EVALUATION OF CAS9 MOLECULE/GUIDE RNA MOLECULE COMPLEXES

RELATED APPLICATIONS

The present application claims priority to U.S. Provisional Application No. 62/152,473 filed April 24, 2015.

FIELD OF THE INVENTION

The present invention(s) relates to the evaluation, selection, and design of Cas9 molecule/guide RNA (gRNA) molecule complexes.

BACKGROUND

Direct delivery of Cas9 ribonucleoprotein (RNP) complexes allows for efficient gene-editing while minimizing off-target activity owing to the rapid turnover of the Cas9 proteins in cells. Gene editing can be achieved in various mammalian cells by cationic lipid delivery of purified Cas9 proteins complexed with *in vitro* transcribed or chemically synthesized guide RNA (gRNA). Efficiency of gene editing mediated by RNP delivery varies by locus and depends on the length of gRNA, as well as the amount and ratio of the Cas9 protein and the gRNA delivered. Given the two-component nature of the RNP complex, precise conditions are required to obtain a complete and productive complex formation between Cas9 protein and gRNA. While the amount of protein and RNA can be quantitated by dye-binding assays, e.g., the Bradford dye binding assay or the riboquant RNA assay, these techniques do not provide a quantitation of the productive RNP complex necessary for gene editing activity.

Structural and biophysical characterization of Cas9/gRNA complexes revealed a large contact area and a high affinity. Thermal melt curves are a useful property to characterize the stability and binding of protein-ligand complexes. Differential Scanning Fluorimetry (DSF) is a biophysical technique where the change in fluorescence of a small molecule dye, e.g., SYPRO® orange, is used to monitor the thermal denaturation of a protein and to determine its thermal melting temperature (). Binding of ligands to the protein tend to stabilize the protein to differing extents and change its T_m. Measurement of the T_m of a protein at different ligand concentrations can allow the measurement of the affinity of protein for that ligand. A wide range of the thermal melting signature can be used to rapidly assay the quality of RNP complexes at a high throughput with a high signal to noise ratio.

Thus, there remains a need for developing assays, such as DSF, that can be used to evaluate the quality of Cas9 molecule/gRNA molecule complexes, e.g., to quantitate productive complex formation, a prerequisite for RNP mediated gene editing.

SUMMARY OF THE INVENTION

Methods are provided herein for screening for a Cas9 molecule/gRNA molecule complex for administration to a subject that includes (a) generating a plurality of samples, each sample comprising a Cas9 molecule/gRNA molecule complex generated by combining a Cas9 molecule and one of a plurality of gRNA molecules; (b) detecting a melting temperature (T_m) value of the Cas9 molecule/gRNA molecule complex in each of the plurality of samples; and (c) selecting at least one sample from the plurality of samples based on one or more of (i) a comparison of the T_m values in the plurality of samples to a T_m value of a reference Cas9 molecule/gRNA molecule complex or a pre-determined threshold T_m value, or (ii) a relative ordering of the T_m values of the plurality of samples. In certain embodiments, the step of detecting the T_m value of the Cas9 molecule/gRNA molecule complex in each sample of the plurality of samples may include assessing each sample in the plurality of samples by differential scanning fluorimetry (DSF). In certain embodiments, the gRNA may be a chimeric gRNA. In certain embodiments, the gRNA may be a modular gRNA. In certain embodiments, the sample may comprise a component comprising an additive, a small molecule, a stabilizing reagent, buffer, pH, salt concentration, glycerol concentration, or other buffer component. In certain embodiments, a sample comprising a Cas9 molecule/gRNA molecule complex having a T_m of at least 8° C greater than a T_m value of the Cas9 molecule absent the gRNA molecule may be selected.

Provided herein in certain embodiments are isolated complexes of a Cas9 molecule and a gRNA molecule having a T_m at least 8° C greater than a T_m value of the Cas9 molecule absent the gRNA molecule selected according to the methods provided herein.

Provided herein in certain embodiments are compositions including an isolated complex of a Cas9 molecule and a gRNA molecule having a T_m at least 8°C greater than a T_m of the Cas9 molecule absent the gRNA molecule selected according to the methods provided herein. In certain embodiments, the difference of the T_m of the non-naturally occurring complex of a Cas9 molecule and a gRNA molecule and the T_m of the Cas9 molecule absent the gRNA molecule may be assessed by DSF. In certain embodiments, the gRNA may be a chimeric or modular gRNA.

Provided herein in certain embodiments are methods of determining the stability of a Cas9 molecule/gRNA molecule complex including (a) generating a plurality of Cas9 molecule/gRNA molecule complexes, each comprising a Cas9 molecule/gRNA molecule complex generated by combining a Cas9 molecule and one of a plurality of gRNA molecules; (b) detecting a T_m value of each of the Cas9 molecule/gRNA molecule complexes of the plurality of Cas9 molecule/gRNA molecule complexes; and (c) determining one or more of the plurality of Cas9 molecule/gRNA molecule complexes is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value.

In certain embodiments of the methods herein, the plurality of gRNA molecules may be a library of candidate gRNA molecules. In certain embodiments, the library of candidate gRNA molecules may comprise a library of tracrRNA molecules or sequences. In certain embodiments, the library of tracrRNA molecules or sequences may be of differing length.

Provided herein in certain embodiments are methods of determining a condition that promotes a stable Cas9 molecule/gRNA molecule complex including (a) combining a Cas9 molecule and a gRNA molecule in a sample to form a Cas9 molecule/gRNA molecule complex; (b) detecting a T_m value of the Cas9 molecule/gRNA molecule complex; and (c) determining the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value.

Provided herein in certain embodiments are methods of screening for a stable Cas9 molecule/gRNA molecule complex including (a) detecting a T_m value of a Cas9 molecule/gRNA molecule complex via DSF; and (b) determining the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value.

Provided herein in certain embodiments are methods for identifying an optimal gRNA to form a stable Cas9 molecule/gRNA molecule complex including (a) combining a Cas9 molecule and a gRNA molecule in a sample to form the Cas9 molecule/gRNA molecule complex; (b) detecting a T_m value of the Cas9 molecule/gRNA molecule complex; and (c) determining the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value by at least 8°C.

Provided herein in certain embodiments are methods of determining the stability of a Cas9 molecule/gRNA molecule complex including (a) combining a Cas9 molecule and a gRNA molecule in a sample to form the Cas9 molecule/gRNA molecule complex; (b) detecting a T_m value of the Cas9 molecule/gRNA molecule complex; (c) measuring an activity value of the Cas9 molecule/gRNA molecule complex; and (d) determining the Cas9 molecule/gRNA molecule complex is stable if (i) the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value and (ii) the activity value of the Cas9 molecule/gRNA molecule complex is greater than an activity value of a reference molecule or an activity reference value.

Provided herein in certain embodiments are methods of optimizing binding of a gRNA molecule to a Cas9 molecule to form a stable Cas9 molecule/gRNA molecule complex including (a) combining the Cas9 molecule and the gRNA molecule in a sample to form a Cas9 molecule/gRNA molecule complex; (b) detecting a T_m value of the Cas9 molecule/gRNA molecule complex; (c) determining a delta value between the T_m value of the Cas9 molecule/gRNA molecule complex and a T_m value of a reference molecule or a T_m reference value; and (d) determining the Cas9 molecule/gRNA molecule complex is stable if the delta value is at least 8°C and the T_m value of the Cas9 molecule/gRNA molecule complex is greater than the T_m value of the reference molecule or the T_m reference value.

Provided herein in certain embodiments are methods of detecting a stable Cas9 molecule/gRNA molecule complex including (a) detecting a thermostability value of a reference molecule; (b) combining a Cas9 molecule and a gRNA molecule in a sample to form a Cas9 molecule/gRNA molecule complex; (c) measuring a thermostability value of the Cas9 molecule/gRNA molecule complex; and (d) determining the Cas9 molecule/gRNA molecule complex is stable if the thermostability value of the Cas9 molecule/gRNA molecule complex is greater than the thermostability value of the reference molecule.

In certain embodiments of the methods herein, the thermostability value may be a denaturation temperature value and the thermostability reference value may be a denaturation temperature reference value. In certain embodiments of the methods herein, the thermostability value may be a T_m value and the thermostability reference value may be a T_m reference value.

In certain embodiments of the methods herein, the Cas9 molecule/gRNA molecule complex may be stable if the T_m value of the Cas9 molecule/gRNA molecule complex is at least 1°C, at least 2°C, at least 3°C, at least 4°C, at least 5°C, at least 6°C, at least 7°C, at

least 8°C, at least 9°C, at least 10°C, at least 11°C, at least 12°C, at least 13°C, at least 14°C, at least 15°C, at least 16°C, at least 17°C, at least 18°C, at least 19°C, or at least 20°C greater than the T_m value of the reference molecule or T_m reference value. For example, in certain embodiments, the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is at least 8°C greater than the T_m value of the reference molecule or T_m reference value.

In certain embodiments of the methods herein, the Cas9 molecule/gRNA molecule complex may be stable if the T_m value of the Cas9 molecule/gRNA molecule complex is about 1°C, about 2°C, about 3°C, about 4°C, about 5°C, about 6°C, about 7°C, about 8°C, about 9°C, about 10°C, about 11°C, about 12°C, about 13°C, about 14°C, about 15°C, about 16°C, about 17°C, about 18°C, about 19°C, or about 20°C greater than the T_m value of the reference molecule or T_m reference value. For example, in certain embodiments, the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is about 8°C greater than the T_m value of the reference molecule or T_m reference value.

In certain embodiments of the methods herein, the Cas9 molecule/gRNA molecule complex may be stable if the T_m value of the Cas9 molecule/gRNA molecule complex is about 1°C to 5°C, about 6°C to 10°C, about 11°C to 15°C, or about 16°C to 20°C greater than the T_m value of the reference molecule or T_m reference value. For example, in certain embodiments, the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is about 6°C to 10°C greater than the T_m value of the reference molecule or T_m reference value.

In certain embodiments of the methods herein, the T_m value may be detected using a thermal shift assay. In certain embodiments, the thermal shift assay may be selected from DSF, differential scanning calorimetry (DSC), or isothermal titration calorimetry (ITC).

In certain embodiments of the methods herein, the gRNA molecule may comprise a chimeric gRNA molecule. In certain embodiments, the gRNA molecule may comprise a modular gRNA molecule.

In certain embodiments of the methods herein, the Cas9 molecule may be any of the Cas9 molecules disclosed herein. For example, the Cas9 molecule may be a Cas9 molecule selected from Table 1. In certain embodiments, the Cas9 molecule may be a chimeric Cas9 molecule, or a synthetic or engineered Cas9 molecule. For example, the Cas9 molecule may

be a Cas9 molecule with a portion or portions deleted. In certain embodiments, the Cas9 molecule may comprise a *S. pyogenes* or a *S. aureus* Cas9 molecule.

In certain embodiments of the methods herein, the reference molecule may be selected from (a) a reference Cas9 molecule in the absence of a gRNA molecule; (b) a reference Cas9 molecule (e.g., the same Cas9 molecule as the Cas9 molecule in the complex being evaluated) complexed with a second gRNA molecule (i.e., a gRNA other than the one in the complex being evaluated); and (c) a reference Cas9 molecule/gRNA molecule complex, wherein the reference Cas9 molecule/gRNA molecule was formed under different conditions, e.g., with different proportions of Cas9 molecule and gRNA molecule, than the Cas9 molecule/gRNA molecule complex, or was formed in a different buffer. In certain embodiments, the reference Cas9 molecule may be the same as the Cas9 molecule of the complex being evaluated. In certain embodiments, the reference Cas9 molecule may be different from the Cas9 molecule of the complex being evaluated. In certain embodiments, the reference Cas9 molecule may differ in primary sequence from the Cas9 molecule of the complex being evaluated. In certain embodiments, the gRNA molecule of the reference Cas9 molecule/gRNA molecule complex may be the same as the gRNA molecule of the complex being evaluated. In certain embodiments, the gRNA molecule of the reference Cas9 molecule/gRNA molecule complex may be different from the gRNA molecule of the complex being evaluated. In certain embodiments, the gRNA molecule of the reference Cas9 molecule/gRNA molecule complex may differ in sequence or differs by a modification from the gRNA molecule of the complex being evaluated.

In certain embodiments of the methods herein, the T_m reference value may comprise a preselected numerical value for T_m. In certain embodiments, the T_m reference value may comprise a value correlated with the T_m of any of the reference molecules described herein.

In certain embodiments, the methods disclosed herein may further include detecting an activity of the Cas9 molecule/gRNA molecule complex; measuring an activity value of the Cas9 molecule/gRNA molecule complex; and determining the Cas9 molecule/gRNA molecule complex is stable if the activity value of the Cas9 molecule/gRNA molecule complex is greater than the activity value of a reference molecule or an activity reference value. In certain embodiments, the activity may comprise one or more of: an ability to induce indels; an ability to modify a target DNA; a propensity of a preselected repair method; an ability of the gRNA molecule to remain hybridized to the DNA target; and an ability of the gRNA molecule to bind to the Cas9 molecule of the Cas9 molecule/gRNA molecule

complex. In certain embodiments, the activity value may be a binding value and the activity may be the ability of the gRNA molecule to bind to the Cas9 molecule comprising: (a) combining the gRNA molecule and the Cas9 molecule in a sample to form the Cas9 molecule/gRNA molecule complex; (b) measuring a binding value of the Cas9 molecule/gRNA molecule complex; and (c) determining the Cas9 molecule/gRNA molecule complex is stable if the binding value of the Cas9 molecule/gRNA molecule complex is greater than the binding value of a reference molecule or the binding reference value. In certain embodiments, the binding value may be measured using a kinetics assay. In certain embodiments, the kinetics assay may be selected from surface plasmon resonance (SPR) assay, Bio-Layer Interferometry (BLI) assay, or gel band shift assay. In certain embodiments, the propensity of a preselected repair method may be HDR or NHEJ. In certain embodiments, the activity of the Cas9 molecule/gRNA molecule complex may be tested using an *in vitro* system; an ex vivo system; an *in vivo* system; a cellular assay; or an animal model.

In certain embodiments, the reference molecule is selected from any of the reference molecules provided herein. In certain embodiments, the reference Cas9 molecule/gRNA molecule complex is formed with different proportions of Cas9 molecule and gRNA molecule than the Cas9 molecule/gRNA molecule complex or is formed in a different buffer than the Cas9 molecule/gRNA molecule complex.

Provided herein in certain embodiments are synthetic Cas9 molecule/gRNA molecule complexes generated using any of the methods described herein.

Provided herein in certain embodiments are compositions comprising Cas9 molecule/gRNA molecule complexes generated using any of the methods described herein.

Provided herein in certain embodiments are vector systems comprising a nucleic acid encoding one or more Cas9 molecule/gRNA molecule complexes generated using the any of the methods described herein.

Provided herein in certain embodiments are methods of delivering a Cas9 molecule/gRNA molecule complex to a target cell comprising delivering any of the Cas9 molecule/gRNA molecule complexes described herein to the target cell. In certain embodiments, the Cas9 molecule/gRNA molecule complexes may be delivered to the cell by RNP cationic lipid transfection, a viral vector, or RNA transfection. In certain embodiments, the viral vector may be an AAV vector.

In certain embodiments, the gRNA molecule in a method, composition, or formulation provided herein may be a unimolecular or chimeric gRNA. In other embodiments, the gRNA molecule may be a modular gRNA.

In certain embodiments, the Cas9 molecule in a method, composition, or formulation provided herein may be an *S. pyogenes*, *S. aureus*, or *S. thermophilus* Cas9 molecule, or a Cas9 molecule derived from an *S. pyogenes*, *S. aureus*, or *S. thermophilus* Cas9 molecule. In certain embodiments, the Cas9 molecule may be labeled, and in certain of these embodiments the label may be a fluorescent dye.

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

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-1I are representations of several exemplary gRNAs.

Fig. 1A depicts a modular gRNA molecule derived in part (or modeled on a sequence in part) from *Streptococcus pyogenes (S. pyogenes)* as a duplexed structure (SEQ ID NOs:39 and 40, respectively, in order of appearance);

Fig. 1B depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:41);

Fig. 1C depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:42);

Fig. 1D depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:43);

Fig. 1E depicts a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:44);

Fig. 1F depicts a modular gRNA molecule derived in part from *Streptococcus thermophilus (S. thermophilus)* as a duplexed structure (SEQ ID NOs:45 and 46, respectively, in order of appearance);

Fig. 1G depicts an alignment of modular gRNA molecules of *S. pyogenes* and *S. thermophilus* (SEQ ID NOs:39, 45, 47, and 46, respectively, in order of appearance).

Figs. 1H-1I depicts additional exemplary structures of unimolecular gRNA molecules.

Fig. 1H shows an exemplary structure of a unimolecular gRNA molecule derived in part from *S. pyogenes* as a duplexed structure (SEQ ID NO:42).

Fig. 1I shows an exemplary structure of a unimolecular gRNA molecule derived in part from *S. aureus* as a duplexed structure (SEQ ID NO:38).

Figs. 2A-2G depict an alignment of Cas9 sequences (Chylinski 2013). The N-terminal RuvC-like domain is boxed and indicated with a "Y." The other two RuvC-like domains are boxed and indicated with a "B." The HNH-like domain is boxed and indicated by a "G." Sm: *S. mutans* (SEQ ID NO:1); Sp: *S. pyogenes* (SEQ ID NO:2); St: *S. thermophilus* (SEQ ID NO:4); and Li: *L. innocua* (SEQ ID NO:5). "Motif" (SEQ ID NO:14) is a consensus sequence based on the four sequences. Residues conserved in all four sequences are indicated by single letter amino acid abbreviation; "*" indicates any amino acid found in the corresponding position of any of the four sequences; and "-" indicates absent.

Figs. 3A-3B show an alignment of the N-terminal RuvC-like domain from the Cas9 molecules disclosed in Chylinski 2013 (SEQ ID NOs:52-95, 120-123). The last line of **Fig. 3B** identifies 4 highly conserved residues.

Figs. 4A-4B show an alignment of the N-terminal RuvC-like domain from the Cas9 molecules disclosed in Chylinski 2013 with sequence outliers removed (SEQ ID NOs:52-123). The last line of **Fig. 4B** identifies 3 highly conserved residues.

Figs. 5A-5C show an alignment of the HNH-like domain from the Cas9 molecules disclosed in Chylinski 2013 (SEQ ID NOs:124-198). The last line of **Fig. 5C** identifies conserved residues.

Figs. 6A-6B show an alignment of the HNH-like domain from the Cas9 molecules disclosed in Chylinski 2013 with sequence outliers removed (SEQ ID NOs:124-141, 148, 149, 151-153, 162, 163, 166-174, 177-187, 194-198). The last line of **Fig. 6B** identifies 3 highly conserved residues.

Fig. 7 illustrates gRNA domain nomenclature using an exemplary gRNA sequence (SEQ ID NO:42).

Figs. 8A and **8B** provide schematic representations of the domain organization of *S. pyogenes* Cas9. **Fig. 8A** shows the organization of the Cas9 domains, including amino acid positions, in reference to the two lobes of Cas9 (recognition (REC) and nuclease (NUC) lobes). **Fig. 8B** shows the percent homology of each domain across 83 Cas9 orthologs.

Fig. 9A depicts the thermal stability of *S. aureus* Cas9 in the absence of gRNA as determined by DSF.

Fig. 9B depicts the thermal stability of *S. pyogenes* Cas9 in the absence of gRNA as determined by DSF.

Fig. 10 depicts the thermal stability of (1) *S. pyogenes* Cas9, (2) *S. pyogenes* Cas9 in the presence of *S. pyogenes* gRNA, and (3) *S. pyogenes* Cas9 in the presence of *S. aureus* gRNA, as determined by DSF.

Fig. 11A depicts the thermal stability of (1) *S. aureus* Cas9 and (2) *S. aureus* Cas9 with gRNA targeting CD3, as determined by DSF.

Fig. 11B depicts exemplary FACS analysis showing the generation CD3 negative population after delivery of *S. aureus* Cas9 and gRNA targeting CD3 to Jurkat T cells.

DETAILED DESCRIPTION

Definitions

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

"Alt-HDR," "alternative homology-directed repair," or "alternative HDR" as used herein refers to the process of repairing DNA damage using a homologous nucleic acid (e.g., an endogenous homologous sequence, e.g., a sister chromatid, or an exogenous nucleic acid, e.g., a template nucleic acid). Alt-HDR is distinct from canonical HDR in that the process utilizes different pathways from canonical HDR, and can be inhibited by the canonical HDR mediators, RAD51 and BRCA2. Also, alt-HDR uses a single-stranded or nicked homologous nucleic acid for repair of the break.

"Canonical HDR" or canonical homology-directed repair as used herein refers to the process of repairing DNA damage using a homologous nucleic acid (e.g., an endogenous homologous sequence, e.g., a sister chromatid, or an exogenous nucleic acid, e.g., a template nucleic acid). Canonical HDR typically acts when there has been significant resection at the double strand break, forming at least one single stranded portion of DNA. In a normal cell, HDR typically involves a series of steps such as recognition of the break, stabilization of the break, resection, stabilization of single stranded DNA, formation of a DNA crossover

intermediate, resolution of the crossover intermediate, and ligation. The process requires RAD51 and BRCA2, and the homologous nucleic acid is typically double-stranded.

Unless indicated otherwise, the term "HDR" as used herein encompasses both canonical HDR and alt-HDR.

"Non-homologous end joining" or "NHEJ" as used herein refers to ligation mediated repair and/or non-template mediated repair including canonical NHEJ (cNHEJ), alternative NHEJ (altNHEJ), microhomology-mediated end joining (MMEJ), single-strand annealing (SSA), and synthesis-dependent microhomology-mediated end joining (SD-MMEJ).

"Acquire" or "acquiring" as the terms are used herein, refer to obtaining possession of a physical entity, or a value, *e.g.*, a numerical value, by one or more or all of: "directly acquiring," "indirectly acquiring" the physical entity or value, or in the case of a value, "acquiring by calculation."

"Directly acquiring" means performing a process (*e.g.*, performing a synthetic or analytical method) to obtain the physical entity or value. "Directly acquiring a physical entity includes performing a process that includes a physical change in a physical substance, *e.g.*, a starting material. Exemplary changes include making a physical entity from two or more starting materials, shearing or fragmenting a substance, separating or purifying a substance, combining two or more separate entities into a mixture, performing a chemical reaction that includes breaking or forming a covalent or noncovalent bond. Directly acquiring a value includes performing a process that includes a physical change in a sample or another substance, *e.g.*, performing an analytical process which includes a physical change in a substance, *e.g.*, a sample, analyte, or reagent (sometimes referred to herein as "physical analysis"), performing an analytical method, *e.g.*, a method which includes one or more of the following: separating or purifying a substance, *e.g.*, an analyte, or a fragment or other derivative thereof, from another substance; combining an analyte, or fragment or other derivative thereof, with another substance, *e.g.*, a buffer, solvent, or reactant; or changing the structure of an analyte, or a fragment or other derivative thereof, *e.g.*, by breaking or forming a covalent or noncovalent bond, between a first and a second atom of the analyte; or by changing the structure of a reagent, or a fragment or other derivative thereof, *e.g.*, by breaking or forming a covalent or noncovalent bond, between a first and a second atom of the reagent.

"Indirectly acquiring" refers to receiving the physical entity or value from another party or source (*e.g.*, a third party laboratory that directly acquired the physical entity or

value). *E.g.*, a first party may acquire a value from a second party (indirectly acquiring) which said second party directly acquired or acquired by calculation.

"Acquiring by calculation" refers to acquiring a value by calculation or computation, *e.g.*, as performed on a machine, *e.g.*, a computer.

"Domain" as used herein is used to describe segments of a protein or nucleic acid. Unless otherwise indicated, a domain is not required to have any specific functional property.

Calculations of homology or sequence identity between two sequences (the terms are used interchangeably herein) are performed as follows. The sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The optimal alignment is determined as the best score using the GAP program in the GCG software package with a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frame shift gap penalty of 5. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences.

"Polypeptide" as used herein refers to a polymer of amino acids having less than 100 amino acid residues. In certain embodiments, it has less than 50, 20, or 10 amino acid residues.

A "reference molecule" as used herein refers to a molecule to which a modified or candidate molecule is compared. For example, a reference Cas9 molecule refers to a Cas9 molecule to which a modified or candidate Cas9 molecule is compared. Likewise, a reference gRNA refers to a gRNA molecule to which a modified or candidate gRNA molecule is compared. Additionally, a reference Cas9 molecule/gRNA molecule complex refers to a Cas9 molecule/gRNA molecule complex to which a Cas9 molecule/gRNA molecule complex is compared. The modified or candidate molecule may be compared to the reference molecule on the basis of sequence (e.g., the modified or candidate molecule may have X% sequence identity or homology with the reference molecule), thermostability, or activity (e.g., the modified or candidate molecule may have X% of the activity of the reference molecule). For example, where the reference molecule is a Cas9 molecule, a

modified or candidate molecule may be characterized as having no more than 10% of the nuclease activity of the reference Cas9 molecule. Examples of reference Cas9 molecules include naturally occurring unmodified Cas9 molecules, e.g., a naturally occurring Cas9 molecule from *S. pyogenes*, *S. aureus*, *S. thermophilus*, or *N. meningitidis*. In certain embodiments, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology with the modified or candidate Cas9 molecule to which it is being compared. In certain embodiments, the reference Cas9 molecule is a parental molecule having a naturally occurring or known sequence on which a mutation has been made to arrive at the modified or candidate Cas9 molecule.

"Reference value" as used herein refers to a reference value that is a preselected numerical value. The preselected numerical value can be a single number or a range. In certain embodiments, the reference value may comprise a value correlated with a value of a reference molecule. In certain embodiments, the reference value may be a thermostability reference value. In certain embodiments, the thermostability reference value is a preselected numerical value for thermostability. In certain embodiments, a thermostability reference value may comprise a value correlated with a thermostability value of a reference molecule. In certain embodiments, a thermostability reference value may comprise a parameter correlated with thermostability. In certain embodiments, the thermostability reference value may be a denaturation temperature reference value or a melting temperature (T_m) reference value. In certain embodiments, the denaturation temperature reference value is a preselected numerical value for denaturation. In certain embodiments, the denaturation temperature reference value may comprise a value correlated with a denaturation temperature value of a reference molecule. In certain embodiments, the denaturation temperature reference value may comprise a parameter correlated with denaturation. In certain embodiments, the T_m reference value may be a preselected numerical value for T_m. In certain embodiments, the T_m reference value may be a pre-determined threshold T_m. In certain embodiments, the T_m reference value may comprise a value correlated with a T_m value of a reference molecule. In certain embodiments, the T_m reference value may comprise a parameter correlated with T_m. In certain embodiments, the reference value may be an activity reference value. In certain embodiments, the activity reference value may comprise a value correlated with activity value of a reference molecule. In certain embodiments, the activity reference value may comprise a parameter correlated with an activity. In certain embodiments, the activity reference value may be a cleavage reference value or a binding reference value. In certain

embodiments, the cleavage reference value may be a preselected numerical value for cleavage of a nucleic acid. In certain embodiments, the binding reference value may be a preselected numerical value for binding of two or more molecules. In certain embodiments the reference value may be a delta reference value. In certain embodiments, the delta reference value may be a preselected numerical value for a delta value.

"Delta value" as used herein is a value representing the difference or shift between two values. For example, in certain embodiments, a delta value may be a value representing the difference between a thermostability value of a Cas9 molecule/gRNA molecule complex being evaluated and a thermostability value of a reference molecule or a thermostability reference value. In certain embodiments, a delta value may be a value representing the difference between the activity value of the Cas9 molecule/gRNA molecule complex being evaluated and the activity value of a reference molecule or an activity reference value.

"Replacement" or "replaced" as used herein with reference to a modification of a molecule does not require a process limitation but merely indicates that the replacement entity is present.

"Subject" as used herein may mean either a human or non-human animal. The term includes, but is not limited to, mammals (e.g., humans, other primates, pigs, rodents (e.g., mice and rats or hamsters), rabbits, guinea pigs, cows, horses, cats, dogs, sheep, and goats). In certain embodiments, the subject is a human. In another embodiment, the subject is poultry. In certain embodiments, the subject is a human, and in certain of these embodiments the human is an infant, child, young adult, or adult.

"X" as used herein in the context of an amino acid sequence refers to any amino acid (e.g., any of the twenty natural amino acids) unless otherwise specified.

"About" as used herein means within 10% of a stated value or a range of values.

A "Cas9 molecule" or "Cas9 polypeptide" as used herein refers to a molecule or polypeptide, respectively, that can interact with a gRNA molecule and, in concert with the gRNA molecule, localize to a site comprising a target domain and, in certain embodiments, a PAM sequence. Cas9 molecules and Cas9 polypeptides include both naturally occurring Cas9 molecules and Cas9 polypeptides and engineered, altered, or modified Cas9 molecules or Cas9 polypeptides that differ, e.g., by at least one amino acid residue, from a reference sequence, e.g., the most similar naturally occurring Cas9 molecule.

Overview

The inventors have discovered that the stability of a Cas9 molecule/gRNA molecule complex as determined by DSF is correlated with a variety of properties of the Cas9 molecule/gRNA molecule complex. As such, a determination of stability can be used to evaluate a Cas9 molecule/gRNA molecule complex (or a component thereof) for a variety of properties, e.g., the ability to cleave a target, the propensity of a cleavage event mediated by a Cas9 molecule/gRNA molecule complex to be resolved by a particular pathway, e.g., HDR or NHEJ, the ability to modulate a target, or suitability for a preselected delivery method. The determination of stability can also be used to evaluate a Cas9 molecule, a gRNA molecule, a method of preparing a complex (e.g., the proportion or stoichiometry of the components), or the addition of an additional component, e.g., on the efficacy or robustness of a Cas9 molecule/gRNA molecule complex, and generally for inclusion in a Cas9 molecule/gRNA molecule complex.

Provided herein based on the disclosed experimental results are methods that include measuring the thermostability of a Cas9/molecule gRNA molecule complex. The thermostability of a protein can increase under favorable conditions such as the addition of a binding RNA molecule, e.g., a gRNA. Thus, information regarding the thermostability of a Cas9/gRNA complex is useful for determining whether the complex is stable. The methods that may include a step of measuring the thermostability of a Cas9/molecule gRNA molecule complex include, without limitation, methods of determining the stability of a Cas9 molecule/gRNA molecule complex, methods of determining a condition that promotes a stable Cas9 molecule/gRNA molecule complex, methods of screening for a stable Cas9 molecule/gRNA molecule complex, methods for identifying an optimal gRNA to form a stable Cas9 molecule/gRNA molecule complex, methods of screening for a Cas9/gRNA complex for administration to a subject, and methods of selecting a Cas9/gRNA complex for administration to a subject. In certain embodiments, the thermostability value of a Cas9 molecule/gRNA molecule complex may be measured. Additionally, in certain embodiments, the thermostability value of a reference molecule may also be measured. In certain embodiments, the Cas9 molecule/gRNA molecule complex may be determined to be stable if the thermostability value of the Cas9 molecule/gRNA molecule complex is greater than the thermostability value of the reference molecule or a thermostability reference value as described herein. In certain embodiments, the reference molecule may be the Cas9 molecule absent the gRNA molecule. In certain embodiments, the thermostability value that is

measured may be a denaturation temperature value. In these embodiments, the thermostability reference value is a denaturation temperature reference value. In certain embodiments, the thermostability value that is measured may be a T_m value. In these embodiments, the thermostability reference value may be a T_m reference value. In certain embodiments, the thermostability value may be measured using a thermal shift assay. As disclosed herein, DSF is a technique that may be used to measure the thermostability of a protein. In certain embodiments, the thermal shift assay used to measure the thermostability may be DSF, differential scanning calorimetry (DSC) or isothermal titration calorimetry (ITC). In certain embodiments, the Cas9 molecule/gRNA molecule complex may be determined to be stable if the thermostability value of the Cas9 molecule/gRNA molecule complex is greater than a thermostability value of the reference molecule or the thermostability reference value. For example, in certain embodiments, the Cas9 molecule/gRNA molecule complex may be determined to be stable if the thermostability value of the Cas9 molecule/gRNA molecule complex is at least 8°C greater than a thermostability value of the reference molecule or the thermostability reference value. In certain embodiments, the reference molecule may be a reference Cas9 molecule in the absence of a gRNA molecule. In certain embodiments, the methods including a step to measure a thermostability value may further comprise steps that include measuring an activity value of the Cas9 molecule/gRNA molecule complex as described herein.

Also provided herein based on the disclosed experimental results are methods that include a step of measuring the activity of a Cas9/molecule gRNA molecule complex, which may also be useful in determining the stability of the complex. The methods that may include a step of measuring the activity of a Cas9/molecule gRNA molecule complex include, without limitation, methods of determining the stability of a Cas9 molecule/gRNA molecule complex, methods of determining a condition that promotes a stable Cas9 molecule/gRNA molecule complex, methods of screening for a stable Cas9 molecule/gRNA molecule complex, methods for identifying an optimal gRNA to form a stable Cas9 molecule/gRNA molecule complex, methods of screening for a Cas9 molecule/gRNA molecule complex for administration to a subject, and methods of selecting a Cas9 molecule/gRNA molecule complex for administration to a subject. In certain embodiments, an activity value of a Cas9 molecule/gRNA molecule complex may be detected. Additionally, in certain embodiments, an activity value of a reference molecule may be detected. In certain embodiments, the Cas9 molecule/gRNA molecule complex may be determined to be stable if the an activity value of

the Cas9 molecule/gRNA molecule complex is greater than an activity value of the reference molecule or the activity reference value. In certain embodiments, the activity that is being detected may be a binding activity. In certain embodiments, a binding activity may include, without limitation, the ability of the gRNA molecule to remain hybridized to the DNA target, the ability of the gRNA molecule to bind to the Cas9 molecule of the Cas9 molecule/gRNA molecule complex, or the ability of the gRNA molecule to bind to the Cas9 molecule of the Cas9 molecule/gRNA molecule complex. In certain embodiments, a binding value of the molecule may be measured. In certain embodiments, the Cas9 molecule/gRNA molecule complex may be selected or determined to be stable if the binding value of the molecule being evaluated is greater than the binding value of a reference molecule or a binding reference value. In certain embodiments, the activity is a cleavage activity. Some examples of a cleavage activity may include, without limitation, any one or more of the ability of a Cas9 molecule/gRNA molecule complex to cleave a target, the propensity of a cleavage event mediated by a Cas9 molecule/gRNA molecule complex to be resolved by a particular pathway, e.g., HDR or NHEJ, the ability of a Cas9 molecule/gRNA molecule complex to modulate a target. In certain embodiments, a cleavage value of the Cas9 molecule/gRNA molecule complex may be measured. In certain embodiments, the Cas9 molecule/gRNA molecule complex may be selected or determined to be stable if the cleavage value of the Cas9 molecule/gRNA molecule complex is greater than the cleavage value of a reference molecule or a cleavage reference value.

The methods disclosed herein may be performed on a plurality of samples. For example, in certain embodiments, the methods may comprise generating a plurality of samples, each sample comprising a Cas9 molecule/gRNA molecule complex generated by combining a Cas9 molecule and one of a plurality of gRNA molecules. In certain embodiments, a thermostability value and/or activity value of the Cas9 molecule/gRNA molecule complex may be detected in each of the plurality of samples. In certain embodiments, at least one sample comprising the Cas9 molecule/gRNA molecule complex may be selected from the plurality of samples. In certain embodiments, the sample comprising the Cas9 molecule/gRNA molecule complex may be selected based on one or more of (i) a comparison of the T_m in the plurality of samples to a T_m of a reference complex or a pre-determined threshold T_m, or (ii) a relative ordering of the T_m values of the plurality of samples.

In certain embodiments, the methods may comprise generating a plurality of Cas9 molecule/gRNA molecule complexes, each comprising a Cas9 molecule/gRNA molecule complex generated by combining a Cas9 molecule and one of a plurality of gRNA molecules. In certain embodiments, a thermostability value of the Cas9 molecule/gRNA molecule complex may be detected for each of the plurality of Cas9 molecule/gRNA molecule complexes. In certain embodiments, a Cas9 molecule/gRNA molecule complex may be determined to be stable if the thermostability value of the Cas9 molecule/gRNA molecule complex is greater than a thermostability value of a reference molecule or a thermostability reference value. In certain embodiments, an activity value of the Cas9 molecule/gRNA molecule complex may be detected for each of the plurality of Cas9 molecule/gRNA molecule complexes. In certain embodiments, a Cas9 molecule/gRNA molecule complex may be determined to be stable if the activity value of the Cas9 molecule/gRNA molecule complex is greater than an activity value of a reference molecule or an activity reference value.

Also provided herein are non-naturally occurring Cas9 molecule/gRNA molecule complexes generated using any of the methods disclosed herein.

Provided herein are compositions that may comprise any of the Cas9 molecule/gRNA molecule complexes generated using the methods described herein. For example, the compositions herein may comprise an isolated complex of a Cas9 molecule and a gRNA molecule having a T_m at least 8°C greater than a T_m of a reference molecule or T_m reference value selected according to the methods provided herein.

Also provided herein are vector systems comprising a nucleic acid encoding a Cas9 molecule/gRNA molecule complex generated using any of the methods described herein.

Provided herein are methods of delivering a Cas9 molecule/gRNA molecule complex to a target cell comprising delivering the Cas9 molecule/gRNA molecule complex generated using any of the methods described herein.

Provided herein based on the disclosed experimental results are methods for evaluating, selecting, optimizing, or designing a Cas9 molecule/gRNA molecule complex or component thereof. In certain embodiments, these methods comprise acquiring or determining a stability value (i.e., value correlated with the stability (e.g., thermostability value or activity value)) of the Cas9 molecule in a Cas9 molecule/gRNA molecule complex or a preparation thereof. In certain embodiments, the stability value may be a thermostability value or activity value of the Cas9 molecule in a Cas9 molecule/gRNA molecule complex or

a preparation thereof. In certain embodiments, the thermostability value may be a T_m value or denaturation temperature value, and in certain embodiments, the thermostability value is acquired or determined using DSF. In certain embodiments, the stability value is acquired for a Cas9 molecule complexed with a gRNA molecule. In other embodiments, the stability value is acquired for a Cas9 molecule in the presence of a gRNA molecule. In certain embodiments, the stability value is compared to a thermostability reference value (i.e., a parameter correlated with thermostability), e.g., T_m reference value, or an activity reference value (i.e., a parameter correlated with an activity), such as cleavage activity, e.g., the ability to cleave a target DNA. For example, wherein the stability value is the T_m value, the T_m value may be compared to a T_m reference value to determine whether the T_m value is identical to, greater than, or less than the T_m reference value.

In certain embodiments, the methods disclosed herein are used to select or design an optimal Cas9 molecule/gRNA pairing, e.g., a pairing with maximum stability or with a stability falling within a desired target range. In certain embodiments, the method is used to select or design one or more gRNAs for pairing with a particular Cas9 molecule, e.g., identifying gRNA molecules that complex with a particular Cas9 molecule with the greatest stability. In other embodiments, the method is used to select a Cas9 molecule for pairing with a particular gRNA or set of gRNAs, e.g., identifying Cas9 molecules that complex with a particular gRNA(s) with the greatest stability. In still other embodiments, the method is used to select both a Cas9 molecule and a gRNA or set of gRNAs for pairing in a Cas9 molecule/gRNA molecule complex.

Provided herein in certain embodiments are Cas9 molecule/gRNA complexes, or components thereof, exhibiting a desired value correlated with the stability of the Cas9 molecule in the complex, as well as compositions and pharmaceutical formulations comprising these complexes or components thereof. In certain embodiments, this value may be the T_m or denaturation temperature, and in certain embodiments the value is acquired or determined using DSF. In certain of these embodiments, the complexes are generated using the methods provided herein.

Provided herein in certain embodiments are methods comprising comparing the thermostability value. In certain embodiments, the methods comprise comparing the thermostability value with a thermostability reference value. In certain embodiments, the thermostability reference value comprises a preselected numerical value (where a value can be a single number or a range), e.g., a preselected numerical value for thermostability, e.g.,

T_m value. In certain embodiments, the thermostability reference value may be a value correlated with thermostability of a) a reference Cas9 molecule, e.g., the same Cas9 molecule as the Cas9 molecule in the complex being evaluated (or a different Cas9 molecule), in the absence of a gRNA molecule; b) a reference Cas9 molecule (e.g., the same Cas9 molecule as the Cas9 molecule in the complex being evaluated) complexed with a second gRNA molecule (i.e., a gRNA other than the one in the complex being evaluated); or c) a reference Cas9 molecule/gRNA molecule complex, wherein the reference Cas9 molecule/gRNA molecule was formed under different conditions, e.g., with different proportions of Cas9 molecule and gRNA molecule, than the Cas9 molecule/gRNA molecule complex, or was formed in a different buffer.

Provided herein in certain embodiments are reference Cas9 molecules. In certain embodiments, the reference Cas9 molecule may be the same as the Cas9 molecule of the complex being evaluated. In certain embodiments, the reference Cas9 molecule may be different, e.g., differs in primary sequence, from the Cas9 molecule of the complex being evaluated. In certain embodiments, the gRNA molecule of a reference Cas9 molecule/gRNA molecule complex may be the same as the gRNA molecule of the complex being evaluated. In certain embodiments, the gRNA molecule of a reference Cas9 molecule/gRNA molecule complex may be different from the gRNA molecule, e.g., differs in sequence or differs by a modification.

Provided herein in certain embodiments, delta values (e.g., a delta value is the difference or shift between two values) may be acquired, e.g., determined. In certain embodiments, the delta values may include a value correlated with the difference in stability of the Cas9 molecule/gRNA molecule complex being evaluated and a reference value. In certain embodiments, the delta value may include a value correlated with: the difference in stability, e.g., denaturation temperature or T_m, of the Cas9 molecule/gRNA molecule complex being evaluated; and the stability, e.g., denaturation temperature or T_m of a reference value, e.g., the value for a reference Cas9 molecule/gRNA molecule complex.

Provided herein in certain embodiments, the methods may include a step of comparing the delta value with a delta reference value (e.g., a reference value for the delta value). In certain embodiments, this may include evaluating if the delta value is equal to or less than the delta reference value; equal to or greater than the delta reference value; or is within a predetermined range of the delta reference value.

Provided herein in certain embodiments, the methods may include selecting the Cas9 molecule/gRNA molecule complex (or a component thereof). In certain embodiments, the method may include selecting the Cas9 molecule/gRNA molecule complex (or a component thereof) based on a comparison of a value of a Cas9 molecule/gRNA molecule complex (e.g., including, but not limited to a thermostability value (e.g., T_m value), activity value, or delta value to a reference value. In certain embodiments, the reference value may be, including but not limited to, a predetermined threshold, an upper bound value, a target value. In certain embodiments, the value may be a delta value. In certain embodiments, the reference value may be a delta reference value.

In certain embodiments, the methods may include evaluating if the delta value is: equal to or less than the delta reference value; equal to or greater than the delta reference value; is within a predetermined range of the delta reference value.

In certain embodiments, the methods may also include evaluating or measuring an activity or property of the selected Cas9 molecule/gRNA molecule complex (or a component thereof). In certain embodiments, the activity may be a cleavage activity, e.g., the ability to cleave. In certain embodiments, the activity may be the ability to be successfully delivered. In certain embodiments, the activity may be a binding activity, e.g., the ability of the gRNA molecule to remain hybridized to the DNA target.

In certain embodiments, the evaluating step may include evaluating or measuring the activity of the selected Cas9 molecule/gRNA molecule complex (or component thereof) in a system, such an *in vitro* system, an *ex vivo* system, or an *in vivo* system, and assay, such as a cellular assay, or a model, such as a cellular or animal model.

In certain embodiments, the evaluating step may include evaluating or measuring an activity of a selected Cas9 molecule/gRNA molecule complex (or component thereof). In certain embodiments, the activity may be a cleavage activity, such as the ability to induce indels, the ability to modify a target DNA, the propensity of a preselected repair method, e.g., a pathway described herein, e.g., HDR or NHEJ, to mediate a cleavage event catalyzed by the Cas9 molecule/gRNA molecule complex. In certain embodiments, the activity may be a binding activity, such as the ability of the gRNA molecule to remain hybridized to the DNA target.

In certain embodiments, the methods may include selecting a Cas9 molecule/gRNA molecule complex (or component thereof) for an activity. For example, the activity may be a cleavage activity such as the ability to induce indels; the ability to modify a target DNA; the

propensity of a preselected repair method, e.g., a pathway described herein, e.g., HDR or NHEJ, to mediate a cleavage event catalyzed by the Cas9 molecule/gRNA molecule complex. In certain embodiments, the activity may be a binding activity, such as the ability of the gRNA molecule to remain hybridized to the DNA target.

In certain embodiments, the methods may include designing or optimizing a Cas9 molecule/gRNA molecule complex (or component thereof) for an activity. For example, the activity may be a cleavage activity such as the ability to induce indels; the ability to modify a target DNA; the propensity of a preselected repair method, e.g., a pathway described herein, e.g., HDR or NHEJ, to mediate a cleavage event catalyzed by the Cas9 molecule/gRNA molecule complex. In certain embodiments, the activity may be a binding activity, such as the ability of the gRNA molecule to remain hybridized to the DNA target.

In certain embodiments, the method may include determining the stability of the Cas9 molecule of at least X Cas9 molecule/gRNA molecule complexes. In certain embodiments, X may be equal to 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 500, or 1,000. In certain embodiments, determining the stability may include measuring the stability value of the Cas 9 molecule. In certain embodiments, the stability value may be the thermostability value. In certain embodiments, the thermostability value may be the T_m value. In certain embodiments, the T_m value may be determined by DSF.

In certain embodiments, determining stability may comprise determining the temperature at which the Cas9 molecule of a Cas9 molecule/gRNA molecule complex denatures, e.g., determining the T_m value. In certain embodiments, the T_m value may be determined by DSF.

In certain embodiments, the methods may comprise determining by DSF, the temperature at which the Cas9 molecule of a Cas9 molecule/gRNA molecule complex denatures, e.g., determining the T_m value, for a first Cas9 molecule/gRNA molecule complex and a second Cas9 molecule/gRNA molecule complex.

In certain embodiments, a first Cas9 molecule/gRNA molecule complex and a second Cas9 molecule/gRNA molecule complex of the at least X Cas9 molecule/gRNA molecule complexes differ by having Cas9 molecules of different sequence, having gRNA molecules that differ in sequence or by modification, by capping or tailing, having been formed under different conditions, e.g., different stoichiometries.

In certain embodiments, responsive to the determination of stability, a Cas9 molecule/gRNA molecule complex (or a component there of) may be selected for optimized

or preselected delivery characteristics, e.g., wherein delivery comprises delivery by RNP cationic lipid transfection, a viral vector (e.g., AAV), or RNA transfection.

In certain embodiments, responsive to the determination of stability, a Cas9 molecule/gRNA molecule complex (or a component there of) is selected for optimized or preselected relationship with a quality control standard.

In certain embodiments, responsive to the determination of stability, a Cas9 molecule/gRNA molecule complex (or a component thereof) may be selected as meeting a quality control or release standard.

In certain embodiments, methods herein may comprise selecting a Cas9 molecule/gRNA molecule complex (or a component thereof) if a thermostability value, e.g., a T_m value or denaturation temperature value, or a delta value, has a preselected relationship with a reference value or a delta reference value.

In certain embodiments, responsive to the determination of stability, a Cas9 molecule/gRNA molecule complex (or a component thereof) is selected for optimized or preselected characteristic, e.g., a cleavage characteristic.

In certain embodiments, responsive to evaluation of a delta value between a thermostability value for a Cas9 molecule/gRNA molecule complex and a thermostability value of a reference molecule, e.g., Cas9 molecule in the absence of gRNA molecule, a Cas9 molecule/gRNA molecule complex (or a component there of) is selected.

In certain embodiments, responsive to evaluation of a delta value between an activity value for a Cas9 molecule/gRNA molecule complex and an activity value of a reference molecule, e.g., Cas9 molecule in the absence of gRNA molecule, a Cas9 molecule/gRNA molecule complex (or a component there of) is selected.

In certain embodiments, methods herein may comprise evaluating a library of (or a single) candidate gRNA molecules, e.g., a library of tracrRNA molecules or sequences, complexed with a Cas9 molecule, and responsive to the determination of stability of a Cas9 molecule/candidate gRNA molecule complex, selecting a candidate gRNA molecule or sequence, e.g., a candidate tracr gRNA molecule or sequence.

In certain embodiments, the library may comprise tracr RNA molecules or sequences of differing structure, e.g., differing length, differing sequence, or having different modifications, e.g., having additional phosphate groups or alternative 5' cap structures.

In certain embodiments, the tracr RNA molecule or sequences may be disposed on a chimeric gRNA.

In certain embodiments, the method s herein may comprise evaluating a component for inclusion in a Cas9 molecule/gRNA molecule preparation, comprising evaluating the stability of the Cas9 molecule of the Cas9 molecule/gRNA molecule complex in a preparation comprising the component. In certain embodiments, the component may comprise an additive, a small molecule, a stabilizing reagent, buffer, pH, salt concentration, glycerol concentration, or other buffer component.

In certain embodiments, the methods herein may comprise evaluating a candidate Cas 9 molecule for inclusion in a Cas9 molecule/gRNA molecule complex, comprising evaluating the stability of the Cas9 molecule of the Cas9 molecule/gRNA molecule complex. In certain embodiments, the candidate Cas9 molecule may comprise a chimeric Cas9 molecule, or a synthetic or engineered Cas9 molecule, e.g., a Cas9 molecule with a portion or portions deleted.

In certain embodiments, determining stability may comprise, determining by differential scanning fluorimetry (DSF), the temperature at which the Cas9 molecule of a Cas9 molecule/gRNA molecule complex denatures, e.g., the T_m, of the Cas9 molecule.

Provided herein are reaction mixtures. The reaction mixtures may comprise a Cas9 molecule/gRNA molecule complex, e.g., a Cas9 molecule/gRNA molecule complex described herein; and a signal emitting compound, e.g., a dye, wherein signal emission is correlated to denaturation of the Cas9 molecule.

Provided herein are differential scanning fluorimeters having disposed therein: a Cas9 molecule/gRNA molecule complex; and a signal emitting compound, e.g., a dye, wherein signal emission is correlated to denaturation of the Cas9 molecule.

Provided herein are Cas9 molecule/gRNA molecule complexes evaluated, selected, optimized, or designed by a method described herein.

Provided herein are compositions comprising a Cas9 molecule/gRNA molecule complex selected or designed by a method described herein. In certain embodiments, the compositions may be a pharmaceutical composition. In certain embodiments, the Cas9 molecule/gRNA molecule complex may be formulated in a pharmaceutically acceptable carrier.

As discussed herein the methods disclosed herein may be used to evaluate, select or design a complex optimized for formulation or delivery. For example, methods described herein can be used in multiple ways for improving complex formation of a chimeric gRNA or a tracrRNA with a Cas9 molecule (e.g., a Cas9 protein) using any delivery method such as,

but not limited to, RNP cationic lipid transfection, viral vectors (e.g., AAV), or RNA transfections.

Methods discussed herein can be used for quality control, or to determine if a release standard has been met for both protein and RNA components. For example, if a standard for a thermal shift is not met, e.g., where there is no thermal shift when incubating Cas9 protein with RNA, a Cas9 molecule or a gRNA molecule of insufficient quality is indicated. In an embodiment, the method is used as a guide, or as a process control, to address such issues, e.g., by the detection and removal of impurities.

Methods described herein can be used for assessing libraries of candidate Cas9 molecules or candidate gRNA molecules for use in a Cas9/gRNA complex. In an embodiment, the method identifies components for optimized binding. This can allow screening of candidates to optimized target cleavage, or other properties. Methods described herein can also be used to evaluate modifications in the length and compositions of gRNA. For example, after purifying a mutant Cas9 protein, a library of gRNA molecules (e.g., tracrRNA molecules) or sequences can be incubated with the protein at a preselected ratio, e.g., at a minimum 1:1 ratio of RNA:protein. The appearance of a thermal shift observed when compared to the Cas9 protein in the absence of a gRNA molecule is indicative of an effective gRNA molecule, e.g., a gRNA molecule capable of mediating one or more CRISPR/Cas-related activities *in vitro, ex vivo*, or *in vivo*.

The methods described herein can be used to screen a library of gRNA molecules (e.g., tracrRNA molecules) or sequences having tracr-regions of different length, combined with cognate guide sequences fused with different linker sequences and assayed for binding using DSF. Well bound complexes could be screened for cutting activity *in vitro*. Methods described herein allow for the evaluation of chemical modifications of RNA, additional phosphate groups, or alternative 5' cap structures, for effect on cleavage, suitability for delivery or other characteristics discussed herein.

The methods described herein can be used to screen a library of components, such as additives, small molecule stabilizing reagents, buffers, salt, e.g., salt molarity, glycerol concentration, and other buffer components for stabilizing the interaction of a Cas9 molecule and a gRNA molecule.

The methods described herein can be used to screen a library of chimeric, engineered, or synthetic Cas9 molecules, e.g., chimeric or engineered Cas9 molecules for stabilizing the interaction of Cas9 molecule and gRNA molecule. Insufficient thermal shift is indicative that

RNA binding is sub-optimal or disrupted. A partial thermal shift would imply that the RNA is binding productively. For chimeric or engineered Cas9 molecules which have no DSF T_m, libraries of gRNA molecules (e.g., tracrRNA molecules) or sequences can be screened for restoring binding, as measured by DSF.

A Cas9 molecule perceived as inactive can be used in a thermostability assay against a nucleic acid molecule library, e.g., a randomized nucleic acid library. This will allow screening for novel molecules that can fulfill the role of tracrRNA. Using this newly discovered nucleic acid molecule (e.g., tracrRNA molecule) one can develop new gRNAs to target genomic DNA, *in vivo* RNA, and/or genetic material from invasive organisms and viruses.

The methods described herein can be applied to any mutated and chimeric forms of Cas9 molecule. It is also understood that the methods described herein can be applied to other Cas molecules, e.g., other Cas molecules described herein.

Guide RNA (gRNA) molecules

A gRNA molecule, as that term is used herein, refers to a nucleic acid that promotes the specific targeting or homing of a gRNA molecule/Cas9 molecule complex to a target nucleic acid. gRNA molecules can be unimolecular (having a single RNA molecule), sometimes referred to herein as "chimeric" gRNAs, or modular (comprising more than one, and typically two, separate RNA molecules). The gRNA molecules provided herein comprise a targeting domain comprising, consisting of, or consisting essentially of a nucleic acid sequence fully or partially complementary to a target nucleic acid sequence. In certain embodiments, the gRNA molecule further comprises one or more additional domains, including for example a first complementarity domain, a linking domain, a second complementarity domain, a proximal domain, a tail domain, and/or a 5' extension domain. Each of these domains is discussed in detail below. In certain embodiments, one or more of the domains in the gRNA molecule comprises an amino acid sequence identical to or sharing sequence homology with a naturally occurring sequence, e.g., from *S. pyogenes, S. aureus,* or *S. thermophilus*.

Several exemplary gRNA structures are provided in **Figs. 1A-1I**. With regard to the three-dimensional form, or intra- or inter-strand interactions of an active form of a gRNA, regions of high complementarity are sometimes shown as duplexes in **Figs. 1A-1I** and other depictions provided herein. **Fig. 7** illustrates gRNA domain nomenclature using the gRNA

sequence of SEQ ID NO:42, which contains one hairpin loop in the tracrRNA-derived region. In certain embodiments, a gRNA may contain more than one (e.g., two, three, or more) hairpin loops in this region (see, e.g., **Figs. 1H-1I**).

In certain embodiments, a unimolecular, or chimeric, gRNA comprises, preferably from 5' to 3':

a targeting domain comprising, consisting of, or consisting essentially of a nucleic acid sequence fully or partially complementary to a target nucleic acid sequence;

a first complementarity domain;

a linking domain;

a second complementarity domain (which is complementary to the first complementarity domain);

a proximal domain; and

optionally, a tail domain.

In certain embodiments, a modular gRNA comprises:

a first strand comprising, preferably from 5' to 3':

a targeting domain comprising, consisting of, or consisting essentially of a nucleic acid sequence fully or partially complementary to a target nucleic acid sequence; and

a first complementarity domain; and

a second strand, comprising, preferably from 5' to 3':

optionally, a 5' extension domain;

a second complementarity domain;

a proximal domain; and

optionally, a tail domain.

Targeting domain

The targeting domain (sometimes referred to alternatively as the guide sequence or complementarity region) comprises, consists of, or consists essentially of a nucleic acid sequence that is complementary or partially complementary to a target nucleic acid. The nucleic acid sequence to which all or a portion of the targeting domain is complementary or partially complementary is referred to herein as the target domain. Methods for selecting targeting domains are known in the art (see, e.g., Fu 2014; Sternberg 2014).

The strand of the target nucleic acid comprising the target domain is referred to herein as the complementary strand because it is complementary to the targeting domain sequence. Since the targeting domain is part of a gRNA molecule, it comprises the base uracil (U) rather than thymine (T); conversely, any DNA molecule encoding the gRNA molecule will comprise thymine rather than uracil. In a targeting domain/target domain pair, the uracil bases in the targeting domain will pair with the adenine bases in the target domain. In certain embodiments, the degree of complementarity between the targeting domain and target domain is sufficient to allow targeting of a gRNA molecule/Cas9 molecule complex to the target nucleic acid.

In certain embodiments, the targeting domain comprises a core domain and an optional secondary domain. In certain of these embodiments, the core domain is located 3' to the secondary domain, and in certain of these embodiments the core domain is located at or near the 3' end of the targeting domain. In certain of these embodiments, the core domain consists of or consists essentially of about 8 to about 13 nucleotides at the 3' end of the targeting domain. In certain embodiments, only the core domain is complementary or partially complementary to the corresponding portion of the target domain, and in certain of these embodiments the core domain is fully complementary to the corresponding portion of the target domain. In other embodiments, the secondary domain is also complementary or partially complementary to a portion of the target domain. In certain embodiments, the core domain is complementary or partially complementary to a core domain target in the target domain, while the secondary domain is complementary or partially complementary to a secondary domain target in the target domain. In certain embodiments, the core domain and secondary domain have the same degree of complementarity with their respective corresponding portions of the target domain. In other embodiments, the degree of complementarity between the core domain and its target and the degree of complementarity between the secondary domain and its target may differ. In certain of these embodiments, the core domain may have a higher degree of complementarity for its target than the secondary domain, whereas in other embodiments the secondary domain may have a higher degree of complementarity than the core domain.

In certain embodiments, the targeting domain and/or the core domain within the targeting domain is 3 to 100, 5 to 100, 10 to 100, or 20 to 100 nucleotides in length, and in certain of these embodiments the targeting domain or core domain is 3 to 15, 3 to 20, 5 to 20, 10 to 20, 15 to 20, 5 to 50, 10 to 50, or 20 to 50 nucleotides in length. In certain

embodiments, the targeting domain and/or the core domain within the targeting domain is 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain embodiments, the targeting domain and/or the core domain within the targeting domain is 6 +/-2, 7+/-2, 8+/-2, 9+/-2, 10+/-2, 10+/-4, 10 +/-5, 11+/-2, 12+/-2, 13+/-2, 14+/-2, 15+/-2, or 16+-2, 20+/-5, 30+/-5, 40+/-5, 50+/-5, 60+/-5, 70+/-5, 80+/-5, 90+/-5, or 100+/-5 nucleotides in length.

In certain embodiments wherein the targeting domain includes a core domain, the core domain is 3 to 20 nucleotides in length, and in certain of these embodiments the core domain 5 to 15 or 8 to 13 nucleotides in length. In certain embodiments wherein the targeting domain includes a secondary domain, the secondary domain is 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 nucleotides in length. In certain embodiments wherein the targeting domain comprises a core domain that is 8 to 13 nucleotides in length, the targeting domain is 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, or 16 nucleotides in length, and the secondary domain is 13 to 18, 12 to 17, 11 to 16, 10 to 15, 9 to 14, 8 to 13, 7 to 12, 6 to 11, 5 to 10, 4 to 9, or 3 to 8 nucleotides in length, respectively.

In certain embodiments, the targeting domain is fully complementary to the target domain. Likewise, where the targeting domain comprises a core domain and/or a secondary domain, in certain embodiments one or both of the core domain and the secondary domain are fully complementary to the corresponding portions of the target domain. In other embodiments, the targeting domain is partially complementary to the target domain, and in certain of these embodiments where the targeting domain comprises a core domain and/or a secondary domain, one or both of the core domain and the secondary domain are partially complementary to the corresponding portions of the target domain. In certain of these embodiments, the nucleic acid sequence of the targeting domain, or the core domain or targeting domain within the targeting domain, is at least 80, 85, 90, or 95% complementary to the target domain or to the corresponding portion of the target domain. In certain embodiments, the targeting domain and/or the core or secondary domains within the targeting domain include one or more nucleotides that are not complementary with the target domain or a portion thereof, and in certain of these embodiments the targeting domain and/or the core or secondary domains within the targeting domain include 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides that are not complementary with the target domain. In certain embodiments, the core domain includes 1, 2, 3, 4, or 5 nucleotides that are not complementary with the corresponding portion of the target domain. In certain embodiments wherein the targeting domain includes

one or more nucleotides that are not complementary with the target domain, one or more of said non-complementary nucleotides are located within five nucleotides of the 5' or 3' end of the targeting domain. In certain of these embodiments, the targeting domain includes 1, 2, 3, 4, or 5 nucleotides within five nucleotides of its 5' end, 3' end, or both its 5' and 3' ends that are not complementary to the target domain. In certain embodiments wherein the targeting domain includes two or more nucleotides that are not complementary to the target domain, two or more of said non-complementary nucleotides are adjacent to one another, and in certain of these embodiments the two or more consecutive non-complementary nucleotides are located within five nucleotides of the 5' or 3' end of the targeting domain. In other embodiments, the two or more consecutive non-complementary nucleotides are both located more than five nucleotides from the 5' and 3' ends of the targeting domain.

In certain embodiments, the targeting domain, core domain, and/or secondary domain do not comprise any modifications. In other embodiments, the targeting domain, core domain, and/or secondary domain, or one or more nucleotides therein, have a modification, including but not limited to the modifications set forth below. In certain embodiments, one or more nucleotides of the targeting domain, core domain, and/or secondary domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the targeting domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the targeting domain, core domain, and/or secondary domain render the targeting domain and/or the gRNA comprising the targeting domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the targeting domain and/or the core or secondary domains include 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the targeting domain and/or core or secondary domains include 1, 2, 3, or 4 modifications within five nucleotides of their respective 5' ends and/or 1, 2, 3, or 4 modifications within five nucleotides of their respective 3' ends. In certain embodiments, the targeting domain and/or the core or secondary domains comprise modifications at two or more consecutive nucleotides.

In certain embodiments wherein the targeting domain includes core and secondary domains, the core and secondary domains contain the same number of modifications. In certain of these embodiments, both domains are free of modifications. In other embodiments, the core domain includes more modifications than the secondary domain, or vice versa.

In certain embodiments, modifications to one or more nucleotides in the targeting domain, including in the core or secondary domains, are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification using a system as set forth below. gRNAs having a candidate targeting domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated using a system as set forth below. The candidate targeting domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

In certain embodiments, all of the modified nucleotides are complementary to and capable of hybridizing to corresponding nucleotides present in the target domain. In another embodiment, 1, 2, 3, 4, 5, 6, 7, or 8 or more modified nucleotides are not complementary to or capable of hybridizing to corresponding nucleotides present in the target domain.

Figs. 1A-1I provide examples of the placement of the targeting domain within a gRNA molecule.

First and second complementarity domains

The first and second complementarity (sometimes referred to alternatively as the crRNA-derived hairpin sequence and tracrRNA-derived hairpin sequences, respectively) domains are fully or partially complementary to one another. In certain embodiments, the degree of complementarity is sufficient for the two domains to form a duplexed region under at least some physiological conditions. In modular gRNA molecules, the two molecules are associated by virtue of the hybridization of the complementarity domains (see e.g., **Fig. 1A**). In certain embodiments, the degree of complementarity between the first and second complementarity domains, together with other properties of the gRNA, is sufficient to allow targeting of a Cas9 molecule to a target nucleic acid. Examples of first and second complementarity domains are set forth in **Figs. 1A-1G**.

In certain embodiments (see, e.g., **Figs. 1A-1B**) the first and/or second complementarity domain includes one or more nucleotides that lack complementarity with the corresponding complementarity domain. In certain embodiments, the first and/or second complementarity domain includes 1, 2, 3, 4, 5, or 6 nucleotides that do not complement with the corresponding complementarity domain. For example, the second complementarity domain may contain 1, 2, 3, 4, 5, or 6 nucleotides that do not pair with corresponding nucleotides in the first complementarity domain. In certain embodiments, the nucleotides on

the first or second complementarity domain that do not complement with the corresponding complementarity domain loop out from the duplex formed between the first and second complementarity domains. In certain of these embodiments, the unpaired loop-out is located on the second complementarity domain, and in certain of these embodiments the unpaired region begins 1, 2, 3, 4, 5, or 6 nucleotides from the 5' end of the second complementarity domain.

In certain embodiments, the first complementarity domain is 5 to 30, 5 to 25, 7 to 25, 5 to 24, 5 to 23, 7 to 22, 5 to 22, 5 to 21, 5 to 20, 7 to 18, 7 to 15, 9 to 16, or 10 to 14 nucleotides in length, and in certain of these embodiments the first complementarity domain is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. In certain embodiments, the second complementarity domain is 5 to 27, 7 to 27, 7 to 25, 5 to 24, 5 to 23, 5 to 22, 5 to 21, 7 to 20, 5 to 20, 7 to 18, 7 to 17, 9 to 16, or 10 to 14 nucleotides in length, and in certain of these embodiments the second complementarity domain is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain embodiments, the first and second complementarity domains are each independently 6 +/-2, 7+/-2, 8+/-2, 9+/-2, 10+/-2, 11+/-2, 12+/-2, 13+/-2, 14+/-2, 15+/-2, 16+/-2, 17+/-2, 18+/-2, 19+/-2, or 20+/-2, 21+/-2, 22+/-2, 23+/-2, or 24+/-2 nucleotides in length. In certain embodiments, the second complementarity domain is longer than the first complementarity domain, e.g., 2, 3, 4, 5, or 6 nucleotides longer.

In certain embodiments, the first and/or second complementarity domains each independently comprise three subdomains, which, in the 5' to 3' direction are: a 5' subdomain, a central subdomain, and a 3' subdomain. In certain embodiments, the 5' subdomain and 3' subdomain of the first complementarity domain are fully or partially complementary to the 3' subdomain and 5' subdomain, respectively, of the second complementarity domain.

In certain embodiments, the 5' subdomain of the first complementarity domain is 4 to 9 nucleotides in length, and in certain of these embodiments the 5' domain is 4, 5, 6, 7, 8, or 9 nucleotides in length. In certain embodiments, the 5' subdomain of the second complementarity domain is 3 to 25, 4 to 22, 4 to 18, or 4 to 10 nucleotides in length, and in certain of these embodiments the 5' domain is 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. In certain embodiments, the central subdomain of the first complementarity domain is 1, 2, or 3 nucleotides in length. In certain embodiments, the central subdomain of the second complementarity domain is 1, 2, 3, 4, or 5

nucleotides in length. In certain embodiments, the 3' subdomain of the first complementarity domain is 3 to 25, 4 to 22, 4 to 18, or 4 to 10 nucleotides in length, and in certain of these embodiments the 3' subdomain is 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. In certain embodiments, the 3' subdomain of the second complementarity domain is 4 to 9, e.g., 4, 5, 6, 7, 8, or 9 nucleotides in length.

The first and/or second complementarity domains can share homology with, or be derived from, naturally occurring or reference first and/or second complementarity domains. In certain of these embodiments, the first and/or second complementarity domains have at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with, or differ by no more than 1, 2, 3, 4, 5, or 6 nucleotides from, the naturally occurring or reference first and/or second complementarity domain. In certain of these embodiments, the first and/or second complementarity domains may have at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with homology with a first and/or second complementarity domain from *S. pyogenes* or *S. aureus*.

In certain embodiments, the first and/or second complementarity domains do not comprise any modifications. In other embodiments, the first and/or second complementarity domains or one or more nucleotides therein have a modification, including but not limited to a modification set forth below. In certain embodiments, one or more nucleotides of the first and/or second complementarity domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the targeting domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the first and/or second complementarity domain render the first and/or second complementarity domain and/or the gRNA comprising the first and/or second complementarity less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the first and/or second complementarity domains each independently include 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the first and/or second complementarity domains each independently include 1, 2, 3, or 4 modifications within five nucleotides of their respective 5' ends, 3' ends, or both their 5' and 3' ends. In other embodiments, the first and/or second complementarity domains each independently contain no modifications within five nucleotides of their respective 5' ends, 3' ends, or both their 5' and 3' ends. In certain embodiments, one or both of the first and second complementarity domains comprise modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the first and/or second complementarity domains are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in the system set forth below. gRNAs having a candidate first or second complementarity domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated using a system as set forth below. The candidate complementarity domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

In certain embodiments, the duplexed region formed by the first and second complementarity domains is, for example, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 bp in length, excluding any looped out or unpaired nucleotides.

In certain embodiments, the first and second complementarity domains, when duplexed, comprise 11 paired nucleotides (see, for e.g., gRNA of SEQ ID NO:48). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 15 paired nucleotides (see, e.g., gRNA of SEQ ID NO:50). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 16 paired nucleotides (see, e.g., gRNA of SEQ ID NO:51). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 21 paired nucleotides (see, e.g., gRNA of SEQ ID NO:29).

In certain embodiments, one or more nucleotides are exchanged between the first and second complementarity domains to remove poly-U tracts. For example, nucleotides 23 and 48 or nucleotides 26 and 45 of the gRNA of SEQ ID NO:48 may be exchanged to generate the gRNA of SEQ ID NOs:49 or 31, respectively. Similarly, nucleotides 23 and 39 of the gRNA of SEQ ID NO:29 may be exchanged with nucleotides 50 and 68 to generate the gRNA of SEQ ID NO:30.

Linking domain

The linking domain is disposed between and serves to link the first and second complementarity domains in a unimolecular or chimeric gRNA. **Figs. 1B-1E** provide examples of linking domains. In certain embodiments, part of the linking domain is from a crRNA-derived region, and another part is from a tracrRNA-derived region.

In certain embodiments, the linking domain links the first and second complementarity domains covalently. In certain of these embodiments, the linking domain

consists of or comprises a covalent bond. In other embodiments, the linking domain links the first and second complementarity domains non-covalently. In certain embodiments, the linking domain is ten or fewer nucleotides in length, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. In other embodiments, the linking domain is greater than 10 nucleotides in length, e.g., 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more nucleotides. In certain embodiments, the linking domain is 2 to 50, 2 to 40, 2 to 30, 2 to 20, 2 to 10, 2 to 5, 10 to 100, 10 to 90, 10 to 80, 10 to 70, 10 to 60, 10 to 50, 10 to 40, 10 to 30, 10 to 20, 10 to 15, 20 to 100, 20 to 90, 20 to 80, 20 to 70, 20 to 60, 20 to 50, 20 to 40, 20 to 30, or 20 to 25 nucleotides in length. In certain embodiments, the linking domain is 10 +/-5, 20+/-5, 20+/-10, 30+/-5, 30+/-10, 40+/-5, 40+/-10, 50+/-5, 50+/-10, 60+/-5, 60+/-10, 70+/-5, 70+/-10, 80+/-5, 80+/-10, 90+/-5, 90+/-10, 100+/-5, or 100+/-10 nucleotides in length.

In certain embodiments, the linking domain shares homology with, or is derived from, a naturally occurring sequence, e.g., the sequence of a tracrRNA that is 5' to the second complementarity domain. In certain embodiments, the linking domain has at least 50%, 60%, 70%, 80%, 90%, or 95% homology with or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from a linking domain disclosed herein, e.g., the linking domains of **Figs. 1B-1E**.

In certain embodiments, the linking domain does not comprise any modifications. In other embodiments, the linking domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth below. In certain embodiments, one or more nucleotides of the linking domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the linking domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the linking domain render the linking domain and/or the gRNA comprising the linking domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the linking domain includes 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the linking domain includes 1, 2, 3, or 4 modifications within five nucleotides of its 5' and/or 3' end. In certain embodiments, the linking domain comprises modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the linking domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification using a system as set forth below. gRNAs having a candidate linking domain having a selected length, sequence, degree of complementarity, or degree of

modification can be evaluated in a system as set forth below. The candidate linking domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

In certain embodiments, the linking domain comprises a duplexed region, typically adjacent to or within 1, 2, or 3 nucleotides of the 3' end of the first complementarity domain and/or the 5' end of the second complementarity domain. In certain of these embodiments, the duplexed region of the linking region is 10+/-5, 15+/-5, 20+/-5, 20+/-10, or 30+/-5 bp in length. In certain embodiments, the duplexed region of the linking domain is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 bp in length. In certain embodiments, the sequences forming the duplexed region of the linking domain are fully complementarity. In other embodiments, one or both of the sequences forming the duplexed region contain one or more nucleotides (e.g., 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides) that are not complementary with the other duplex sequence.

5' extension domain

In certain embodiments, a modular gRNA as disclosed herein comprises a 5' extension domain, i.e., one or more additional nucleotides 5' to the second complementarity domain (see, e.g., **Fig. 1A**). In certain embodiments, the 5' extension domain is 2 to 10 or more, 2 to 9, 2 to 8, 2 to 7, 2 to 6, 2 to 5, or 2 to 4 nucleotides in length, and in certain of these embodiments the 5' extension domain is 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more nucleotides in length.

In certain embodiments, the 5' extension domain nucleotides do not comprise modifications, e.g., modifications of the type provided below. However, in certain embodiments, the 5' extension domain comprises one or more modifications, e.g., modifications that it render it less susceptible to degradation or more bio-compatible, e.g., less immunogenic. By way of example, the backbone of the 5' extension domain can be modified with a phosphorothioate, or other modification(s) as set forth below. In certain embodiments, a nucleotide of the 5' extension domain can comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation, or other modification(s) as set forth below.

In certain embodiments, the 5' extension domain can comprise as many as 1, 2, 3, 4, 5, 6, 7, or 8 modifications. In certain embodiments, the 5' extension domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 5' end, e.g., in a modular

gRNA molecule. In certain embodiments, the 5' extension domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 3' end, e.g., in a modular gRNA molecule.

In certain embodiments, the 5' extension domain comprises modifications at two consecutive nucleotides, e.g., two consecutive nucleotides that are within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension domain, or more than 5 nucleotides away from one or both ends of the 5' extension domain. In certain embodiments, no two consecutive nucleotides are modified within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension domain, or within a region that is more than 5 nucleotides away from one or both ends of the 5' extension domain. In certain embodiments, no nucleotide is modified within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension domain, or within a region that is more than 5 nucleotides away from one or both ends of the 5' extension domain.

Modifications in the 5' extension domain can be selected so as to not interfere with gRNA molecule efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate 5' extension domain having a selected length, sequence, degree of complementarity, or degree of modification, can be evaluated in a system as set forth below. The candidate 5' extension domain can be placed, either alone, or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

In certain embodiments, the 5' extension domain has at least 60, 70, 80, 85, 90 or 95% homology with, or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from, a reference 5' extension domain, e.g., a naturally occurring, e.g., an *S. pyogenes*, *S. aureus*, or *S. thermophilus*, 5' extension domain, or a 5' extension domain described herein, e.g., from **Figs. 1A-1G**.

Proximal domain

Figs. 1A-1G provide examples of proximal domains.

In certain embodiments, the proximal domain is 5 to 20 or more nucleotides in length, e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain of these embodiments, the proximal domain is 6 +/-2, 7+/-2, 8+/-2, 9+/-2, 10+/-2, 11+/-2, 12+/-2, 13+/-2, 14+/-2, 14+/-2, 16+/-2, 17+/-2, 18+/-2, 19+/-2,

or 20+/-2 nucleotides in length. In certain embodiments, the proximal domain is 5 to 20, 7, to 18, 9 to 16, or 10 to 14 nucleotides in length.

In certain embodiments, the proximal domain can share homology with or be derived from a naturally occurring proximal domain. In certain of these embodiments, the proximal domain has at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from a proximal domain disclosed herein, e.g., an *S. pyogenes*, *S. aureus*, or *S. thermophilus* proximal domain, including those set forth in **Figs. 1A-1G**.

In certain embodiments, the proximal domain does not comprise any modifications. In other embodiments, the proximal domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth in herein. In certain embodiments, one or more nucleotides of the proximal domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the proximal domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the proximal domain render the proximal domain and/or the gRNA comprising the proximal domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the proximal domain includes 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the proximal domain includes 1, 2, 3, or 4 modifications within five nucleotides of its 5' and/or 3' end. In certain embodiments, the proximal domain comprises modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the proximal domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate proximal domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated in a system as set forth below. The candidate proximal domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

Tail domain

A broad spectrum of tail domains are suitable for use in the gRNA molecules disclosed herein. **Figs. 1A** and **1C-1G** provide examples of such tail domains.

In certain embodiments, the tail domain is absent. In other embodiments, the tail domain is 1 to 100 or more nucleotides in length, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides in length. In certain embodiments, the tail domain is 1 to 5, 1 to 10, 1 to 15, 1 to 20, 1 to 50, 10 to 100, 20 to 100, 10 to 90, 20 to 90, 10 to 80, 20 to 80, 10 to 70, 20 to 70, 10 to 60, 20 to 60, 10 to 50, 20 to 50, 10 to 40, 20 to 40, 10 to 30, 20 to 30, 20 to 25, 10 to 20, or 10 to 15 nucleotides in length. In certain embodiments, the tail domain is 5 +/-5, 10 +/-5, 20+/-10, 20+/-5, 25+/-10, 30+/-10, 30+/-5, 40+/-10, 40+/-5, 50+/-10, 50+/-5, 60+/-10, 60+/-5, 70+/-10, 70+/-5, 80+/-10, 80+/-5, 90+/-10, 90+/-5, 100+/-10, or 100+/-5 nucleotides in length.

In certain embodiments, the tail domain can share homology with or be derived from a naturally occurring tail domain or the 5' end of a naturally occurring tail domain. In certain of these embodiments, the proximal domain has at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from a naturally occurring tail domain disclosed herein, e.g., an *S. pyogenes*, *S. aureus*, or *S. thermophilus* tail domain, including those set forth in **Figs. 1A** and **1C-1G**.

In certain embodiments, the tail domain includes sequences that are complementary to each other and which, under at least some physiological conditions, form a duplexed region. In certain of these embodiments, the tail domain comprises a tail duplex domain which can form a tail duplexed region. In certain embodiments, the tail duplexed region is 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 bp in length. In certain embodiments, the tail domain comprises a single stranded domain 3' to the tail duplex domain that does not form a duplex. In certain of these embodiments, the single stranded domain is 3 to 10 nucleotides in length, e.g., 3, 4, 5, 6, 7, 8, 9, 10, or 4 to 6 nucleotides in length.

In certain embodiments, the tail domain does not comprise any modifications. In other embodiments, the tail domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth herein. In certain embodiments, one or more nucleotides of the tail domain may comprise a 2' modification (e.g., a modification at the 2' position on ribose), e.g., a 2-acetylation, e.g., a 2' methylation. In certain embodiments, the backbone of the tail domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the tail domain render the tail domain and/or the gRNA comprising the tail domain less susceptible to degradation or more bio-compatible, e.g., less immunogenic. In certain embodiments, the tail domain includes 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments

the tail domain includes 1, 2, 3, or 4 modifications within five nucleotides of its 5' and/or 3' end. In certain embodiments, the tail domain comprises modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the tail domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification as set forth below. gRNAs having a candidate tail domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated using a system as set forth below. The candidate tail domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

In certain embodiments, the tail domain includes nucleotides at the 3' end that are related to the method of *in vitro* or *in vivo* transcription. When a T7 promoter is used for *in vitro* transcription of the gRNA, these nucleotides may be any nucleotides present before the 3' end of the DNA template. When a U6 promoter is used for *in vivo* transcription, these nucleotides may be the sequence UUUUUU. When an H1 promoter is used for transcription, these nucleotides may be the sequence UUUU. When alternate pol-III promoters are used, these nucleotides may be various numbers of uracil bases depending on, e.g., the termination signal of the pol-III promoter, or they may include alternate bases.

In certain embodiments, the proximal and tail domain taken together comprise, consist of, or consist essentially of the sequence set forth in SEQ ID NOs:32, 33, 34, 35, 36, or 37.

Exemplary unimolecular/chimeric gRNAs

In certain embodiments, a unimolecular or chimeric gRNA as disclosed herein has the structure: 5' [targeting domain]-[first complementarity domain]-[linking domain]-[second complementarity domain]-[proximal domain]-[tail domain]-3', wherein:

the targeting domain comprises a core domain and optionally a secondary domain, and is 10 to 50 nucleotides in length;

the first complementarity domain is 5 to 25 nucleotides in length and, in certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference first complementarity domain disclosed herein;

the linking domain is 1 to 5 nucleotides in length;

the second complementarity domain is 5 to 27 nucleotides in length and, in certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference second complementarity domain disclosed herein;

the proximal domain is 5 to 20 nucleotides in length and, in certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference proximal domain disclosed herein; and

the tail domain is absent or a nucleotide sequence is 1 to 50 nucleotides in length and, in certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference tail domain disclosed herein.

In certain embodiments, a unimolecular gRNA as disclosed herein comprises, preferably from 5' to 3':

a targeting domain, e.g., comprising 10-50 nucleotides;

a first complementarity domain, e.g., comprising 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides;

a linking domain;

a second complementarity domain;

a proximal domain; and

a tail domain,

wherein,

(a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides;

(b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; or

(c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the sequence from (a), (b), and/or (c) has at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% homology with the corresponding sequence of a naturally occurring gRNA, or with a gRNA described herein.

In certain embodiments, the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that are complementary to the corresponding nucleotides of the first complementarity domain.

In certain embodiments, the targeting domain consists of, consists essentially of, or comprises 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 consecutive nucleotides) complementary or partially complementary to the target domain or a portion thereof, e.g., the targeting domain is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain of these embodiments, the targeting domain is complementary to the target domain over the entire length of the targeting domain, the entire length of the target domain, or both.

In certain embodiments, a unimolecular or chimeric gRNA molecule disclosed herein (comprising a targeting domain, a first complementary domain, a linking domain, a second complementary domain, a proximal domain and, optionally, a tail domain) comprises the amino acid sequence set forth in SEQ ID NO:42, wherein the targeting domain is listed as 20 Ns (residues 1-20) but may range in length from 16 to 26 nucleotides and wherein the final six residues (residues 97-102) represent a termination signal for the U6 promoter but may be absent or fewer in number. In certain embodiments, the unimolecular, or chimeric, gRNA molecule is a *S. pyogenes* gRNA molecule.

In certain embodiments, a unimolecular or chimeric gRNA molecule disclosed herein (comprising a targeting domain, a first complementary domain, a linking domain, a second complementary domain, a proximal domain and, optionally, a tail domain) comprises the amino acid sequence set forth in SEQ ID NO:38, wherein the targeting domain is listed as 20 Ns (residues 1-20) but may range in length from 16 to 26 nucleotides, and wherein the final six residues (residues 97-102) represent a termination signal for the U6 promoter but may be absent or fewer in number. In certain embodiments, the unimolecular or chimeric gRNA molecule is an *S. aureus* gRNA molecule.

The sequences and structures of exemplary chimeric gRNAs are also shown in **Figs. 1H-1I**.

Exemplary modular gRNAs

In certain embodiments, a modular gRNA disclosed herein comprises:

a first strand comprising, preferably from 5' to 3';

a targeting domain, e.g., comprising 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides;

a first complementarity domain; and

a second strand, comprising, preferably from 5' to 3':

optionally a 5' extension domain;

a second complementarity domain;

a proximal domain; and

a tail domain,

wherein:

(a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides;

(b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; or

(c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the sequence from (a), (b), or (c), has at least 60, 75, 80, 85, 90, 95, or 99% homology with the corresponding sequence of a naturally occurring gRNA, or with a gRNA described herein.

In certain embodiments, the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length.

In certain embodiments, the targeting domain consists of, consists essentially of, or comprises 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides (e.g., 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 consecutive nucleotides) complementary to the target domain or a portion thereof. In certain of these embodiments, the targeting domain is complementary to the target domain over the entire length of the targeting domain, the entire length of the target domain, or both.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 16 nucleotides (e.g., 16 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 16 nucleotides in length. In certain embodiments of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 17 nucleotides (e.g., 17 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 17 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 18 nucleotides (e.g., 18 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 18 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50,

51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 19 nucleotides (e.g., 19 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 19 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 20 nucleotides (e.g., 20 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 20 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 21 nucleotides (e.g., 21 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 21 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 22 nucleotides (e.g., 22 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 22 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15,

18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 23 nucleotides (e.g., 23 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 23 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 24 nucleotides (e.g., 24 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 24 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 25 nucleotides (e.g., 25 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 25 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, consists of, or consists essentially of 26 nucleotides (e.g., 26 consecutive nucleotides) having complementarity with the target domain, e.g., the targeting domain is 26 nucleotides in length. In certain of these embodiments, (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides; (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; and/or there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

Methods for designing gRNAs

Methods for designing gRNAs are described herein, including methods for selecting, designing, and validating targeting domains for use in the gRNAs described herein. Exemplary targeting domains for incorporation into gRNAs are also provided herein. It is contemplated herein that in certain embodiments the targeting domain hybridizes to the target domain through complementary base pairing.

Methods for selection and validation of target sequences as well as off-target analyses have been described previously (see, e.g., Mali 2013; Hsu 2013; Fu 2014; Heigwer 2014; Bae 2014; Xiao 2014). For example, a software tool can be used to optimize the choice of potential targeting domains corresponding to a user's target sequence, e.g., to minimize total off-target activity across the genome. Off-target activity may be other than cleavage. For each possible targeting domain choice using *S. pyogenes* Cas9, the tool can identify all off-target sequences (preceding either NAG or NGG PAMs) across the genome that contain up to certain number (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of mismatched base-pairs. The cleavage efficiency at each off-target sequence can be predicted, e.g., using an experimentally-derived weighting scheme. Each possible targeting domain is then ranked according to its total predicted off-target cleavage; the top-ranked targeting domains represent those that are likely to have the greatest on-target cleavage and the least off-target cleavage. Other functions, e.g., automated reagent design for CRISPR construction, primer design for the on-target Surveyor assay, and primer design for high-throughput detection and quantification of off-target cleavage via next-gen sequencing, can also be included in the tool. Candidate targeting domains and gRNAs comprising those targeting domains can be functionally evaluated using methods known in the art and/or as set forth herein.

As a non-limiting example, targeting domains for use in gRNAs for use with *S. pyogenes* and *S. aureus* Cas9s were identified using a DNA sequence searching algorithm. gRNA design was carried out using custom gRNA design software based on the public tool cas-offinder (Bae 2014). This software scores guides after calculating their genome-wide off-target propensity. Typically matches ranging from perfect matches to 7 mismatches are considered for guides ranging in length from 17 to 24. Once the off-target sites are computationally determined, an aggregate score is calculated for each guide and summarized in a tabular output using a web-interface. In addition to identifying potential target sites adjacent to PAM sequences, the software also identifies all PAM adjacent sequences that differ by 1, 2, 3, or more than 3 nucleotides from the selected target sites. Genomic DNA sequences for each gene were obtained from the UCSC Genome browser, and sequences were screened for repeat elements using the publically available RepeaT$_m$asker program. RepeaT$_m$asker searches input DNA sequences for repeated elements and regions of low complexity. The output is a detailed annotation of the repeats present in a given query sequence.

Following identification, targeting domain were ranked into tiers based on their orthogonality and the presence of a 5' G (based on identification of close matches in the human genome containing a relevant PAM, e.g., an NGG PAM for *S. pyogenes*, or an NNGRRT (SEQ ID NO:204) or NNGRRV (SEQ ID NO:205) PAM for *S. aureus*). Orthogonality refers to the number of sequences in the human genome that contain a minimum number of mismatches to the target sequence. A "high level of orthogonality" or "good orthogonality" may, for example, refer to 20-mer targeting domain that have no identical sequences in the human genome besides the intended target, nor any sequences that contain one or two mismatches in the target sequence. Targeting domains with good orthogonality are selected to minimize off-target DNA cleavage.

Targeting domains were identified for both single-gRNA nuclease cleavage and for a dual-gRNA paired "nickase" strategy. Criteria for selecting targeting domains and the determination of which targeting domains can be used for the dual-gRNA paired "nickase" strategy is based on two considerations:

(1) Targeting domain pairs should be oriented on the DNA such that PAMs are facing out and cutting with the D10A Cas9 nickase will result in 5' overhangs; and

(2) An assumption that cleaving with dual nickase pairs will result in deletion of the entire intervening sequence at a reasonable frequency. However, cleaving with dual nickase pairs can also result in indel mutations at the site of only one of the gRNAs. Candidate pair members can be tested for how efficiently they remove the entire sequence versus causing indel mutations at the target site of one targeting domain.

Cas9 molecules

Cas9 molecules of a variety of species can be used in the methods and compositions described herein. While *S. pyogenes, S. aureus,* and *S. thermophilus* Cas9 molecules are the subject of much of the disclosure herein, Cas9 molecules of, derived from, or based on the Cas9 proteins of other species listed herein can be used as well. These include, for example, Cas9 molecules from *Acidovorax avenae, Actinobacillus pleuropneumoniae, Actinobacillus succinogenes, Actinobacillus suis, Actinomyces sp., cycliphilus denitrificans, Aminomonas paucivorans, Bacillus cereus, Bacillus smithii, Bacillus thuringiensis, Bacteroides sp., Blastopirellula marina, Bradyrhizobium sp., Brevibacillus laterosporus, Campylobacter coli, Campylobacter jejuni, Campylobacter lari, Candidatus puniceispirillum, Clostridium cellulolyticum, Clostridium perfringens, Corynebacterium accolens, Corynebacterium diphtheria, Corynebacterium matruchotii, Dinoroseobacter shibae, Eubacterium dolichum, gamma proteobacterium, Gluconacetobacter diazotrophicus, Haemophilus parainfluenzae, Haemophilus sputorum, Helicobacter canadensis, Helicobacter cinaedi, Helicobacter mustelae, Ilyobacter polytropus, Kingella kingae, Lactobacillus crispatus, Listeria ivanovii, Listeria monocytogenes, Listeriaceae bacterium, Methylocystis sp., Methylosinus trichosporium, Mobiluncus mulieris, Neisseria bacilliformis, Neisseria cinerea, Neisseria flavescens, Neisseria lactamica, Neisseria sp., Neisseria wadsworthii, Nitrosomonas sp., Parvibaculum lavamentivorans, Pasteurella multocida, Phascolarctobacterium succinatutens, Ralstonia syzygii, Rhodopseudomonas palustris, Rhodovulum sp., Simonsiella muelleri, Sphingomonas sp., Sporolactobacillus vineae, Staphylococcus lugdunensis, Streptococcus sp., Subdoligranulum sp., Tistrella mobilis, Treponema sp.,* or *Verminephrobacter eiseniae.*

Cas9 domains

Crystal structures have been determined for two different naturally occurring bacterial Cas9 molecules (Jinek 2014) and for *S. pyogenes* Cas9 with a guide RNA (e.g., a synthetic fusion of crRNA and tracrRNA) (Nishimasu 2014; Anders 2014).

A naturally occurring Cas9 molecule comprises two lobes: a recognition (REC) lobe and a nuclease (NUC) lobe; each of which further comprise domains described herein. **Figs. 8A-8B** provide a schematic of the organization of important Cas9 domains in the primary structure. The domain nomenclature and the numbering of the amino acid residues encompassed by each domain used throughout this disclosure is as described previously (Nishimasu 2014). The numbering of the amino acid residues is with reference to Cas9 from *S. pyogenes*.

The REC lobe comprises the arginine-rich bridge helix (BH), the REC1 domain, and the REC2 domain. The REC lobe does not share structural similarity with other known proteins, indicating that it is a Cas9-specific functional domain. The BH domain is a long α helix and arginine rich region and comprises amino acids 60-93 of *S. pyogenes* Cas9 (SEQ ID NO:2). The REC1 domain is important for recognition of the repeat:anti-repeat duplex, e.g., of a gRNA or a tracrRNA, and is therefore critical for Cas9 activity by recognizing the target sequence. The REC1 domain comprises two REC1 motifs at amino acids 94 to 179 and 308 to 717 of *S. pyogenes* Cas9 (SEQ ID NO:2). These two REC1 domains, though separated by the REC2 domain in the linear primary structure, assemble in the tertiary structure to form the REC1 domain. The REC2 domain, or parts thereof, may also play a role in the recognition of the repeat:anti-repeat duplex. The REC2 domain comprises amino acids 180-307 of *S. pyogenes* Cas9 (SEQ ID NO:2).

The NUC lobe comprises the RuvC domain, the HNH domain, and the PAM-interacting (PI) domain. The RuvC domain shares structural similarity to retroviral integrase superfamily members and cleaves a single strand, e.g., the non-complementary strand of the target nucleic acid molecule. The RuvC domain is assembled from the three split RuvC motifs (RuvCI, RuvCII, and RuvCIII, which are often commonly referred to in the art as RuvCI domain or N-terminal RuvC domain, RuvCII domain, and RuvCIII domain, respectively) at amino acids 1-59, 718-769, and 909-1098, respectively, of *S. pyogenes* Cas9 (SEQ ID NO:2). Similar to the REC1 domain, the three RuvC motifs are linearly separated by other domains in the primary structure. However, in the tertiary structure, the three RuvC motifs assemble and form the RuvC domain. The HNH domain shares structural similarity

with HNH endonucleases and cleaves a single strand, e.g., the complementary strand of the target nucleic acid molecule. The HNH domain lies between the RuvC II-III motifs and comprises amino acids 775-908 of *S. pyogenes* Cas9 (SEQ ID NO:2). The PI domain interacts with the PAM of the target nucleic acid molecule, and comprises amino acids 1099-1368 of *S. pyogenes* Cas9 (SEQ ID NO:2).

RuvC-like domain and HNH-like domain

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain and a RuvC-like domain, and in certain of these embodiments cleavage activity is dependent on the RuvC-like domain and the HNH-like domain. A Cas9 molecule or Cas9 polypeptide can comprise one or more of a RuvC-like domain and an HNH-like domain. In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises a RuvC-like domain, e.g., a RuvC-like domain described below, and/or an HNH-like domain, e.g., an HNH-like domain described below.

RuvC-like domains

In certain embodiments, a RuvC-like domain cleaves a single strand, e.g., the non-complementary strand of the target nucleic acid molecule. The Cas9 molecule or Cas9 polypeptide can include more than one RuvC-like domain (e.g., one, two, three or more RuvC-like domains). In certain embodiments, a RuvC-like domain is at least 5, 6, 7, 8 amino acids in length but not more than 20, 19, 18, 17, 16 or 15 amino acids in length. In certain embodiments, the Cas9 molecule or Cas9 polypeptide comprises an N-terminal RuvC-like domain of about 10 to 20 amino acids, e.g., about 15 amino acids in length.

N-terminal RuvC-like domains

Some naturally occurring Cas9 molecules comprise more than one RuvC-like domain with cleavage being dependent on the N-terminal RuvC-like domain. Accordingly, a Cas9 molecule or Cas9 polypeptide can comprise an N-terminal RuvC-like domain. Exemplary N-terminal RuvC-like domains are described below.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an N-terminal RuvC-like domain comprising an amino acid sequence of Formula I:

$D\text{-}X_1\text{-}G\text{-}X_2\text{-}X_3\text{-}X_4\text{-}X_5\text{-}G\text{-}X_6\text{-}X_7\text{-}X_8\text{-}X_9$ (SEQ ID NO:20),

wherein

X_1 is selected from I, V, M, L, and T (e.g., selected from I, V, and L);

X_2 is selected from T, I, V, S, N, Y, E, and L (e.g., selected from T, V, and I);

X_3 is selected from N, S, G, A, D, T, R, M, and F (e.g., A or N);

X_4 is selected from S, Y, N, and F (e.g., S);

X_5 is selected from V, I, L, C, T, and F (e.g., selected from V, I and L);

X_6 is selected from W, F, V, Y, S, and L (e.g., W);

X_7 is selected from A, S, C, V, and G (e.g., selected from A and S);

X_8 is selected from V, I, L, A, M, and H (e.g., selected from V, I, M and L); and

X_9 is selected from any amino acid or is absent (e.g., selected from T, V, I, L, Δ, F, S, A, Y, M, and R, or, e.g., selected from T, V, I, L, and Δ).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:20 by as many as 1 but no more than 2, 3, 4, or 5 residues.

In certain embodiments, the N-terminal RuvC-like domain is cleavage competent. In other embodiments, the N-terminal RuvC-like domain is cleavage incompetent.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an N-terminal RuvC-like domain comprising an amino acid sequence of Formula II:

D-X_1-G-X_2-X_3-S-X_5-G-X_6-X_7-X_8-X_9, (SEQ ID NO:21),

wherein

X_1 is selected from I, V, M, L, and T (e.g., selected from I, V, and L);

X_2 is selected from T, I, V, S, N, Y, E, and L (e.g., selected from T, V, and I);

X_3 is selected from N, S, G, A, D, T, R, M and F (e.g., A or N);

X_5 is selected from V, I, L, C, T, and F (e.g., selected from V, I and L);

X_6 is selected from W, F, V, Y, S, and L (e.g., W);

X_7 is selected from A, S, C, V, and G (e.g., selected from A and S);

X_8 is selected from V, I, L, A, M, and H (e.g., selected from V, I, M and L); and

X_9 is selected from any amino acid or is absent (e.g., selected from T, V, I, L, Δ, F, S, A, Y, M, and R or selected from e.g., T, V, I, L, and Δ).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:21 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiments, the N-terminal RuvC-like domain comprises an amino acid sequence of Formula III:

D-I-G-X_2-X_3-S-V-G-W-A-X_8-X_9 (SEQ ID NO:22),

wherein

X_2 is selected from T, I, V, S, N, Y, E, and L (e.g., selected from T, V, and I);

X_3 is selected from N, S, G, A, D, T, R, M, and F (e.g., A or N);

X_8 is selected from V, I, L, A, M, and H (e.g., selected from V, I, M and L); and

X_9 is selected from any amino acid or is absent (e.g., selected from T, V, I, L, Δ, F, S, A, Y, M, and R or selected from e.g., T, V, I, L, and Δ).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:22 by as many as 1 but not more than, 2, 3, 4, or 5 residues.

In certain embodiments, the N-terminal RuvC-like domain comprises an amino acid sequence of Formula IV:

D-I-G-T-N-S-V-G-W-A-V-X (SEQ ID NO:23),

wherein

X is a non-polar alkyl amino acid or a hydroxyl amino acid, e.g., X is selected from V, I, L, and T (e.g., the Cas9 molecule can comprise an N-terminal RuvC-like domain shown in **Figs. 2A-2G** (depicted as Y)).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:23 by as many as 1 but not more than, 2, 3, 4, or 5 residues.

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of an N-terminal RuvC like domain disclosed herein, e.g., in **Figs. 3A-3B**, as many as 1 but no more than 2, 3, 4, or 5 residues. In an embodiment, 1, 2, 3 or all of the highly conserved residues identified in **Figs. 3A-3B** are present.

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of an N-terminal RuvC-like domain disclosed herein, e.g., in **Figs. 4A-4B**, as many as 1 but no more than 2, 3, 4, or 5 residues. In an embodiment, 1, 2, or all of the highly conserved residues identified in **Figs. 4A-4B** are present.

Additional RuvC-like domains

In addition to the N-terminal RuvC-like domain, the Cas9 molecule or Cas9 polypeptide can comprise one or more additional RuvC-like domains. In certain embodiments, the Cas9 molecule or Cas9 polypeptide can comprise two additional RuvC-like domains. Preferably, the additional RuvC-like domain is at least 5 amino acids in length and, e.g., less than 15 amino acids in length, e.g., 5 to 10 amino acids in length, e.g., 8 amino acids in length.

An additional RuvC-like domain can comprise an amino acid sequence of Formula V:

I-X_1-X_2-E-X_3-A-R-E (SEQ ID NO:15),

wherein

X_1 is V or H;

X_2 is I, L or V (e.g., I or V); and

X_3 is M or T.

In certain embodiments, the additional RuvC-like domain comprises an amino acid sequence of Formula VI:

I-V-X_2-E-M-A-R-E (SEQ ID NO:16),

wherein

X_2 is I, L or V (e.g., I or V) (e.g., the Cas9 molecule or Cas9 polypeptide can comprise an additional RuvC-like domain shown in **Fig. 2A-2G** (depicted as B)).

An additional RuvC-like domain can comprise an amino acid sequence of Formula VII:

H-H-A-X_1-D-A-X_2-X_3 (SEQ ID NO:17),

wherein

X_1 is H or L;

X_2 is R or V; and

X_3 is E or V.

In certain embodiments, the additional RuvC-like domain comprises the amino acid sequence: H-H-A-H-D-A-Y-L (SEQ ID NO:18).

In certain embodiments, the additional RuvC-like domain differs from a sequence of SEQ ID NOs:15-18 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiments, the sequence flanking the N-terminal RuvC-like domain has the amino acid sequence of Formula VIII:

K-X_1'-Y-X_2'-X_3'-X_4'-Z-T-D-X_9'-Y (SEQ ID NO:19),

wherein

X_1' is selected from K and P;

X_2' is selected from V, L, I, and F (e.g., V, I and L);

X_3' is selected from G, A and S (e.g., G);

X_4' is selected from L, I, V, and F (e.g., L);

X_9' is selected from D, E, N, and Q; and

Z is an N-terminal RuvC-like domain, e.g., as described above.

HNH-like domains

In certain embodiments, an HNH-like domain cleaves a single stranded complementary domain, e.g., a complementary strand of a double stranded nucleic acid molecule. In certain embodiments, an HNH-like domain is at least 15, 20, or 25 amino acids in length but not more than 40, 35, or 30 amino acids in length, e.g., 20 to 35 amino acids in length, e.g., 25 to 30 amino acids in length. Exemplary HNH-like domains are described below.

In an embodiment, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain having an amino acid sequence of Formula IX:

X_1-X_2-X_3-H-X_4-X_5-P-X_6-X_7-X_8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-N-X^{16}-X^{17}-X^{18}-X^{19}-X_{20}-X_{21}-X_{22}-X_{23}-N (SEQ ID NO:25),

wherein

X_1 is selected from D, E, Q, and N (e.g., D and E);

X^2 is selected from L, I, R, Q, V, M, and K;

X_3 is selected from D and E;

X_4 is selected from I, V, T, A, and L (e.g., A, I, and V);

X_5 is selected from V, Y, I, L, F, and W (e.g., V, I, and L);

X_6 is selected from Q, H, R, K, Y, I, L, F, and W;

X_7 is selected from S, A, D, T, and K (e.g., S and A);

X_8 is selected from F, L, V, K, Y, M, I, R, A, E, D, and Q (e.g., F);

X_9 is selected from L, R, T, I, V, S, C, Y, K, F, and G;

X_{10} is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

X_{11} is selected from D, S, N, R, L, and T (e.g., D);

X_{12} is selected from D, N and S;

X_{13} is selected from S, A, T, G, and R (e.g., S);

X_{14} is selected from I, L, F, S, R, Y, Q, W, D, K, and H (e.g., I, L, and F);

X_{15} is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y, and V;

X_{16} is selected from K, L, R, M, T, and F (e.g., L, R, and K);

X_{17} is selected from V, L, I, A, and T;

X_{18} is selected from L, I, V, and A (e.g., L and I);

X_{19} is selected from T, V, C, E, S, and A (e.g., T and V);

X_{20} is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H, and A;

X_{21} is selected from S, P, R, K, N, A, H, Q, G, and L;

X_{22} is selected from D, G, T, N, S, K, A, I, E, L, Q, R, and Y; and

X_{23} is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D, and F.

In certain embodiments, a HNH-like domain differs from a sequence of SEQ ID NO:25 by at least one but not more than, 2, 3, 4, or 5 residues.

In certain embodiments, the HNH-like domain is cleavage competent. In other embodiments, the HNH-like domain is cleavage incompetent.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain comprising an amino acid sequence of Formula X:

X_1-X_2-X_3-H-X_4-X_5-P-X_6-S-X_8-X_9-X_{10}-D-D-S-X_{14}-X_{15}-N-K-V-L-X_{19}-X_{20}-X_{21}-X_{22}-X_{23}-N (SEQ ID NO:26),

wherein

X_1 is selected from D and E;

X_2 is selected from L, I, R, Q, V, M, and K;

X_3 is selected from D and E;

X_4 is selected from I, V, T, A, and L (e.g., A, I, and V);

X_5 is selected from V, Y, I, L, F, and W (e.g., V, I, and L);

X_6 is selected from Q, H, R, K, Y, I, L, F, and W;

X_8 is selected from F, L, V, K, Y, M, I, R, A, E, D, and Q (e.g., F);

X_9 is selected from L, R, T, I, V, S, C, Y, K, F, and G;

X_{10} is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

X_{14} is selected from I, L, F, S, R, Y, Q, W, D, K, and H (e.g., I, L, and F);

X_{15} is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y, and V;

X_{19} is selected from T, V, C, E, S, and A (e.g., T and V);

X_{20} is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H, and A;

X_{21} is selected from S, P, R, K, N, A, H, Q, G, and L;

X_{22} is selected from D, G, T, N, S, K, A, I, E, L, Q, R, and Y; and

X_{23} is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D, and F.

In certain embodiment, the HNH-like domain differs from a sequence of SEQ ID NO:26 by 1, 2, 3, 4, or 5 residues.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain comprising an amino acid sequence of Formula XI:

X_1-V-X_3-H-I-V-P-X_6-S-X_8-X_9-X_{10}-D-D-S-X_{14}-X_{15}-N-K-V-L-T-X_{20}-X_{21}-X_{22}-X_{23}-N (SEQ ID NO:27),

wherein

X_1 is selected from D and E;

X_3 is selected from D and E;

X_6 is selected from Q, H, R, K, Y, I, L, and W;

X_8 is selected from F, L, V, K, Y, M, I, R, A, E, D, and Q (e.g., F);

X_9 is selected from L, R, T, I, V, S, C, Y, K, F, and G;

X_{10} is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

X_{14} is selected from I, L, F, S, R, Y, Q, W, D, K, and H (e.g., I, L, and F);

X_{15} is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y, and V;

X_{20} is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H, and A;

X_{21} is selected from S, P, R, K, N, A, H, Q, G, and L;

X_{22} is selected from D, G, T, N, S, K, A, I, E, L, Q, R, and Y; and

X_{23} is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D, and F.

In certain embodiments, the HNH-like domain differs from a sequence of SEQ ID NO:27 by 1, 2, 3, 4, or 5 residues.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain having an amino acid sequence of Formula XII:

D-X_2-D-H-I-X_5-P-Q-X_7-F-X_9-X_{10}-D-X_{12}-S-I-D-N-X_{16}-V-L-X_{19}-X_{20}-S-X_{22}-X_{23}-N (SEQ ID NO:28),

wherein

X_2 is selected from I and V;

X_5 is selected from I and V;

X_7 is selected from A and S;

X_9 is selected from I and L;

X_{10} is selected from K and T;

X_{12} is selected from D and N;

X_{16} is selected from R, K, and L;

X_{19} is selected from T and V;

X_{20} is selected from S, and R;

X_{22} is selected from K, D, and A; and

X_{23} is selected from E, K, G, and N (e.g., the Cas9 molecule or Cas9 polypeptide can comprise an HNH-like domain as described herein).

In an embodiment, the HNH-like domain differs from a sequence of SEQ ID NO:28 by as many as 1 but no more than 2, 3, 4, or 5 residues.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises the amino acid sequence of Formula XIII:

L-Y-Y-L-Q-N-G-X_1'-D-M-Y-X_2'-X_3'-X_4'-X_5'-L-D-I-X_6'-X_7'-L-S-X_8'-Y-Z-N-R-X_9'-K-X_{10}'-D-X_{11}'-V-P (SEQ ID NO:24),

wherein

X_1' is selected from K and R;

X_2' is selected from V and T;

X_3' is selected from G and D;

X_4' is selected from E, Q and D;

X_5' is selected from E and D;

X_6' is selected from D, N, and H;

X_7' is selected from Y, R, and N;

X_8' is selected from Q, D, and N;

X_9' is selected from G and E;

X_{10}' is selected from S and G;

X_{11}' is selected from D and N; and

Z is an HNH-like domain, e.g., as described above.

In certain embodiments, the Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence that differs from a sequence of SEQ ID NO:24 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiments, the HNH-like domain differs from a sequence of an HNH-like domain disclosed herein, e.g., in **Figs. 5A-5C**, by as many as 1 but not more than 2, 3, 4, or 5 residues. In certain embodiments, 1 or both of the highly conserved residues identified in **Figs. 5A-5C** are present.

In certain embodiments, the HNH -like domain differs from a sequence of an HNH-like domain disclosed herein, e.g., in **Figs. 6A-6B**, by as many as 1 but not more than 2, 3, 4, or 5 residues. In an embodiment, 1, 2, or all 3 of the highly conserved residues identified in **Figs. 6A-6B** are present.

Cas9 Activities

In certain embodiments, the Cas9 molecule or Cas9 polypeptide is capable of cleaving a target nucleic acid molecule. Typically, wild-type Cas9 molecules cleave both strands of a target nucleic acid molecule. Cas9 molecules and Cas9 polypeptides can be engineered to alter nuclease cleavage (or other properties), e.g., to provide a Cas9 molecule or Cas9 polypeptide which is a nickase, or which lacks the ability to cleave target nucleic acid. A Cas9 molecule or Cas9 polypeptide that is capable of cleaving a target nucleic acid molecule is referred to herein as an eaCas9 (an enzymatically active Cas9) molecule or eaCas9 polypeptide.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide comprises one or more of the following enzymatic activities:

(1) nickase activity, i.e., the ability to cleave a single strand, e.g., the non-complementary strand or the complementary strand, of a nucleic acid molecule;

(2) double stranded nuclease activity, i.e., the ability to cleave both strands of a double stranded nucleic acid and create a double stranded break, which in an embodiment is the presence of two nickase activities;

(3) endonuclease activity;

(4) exonuclease activity; and

(5) helicase activity, i.e., the ability to unwind the helical structure of a double stranded nucleic acid.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide cleaves both DNA strands and results in a double stranded break. In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide cleaves only one strand, e.g., the strand to which the gRNA hybridizes to, or the strand complementary to the strand the gRNA hybridizes with. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with a RuvC domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH domain and cleavage activity associated with a RuvC domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an active, or cleavage competent, HNH domain and an inactive, or cleavage incompetent, RuvC domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an inactive, or cleavage incompetent, HNH domain and an active, or cleavage competent, RuvC domain.

Some Cas9 molecules or Cas9 polypeptides have the ability to interact with a gRNA molecule, and in conjunction with the gRNA molecule to localize to a core target domain, but are incapable of cleaving the target nucleic acid or of cleaving at efficient rates. Cas9 molecules having no, or no substantial, cleavage activity are referred to herein as eiCas9 molecules or eiCas9 polypeptides. For example, an eiCas9 molecule or eiCas9 polypeptide can lack cleavage activity or have substantially less, e.g., less than 20%, 10%, 5%, 1%, or 0.1 % of the cleavage activity of a reference Cas9 molecule or eiCas9 polypeptide, as measured by an assay described herein.

Targeting and PAMs

A Cas9 molecule or Cas9 polypeptide can interact with a gRNA molecule and, in concert with the gRNA molecule, localize to a site which comprises a target domain, and in certain embodiments, a PAM sequence.

In certain embodiments, the ability of an eaCas9 molecule or eaCas9 polypeptide to interact with and cleave a target nucleic acid is PAM sequence dependent. A PAM sequence is a sequence in the target nucleic acid. In an embodiment, cleavage of the target nucleic acid occurs upstream from the PAM sequence. EaCas9 molecules from different bacterial species can recognize different sequence motifs (e.g., PAM sequences). In an embodiment, an eaCas9 molecule of *S. pyogenes* recognizes the sequence motif NGG and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence (see, e.g., Mali 2013). In an embodiment, an eaCas9 molecule of *S. thermophilus* recognizes the sequence motif NGGNG (SEQ ID NO:199) and/or NNAGAAW (W = A or T) (SEQ ID NO:200) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from these sequences (see, e.g., Horvath 2010; Deveau 2008). In an embodiment, an eaCas9 molecule of *S. mutans* recognizes the sequence motif NGG and/or NAAR (R = A or G) (SEQ ID NO:201) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5 bp, upstream from this sequence (see, e.g., Deveau 2008). In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R = A or G) (SEQ ID NO:202) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R = A or G) (SEQ ID NO:203) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif

NNGRRT (R = A or G) (SEQ ID NO:204) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. In an embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRV (R = A or G, V = A, G, or C) (SEQ ID NO:205) and directs cleavage of a target nucleic acid sequence 1 to 10, e.g., 3 to 5, bp upstream from that sequence. The ability of a Cas9 molecule to recognize a PAM sequence can be determined, e.g., using a transformation assay as described previously (Jinek 2012). In each of the aforementioned embodiments (i.e., SEQ ID NOs:199-205), N can be any nucleotide residue, e.g., any of A, G, C, or T.

As is discussed herein, Cas9 molecules can be engineered to alter the PAM specificity of the Cas9 molecule.

Exemplary naturally occurring Cas9 molecules have been described previously (see, e.g., Chylinski 2013). Such Cas9 molecules include Cas9 molecules of a cluster 1 bacterial family, cluster 2 bacterial family, cluster 3 bacterial family, cluster 4 bacterial family, cluster 5 bacterial family, cluster 6 bacterial family, a cluster 7 bacterial family, a cluster 8 bacterial family, a cluster 9 bacterial family, a cluster 10 bacterial family, a cluster 11 bacterial family, a cluster 12 bacterial family, a cluster 13 bacterial family, a cluster 14 bacterial family, a cluster 15 bacterial family, a cluster 16 bacterial family, a cluster 17 bacterial family, a cluster 18 bacterial family, a cluster 19 bacterial family, a cluster 20 bacterial family, a cluster 21 bacterial family, a cluster 22 bacterial family, a cluster 23 bacterial family, a cluster 24 bacterial family, a cluster 25 bacterial family, a cluster 26 bacterial family, a cluster 27 bacterial family, a cluster 28 bacterial family, a cluster 29 bacterial family, a cluster 30 bacterial family, a cluster 31 bacterial family, a cluster 32 bacterial family, a cluster 33 bacterial family, a cluster 34 bacterial family, a cluster 35 bacterial family, a cluster 36 bacterial family, a cluster 37 bacterial family, a cluster 38 bacterial family, a cluster 39 bacterial family, a cluster 40 bacterial family, a cluster 41 bacterial family, a cluster 42 bacterial family, a cluster 43 bacterial family, a cluster 44 bacterial family, a cluster 45 bacterial family, a cluster 46 bacterial family, a cluster 47 bacterial family, a cluster 48 bacterial family, a cluster 49 bacterial family, a cluster 50 bacterial family, a cluster 51 bacterial family, a cluster 52 bacterial family, a cluster 53 bacterial family, a cluster 54 bacterial family, a cluster 55 bacterial family, a cluster 56 bacterial family, a cluster 57 bacterial family, a cluster 58 bacterial family, a cluster 59 bacterial family, a cluster 60 bacterial family, a cluster 61 bacterial family, a cluster 62 bacterial family, a cluster 63 bacterial family, a cluster 64 bacterial family, a cluster 65 bacterial family, a

cluster 66 bacterial family, a cluster 67 bacterial family, a cluster 68 bacterial family, a cluster 69 bacterial family, a cluster 70 bacterial family, a cluster 71 bacterial family, a cluster 72 bacterial family, a cluster 73 bacterial family, a cluster 74 bacterial family, a cluster 75 bacterial family, a cluster 76 bacterial family, a cluster 77 bacterial family, or a cluster 78 bacterial family.

Exemplary naturally occurring Cas9 molecules include a Cas9 molecule of a cluster 1 bacterial family. Examples include a Cas9 molecule of: *S. aureus*, *S. pyogenes* (e.g., strains SF370, MGAS10270, MGAS10750, MGAS2096, MGAS315, MGAS5005, MGAS6180, MGAS9429, NZ131, SSI-1), *S. thermophilus* (e.g., strain LMD-9), *S. pseudoporcinus* (e.g., strain SPIN 20026), *S. mutans* (e.g., strains UA159, NN2025), *S. macacae* (e.g., strain NCTC11558), *S. gallolyticus* (e.g., strains UCN34, ATCC BAA-2069), *S. equines* (e.g., strains ATCC 9812, MGCS 124), *S. dysdalactiae* (e.g., strain GGS 124), *S. bovis* (e.g., strain ATCC 700338), *S. anginosus* (e.g., strain F0211), *S. agalactiae* (e.g., strains NEM316, A909), *Listeria monocytogenes* (e.g., strain F6854), *Listeria innocua* (*L. innocua*, e.g., strain Clip11262), *Enterococcus italicus* (e.g., strain DSM 15952), or *Enterococcus faecium* (e.g., strain 1,231,408).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence:

having 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with;

differs at no more than, 2, 5, 10, 15, 20, 30, or 40% of the amino acid residues when compared with;

differs by at least 1, 2, 5, 10 or 20 amino acids, but by no more than 100, 80, 70, 60, 50, 40, or 30 amino acids from; or

identical to any Cas9 molecule sequence described herein, or to a naturally occurring Cas9 molecule sequence, e.g., a Cas9 molecule from a species listed herein (e.g., SEQ ID NOs:1, 2, 4-6, or 12) or described in Chylinski 2013. In an embodiment, the Cas9 molecule or Cas9 polypeptide comprises one or more of the following activities: a nickase activity; a double stranded cleavage activity (e.g., an endonuclease and/or exonuclease activity); a helicase activity; or the ability, together with a gRNA molecule, to localize to a target nucleic acid.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises any of the amino acid sequence of the consensus sequence of **Figs. 2A-2G**, wherein "*" indicates any

amino acid found in the corresponding position in the amino acid sequence of a Cas9 molecule of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua*, and "-" indicates absent. In an embodiment, a Cas9 molecule or Cas9 polypeptide differs from the sequence of the consensus sequence disclosed in **Figs. 2A-2G** by at least 1, but no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues. In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises the amino acid sequence of SEQ ID NO:2. In other embodiments, a Cas9 molecule or Cas9 polypeptide differs from the sequence of SEQ ID NO:2 by at least 1, but no more than 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid residues.

A comparison of the sequence of a number of Cas9 molecules indicate that certain regions are conserved. These are identified below as:

region 1 (residues 1 to 180, or in the case of region 1' residues 120 to 180)

region 2 (residues 360 to 480);

region 3 (residues 660 to 720);

region 4 (residues 817 to 900); and

region 5 (residues 900 to 960).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises regions 1-5, together with sufficient additional Cas9 molecule sequence to provide a biologically active molecule, e.g., a Cas9 molecule having at least one activity described herein. In certain embodiments, regions 1-5 each independently have 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with the corresponding residues of a Cas9 molecule or Cas9 polypeptide described herein, e.g., a sequence from **Figs. 2A-2G** (SEQ ID NOs:1, 2, 4, 5, 14).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 1:

having 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 1-180 (the numbering is according to the motif sequence in **Fig. 2**; 52% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes* (SEQ ID NO:2);

differs by at least 1, 2, 5, 10 or 20 amino acids but by no more than 90, 80, 70, 60, 50, 40, or 30 amino acids from amino acids 1-180 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *Listeria innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 1-180 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 1':

having 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 120-180 (55% of residues in the four Cas9 sequences in **Fig. 2** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively);

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20, or 10 amino acids from amino acids 120-180 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 120-180 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 2:

having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 360-480 (52% of residues in the four Cas9 sequences in **Fig. 2** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively);

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20, or 10 amino acids from amino acids 360-480 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 360-480 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 3:

having 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 660-720 (56% of residues in the four Cas9 sequences in **Fig. 2** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively);

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20, or 10 amino acids from amino acids 660-720 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 660-720 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 4:

having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 817-900 (55% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively);

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20, or 10 amino acids from amino acids 817-900 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 817-900 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 5:

having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 900-960 (60% of residues in the four Cas9 sequences in **Figs. 2A-2G** are conserved) of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively);

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20, or 10 amino acids from amino acids 900-960 of the amino acid sequence of Cas9 of *S. pyogenes, S. thermophilus, S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively); or

is identical to amino acids 900-960 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua* (SEQ ID NOs:2, 4, 1, and 5, respectively).

Engineered or altered Cas9

Cas9 molecules and Cas9 polypeptides described herein can possess any of a number of properties, including nuclease activity (e.g., endonuclease and/or exonuclease activity); helicase activity; the ability to associate functionally with a gRNA molecule; and the ability to target (or localize to) a site on a nucleic acid (e.g., PAM recognition and specificity). In certain embodiments, a Cas9 molecule or Cas9 polypeptide can include all or a subset of these properties. In a typical embodiment, a Cas9 molecule or Cas9 polypeptide has the ability to interact with a gRNA molecule and, in concert with the gRNA molecule, localize to a site in a nucleic acid. Other activities, e.g., PAM specificity, cleavage activity, or helicase activity can vary more widely in Cas9 molecules and Cas9 polypeptides.

Cas9 molecules include engineered Cas9 molecules and engineered Cas9 polypeptides (engineered, as used in this context, means merely that the Cas9 molecule or Cas9 polypeptide differs from a reference sequences, and implies no process or origin limitation). An engineered Cas9 molecule or Cas9 polypeptide can comprise altered enzymatic properties, e.g., altered nuclease activity (as compared with a naturally occurring or other reference Cas9 molecule) or altered helicase activity. As discussed herein, an engineered Cas9 molecule or Cas9 polypeptide can have nickase activity (as opposed to double strand nuclease activity). In certain embodiments, an engineered Cas9 molecule or Cas9 polypeptide can have an alteration that alters its size, e.g., a deletion of amino acid sequence that reduces its size, e.g., without significant effect on one or more Cas9 activities. In certain embodiments, an engineered Cas9 molecule or Cas9 polypeptide can comprise an alteration that affects PAM recognition, e.g., an engineered Cas9 molecule can be altered to recognize a PAM sequence other than that recognized by the endogenous wild-type PI domain. In certain embodiments, a Cas9 molecule or Cas9 polypeptide can differ in sequence from a naturally occurring Cas9 molecule but not have significant alteration in one or more Cas9 activities.

Cas9 molecules or Cas9 polypeptides with desired properties can be made in a number of ways, e.g., by alteration of a parental, e.g., naturally occurring, Cas9 molecules or Cas9 polypeptides, to provide an altered Cas9 molecule or Cas9 polypeptide having a desired property. For example, one or more mutations or differences relative to a parental Cas9 molecule, e.g., a naturally occurring or engineered Cas9 molecule, can be introduced. Such mutations and differences comprise: substitutions (e.g., conservative substitutions or substitutions of non-essential amino acids); insertions; or deletions. In an embodiment, a

Cas9 molecule or Cas9 polypeptide can comprises one or more mutations or differences, e.g., at least 1, 2, 3, 4, 5, 10, 15, 20, 30, 40, or 50 mutations but less than 200, 100, or 80 mutations relative to a reference, e.g., a parental, Cas9 molecule.

In certain embodiments, a mutation or mutations do not have a substantial effect on a Cas9 activity, e.g., a Cas9 activity described herein. In other embodiments, a mutation or mutations have a substantial effect on a Cas9 activity, e.g., a Cas9 activity described herein.

Non-cleaving and modified-cleavage Cas9

In an embodiment, a Cas9 molecule or Cas9 polypeptide comprises a cleavage property that differs from naturally occurring Cas9 molecules, e.g., that differs from the naturally occurring Cas9 molecule having the closest homology. For example, a Cas9 molecule or Cas9 polypeptide can differ from naturally occurring Cas9 molecules, e.g., a Cas9 molecule of *S. pyogenes*, as follows: its ability to modulate, e.g., decreased or increased, cleavage of a double stranded nucleic acid (endonuclease and/or exonuclease activity), e.g., as compared to a naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S. pyogenes)*; its ability to modulate, e.g., decreased or increased, cleavage of a single strand of a nucleic acid, e.g., a non-complementary strand of a nucleic acid molecule or a complementary strand of a nucleic acid molecule (nickase activity), e.g., as compared to a naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S. pyogenes)*; or the ability to cleave a nucleic acid molecule, e.g., a double stranded or single stranded nucleic acid molecule, can be eliminated.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide comprises one or more of the following activities: cleavage activity associated with an N-terminal RuvC-like domain; cleavage activity associated with an HNH-like domain; cleavage activity associated with an HNH-like domain and cleavage activity associated with an N-terminal RuvC-like domain.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide comprises an active, or cleavage competent, HNH-like domain (e.g., an HNH-like domain described herein, e.g., SEQ ID NOs:24-28) and an inactive, or cleavage incompetent, N-terminal RuvC-like domain. An exemplary inactive, or cleavage incompetent N-terminal RuvC-like domain can have a mutation of an aspartic acid in an N-terminal RuvC-like domain, e.g., an aspartic acid at position 9 of the consensus sequence disclosed in **Figs. 2A-2G** or an aspartic acid at position 10 of SEQ ID NO:2, e.g., can be substituted with an alanine. In an embodiment, the eaCas9 molecule or eaCas9 polypeptide differs from wild-type in the N-terminal RuvC-like

domain and does not cleave the target nucleic acid, or cleaves with significantly less efficiency, e.g., less than 20, 10, 5, 1, or 0.1% of the cleavage activity of a reference Cas9 molecule, e.g., as measured by an assay described herein. The reference Cas9 molecule can by a naturally occurring unmodified Cas9 molecule, e.g., a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, or *S. thermophilus*. In an embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology.

In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an inactive, or cleavage incompetent, HNH domain and an active, or cleavage competent, N-terminal RuvC-like domain (e.g., a RuvC-like domain described herein, e.g., SEQ ID NOs:15-23). Exemplary inactive, or cleavage incompetent HNH-like domains can have a mutation at one or more of: a histidine in an HNH-like domain, e.g., a histidine shown at position 856 of the consensus sequence disclosed in **Figs. 2A-2G**, e.g., can be substituted with an alanine; and one or more asparagines in an HNH-like domain, e.g., an asparagine shown at position 870 of the consensus sequence disclosed in **Figs. 2A-2G** and/or at position 879 of the consensus sequence disclosed in **Figs. 2A-2G**, e.g., can be substituted with an alanine. In an embodiment, the eaCas9 differs from wild-type in the HNH-like domain and does not cleave the target nucleic acid, or cleaves with significantly less efficiency, e.g., less than 20, 10, 5, 1, or 0.1% of the cleavage activity of a reference Cas9 molecule, e.g., as measured by an assay described herein. The reference Cas9 molecule can by a naturally occurring unmodified Cas9 molecule, e.g., a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes* or *S. thermophilus*. In an embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology.

In certain embodiments, exemplary Cas9 activities comprise one or more of PAM specificity, cleavage activity, and helicase activity. A mutation(s) can be present, e.g., in: one or more RuvC domains, e.g., an N-terminal RuvC domain; an HNH domain; a region outside the RuvC domains and the HNH domain. In an embodiment, a mutation(s) is present in a RuvC domain. In an embodiment, a mutation(s) is present in an HNH domain. In an embodiment, mutations are present in both a RuvC domain and an HNH domain.

Exemplary mutations that may be made in the RuvC domain or HNH domain with reference to the *S. pyogenes* sequence include: D10A, E762A, H840A, N854A, N863A, and/or D986A.

In certain embodiments, a Cas9 molecule may be an eiCas9 molecule comprising one or more differences in a RuvC domain and/or HNH domain as compared to a reference Cas9 molecule, wherein the eiCas9 molecule does not cleave a nucleic acid or cleaves with significantly less efficiency than the reference Cas9 molecule, e.g., in a cleavage assay as described herein, e.g., the eiCas9 molecule cuts with 50%, 25%, 10%, or 1% less efficiency than a reference Cas9 molecule, e.g., the corresponding wild-type Cas9 molecule.

Whether or not a particular sequence, e.g., a substitution, may affect one or more activity, such as targeting activity, cleavage activity, etc., can be evaluated or predicted, e.g., by evaluating whether the mutation is conservative. In an embodiment, a "non-essential" amino acid residue, as used in the context of a Cas9 molecule, is a residue that can be altered from the wild-type sequence of a Cas9 molecule, e.g., a naturally occurring Cas9 molecule, e.g., an eaCas9 molecule, without abolishing or more preferably, without substantially altering a Cas9 activity (e.g., cleavage activity), whereas changing an "essential" amino acid residue results in a substantial loss of activity (e.g., cleavage activity).

In an embodiment, a Cas9 molecule comprises a cleavage property that differs from naturally occurring Cas9 molecules, e.g., that differs from the naturally occurring Cas9 molecule having the closest homology. For example, a Cas9 molecule can differ from naturally occurring Cas9 molecules, e.g., a Cas9 molecule of *S aureus*, *S. pyogenes*, or *C. jejuni* as follows: its ability to modulate, e.g., decreased or increased, cleavage of a double stranded break (endonuclease and/or exonuclease activity), e.g., as compared to a naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S aureus*, *S. pyogenes,* or *C. jejuni)*; its ability to modulate, e.g., decreased or increased, cleavage of a single strand of a nucleic acid, e.g., a non-complimentary strand of a nucleic acid molecule or a complementary strand of a nucleic acid molecule (nickase activity), e.g., as compared to a naturally occurring Cas9 molecule (e.g., a Cas9 molecule of *S aureus*, *S. pyogenes,* or *C. jejuni*); or the ability to cleave a nucleic acid molecule, e.g., a double stranded or single stranded nucleic acid molecule, can be eliminated.

In an embodiment, the altered Cas9 molecule is an eaCas9 molecule comprising one or more of the following activities: cleavage activity associated with a RuvC domain; cleavage activity associated with an HNH domain; cleavage activity associated with an HNH domain and cleavage activity associated with a RuvC domain.

In certain embodiments, the altered Cas9 molecule is an eiCas9 molecule which does not cleave a nucleic acid molecule (either a double-stranded or single-stranded nucleic acid

molecule) or cleaves a nucleic acid molecule with significantly less efficiency, e.g., less than 20%, 10%, 5%, 1%, or 0.1% of the cleavage activity of a reference Cas9 molecule, e.g., as measured by an assay described herein. The reference Cas9 molecule can be a naturally occurring unmodified Cas9 molecule, e.g., a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, *S. thermophilus, S. aureus, C. jejuni* or *N. meningitidis*. In certain embodiments, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology. In certain embodiments, the eiCas9 molecule lacks substantial cleavage activity associated with a RuvC domain and/or cleavage activity associated with an HNH domain.

In certain embodiments, the altered Cas9 molecule or Cas9 polypeptide comprises a sequence in which:

the sequence corresponding to the fixed sequence of the consensus sequence disclosed in **Figs. 2A-2G** differs at no more than 1, 2, 3, 4, 5, 10, 15, or 20% of the fixed residues in the consensus sequence disclosed in **Figs. 2A-2G**; and

the sequence corresponding to the residues identified by "*" in the consensus sequence disclosed in **Figs. 2A-2G** differs at no more than 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, or 40% of the "*" residues from the corresponding sequence of naturally occurring Cas9 molecule, e.g., an *S. thermophilus*, *S. mutans*, *S. pyogenes*, or *L. innocua* Cas9 molecule.

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the amino acid sequence of *S. thermophilus* Cas9 disclosed in **Figs. 2A-2G** (SEQ ID NO:4) with one or more amino acids that differ from the sequence of *S. thermophilus* (e.g., substitutions) at one or more residues (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, or 200 amino acid residues) represented by an "*" in the consensus sequence disclosed in **Figs. 2A-2G** (SEQ ID NO:14).

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the amino acid sequence of *S. mutans* Cas9 disclosed in **Figs. 2A-2G** (SEQ ID NO:1) with one or more amino acids that differ from the sequence of *S. mutans* (e.g., substitutions) at one or more residues (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, or 200 amino acid residues) represented by an "*" in the consensus sequence disclosed in **Figs. 2A-2G** (SEQ ID NO:14).

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the amino acid sequence of *S. pyogenes* Cas9 disclosed in **Figs. 2A-2G** (SEQ ID NO:2) with one or more amino acids that differ from the

sequence of *S. pyogenes* (e.g., substitutions) at one or more residues (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, or 200 amino acid residues) represented by an "*" in the consensus sequence disclosed in **Figs. 2A-2G** (SEQ ID NO:14).

In an embodiment, the altered Cas9 molecule or Cas9 polypeptide is an eaCas9 molecule or eaCas9 polypeptide comprising the amino acid sequence of *L. innocua* Cas9 disclosed in **Figs. 2A-2G** (SEQ ID NO:5) with one or more amino acids that differ from the sequence of *L. innocua* (e.g., substitutions) at one or more residues (e.g., 2, 3, 5, 10, 15, 20, 30, 50, 70, 80, 90, 100, or 200 amino acid residues) represented by an "*" in the consensus sequence disclosed in **Figs. 2A-2G** (SEQ ID NO:14).

In certain embodiments, the altered Cas9 molecule or Cas9 polypeptide, e.g., an eaCas9 molecule or eaCas9 polypeptide, can be a fusion, e.g., of two of more different Cas9 molecules, e.g., of two or more naturally occurring Cas9 molecules of different species. For example, a fragment of a naturally occurring Cas9 molecule of one species can be fused to a fragment of a Cas9 molecule of a second species. As an example, a fragment of a Cas9 molecule of *S. pyogenes* comprising an N-terminal RuvC-like domain can be fused to a fragment of Cas9 molecule of a species other than *S. pyogenes* (e.g., *S. thermophilus*) comprising an HNH-like domain.

Cas9 with altered or no PAM recognition

Naturally occurring Cas9 molecules can recognize specific PAM sequences, for example the PAM recognition sequences described above for, e.g., *S. pyogenes*, *S. thermophilus, S. mutans*, and *S. aureus*.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide has the same PAM specificities as a naturally occurring Cas9 molecule. In other embodiments, a Cas9 molecule or Cas9 polypeptide has a PAM specificity not associated with a naturally occurring Cas9 molecule, or a PAM specificity not associated with the naturally occurring Cas9 molecule to which it has the closest sequence homology. For example, a naturally occurring Cas9 molecule can be altered, e.g., to alter PAM recognition, e.g., to alter the PAM sequence that the Cas9 molecule or Cas9 polypeptide recognizes in order to decrease off-target sites and/or improve specificity; or eliminate a PAM recognition requirement. In certain embodiments, a Cas9 molecule or Cas9 polypeptide can be altered, e.g., to increase length of PAM recognition sequence and/or improve Cas9 specificity to high level of identity (e.g., 98%, 99%, or 100% match between gRNA and a PAM sequence), e.g., to decrease off-target sites

and/or increase specificity. In certain embodiments, the length of the PAM recognition sequence is at least 4, 5, 6, 7, 8, 9, 10, or 15 amino acids in length. In an embodiment, the Cas9 specificity requires at least 90%, 95%, 96%, 97%, 98%, 99% or more homology between the gRNA and the PAM sequence. Cas9 molecules or Cas9 polypeptides that recognize different PAM sequences and/or have reduced off-target activity can be generated using directed evolution. Exemplary methods and systems that can be used for directed evolution of Cas9 molecules are described (see, e.g., Esvelt 2011). Candidate Cas9 molecules can be evaluated, e.g., by methods described herein.

Size-optimized Cas9

Engineered Cas9 molecules and engineered Cas9 polypeptides described herein include a Cas9 molecule or Cas9 polypeptide comprising a deletion that reduces the size of the molecule while still retaining desired Cas9 properties, e.g., essentially native conformation, Cas9 nuclease activity, and/or target nucleic acid molecule recognition. Provided herein are Cas9 molecules or Cas9 polypeptides comprising one or more deletions and optionally one or more linkers, wherein a linker is disposed between the amino acid residues that flank the deletion. Methods for identifying suitable deletions in a reference Cas9 molecule, methods for generating Cas9 molecules with a deletion and a linker, and methods for using such Cas9 molecules will be apparent to one of ordinary skill in the art upon review of this document.

A Cas9 molecule, e.g., a *S. aureus*, *S. pyogenes*, or *C. jejuni,* Cas9 molecule, having a deletion is smaller, e.g., has reduced number of amino acids, than the corresponding naturally-occurring Cas9 molecule. The smaller size of the Cas9 molecules allows increased flexibility for delivery methods, and thereby increases utility for genome-editing. A Cas9 molecule can comprise one or more deletions that do not substantially affect or decrease the activity of the resultant Cas9 molecules described herein. Activities that are retained in the Cas9 molecules comprising a deletion as described herein include one or more of the following:

a nickase activity, i.e., the ability to cleave a single strand, e.g., the non-complementary strand or the complementary strand, of a nucleic acid molecule; a double stranded nuclease activity, i.e., the ability to cleave both strands of a double stranded nucleic acid and create a double stranded break, which in an embodiment is the presence of two nickase activities;

an endonuclease activity;

an exonuclease activity;

a helicase activity, i.e., the ability to unwind the helical structure of a double stranded nucleic acid;

and recognition activity of a nucleic acid molecule, e.g., a target nucleic acid or a gRNA.

Activity of the Cas9 molecules described herein can be assessed using the activity assays described herein or in the art.

Identifying regions suitable for deletion

Suitable regions of Cas9 molecules for deletion can be identified by a variety of methods. Naturally-occurring orthologous Cas9 molecules from various bacterial species, e.g., any one of those listed in **Table 1**, can be modeled onto the crystal structure of *S. pyogenes* Cas9 (Nishimasu 2014) to examine the level of conservation across the selected Cas9 orthologs with respect to the three-dimensional conformation of the protein. Less conserved or unconserved regions that are spatially located distant from regions involved in Cas9 activity, e.g., interface with the target nucleic acid molecule and/or gRNA, represent regions or domains are candidates for deletion without substantially affecting or decreasing Cas9 activity.

Nucleic acids encoding Cas9 molecules

Nucleic acids encoding the Cas9 molecules or Cas9 polypeptides, e.g., an eaCas9 molecule or eaCas9 polypeptides are provided herein. Exemplary nucleic acids encoding Cas9 molecules or Cas9 polypeptides have been described previously (see, e.g., Cong 2013; Wang 2013; Mali 2013; Jinek 2012).

In an embodiment, a nucleic acid encoding a Cas9 molecule or Cas9 polypeptide can be a synthetic nucleic acid sequence. For example, the synthetic nucleic acid molecule can be chemically modified, e.g., as described herein. In an embodiment, the Cas9 mRNA has one or more (e.g., all of the following properties: it is capped, polyadenylated, substituted with 5-methylcytidine and/or pseudouridine.

In addition, or alternatively, the synthetic nucleic acid sequence can be codon optimized, e.g., at least one non-common codon or less-common codon has been replaced by a common codon. For example, the synthetic nucleic acid can direct the synthesis of an

optimized messenger mRNA, e.g., optimized for expression in a mammalian expression system, e.g., described herein.

In addition, or alternatively, a nucleic acid encoding a Cas9 molecule or Cas9 polypeptide may comprise a nuclear localization sequence (NLS). Nuclear localization sequences are known in the art.

An exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. pyogenes* is set forth in SEQ ID NO:3. The corresponding amino acid sequence of an *S. pyogenes* Cas9 molecule is set forth in SEQ ID NO:2.

An exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. aureus* is set forth in SEQ ID NO:7. An amino acid sequence of an *S. aureus* Cas9 molecule is set forth in SEQ ID NO:6.

Provided below is an exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. aureus* Cas9.

If any of the above Cas9 sequences are fused with a peptide or polypeptide at the C-terminus, it is understood that the stop codon will be removed.

Other Cas molecules and Cas polypeptides

Various types of Cas molecules or Cas polypeptides can be used to practice the inventions disclosed herein. In some embodiments, Cas molecules of Type II Cas systems are used. In other embodiments, Cas molecules of other Cas systems are used. For example, Type I or Type III Cas molecules may be used. Exemplary Cas molecules (and Cas systems) have been described previously (see, e.g., Haft 2005 and Makarova 2011). Exemplary Cas molecules (and Cas systems) are also shown in **Table 2.**

Functional analyses of candidate molecules

Candidate Cas9 molecules, candidate gRNA molecules, and candidate Cas9 molecule/gRNA molecule complexes can be evaluated by art-known techniques or as described herein. Each technique described herein may be used alone or in combination with one or more techniques to evaluate the candidate molecule. The techniques disclosed herein may be used for a variety of methods including, without limitation, methods of determining the stability of a Cas9 molecule/gRNA molecule complex, methods of determining a condition that promotes a stable Cas9 molecule/gRNA molecule complex, methods of screening for a stable Cas9 molecule/gRNA molecule complex, methods of identifying an

optimal gRNA to form a stable Cas9 molecule/gRNA molecule complex, methods of screening for a Cas9 molecule/gRNA molecule complex for administration to a subject, and methods of selecting a Cas9 molecule/gRNA molecule complex for administration to a subject.

Techniques for measuring thermostability of Cas9/gRNA complexes

The thermostability of Cas9-gRNA ribonucleoprotein (RNP) complexes can be detected by differential scanning fluorimetry (DSF) and other techniques. The thermostability of a protein can increase under favorable conditions such as the addition of a binding RNA molecule, e.g., a gRNA. Thus, information regarding the thermostability of a Cas9/gRNA complex is useful for determining whether the complex is stable.

Differential Scanning Fluorimetry (DSF)

DSF is a technique that may be used to measure the thermostability of a protein. The assay can be applied in a number of ways. Exemplary protocols include, but are not limited to, a protocol to determine the desired solution conditions for RNP formation (assay 1, see below), a protocol to test the desired stoichiometric ratio of gRNA:Cas9 protein (assay 2, see below), a protocol to screen for effective gRNA molecules for Cas9 molecules, e.g., wild-type or mutant Cas9 molecules (assay 3, see below), and a protocol to examine RNP formation in the presence of target DNA (assay 4).

Assay 1

To determine the desired solution to form RNP complexes, a 2 μM solution of Cas9 is made in water with 10x SYPRO Orange® (Life Technologies cat#S-6650) and dispensed into a 384 well plate. An equimolar amount of gRNA diluted in solutions with varied pH and salt is then added. After incubating at room temperature for 10 minutes and centrifugation at 2000 rpm to remove any bubbles, a Bio-Rad CFX384™ Real-Time System C1000 Touch™ Thermal Cycler with the Bio-Rad CFX Manager software is used to run a gradient from 20°C to 90°C with a 1°C increase in temperature every 10 seconds.

Assay 2

The second assay includes mixing various concentrations of gRNA molecules with 2 μM Cas9 in the buffer from assay 1 above and incubating at RT for 10 minutes in a 384 well

plate. An equal volume of optimal buffer with 10x SYPRO Orange® (Life Technologies cat#S-6650) is added and the plate is sealed with Microseal® B adhesive (MSB-1001). Following centrifugation at 2000 rpm to remove any bubbles, a Bio-Rad CFX384™ Real-Time System C1000 Touch™ Thermal Cycler with the Bio-Rad CFX Manager software is used to run a gradient from 20°C to 90°C with a 1°C increase in temperature every 10 seconds.

Assay 3

In the third assay, a Cas9 molecule (e.g., a Cas9 protein, e.g., a Cas9 variant protein) of interest is purified. A library of variant gRNA molecules is synthesized and resuspended to a concentration of 20 μM. The Cas9 molecule is incubated with the gRNA molecule at a final concentration of 1 μM each in a predetermined buffer in the presence of 5x SYPRO Orange® (Life Technologies cat#S-6650). After incubating at room temperature for 10 minutes and centrifugation at 2000 rpm for 2 minutes to remove any bubbles, a Bio-Rad CFX384™ Real-Time System C1000 Touch™ Thermal Cycler with the Bio-Rad CFX Manager software is used to run a gradient from 20°C to 90°C with an increase of 1°C in temperature every 10 seconds.

Assay 4

In the fourth assay, a DSF experiment is performed with the following samples: Cas9 protein alone, Cas9 protein with gRNA, Cas9 protein with gRNA and target DNA, and Cas9 protein with target DNA. The order of mixing components is: reaction solution, Cas9 protein, gRNA, DNA, and SYPRO Orange. The reaction solution contains 10 mM HEPES pH 7.5, 100 mM NaCl, in the absence or presence of $MgCl_2$. Following centrifugation at 2000 rpm for 2 minutes to remove any bubbles, a Bio-Rad CFX384™ Real-Time System C1000 Touch™ Thermal Cycler with the Bio-Rad CFX Manager software is used to run a gradient from 20°C to 90°C with a 1° increase in temperature every 10 seconds.

Examples 1 and 2 as described herein disclose exemplary results using DSF to evaluate and determine the stability of Cas9 molecules and Cas9/gRNA complexes. As shown herein, a higher T_m value of a Cas9/gRNA complex compared to the T_m value of the Cas9 molecule in the absence of the gRNA molecule is indicative of a tighter complex between Cas9 and gRNA. Thus, information regarding the T_m of Cas9 molecules and Cas9/gRNA complexes is useful for determining whether the Cas9/gRNA complex is stable.

In addition to DSF, there are a number of additional techniques known in the art that may be used for determining the stability of the Cas9 molecule in a Cas9 molecule/gRNA molecule complex or a preparation thereof. These include alternative methods to DSF that measure thermostability, including but not limited to, differential scanning calorimetry (DSC) and isothermal titration calorimetry (ITC).

DSC

DSC is a technique that is highly precise in measuring thermostability of material in different buffers as well as apo vs. complex. An advantage to DSC is that it can also provide differences in enthalpy of transitions between samples. However, DSC requires significantly larger quantities of material to run than DSF (>50 fold more). DSC is lower throughput since it can only run a single sample at a time.

ITC

ITC can measure both the thermostability and kinetic rates of interactions of two molecules. The advantage of ITC versus other techniques is that it provides more precise measurements and kinetic information. However, it requires lower throughput and larger quantities of material.

The thermostability techniques disclosed herein may be used to measure the thermostability of a molecule (e.g., Cas9 molecule), which can increase under favorable conditions such as the addition of a binding RNA molecule. In addition, the thermostability of a molecule may increase under favorable conditions such as the presence of a component. In certain embodiments, the component may comprise an additive, a small molecule, a stabilizing reagent, buffer, pH, salt concentration, glycerol concentration, or other buffer component.

In certain embodiments, a molecule (e.g., Cas9/gRNA complex) may be selected or determined to be stable if the thermostability value of the molecule is greater than the thermostability value of a reference molecule or a thermostability reference value. In certain embodiments, the thermostability value being measured is the denaturation temperature value of the molecule. For example, in certain embodiments, a molecule (e.g., Cas9/gRNA complex) may be selected or determined to be stable if the denaturation temperature value of the molecule is greater than the denaturation temperature value of a reference molecule or a denaturation temperature reference value. In certain embodiments, the thermostability value being measured is the T_m value of the molecule. For example, in certain embodiments, a molecule (e.g., Cas9/gRNA complex) may be selected or determined to be stable if the T_m

value of the molecule is greater than the T_m value of a reference molecule or a T_m reference value. In certain embodiments, the reference molecule may be the Cas9 molecule in the absence of the gRNA molecule.

In certain embodiments, the molecule being evaluated may be selected or determined to be stable if the T_m value of the molecule is at least 1°C, at least 2°C, at least 3°C, at least 4°C, at least 5°C, at least 6°C, at least 7°C, at least 8°C, at least 9°C, at least 10°C, at least 11°C, at least 12°C, at least 13°C, at least 14°C, at least 15°C, at least 16°C, at least 17°C, at least 18°C, at least 19°C, at least 20°C, at least 21°C, at least 22°C, at least 23°C, at least 24°C, at least 25°C, at least 26°C, at least 27°C, at least 28°C, at least 29°C, at least 30°C, at least 31°C, at least 32°C, at least 33°C, at least 34°C, at least 35°C, at least 36°C, at least 37°C, at least 38°C, at least 39°C, at least 40°C, at least 41°C, at least 42°C, at least 43°C, at least 44°C, at least 45°C, at least 46°C, at least 47°C, at least 48°C, at least 49°C, or at least 50°C greater than the T_m value of the reference molecule or T_m reference value. For example, the molecule being evaluated may be selected or determined to be stable if the T_m value of the molecule is at least 8°C greater than the reference molecule or T_m reference value.

In certain embodiments, the molecule being evaluated may be selected or determined to be stable if the T_m value of the molecule is about 1°C, about 2°C, about 3°C, about 4°C, about 5°C, about 6°C, about 7°C, about 8°C, about 9°C, about 10°C, about 11°C, about 12°C, about 13°C, about 14°C, about 15°C, about 16°C, about 17°C, about 18°C, about 19°C, about 20°C, about 21°C, about 22°C, about 23°C, about 24°C, about 25°C, about 26°C, about 27°C, about 28°C, about 29°C, about 30°C, about 31°C, about 32°C, about 33°C, about 34°C, about 35°C, about 36°C, about 37°C, about 38°C, about 39°C, about 40°C, about 41°C, about 42°C, about 43°C, about 44°C, about 45°C, about 46°C, about 47°C, about 48°C, about 49°C, or about 50°C greater than the T_m value of the reference molecule or T_m reference value. For example, the molecule being evaluated may be selected or determined to be stable if the T_m value of the molecule is about 8°C greater than the reference molecule or T_m reference value.

In certain embodiments, the molecule being evaluated may be selected or determined to be stable if the T_m value of the molecule is about 1°C to about 5°C, about 6°C to about 10°C, about 11°C to about 15°C, about 16°C to about 20°C, about 21°C to about 25°C, about 26°C to about 30°C, about 31°C to about 35°C, about 36°C to about 40°C, about 41°C to about 45°C, about 46°C to about 50°C greater than the T_m value of the reference molecule or T_m reference value. For example, the molecule being evaluated may be selected or

determined to be stable if the T_m value of the molecule is about 6°C to about 10°C greater than the reference molecule or T_m reference value. In certain embodiments, the molecule being evaluated may be selected or determined to be stable if the T_m value of the molecule is about 8°C to about 9°C greater than the reference molecule or T_m reference value.

Provided herein in certain embodiments, the methods herein may include steps of detecting a T_m value of a Cas9 molecule/gRNA molecule complex, determining a delta value between the T_m value of the Cas9 molecule/gRNA molecule complex and a T_m value of a reference molecule or a T_m reference value, and determining the Cas9 molecule/gRNA molecule complex is stable if the delta value is at least 8°C, and the T_m value of the Cas9 molecule/gRNA molecule complex is greater than the T_m value of the reference molecule or the T_m reference value. In certain embodiments, the reference molecule may be the Cas9 molecule absent the gRNA molecule.

In certain embodiments, the molecule being evaluated may be selected or determined to be stable if a delta value between a T_m value of the molecule being evaluated and a T_m value of a reference molecule or a T_m reference value is at least 1°C, at least 2°C, at least 3°C, at least 4°C, at least 5°C, at least 6°C, at least 7°C, at least 8°C, at least 9°C, at least 10°C, at least 11°C, at least 12°C, at least 13°C, at least 14°C, at least 15°C, at least 16°C, at least 17°C, at least 18°C, at least 19°C, at least 20°C, at least 21°C, at least 22°C, at least 23°C, at least 24°C, at least 25°C, at least 26°C, at least 27°C, at least 28°C, at least 29°C, at least 30°C, at least 31°C, at least 32°C, at least 33°C, at least 34°C, at least 35°C, at least 36°C, at least 37°C, at least 38°C, at least 39°C, at least 40°C, at least 41°C, at least 42°C, at least 43°C, at least 44°C, at least 45°C, at least 46°C, at least 47°C, at least 48°C, at least 49°C, or at least 50°C, and the T_m value of the molecule being evaluated is greater than the T_m value of the reference molecule or the T_m reference value. In certain embodiments, the molecule being evaluated may be a CA9 molecule/gRNA molecule complex and the reference molecule may be the Cas9 molecule absent the gRNA molecule.

In certain embodiments, the molecule being evaluated may be selected or determined to be stable if a delta value between a T_m value of the molecule being evaluated and a T_m value of a reference molecule or a T_m reference value is about 1°C, about 2°C, about 3°C, about 4°C, about 5°C, about 6°C, about 7°C, about 8°C, about 9°C, about 10°C, about 11°C, about 12°C, about 13°C, about 14°C, about 15°C, about 16°C, about 17°C, about 18°C, about 19°C, about 20°C, about 21°C, about 22°C, about 23°C, about 24°C, about 25°C, about 26°C, about 27°C, about 28°C, about 29°C, about 30°C, about 31°C, about 32°C, about 33°C, about

34°C, about 35°C, about 36°C, about 37°C, about 38°C, about 39°C, about 40°C, about 41°C, about 42°C, about 43°C, about 44°C, about 45°C, about 46°C, about 47°C, about 48°C, about 49°C, or about 50°C, and the T_m value of the molecule being evaluated is greater than the T_m value of the reference molecule or the T_m reference value. For example, a Cas9 molecule/gRNA molecule complex may be selected or determined to be stable if the delta value is about 8°C and the T_m value of the Cas9 molecule/gRNA molecule complex is greater than the T_m value of the reference molecule or the T_m reference value. In certain embodiments, the reference molecule may be the Cas9 molecule absent the gRNA molecule.

In certain embodiments, the molecule being evaluated may be selected or determined to be stable if a delta value between a T_m value of the molecule being evaluated and a T_m value of a reference molecule or a T_m reference value is about 1°C to about 5°C, about 6°C to about 10°C, about 11°C to about 15°C, about 16°C to about 20°C, about 21°C to about 25°C, about 26°C to about 30°C, about 31°C to about 35°C, about 36°C to about 40°C, about 41°C to about 45°C, about 46°C to about 50°C, and the T_m value of the molecule being evaluated is greater than the T_m value of the reference molecule or the T_m reference value. For example, a Cas9 molecule/gRNA molecule complex may be selected or determined to be stable if the delta value is about 6°C to about 10°C and the T_m value of the Cas9 molecule/gRNA molecule complex is greater than the T_m value of the reference molecule or the T_m reference value. In certain embodiments, the reference molecule may be the Cas9 molecule absent the gRNA molecule.

In certain embodiments, the methods used herein may be used to evaluate a plurality of gRNAs having different lengths complexed with Cas9 molecules to determine which Cas9/gRNA complex forms a stable Cas9/gRNA complex. These methods may also be used to evaluate different stoichiometries of Cas9 molecules and gRNA molecules to determine which Cas9/gRNA complex forms a stable Cas9/gRNA complex.

In certain embodiments, a plurality of samples, each sample comprising a Cas9/gRNA complex may be generated by combining a Cas9 molecule and one of a plurality of gRNA molecules. In certain embodiments, a T_m value of the Cas9/gRNA complex may be detected in each of the plurality of samples. In certain embodiments, at least one sample may be selected from the plurality of samples based on one or more of (i) a comparison of the T_m values in the plurality of samples to a T_m value of a reference Cas9/gRNA complex or a pre-determined threshold T_m value, or (ii) a relative ordering of the T_m values of the plurality of samples. In certain embodiments, the T_m value may be detected by DSF. In certain

embodiments, the at least one sample may be selected if the T_m value of the Cas9 molecule/gRNA molecule complex is at least 8°C greater than the T_m value of the Cas9 molecule absent the gRNA molecule.

Techniques for measuring the activity of Cas9 and Cas9/gRNA complexes

In addition to thermostability techniques, there are a variety of other techniques known in the art that may be used with the methods herein. For example, certain activities of a Cas9 molecule/gRNA molecule complex or molecules thereof can be measured to select or determine whether the Cas9 molecule/gRNA molecule complex or molecules thereof are stable. These techniques may be used alone or in conjunction with the thermostability techniques described herein to determine whether a Cas9 molecule/gRNA molecule complex or molecules thereof are stable. The techniques disclosed herein may be used to detect an activity of the molecule being evaluated (e.g., Cas9 molecule in a Cas9 molecule/gRNA molecule complex, gRNA molecule in a Cas9 molecule/gRNA molecule complex, or Cas9 molecule/gRNA molecule complex or preparation thereof). In certain embodiments, an activity value of the molecule being evaluated may be measured. In certain embodiments, the molecule being evaluated may be selected or determined to be stable if the activity value of the molecule being evaluated is greater than the activity value of a reference molecule or an activity reference value.

The various activities of the molecule to be evaluated that may be detected include the binding activity and cleavage activity of the molecule. The binding activity and cleavage activity of the molecule to be evaluated may be detected using the techniques described herein.

Some examples of a binding activity of a molecule include, without limitation, the ability of a gRNA molecule to remain hybridized to the DNA target, the ability of a gRNA molecule to bind to the Cas9 molecule of the Cas9 molecule/gRNA molecule complex, or the ability of a gRNA molecule to bind to the Cas9 molecule of the Cas9 molecule/gRNA molecule complex. In certain embodiments, when a binding activity of a molecule is being detected, a binding value may be measured. In certain embodiments, the molecule being evaluated may be selected or determined to be stable if the binding value of the molecule being evaluated is greater than the binding value of a reference molecule or a binding reference value.

Some examples of a cleavage activity may include, without limitation, the ability to induce indels, the ability to modify a target DNA, and a propensity of a preselected repair method. In certain embodiments, when the cleavage activity is being detected, a cleavage value may be measured. In certain embodiments, the molecule being evaluated may be selected or determined to be stable if the cleavage value of the molecule being evaluated is greater than the cleavage value of a reference molecule or a cleavage reference value.

Techniques for measuring the binding activity (kinetics) of Cas9 and Cas9/gRNA complexes

The binding activity of the molecule being evaluated (e.g., Cas9 molecule in a Cas9 molecule/gRNA molecule complex, gRNA molecule in a Cas9 molecule/gRNA molecule complex, or Cas9 molecule/gRNA molecule complex or preparation thereof) can be detected using various techniques. The kinetics of binding between two molecules may be more favorable under certain conditions, such as the presence of a component. In certain embodiments, the component may comprise an additive, a small molecule, a stabilizing reagent, buffer, pH, salt concentration, glycerol concentration, or other buffer component. or the addition of a particular component.

Methods that include detecting the binding activity of Cas9/gRNA complexes include, without limitation, detecting the ability of the gRNA molecule to bind to the Cas9 molecule of the Cas9 molecule/gRNA molecule complex and detecting the ability of Cas9 molecules and Cas9/gRNA complexes to bind to target DNA. These methods may be performed using techniques such as kinetics assays that provide biophysical information about the binding of the molecules being evaluated. Some examples of kinetics assays that may be used are, without limitation, surface plasmon resonance (SPR), BioLayer Interferometry (BLI), and gel band shift assay as described below.

SPR

SPR requires the use of either a BiaCore or ProteOn XPR system. In this technique, one molecule is attached either covalently or via an affinity method to the surface of a chip. The second molecule is injected into a flow cell and is pushed through via buffer. Changes in the angle of reflected light lead to changes in the amount of plasmon resonance. From this, kinetic association and disassociation can be measured.

BLI

BLI requires an instrument called the Octet by fortéBio. Similar to SPR, BLI is capable of determining kinetic rates of interaction between two molecules.

Gel band shift assay

Gel band shift assay (e.g., electrophoretic mobility shift assay) is another method to determine the K_D of two interacting molecules. This determination is more crude than other available techniques, but has the advantage that it can be performed with relatively inexpensive reagents.

Binding assay: testing the binding of Cas9 molecules and Cas9/gRNA complexes to target DNA

Exemplary methods for evaluating the binding of Cas9 molecule to target DNA have been described previously (Jinek 2012). The techniques described herein, such as SPR, BLI and gel band shift assays may be used to measure, for example, the ability of a gRNA molecule to bind to the Cas9 molecule of the Cas9 molecule/gRNA molecule complex or the ability of Cas9 molecules and Cas9/gRNA complexes to bind to target DNA.

For example, in an electrophoretic mobility shift assay, target DNA duplexes are formed by mixing of each strand (10 nmol) in deionized water, heating to 95°C for 3 minutes, and slow cooling to room temperature. All DNAs are purified on 8% native gels containing 1X TBE. DNA bands are visualized by UV shadowing, excised, and eluted by soaking gel pieces in DEPC-treated H_2O. Eluted DNA is ethanol precipitated and dissolved in DEPC-treated H_2O. DNA samples are 5' end labeled with [γ-^{32}P]-ATP using T4 polynucleotide kinase for 30 minutes at 37°C. Polynucleotide kinase is heat denatured at 65°C for 20 minutes, and unincorporated radiolabel is removed using a column. Binding assays are performed in buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, and 10% glycerol in a total volume of 10 µL. Cas9 protein molecules are programmed with equimolar amounts of pre-annealed gRNA molecule and titrated from 100 pM to 1 µM. Radiolabeled DNA is added to a final concentration of 20 pM. Samples are incubated for 1 hour at 37°C and resolved at 4°C on an 8% native polyacrylamide gel containing 1X TBE and 5 mM $MgCl_2$. Gels are dried and DNA visualized by phosphorimaging.

Techniques for measuring cleavage activity of Cas9/gRNA complexes

Methods described herein may include detecting the cleavage activity of the Cas9/gRNA complex. This may include detecting the ability of the Cas9/gRNA complex to modify a target DNA, for example, the ability of the Cas9/gRNA complex to cleave a target

nucleic acid. Some examples of techniques that may be used to detect the cleavage activity of Cas9/gRNA complexes are described herein.

Cleavage assay: testing the endonuclease activity of Cas9 molecule/gRNA molecule complexes

Additional activities that can be tested to determine the stability of a Cas9/gRNA complex include the ability of the Cas9/gRNA complex to modify a target DNA, for example, the ability of the Cas9/gRNA complex to cleave a target nucleic acid. The endonuclease activity of a Cas9/gRNA complex may be measured as disclosed herein. For example, exemplary methods for evaluating the endonuclease activity of Cas9 molecule have been described previously (Jinek 2012).

The ability of a Cas9 molecule/gRNA molecule complex to bind to and cleave a target nucleic acid can be evaluated in a plasmid cleavage assay. In this assay, a synthetic or *in vitro*-transcribed gRNA molecule is pre-annealed prior to the reaction by heating to 95°C and slowly cooling down to room temperature. Native or restriction digest-linearized plasmid DNA (300 ng (~8 nM)) is incubated for 60 minutes at 37°C with purified Cas9 protein molecule (50-500 nM) and gRNA (50-500 nM, 1:1) in a Cas9 plasmid cleavage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA) with or without 10 mM $MgCl_2$. The reactions are stopped with 5X DNA loading buffer (30% glycerol, 1.2% SDS, 250 mM EDTA), resolved by a 0.8 or 1% agarose gel electrophoresis and visualized by ethidium bromide staining. The resulting cleavage products indicate whether the Cas9 molecule cleaves both DNA strands, or only one of the two strands. For example, linear DNA products indicate the cleavage of both DNA strands, while nicked open circular products indicate that only one of the two strands is cleaved.

Alternatively, the ability of a Cas9 molecule/gRNA molecule complex to bind to and cleave a target nucleic acid can be evaluated in an oligonucleotide DNA cleavage assay. In this assay, DNA oligonucleotides (10 pmol) are radiolabeled by incubating with 5 units T4 polynucleotide kinase and ~3–6 pmol (~20–40 mCi) [γ-^{32}P]-ATP in 1X T4 polynucleotide kinase reaction buffer at 37°C for 30 minutes, in a 50 µL reaction. After heat inactivation (65°C for 20 min), reactions are purified through a column to remove unincorporated label. Duplex substrates (100 nM) are generated by annealing labeled oligonucleotides with equimolar amounts of unlabeled complementary oligonucleotide at 95°C for 3 minutes, followed by slow cooling to room temperature. For cleavage assays, gRNA molecules are

annealed by heating to 95°C for 30 seconds, followed by slow cooling to room temperature. Cas9 (500 nM final concentration) is pre-incubated with the annealed gRNA molecules (500 nM) in cleavage assay buffer (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM MgCl2, 1 mM DTT, 5% glycerol) in a total volume of 9 μL. Reactions are initiated by the addition of 1 μL target DNA (10 nM) and incubated for 1 hour at 37°C. Reactions are quenched by the addition of 20 μL of loading dye (5 mM EDTA, 0.025% SDS, 5% glycerol in formamide) and heated to 95°C for 5 minutes. Cleavage products are resolved on 12% denaturing polyacrylamide gels containing 7 M urea and visualized by phosphorimaging. The resulting cleavage products indicate that whether the complementary strand, the non-complementary strand, or both are cleaved.

One or both of these assays can be used to determine the stability of a Cas9/gRNA complex and evaluate the suitability of a candidate gRNA molecule or candidate Cas9 molecule.

Genome editing approaches

The methods described herein can be used for evaluating Cas9 molecule/gRNA molecule complexes. These Cas9 molecule/gRNA molecule complexes can be used to target genes using one or more of the approaches or pathways discussed herein. In certain embodiments, a mutation in a target gene is corrected by HDR using an exogenously provided template nucleic acid. In other embodiments, a mutation in a target gene is corrected by HDR without using an exogenously provided template nucleic acid. In certain embodiments, one or both alleles of a target gene are knocked out using NHEJ. In certain embodiments, expression of a target gene is knocked down. The methods described herein can be used to evaluate whether a Cas9 molecule/gRNA molecule complex is desirable for one or more of the approaches or pathways discussed herein.

HDR repair and template nucleic acids

As described herein, nuclease-induced HDR can be used to alter a target position within a target sequence (e.g., correct, e.g., repair or edit, a mutation in the genome).

In certain embodiments, HDR-based methods for altering a target position utilize an exogenously provided template nucleic acid (also referred to herein as a donor construct or donor template). While not wishing to be bound by theory, it is believed that alteration of the target position occurs by HDR with the exogenously provided donor template or template

nucleic acid. It is contemplated that a plasmid donor template can be used as a template for homologous recombination. It is further contemplated that a single stranded donor template can be used as a template for alteration of a target position by alternate methods of HDR (e.g., single strand annealing) between the target sequence and the donor template. Donor template-effected alteration of a target position depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a double-strand break or two single-strand breaks.

In other embodiments, HDR-based methods for altering a target position do not utilize an exogenously provided template nucleic acid. While not wishing to be bound by theory, it is believed that alteration of the target position occurs by HDR with an endogenous genomic donor sequence. In certain embodiments, the endogenous genomic donor sequence is located on the same chromosome as the target position. In other embodiments, the endogenous genomic donor sequence is located on a different chromosome from the target sequence. Alteration of a target position by an endogenous genomic donor sequence depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a double-strand break or two single-strand breaks.

Mutations that can be corrected by HDR using a template nucleic acid, or using endogenous genomic donor sequence, include point mutations. In certain embodiments, a point mutation can be corrected using either one double-strand break or two single-strand breaks. In certain embodiments, a point mutation can be corrected by (1) one double-strand break, (2) two single-strand breaks, (3) two double-strand breaks with a break occurring on each side of the target position, (4) one double-strand break and two single-strand breaks with the double-strand break and two single-strand breaks occurring on each side of the target position, (5) four single-strand breaks with a pair of single-strand breaks occurring on each side of the target position, or (6) one single-strand break.

In certain embodiments wherein a single-stranded template nucleic acid is used, the target position can be altered by alternative HDR.

Donor template-effected alteration of a target position depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a nick, a double-strand break, or two single-strand breaks, e.g., one on each strand of the target nucleic acid. After introduction of the breaks on the target nucleic acid, resection occurs at the break ends resulting in single stranded overhanging DNA regions.

In canonical HDR, a double-stranded donor template is introduced, comprising homologous sequence to the target nucleic acid that will either be directly incorporated into

the target nucleic acid or used as a template to correct the sequence of the target nucleic acid. After resection at the break, repair can progress by different pathways, e.g., by the double Holliday junction model (or double strand break repair, DSBR, pathway) or the synthesis-dependent strand annealing (SDSA) pathway. In the double Holliday junction model, strand invasion by the two single stranded overhangs of the target nucleic acid to the homologous sequences in the donor template occurs, resulting in the formation of an intermediate with two Holliday junctions. The junctions migrate as new DNA is synthesized from the ends of the invading strand to fill the gap resulting from the resection. The end of the newly synthesized DNA is ligated to the resected end, and the junctions are resolved, resulting in the correction of the target nucleic acid, e.g., incorporation of the correct sequence of the donor template at the corresponding target position. Crossover with the donor template may occur upon resolution of the junctions. In the SDSA pathway, only one single stranded overhang invades the donor template and new DNA is synthesized from the end of the invading strand to fill the gap resulting from resection. The newly synthesized DNA then anneals to the remaining single stranded overhang, new DNA is synthesized to fill in the gap, and the strands are ligated to produce the corrected DNA duplex.

In alternative HDR, a single strand donor template, e.g., template nucleic acid, is introduced. A nick, single-strand break, or double-strand break at the target nucleic acid, for altering a desired A1AT target position, is mediated by a Cas9 molecule, e.g., described herein, and resection at the break occurs to reveal single stranded overhangs. Incorporation of the sequence of the template nucleic acid to correct or alter the target position of the target nucleic acid typically occurs by the SDSA pathway, as described above.

Additional details on template nucleic acids are provided in Section IV entitled "Template nucleic acids" in International Application PCT/US2014/057905.

NHEJ approaches for gene targeting

As described herein, nuclease-induced NHEJ can be used to target gene-specific knockouts and remove (e.g., delete) sequences in a gene of interest.

While not wishing to be bound by theory, it is believed that, in certain embodiments, the genomic alterations associated with the methods described herein rely on nuclease-induced NHEJ and the error-prone nature of the NHEJ repair pathway. NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the

double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. Two-thirds of these mutations typically alter the reading frame and, therefore, produce a non-functional protein. Additionally, mutations that maintain the reading frame, but which insert or delete a significant amount of sequence, can destroy functionality of the protein. This is locus dependent as mutations in critical functional domains are likely less tolerable than mutations in non-critical regions of the protein.

The indel mutations generated by NHEJ are unpredictable in nature; however, at a given break site certain indel sequences are favored and are over represented in the population, likely due to small regions of microhomology. The lengths of deletions can vary widely; they are most commonly in the 1-50 bp range, but can reach greater than 100-200 bp. Insertions tend to be shorter and often include short duplications of the sequence immediately surrounding the break site. However, it is possible to obtain large insertions, and in these cases, the inserted sequence has often been traced to other regions of the genome or to plasmid DNA present in the cells.

Because NHEJ is a mutagenic process, it can also be used to delete small sequence motifs (e.g., motifs less than or equal to 50 nucleotides in length) as long as the generation of a specific final sequence is not required. If a double-strand break is targeted near to a target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. In this way, DNA segments as large as several hundred kilobases can be deleted. Both of these approaches can be used to delete specific DNA sequences; however, the error-prone nature of NHEJ may still produce indel mutations at the site of repair.

Both double strand cleaving eaCas9 molecules and single strand, or nickase, eaCas9 molecules can be used in the methods and compositions described herein to generate NHEJ-mediated indels. NHEJ-mediated indels targeted to the gene, e.g., a coding region, e.g., an early coding region of a gene, of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a start codon, within a first exon of the coding sequence, or

within 500 bp of the start codon (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100, or 50 bp).

Targeted knockdown

Unlike CRISPR/Cas-mediated gene knockout, which permanently eliminates expression by mutating the gene at the DNA level, CRISPR/Cas knockdown allows for temporary reduction of gene expression through the use of artificial transcription factors. Mutating key residues in both DNA cleavage domains of the Cas9 protein (e.g., the D10A and H840A mutations) results in the generation of an enzymatically inactive Cas9 (eiCas9, also known as dead Cas9 or dCas9) molecule. An eiCas9 molecule complexes with a gRNA and localizes to the DNA sequence specified by that gRNA's targeting domain, but does cleave the target DNA. Fusion of eiCas9 to an effector domain, e.g., a transcription repression domain, enables recruitment of the effector to any DNA site specified by the gRNA. Although the eiCas9 molecule itself can block transcription when recruited to early regions in the coding sequence, more robust repression can be achieved by fusing a transcriptional repression domain (e.g., KRAB, SID, or ERD) to the eiCas9 and recruiting it to the target knockdown position, e.g., within 1000 bp of sequence 3' to the start codon or within 500 bp of a promoter region 5' to a gene start codon. It is likely that targeting DNAse I hypersensitive sites (DHSs) of the promoter may yield more efficient gene repression or activation because these regions are more likely to be accessible to the Cas9 protein and are also more likely to harbor sites for endogenous transcription factors. Especially for gene repression, it is contemplated herein that blocking the binding site of an endogenous transcription factor would aid in downregulating gene expression. In certain embodiments, one or more eiCas9 molecules may be used to block binding of one or more endogenous transcription factors. In other embodiments, an eiCas9 molecule can be fused to a chromatin modifying protein. Altering chromatin status can result in decreased expression of the target gene. One or more eiCas9 molecules fused to one or more chromatin modifying proteins may be used to alter chromatin status.

In certain embodiments, a gRNA molecule can be targeted to a known transcription response element (e.g., promoter, enhancer, etc.), a known upstream activating sequence (UAS), and/or a sequence of unknown or known function suspected of being able to control expression of the target DNA.

CRISPR/Cas-mediated gene knockdown can be used to reduce expression of an unwanted allele or transcript. Contemplated herein are scenarios wherein permanent destruction of the gene is not ideal. In these scenarios, site-specific repression may be used to temporarily reduce or eliminate expression. It is also contemplated herein that the off-target effects of a Cas-repressor may be less severe than those of a Cas-nuclease as a nuclease can cleave any DNA sequence and cause mutations whereas a Cas-repressor may only have an effect if it targets the promoter region of an actively transcribed gene. However, while nuclease-mediated knockout is permanent, repression may only persist as long as the Cas-repressor is present in the cells. Once the repressor is no longer present, it is likely that endogenous transcription factors and gene regulatory elements would restore expression to its natural state.

Single-strand annealing

Single strand annealing (SSA) is another DNA repair process that repairs a double-strand break between two repeat sequences present in a target nucleic acid. Repeat sequences utilized by the SSA pathway are generally greater than 30 nucleotides in length. Resection at the break ends occurs to reveal repeat sequences on both strands of the target nucleic acid. After resection, single strand overhangs containing the repeat sequences are coated with RPA protein to prevent the repeats sequences from inappropriate annealing, e.g., to themselves. RAD52 binds to and each of the repeat sequences on the overhangs and aligns the sequences to enable the annealing of the complementary repeat sequences. After annealing, the single-strand flaps of the overhangs are cleaved. New DNA synthesis fills in any gaps, and ligation restores the DNA duplex. As a result of the processing, the DNA sequence between the two repeats is deleted. The length of the deletion can depend on many factors including the location of the two repeats utilized, and the pathway or processivity of the resection.

In contrast to HDR pathways, SSA does not require a template nucleic acid to alter or correct a target nucleic acid sequence. Instead, the complementary repeat sequence is utilized.

Other DNA repair pathways

SSBR (single strand break repair)

Single-stranded breaks (SSB) in the genome are repaired by the SSBR pathway, which is a distinct mechanism from the DSB repair mechanisms discussed above. The SSBR

pathway has four major stages: SSB detection, DNA end processing, DNA gap filling, and DNA ligation. A more detailed explanation is given in Caldecott 2008, and a summary is given here.

In the first stage, when a SSB forms, PARP1 and/or PARP2 recognize the break and recruit repair machinery. The binding and activity of PARP1 at DNA breaks is transient and it seems to accelerate SSBR by promoting the focal accumulation or stability of SSBR protein complexes at the lesion. Arguably the most important of these SSBR proteins is XRCC1, which functions as a molecular scaffold that interacts with, stabilizes, and stimulates multiple enzymatic components of the SSBR process including the protein responsible for cleaning the DNA 3' and 5' ends. For instance, XRCC1 interacts with several proteins (DNA polymerase beta, PNK, and three nucleases, APE1, APTX, and APLF) that promote end processing. APE1 has endonuclease activity. APLF exhibits endonuclease and 3' to 5' exonuclease activities. APTX has endonuclease and 3' to 5' exonuclease activity.

This end processing is an important stage of SSBR since the 3'- and/or 5'-termini of most, if not all, SSBs are 'damaged.' End processing generally involves restoring a damaged 3'-end to a hydroxylated state and and/or a damaged 5' end to a phosphate moiety, so that the ends become ligation-competent. Enzymes that can process damaged 3' termini include PNKP, APE1, and TDP1. Enzymes that can process damaged 5' termini include PNKP, DNA polymerase beta, and APTX. LIG3 (DNA ligase III) can also participate in end processing. Once the ends are cleaned, gap filling can occur.

At the DNA gap filling stage, the proteins typically present are PARP1, DNA polymerase beta, XRCC1, FEN1 (flap endonuclease 1), DNA polymerase delta/epsilon, PCNA, and LIG1. There are two ways of gap filling, the short patch repair and the long patch repair. Short patch repair involves the insertion of a single nucleotide that is missing. At some SSBs, "gap filling" might continue displacing two or more nucleotides (displacement of up to 12 bases have been reported). FEN1 is an endonuclease that removes the displaced 5'-residues. Multiple DNA polymerases, including Polβ, are involved in the repair of SSBs, with the choice of DNA polymerase influenced by the source and type of SSB.

In the fourth stage, a DNA ligase such as LIG1 (Ligase I) or LIG3 (Ligase III) catalyzes joining of the ends. Short patch repair uses Ligase III and long patch repair uses Ligase I.

Sometimes, SSBR is replication-coupled. This pathway can involve one or more of CtIP, MRN, ERCC1, and FEN1. Additional factors that may promote SSBR include: aPARP, PARP1, PARP2, PARG, XRCC1, DNA polymerase b, DNA polymerase d, DNA polymerase e, PCNA, LIG1, PNK, PNKP, APE1, APTX, APLF, TDP1, LIG3, FEN1, CtIP, MRN, and ERCC1.

MMR (mismatch repair)

Cells contain three excision repair pathways: MMR, BER, and NER. The excision repair pathways have a common feature in that they typically recognize a lesion on one strand of the DNA, then exo/endonucleases remove the lesion and leave a 1-30 nucleotide gap that is sub-sequentially filled in by DNA polymerase and finally sealed with ligase. A more complete picture is given in Li 2008, and a summary is provided here.

Mismatch repair (MMR) operates on mispaired DNA bases.

The MSH2/6 or MSH2/3 complexes both have ATPases activity that plays an important role in mismatch recognition and the initiation of repair. MSH2/6 preferentially recognizes base-base mismatches and identifies mispairs of 1 or 2 nucleotides, while MSH2/3 preferentially recognizes larger ID mispairs.

hMLH1 heterodimerizes with hPMS2 to form hMutLα which possesses an ATPase activity and is important for multiple steps of MMR. It possesses a PCNA/replication factor C (RFC)-dependent endonuclease activity which plays an important role in 3' nick-directed MMR involving EXO1 (EXO1 is a participant in both HR and MMR.) It regulates termination of mismatch-provoked excision. Ligase I is the relevant ligase for this pathway. Additional factors that may promote MMR include: EXO1, MSH2, MSH3, MSH6, MLH1, PMS2, MLH3, DNA Pol d, RPA, HMGB1, RFC, and DNA ligase I.

Base excision repair (BER)

The base excision repair (BER) pathway is active throughout the cell cycle; it is responsible primarily for removing small, non-helix-distorting base lesions from the genome. In contrast, the related Nucleotide Excision Repair pathway (discussed in the next section) repairs bulky helix-distorting lesions. A more detailed explanation is given in Caldecott 2008, and a summary is given here.

Upon DNA base damage, base excision repair (BER) is initiated and the process can be simplified into five major steps: (a) removal of the damaged DNA base; (b) incision of the

subsequent a basic site; (c) clean-up of the DNA ends; (d) insertion of the correct nucleotide into the repair gap; and (e) ligation of the remaining nick in the DNA backbone. These last steps are similar to the SSBR.

In the first step, a damage-specific DNA glycosylase excises the damaged base through cleavage of the N-glycosidic bond linking the base to the sugar phosphate backbone. Then AP endonuclease-1 (APE1) or bifunctional DNA glycosylases with an associated lyase activity incised the phosphodiester backbone to create a DNA single strand break (SSB). The third step of BER involves cleaning-up of the DNA ends. The fourth step in BER is conducted by Polβ that adds a new complementary nucleotide into the repair gap and in the final step XRCC1/Ligase III seals the remaining nick in the DNA backbone. This completes the short-patch BER pathway in which the majority (~80%) of damaged DNA bases are repaired. However, if the 5'-ends in step 3 are resistant to end processing activity, following one nucleotide insertion by Pol β there is then a polymerase switch to the replicative DNA polymerases, Pol δ/ε, which then add ~2–8 more nucleotides into the DNA repair gap. This creates a 5' flap structure, which is recognized and excised by flap endonuclease-1 (FEN-1) in association with the processivity factor proliferating cell nuclear antigen (PCNA). DNA ligase I then seals the remaining nick in the DNA backbone and completes long-patch BER. Additional factors that may promote the BER pathway include: DNA glycosylase, APE1, Polb, Pold, Pole, XRCC1, Ligase III, FEN-1, PCNA, RECQL4, WRN, MYH, PNKP, and APTX.

Nucleotide excision repair (NER)

Nucleotide excision repair (NER) is an important excision mechanism that removes bulky helix-distorting lesions from DNA. Additional details about NER are given in Marteijn 2014, and a summary is given here. NER a broad pathway encompassing two smaller pathways: global genomic NER (GG-NER) and transcription coupled repair NER (TC-NER). GG-NER and TC-NER use different factors for recognizing DNA damage. However, they utilize the same machinery for lesion incision, repair, and ligation.

Once damage is recognized, the cell removes a short single-stranded DNA segment that contains the lesion. Endonucleases XPF/ERCC1 and XPG (encoded by ERCC5) remove the lesion by cutting the damaged strand on either side of the lesion, resulting in a single-strand gap of 22–30 nucleotides. Next, the cell performs DNA gap filling synthesis and ligation. Involved in this process are: PCNA, RFC, DNA Pol δ, DNA Pol ε or DNA Pol κ,

and DNA ligase I or XRCC1/Ligase III. Replicating cells tend to use DNA pol ε and DNA ligase I, while non-replicating cells tend to use DNA Pol δ, DNA Pol κ, and the XRCC1/ Ligase III complex to perform the ligation step.

NER can involve the following factors: XPA-G, POLH, XPF, ERCC1, XPA-G, and LIG1. Transcription-coupled NER (TC-NER) can involve the following factors: CSA, CSB, XPB, XPD, XPG, ERCC1, and TTDA. Additional factors that may promote the NER repair pathway include XPA-G, POLH, XPF, ERCC1, XPA-G, LIG1, CSA, CSB, XPA, XPB, XPC, XPD, XPF, XPG, TTDA, UVSSA, USP7, CETN2, RAD23B, UV-DDB, CAK subcomplex, RPA, and PCNA.

Interstrand crosslink (ICL)

A dedicated pathway called the ICL repair pathway repairs interstrand crosslinks. Interstrand crosslinks, or covalent crosslinks between bases in different DNA strand, can occur during replication or transcription. ICL repair involves the coordination of multiple repair processes, in particular, nucleolytic activity, translesion synthesis (TLS), and HDR. Nucleases are recruited to excise the ICL on either side of the crosslinked bases, while TLS and HDR are coordinated to repair the cut strands. ICL repair can involve the following factors: endonucleases, e.g., XPF and RAD51C, endonucleases such as RAD51, translesion polymerases, e.g., DNA polymerase zeta and Rev1), and the Fanconi anemia (FA) proteins, e.g., FancJ.

Other pathways

Several other DNA repair pathways exist in mammals.

Translesion synthesis (TLS) is a pathway for repairing a single stranded break left after a defective replication event and involves translesion polymerases, e.g., DNA polβ and Rev1.

Error-free postreplication repair (PRR) is another pathway for repairing a single stranded break left after a defective replication event.

Target cells

Cas9 molecules and gRNA molecules, e.g., a Cas9 molecule/gRNA molecule complex, can be used to alter (e.g., introduce a mutation in) a target nucleic acid in a wide

variety of cells. This alteration may be carried out *in vitro*, *ex vivo*, or *in vivo*. In certain embodiments, this alteration results in modulation of gene expression.

The Cas9 and gRNA molecules described herein can be delivered to a target cell. Exemplary target cells include, but are not limited to, blood cells, neural cells, immune cells, muscle cells, kidney cells, mammary cells, GI tract cells, vascular cells, lung cells, bone cells, pancreatic cells, skin cells, adipocytes, hormone secreting cells, liver cells, epithelial cells, and fibroblasts. In certain embodiments, the target cell is a normal cell. In other embodiments, the target cell is a diseased cell. In certain of these embodiments, the target cell is a cancer cell.

A suitable cell may include a stem cell such as, e.g., an embryonic stem cell, induced pluripotent stem cell, hematopoietic stem cell, neuronal stem cell, or mesenchymal stem cell. In certain embodiments, the cell is an induced pluripotent stem (iPS) cell or a cell derived from an iPS cell, e.g., an iPS cell generated from the subject, modified to correct the mutation, and differentiated into a clinically relevant cell such as, e.g., a hepatocyte, macrophage, mononuclear phagocyte, alveolar macrophage, myeloid progenitor cell, lung epithelial cell, or hematopoietic stem cell. In certain embodiments, AAV is used to transduce the target cells.

Cells produced by the methods described herein may be used immediately. Alternatively, the cells may be frozen (e.g., in liquid nitrogen) and stored for later use. The cells will usually be frozen in 10% dimethylsulfoxide (DMSO), 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperature, and thawed in such a manner as commonly known in the art for thawing frozen cultured cells.

Delivery, formulations, and routes of administration

Cas system components, e.g., a Cas9 molecule, gRNA molecule (e.g., a Cas9 molecule/gRNA molecule complex), a donor template nucleic acid, or all three, can be delivered, formulated, or administered in a variety of forms, see, e.g., **Tables 3** and **4**. **Table 3** provides examples of how the components can be formulated, delivered, or administered. **Table 4** summarizes various delivery methods for the components of a Cas system, e.g., the Cas9 molecule component and the gRNA molecule component, as described herein.

DNA-based delivery of Cas9, gRNA, and/or donor templates

DNA encoding a Cas9 molecule (e.g., a eaCas9 molecule) or gRNA molecule, a donor template, or any combination thereof (e.g., two or all) can be administered to subjects or delivered into cells by art-known methods or as described herein. For example, DNA encoding Cas9 and/or gRNA, as well as donor templates, can be delivered by, e.g., vectors (e.g., viral or non-viral vectors), non-vector based methods (e.g., using naked DNA or DNA complexes), or a combination thereof. Similarly, donor templates can be delivered by, e.g., vectors (e.g., viral or non-viral vectors), non-vector based methods (e.g., using naked DNA or DNA complexes), or a combination thereof.

DNA encoding Cas9 molecules (e.g., eaCas9 molecules) and/or gRNA molecules can be conjugated to molecules (e.g., N-acetylgalactosamine) promoting uptake by the target cells (e.g., the target cells described herein). Donor templates can likewise be conjugated to molecules (e.g., N-acetylgalactosamine) promoting uptake by the target cells (e.g., the target cells described herein).

In certain embodiments, a DNA encoding Cas9 and/or gRNA is delivered by a vector (e.g., viral vector/virus or plasmid).

In certain embodiments, the vector may comprise a DNA sequence that encodes a Cas9 molecule and/or a gRNA molecule. In certain embodiments, the vector may comprise a donor template with high homology to the region (e.g., target sequence) being targeted. In certain of these embodiments, the donor template comprises all or part of a target sequence. Exemplary donor templates are a repair template, e.g., a gene correction template, or a gene mutation template, e.g., point mutation (e.g., single nucleotide (nt) substitution) template).

In certain embodiments, the vector may comprise a sequence encoding a signal peptide (e.g., for nuclear localization, nucleolar localization, or mitochondrial localization), fused, e.g., to a Cas9 molecule sequence. In certain embodiments, the vector may comprise one or more regulatory/control elements, e.g., promoters, enhancers, introns, polyadenylation signals, Kozak consensus sequences, internal ribosome entry sites (IRES), 2A sequences, and/or splice acceptors or donors. In certain of these embodiments wherein the vector comprises a promoter, the promoter is recognized by RNA polymerase II (e.g., a CMV promoter). In other embodiments, the promoter is recognized by RNA polymerase III (e.g., a U6 promoter).

In certain embodiments, the vector is a viral vector (e.g., for generation of recombinant viruses). In certain of these embodiments, the virus is a DNA virus (e.g.,

dsDNA or ssDNA virus). In other embodiments, the virus is an RNA virus (e.g., an ssRNA virus). Exemplary viral vectors/viruses include, e.g., retroviruses, lentiviruses, adenoviruses, adeno-associated viruses (AAV), vaccinia viruses, poxviruses, and herpes simplex viruses. In certain embodiments, a DNA encoding Cas9 and/or gRNA is delivered by a recombinant AAV. In certain embodiments, a donor template nucleic acid is delivered by a recombinant AAV. In certain embodiments, the viral vector is capable of cell type and/or tissue type recognition. In certain embodiments, the viral vector achieves cell type-specific expression. In certain embodiments, the viral vector has increased efficiency of fusion of the viral vector and a target cell membrane.

In certain embodiments, a DNA encoding Cas9 and/or gRNA is delivered by a non-vector based method (e.g., using naked DNA or DNA complexes). For example, the DNA can be delivered, e.g., by organically modified silica or silicate (Ormosil), electroporation, transient cell compression or squeezing (see, e.g., Lee 2012), gene gun, sonoporation, magnetofection, lipid-mediated transfection, dendrimers, inorganic nanoparticles, calcium phosphates, or a combination thereof.

In certain embodiments, a DNA encoding Cas9 and/or gRNA is delivered by a combination of vector and non-vector based methods. In certain embodiments, a donor template is delivered by a combination of vector and non-vector based methods.

Exemplary lipids for gene transfer are shown below in **Table 1**. Exemplary polymers for gene transfer are shown below in **Table 5**.

In certain embodiments, a non-vector delivery vehicle has targeting modifications to increase target cell update of nanoparticles and liposomes, e.g., cell specific antigens, monoclonal antibodies, single chain antibodies, aptamers, polymers, sugars (e.g., N-acetylgalactosamine (GalNAc)), and cell penetrating peptides. In certain embodiments, the vehicle uses fusogenic and endosome-destabilizing peptides/polymers. In certain embodiments, the vehicle undergoes acid-triggered conformational changes (e.g., to accelerate endosomal escape of the cargo). In certain embodiments, a stimuli-cleavable polymer is used, e.g., for release in a cellular compartment. In certain embodiments, the delivery vehicle is a biological non-viral delivery vehicle.

In certain embodiments, one or more nucleic acid molecules (e.g., DNA molecules) other than components of a Cas system (i.e., other than DNA encoding Cas9 molecules and/or gRNA molecules, or donor templates) are delivered. In certain of these embodiments, these other nucleic acid molecules are delivered at the same time as one or more of the

components of the Cas system. In other embodiments, these other nucleic acid molecules are delivered before or after (e.g., less than about 30 minutes, 1 hour, 2 hours, 3 hours, 6 hours, 9 hours, 12 hours, 1 day, 2 days, 3 days, 1 week, 2 weeks, or 4 weeks) one or more of the components of the Cas system. In certain embodiments, these other nucleic acid molecules are delivered by a different means than the one or more of the components of the Cas system. The other nucleic acid molecules can be delivered by any of the delivery methods described herein.

RNA-based delivery of Cas9 and/or gRNA

gRNA molecules and/or RNA molecules encoding Cas9 molecules can be delivered into cells, e.g., target cells described herein, by art-known methods or as described herein. For example, gRNA molecules and/or RNA molecules encoding Cas9 molecules can be delivered, e.g., by microinjection, electroporation, transient cell compression or squeezing (see, e.g., Lee 2012), lipid-mediated transfection, peptide-mediated delivery, or a combination thereof. gRNA molecules and/or RNA molecules encoding Cas9 molecules can be conjugated to molecules promoting uptake by the target cells (e.g., target cells described herein).

In certain embodiments, delivery via electroporation comprises mixing the cells with the gRNA molecules and/or RNA molecules encoding Cas9 molecules, with or without donor template nucleic acid molecules, in a cartridge, chamber, or cuvette, and applying one or more electrical impulses of defined duration and amplitude. In certain embodiments, delivery via electroporation is performed using a system in which cells are mixed with the gRNA molecules and/or RNA molecules encoding Cas9 molecules, with or without donor template nucleic acid molecules, in a vessel connected to a device (e.g., a pump) which feeds the mixture into a cartridge, chamber, or cuvette wherein one or more electrical impulses of defined duration and amplitude are applied, after which the cells are delivered to a second vessel. gRNA molecules and/or RNA molecules encoding Cas9 molecules can be conjugated to molecules to promote uptake by the target cells (e.g., target cells described herein).

Delivery of Cas9 molecules

Cas9 molecules can be delivered into cells by art-known methods or as described herein. For example, Cas9 protein molecules can be delivered, e.g., by microinjection,

electroporation, transient cell compression or squeezing (see, e.g., Lee 2012), lipid-mediated transfection, peptide-mediated delivery, or a combination thereof. Delivery can be accompanied by DNA encoding a gRNA or by a gRNA. Cas9 protein can be conjugated to molecules promoting uptake by the target cells (e.g., target cells described herein).

In certain embodiments, a Cas9 protein may be combined with a gRNA molecule to form a ribonucleoprotein (RNP) complex to be administered to a subject or delivered into a cell by art-known methods or as described herein. Direct delivery of Cas9/gRNA RNP complexes to cells eliminates the need to express from nucleic acid (e.g., transfection of plasmids encoding Cas9 and gRNA). It also eliminates unwanted integration of DNA segments derived from nucleic acid delivery (e.g., transfection of plasmids encoding Cas9 and gRNA). Therefore it is an alternative delivery approach which provides rapid action, fast turnover, high rate of on-target modification, reduced off-target effects, and less toxicity to cells. It can also be utilized to deliver the Cas9/gRNA complex to cells that are difficult to transfect (e.g., primary and pluripotent stem cells that are difficult to transfect). In certain embodiments, a Cas9/gRNA RNP complex may be formed prior to administration (i.e., pre-formed). In certain embodiments, multiple (e.g., more than one) Cas9/gRNA RNP complexes may be delivered (e.g., administered) simultaneously or sequentially. In certain embodiments, Cas9/gRNA RNP complexes may be delivered to cells by electroporation.

In certain embodiments, delivery via electroporation comprises mixing the cells with the Cas9 molecules, with or without gRNA molecules and/or donor template nucleic acids, in a cartridge, chamber, or cuvette, and applying one or more electrical impulses of defined duration and amplitude. In certain embodiments, delivery via electroporation is performed using a system in which cells are mixed with the Cas9 molecules with or without gRNA and/or donor template nucleic acids in a vessel connected to a device (e.g., a pump) which feeds the mixture into a cartridge, chamber, or cuvette wherein one or more electrical impulses of defined duration and amplitude are applied, after which the cells are delivered to a second vessel.

Route of administration of Cas system components

Systemic modes of administration include oral and parenteral routes. Parenteral routes include, by way of example, intravenous, intrarterial, intramuscular, intradermal, subcutaneous, intranasal, and intraperitoneal routes. Components administered systemically may be modified or formulated to target the components to cells of the blood and bone

marrow.

Local modes of administration include, by way of example, intra-bone marrow, intrathecal, and intra-cerebroventricular routes. In certain embodiments, significantly smaller amounts of the components (compared with systemic approaches) may exert an effect when administered locally compared to when administered systemically (for example, intravenously). Local modes of administration can reduce or eliminate the incidence of potentially toxic side effects that may occur when therapeutically effective amounts of a component are administered systemically.

In addition, components may be formulated to permit release over a prolonged period of time.

Ex vivo delivery of Cas system components

In certain embodiments, Cas system components described in **Table 3** are introduced into cells which are then introduced into a subject, e.g., the cells are removed from a subject, manipulated *ex vivo*, and reintroduced into the subject. Methods of introducing the components can include, e.g., any of the delivery methods described in **Table 4**.

Modified nucleosides, nucleotides, and nucleic acids

Modified nucleosides and modified nucleotides can be present in nucleic acids, e.g., particularly gRNA, but also other forms of RNA, e.g., mRNA, RNAi, or siRNA. As described herein, "nucleoside" is defined as a compound containing a five-carbon sugar molecule (a pentose or ribose) or derivative thereof, and an organic base, purine or pyrimidine, or a derivative thereof. As described herein, "nucleotide" is defined as a nucleoside further comprising a phosphate group.

Modified nucleosides and nucleotides can include one or more of:

(i) alteration, e.g., replacement, of one or both of the non-linking phosphate oxygens and/or of one or more of the linking phosphate oxygens in the phosphodiester backbone linkage;

(ii) alteration, e.g., replacement, of a constituent of the ribose sugar, e.g., of the 2' hydroxyl on the ribose sugar;

(iii) wholesale replacement of the phosphate moiety with "dephospho" linkers;

(iv) modification or replacement of a naturally occurring nucleobase;

(v) replacement or modification of the ribose-phosphate backbone;

(vi) modification of the 3' end or 5' end of the oligonucleotide, e.g., removal, modification or replacement of a terminal phosphate group or conjugation of a moiety; and

(vii) modification of the sugar.

The modifications listed above can be combined to provide modified nucleosides and nucleotides that can have two, three, four, or more modifications. For example, a modified nucleoside or nucleotide can have a modified sugar and a modified nucleobase. In an embodiment, every base of a gRNA is modified, e.g., all bases have a modified phosphate group, e.g., all are phosphorothioate groups. In an embodiment, all, or substantially all, of the phosphate groups of a unimolecular or modular gRNA molecule are replaced with phosphorothioate groups.

In an embodiment, modified nucleotides, e.g., nucleotides having modifications as described herein, can be incorporated into a nucleic acid, e.g., a "modified nucleic acid." In an embodiment, the modified nucleic acids comprise one, two, three or more modified nucleotides. In an embodiment, at least 5% (e.g., at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or about 100%) of the positions in a modified nucleic acid are a modified nucleotides.

Unmodified nucleic acids can be prone to degradation by, e.g., cellular nucleases. For example, nucleases can hydrolyze nucleic acid phosphodiester bonds. Accordingly, in one aspect the modified nucleic acids described herein can contain one or more modified nucleosides or nucleotides, e.g., to introduce stability toward nucleases.

In an embodiment, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can exhibit a reduced innate immune response when introduced into a population of cells, both *in vivo* and *ex vivo*. The term "innate immune response" includes a cellular response to exogenous nucleic acids, including single stranded nucleic acids, generally of viral or bacterial origin, which involves the induction of cytokine expression and release, particularly the interferons, and cell death. In an embodiment, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can disrupt binding of a major groove interacting partner with the nucleic acid. In an embodiment, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can exhibit a reduced innate immune response when introduced into a

population of cells, both *in vivo* and *ex vivo*, and also disrupt binding of a major groove interacting partner with the nucleic acid.

Definitions of chemical groups

As used herein, "alkyl" is meant to refer to a saturated hydrocarbon group which is straight-chained or branched. Example alkyl groups include methyl (Me), ethyl (Et), propyl (*e.g.*, n-propyl and isopropyl), butyl (*e.g.*, n-butyl, isobutyl, t-butyl), pentyl (*e.g.*, n-pentyl, isopentyl, neopentyl), and the like. An alkyl group can contain from 1 to about 20, from 2 to about 20, from 1 to about 12, from 1 to about 8, from 1 to about 6, from 1 to about 4, or from 1 to about 3 carbon atoms.

As used herein, "aryl" refers to monocyclic or polycyclic (*e.g.*, having 2, 3 or 4 fused rings) aromatic hydrocarbons such as, for example, phenyl, naphthyl, anthracenyl, phenanthrenyl, indanyl, indenyl, and the like. In an embodiment, aryl groups have from 6 to about 20 carbon atoms.

As used herein, "alkenyl" refers to an aliphatic group containing at least one double bond.

As used herein, "alkynyl" refers to a straight or branched hydrocarbon chain containing 2-12 carbon atoms and characterized in having one or more triple bonds. Examples of alkynyl groups include, but are not limited to, ethynyl, propargyl, and 3-hexynyl.

As used herein, "arylalkyl" or "aralkyl" refers to an alkyl moiety in which an alkyl hydrogen atom is replaced by an aryl group. Aralkyl includes groups in which more than one hydrogen atom has been replaced by an aryl group. Examples of "arylalkyl" or "aralkyl" include benzyl, 2-phenylethyl, 3-phenylpropyl, 9-fluorenyl, benzhydryl, and trityl groups.

As used herein, "cycloalkyl" refers to a cyclic, bicyclic, tricyclic, or polycyclic non-aromatic hydrocarbon groups having 3 to 12 carbons. Examples of cycloalkyl moieties include, but are not limited to, cyclopropyl, cyclopentyl, and cyclohexyl.

As used herein, "heterocyclyl" refers to a monovalent radical of a heterocyclic ring system. Representative heterocyclyls include, without limitation, tetrahydrofuranyl, tetrahydrothienyl, pyrrolidinyl, pyrrolidonyl, piperidinyl, pyrrolinyl, piperazinyl, dioxanyl, dioxolanyl, diazepinyl, oxazepinyl, thiazepinyl, and morpholinyl.

As used herein, "heteroaryl" refers to a monovalent radical of a heteroaromatic ring system. Examples of heteroaryl moieties include, but are not limited to, imidazolyl, oxazolyl,

thiazolyl, triazolyl, pyrrolyl, furanyl, indolyl, thiophenyl pyrazolyl, pyridinyl, pyrazinyl, pyridazinyl, pyrimidinyl, indolizinyl, purinyl, naphthyridinyl, quinolyl, and pteridinyl.

Phosphate backbone modifications

Phosphate group

In an embodiment, the phosphate group of a modified nucleotide can be modified by replacing one or more of the oxygens with a different substituent. Further, the modified nucleotide, e.g., modified nucleotide present in a modified nucleic acid, can include the wholesale replacement of an unmodified phosphate moiety with a modified phosphate as described herein. In an embodiment, the modification of the phosphate backbone can include alterations that result in either an uncharged linker or a charged linker with unsymmetrical charge distribution.

Examples of modified phosphate groups include, phosphorothioate, phosphoroselenates, borano phosphates, borano phosphate esters, hydrogen phosphonates, phosphoroamidates, alkyl or aryl phosphonates and phosphotriesters. In an embodiment, one of the non-bridging phosphate oxygen atoms in the phosphate backbone moiety can be replaced by any of the following groups: sulfur (S), selenium (Se), BR_3 (wherein R can be, e.g., hydrogen, alkyl, or aryl), C (e.g., an alkyl group, an aryl group, and the like), H, NR_2 (wherein R can be, e.g., hydrogen, alkyl, or aryl), or OR (wherein R can be, e.g., alkyl or aryl). The phosphorous atom in an unmodified phosphate group is achiral. However, replacement of one of the non-bridging oxygens with one of the above atoms or groups of atoms can render the phosphorous atom chiral; that is to say that a phosphorous atom in a phosphate group modified in this way is a stereogenic center. The stereogenic phosphorous atom can possess either the "R" configuration (herein Rp) or the "S" configuration (herein Sp).

Phosphorodithioates have both non-bridging oxygens replaced by sulfur. The phosphorus center in the phosphorodithioates is achiral which precludes the formation of oligoribonucleotide diastereomers. In an embodiment, modifications to one or both non-bridging oxygens can also include the replacement of the non-bridging oxygens with a group independently selected from S, Se, B, C, H, N, and OR (R can be, e.g., alkyl or aryl).

The phosphate linker can also be modified by replacement of a bridging oxygen, (i.e., the oxygen that links the phosphate to the nucleoside), with nitrogen (bridged phosphoroamidates), sulfur (bridged phosphorothioates) and carbon (bridged

methylenephosphonates). The replacement can occur at either linking oxygen or at both of the linking oxygens.

Replacement of the phosphate group

The phosphate group can be replaced by non-phosphorus containing connectors. In an embodiment, the charge phosphate group can be replaced by a neutral moiety.

Examples of moieties which can replace the phosphate group can include, without limitation, e.g., methyl phosphonate, hydroxylamino, siloxane, carbonate, carboxymethyl, carbamate, amide, thioether, ethylene oxide linker, sulfonate, sulfonamide, thioformacetal, formacetal, oxime, methyleneimino, methylenemethylimino, methylenehydrazo, methylenedimethylhydrazo and methyleneoxymethylimino.

Replacement of the ribophosphate backbone

Scaffolds that can mimic nucleic acids can also be constructed wherein the phosphate linker and ribose sugar are replaced by nuclease resistant nucleoside or nucleotide surrogates. In an embodiment, the nucleobases can be tethered by a surrogate backbone. Examples can include, without limitation, the morpholino, cyclobutyl, pyrrolidine and peptide nucleic acid (PNA) nucleoside surrogates.

Sugar modifications

The modified nucleosides and modified nucleotides can include one or more modifications to the sugar group. For example, the 2' hydroxyl group (OH) can be modified or replaced with a number of different "oxy" or "deoxy" substituents. In an embodiment, modifications to the 2' hydroxyl group can enhance the stability of the nucleic acid since the hydroxyl can no longer be deprotonated to form a 2'-alkoxide ion. The 2'-alkoxide can catalyze degradation by intramolecular nucleophilic attack on the linker phosphorus atom.

Examples of "oxy"-2' hydroxyl group modifications can include alkoxy or aryloxy (OR, wherein "R" can be, e.g., alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or a sugar); polyethyleneglycols (PEG), $O(CH_2CH_2O)_nCH_2CH_2OR$ wherein R can be, e.g., H or optionally substituted alkyl, and n can be an integer from 0 to 20 (e.g., from 0 to 4, from 0 to 8, from 0 to 10, from 0 to 16, from 1 to 4, from 1 to 8, from 1 to 10, from 1 to 16, from 1 to 20, from 2 to 4, from 2 to 8, from 2 to 10, from 2 to 16, from 2 to 20, from 4 to 8, from 4 to 10, from 4 to 16, and from 4 to 20). In an embodiment, the "oxy"-2' hydroxyl group modification can include "locked" nucleic acids (LNA) in which the 2' hydroxyl can be connected, e.g., by a C_{1-6} alkylene or C_{1-6} heteroalkylene bridge, to the 4' carbon of the same

ribose sugar, where exemplary bridges can include methylene, propylene, ether, or amino bridges; O-amino (wherein amino can be, e.g., NH_2; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino) and aminoalkoxy, $O(CH_2)_n$-amino, (wherein amino can be, e.g., NH_2; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino). In an embodiment, the "oxy"-2' hydroxyl group modification can include the methoxyethyl group (MOE), ($OCH_2CH_2OCH_3$, e.g., a PEG derivative).

"Deoxy" modifications can include hydrogen (i.e., deoxyribose sugars, e.g., at the overhang portions of partially ds RNA); halo (e.g., bromo, chloro, fluoro, or iodo); amino (wherein amino can be, e.g., NH_2; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, diheteroarylamino, or amino acid); $NH(CH_2CH_2NH)_nCH_2CH_2$-amino (wherein amino can be, e.g., as described herein), -NHC(O)R (wherein R can be, e.g., alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar), cyano; mercapto; alkyl-thio-alkyl; thioalkoxy; and alkyl, cycloalkyl, aryl, alkenyl and alkynyl, which may be optionally substituted with e.g., an amino as described herein.

The sugar group can also contain one or more carbons that possess the opposite stereochemical configuration than that of the corresponding carbon in ribose. Thus, a modified nucleic acid can include nucleotides containing e.g., arabinose, as the sugar. The nucleotide "monomer" can have an alpha linkage at the 1' position on the sugar, e.g., alpha-nucleosides. The modified nucleic acids can also include "abasic" sugars, which lack a nucleobase at C-1'. These abasic sugars can also be further modified at one or more of the constituent sugar atoms. The modified nucleic acids can also include one or more sugars that are in the L form, e.g., L-nucleosides.

Generally, RNA includes the sugar group ribose, which is a 5-membered ring having an oxygen. Exemplary modified nucleosides and modified nucleotides can include, without limitation, replacement of the oxygen in ribose (e.g., with sulfur (S), selenium (Se), or alkylene, such as, e.g., methylene or ethylene); addition of a double bond (e.g., to replace ribose with cyclopentenyl or cyclohexenyl); ring contraction of ribose (e.g., to form a 4-membered ring of cyclobutane or oxetane); ring expansion of ribose (e.g., to form a 6- or 7-membered ring having an additional carbon or heteroatom, such as for example, anhydrohexitol, altritol, mannitol, cyclohexanyl, cyclohexenyl, and morpholino that also has a phosphoramidate backbone). In an embodiment, the modified nucleotides can include

multicyclic forms (e.g., tricyclo; and "unlocked" forms, such as glycol nucleic acid (GNA) (e.g., R-GNA or S-GNA, where ribose is replaced by glycol units attached to phosphodiester bonds), threose nucleic acid (TNA, where ribose is replaced with α-L-threofuranosyl-(3'→2')).

Modifications on the nucleobase

The modified nucleosides and modified nucleotides described herein, which can be incorporated into a modified nucleic acid, can include a modified nucleobase. Examples of nucleobases include, but are not limited to, adenine (A), guanine (G), cytosine (C), and uracil (U). These nucleobases can be modified or wholly replaced to provide modified nucleosides and modified nucleotides that can be incorporated into modified nucleic acids. The nucleobase of the nucleotide can be independently selected from a purine, a pyrimidine, a purine or pyrimidine analog. In an embodiment, the nucleobase can include, for example, naturally-occurring and synthetic derivatives of a base.

Uracil

In an embodiment, the modified nucleobase is a modified uracil. Exemplary nucleobases and nucleosides having a modified uracil include without limitation pseudouridine (ψ), pyridin-4-one ribonucleoside, 5-aza-uridine, 6-aza-uridine, 2-thio-5-aza-uridine, 2-thio-uridine (s2U), 4-thio-uridine (s4U), 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxy-uridine (ho^5U), 5-aminoallyl-uridine, 5-halo-uridine (e.g., 5-iodo-uridine or 5-bromo-uridine), 3-methyl-uridine (m^3U), 5-methoxy-uridine (mo^5U), uridine 5-oxyacetic acid (cmo^5U), uridine 5-oxyacetic acid methyl ester ($mcmo^5U$), 5-carboxymethyl-uridine (cm^5U), 1-carboxymethyl-pseudouridine, 5-carboxyhydroxymethyl-uridine (chm^5U), 5-carboxyhydroxymethyl-uridine methyl ester ($mchm^5U$), 5-methoxycarbonylmethyl-uridine (mcm^5U), 5-methoxycarbonylmethyl-2-thio-uridine (mcm^5s2U), 5-aminomethyl-2-thio-uridine (nm^5s2U), 5-methylaminomethyl-uridine (mnm^5U), 5-methylaminomethyl-2-thio-uridine (mnm^5s2U), 5-methylaminomethyl-2-seleno-uridine (mnm^5se^2U), 5-carbamoylmethyl-uridine (ncm^5U), 5-carboxymethylaminomethyl-uridine ($cmnm^5U$), 5-carboxymethylaminomethyl-2-thio-uridine ($cmnm^5s2U$), 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-taurinomethyl-uridine (τcm^5U), 1-taurinomethyl-pseudouridine, 5-taurinomethyl-2-thio-uridine(τm^5s2U), 1-taurinomethyl-4-thio-pseudouridine, 5-methyl-uridine (m^5U, i.e., having the nucleobase deoxythymine), 1-methyl-pseudouridine ($m^1\psi$), 5-methyl-2-thio-uridine (m^5s2U), 1-methyl-4-thio-pseudouridine ($m^1s^4\psi$), 4-thio-1-methyl-

pseudouridine, 3-methyl-pseudouridine ($m^3\psi$), 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine (D), dihydropseudouridine, 5,6-dihydrouridine, 5-methyl-dihydrouridine (m^5D), 2-thio-dihydrouridine, 2-thio-dihydropseudouridine, 2-methoxy-uridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, 4-methoxy-2-thio-pseudouridine, N1-methyl-pseudouridine, 3-(3-amino-3-carboxypropyl)uridine (acp^3U), 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine ($acp^3\psi$), 5-(isopentenylaminomethyl)uridine (inm^5U), 5-(isopentenylaminomethyl)-2-thio-uridine (inm^5s2U), α-thio-uridine, 2'-O-methyl-uridine (Um), 5,2'-O-dimethyl-uridine (m^5Um), 2'-O-methyl-pseudouridine (ψm), 2-thio-2'-O-methyl-uridine (s2Um), 5-methoxycarbonylmethyl-2'-O-methyl-uridine (mcm^5Um), 5-carbamoylmethyl-2'-O-methyl-uridine (ncm^5Um), 5-carboxymethylaminomethyl-2'-O-methyl-uridine ($cmnm^5Um$), 3,2'-O-dimethyl-uridine (m^3Um), 5-(isopentenylaminomethyl)-2'-O-methyl-uridine (inm^5Um), 1-thio-uridine, deoxythymidine, 2'-F-ara-uridine, 2'-F-uridine, 2'-OH-ara-uridine, 5-(2-carbomethoxyvinyl) uridine, 5-[3-(1-E-propenylamino)uridine, pyrazolo[3,4-d]pyrimidines, xanthine, and hypoxanthine.

Cytosine

In an embodiment, the modified nucleobase is a modified cytosine. Exemplary nucleobases and nucleosides having a modified cytosine include without limitation 5-aza-cytidine, 6-aza-cytidine, pseudoisocytidine, 3-methyl-cytidine (m^3C), N4-acetyl-cytidine (act), 5-formyl-cytidine (f^5C), N4-methyl-cytidine (m^4C), 5-methyl-cytidine (m^5C), 5-halo-cytidine (e.g., 5-iodo-cytidine), 5-hydroxymethyl-cytidine (hm^5C), 1-methyl-pseudoisocytidine, pyrrolo-cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine (s2C), 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxy-cytidine, 2-methoxy-5-methyl-cytidine, 4-methoxy-pseudoisocytidine, 4-methoxy-1-methyl-pseudoisocytidine, lysidine (k^2C), α-thio-cytidine, 2'-O-methyl-cytidine (Cm), 5,2'-O-dimethyl-cytidine (m^5Cm), N4-acetyl-2'-O-methyl-cytidine (ac^4Cm), N4,2'-O-dimethyl-cytidine (m^4Cm), 5-formyl-2'-O-methyl-cytidine (f^5Cm), N4,N4,2'-O-trimethyl-cytidine ($m^4{}_2Cm$), 1-thio-cytidine, 2'-F-ara-cytidine, 2'-F-cytidine, and 2'-OH-ara-cytidine.

Adenine

In an embodiment, the modified nucleobase is a modified adenine. Exemplary nucleobases and nucleosides having a modified adenine include without limitation 2-amino-

purine, 2,6-diaminopurine, 2-amino-6-halo-purine (e.g., 2-amino-6-chloro-purine), 6-halo-purine (e.g., 6-chloro-purine), 2-amino-6-methyl-purine, 8-azido-adenosine, 7-deaza-adenosine, 7-deaza-8-aza-adenosine, 7-deaza-2-amino-purine, 7-deaza-8-aza-2-amino-purine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyl-adenosine (m^1A), 2-methyl-adenosine (m^2A), N6-methyl-adenosine (m^6A), 2-methylthio-N6-methyl-adenosine ($ms2m^6A$), N6-isopentenyl-adenosine (i^6A), 2-methylthio-N6-isopentenyl-adenosine (ms^2i^6A), N6-(cis-hydroxyisopentenyl)adenosine (io^6A), 2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine ($ms2io^6A$), N6-glycinylcarbamoyl-adenosine (g^6A), N6-threonylcarbamoyl-adenosine (t^6A), N6-methyl-N6-threonylcarbamoyl-adenosine (m^6t^6A), 2-methylthio-N6-threonylcarbamoyl-adenosine (ms^2g^6A), N6,N6-dimethyl-adenosine ($m^6{}_2A$), N6-hydroxynorvalylcarbamoyl-adenosine (hn^6A), 2-methylthio-N6-hydroxynorvalylcarbamoyl-adenosine ($ms2hn^6A$), N6-acetyl-adenosine (ac^6A), 7-methyl-adenosine, 2-methylthio-adenosine, 2-methoxy-adenosine, α-thio-adenosine, 2'-O-methyl-adenosine (Am), N^6,2'-O-dimethyl-adenosine (m^6Am), N^6-Methyl-2'-deoxyadenosine, N6,N6,2'-O-trimethyl-adenosine ($m^6{}_2Am$), 1,2'-O-dimethyl-adenosine (m^1Am), 2'-O-ribosyladenosine (phosphate) (Ar(p)), 2-amino-N6-methyl-purine, 1-thio-adenosine, 8-azido-adenosine, 2'-F-ara-adenosine, 2'-F-adenosine, 2'-OH-ara-adenosine, and N6-(19-amino-pentaoxanonadecyl)-adenosine.

Guanine

In an embodiment, the modified nucleobase is a modified guanine. Exemplary nucleobases and nucleosides having a modified guanine include without limitation inosine (I), 1-methyl-inosine (m^1I), wyosine (imG), methylwyosine (mimG), 4-demethyl-wyosine (imG-14), isowyosine (imG2), wybutosine (yW), peroxywybutosine (o_2yW), hydroxywybutosine (OHyW), undermodified hydroxywybutosine (OHyW*), 7-deaza-guanosine, queuosine (Q), epoxyqueuosine (oQ), galactosyl-queuosine (galQ), mannosyl-queuosine (manQ), 7-cyano-7-deaza-guanosine ($preQ_0$), 7-aminomethyl-7-deaza-guanosine ($preQ_1$), archaeosine (G^+), 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine (m^7G), 6-thio-7-methyl-guanosine, 7-methyl-inosine, 6-methoxy-guanosine, 1-methyl-guanosine (m'G), N2-methyl-guanosine (m^2G), N2,N2-dimethyl-guanosine ($m^2{}_2G$), N2,7-dimethyl-guanosine (m^2,7G), N2, N2,7-dimethyl-guanosine (m^2,2,7G), 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-methyl-6- thio-guanosine, N2-methyl-6-thio-guanosine, N2,N2-dimethyl-6-thio-guanosine, α-thio-guanosine, 2'-O-methyl-guanosine (Gm), N2-methyl-2'-O-methyl-guanosine (m^2Gm),

N2,N2-dimethyl-2'-O-methyl-guanosine ($m^2{}_2$Gm), 1-methyl-2'-O-methyl-guanosine (m'Gm), N2,7-dimethyl-2'-O-methyl-guanosine (m^2,7Gm), 2'-O-methyl-inosine (Im), 1,2'-O-dimethyl-inosine (m'Im), O^6-phenyl-2'-deoxyinosine, 2'-O-ribosylguanosine (phosphate) (Gr(p)), 1-thio-guanosine, O^6-methyl-guanosine, O^6-Methyl-2'-deoxyguanosine, 2'-F-ara-guanosine, and 2'-F-guanosine.

Exemplary modified gRNAs

In certain embodiments, modified nucleic acids as described herein can be modified gRNAs. It is to be understood that any of the gRNAs described herein can be modified as described herein.

Through experimentation (results not shown), it has been found that the gRNA component of the CRISPR/Cas system is more efficient at editing genes in T cells when the gRNA is modified at or near its 5' end (e.g., when the 5' end of the gRNA is modified by inclusion of a eukaryotic mRNA cap structure or cap analog. While not wishing to be bound by theory, it is believed that these and other modified gRNAs described herein elicit a reduced innate immune response from certain circulatory cell types (e.g., T cells), and that this reduced response may be responsible for the observed improvements. The present invention encompasses the realization that minimizing the innate immune response of circulating cells (e.g., T cells) to gRNAs could be advantageous when using gRNAs to edit circulating cells (whether *ex vivo* or *in vivo*), and could also be advantageous when using gRNAs to edit non-circulating cells, e.g., when a gRNA is administered systemically or locally for *in vivo* gene editing purposes. The present invention also encompasses the realization that the improvements observed with a 5' capped gRNA can be extended to gRNAs that have been modified in other ways to achieve the same type of structural or functional result (e.g., by the inclusion of modified nucleosides or nucleotides, or when an *in vitro* transcribed gRNA is modified by treat$_m$ent with a phosphatase such as calf intestinal alkaline phosphatase to remove the 5' triphosphate group). While not wishing to be bound by theory, in certain embodiments the modified gRNAs described herein may contain one or more modifications (e.g., modified nucleosides or nucleotides) which introduce stability toward nucleases (e.g., by the inclusion of modified nucleosides or nucleotides and/or a 3' polyA tract).

Accordingly, in certain embodiments the compositions and methods provided herein utilize gRNAs that include one or more modified nucleosides or nucleotides as described herein. In certain of these embodiments, the inclusion of the one or more modified

nucleosides or nucleotides causes the gRNA to elicit a reduced innate immune response in certain circulating cell types (e.g., T cells, macrophages, dendritic cells, and/or B cells) as compared to an otherwise unmodified gRNA.

In certain embodiments, a gRNA for use in the compositions and methods provided herein is modified at or near its 5' end (e.g., within 1-10, 1-5, or 1-2 nucleotides of its 5' end). In certain embodiments, the gRNA is modified by inclusion of a eukaryotic mRNA cap structure or cap analog (e.g., a *G(5')ppp(5')G* cap analog, a *m7G(5')ppp(5')G* cap analog, or a *3'-O-Me-m7G(5')ppp(5')G* anti reverse cap analog (ARCA)). The cap or cap analog can incorporated during chemical synthesis or *in vitro* transcription of the gRNA. In certain embodiments, an *in vitro* transcribed gRNA is modified by treat$_m$ent with a phosphatase (e.g., calf intestinal alkaline phosphatase) to remove the 5' triphosphate group.

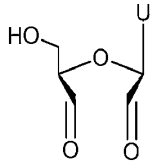
In certain embodiments, a gRNA for use in the compositions and methods provided herein is modified at or near its 3' end (e.g., within 1-10, 1-5, or 1-2 nucleotides of its 3' end).

In an embodiment, the 3' end of a gRNA is modified by the addition of one or more (e.g., 25-200) adenine (A) residues. The polyA tract can be contained in the nucleic acid (e.g., plasmid, PCR product, viral genome) encoding the gRNA, or can be added to the gRNA during chemical synthesis, or following *in vitro* transcription using a polyadenosine polymerase (e.g., *E. coli* Poly(A)Polymerase).

In certain embodiments, a gRNA for use in the compositions and methods provided herein comprises both a modification at or near its 5' end and a modification at or near its 3' end.

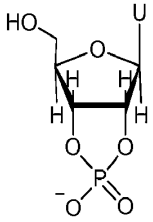
In certain embodiments, *in vitro* transcribed gRNA contains both a 5' cap structure or cap analog and a 3' polyA tract. In an embodiment, an *in vitro* transcribed gRNA is modified by treat$_m$ent with a phosphatase (e.g., calf intestinal alkaline phosphatase) to remove the 5' triphosphate group and comprises a 3' polyA tract.

In some embodiments, gRNAs can be modified at a 3' terminal U ribose. For example, the two terminal hydroxyl groups of the U ribose can be oxidized to aldehyde groups and a concomitant opening of the ribose ring to afford a modified nucleoside as shown below:



wherein "U" can be an unmodified or modified uridine.

In another embodiment, the 3' terminal U can be modified with a 2'3' cyclic phosphate as shown below:



wherein "U" can be an unmodified or modified uridine.

In some embodiments, the gRNA molecules may contain 3' nucleotides which can be stabilized against degradation, e.g., by incorporating one or more of the modified nucleotides described herein. In this embodiment, e.g., uridines can be replaced with modified uridines, e.g., 5-(2-amino)propyl uridine, and 5-bromo uridine, or with any of the modified uridines described herein; adenosines and guanosines can be replaced with modified adenosines and guanosines, e.g., with modifications at the 8-position, e.g., 8-bromo guanosine, or with any of the modified adenosines or guanosines described herein.

In some embodiments, sugar-modified ribonucleotides can be incorporated into the gRNA, e.g., wherein the 2' OH-group is replaced by a group selected from H, -OR, -R (wherein R can be, e.g., alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar), halo, -SH, -SR (wherein R can be, e.g., alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar), amino (wherein amino can be, e.g., NH_2; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, diheteroarylamino, or amino acid); or cyano (-CN). In some embodiments, the phosphate backbone can be modified as described herein, e.g., with a phosphothioate group. In some embodiments, one or more of the nucleotides of the gRNA can each independently be a modified or unmodified nucleotide including, but not limited to 2'-sugar modified, such as, 2'-O-methyl, 2'-O-methoxyethyl, or 2'-Fluoro modified including, e.g., 2'-F or 2'-O-methyl, adenosine (A), 2'-F or 2'-O-methyl, cytidine (C), 2'-F or 2'-O-methyl, uridine (U), 2'-F or 2'-O-methyl, thymidine (T), 2'-F or 2'-O-methyl, guanosine (G), 2'-O-methoxyethyl-5-methyluridine (Teo), 2'-O-methoxyethyladenosine (Aeo), 2'-O-methoxyethyl-5-methylcytidine (m5Ceo), and any combinations thereof.

In some embodiments, a gRNA can include "locked" nucleic acids (LNA) in which the 2' OH-group can be connected, e.g., by a C1-6 alkylene or C1-6 heteroalkylene bridge, to the 4' carbon of the same ribose sugar, where exemplary bridges can include methylene, propylene, ether, or amino bridges; O-amino (wherein amino can be, e.g., NH_2; alkylamino,

dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino) and aminoalkoxy or $O(CH_2)_n$-amino (wherein amino can be, e.g., NH_2; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino).

In some embodiments, a gRNA can include a modified nucleotide which is multicyclic (e.g., tricyclo; and "unlocked" forms, such as glycol nucleic acid (GNA) (e.g., R-GNA or S-GNA, where ribose is replaced by glycol units attached to phosphodiester bonds), or threose nucleic acid (TNA, where ribose is replaced with α-L-threofuranosyl-(3'→2')).

Generally, gRNA molecules include the sugar group ribose, which is a 5-membered ring having an oxygen. Exemplary modified gRNAs can include, without limitation, replacement of the oxygen in ribose (e.g., with sulfur (S), selenium (Se), or alkylene, such as, e.g., methylene or ethylene); addition of a double bond (e.g., to replace ribose with cyclopentenyl or cyclohexenyl); ring contraction of ribose (e.g., to form a 4-membered ring of cyclobutane or oxetane); ring expansion of ribose (e.g., to form a 6- or 7-membered ring having an additional carbon or heteroatom, such as for example, anhydrohexitol, altritol, mannitol, cyclohexanyl, cyclohexenyl, and morpholino that also has a phosphoramidate backbone). Although the majority of sugar analog alterations are localized to the 2' position, other sites are amenable to modification, including the 4' position. In an embodiment, a gRNA comprises a 4'-S, 4'-Se or a 4'-C-aminomethyl-2'-O-Me modification.

In some embodiments, deaza nucleotides, e.g., 7-deaza-adenosine, can be incorporated into the gRNA. In some embodiments, O- and N-alkylated nucleotides, e.g., N6-methyl adenosine, can be incorporated into the gRNA. In some embodiments, one or more or all of the nucleotides in a gRNA molecule are deoxynucleotides.

miRNA binding sites

microRNAs (or miRNAs) are naturally occurring cellular 19-25 nucleotide long noncoding RNAs. They bind to nucleic acid molecules having an appropriate miRNA binding site, e.g., in the 3' UTR of an mRNA, and down-regulate gene expression. While not wishing to be bound by theory, it is believed that this down regulation occurs by either reducing nucleic acid molecule stability or inhibiting translation. An RNA species disclosed herein, e.g., an mRNA encoding Cas9, can comprise an miRNA binding site, e.g., in its 3'UTR. The miRNA binding site can be selected to promote down regulation of expression is a selected cell type.

EXAMPLES

The following Examples are merely illustrative and are not intended to limit the scope or content of the invention in any way.

Example 1: Biophysical characterization and direct delivery of Cas9 ribonucleoprotein complexes

Direct delivery of Cas9 ribonucleoprotein (RNP) complexes allows for efficient gene editing while minimizing off-target activity owing to the rapid turnover of the Cas9 protein in cells. The efficiency of gene editing mediated by RNP delivery varies by locus, and depends on the length of gRNA and the amount and ratio of Cas9 protein and gRNA delivered.

Structural and biophysical characterization of Cas9 complexes with gRNA revealed a large contact area and a high affinity. Thermal melt curves are a useful property to detect the binding and stability of complexes. The large increase in melting temperature from an apo-Cas9 molecule (i.e., a Cas9 molecule in the absence of gRNA molecule) to a Cas9 molecule complexed with gRNA was used to characterize the affinity of Cas9 for gRNA. Multiple gRNAs of differing lengths were complexed with Cas9 at different stoichiometries, and the interaction was measured using thermal shift (e.g., the shift of melting temperature). These biophysically characterized complexes were then transfected into 293T cells, and the efficiency of indel generated was measured. Subtle differences in gRNA length and base composition was shown to affect the binding and formation of RNP complex. Correlating binding affinity with efficiency of genome editing allows for the design of an optimal composition of RNPs, e.g., for cationic lipid mediated direct delivery.

Evaluation of Cas9 molecule and Cas9 molecule/gRNA molecule complexes by DSF

S. aureus and *S. pyogenes* Cas9 molecules were recombinantly expressed and purified using Ni affinity chromatography, SP Sepharose, and Superdex 200.

DSF was used to examine the stability of purified *S. aureus* and *S. pyogenes* Cas9 molecules in the absence of gRNA molecules. The reaction mix contained 5 μM Cas9 molecules and 5x SYPRO Orange® (Life Technologies cat#S-6650) in 10 μL volume. The gradient was run at 20°C for 1 minute, then from 20°C to 95°C with 1°C increments every 10 seconds. The derivative of the fluorescent signal was plotted against temperature, and the temperature midpoint for the unfolding transition (T_m) was determined. As shown in **Figs.**

9A and **9B**, the *S. aureus* Cas9 molecule (T_m = 36°C) was less stable than the *S. pyogenes* Cas9 molecule (T_m = 40°C).

DSF was also used to examine Cas9 molecule/gRNA molecule complexes. *S. pyogenes* Cas9 alone, *S. pyogenes* Cas9 with 1 μM *S. pyogenes* gRNA, and *S. pyogenes* Cas9 with 1 μM *S. aureus* gRNA were tested in H150 (10 mM Hepes pH 7.5, 150 mM NaCl). The DNA sequences encoding the gRNA molecules used in these experiments were:

S. pyogenes gRNA:

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTT (SEQ ID NO:206); and

S. aureus gRNA:

GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTAT CTCGTCAACTTGTTGGCGAGATTTT (SEQ ID NO:207).

As shown in **Fig. 10**, thermal denaturation of the *S. pyogenes* Cas9 was 42°C, and a thermal shift to 50°C was observed in the presence of the concomitant gRNA molecule. However, a shift was not observed when the *S. pyogenes* Cas9 was incubated with an alien gRNA molecule. These results suggest that Cas9 molecule/gRNA molecule complex formation may correlate with thermal shift.

Delivery of Cas9 molecule/gRNA molecule complexes to Jurkat T cells

Cas9 and gRNA were at a 1:25 ratio. gRNA construct was generated using a PCR protocol. gRNA was *in vitro* transcribed and capped (e.g., 5' Anti-Reverse Cap Analog (ARCA) cap) and tailed (e.g., a 3' polyA tail). As shown in **Fig. 11A**, gRNA binding induced T_m shift from 32°C to 46°C. The Cas9/gRNA complex was delivered to Jurkat T cells. **Fig. 11B** indicates that about 20% of the cells loss the CD3 marker.

Example 2: Comparison of gRNA molecules from *in vitro* transcription or chemical synthesis

DSF was used to assay the successful formation of Cas9 molecule/gRNA molecule complex, and the quality and integrity of gRNA molecules obtained from three methods or suppliers were compared.

gRNA molecules with sequences corresponding to those in **Table 6** were obtained at the 15-50 μg scale.

Purified Cas9 protein at 4 μM was complexed with an equimolar amount of gRNA at room temperature in H150 buffer (10mM Hepes pH7.5, 150mM NaCl). Part of this reaction

containing Cas9 molecule/gRNA molecule complex (RNP) was then transfected using lipofectamine-2000 transfection reagent into HEK293FT cells. RNP at 100 nM was used in all cases, regardless of guide source. In each case, the remainder of RNP was diluted to 1 μM in H150 buffer, SYPRO orange was added to a final concentration of 5X from a 5000X stock, and the DSF assay was performed according to assay 1 described herein. Indel quantitation was performed according to protocols described herein.

Results of the experiment are summarized in **Table 6**. Delta T_m*s* were compared to Cas9 protein melting at 42°C. The results showed that samples which indicated complete RNP complex formation as evidenced by a Delta T_m of 8-9°C all showed good NHEJ activity in HEK293FT cells.

Chemically synthesized gRNAs from Company 2 were found inadequate, with negligible Delta T_m*s* and lower NHEJ activity (11% indels).

In vitro transcribed gRNAs by the MEGAshortscript T7 kit and gRNAs purchased from Company 1 were of sufficient quality and integrity as demonstrated by 7-8°C T_m*s* and NHEJ activity in the 22-27°C range.

Example 3: Cloning and initial screening of gRNAs

This example discloses a method for evaluating chimeric gRNAs. The same approach may also be used to evaluate modular gRNAs.

Cloning gRNAs into vectors

For each gRNA, a pair of overlapping oligonucleotides is designed and obtained. Oligonucleotides are annealed and ligated into a digested vector backbone containing an upstream U6 promoter and the remaining sequence of a long chimeric gRNA. Plasmids are sequence-verified and prepped to generate sufficient amounts of transfection-quality DNA. In certain embodiments, the U6 promoter may be replaced with an alternate promoter to drive *in vivo* transcription (e.g., H1 promoter) or *in vitro* transcription (e.g., a T7 promoter).

Cloning gRNAs into linear dsDNA molecules (STITCHR)

A single oligonucleotide is designed and obtained for each gRNA. The U6 promoter and the gRNA scaffold (e.g., including everything except the targeting domain, e.g., including sequences derived from the crRNA and tracrRNA, e.g., including a first complementarity domain; a linking domain; a second complementarity domain; a proximal domain; and a tail domain) are separately PCR amplified and purified as dsDNA molecules. The gRNA-specific oligonucleotide is used in a PCR reaction to stitch together the U6 and

the gRNA scaffold, linked by the targeting domain specified in the oligonucleotide. The resulting dsDNA molecules (STITCHR products) are purified for transfection. Any gRNA scaffold may be used to create gRNAs compatible with Cas9s from any bacterial species. In certain embodiments, the U6 promoter may be replaced with an alternate promoter to drive *in vivo* transcription (e.g., H1 promoter) or *in vitro* transcription (e.g., T7 promoter).

Initial gRNA screen

Each gRNA to be tested is transfected, along with a plasmid expressing Cas9 and a small amount of a GFP-expressing plasmid, into human cells. In preliminary experiments, these cells can be immortalized human cell lines such as 293T, K562 or U2OS. Alternatively, primary human cells may be used. The cells used for screening may be relevant to the eventual therapeutic cell target (e.g., an erythroid cell). The use of primary cells similar to the potential therapeutic target cell population may provide important information on gene targeting rates in the context of endogenous chromatin and gene expression.

Transfection may be performed using lipid transfection (such as Lipofectamine or Fugene) or by electroporation (such as Lonza Nucleofection). Following transfection, GFP expression can be determined either by fluorescence microscopy or by flow cytometry to confirm consistent and high levels of transfection. Preliminary transfections can comprise different gRNAs and different targeting approaches (e.g., 17-mers, 20-mers, nuclease, dual-nickase, etc.) to determine which gRNAs/combinations of gRNAs give the greatest activity.

Efficiency of cleavage with each gRNA may be assessed by measuring NHEJ-induced indel formation at the target locus by T7E1 endonuclease assay. For this assay, PCR amplicons are approximately 500-700 bp, with the intended cut site placed asymmetrically in the amplicon. Following amplification, purification, and size-verification of PCR products, DNA is denatured and re-hybridized by heating to 95°C and then slowly cooling. Hybridized PCR products are then digested with T7 Endonuclease I (or other mismatch-sensitive enzyme), which recognizes and cleaves non-perfectly matched DNA. If indels are present in the original template DNA, denaturation and re-annealing of the amplicons results in hybridization of DNA strands harboring different indels, leading to double-stranded DNA that is not perfectly matched. Digestion products may be visualized by gel electrophoresis or capillary electrophoresis. The fraction of DNA that is cleaved (density of cleavage products divided by the density of cleaved and uncleaved) may be used to estimate percent NHEJ

using the following equation: %NHEJ = (1-(1-fraction cleaved)$^{1/2}$). The T7E1 assay is sensitive down to about 2-5% NHEJ.

In certain embodiments, other methods may be used to assess cleavage efficiency, including for example sequencing and use of mismatch-sensitive enzymes, e.g., CelI/Surveyor nuclease. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed, miniprepped and sequenced with a single primer. Sanger sequencing may be used for determining the exact nature of indels after determining the NHEJ rate by T7E1. For next-generation sequencing, amplicons may be 300-500 bp, with the intended cut site placed asymmetrically. Following PCR, next-generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, e.g., for use in high throughput sequencing (for example on an Illumina MiSeq). This method allows for detection of very low NHEJ rates.

Example 4: Assessment of gene targeting by NHEJ

The gRNAs that induce the greatest levels of NHEJ in initial tests can be selected for further evaluation of gene targeting efficiency. In this case, cells are derived from disease subjects and, therefore, harbor the relevant mutation.

Following transfection (usually 2-3 days post-transfection,) genomic DNA may be isolated from a bulk population of transfected cells and PCR may be used to amplify the target region. Following PCR, gene targeting efficiency to generate the desired mutations (either knockout of a target gene or removal of a target sequence motif) may be determined by sequencing. For Sanger sequencing, PCR amplicons may be 500-700 bp long. For next generation sequencing, PCR amplicons may be 300-500 bp long. If the goal is to knockout gene function, sequencing may be used to assess what percent of alleles have undergone NHEJ-induced indels that result in a frameshift or large deletion or insertion that would be expected to destroy gene function. If the goal is to remove a specific sequence motif, sequencing may be used to assess what percent of alleles have undergone NHEJ-induced deletions that span this sequence.

Example 5: Assessment of gene targeting by HDR

The gRNAs that induce the greatest levels of NHEJ in initial tests can be selected for further evaluation of gene targeting efficiency. In this case, cells are derived from disease subjects and, therefore, harbor the relevant mutation.

Following transfection (usually 2-3 days post-transfection,) genomic DNA may be isolated from a bulk population of transfected cells and PCR may be used to amplify the target region. Following PCR, gene targeting efficiency can be determined by several methods.

Determination of gene targeting frequency involves measuring the percentage of alleles that have undergone homologous directed repair (HDR) with the exogenously provided donor template or endogenous genomic donor sequence and which therefore have incorporated the desired correction. If the desired HDR event creates or destroys a restriction enzyme site, the frequency of gene targeting may be determined by a RFLP assay. If no restriction site is created or destroyed, sequencing may be used to determine gene targeting frequency. If a RFLP assay is used, sequencing may still be used to verify the desired HDR event and ensure that no other mutations are present. If an exogenously provided donor template is employed, at least one of the primers is placed in the endogenous gene sequence outside of the region included in the homology arms, which prevents amplification of donor template still present in the cells. Therefore, the length of the homology arms present in the donor template may affect the length of the PCR amplicon. PCR amplicons can either span the entire donor region (both primers placed outside the homology arms) or they can span only part of the donor region and a single junction between donor and endogenous DNA (one internal and one external primer). If the amplicons span less than the entire donor region, two different PCRs should be used to amplify and sequence both the 5' and the 3' junction.

If the PCR amplicon is short (less than 600 bp) it is possible to use next generation sequencing. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, e.g., for use in high throughput sequencing (for example on an Illumina MiSeq). This method allows for detection of very low gene targeting rates.

If the PCR amplicon is too long for next generation sequencing, Sanger sequencing can be performed. For Sanger sequencing, purified PCR amplicons will be cloned into a plasmid backbone (for example, TOPO cloned using the LifeTech Zero Blunt® TOPO® cloning kit), transformed, miniprepped and sequenced.

The same or similar assays described above can be used to measure the percentage of alleles that have undergone HDR with endogenous genomic donor sequence and which therefore have incorporated the desired correction.

Incorporation by Reference

All publications, patents, and patent applications mentioned herein are hereby incorporated by reference in their entirety as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

References

Anders et al. Nature 513(7519):569-573 (2014)

Bae et al. Bioinformatics 30(10):1473-1475 (2014)

Caldecott Nat Rev Genet 9(8):619-631 (2008)

Chylinski et al. RNA Biol 10(5):726–737 (2013)

Cong et al. Science 399(6121):819-823 (2013)

Deveau et al. J Bacterial 190(4):1390-1400 (2008)

Esvelt et al. Nature 472(7344):499-503 (2011)

Fu et al. Nat Biotechnol 32:279-284 (2014)

Haft PLoS Comput Biol 1(6):e60 (2005)

Heigwer Nat Methods 11(2):122-123 (2014)

Horvath & Barrangou Science 327(5962):167-170 (2010)

Hsu et al. Nat Biotechnol 31(9):827-832 (2013)

Jinek et al. Science 337(6096):816-821 (2012)

Jinek et al. Science 343(6176):1247997 (2014)

Lee et al. Nano Lett 12(12):6322-6327 (2012)

Li Cell Res 18(1):85-98 (2008)

Makarova et al. Nat Rev Microbiol 9(6):467-477 (2011)

Mali et al. Science 339(6121):823-826 (2013)

Marteijn et al. Nat Rev Mol Cell Biol 15(7):465-481 (2014)

Nishimasu et al. Cell 156(5):935-949 (2014)

Sternberg et al. Nature 507(7490):62-67 (2014)

Wang et al. Cell 153(4):910-918 (2013)

Xiao et al. "CasOT: a genome-wide Cas9/gRNA off-target searching tool." Bioinformatics (epub Jan. 21, 2014)

Table 1: Lipids used for gene transfer

Lipid	Abbreviation	Feature
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
N-[1-(2,3-Dioleyloxy)propyl]*N,N,N*-trimethylammonium chloride	DOT_mA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Dioctadecylamidoglycylspermine	DOGS	Cationic
N-(3-Aminopropyl)-*N,N*-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleyloxy-*N*-[2(sperminecarboxamido-ethyl]-*N,N*-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleyl-3-trimethylammonium-propane	DOPA	Cationic
N-(2-Hydroxyethyl)-*N,N*-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[*N*-(*N',N'*-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Diodeoxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Dioctadecylamidoglicylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammonium bromide	CLIP-6	Cationic
Ethyldimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearyloxy-*N,N*-dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
O,O'-Dimyristyl-N-lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylphosphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
N-t-Butyl-*N*0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl] imidazolinium chloride	DOTIM	Cationic
*N*1-Cholesteryloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-*N*-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleyloxy-3- dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]- dioxolane	DLin-KC2-DMA	Cationic
dilinoleyl- methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Table 2: Cas systems

Gene name[‡]	System type or subtype	Name from Haft 2005[§]	Structure of encoded protein (PDB accessions)[¶]	Families (and superfamily) of encoded protein[#**]	Representatives
cas1	• Type I • Type II • Type III	*cas1*	3GOD, 3LFX and 2YZS	COG1518	SERP2463, SPy1047 and *ygbT*
cas2	• Type I • Type II • Type III	*cas2*	2IVY, 2I8E and 3EXC	COG1343 and COG3512	SERP2462, SPy1048, SPy1723 (N-terminal domain) and *ygbF*
cas3'	• Type I[‡‡]	*cas3*	NA	COG1203	APE1232 and *ygcB*
cas3''	• Subtype I-A • Subtype I-B	NA	NA	COG2254	APE1231 and BH0336
cas4	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-D • Subtype II-B	*cas4* and *csa1*	NA	COG1468	APE1239 and BH0340
cas5	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-E	*cas5a*, *cas5d*, *cas5e*, *cas5h*, *cas5p*, *cas5t* and *cmx5*	3KG4	COG1688 (RAMP)	APE1234, BH0337, *devS* and *ygcI*
cas6	• Subtype I-A • Subtype I-B • Subtype I-D • Subtype III-A• Subtype III-B	*cas6* and *cmx6*	3I4H	COG1583 and COG5551 (RAMP)	PF1131 and slr7014
cas6e	• Subtype I-E	*cse3*	1WJ9	(RAMP)	*ygcH*
cas6f	• Subtype I-F	*csy4*	2XLJ	(RAMP)	y1727
cas7	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-E	*csa2*, *csd2*, *cse4*, *csh2*, *csp1* and *cst2*	NA	COG1857 and COG3649 (RAMP)	*devR* and *ygcJ*
cas8a1	• Subtype I-A[‡‡]	*cmx1*, *cst1*, *csx8*, *csx13* and CXXC-CXXC	NA	BH0338-like	LA3191[§§] and PG2018[§§]
cas8a2	• Subtype I-A[‡‡]	*csa4* and *csx9*	NA	PH0918	AF0070, AF1873, MJ0385, PF0637, PH0918 and SSO1401
cas8b	• Subtype I-B[‡‡]	*csh1* and T_m1802	NA	BH0338-like	MTH1090 and T_m1802
cas8c	• Subtype I-C[‡‡]	*csd1* and *csp2*	NA	BH0338-like	BH0338
cas9	• Type II[‡‡]	*csn1* and *csx12*	NA	COG3513	FTN_0757 and SPy1046

Gene name[‡]	System type or subtype	Name from Haft 2005[§]	Structure of encoded protein (PDB accessions)[¶]	Families (and superfamily) of encoded protein[#**]	Representatives
cas10	• Type III[‡‡]	*cmr2*, *csm1* and *csx11*	NA	COG1353	MTH326, Rv2823c[§§] and T_m1794[§§]
cas10d	• Subtype I-D[‡‡]	*csc3*	NA	COG1353	slr7011
csy1	• Subtype I-F[‡‡]	*csy1*	NA	y1724-like	y1724
csy2	• Subtype I-F	*csy2*	NA	(RAMP)	y1725
csy3	• Subtype I-F	*csy3*	NA	(RAMP)	y1726
cse1	• Subtype I-E[‡‡]	*cse1*	NA	YgcL-like	*ygcL*
cse2	• Subtype I-E	*cse2*	2ZCA	YgcK-like	*ygcK*
csc1	• Subtype I-D	*csc1*	NA	alr1563-like (RAMP)	alr1563
csc2	• Subtype I-D	*csc1* and *csc2*	NA	COG1337 (RAMP)	slr7012
csa5	• Subtype I-A	*csa5*	NA	AF1870	AF1870, MJ0380, PF0643 and SSO1398
csn2	• Subtype II-A	*csn2*	NA	SPy1049-like	SPy1049
csm2	• Subtype III-A[‡‡]	*csm2*	NA	COG1421	MTH1081 and SERP2460
csm3	• Subtype III-A	*csc2* and *csm3*	NA	COG1337 (RAMP)	MTH1080 and SERP2459
csm4	• Subtype III-A	*csm4*	NA	COG1567 (RAMP)	MTH1079 and SERP2458
csm5	• Subtype III-A	*csm5*	NA	COG1332 (RAMP)	MTH1078 and SERP2457
csm6	• Subtype III-A	APE2256 and *csm6*	2WTE	COG1517	APE2256 and SSO1445
cmr1	• Subtype III-B	*cmr1*	NA	COG1367 (RAMP)	PF1130
cmr3	• Subtype III-B	*cmr3*	NA	COG1769 (RAMP)	PF1128
cmr4	• Subtype III-B	*cmr4*	NA	COG1336 (RAMP)	PF1126
cmr5	• Subtype III-B[‡‡]	*cmr5*	2ZOP and 2OEB	COG3337	MTH324 and PF1125
cmr6	• Subtype III-B	*cmr6*	NA	COG1604 (RAMP)	PF1124
csb1	• Subtype I-U	GSU0053	NA	(RAMP)	Balac_1306 and GSU0053
csb2	• Subtype I-U[§§]	NA	NA	(RAMP)	Balac_1305 and GSU0054

Gene name[‡]	System type or subtype	Name from Haft 2005[§]	Structure of encoded protein (PDB accessions)[¶]	Families (and superfamily) of encoded protein[#,**]	Representatives
csb3	• Subtype I-U	NA	NA	(RAMP)	Balac_1303[§§]
csx17	• Subtype I-U	NA	NA	NA	Btus_2683
csx14	• Subtype I-U	NA	NA	NA	GSU0052
csx10	• Subtype I-U	*csx10*	NA	(RAMP)	Caur_2274
csx16	• Subtype III-U	VVA1548	NA	NA	VVA1548
csaX	• Subtype III-U	*csaX*	NA	NA	SSO1438
csx3	• Subtype III-U	*csx3*	NA	NA	AF1864
csx1	• Subtype III-U	*csa3*, *csx1*, *csx2*, DXTHG, NE0113 and TIGR02710	1XMX and 2I71	COG1517 and COG4006	MJ1666, NE0113, PF1127 and T_m1812
csx15	• Unknown	NA	NA	TTE2665	TTE2665
csf1	• Type U	*csf1*	NA	NA	AFE_1038
csf2	• Type U	*csf2*	NA	(RAMP)	AFE_1039
csf3	• Type U	*csf3*	NA	(RAMP)	AFE_1040
csf4	• Type U	*csf4*	NA	NA	AFE_1037

Table 3: Component formulation, delivery, and administration strategies

Elements			
Cas9 Molecule(s)	**gRNA Molecule(s)**	**Donor Template Nucleic Acid**	**Comments**
DNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
DNA		DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA, here from a single molecule. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided on the same DNA molecule that encodes the Cas9.
DNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the Cas9.

mRNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is provided as in vitro transcribed or synthesized RNA. In this embodiment, the donor template is provided as a DNA molecule.
mRNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided as a separate DNA molecule.
mRNA	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from in vitro transcribed mRNA, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
Protein	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is provided as a protein, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided as a separate DNA molecule.
Protein	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is provided as a protein, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA.
Protein	RNA	DNA	In this embodiment, an eaCas9 molecule is provided as a protein, and a gRNA is provided as transcribed or synthesized RNA. In this embodiment, the donor template is provided as a DNA molecule.

Table 4: Delivery methods for Cas system components

Delivery Vector/Mode		Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical (e.g., electroporation, particle gun, Calcium Phosphate transfection, cell compression or squeezing)		YES	Transient	NO	Nucleic Acids and Proteins
Viral	**Retrovirus**	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modifications	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
Non-Viral	**Cationic Liposomes**	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
	Polymeric Nanoparticles	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
Biological Non-Viral Delivery Vehicles	**Attenuated Bacteria**	YES	Transient	NO	Nucleic Acids
	Engineered Bacteriophages	YES	Transient	NO	Nucleic Acids
	Mammalian Virus-like Particles	YES	Transient	NO	Nucleic Acids
	Biological liposomes: Erythrocyte Ghosts and Exosomes	YES	Transient	NO	Nucleic Acids

Table 5: Polymers used for gene transfer

Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis(succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimidate	DTBP
Poly(ethylene imine) biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(*N*-vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM
Poly(amido ethylenimine)	SS-PAEI
Triethylenetetramine	TETA
Poly(β-aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	
Poly(α-[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(*N*-ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(*N*-2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
N-Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

Table 6

Synthesis or manufacturer route	**Sequence of gRNA molecule**	T_m °C	**Delta** T_m	**% NHEJ**
Purchased from Company 1	GUAACGGCAGACUUCUCCUCGUU UUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAA CUUGAAAAAGUGGCACCGAGUCG GUGCUUUU (SEQ ID NO:208)	50	8	26.14
Synthesized from Company 2	GUAACGGCAGACUUCUCCUCGUU UUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAA CUUGAAAAAGUGGCACCGAGUCG GUGCUUUU (SEQ ID NO:208)	44	2	11.64
In vitro transcribed using MEGAshortscript™ T7 Kit	GUAACGGCAGACUUCUCCUCGUU UUAGAGCUAGAAAUAGCAAGUUA AAAUAAGGCUAGUCCGUUAUCAA CUUGAAAAAGUGGCACCGAGUCG GUGCUUUU (SEQ ID NO:208)	50	8	27.33
Synthesized from Company 2	GGUAACGGCAGACUUCUCCUCGU UUUAGAGCUAGAAAUAGCAAGUU AAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUC GGUGCUUUU (SEQ ID NO:209)	47	5	11.43
In vitro transcribed using MEGAshortscript™ T7 Kit	GGGUAACGGCAGACUUCUCCUCG UUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGU CGGUGCUUUU (SEQ ID NO:210)	51	9	22.74

What is claimed is:

1. A method of screening for a Cas9 molecule/gRNA molecule complex for administration to a subject, the method comprising:

(a) generating a plurality of samples, each sample comprising a Cas9 molecule/gRNA molecule complex generated by combining a Cas9 molecule and one of a plurality of gRNA molecules;

(b) detecting a melting temperature (T_m) value of the Cas9 molecule/gRNA molecule complex in each of the plurality of samples; and

(c) selecting at least one sample from the plurality of samples based on one or more of (i) a comparison of the T_m values in the plurality of samples to a T_m value of a reference Cas9 molecule/gRNA molecule complex or a pre-determined threshold T_m value, or (ii) a relative ordering of the T_m values of the plurality of samples.

2. The method according to claim 1, wherein the step of detecting the T_m value of the Cas9 molecule/gRNA molecule complex in each sample of the plurality of samples includes assessing each sample in the plurality of samples by differential scanning fluorimetry (DSF).

3. The method according to claim 1, wherein the gRNA is a chimeric gRNA.

4. The method according to claim 1, wherein the gRNA is a modular gRNA.

5. An isolated complex of a Cas9 molecule and a gRNA molecule having a melting temperature (T_m) at least 8° C greater than a T_m value of the Cas9 molecule in the absence of the gRNA molecule selected according to the method of one of claims 1-2.

6. A composition comprising an isolated complex of a Cas9 molecule and a gRNA molecule having a melting temperature (T_m) at least 8°C greater than a T_m of the Cas9 molecule in the absence of the gRNA molecule selected according to the method of one of claims 1-2.

7. The composition of claim 6, wherein the difference of the melting temperature (T_m) of the non-naturally occurring complex of a Cas9 molecule and a gRNA molecule and the melting temperature (T_m) of the Cas9 molecule in the absence of the gRNA molecule is assessed by differential scanning fluorimetry (DSF).

8. The composition of claim 6, wherein the gRNA is a chimeric or modular gRNA.

9. A method of determining the stability of a Cas9 molecule/gRNA molecule complex, the method comprising:

(a) generating a plurality of Cas9 molecule/gRNA molecule complexes, each comprising a Cas9 molecule/gRNA molecule complex generated by combining a Cas9 molecule and one of a plurality of gRNA molecules;

(b) detecting a melting temperature (T_m) value of each of the Cas9 molecule/gRNA molecule complexes of the plurality of Cas9 molecule/gRNA molecule complexes; and

(c) determining one or more of the plurality of Cas9 molecule/gRNA molecule complexes is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value.

10. A method of determining a condition that promotes a stable Cas9 molecule/gRNA molecule complex, the method comprising:

(a) combining a Cas9 molecule and a gRNA molecule in a sample to form a Cas9 molecule/gRNA molecule complex;

(b) detecting a melting temperature (T_m) value of the Cas9 molecule/gRNA molecule complex; and

(c) determining the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value.

11. A method of screening for a stable Cas9 molecule/gRNA molecule complex, the method comprising:

(a) detecting a melting temperature (T_m) value of a Cas9 molecule/gRNA molecule complex by differential scanning fluorimetry (DSF); and

(b) determining the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value.

12. A method for identifying an optimal gRNA to form a stable Cas9 molecule/gRNA molecule complex, the method comprising:

(a) combining a Cas9 molecule and a gRNA molecule in a sample to form the Cas9 molecule/gRNA molecule complex;

(b) detecting a melting temperature (T_m) value of the Cas9 molecule/gRNA molecule complex; and

(c) determining the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value by at least 8°C.

13. A method of determining the stability of a Cas9 molecule/gRNA molecule complex, the method comprising:

(a) combining a Cas9 molecule and a gRNA molecule in a sample to form the Cas9 molecule/gRNA molecule complex;

(b) detecting a melting temperature (T_m) value of the Cas9 molecule/gRNA molecule complex;

(c) measuring an activity value of the Cas9 molecule/gRNA molecule complex; and

(d) determining the Cas9 molecule/gRNA molecule complex is stable if (i) the T_m value of the Cas9 molecule/gRNA molecule complex is greater than a T_m value of a reference molecule or a T_m reference value and (ii) the activity value of the Cas9 molecule/gRNA molecule complex is greater than an activity value of a reference molecule or an activity reference value.

14. A method of optimizing binding of a gRNA molecule to a Cas9 molecule to form a stable Cas9 molecule/gRNA molecule complex, the method comprising:

(a) combining the Cas9 molecule and the gRNA molecule in a sample to form a Cas9 molecule/gRNA molecule complex;

(b) detecting a melting temperature (T_m) of the Cas9 molecule/gRNA molecule complex;

(c) determining a delta value between the T_m value of the Cas9 molecule/gRNA molecule complex and a T_m value of a reference molecule or a T_m reference value; and

(d) determining the Cas9 molecule/gRNA molecule complex is stable if the delta value is at least 8°C and the T_m value of the Cas9 molecule/gRNA molecule complex is greater than the T_m value of the reference molecule or the T_m reference value.

15. A method of detecting a stable Cas9 molecule/gRNA molecule complex, the method comprising:

(a) detecting a thermostability value of a reference molecule;

(b) combining a Cas9 molecule and a gRNA molecule in a sample to form a Cas9 molecule/gRNA molecule complex;

(c) measuring a thermostability value of a Cas9 molecule/gRNA molecule complex; and

(d) determining the Cas9 molecule/gRNA molecule complex is stable if the thermostability value of the Cas9 molecule/gRNA molecule complex is greater than the thermostability value of the reference molecule.

16. The method of claim 15, wherein the thermostability value is a denaturation temperature value and the thermostability reference value is a denaturation temperature reference value.

17. The method of claim 15, wherein the thermostability value is a melting temperature (T_m) value and the thermostability reference value is a T_m reference value.

18. The method of any of claims 9-11, 13 and 17, wherein the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is at least 1°C, at least 2°C, at least 3°C, at least 4°C, at least 5°C, at least 6°C, at least 7°C, at least 8°C, at least 9°C, at least 10°C, at least 11°C, at least 12°C, at least 13°C, at least 14°C, at least 15°C, at least 16°C, at least 17°C, at least 18°C, at least 19°C, or at least 20°C greater than the T_m value of the reference molecule or T_m reference value.

19. The method of claim 18, wherein the Cas9 molecule/gRNA molecule complex is stable if the T_m of the Cas9 molecule/gRNA molecule complex is at least 8°C greater than the T_m value of the reference molecule or T_m reference value.

20. The method of any of claims 9-11, 13 and 17, wherein the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is about 1°C, about 2°C, about 3°C, about 4°C, about 5°C, about 6°C, about 7°C, about 8°C, about 9°C, about 10°C, about 11°C, about 12°C, about 13°C, about 14°C, about 15°C, about 16°C, about 17°C, about 18°C, about 19°C, or about 20°C greater than the T_m value of the reference molecule or T_m reference value.

21. The method of any of claims 9-11, 13 and 17, wherein the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is about 1°C to 5°C, about 6°C to 10°C, about 11°C to 15°C, or about 16°C to 20°C greater than the T_m value of the reference molecule or T_m reference value.

22. The method of claim 21, wherein the Cas9 molecule/gRNA molecule complex is stable if the T_m value of the Cas9 molecule/gRNA molecule complex is about 6°C to 10°C greater than the T_m value of the reference molecule or T_m reference value.

23. The method of any of claims 9-10, 12-14, and 17-22, wherein the T_m value is detected using a thermal shift assay.

24. The method of claim 23, wherein the thermal shift assay is selected from differential scanning fluorimetry (DSF), differential scanning calorimetry (DSC), or isothermal titration calorimetry (ITC).

25. The method of any of claims 9-24, wherein the gRNA molecule comprises a chimeric gRNA molecule.

26. The method of any of claims 9-24, wherein the gRNA molecule comprises a modular gRNA molecule.

27. The method of any of claims 1-4 and 9-26, wherein the Cas9 molecule is a Cas9 molecule selected from Table 1.

28. The method of any of claims 1-4 and 9-26, wherein the Cas9 molecule is a chimeric Cas9 molecule, or a synthetic or engineered Cas9 molecule, e.g., a Cas9 molecule with a portion or portions deleted.

29. The method of any of claims 1-4 and 9-26, wherein the Cas9 molecule comprises a *S. pyogenes* or a *S. aureus* Cas9 molecule.

30. The method of any of claims 9-29, wherein the reference molecule is selected from:

(a) a reference Cas9 molecule in the absence of a gRNA molecule;

(b) a reference Cas9 molecule (e.g., the same Cas9 molecule as the Cas9 molecule in the complex being evaluated) complexed with a second gRNA molecule (i.e., a gRNA other than the one in the complex being evaluated); and

(c) a reference Cas9 molecule/gRNA molecule complex, wherein the reference Cas9 molecule/gRNA molecule was formed under different conditions, e.g., with different proportions of Cas9 molecule and gRNA molecule, than the Cas9 molecule/gRNA molecule complex, or was formed in a different buffer.

31. The method of claim 30, wherein the reference Cas9 molecule is the same as the Cas9 molecule of the complex being evaluated.

32. The method of claim 30, wherein the reference Cas9 molecule is different from the Cas9 molecule of the complex being evaluated.

33. The method of claim 32, wherein the reference Cas9 molecule differs in primary sequence from the Cas9 molecule of the complex being evaluated.

34. The method of any of claims 30-33, wherein the gRNA molecule of the reference Cas9 molecule/gRNA molecule complex is the same as the gRNA molecule of the complex being evaluated.

35. The method of any of claims 30-33, wherein the gRNA molecule of the reference Cas9 molecule/gRNA molecule complex is different from the gRNA molecule of the complex being evaluated.

36. The method of claim 35, wherein the gRNA molecule of the reference Cas9 molecule/gRNA molecule complex differs in sequence or differs by a modification from the gRNA molecule of the complex being evaluated.

37. The method of any of claims 9-14 and 17-36, wherein the T_m reference value comprises a preselected numerical value for T_m.

38. The method of any of claims 9-14 and 17-36, wherein the T_m reference value comprises a value correlated with the T_m of any of the reference molecules of claims 22-27.

39. The method of any of claims 1-4, 9-12, and 14-38, further comprising:

detecting an activity of the Cas9 molecule/gRNA molecule complex;

measuring an activity value of the Cas9 molecule/gRNA molecule complex; and

determining the Cas9 molecule/gRNA molecule complex is stable if the activity value of the Cas9 molecule/gRNA molecule complex is greater than the activity value of a reference molecule or an activity reference value.

40. The method of claims 13 and 39, wherein the activity comprises one or more of:

an ability to induce indels;

an ability to modify a target DNA;

a propensity of a preselected repair method;

an ability of the gRNA molecule to remain hybridized to the DNA target; and

an ability of the gRNA molecule to bind to the Cas9 molecule of the Cas9 molecule/gRNA molecule complex.

41. The method of claim 40, wherein the activity value is a binding value and the activity is the ability of the gRNA molecule to bind to the Cas9 molecule comprising:

combining the gRNA molecule and the Cas9 molecule in a sample to form the Cas9 molecule/gRNA molecule complex;

measuring a binding value of the Cas9 molecule/gRNA molecule complex; and

determining the Cas9 molecule/gRNA molecule complex is stable if the binding value of the Cas9 molecule/gRNA molecule complex is greater than the binding value of a reference molecule or the binding reference value

42. The method of claim 41, wherein the binding value is measured using a kinetics assay.

43. The method of claim 42, wherein the kinetics assay is selected from surface plasmon resonance (SPR) assay, Bio-Layer Interferometry (BLI) assay, or gel band shift assay.

44. The method of claim 40, wherein the propensity of a preselected repair method is HDR or NHEJ.

45. The method of any of claims 13 and 39-44, wherein the activity of the Cas9 molecule/gRNA molecule complex is tested using:

an *in vitro* system;

an ex vivo system;

an *in vivo* system;

a cellular assay; or

an animal model.

46. The method of any of claims 41-45, wherein the reference molecule is selected from any one of the reference molecules in claims 30-33.

47. The method of any of claims 41-45, wherein the reference Cas9 molecule/gRNA molecule complex is formed with different proportions of Cas9 molecule and gRNA molecule than the Cas9 molecule/gRNA molecule complex or is formed in a different buffer than the Cas9 molecule/gRNA molecule complex.

48. The method of claim 10, wherein the sample comprises a component comprising an additive, a small molecule, a stabilizing reagent, buffer, pH, salt concentration, glycerol concentration, or other buffer component.

49. A synthetic Cas9 molecule/gRNA molecule complex generated using the methods of any of claims 1-4 and 9-48.

50. A composition comprising the stable Cas9 molecule/gRNA molecule complex generated using the methods of any of claims 1-3 and 9-48.

51. A vector system comprising a nucleic acid encoding the stable Cas9 molecule/gRNA molecule complex generated using the methods of any of claims 1-4 and 9-48.

52. A method of delivering a Cas9 molecule/gRNA molecule complex to a target cell comprising delivering the stable Cas9 molecule/gRNA molecule complex of any of claims 9-48 to the target cell.

53. The method of claim 52, wherein the stable Cas9 molecule/gRNA molecule complex is delivered to the cell by RNP cationic lipid transfection, a viral vector, or RNA transfection.

54. The method of claim 53, wherein the viral vector is an AAV vector.

55. The methods of any of claims 1 and 9, wherein the plurality of gRNA molecules is a library of candidate gRNA molecules.

56. The method of claim 55, wherein the library of candidate gRNA molecules comprises a library of tracrRNA molecules or sequences.

57. The method of claim 56, wherein the library of tracrRNA molecules or sequences are of differing length.