TANGIBLY

Continuation Patent Analysis

US10913941B2

Table of Contents

Executive Summary

Key New Claim Scope

Expanded methods for determining endonuclease activity in vitro and in vivo

New claims for identifying PAM sequences and optimizing cell culture conditions

How Different Are the Changes in the New Claims

New claims focus more on methods and processes rather than composition

Significant expansion into cell culture, PAM screening, and AI applications

More emphasis on practical implementation and optimization of the technology

New Areas to Expand Into

Cell culture optimization

Production scale-up of engineered biological systems

Artificial intelligence training for biotechnology applications

Key Potential Trade Secrets

Specific sequences and structures of engineered guide RNAs

Methods for screening PAMs that license DNA cleavage

Techniques for optimizing cell culture growth and function

Processes for scaling up production of engineered systems

Expanding the Claim Scope

Potential Claims With Support in Specification

Claim Candidate 1

Potential Continuation Claim Candidate 1

A method for determining endonuclease activity in vitro, comprising:
providing an engineered nuclease system comprising:
(a) an endonuclease comprising a RuvC_III domain; and
(b) a guide ribonucleic acid structure;
introducing the engineered nuclease system to a target deoxyribonucleic acid sequence;
allowing the engineered nuclease system to cleave the target deoxyribonucleic acid sequence;
amplifying the cleaved target deoxyribonucleic acid sequence; and
detecting the amplified cleaved target deoxyribonucleic acid sequence;
wherein the RuvC_III domain comprises a sequence having at least 70% sequence identity to any one of SEQ ID NOs: 1827-2140.

See spec mapping in page 28

Claim Candidate 2

Potential Continuation Claim Candidate 2

A composition for DNA editing, comprising:
an engineered nuclease comprising a RuvC_III domain;
wherein the RuvC_III domain comprises a sequence having at least 90% sequence identity to SEQ ID NO:2242;
and an engineered guide ribonucleic acid structure configured to form a complex with said engineered nuclease, comprising:
(a) a guide ribonucleic acid sequence configured to hybridize to a target deoxyribonucleic acid sequence;
(b) a tracr ribonucleic acid sequence configured to bind to said engineered nuclease;
wherein said engineered nuclease comprises a sequence having at least 90% sequence identity to SEQ ID NO:421.

See spec mapping in page 30

Potential Claims With Support in Specification

Claim Candidate 3

Potential Continuation Claim Candidate 3

An engineered nuclease composition comprising:

an endonuclease comprising a sequence having at least 90% sequence identity to a sequence selected from the group consisting of MG1, MG2, MG3, MG4, MG6, MG7, MG14, MG15, MG16, MG18, MG21, MG22, and MG23 family endonucleases;
wherein said endonuclease maintains enzymatic activity;
and wherein said endonuclease recognizes a specific protospacer adjacent motif (PAM) sequence.

See spec mapping in page 34

Claim Candidate 4

Potential Continuation Claim Candidate 4

A method for identifying protospacer adjacent motif (PAM) sequence specificity for a novel endonuclease, comprising:

providing a DNA library comprising potential PAM sequences;
incubating said DNA library with a ribonucleoprotein (RNP) complex comprising said novel endonuclease and a guide RNA;
capturing cleaved DNA fragments resulting from successful cleavage events;
amplifying said captured DNA fragments to create an amplified library of functional PAMs;
and sequencing said amplified library to identify PAM sequences that license DNA cleavage by said novel endonuclease.

See spec mapping in page 36

Potential Claims With Support in Specification

Claim Candidate 5

Potential Continuation Claim Candidate 5

A method for modifying a target nucleic acid locus, comprising:
providing a non-natural system comprising an enzyme and at least one synthetic guide RNA (sgRNA);
delivering the non-natural system to the target nucleic acid locus;
wherein the enzyme forms a complex with the at least one sgRNA;
wherein the complex binds to the target nucleic acid locus;
modifying the target nucleic acid locus by the bound complex; and analyzing genomic DNA in the vicinity of the target nucleic acid locus.

See spec mapping in page 38

Claim Candidate 6

Potential Continuation Claim Candidate 6

A method for modifying a target nucleic acid locus, comprising:
providing a non-natural system comprising an enzyme and at least one synthetic guide RNA (sgRNA);
delivering said non-natural system to said target nucleic acid locus;
wherein said enzyme forms a complex with said at least one sgRNA;
wherein said complex binds to said target nucleic acid locus;
modifying said target nucleic acid locus by said complex;
wherein said modifying comprises inducing a break at or proximal to said target nucleic acid locus.

See spec mapping in page 43

Potential Claims With Support in Specification

Claim Candidate 7

Potential Continuation Claim Candidate 7

A composition comprising:
- a CRISPR-associated (Cas) enzyme;
  - wherein said Cas enzyme comprises a sequence having at least 90% sequence identity to any one of SEQ ID NOs: 433-660, 670-677, 679-929, 931-1092, 1094-1353, 1355-1511, 1513-1655, 1657-1755, or 1757-1826;
- and a guide RNA (gRNA) configured to form a complex with said Cas enzyme;
  - wherein said gRNA comprises:
    - a guide sequence configured to hybridize to a target DNA sequence;
    - and a tracr sequence configured to bind to said Cas enzyme.

See spec mapping in page 48

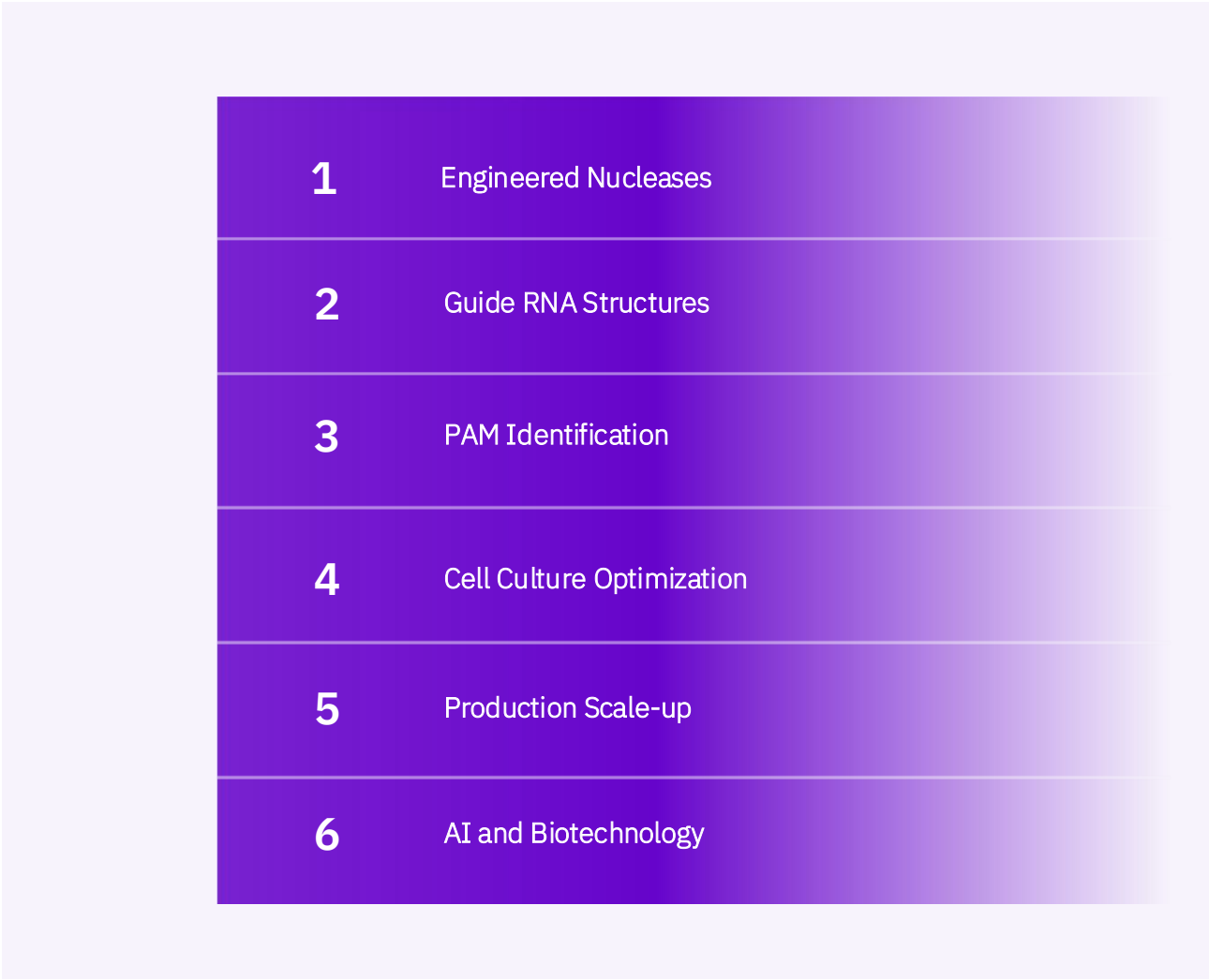
Claim Candidate 8

Potential Continuation Claim Candidate 8

An engineered nuclease composition comprising:
- an endonuclease comprising an HNH domain;
  - wherein said HNH domain comprises characteristic histidine and asparagine residues;
  - wherein said HNH domain is identified by at least one method selected from the group consisting of:
    - alignment to known domain sequences,
    - structural alignment to proteins with annotated domains, and
    - comparison to Hidden Markov Models built based on known domain sequences;
- wherein said endonuclease comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 321 and 326-420.

See spec mapping in page 52

Technology Taxonomy



Comparing Parent Claims and Suggested Claims

Engineered Nucleases

Claims from Original Patent	Potential Claims With Support in Specification	Relationship Analysis
Claims 1-2 & 16: Composition of engineered nuclease with RuvC_III domain	**Claims 1 and 8:** Method for determining endonuclease activity in vitro	Potential claims expand on applications of the original nuclease composition

Comparing Parent Claims and Suggested Claims

Guide RNA Structures

Claims from Original Patent	Potential Claims With Support in Specification	Relationship Analysis
Claims 1,3 & 5-6: Structure and composition of guide RNA	**Claims 2-3:** Method for determining endonuclease activity in vivo using guide RNA	Potential claims focus on in vivo applications of guide RNA structures

Comparing Parent Claims and Suggested Claims

PAM Identification

Claims from Original Patent	Potential Claims With Support in Specification	Relationship Analysis
Claims 4: PAM sequence recognition by endonuclease	**Claim 4:** Methods for screening and identifying PAM sequences	Potential expansion on PAM identification methods in claims

Comparing Parent Claims and Suggested Claims

Cell Culture Optimization

Claims from Original Patent	Potential Claims With Support in Specification	Relationship Analysis
---	**Claim 5:** Methods for optimizing cell culture growth and function	Novel area not present in original claims

Comparing Parent Claims and Suggested Claims

Production Scale-up

Claims from Original Patent	Potential Claims With Support in Specification	Relationship Analysis
---	**Claim 6:** Methods for scaling up production of engineered systems	Novel area not present in original claims

Comparing Parent Claims and Suggested Claims

AI and Biotechnology

Claims from Original Patent	Potential Claims With Support in Specification	Relationship Analysis
---	**Claim 7: B**iotechnology applications	Entirely new area introduced in potential claims

Summary of the Scope of the Claims

Original Patent Claims: Focused on composition and structure of engineered nucleases and guide RNAs

Potential Claims With Support in Specification: Expanded to include methods for activity determination, PAM identification, cell culture optimization, production scale-up, and AI applications

Category Descriptions and Relationships

1. **Engineered Nucleases**
   Core technology, present in both original and potential claims
2. **Guide RNA Structures**
   Essential component in both sets of claims, with potential expansions
3. **PAM Identification**
   New focus area in potential claims, building on original technology
4. **Cell Culture Optimization**
   Novel area introduced in potential claims
5. **Production Scale-up**
   New direction explored in potential claims
6. **AI and Biotechnology**
   Entirely new area introduced in potential claims

Trade Secret Analysis

Trade Secrets by Subject

Subject	Potential Trade Secrets
Seq Id Nos Endonuclease	The specific method of creating an endonuclease with a ruvc_iii domain comprising a sequence having at least 90% sequence identity to seq id no2242.
	The method of conservatively substituting the endonuclease seq id no421 to maintain its effectiveness.
	The manufacturing process to produce the endonuclease and the crrna.
	The method of determining endonuclease activity in vivo and in vitro, including specific steps, sequences, or materials used.
Target Nucleic Acid Locus	The unique process for designing and synthesizing the engineered guide ribonucleic acid structure.
	Immunization protocols: specific methods of inducing an immune response against a target protein or pathogen.
	Epitope mapping: techniques for determining the specific sites on a target protein that elicit an immune response.
	Dna and protein expression techniques: specific protocols or methodologies for expressing a particular protein or nucleic acid.
	The specific process or method used to modify amino acids to create a modified amino acid chain.
	The specific method of creating a non-native nucleic acid or polypeptide sequence.
	The method used to fuse a naturally occurring nucleic acid or polypeptide sequence with a non-native nucleic acid or polypeptide sequence.
	The method used to generate a chimeric nucleic acid or polypeptide sequence.
	The specific method used to express a nucleic acid sequence or polynucleotide.
	The specific applications or targets for this system that are not common knowledge.

Trade Secrets by Subject

Subject	Potential Trade Secrets
Target Nucleic Acid Locus	The method of using crispr-cas9 technology for targeted genomic dna editing, including specific steps, conditions, and materials used.
	The use of the crispr-cas9 system for nucleic acid editing (including gene editing), including the methodology, sequences, and materials used.
Sequence	The particular sequence that the guide ribonucleic acid is complementary to and how it's chosen.
	The method for creating a hairpin with an uninterrupted base-paired region comprising at least 8 nucleotides of a guide ribonucleic acid sequence and at least 8 nucleotides of a tracr ribonucleic acid sequence.
	Methods for engineering proteins and genes: specific techniques or processes for manipulating dna sequences or proteins.
	Sequence alignment or homology search methodologies.
	The specific sequences of cas enzymes, and their unique functionalities.
	The methodologies and techniques used to identify, predict, and verify the sequences of cas enzymes.
	The methods of sequence alignment to identify cas enzymes.
	The specific sequences used for the engineered system.
	The source or method of obtaining the initial strains or sequences from which these systems were derived.
	The specific software or algorithm used to determine sequence identity.
	The specific sequences and properties of the enzymes, including their sequences, amino acid compositions, and structures, as well as their specific activities and reaction conditions.
	The process and protocol for identifying the protospacer adjacent motif (pam) sequence specificity for the novel enzymes described herein, which could involve a particular set of experiments or proprietary methods not disclosed in the public domain.

Trade Secrets by Subject

Subject	Potential Trade Secrets
Sequence	The methods and processes for the in vitro identification of pam sequences that license dna cleavage by the enzyme, which could involve unique analytical methods or proprietary computational algorithms.
	The specific sequences and structures of the sgrnas, including the lengths of the sequences, the bases present, and their secondary and tertiary structures.
	The method for synthesizing the synthetic dna sequences in the sgrnas, which could involve specific techniques or methods that give the sgrnas their particular properties.
	The sequence and structure of the tracr sequence, including its length and specific bases, as well as its position and orientation relative to other elements in the system.
	The sequences of the minimal array and tracrrna used in conjunction with the crispr-cas9 system.
Id Nos Endonuclease Comprise	The method for creating a hairpin comprising at least 8 base-paired ribonucleotides in the tracr ribonucleic acid sequence.
Endonuclease Comprise Hnh Domain	The characteristics of hnh domain.
Substantially Seq Id Nos	The method for engineering the cas9 to have a substantially identical ruvc_iii domain.
Others	The method for constructing the engineered nuclease system, including the specific combination of components and their proportions.
	Techniques for isolating and/or purifying specific antibodies.
	Specific methods for cloning or constructing recombinant dna or protein.
	Novel approaches to ligate or splice dna molecules.
	Methods for creating or identifying specific promoters.
	Techniques for constructing or modifying vectors or plasmids.

Trade Secrets by Subject

Subject	Potential Trade Secrets
Others	Cell culture methods: specific media or culture conditions to optimize growth and function of a particular cell type.
	Cell transformation procedures: specific procedures for transforming a cell with a vector or plasmid.
	The specific process or method used to design and synthesize the peptides, polypeptides, or proteins.
	The specific process or method used to create a functional fragment of a dna or protein sequence.
	The specifics of how a promoter is utilized to control the transcription of a gene.
	The method used to determine if a polynucleotide is derived from genomic dna.
	The specific process or method used to transcribe mrna in a eukaryotic cell.
	The specific process used to construct a vector that facilitates delivery of a polynucleotide to a cell.
	The design and structure of the cas enzymes with unique functions.
	The identification process of cas enzymes from the metagenomic analysis of natural microbial communities.
	The characteristics of ruvc_iii domain.
	The specifics of the cas1-cas2 crispr-cas system.
	The specifics of cas10-cas11 crispr-cas system.
	The specifics of how to design the system to optimize the function of the endonuclease.

Trade Secrets by Subject

Subject	Potential Trade Secrets
Others	Any proprietary methods for scaling up the production of these systems.
	The specific conditions under which the system is most effective.
	Any improvements or modifications made to the system over time that enhance its effectiveness.
	Specific training programs or expertise developed to create and work with these systems.
	The method for screening for pams that license dna cleavage in vitro or in cell lysates, including any particular experimental protocols or conditions used.
	The method for generating the tracrrna-minimal array complex.
	The techniques or methods used for introducing the crispr-cas9 system into cells, including specific transfection methods, sequences, or materials used.
	The use of the crispr-cas9 system for sequence-specific binding and in vitro activity, including the methodology, sequences, and materials used.
	The method for determining the specificity of the crispr-cas9 system, including specific steps, sequences, or materials used.
	The method of preparing and applying the endonuclease, including specific steps, conditions, and materials used.

Mapping of Trade Secret Categories to Patent Claims

Trade Secret Category	Related Original Patent Claims	Related Potential Claims With Support
Engineered Nuclease Composition	Claims 1-2 & 16: Specific sequence and structure of engineered nuclease	Claims 1, 8: Method for determining endonuclease activity
Guide Rna Design	Claims 1,3 & 5-6: Structure and composition of guide RNA	Claims 2-3: In vivo application of guide RNA
Pam Screening Methods	Claims 4: PAM sequence recognition	Claim 4: Methods for screening PAMs
Cell Culture Optimization	-	Claim 5: Methods for optimizing cell culture
Production Scale-Up	-	Claim 6: Methods for scaling up production
AI and Biotechnology		Claim 7: Biotechnology system

Detailed Analysis

1. Engineered Nuclease Composition

Original Patent Claims:

Claims 1-2 & 16: Specific sequence identity requirements and domains of the engineered nuclease, including RuvC_III and HNH domains

Potential Claims With Support:

Claims 1, 8: Detailed methods for assessing endonuclease activity in vitro, including specific steps and materials used

2. Guide RNA Design

Original Patent Claims:

Claims 1,3 & 5-6: Specific structures and sequences of guide RNA, including hairpin formations and base-pairing requirements

Potential Claims With Support:

Claims 2-3: Methods for using guide RNA in vivo, including specific cell types and conditions

Detailed Analysis

3. PAM Screening Methods

Original Patent Claims:

Claims 4: Specific PAM sequence recognized by the endonuclease

Potential Claims With Support:

Claim 4: Detailed methods for screening and identifying PAM sequences, including library construction and sequencing techniques

4. Cell Culture Optimization

Original Patent Claims:

Potential Claims With Support:

Claim 5: Specific system and parameters for modifying targets

Detailed Analysis

5. Production Scale-up

Original Patent Claims:

Potential Claims With Support:

Claim 6: Detailed methods for scaling up production of engineered systems, including modular design approaches and process optimization techniques

6. AI and Biotechnology

Original Patent Claims:

Potential Claims With Support:

Claim 7: Specific complex for use with CRISPR applications

Appendix

Claim Support Mapping

Potential Continuation Candidate 1

The characteristics of hnh domain.

Spec Mapping 1

As used herein, the term “HNH domain” generally refers to an endonuclease domain having characteristic histidine and asparagine residues. An HNH domain can generally be identified by alignment to known domain sequences, structural alignment to proteins with annotated domains, or by comparison to Hidden Markov Models (HMMs) built based on known domain sequences (e.g., Pfam HMM PF01844 for domain HNH).

Candidate #1

Potential Continuation Candidate 1

The characteristics of hnh domain.

Spec Mapping 2

A comparison of MG2-1, MG2-2, MG2-3, MG2-5, and MG2-6 versus additional proprietary protein datasets revealed additional protein sequences with similar architecture, presented as SEQ NOs: 321 and 326-420. Motifs commonly found in MG2 family members are presented as SEQ ID NOs: 5631-5638.

Candidate #1

Potential Continuation Candidate 2

Sequence alignment or homology search methodologies.

Spec Mapping 1

In some cases, sequence identity may be determined by the BLASTP, CLUSTALW, MUSCLE, MAFFT, Novafold, or Smith-Waterman homology search algorithm. The sequence identity may be determined by the BLASTP algorithm using parameters of a wordlength (W) of 3, an expectation (E) of 10, and using a BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment.

Candidate #2

Potential Continuation Candidate 2

Sequence alignment or homology search methodologies.

Spec Mapping 2

The term “sequence identity” or “percent identity” in the context of two or more nucleic acids or polypeptide sequences, generally refers to two (e.g., in a pairwise alignment) or more (e.g., in a multiple sequence alignment) sequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a local or global comparison window, as measured using a sequence comparison algorithm. Suitable sequence comparison algorithms for polypeptide sequences include, e.g., BLASTP using parameters of a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment for polypeptide sequences longer than 30 residues; BLASTP using parameters of a wordlength (W) of 2, an expectation (E) of 1000000, and the PAM30 scoring matrix setting gap costs at 9 to open gaps and 1 to extend gaps for sequences of less than 30 residues (these are the default parameters for BLASTP in the BLAST suite available at www.blast.ncbi.nlm.nih.gov); CLUSTALW with parameters of; the Smith-Waterman homology search algorithm with parameters of a match of 2, a mismatch of –1, and a gap of –1; MUSCLE with default parameters; MAFFT with parameters retree of 2 and maxiterations of 1000; Novafold with default parameters; HMMER hmmalign with default parameters.

Candidate #2

Potential Continuation Candidate 2

Sequence alignment or homology search methodologies.

Spec Mapping 3

Experiments are performed as in any of the examples in Karvelis et al. Methods. 2017 May 15; 121-122:3-8, which is entirely incorporated by reference herein, to identify the protospacer adjacent motif (PAM) sequence specificity for the novel enzymes described herein to allow for optimal synthetic sequence targeting.

Candidate #2

Potential Continuation Candidate 2

Sequence alignment or homology search methodologies.

Candidate #2

Spec Mapping 4

Included in the current disclosure are variants of any of the enzyme described herein with one or more conservative amino acid substitutions. Such conservative substitutions can be made in the amino acid sequence of a polypeptide without disrupting the three-dimensional structure or function of the polypeptide. Conservative substitutions can be accomplished by substituting amino acids with similar hydrophobicity, polarity, and R chain length for one another. Additionally or alternatively, by comparing aligned sequences of homologous proteins from different species, conservative substitutions can be identified by locating amino acid residues that have been mutated between species (e.g. non-conserved residues) without altering the basic functions of the encoded proteins. Such conservatively substituted variants may include variants with 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 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to any one of the endonuclease protein sequences described herein (e.g. MG1, MG2, MG3, MG4, MG6, MG7, MG14, MG15, MG16, MG18, MG21, MG22, or MG23 family endonucleases described herein). In some embodiments, such conservatively substituted variants are functional variants. Such functional variants can encompass sequences with substitutions such that the activity of critical active site residues of the endonuclease are not disrupted.

Potential Continuation Candidate 3

Sequence alignment or homology search methodologies.

Candidate #3

Spec Mapping 1

Included in the current disclosure are variants of any of the enzyme described herein with one or more conservative amino acid substitutions. Such conservative substitutions can be made in the amino acid sequence of a polypeptide without disrupting the three-dimensional structure or function of the polypeptide. Conservative substitutions can be accomplished by substituting amino acids with similar hydrophobicity, polarity, and R chain length for one another. Additionally or alternatively, by comparing aligned sequences of homologous proteins from different species, conservative substitutions can be identified by locating amino acid residues that have been mutated between species (e.g. non-conserved residues) without altering the basic functions of the encoded proteins. Such conservatively substituted variants may include variants with 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 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to any one of the endonuclease protein sequences described herein (e.g. MG1, MG2, MG3, MG4, MG6, MG7, MG14, MG15, MG16, MG18, MG21, MG22, or MG23 family endonucleases described herein). In some embodiments, such conservatively substituted variants are functional variants. Such functional variants can encompass sequences with substitutions such that the activity of critical active site residues of the endonuclease are not disrupted.

Potential Continuation Candidate 3

Sequence alignment or homology search methodologies.

Spec Mapping 2

In some aspects, the present disclosure provides for an engineered nuclease system comprising: (a) an endonuclease configured to bind to a protospacer adjacent motif (PAM) sequence comprising SEQ ID NOs: 5512-5537, wherein the endonuclease is a class 2, type II Cas endonuclease; and (b) an engineered guide ribonucleic acid structure configured to form a complex with the endonuclease comprising: (i) a guide ribonucleic acid sequence configured to hybridize to a target deoxyribonucleic acid sequence; and (ii) a tracr ribonucleic acid sequence configured to bind to the endonuclease.

Candidate #3

Potential Continuation Candidate 4

Cell culture methods: specific media or culture conditions to optimize growth and function of a particular cell type.

Spec Mapping 1

The practice of some methods disclosed herein employ, unless otherwise indicated, techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA. See for example Sambrook and Green, Molecular Cloning: A Laboratory Manual, 4th Edition (2012); the series Current Protocols in Molecular Biology (F. M. Ausubel, et al. eds.); the series Methods In Enzymology (Academic Press, Inc.), PCR 2: A Practical Approach (M. J. MacPherson, B. D. Hames and G. R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) Antibodies, A Laboratory Manual, and Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition (R. I. Freshney, ed. (2010)) (which is entirely incorporated by reference herein).

Candidate #4

Potential Continuation Candidate 4

Cell culture methods: specific media or culture conditions to optimize growth and function of a particular cell type.

Spec Mapping 2

In Vitro Activity

Candidate #4

Potential Continuation Candidate 5

Methods used to modify a target nucleic acid locus

Spec Mapping 1

The term “expression”, as used herein, generally refers to the process by which a nucleic acid sequence or a polynucleotide is transcribed from a DNA template (such as into mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as “gene product.” If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

Candidate #5

Potential Continuation Candidate 5

Methods used to modify a target nucleic acid locus

Candidate #5

Spec Mapping 2

A library of target plasmids containing a spacer sequence matching that in the minimal array followed by 8N mixed bases (putative PAM sequences) was incubated with the output of the TXTL reaction. After 1-3 hr, the reaction was stopped and the DNA was recovered via a DNA clean-up kit, e.g., Zymo DCC, AMPure XP beads, QiaQuick etc. Adapter sequences were blunt-end ligated to DNA with active PAM sequences that had been cleaved by the endonuclease, whereas DNA that had not been cleaved was inaccessible for ligation. DNA segments comprising active PAM sequences were then amplified by PCR with primers specific to the library and the adapter sequence. The PCR amplification products were resolved on a gel to identify amplicons that corresponded to cleavage events. The amplified segments of the cleavage reaction were also used as template for preparation of an NGS library. Sequencing this resulting library, which was a subset of the starting 8N library, revealed the sequences which contain the correct PAM for the active CRISPR complex. For PAM testing with a single RNA construct, the same procedure was repeated except that an in vitro transcribed RNA was added along with the plasmid library and the tracr/minimal CRISPR array template was omitted. For endonucleases where NGS libraries were prepared, seqLogo (see e.g., Huber et al. Nat Methods. 2015 February; 12(2):115-21) representations were constructed and are presented in FIGS. 27, 38, 29, 30, 31, 32, 33, 34, and 35. The seqLogo module used to construct these representations takes the position weight matrix of a DNA sequence motif (e.g. a PAM sequence) and plots the corresponding sequence logo as introduced by Schneider and Stephens (see e.g. Schneider et al. Nucleic Acids Res. 1990 Oct. 25; 18(20):6097-100. The characters representing the sequence in the seqLogo representations have been stacked on top of each other for each position in the aligned sequences (e.g. PAM sequences). The height of each letter is proportional to its frequency, and the letters have been sorted so the most common one is on top.

Potential Continuation Candidate 5

Methods used to modify a target nucleic acid locus

Candidate #5

Spec Mapping 3

In one aspect, the present disclosure provides for an engineered nuclease system comprising (a) an endonuclease. In some cases, the endonuclease is a Cas endonuclease. In some cases, the endonuclease is a Type II, Class II Cas endonuclease. The endonuclease may comprise a RuvC_III domain, wherein said RuvC_III domain has at least about 70% sequence identity to any one of SEQ ID NOs: 1827-2140. In some cases, the endonuclease may comprise a RuvC_III domain, wherein the RuvC_III domain has 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 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to any one of SEQ ID NOs: 1827-2140. In some cases, the endonuclease may comprise a RuvC_III domain, wherein the substantially identical to any one of SEQ ID NOs: 1827-2140. The endonuclease may comprise a RuvC_III domain having at least about 70% sequence identity to any one of SEQ ID NOs: 1827-1831. In some cases, the endonuclease may comprise a RuvC_III domain having 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 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to any one of SEQ ID NOs: 1827-1831. In some cases, the endonuclease may comprise a RuvC_III domain substantially identical to any one of SEQ ID NOs: 1827-1831. In some cases, the endonuclease may comprise a RuvC_III domain having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to SEQ ID NO: 1827. In some cases, the endonuclease may comprise a RuvC_III domain having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to SEQ ID NO: 1828. In

Potential Continuation Candidate 5

Methods used to modify a target nucleic acid locus

Spec Mapping 4

The practice of some methods disclosed herein employ, unless otherwise indicated, techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA. See for example Sambrook and Green, Molecular Cloning: A Laboratory Manual, 4th Edition (2012); the series Current Protocols in Molecular Biology (F. M. Ausubel, et al. eds.); the series Methods In Enzymology (Academic Press, Inc.), PCR 2: A Practical Approach (M. J. MacPherson, B. D. Hames and G. R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) Antibodies, A Laboratory Manual, and Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition (R. I. Freshney, ed. (2010)) (which is entirely incorporated by reference herein).

Candidate #5

Potential Continuation Candidate 5

The method used to determine if a polynucleotide is derived from genomic dna.

Spec Mapping 5

Engineered strains with PAM sequences integrated into their genomic DNA were transformed with DNA encoding the endonuclease. Transformants were then made chemocompetent and transformed with 50 ng of single guide RNAs either specific to the target sequence (“on target”), or non-specific to the target (“non target”). After heat shock, transformations were recovered in SOC for 2 hrs at 37° C. Nuclease efficiency was then determined by a 5-fold dilution series grown on induction media. Colonies were quantified from the dilution series in triplicate.

Candidate #5

Potential Continuation Candidate 6

The specific process or method used to transcribe mrna in a eukaryotic cell.

Spec Mapping 1

The term “expression”, as used herein, generally refers to the process by which a nucleic acid sequence or a polynucleotide is transcribed from a DNA template (such as into mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as “gene product.” If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

Candidate #6

Potential Continuation Candidate 6

The specific process or method used to transcribe mrna in a eukaryotic cell.

Candidate #6

Spec Mapping 2

In another aspect, the present disclosure provides a method for modifying a target nucleic acid locus. The method may comprise delivering to the target nucleic acid locus any of the non-natural systems disclosed herein, including an enzyme and at least one synthetic guide RNA (sgRNA) disclosed herein. The enzyme may form a complex with the at least one sgRNA, and upon binding of the complex to the target nucleic acid locus, may modify the target nucleic acid locus. Delivering the enzyme to said locus may comprise transfecting a cell with the system or nucleic acids encoding the system. Delivering the nuclease to said locus may comprise electroporating a cell with the system or nucleic acids encoding the system. Delivering the nuclease to said locus may comprise incubating the system in a buffer with a nucleic acid comprising the locus of interest. In some cases, the target nucleic acid locus comprises deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The target nucleic acid locus may comprise genomic DNA, viral DNA, viral RNA, or bacterial DNA. The target nucleic acid locus may be within a cell. The target nucleic acid locus may be in vitro. The target nucleic acid locus may be within a eukaryotic cell or a prokaryotic cell. The cell may be an animal cell, a human cell, bacterial cell, archaeal cell, or a plant cell. The enzyme may induce a single or double-stranded break at or proximal to the target locus of interest.

Potential Continuation Candidate 6

The specific process or method used to transcribe mrna in a eukaryotic cell.

Spec Mapping 3

Whichever transfection method is selected, (i) and (ii) are introduced into cells. A period of incubation is allowed to pass so that the enzyme and/or sgRNA can be transcribed and/or translated into active form. After the incubation period, genomic DNA in the vicinity of the targeting sequence is analyzed (e.g., by sequencing). An indel is introduced into the genomic DNA in the vicinity of the targeting sequence as a result of enzyme-mediated cleavage and non-homologous end joining.

Candidate #6

Potential Continuation Candidate 6

The specific process or method used to transcribe mrna in a eukaryotic cell.

Spec Mapping 4

The terms “transfection” or “transfected” generally refer to introduction of a nucleic acid into a cell by non-viral or viral-based methods. The nucleic acid molecules may be gene sequences encoding complete proteins or functional portions thereof. See, e.g., Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 18.1-18.88.

Candidate #6

Potential Continuation Candidate 6

The specific process or method used to transcribe mrna in a eukaryotic cell.

Spec Mapping 5

The practice of some methods disclosed herein employ, unless otherwise indicated, techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA. See for example Sambrook and Green, Molecular Cloning: A Laboratory Manual, 4th Edition (2012); the series Current Protocols in Molecular Biology (F. M. Ausubel, et al. eds.); the series Methods In Enzymology (Academic Press, Inc.), PCR 2: A Practical Approach (M. J. MacPherson, B. D. Hames and G. R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) Antibodies, A Laboratory Manual, and Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition (R. I. Freshney, ed. (2010)) (which is entirely incorporated by reference herein).

Candidate #6

Potential Continuation Candidate 7

The identification process of cas enzymes from the metagenomic analysis of natural microbial communities.

Candidate #7

Spec Mapping 1

The discovery of new Cas enzymes with unique functionality and structure may offer the potential to further disrupt deoxyribonucleic acid (DNA) editing technologies, improving speed, specificity, functionality, and ease of use. Relative to the predicted prevalence of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems in microbes and the sheer diversity of microbial species, relatively few functionally characterized CRISPR/Cas enzymes exist in the literature. This is partly because a huge number of microbial species may not be readily cultivated in laboratory conditions. Metagenomic sequencing from natural environmental niches that represent large numbers of microbial species may offer the potential to drastically increase the number of new CRISPR/Cas systems known and speed the discovery of new oligonucleotide editing functionalities. A recent example of the fruitfulness of such an approach is demonstrated by the 2016 discovery of CasX/CasY CRISPR systems from metagenomic analysis of natural microbial communities.

Potential Continuation Candidate 7

The identification process of cas enzymes from the metagenomic analysis of natural microbial communities.

Spec Mapping 2

Type IV CRISPR-Cas systems possess an effector complex that consists of a highly reduced large subunit nuclease (csf1), two genes for RAMP proteins of the Cas5 (csf3) and Cas7 (csf2) groups, and, in some cases, a gene for a predicted small subunit; such systems are commonly found on endogenous plasmids.

Candidate #7

Potential Continuation Candidate 7

The identification process of cas enzymes from the metagenomic analysis of natural microbial communities.

Spec Mapping 3

Type I CRISPR-Cas systems are considered of moderate complexity in terms of components. In Type I CRISPR-Cas systems, the array of RNA-targeting elements is transcribed as a long precursor crRNA (pre-crRNA) that is processed at repeat elements to liberate short, mature crRNAs that direct the nuclease complex to nucleic acid targets when they are followed by a suitable short consensus sequence called a protospacer-adjacent motif (PAM). This processing occurs via an endoribonuclease subunit (Cas6) of a large endonuclease complex called Cascade, which also comprises a nuclease (Cas3) protein component of the crRNA-directed nuclease complex. Cas I nucleases function primarily as DNA nucleases.

Candidate #7

Potential Continuation Candidate 7

The identification process of cas enzymes from the metagenomic analysis of natural microbial communities.

Candidate #7

Spec Mapping 4

In one aspect, the present disclosure provides for an engineered nuclease system comprising (a) an endonuclease. In some cases, the endonuclease is a Cas endonuclease. In some cases, the endonuclease is a Type II, Class II Cas endonuclease. The endonuclease may comprise a RuvC_III domain, wherein said RuvC_III domain has at least about 70% sequence identity to any one of SEQ ID NOs: 1827-2140. In some cases, the endonuclease may comprise a RuvC_III domain, wherein the RuvC_III domain has 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 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to any one of SEQ ID NOs: 1827-2140. In some cases, the endonuclease may comprise a RuvC_III domain, wherein the substantially identical to any one of SEQ ID NOs: 1827-2140. The endonuclease may comprise a RuvC_III domain having at least about 70% sequence identity to any one of SEQ ID NOs: 1827-1831. In some cases, the endonuclease may comprise a RuvC_III domain having 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 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to any one of SEQ ID NOs: 1827-1831. In some cases, the endonuclease may comprise a RuvC_III domain substantially identical to any one of SEQ ID NOs: 1827-1831. In some cases, the endonuclease may comprise a RuvC_III domain having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to SEQ ID NO: 1827. In some cases, the endonuclease may comprise a RuvC_III domain having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to SEQ ID NO: 1828. In

Potential Continuation Candidate 8

Engineered nuclease compositions

Spec Mapping 1

As used herein, the term “HNH domain” generally refers to an endonuclease domain having characteristic histidine and asparagine residues. An HNH domain can generally be identified by alignment to known domain sequences, structural alignment to proteins with annotated domains, or by comparison to Hidden Markov Models (HMMs) built based on known domain sequences (e.g., Pfam HMM PF01844 for domain HNH).

Candidate #8

Potential Continuation Candidate 8

Engineered nuclease compositions

Spec Mapping 2

A comparison of MG2-1, MG2-2, MG2-3, MG2-5, and MG2-6 versus additional proprietary protein datasets revealed additional protein sequences with similar architecture, presented as SEQ NOs: 321 and 326-420.

In some embodiments, the endonuclease comprises a sequence at least 70%, 80%, or 90% identical to a sequence selected from the group consisting of SEQ ID NOs: 320-321 or SEQ ID NOs: 320-420

Candidate #8