

# RNA-GUIDED NUCLEASES AND ACTIVE FRAGMENTS AND VARIANTS THEREOF AND METHODS OF USE

## CROSS-REFERENCE TO RELATED APPLICATIONS

5 This application claims priority to U.S. Provisional Application No.63/014,970, filed April 24, 2020, and U.S. Provisional Application No. 63/077,211, filed September 11, 2020, each of which application is incorporated by reference herein in its entirety.

## STATEMENT REGARDING THE SEQUENCE LISTING

10 The Sequence Listing associated with this application is provided in ASCII format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The ASCII copy named L103438\_1200WO\_0084\_1\_SL is 878,362 bytes in size, was created on April 21, 2021, and is being submitted electronically via EFS-Web.

## FIELD OF THE INVENTION

15 The present invention relates to the field of molecular biology and gene editing.

## BACKGROUND OF THE INVENTION

20 Targeted genome editing or modification is rapidly becoming an important tool for basic and applied research. Initial methods involved engineering nucleases such as meganucleases, zinc finger fusion proteins or TALENs, requiring the generation of chimeric nucleases with engineered, programmable, sequence-specific DNA-binding domains specific for each particular target sequence. RNA-guided nucleases, such as the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) proteins of the CRISPR-Cas bacterial system, allow for the targeting of  
25 specific sequences by complexing the nucleases with guide RNA that specifically hybridizes with a particular target sequence. Producing target-specific guide RNAs is less costly and more efficient than generating chimeric nucleases for each target sequence. Such RNA-guided nucleases can be used to edit genomes optionally through the introduction of a sequence-specific, double-stranded break that is repaired via error-prone non-homologous end-joining (NHEJ) to introduce a mutation at  
30 a specific genomic location. Alternatively, heterologous DNA may be introduced into the genomic site via homology-directed repair. RNA-guided nucleases (RGNs) can also be used for base editing when fused with a deaminase.

## BRIEF SUMMARY

35 Compositions and methods for binding a target sequence of interest are provided. The compositions find use in cleaving or modifying a target sequence of interest, detection of a target sequence of interest, and modifying the expression of a sequence of interest. Compositions comprise

RNA-guided nuclease (RGN) polypeptides, CRISPR RNAs (crRNAs), trans-activating CRISPR RNAs (tracrRNAs), guide RNAs (gRNAs), nucleic acid molecules encoding the same, and vectors and host cells comprising the nucleic acid molecules. Also provided are RGN systems for binding a target sequence of interest, wherein the RGN system comprises an RNA-guided nuclease polypeptide and one or more guide RNAs. Thus, methods disclosed herein are drawn to binding a target sequence of interest, and in some examples, cleaving or modifying the target sequence of interest. The target sequence of interest can be modified, for example, as a result of non-homologous end joining, homology-directed repair with an introduced donor sequence, or base editing.

In one aspect, the present disclosure provides a nucleic acid molecule comprising a polynucleotide encoding an RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 90% sequence identity to, or having at least 95% sequence identity to, SEQ ID NO: 83;

wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82),

wherein said polynucleotide encoding an RGN polypeptide is operably linked to a promoter heterologous to said polynucleotide and/or codon optimized for expression in a eukaryotic cell.

In another aspect, the present disclosure provides a vector comprising the nucleic acid molecule of the present disclosure.

In yet another aspect, the present disclosure provides a cell comprising the nucleic acid molecule of the present disclosure or the vector of the present disclosure.

In yet another aspect, the present disclosure provides an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to, or having at least 95% sequence identity to, SEQ ID NO: 83;

wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82), and wherein said RGN polypeptide is operably fused to one or more nuclear localization signals (NLS) heterologous to said RGN polypeptide.

In yet another aspect, the present disclosure provides a system for binding a target DNA sequence of a DNA molecule, said system comprising:

a) one or more guide RNAs capable of hybridizing to said target DNA sequence, or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more guide RNAs (gRNAs); and

b) an RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 90% sequence identity to, or having at least 95% sequence identity to, SEQ ID NO: 83, or a polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide, wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82);

wherein at least one of said nucleotide sequences encoding the one or more guide RNAs and said nucleotide sequence encoding the RGN polypeptide is operably linked to a promoter heterologous to said nucleotide sequence.

In yet another aspect, the present disclosure provides a system for binding a target DNA sequence of a DNA molecule, said system comprising:

a) an RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as SEQ ID NO: 83, or a polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide, wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82); and

b) one or more guide RNAs comprising (i) a spacer sequence that is complementary to said target DNA sequence and (ii) a CRISPR repeat sequence recognized by the RGN; or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more guide RNAs.

In yet another aspect, the present disclosure provides a complex comprising (i) an isolated RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as SEQ ID NO: 83 bound to (ii) a non-natural guide RNA (gRNA), wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82).

In yet another aspect, the present disclosure provides a method for cleaving and/or modifying a target DNA sequence of a DNA molecule, comprising contacting the DNA molecule with:

a) an RNA-guided nuclease (RGN) polypeptide, wherein said RGN comprises an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as, SEQ ID NO: 83, wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82); and

b) one or more guide RNAs capable of targeting the RGN of (a) to the target DNA sequence; wherein the one or more guide RNAs hybridize to the target DNA sequence, thereby directing said RGN polypeptide to bind to said target DNA sequence and cleavage and/or modification of said target DNA sequence occurs.

## DETAILED DESCRIPTION

Many modifications and other embodiments of the inventions set forth herein will come to mind to one skilled in the art to which these inventions pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood

that the inventions are not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended embodiments. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

5

### *I. Overview*

RNA-guided nucleases (RGNs) allow for the targeted manipulation of specific site(s) within a genome and are useful in the context of gene targeting for therapeutic and research applications. In a variety of organisms, including mammals, RNA-guided nucleases have been used for genome engineering by stimulating non-homologous end joining and homologous recombination, for example. The compositions and methods described herein are useful for creating single- or double-stranded breaks in polynucleotides, modifying polynucleotides, detecting a particular site within a polynucleotide, or modifying the expression of a particular gene.

The RNA-guided nucleases disclosed herein can alter gene expression by modifying a target sequence. In specific embodiments, the RNA-guided nucleases are directed to the target sequence by a guide RNA (gRNA) as part of a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNA-guided nuclease system. The RGNs are considered “RNA-guided” because guide RNAs form a complex with the RNA-guided nucleases to direct the RNA-guided nuclease to bind to a target sequence and in some embodiments, introduce a single-stranded or double-stranded break at the target sequence. After the target sequence has been cleaved, the break can be repaired such that the DNA sequence of the target sequence is modified during the repair process. Thus, provided herein are methods for using the RNA-guided nucleases to modify a target sequence in the DNA of host

## Claims

1. A nucleic acid molecule comprising a polynucleotide encoding an RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 90% sequence identity to, or having at least 95% sequence identity to, SEQ ID NO: 83;  
wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82),  
wherein said polynucleotide encoding an RGN polypeptide is operably linked to a promoter heterologous to said polynucleotide and/or codon optimized for expression in a eukaryotic cell.
2. The nucleic acid molecule of claim 1, wherein said RGN polypeptide comprises the amino acid sequence set forth as SEQ ID NO: 83.
3. The nucleic acid molecule of claim 1, wherein said RGN polypeptide is nuclease inactive or is a nickase.
4. The nucleic acid molecule of any one of claims 1-3, wherein the RGN polypeptide is operably fused to a base-editing polypeptide or an effector domain, optionally wherein said RGN polypeptide is operably fused to one or more nuclear localization signals heterologous to said RGN polypeptide.
5. The nucleic acid molecule of claim 4,  
wherein the base-editing polypeptide is a deaminase; or  
wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain, optionally wherein said expression modulator domain is an epigenetic modification domain, a transcriptional repressor domain, or a transcriptional activation domain.
6. The nucleic acid molecule of claim 4 or 5, wherein said effector domain is operably fused at the N-terminus, at the C-terminus, or at an internal location of said RGN polypeptide.
7. A vector comprising the nucleic acid molecule of any one of claims 1-6.
8. The vector of claim 7, further comprising at least one nucleotide sequence encoding a guide RNA capable of hybridizing to a target DNA sequence.

9. The vector of claim 8, wherein the guide RNA is selected from the group consisting of:  
a) a guide RNA comprising:

i) a CRISPR RNA comprising a CRISPR repeat comprising a nucleotide sequence having at least 90% sequence identity to, a nucleotide sequence having at least 95% sequence identity to, or the nucleotide sequence set forth as, SEQ ID NO: 84; and

ii) a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to, a nucleotide sequence having at least 95% sequence identity to, or the nucleotide sequence set forth as, SEQ ID NO: 85;

wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as, SEQ ID NO: 83; and

b) a guide RNA comprising:

i) a CRISPR RNA comprising a CRISPR repeat comprising a nucleotide sequence having at least 90% sequence identity to, a nucleotide sequence having at least 95% sequence identity to, or the nucleotide sequence set forth as, SEQ ID NO: 84; and

ii) a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to, a nucleotide sequence having at least 95% sequence identity to, or the nucleotide sequence set forth as, SEQ ID NO: 78;

wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as, SEQ ID NO: 83.

10. A cell comprising the nucleic acid molecule of any one of claims 1-6 or the vector of any one of claims 7-9.

11. An RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to, or having at least 95% sequence identity to, SEQ ID NO: 83;

wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82), and wherein said RGN polypeptide is operably fused to one or more nuclear localization signals (NLS) heterologous to said RGN polypeptide.

12. The RGN polypeptide of claim 11, wherein said RGN polypeptide comprises the amino acid sequence set forth as SEQ ID NO: 83.

13. The RGN polypeptide of claim 11, wherein said RGN polypeptide is nuclease inactive or a nickase.

14. The RGN polypeptide of any one of claims 11-13, wherein the RGN polypeptide is operably fused to a base-editing polypeptide or an effector domain.

15. The RGN polypeptide of claim 14, wherein the base-editing polypeptide is a deaminase; or wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain, optionally wherein said expression modulator domain is an epigenetic modification domain, a transcriptional repressor domain, or a transcriptional activation domain.

16. The RGN polypeptide of claim 14 or 15, wherein said effector domain is operably fused at the N-terminus, at the C-terminus, or at an internal location of said RGN polypeptide.

17. A system for binding a target DNA sequence of a DNA molecule, said system comprising:

a) one or more guide RNAs capable of hybridizing to said target DNA sequence, or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more guide RNAs (gRNAs); and

b) an RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 90% sequence identity to, or having at least 95% sequence identity to, SEQ ID NO: 83, or a polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide, wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82); wherein at least one of said nucleotide sequences encoding the one or more guide RNAs and said nucleotide sequence encoding the RGN polypeptide is operably linked to a promoter heterologous to said nucleotide sequence.

18. The system of claim 17, wherein said RGN polypeptide comprises the amino acid sequence set forth as SEQ ID NO: 83.

19. The system of claim 17 or 18, wherein said target DNA sequence is a eukaryotic target DNA sequence, optionally wherein nucleotide sequences encoding the one or more guide RNAs and the nucleotide sequence encoding an RGN polypeptide are located on one vector.

20. The system of any one of claims 17-19, wherein said gRNA is selected from the group consisting of:

a) a gRNA comprising a CRISPR repeat comprising a nucleotide sequence having at least 90% sequence identity to, a nucleotide sequence having at least 95% sequence identity to, or the nucleotide sequence set forth as, SEQ ID NO: 84 and a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to, a nucleotide sequence having at least 95% sequence identity to, or the nucleotide sequence set forth as, SEQ ID NO: 85, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as, SEQ ID NO: 83; and

b) a gRNA comprising a CRISPR repeat comprising a nucleotide sequence having at least 90% sequence identity to, a nucleotide sequence having at least 95% sequence identity to, or the nucleotide sequence set forth as, SEQ ID NO: 84 and a tracrRNA having at least 90% sequence identity to, a nucleotide sequence having at least 95% sequence identity to, or the nucleotide sequence set forth as, SEQ ID NO: 78, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as, SEQ ID NO: 83.

21. The system of any one of claims 17, 19, and 20, wherein said RGN polypeptide is nuclease inactive or is a nickase.

22. The system of any one of claims 17-21, wherein the RGN polypeptide is operably linked to a base-editing polypeptide or an effector domain, optionally wherein said RGN polypeptide is operably fused to one or more nuclear localization signals (NLS) heterologous to said RGN polypeptide.

23. The system of claim 22,  
wherein the base-editing polypeptide is a deaminase; or  
wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain, optionally wherein said expression modulator domain is an epigenetic modification domain, a transcriptional repressor domain, or a transcriptional activation domain.

24. The system of claim 22 or 23, wherein said effector domain is operably fused at the N-terminus, at the C-terminus, or at an internal location of said RGN polypeptide.

25. A system for binding a target DNA sequence of a DNA molecule, said system comprising:

a) an RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as SEQ ID NO: 83, or a polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide, wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82); and

b) one or more guide RNAs comprising (i) a spacer sequence that is complementary to said target DNA sequence and (ii) a CRISPR repeat sequence recognized by the RGN; or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more guide RNAs.

26. A complex comprising (i) an isolated RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as SEQ ID NO: 83 bound to (ii) a non-natural guide RNA (gRNA), wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82).

27. The complex of claim 26, wherein the non-natural gRNA is selected from: (a) a gRNA comprising one or more synthetic ribonucleotide analogues; (b) a single guide RNA; and (c) a gRNA comprising one or more peptide nucleic acid (PNA), PNA-DNA chimera, locked nucleic acid (LNA), and/or phosphorothioate linkage.

28. A method for cleaving and/or modifying a target DNA sequence of a DNA molecule, comprising contacting the DNA molecule with:

a) an RNA-guided nuclease (RGN) polypeptide, wherein said RGN comprises an amino acid sequence having at least 90% sequence identity to, an amino acid sequence having at least 95% sequence identity to, or the amino acid sequence set forth as, SEQ ID NO: 83, wherein said RGN polypeptide is capable of recognizing a protospacer adjacent motif (PAM) comprising a nucleotide sequence of NNGRR (SEQ ID NO: 82); and

b) one or more guide RNAs capable of targeting the RGN of (a) to the target DNA sequence;

wherein the one or more guide RNAs hybridize to the target DNA sequence, thereby directing said RGN polypeptide to bind to said target DNA sequence and cleavage and/or modification of said target DNA sequence occurs.

29. The method of claim 28, wherein cleavage by said RGN polypeptide generates a double-stranded break or a single-stranded break.

30. The method of claim 28 or 29, wherein said modified target DNA sequence comprises:  
insertion of heterologous DNA into the target DNA sequence;  
deletion of at least one nucleotide from the target DNA sequence; or  
mutation of at least one nucleotide in the target DNA sequence.