUAGCAAGUUAUUUUAA
GGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'

44. The method of any one of claims 35-43, wherein the fluorophore molecule is one or more of 4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1*H*-imidazol-5-one (DFHBI).

45. The method of any one of claims 35-44, wherein the reference is the fluorescence in the absence of the RNA guided endonuclease-guide RNA complex.

46. The method of any one of claims 1-45, wherein the agent or small molecule has the structure:



wherein each wavy bond may be in either the R or the S configuration, with the proviso that two wavy bonds of membered of a five or six membered fused ring have the same stereochemistry,

R^1 - R^4 are independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; where

X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen

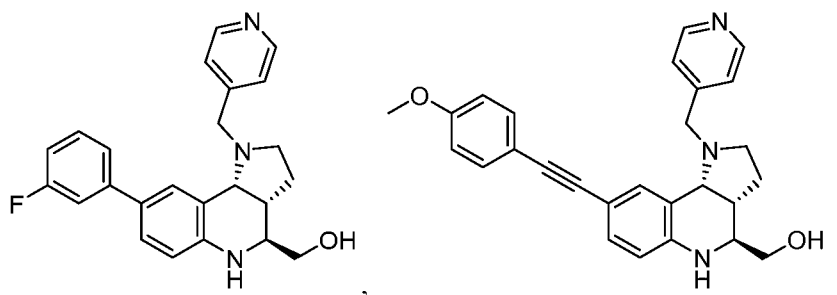
Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-R^L-NH-$

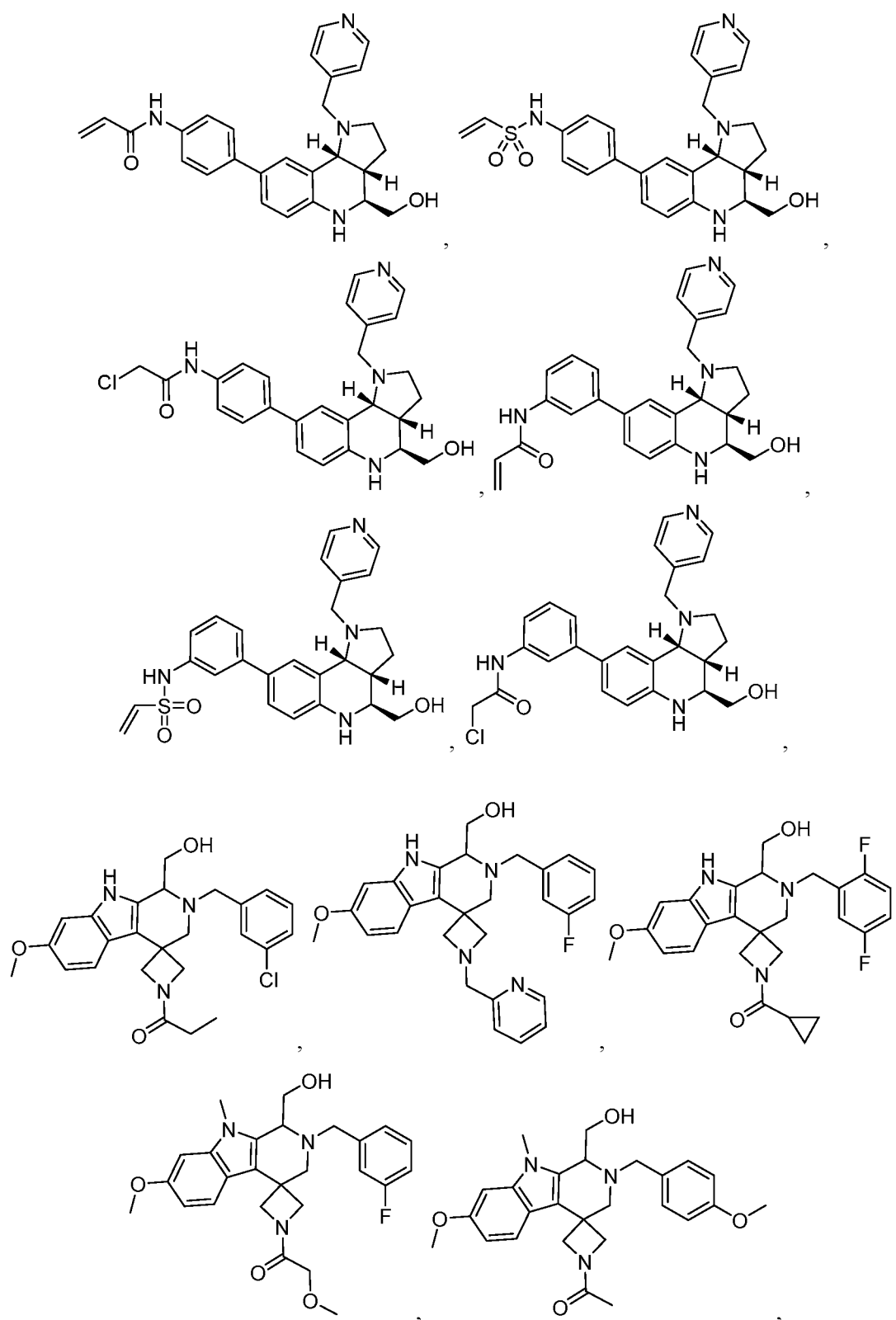
$C(0)-(CH_2)_n-$, $-R^L-NH-S(0)_2-(CH_2)_n-$, $-(CH_2)_n-R^L-$, $-R^L-C(0)-O-$, $-S-$, $-S(O)-$, $-S(0)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

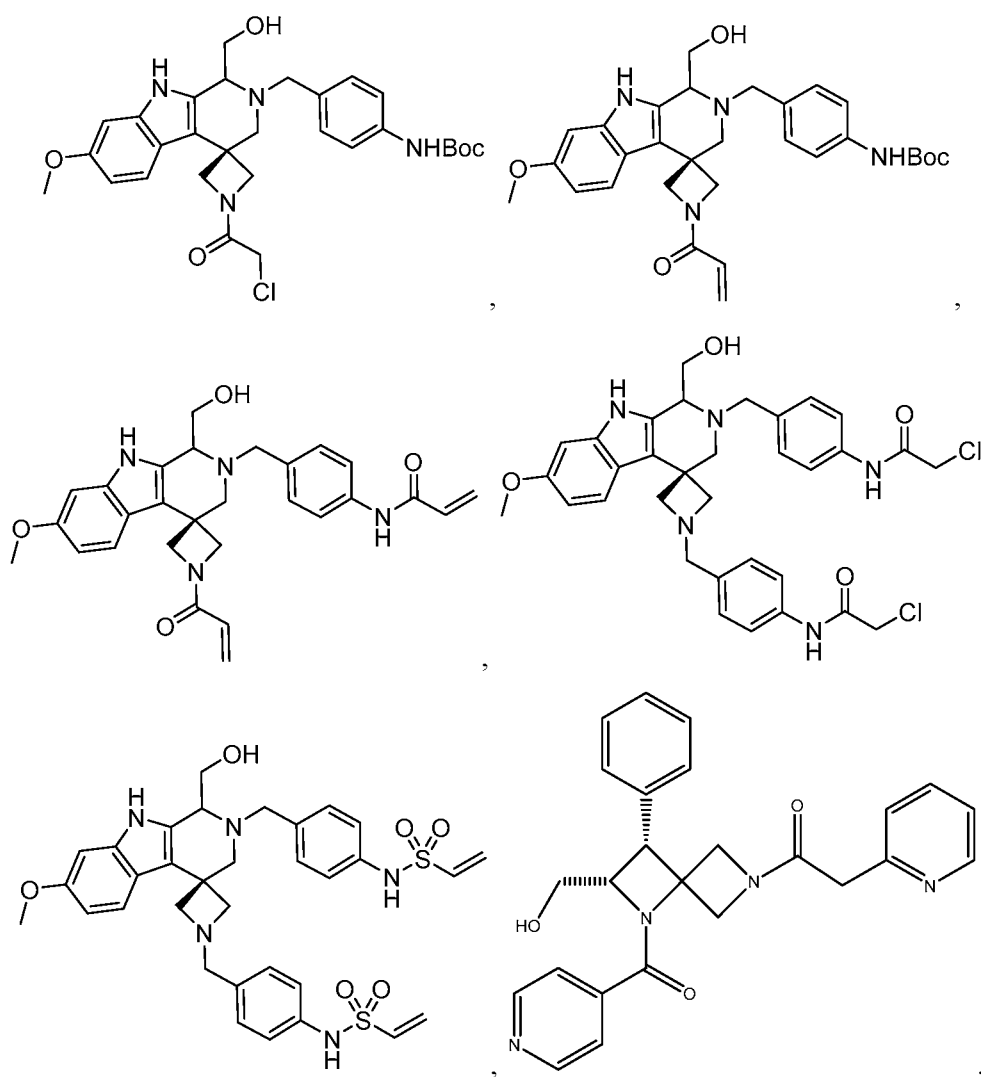
R^L is independently selected at each occurrence from C_1-C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

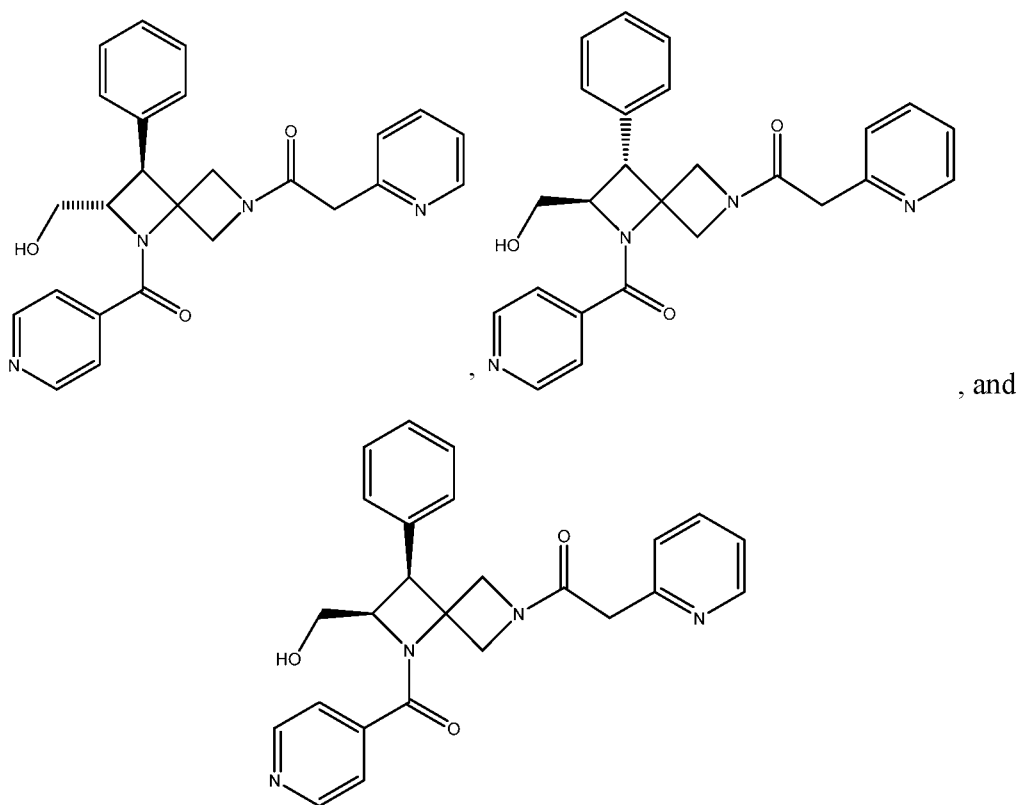
R is selected from C_1-C_{12} hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

47. The method of claim 46, wherein the agent or small molecule is selected from







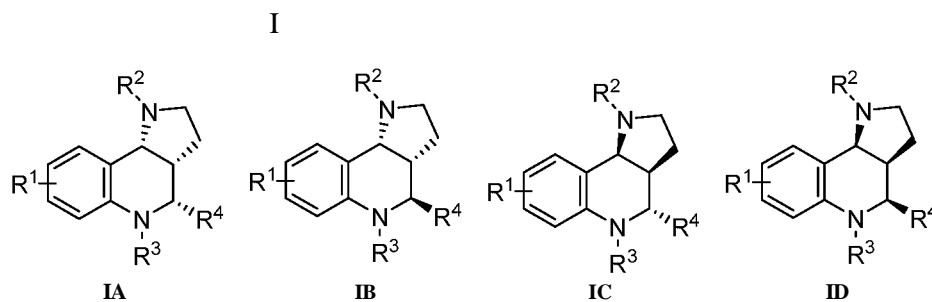


48. A method of identifying an agent that inhibits an RNA guided endonuclease-guide RNA complex, the method comprising one or more of methods 18-47.
49. The method of claim 48, further comprising assaying the agent or small compound in one or more of an EGFP disruption assay, DNA cleavage assay, or Surveyor assay.
50. An oligonucleotide comprising at least 4-13 PAM sites and a fluorescent reporter at the 5' or 3' terminus.
51. The oligonucleotide of claim 50, wherein the PAM sites are 3' or 5' of a target nucleic acid.
52. The oligonucleotide of claim 50 or 51, wherein the PAM sites have the nucleic acid sequence 5'-NGG-3', 5'-NNGGGT-3', 5'-TTN-3', or 5'-TTTN-3'.

53. The oligonucleotide of any one of claims 50-52, wherein the fluorescent reporter is Fluorescein isothiocyanate (FITC).

54. A kit comprising the oligonucleotide of any one of claims 50-53.

55. A compound having the structure of Formula IA, IB, IC, or ID:



wherein R^1 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

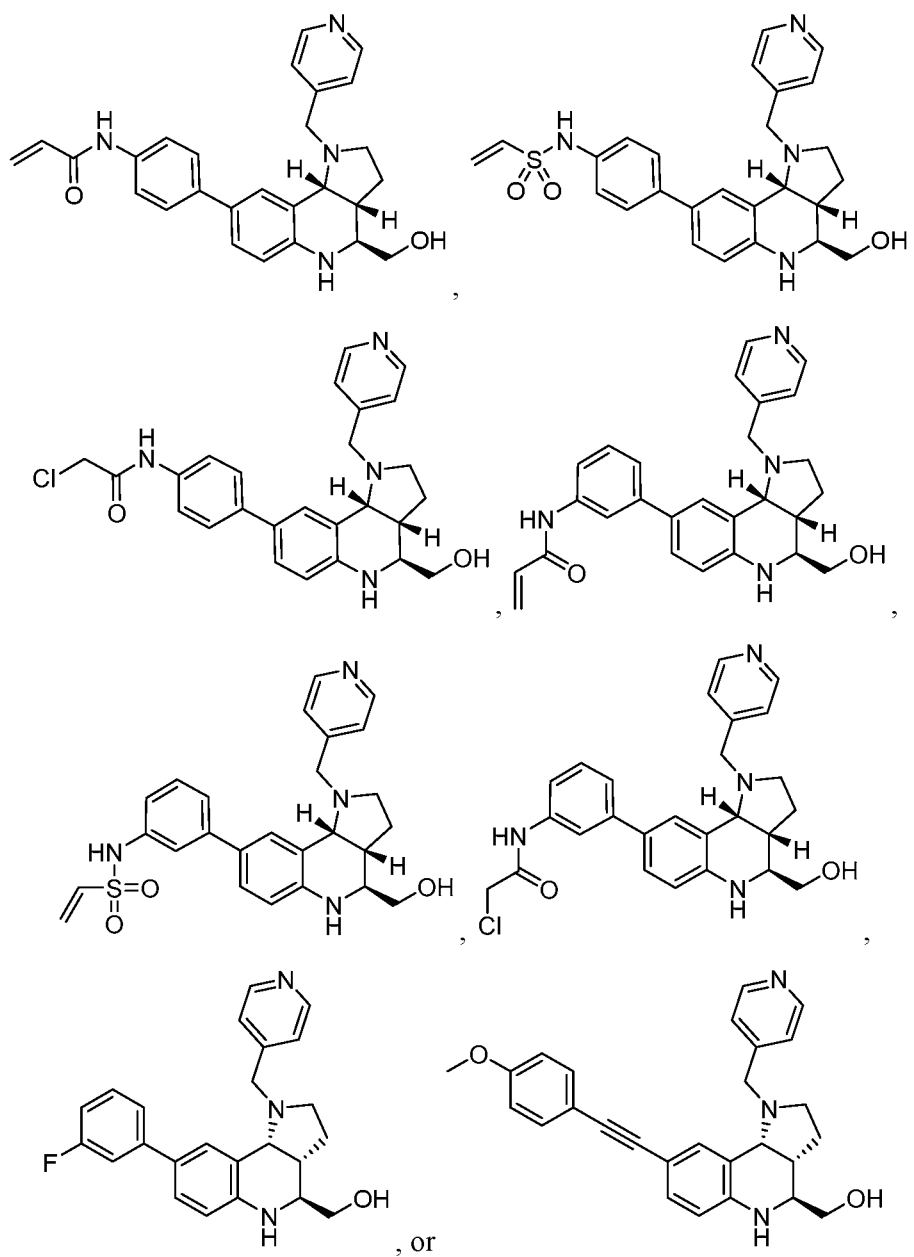
R^2 - R^4 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen;

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-R^L-NH-C(=O)-(CH_2)_n-$, $-R^L-NH-S(=O)_2-(CH_2)_n-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)-O-$, $-S-$, $-S(O)-$, $-S(=O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

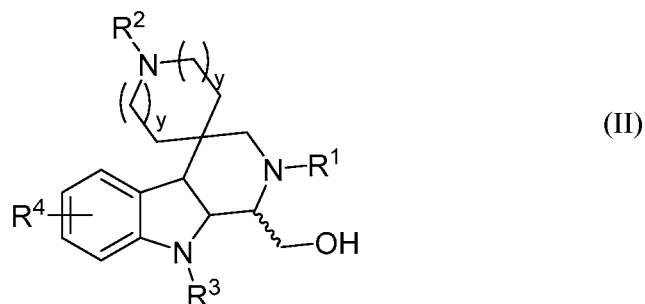
R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1 - C_{12} hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

56. The compound of claim 55 having the structure:



57. A compound having the structure of Formula (II):



wherein "y" is 0 or 1,

R^4 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R;

and

R^1 - R^3 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

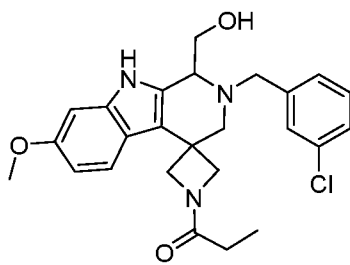
X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen;

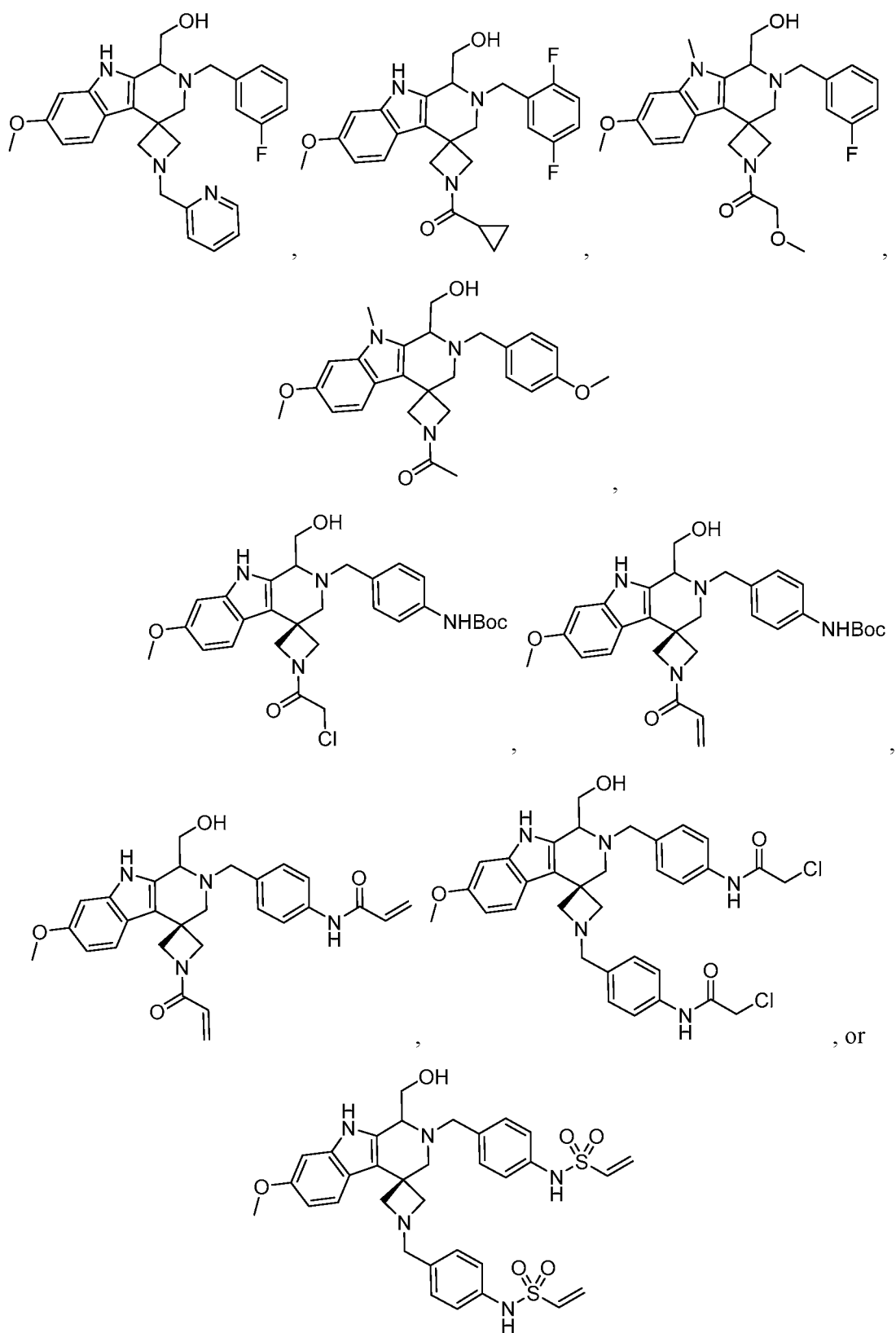
Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-R^L-NH-C(=O)-(CH_2)_n-$, $-R^L-NH-S(=O)_2-(CH_2)_n-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)-O-$, $-S-$, $-S(O)-$, $-S(=O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C_1 - C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof); optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

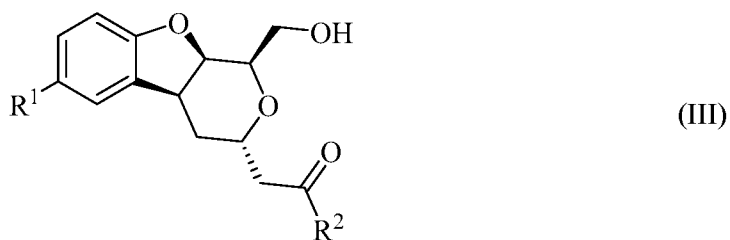
R is selected from C_1 - C_{12} hydrocarbons (e.g., alkyl, alkenyl, alkynyl, aryl, alkyl-aryl, aryl-alkyl, and combinations thereof), optionally substituted with one or more (e.g., 1-5) groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

58. The compound of claim 57 having the structure:





59. A compound having the structure of Formula (III):



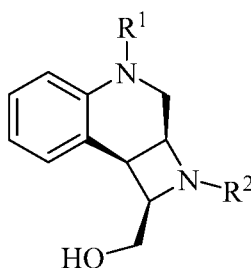
wherein R^1 - R^2 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen (*e.g.*, -Cl, -F, -Br, etc.);

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L$, $-R^L-C(=O)-O-$, $-S-$, $-S(O)-$, $-S(O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from $Ci-Ci_2$ linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from $Ci-i_2$ hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

60. A compound having the structure:



wherein each wavy bond may be in the R or the S configuration,

R^1 - R^2 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

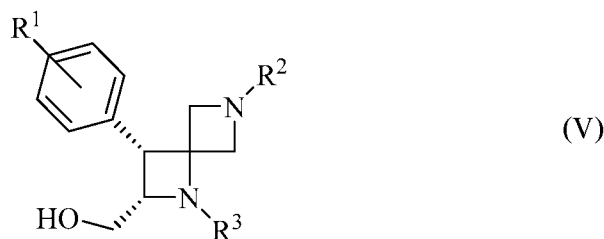
X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen;

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)O-$, $-S-$, $-S(O)-$, $-S(O)_2-$; wherein n is independently at each occurrence from 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C_1-C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1-C_{12} hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

61. A compound having the structure of formula (V):



wherein R^1 is independently selected at each occurrence from hydrogen, -X, -R, -Li-X, or -Li-R; and

R^2 - R^3 are independently selected from hydrogen, -X, -R, -Li-X, or -Li-R; where

X is independently selected at each occurrence from CN, OH, CF_3 , COOH, OR, OR, NR_2 , or halogen;

Li is selected from $-(CH_2)_n-$, $-(CH_2)_n-C(=O)O-$, $-(CH_2)_n-C(=O)-NH-$, $-(CH_2)_n-NH-C(=O)-$, $-(CH_2)_n-NH-SO_2-$, $-(CH_2)_n-SO_2-NH-$, $-(CH_2)_n-SO_2-$, $-(CH_2)_n-SO_2-NH-C(=O)-$, $-(CH_2)_n-R^L-$, $-R^L-C(=O)O-$, $-S-$, $-S(O)-$, $-S(O)_2-$; wherein n is independently at each occurrence 0, 1, 2, 3, 4, 5, or 6;

R^L is independently selected at each occurrence from C_1-C_{12} linear and/or branched and/or cyclic and/or aromatic bivalent radicals; optionally substituted with one or more groups X and/or with 1-6 heteroatoms selected from O, S, N, P, F, Cl, Br, I; and

R is selected from C_1-C_{12} hydrocarbons, optionally substituted with one or more groups X and/or with 1-10 heteroatoms selected from O, S, N, P, F, Cl, Br, I, and combinations thereof.

62. The compound according to claim 61 having the structure:

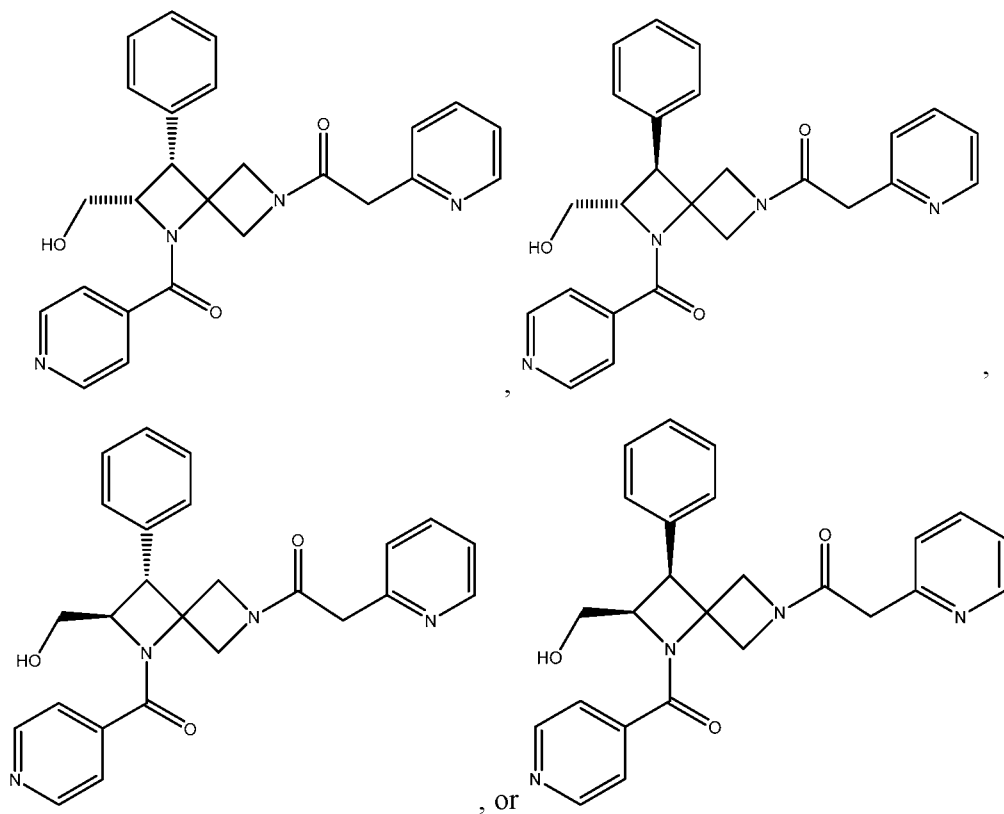


FIG. 1A

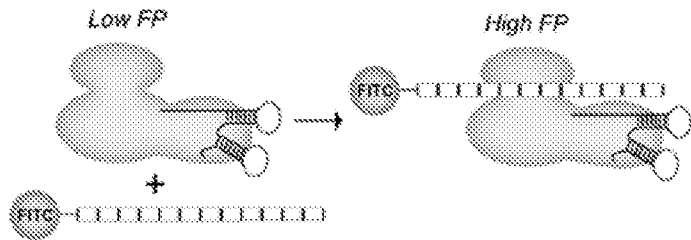


FIG. 1B

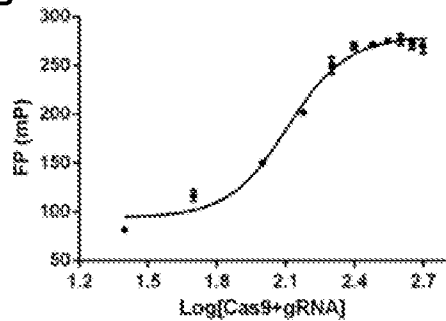


FIG. 1C

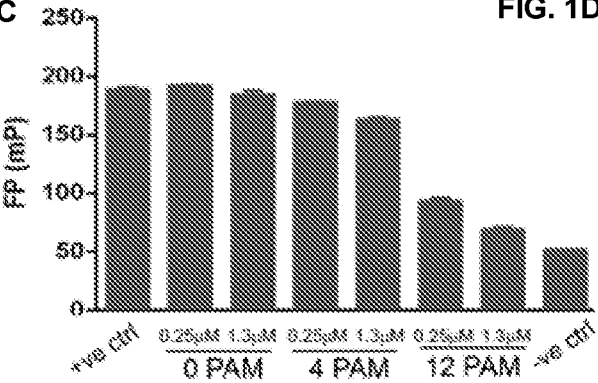


FIG. 1D

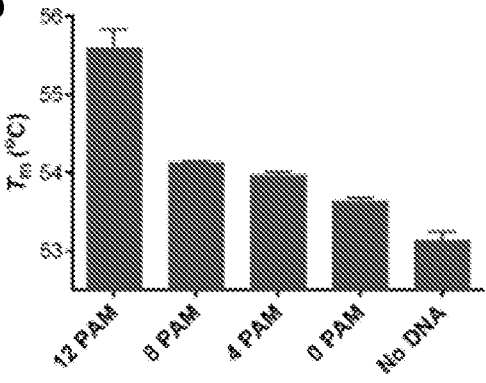


FIG. 1E

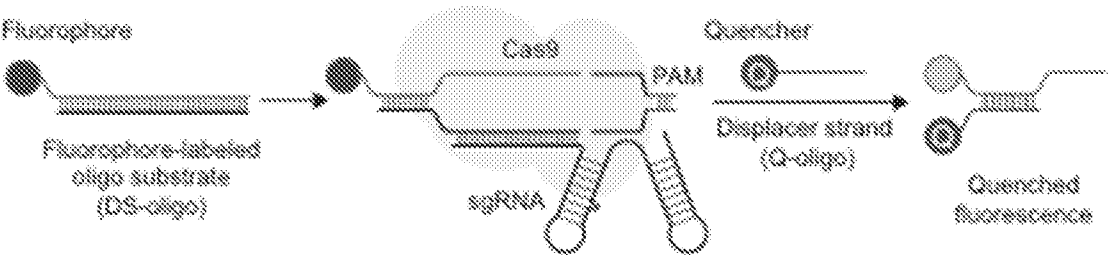


FIG. 1G

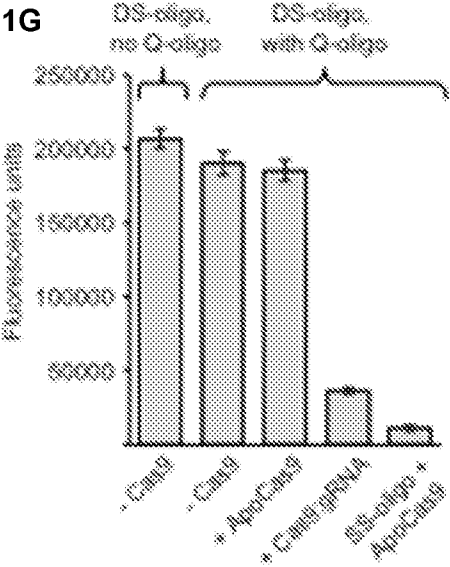


FIG. 1F

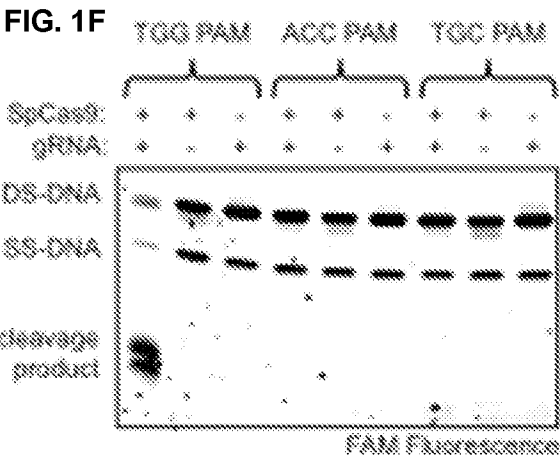


FIG. 1H

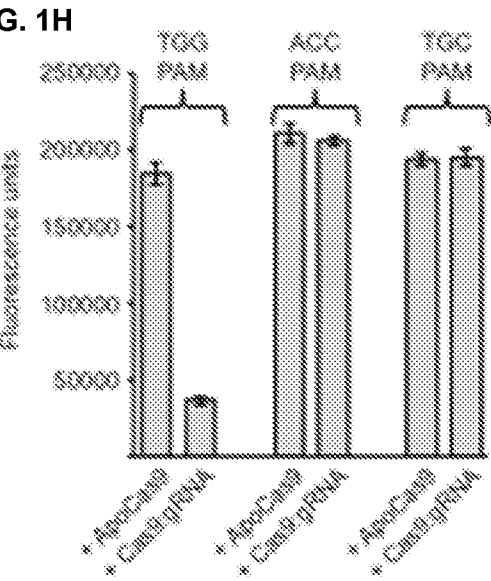


FIG. 1I

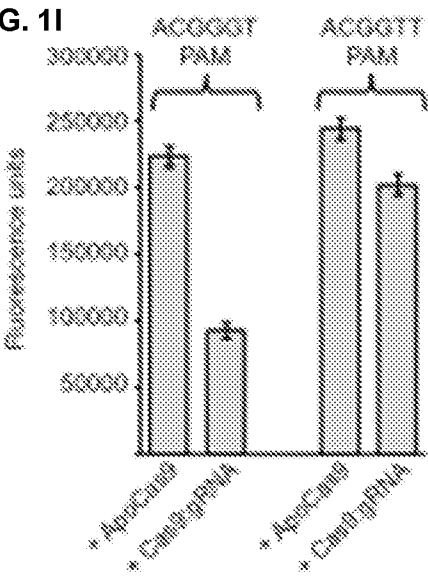


FIG. 1J

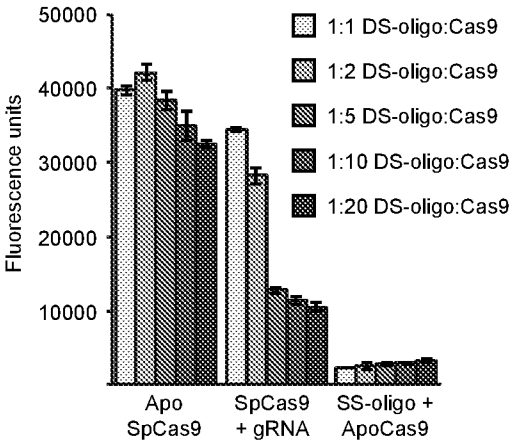


FIG. 1K

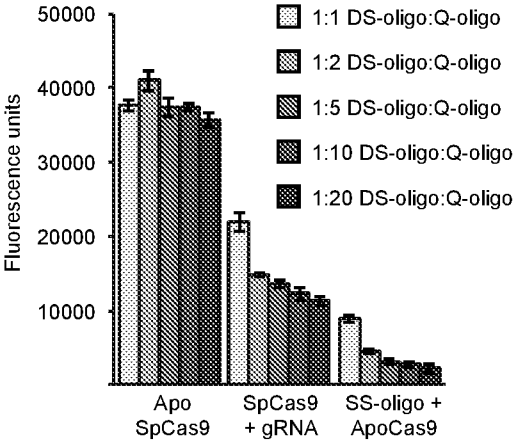


FIG. 1L

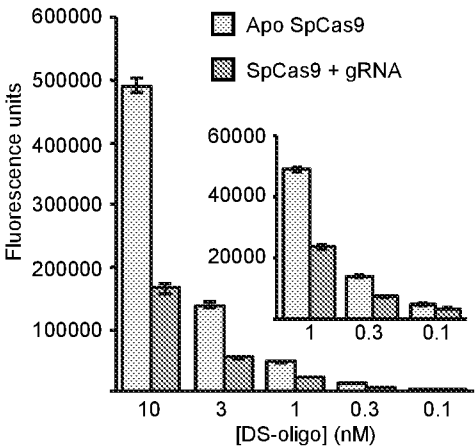


FIG. 1M

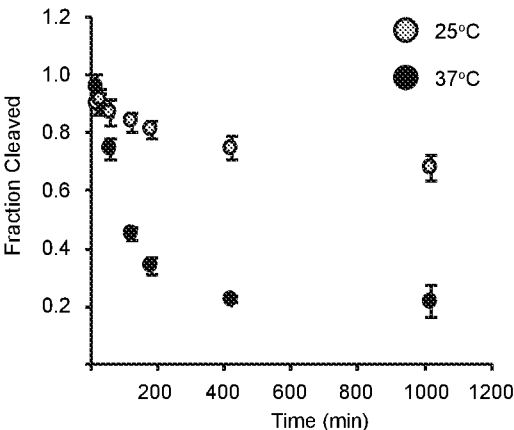


FIG. 2A

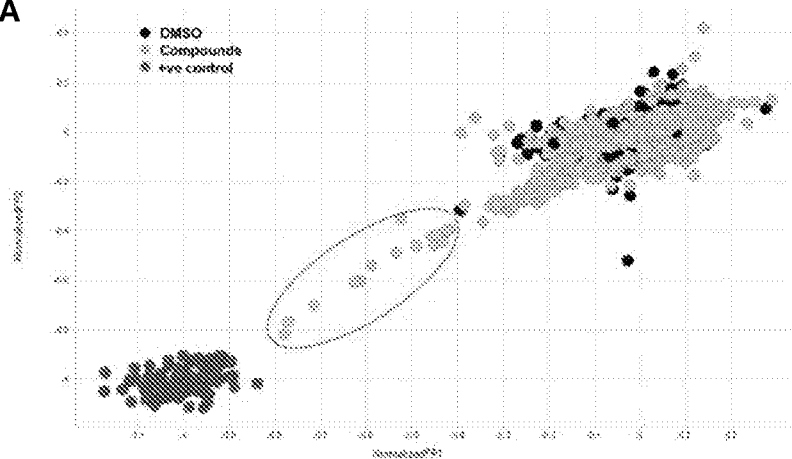


FIG. 2B

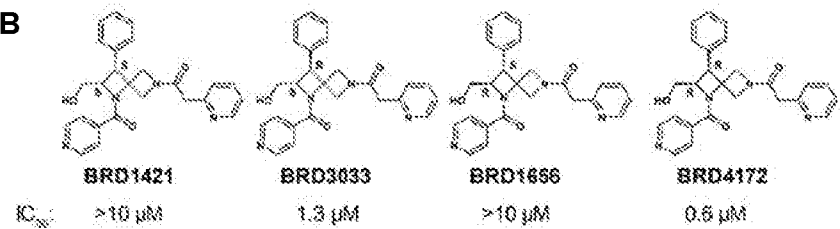


FIG. 2C

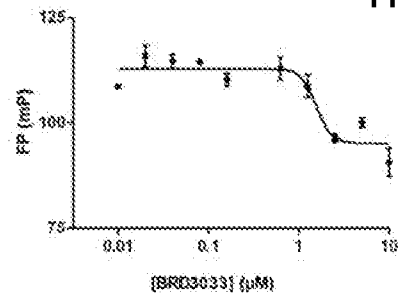


FIG. 2D

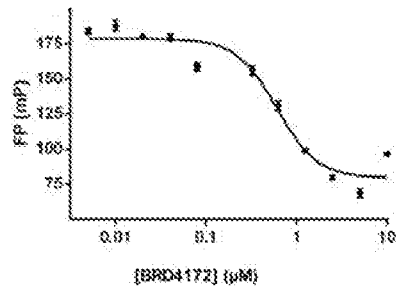


FIG. 2E

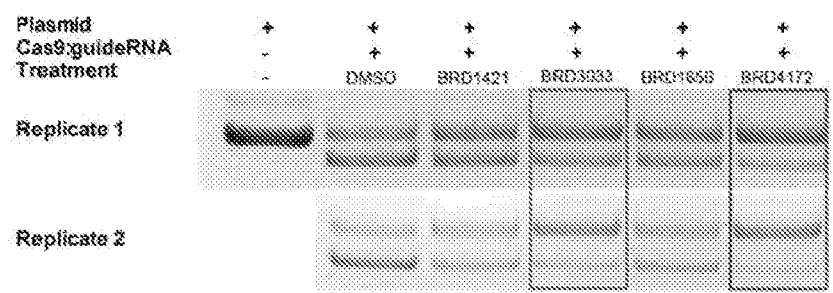


FIG. 2F

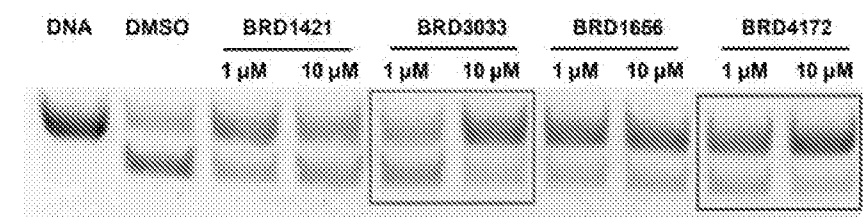


FIG. 3A

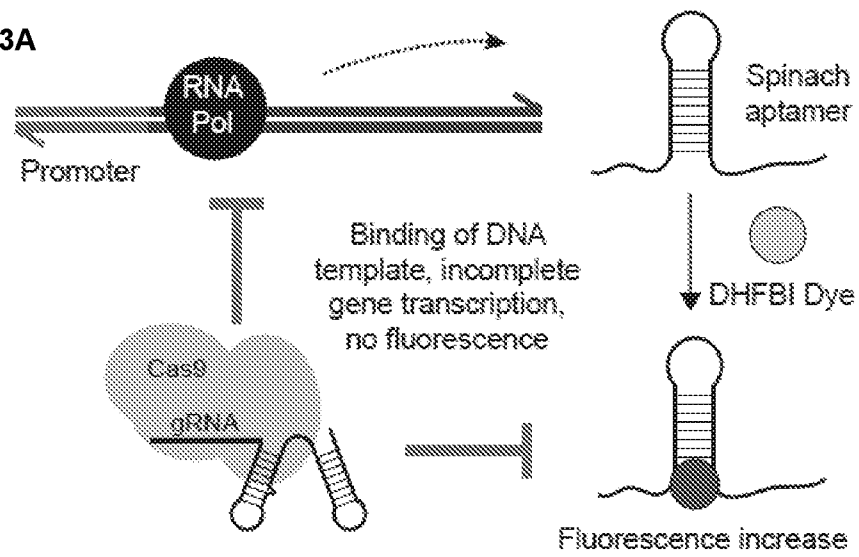


FIG. 3B

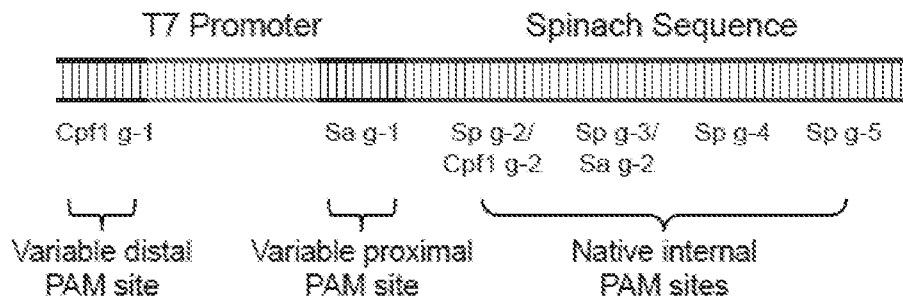


FIG. 3C

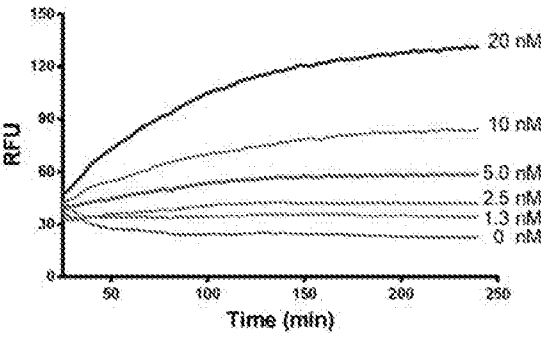


FIG. 3D

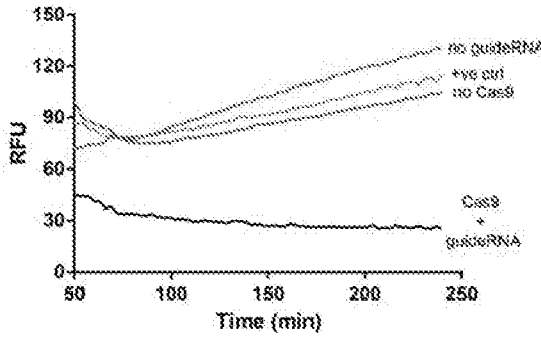


FIG. 3E

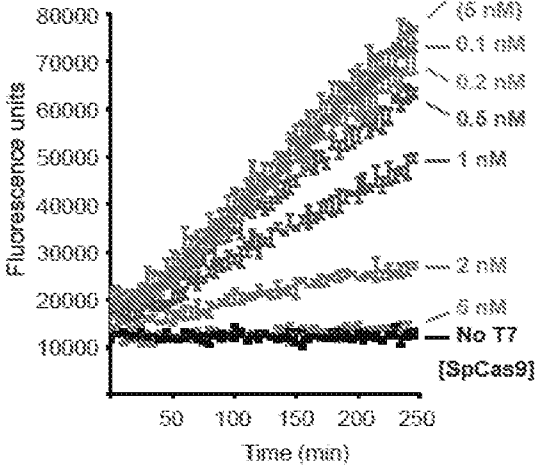


FIG. 3F

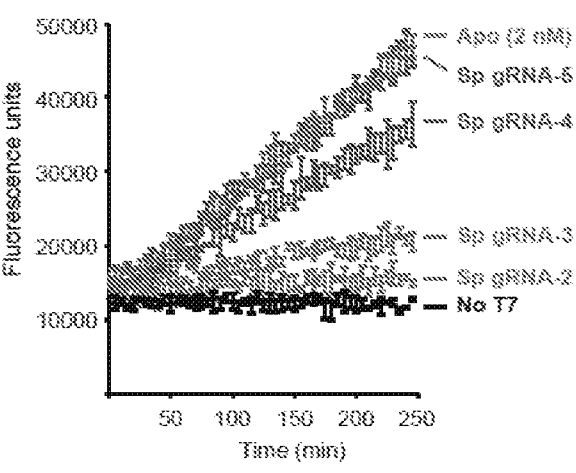


FIG. 3G

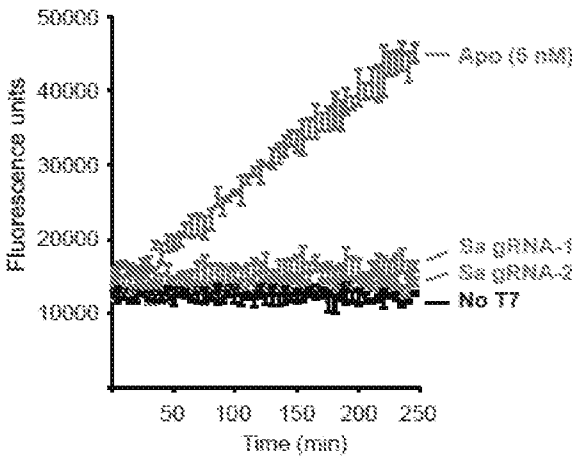


FIG. 3H

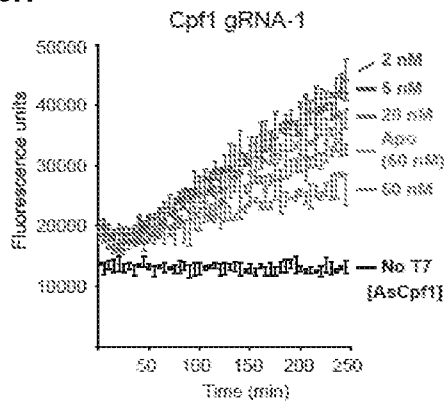


FIG. 3I

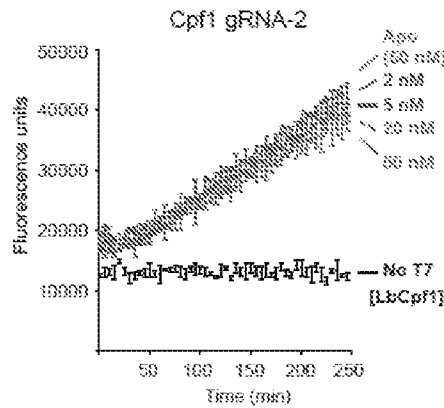
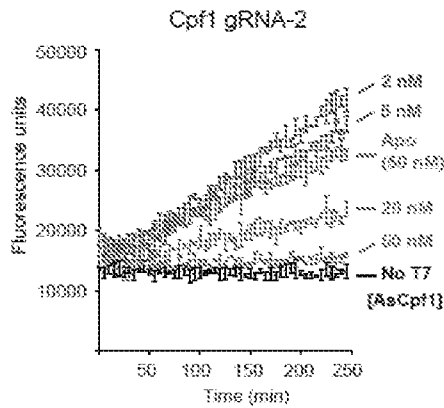
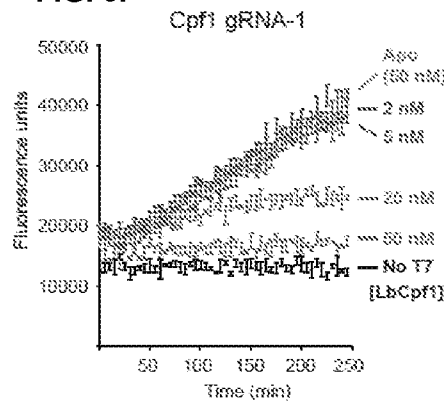


FIG. 3J

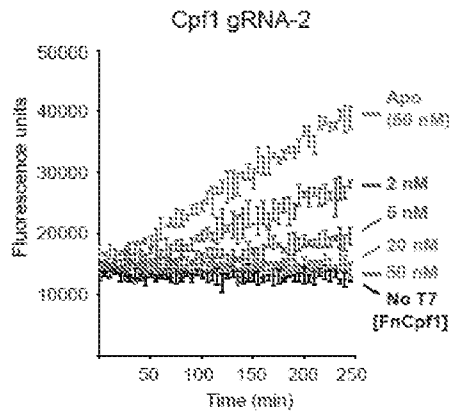
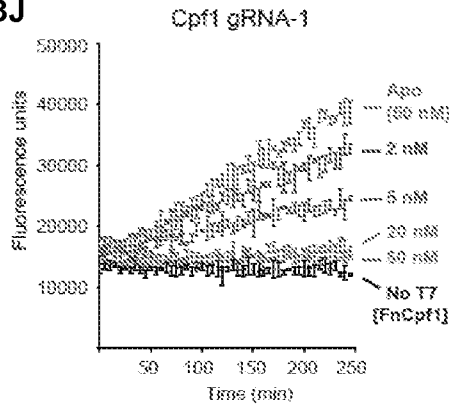


FIG. 3K



FIG. 3L

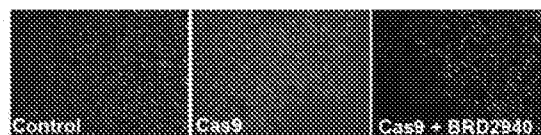


FIG. 3M

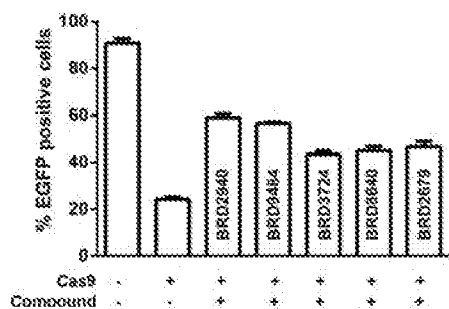


FIG. 3N

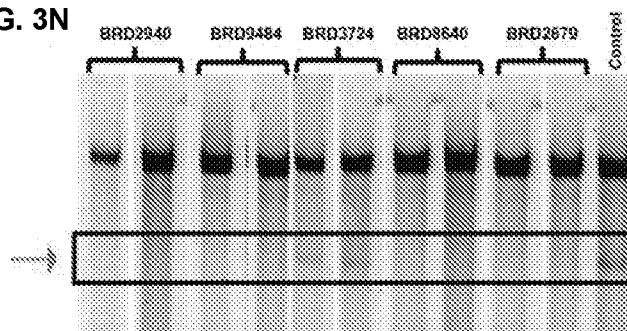


FIG. 3O

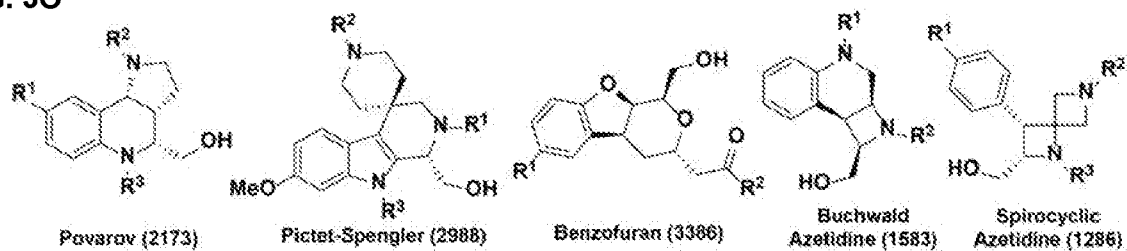


FIG. 3P

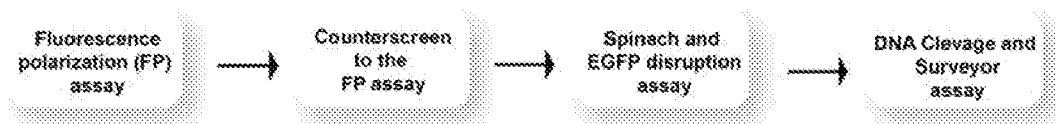


FIG. 3Q

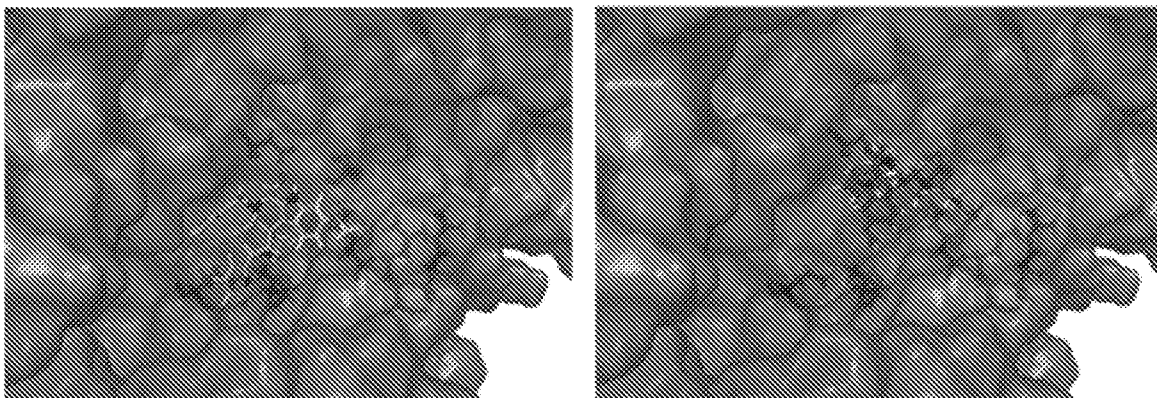
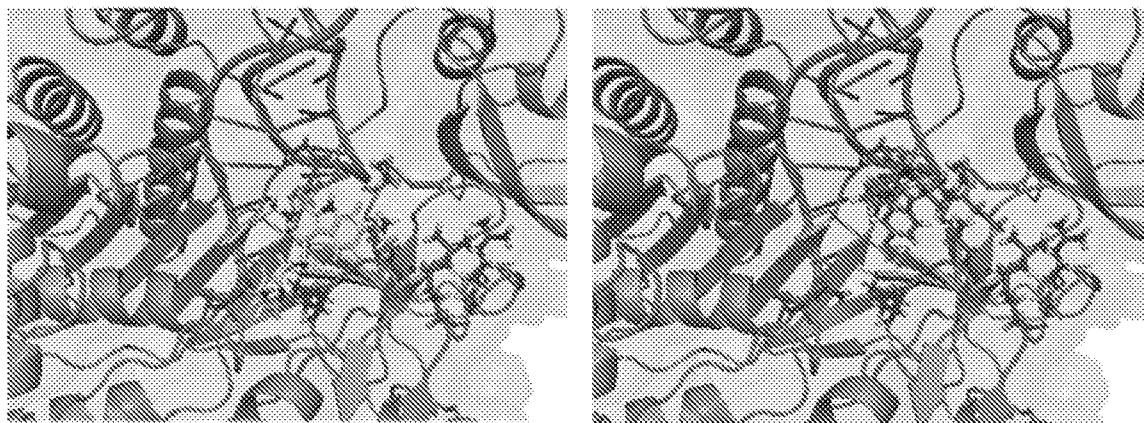


FIG. 3R



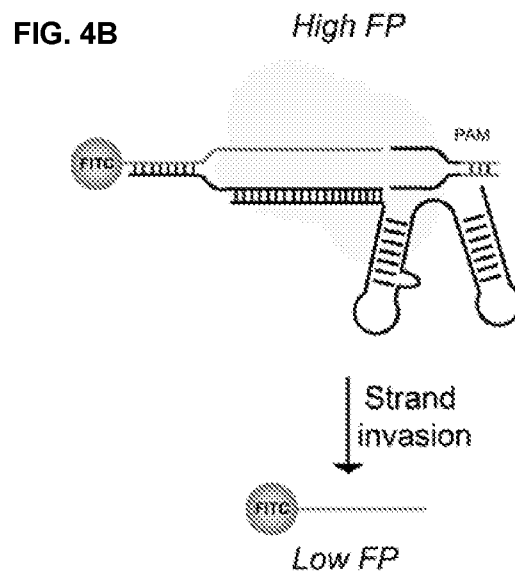
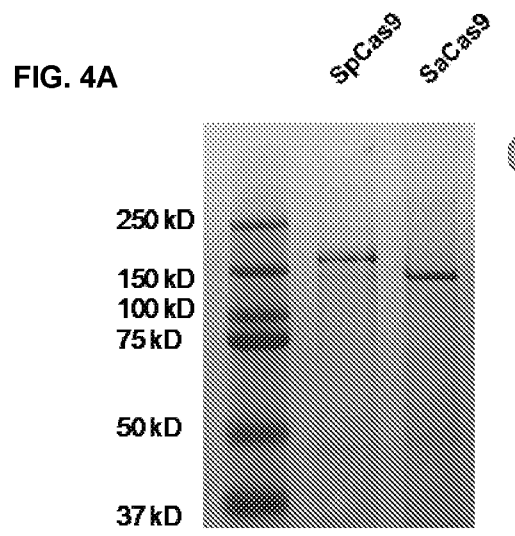


FIG. 4C

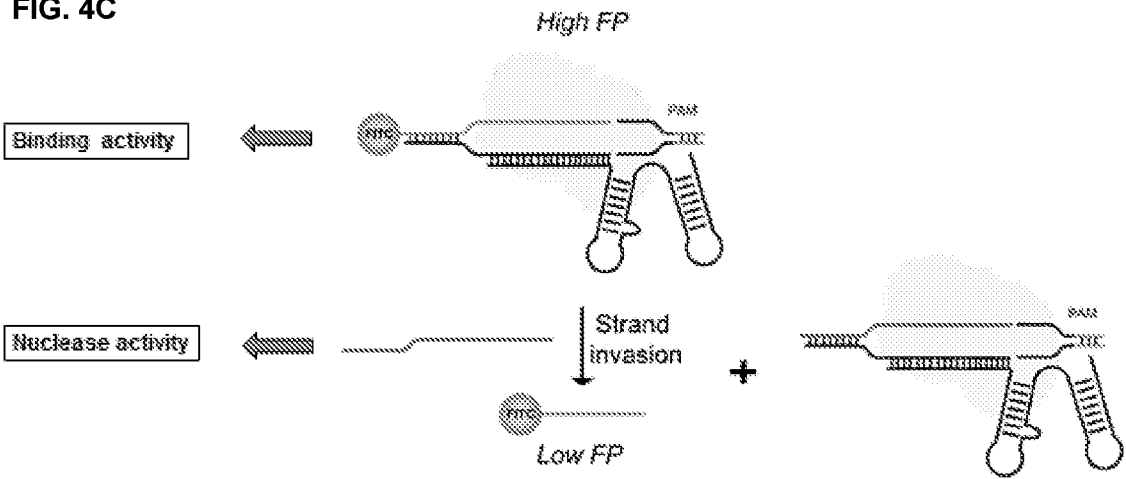


FIG. 4D

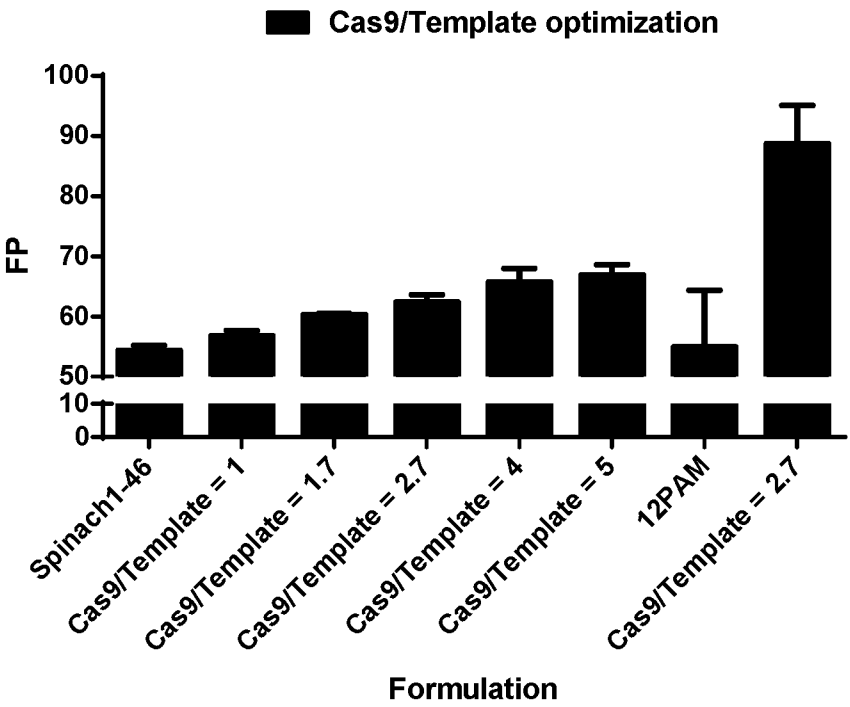


FIG. 4E

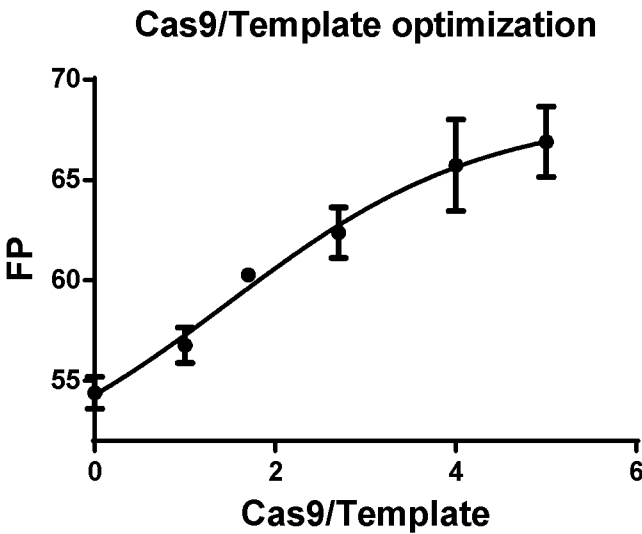


FIG. 4F

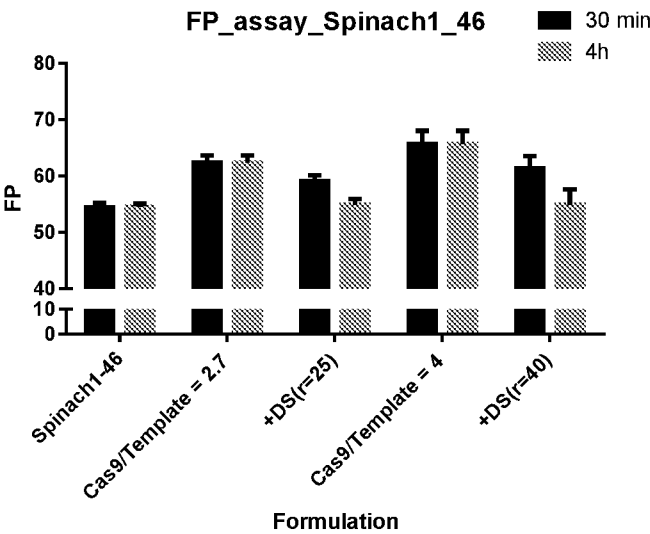


FIG. 4G

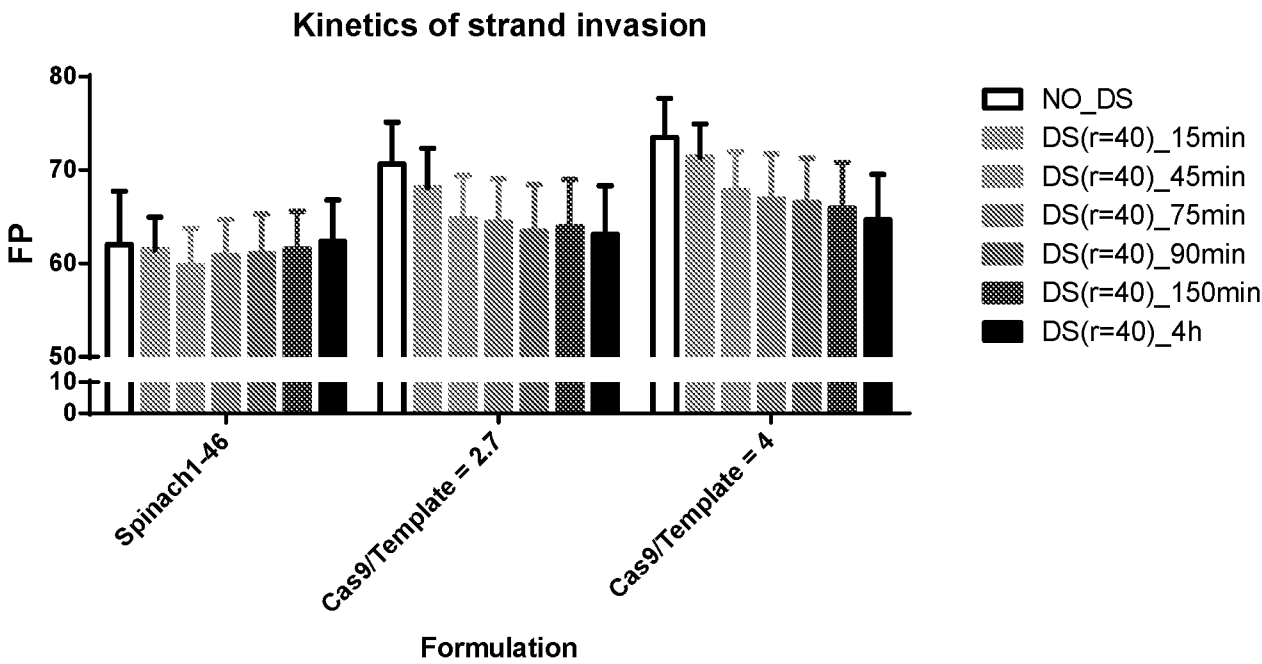


FIG. 5A

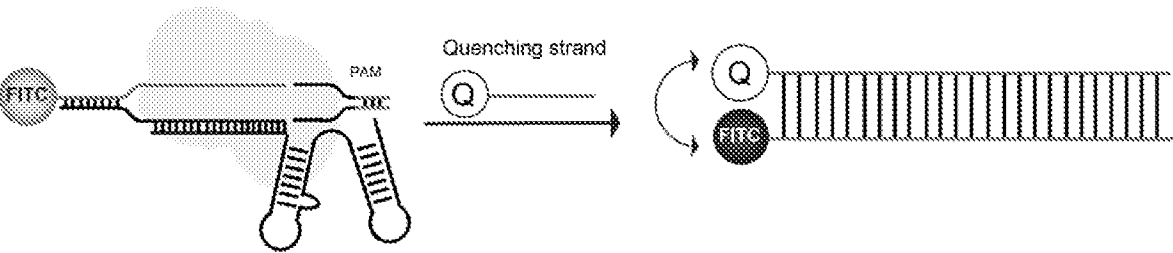


FIG. 5B

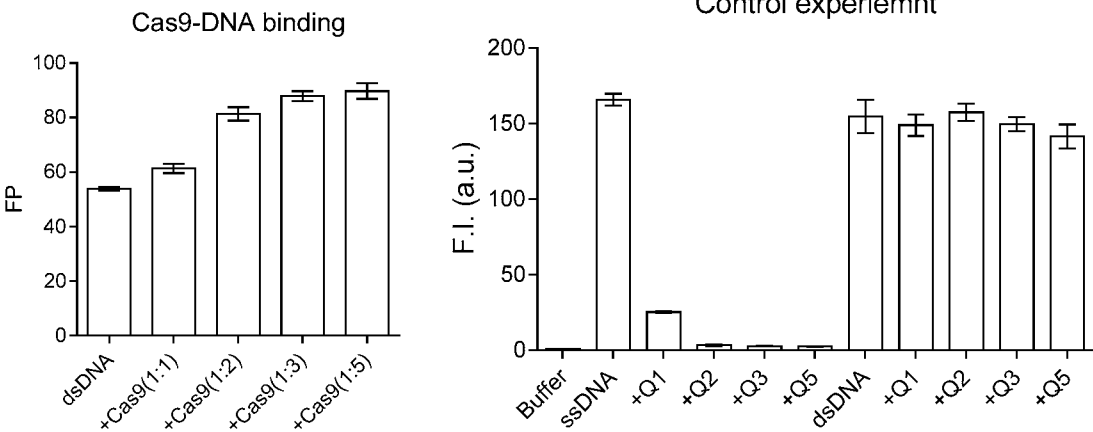


FIG. 5D

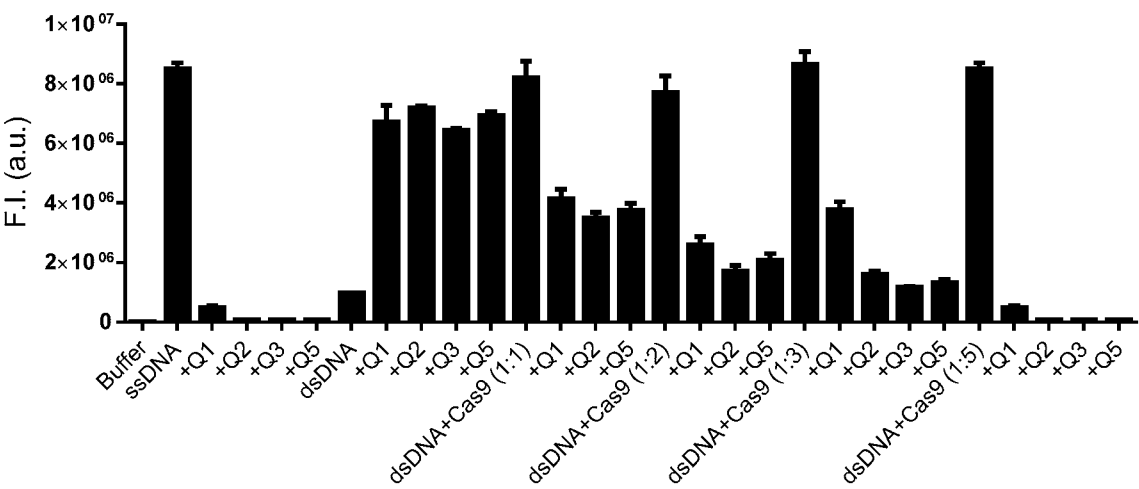


FIG. 5E

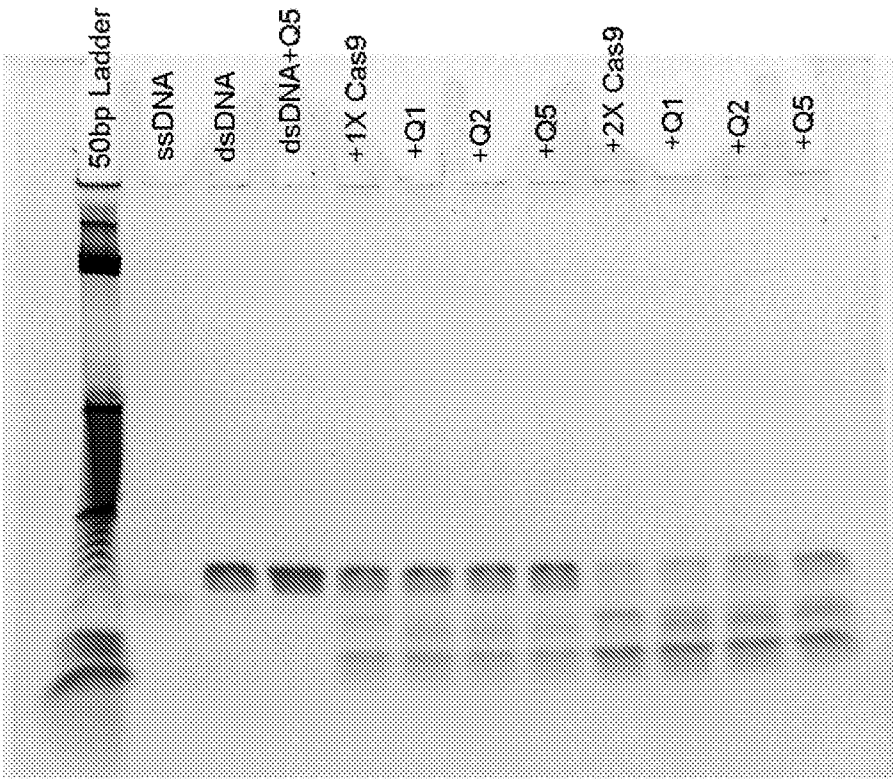
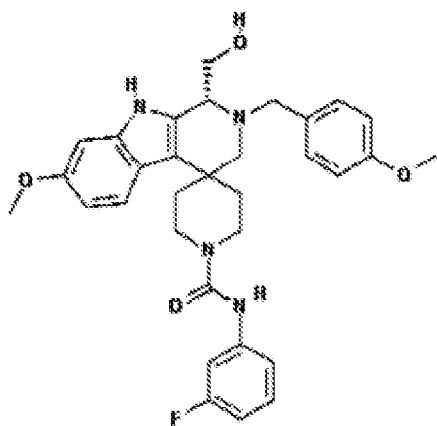
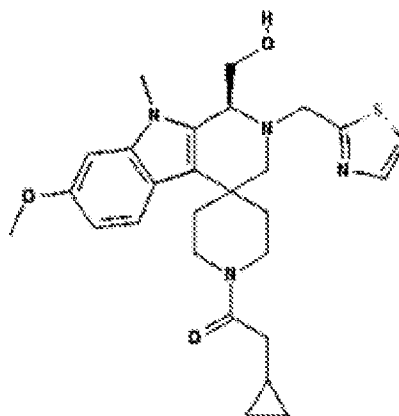


FIG. 6A

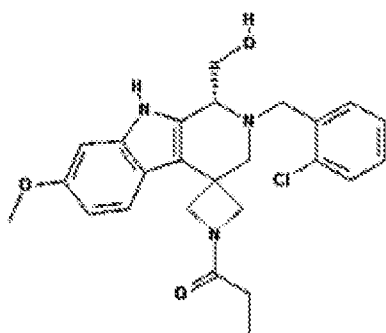
Chemical structure of hit compounds



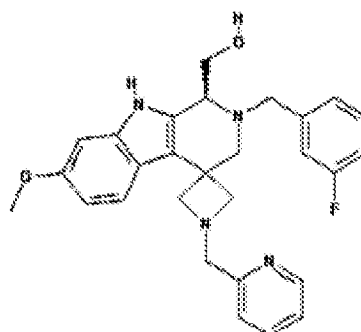
BRD-K33182911-001-01-8
(BRD2911)



BRD-K36151368-001-01-8
(BRD1368)



BRD-K86027682-001-01-8
(BRD7682)



BRD-K96745813-001-01-5
(BRD5813)

FIG. 6A continued

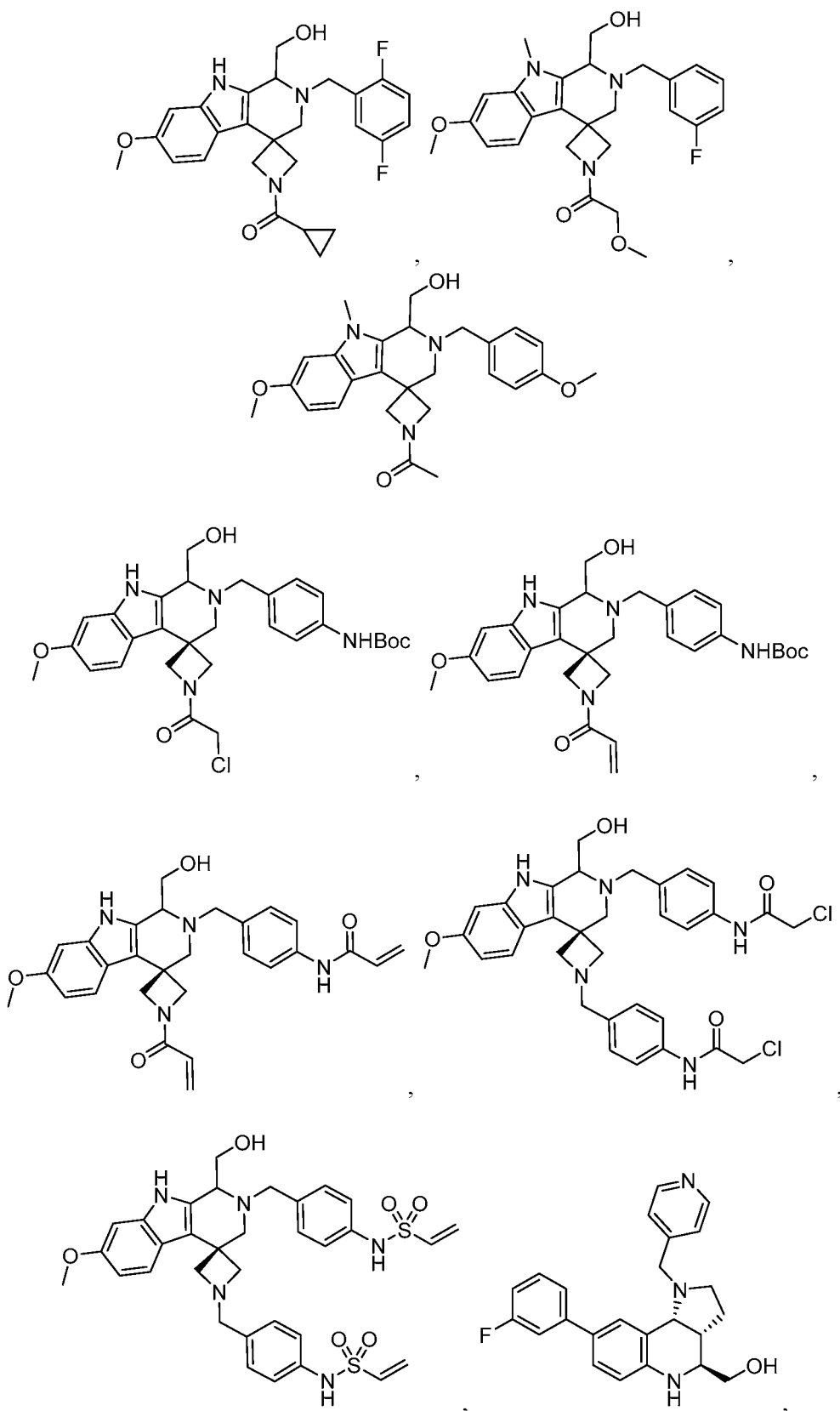


FIG. 6A continued

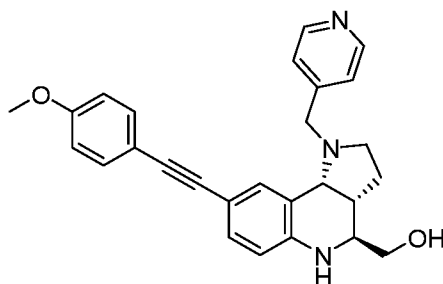


FIG. 6B

Compound mediated inhibition of Cas9 activity in egfp-U2OS cells (48h treatment)

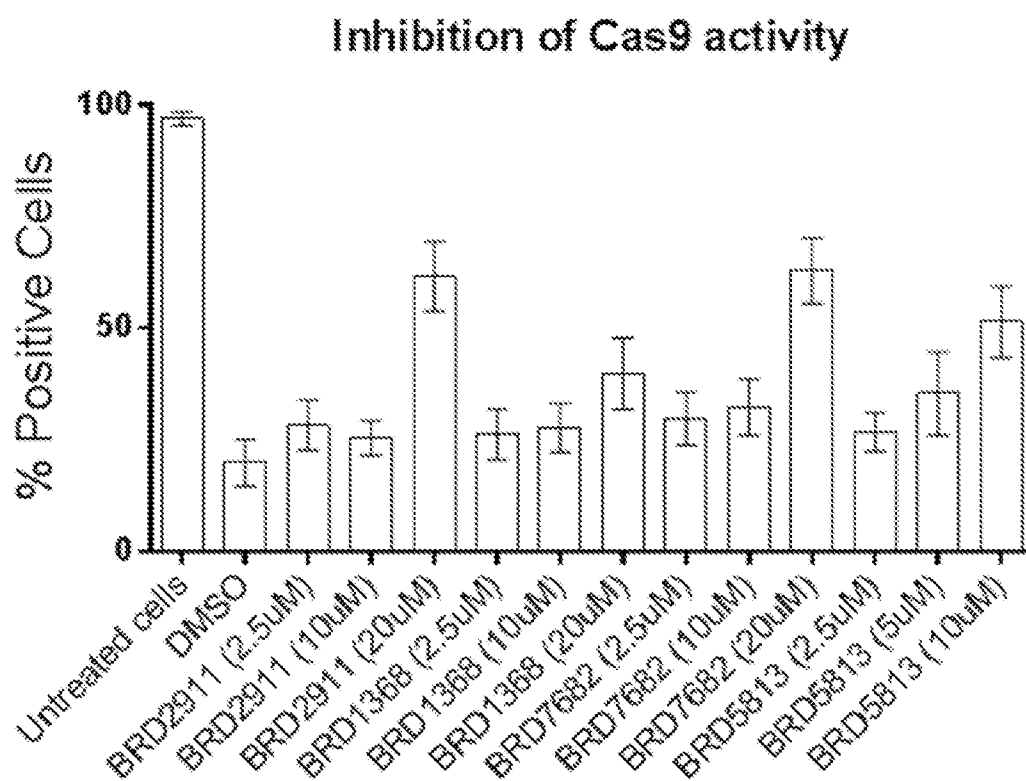
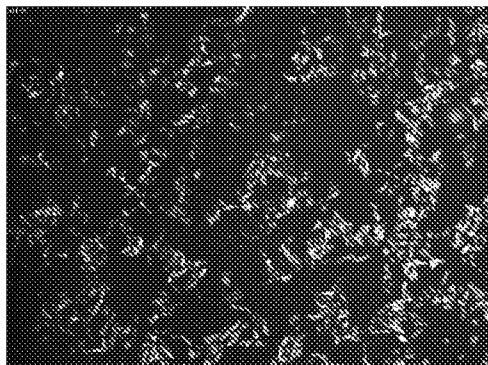


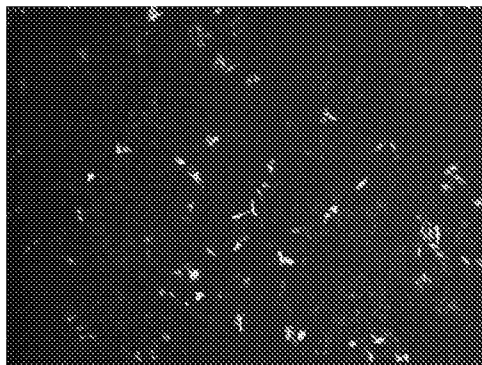
FIG. 6C

Cell images for compound mediated inhibition of Cas9
activity in egfp-U2OS cells (48h treatment)

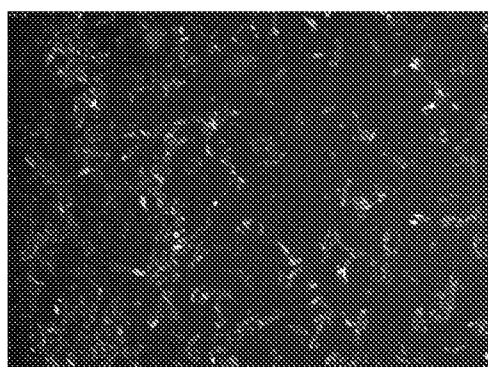
Untreated cells



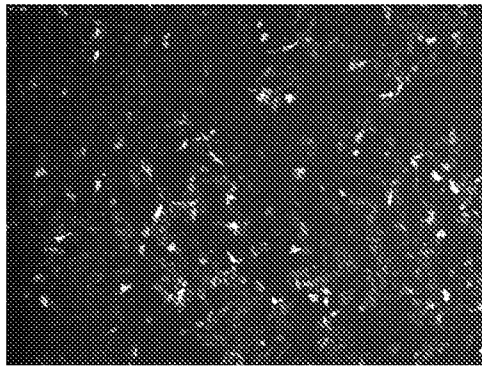
DMSO control



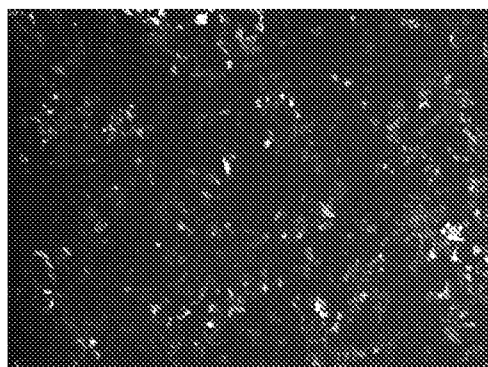
BRD2911



BRD1368



BRD7682



BRD5813

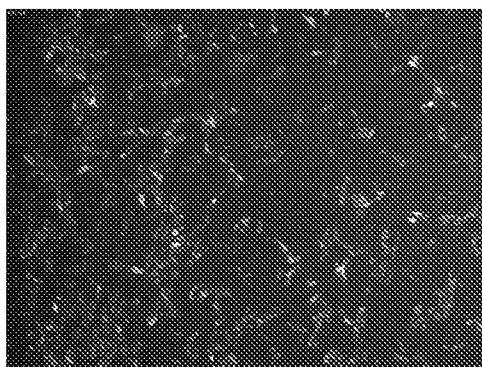


FIG. 6D

Compound mediated inhibition of Cas9 activity in HEK293T cells (48h treatment)
as measured by Surveyor assay

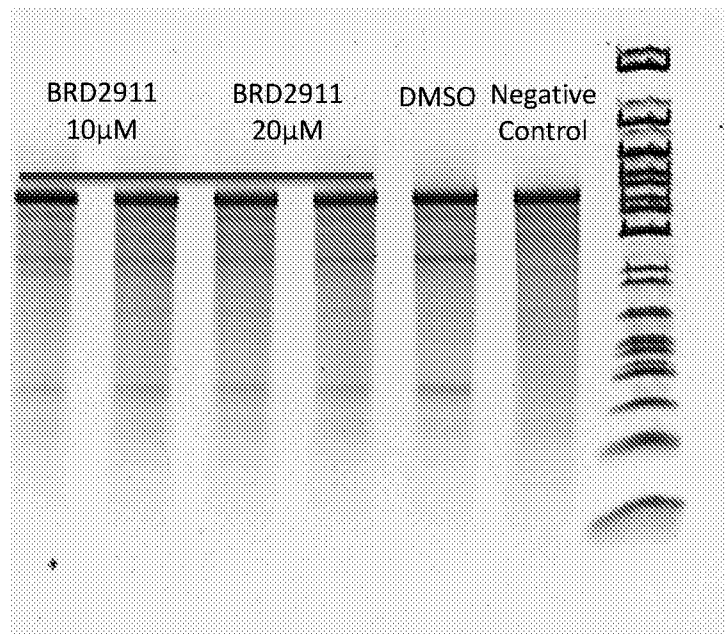
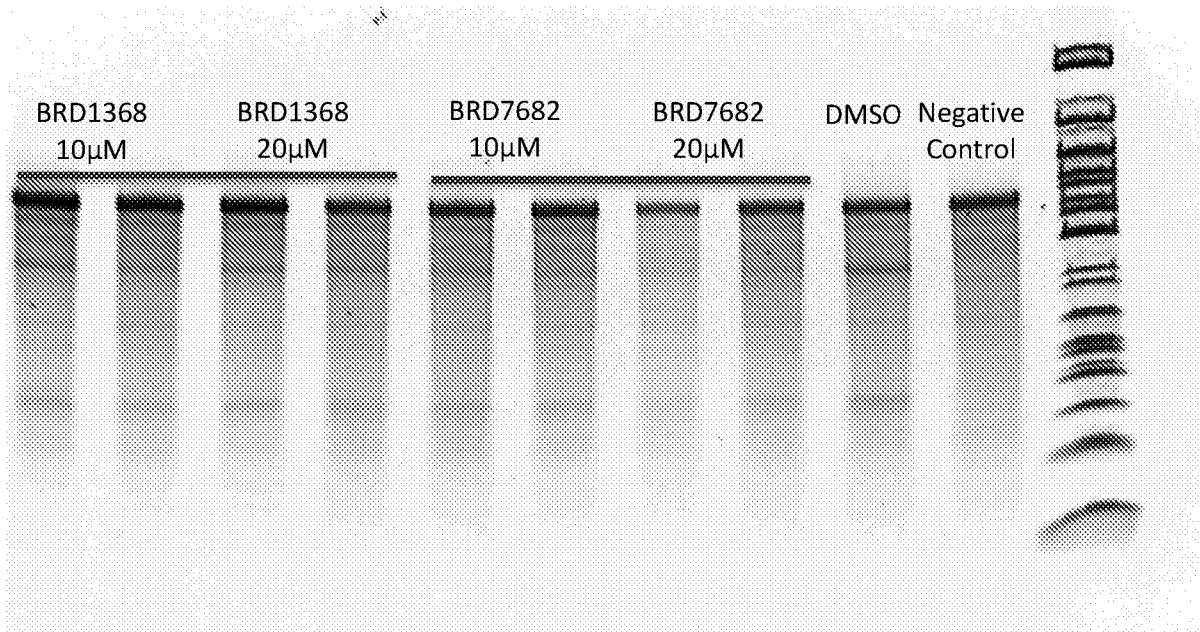


FIG. 7A

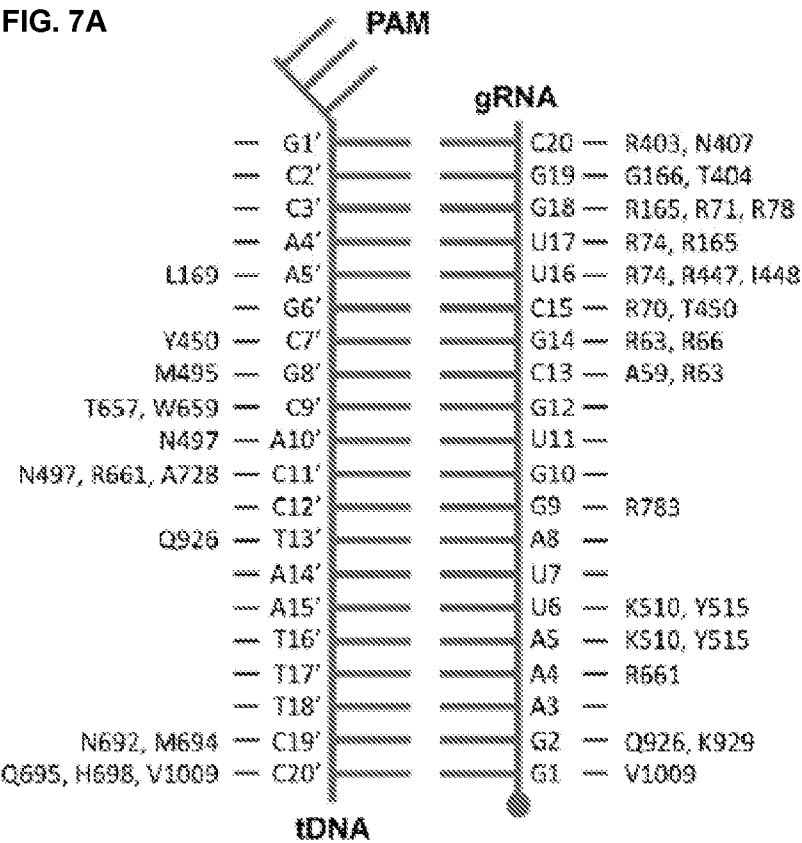


FIG. 7B

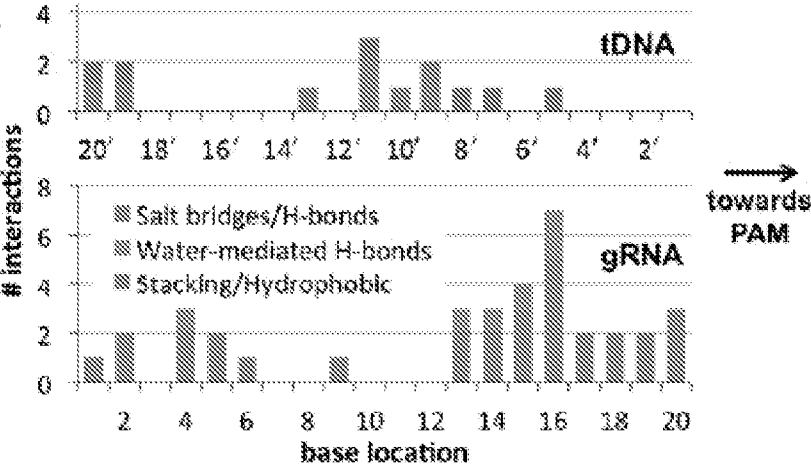
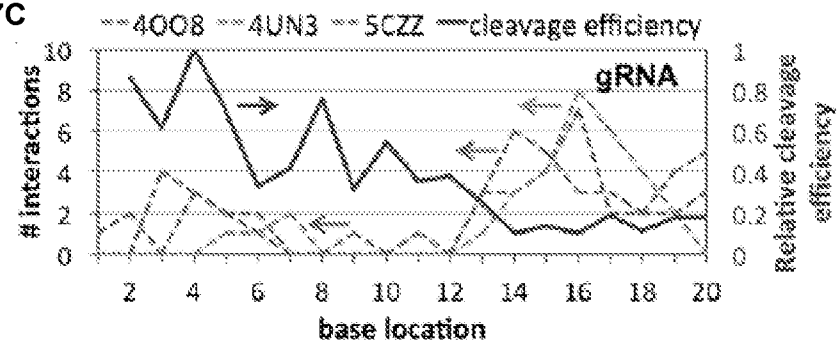


FIG. 7C



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US1 7/59365

A. CLASSIFICATION OF SUBJECT MATTER
IPC - C07D 471/04; C12N 9/22 (2018.01)
CPC - C07D 471/04; C12N 9/22

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2016/022363 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 11 February 2016; abstract; paragraphs [00100], [00103]	1 ----- 2-3, 11/1-3
X -- Y	US 2009/0275568 A1 (KAJINO, M et al.) 05 November 2009; paragraphs [0010], [0023], [0094], [0120]	55 ----- 2-3, 11/1-3

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

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

Date of the actual completion of the international search

08 January 2018 (08.01.2018)

Date of mailing of the international search report

12 MAR '2018

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/59365

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 12-17, 23-32, 38-49, 53-54
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

---Continued Within the Next Supplemental Box---

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-3, 11/1-3, 55

Remark on Protest

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

---Continued from Box No. III Observations where unity of invention is lacking ---

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups I+, Claims 1-11 (in-part) and 55-56 (in-part) disclose a method of inhibiting the activity of an RNA guided endonuclease-guide RNA complex, the method comprising contacting the RNA guided endonuclease-guide RNA complex with a small molecule having the formula of Formula I: as shown, wherein R1 is hydrogen; R2 is R wherein R is C1-12 hydrocarbon comprising 1-10 N atoms; and R3-R4 are hydrogen (first exemplary compound structure).

The compound, method and composition will be searched to the extent they encompass a compound of Formula I, as shown, wherein R1 is hydrogen; R2 is R wherein R is C1-12 hydrocarbon comprising 1-10 N atoms; and R3-R4 are hydrogen (first exemplary compound structure). Applicant is invited to elect additional compound(s), with specified structural formula(ae) for each, to be searched. Additional compound(s) will be searched upon the payment of additional fees. It is believed that claims 1 (in-part), 2 (in-part), 3 (in-part), 11 (in-part), and 55 (in-part) encompass this first named invention and thus these claims will be searched without fee to the extent that they encompass a compound of structure I wherein R1 is hydrogen; R2 is R wherein R is C1-12 hydrocarbon comprising 1-10 N atoms; and R3-R4 are hydrogen (first exemplary compound structure). Applicants must specify the claims that encompass any additionally elected compound(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be a compound of structure I wherein R1 is hydrogen; R2 is hydrogen; and R3-R4 are hydrogen (first exemplary elected compound structure).

Groups I+ share the technical features including a method of inhibiting the activity of an RNA guided endonuclease-guide RNA complex, the method comprising contacting the RNA guided endonuclease-guide RNA complex with a small molecule having the formula of Formula I: as shown, wherein R1 is hydrogen; R2 is R wherein R is C1-12 hydrocarbon comprising 1-10 N atoms; and R3-R4 are hydrogen.

However, these shared technical features are previously disclosed by WO 2016/022636 A2 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) (hereinafter 'Harvard') in view of US 2009/0275568 A1 to Kajino, et al. (hereinafter 'Kajino').

Harvard discloses a method of inhibiting the activity of an RNA guided endonuclease-guide RNA complex, the method comprising contacting the RNA guided endonuclease-guide RNA complex with a small molecule (method of inhibiting the gRNA (guided RNA) binding activity of RNA endonucleases, such as Cas9 (RNA complex), wherein the method comprises attaching (contacting) a compound (small molecule) to Cas9; abstract), but Harvard does not disclose wherein the small molecule has the formula of Formula I: as shown, wherein R1 is hydrogen; R2 is R wherein R is C1-12 hydrocarbon comprising 1-10 N atoms; and R3-R4 are hydrogen.

However, Kajino discloses wherein a small molecule has the formula of Formula I: as shown, wherein R1 is hydrogen; R2 is R wherein R is C1-12 hydrocarbon comprising 1-10 N atoms; and R3-R4 are hydrogen (compound of formula I wherein the dashed lines are absent, n is 2, R4-R5 are hydrogen, R7-R10 are hydrogen (R1 is hydrogen), R6 is a heterocyclic group (R2 is R wherein R is C1-C12 hydrocarbon comprising 1 or more nitrogen atoms), R1 is hydrogen (R3 is hydrogen, and R2 is methyl (R4 is hydrogen); paragraphs [0010], [0023]). It would have been obvious to a person of ordinary skill in the art, at the time of the invention, to have modified the method, as previously disclosed by Harvard, in order to have provided wherein a small molecule has the formula of Formula I: as shown, wherein R1 is hydrogen; R2 is R wherein R is C1-12 hydrocarbon comprising 1-10 N atoms; and R3-R4 are hydrogen, as previously disclosed by Kajino, for providing methods of treating gastrointestinal disorders, such as inflammatory bowel disease (Harvard; paragraph [00100]; Kajino; paragraph [0120]).

Since none of the special technical features of the Groups I+ inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Harvard and Kajino references, unity of invention is lacking.

Group II, Claims 18-22, 33-37 and 50-52 are directed toward a method of detecting nucleic acid binding of an RNA guided endonuclease-guide RNA complex comprising contacting the RNA guided endonuclease-guide RNA complex with a probe comprising an oligonucleotide comprising at least 4-12 PAM sites and a fluorescent reporter at the 5' terminus.

Group III, Claims 57-58 are directed toward a compound having the structure of Formula (II): as shown.

Group IV, Claim 59 is directed toward a compound having the structure of Formula (III): as shown.

Group V, Claim 60 is directed toward a compound having the structure: as shown.

Group VI, Claims 61-62 are directed toward a compound having the structure of Formula (V): as shown.

The inventions listed as Groups I+ and II-VI do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Groups I+ include a method of inhibiting the activity of an RNA guided endonuclease-guide RNA complex, the method comprising contacting the RNA guided endonuclease-guide RNA complex with a small molecule having the formula of Formula I: as shown, wherein R1 is hydrogen; R2 is R wherein R is C1-12 hydrocarbon comprising 1-10 N atoms; and R3-R4 are hydrogen, which is not present in Groups II-VI; the special technical features of Group II include a method of detecting nucleic acid binding of an RNA guided endonuclease-guide RNA complex comprising contacting the RNA guided endonuclease-guide RNA complex with a probe comprising an oligonucleotide comprising at least 4-12 PAM sites and a fluorescent reporter at the 5' terminus, which is not present in Groups I+ and III-VI; the special technical features of Group III include a compound having the structure of Formula (II): as shown, wherein the compound comprises a spiro heterocyclic ring, which is not present in Groups I+, II and IV-VI; and the special technical features of Group IV include a compound having the structure of Formula (III): as shown, wherein the compound comprises a tri-fused ring system comprising tRtrahydrofuran and tetrahydropyran, which is not present in Groups I+, II-III and V-VI; the special technical features of Group V include a compound having the structure: as shown, wherein the compound comprises a tri-fused ring system comprising piperidine and azetidine, which is not present in Groups I+, II-IV and VI; and special technical features of Group VI include a compound having the structure of Formula (V): as shown, wherein the compound comprises two spiro azetidine rings, which is not present in Groups I+ and II-V.

The common technical features of Groups I+ and III-IV are a compound comprising a multi-cyclic ring structure comprising a phenyl ring and one or more heterocyclic rings.

These common technical features are disclosed by US 2007/0161619 A1 to Hu, et al. (hereinafter 'Hu').

Hu discloses a compound comprising a multi-cyclic ring structure comprising a phenyl ring and one or more heterocyclic rings (compound of formula I; paragraph [0055]).

Since the common technical features are previously disclosed by Hu, these common features are not special and so Groups I+ and III-VI lack unity.

No technical features are shared between Groups I+ and II, accordingly, these groups lack unity a priori.

No technical features are shared between Groups II and III-VI, accordingly, these groups lack unity a priori.