



US 20180282715A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2018/0282715 A1**

Carter et al.

(43) **Pub. Date:** **Oct. 4, 2018**

(54) **NOVEL CRISPR-ASSOCIATED (CAS)
PROTEIN**

(71) Applicant: **Caribou Biosciences, Inc.**, Berkeley, CA (US)

(72) Inventors: **Matthew Merrill Carter**, Berkeley, CA (US); **Paul Daniel Donohoe**, Berkeley, CA (US)

(21) Appl. No.: **15/937,840**

(22) Filed: **Mar. 27, 2018**

Related U.S. Application Data

(60) Provisional application No. 62/477,494, filed on Mar. 28, 2017, provisional application No. 62/629,641, filed on Feb. 12, 2018.

Publication Classification

(51) **Int. Cl.**

C12N 9/22 (2006.01)
C12N 15/11 (2006.01)
C12N 15/85 (2006.01)
C12N 15/113 (2006.01)

(52) **U.S. Cl.**

CPC **C12N 9/22** (2013.01); **C12N 15/11** (2013.01); **C12N 2800/22** (2013.01); **C12N 15/1136** (2013.01); **C12N 2310/20** (2017.05); **C12N 15/85** (2013.01)

(57)

ABSTRACT

A new CRISPR-associated (Cas) protein, termed “CasM,” is described, as well as polynucleotides encoding the same and methods of using CasM for site-specific genome engineering. CasM proteins are capable of targeting and cleaving single-stranded RNA.

Specification includes a Sequence Listing.

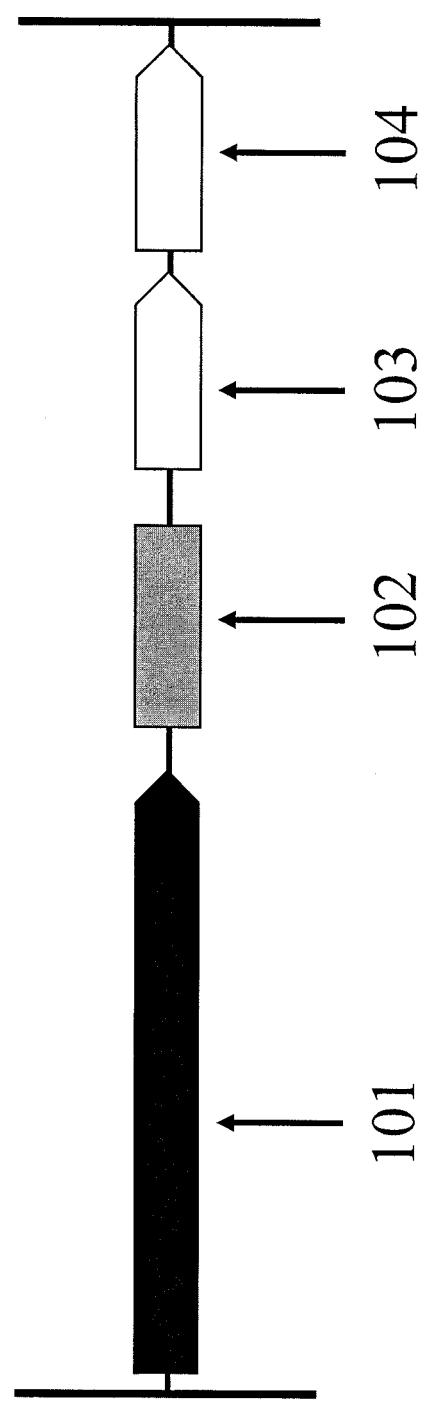


FIG. 1

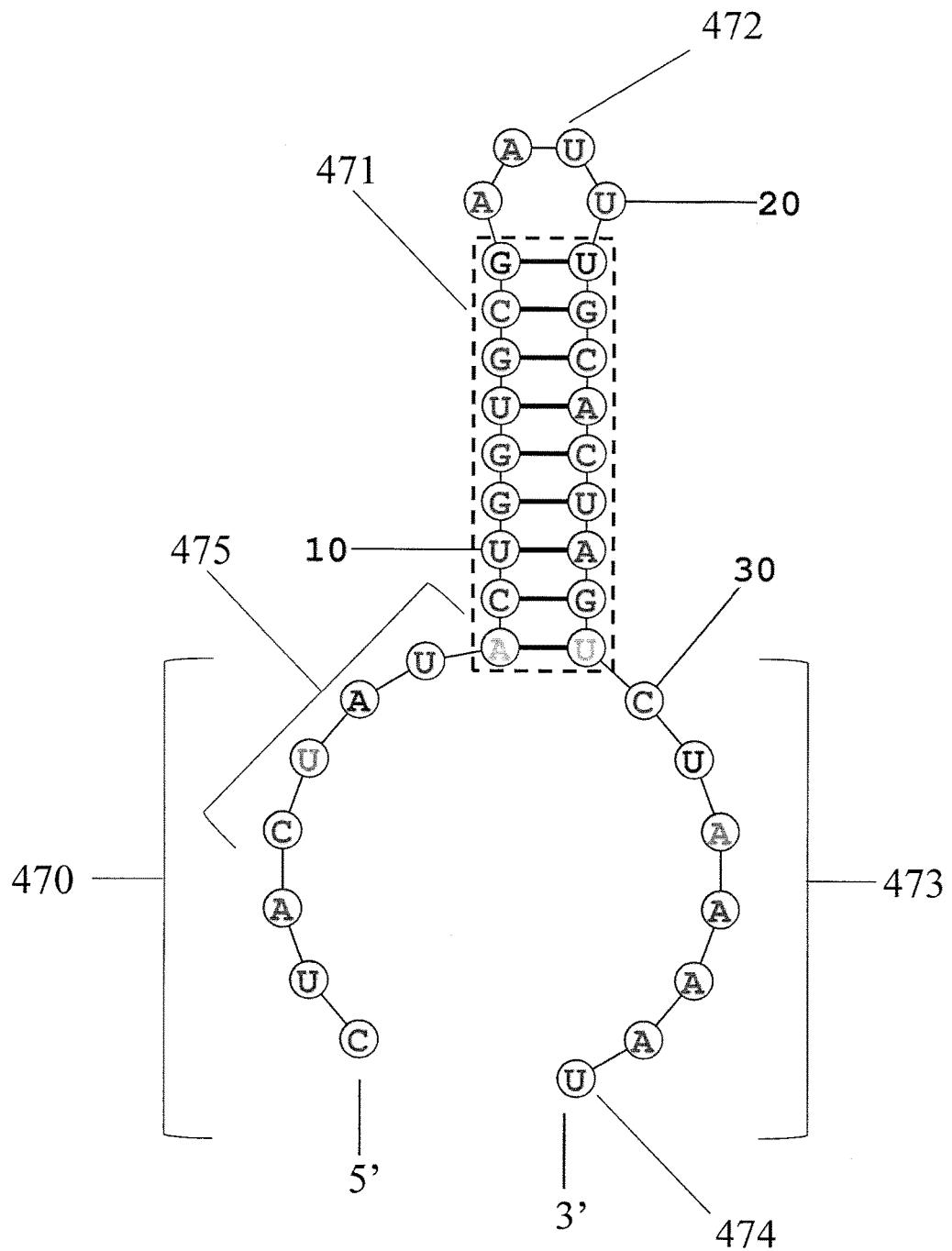


FIG. 2

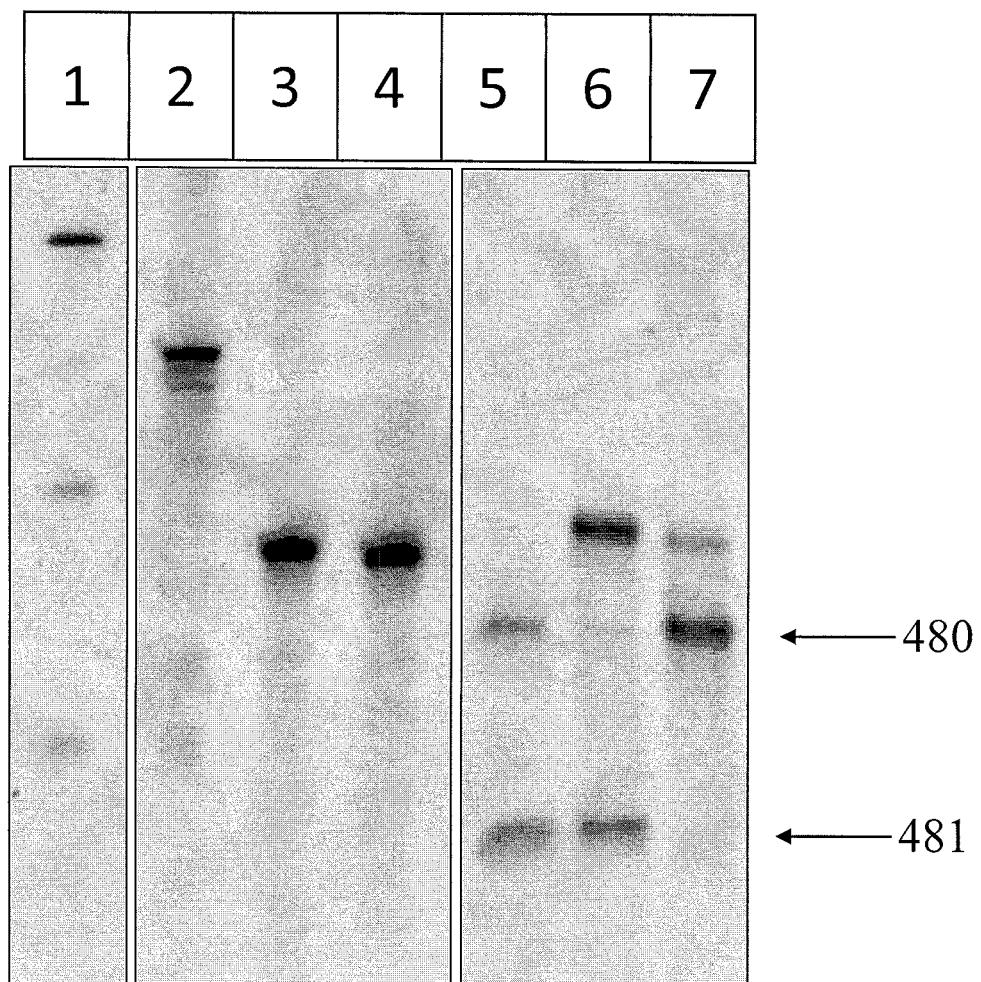


FIG. 3

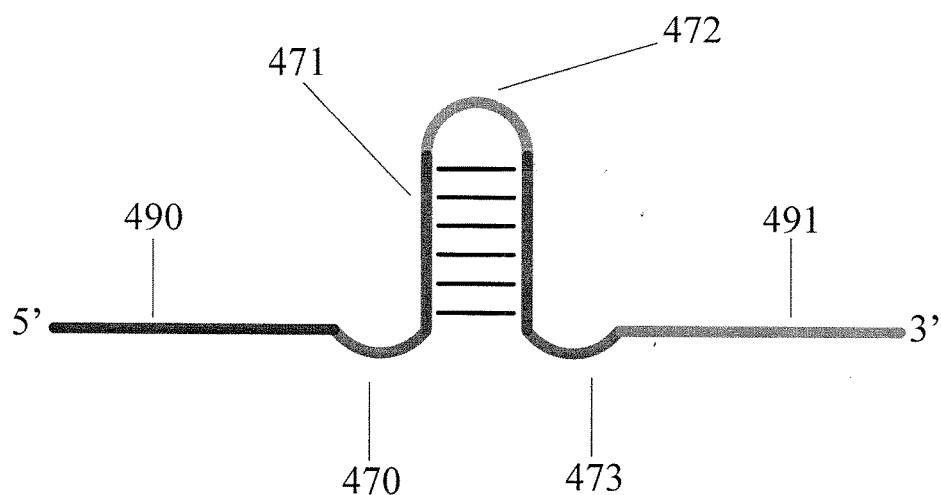


FIG. 4

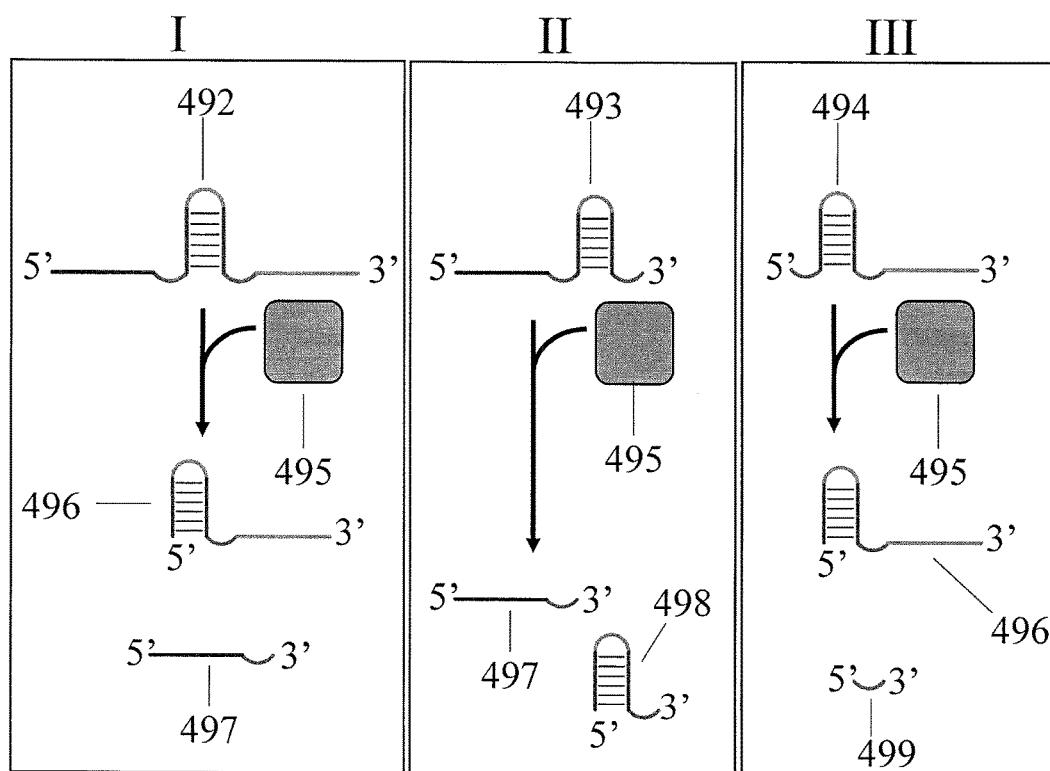


FIG. 5

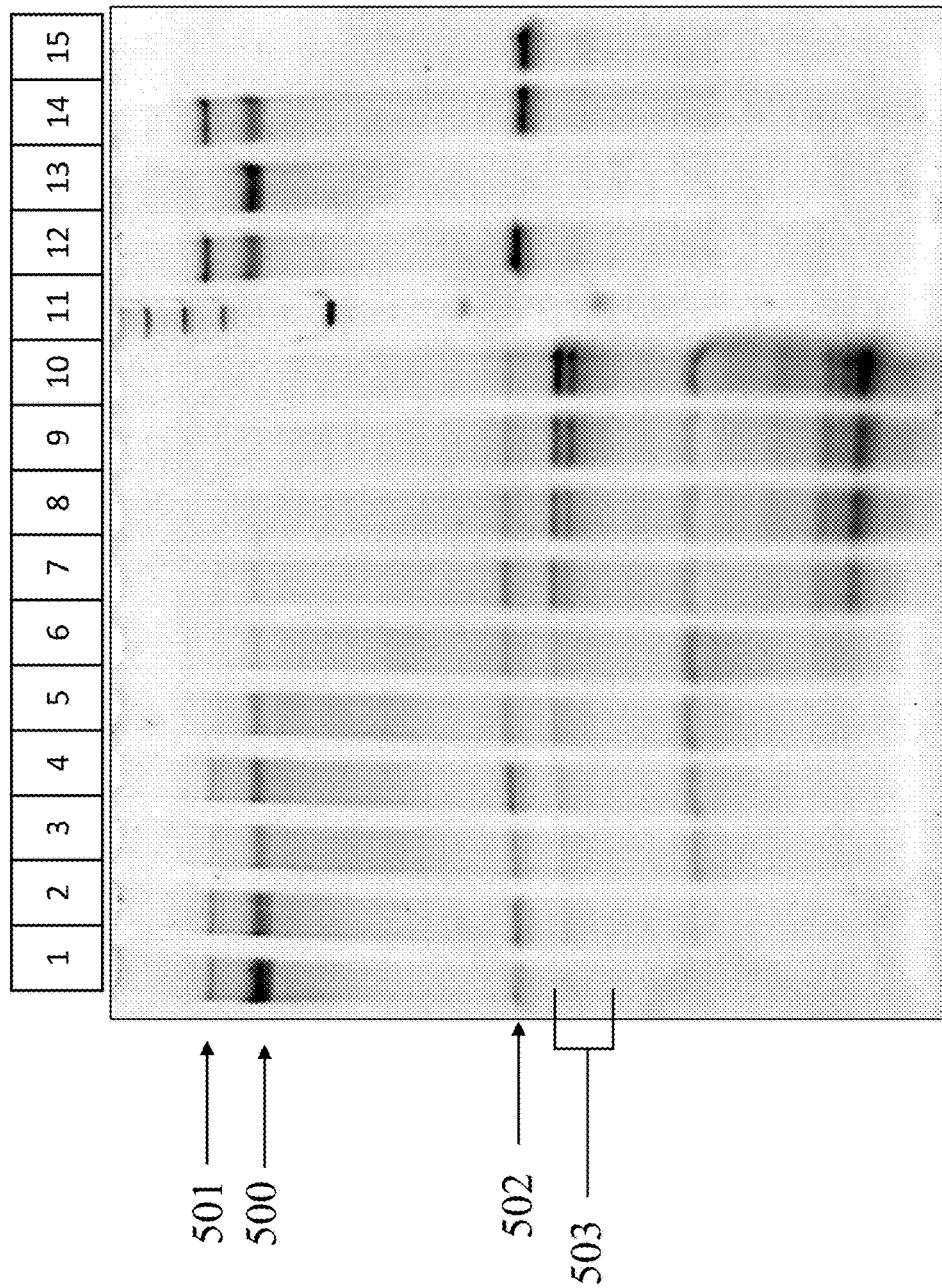


FIG. 6

NOVEL CRISPR-ASSOCIATED (CAS) PROTEIN**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims the benefit under 35 U.S.C. § 119(e)(1) of U.S. Provisional Application Nos. 62/477,494, filed 28 Mar. 2017, and 62/629,641, filed 12 Feb. 2018, which applications are incorporated herein by reference in their entireties.

TECHNICAL FIELD

[0002] The present invention relates to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems. In particular, the invention relates to a new CRISPR-associated (Cas) protein, termed "CasM," and the uses of CasM for site-specific nucleic acid engineering.

BACKGROUND OF THE INVENTION

[0003] Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) proteins are found in prokaryotic immune systems. These systems provide resistance against exogenous genetic elements, such as viruses and plasmids, by targeting their nucleic acids for degradation, in a sequence-specific manner.

[0004] There are several different CRISPR-Cas systems and the nomenclature and classification of these have changed as the systems have been characterized. In particular, CRISPR-Cas systems have now been reclassified into two classes, containing several types and subtypes (Makarova et al., *Nature Reviews Microbiology* (2015) 13:1-15; Shmakov et al., *Nature Reviews Microbiology* (2017) 15:169-182). This classification is based upon identifying all cas genes in a CRISPR-Cas locus and then determining the signature genes in each CRISPR-Cas locus, thereby determining whether the CRISPR-Cas systems should be placed in either Class 1 or Class 2 based upon the genes encoding the effector module, i.e., the proteins involved in the interference stage.

[0005] There remains a need to discover and characterize new CRISPR-associated (Cas) proteins, and their potential use for site-specific nucleic acid engineering.

SUMMARY

[0006] The present invention is based on the discovery of a new Cas protein, termed "CasM" herein. This protein shares no homology to any known Cas protein or to any known protein family.

[0007] Accordingly, in one aspect, the invention is directed to an isolated CasM protein capable of producing a single-strand break at an RNA target site when guided to the RNA target site by a cognate nucleic acid guide. In certain embodiments, the cognate nucleic acid guide comprises RNA, such as crRNA. In additional embodiments, the CasM protein comprises an amino acid sequence selected from the group consisting of the amino acid sequence of SEQ ID NOS:37-44 or 45; an ortholog of the amino acid sequence of SEQ ID NOS:37-44 or 45, i.e., a CasM sequence from a species other than the species producing the reference sequence; and a variant of the amino acid sequence of SEQ ID NOS:37-44 or 45, e.g., an active homolog of the reference amino acid sequence.

[0008] In further embodiments, the invention is directed to a complex comprising a CasM protein, and a cognate nucleic acid guide. In certain embodiments, the cognate nucleic acid guide in the complex comprises a repeat sequence and a spacer sequence, wherein the repeat sequence and the spacer sequence do not naturally occur together. In certain embodiments, the cognate nucleic acid guide comprises a modified base analog.

[0009] In additional embodiments, the cognate nucleic acid guide comprises RNA, such as, but not limited to, crRNA. In some embodiments, the cognate nucleic acid guide, such as crRNA, comprises a spacer sequence that is complementary to a DNA or RNA target sequence that occurs in a prokaryotic or eukaryotic cell.

[0010] In further embodiments, the crRNA/CasM protein complex is capable of binding to a first RNA target sequence complementary to the crRNA spacer sequence, wherein binding of the crRNA/CasM protein complex results in the cleavage of a first RNA target. In additional embodiments, after cleavage of the first RNA target sequence by the crRNA/CasM protein complex, the complex is capable of non-specific endonuclease activity toward any single-stranded RNA in a sequence independent manner.

[0011] In further embodiments, the complex modifies the transcription or translation of a target locus in cell.

[0012] In additional embodiments, the invention is directed to an isolated polynucleotide encoding a CasM protein, wherein the CasM protein is capable of producing a single-strand break at an RNA target site when guided to the RNA target site by a cognate nucleic acid guide. In certain embodiments, the cognate nucleic acid guide comprises RNA, such as crRNA. In additional embodiments, the CasM protein encoded by the polynucleotide comprises an amino acid sequence selected from the group consisting of the amino acid sequence of SEQ ID NOS:37-44 or 45; an ortholog of the amino acid sequence of SEQ ID NOS:37-44 or 45, i.e., a CasM sequence from a species other than the species producing the reference sequence; and a variant of the amino acid sequence of SEQ ID NOS:37-44 or 45, e.g., an active homolog of the reference amino acid sequence.

[0013] In further embodiments, the invention is directed to a modified polynucleotide encoding a CasM protein, wherein the CasM protein is capable of producing a single-strand break at an RNA target site when guided to the RNA target site by a cognate nucleic acid guide, wherein the polynucleotide is modified relative to its native sequence, such as modified for expression in a selected host cell. In additional embodiments, the CasM protein encoded by the polynucleotide comprises an amino acid sequence selected from the group consisting of the amino acid sequence of SEQ ID NOS:37-44 or 45; an ortholog of the amino acid sequence of SEQ ID NOS:37-44 or 45, i.e., a CasM sequence from a species other than the species producing the reference sequence; and a variant of the amino acid sequence of SEQ ID NOS:37-44 or 45, e.g., an active homolog of the reference amino acid sequence.

[0014] In certain embodiments, the polynucleotide is modified for expression in a bacterial cell, such as for expression in an *Escherichia coli* cell. In certain embodiments, the polynucleotide comprises the sequence of SEQ ID NOS:2-8 or 9.

[0015] In other embodiments, the polynucleotide is modified for expression in a eukaryotic cell, e.g., a mammalian

cell, such as a human cell. In certain embodiments, the polynucleotide comprises the sequence of SEQ ID NOS:10-17 or 18.

[0016] In additional embodiments, the polynucleotide is modified for expression in a plant cell, such as for expression in a *Zea mays* (corn) cell. In certain embodiments the polynucleotide comprises the sequence of SEQ ID NOS:19-26 or 27.

[0017] In further embodiments, the invention is directed to a recombinant vector comprising a polynucleotide or modified polynucleotide as described herein, and at least one control element operably linked to the polynucleotide, whereby a CasM coding sequence in the polynucleotide is capable of being transcribed and translated in a host cell. In certain embodiments, at least one of the control elements is heterologous to the coding system.

[0018] In additional embodiments, the CasM protein encoded by the polynucleotide comprises an amino acid sequence selected from the group consisting of the amino acid sequence of SEQ ID NOS:37-44 or 45; an ortholog of the amino acid sequence of SEQ ID NOS:37-44 or 45; and a variant of the amino acid sequence of SEQ ID NOS:37-44 or 45.

[0019] In further embodiments, the invention is directed to a host cell transformed with a recombinant vector described herein. In certain embodiments, the host cell is a prokaryotic or eukaryotic cell.

[0020] In additional embodiments, the invention is directed to a method of producing a CasM protein comprising providing a population of host cells transformed with a recombinant vector as described herein; and culturing the population of cells under conditions whereby the CasM protein encoded by the polynucleotide present in the recombinant vector is expressed.

[0021] In further embodiments, the invention is directed to a eukaryotic host cell comprising a CasM protein of a complex comprising the CasM protein, as described herein.

[0022] In additional embodiments, the invention is directed to a method of directing a CasM protein to a selected nucleic acid target sequence, comprising contacting the selected nucleic acid target sequence with a cognate nucleic acid guide/CasM complex that targets said selected nucleic acid target sequence, whereby the CasM protein is delivered to the nucleic acid target sequence. In certain embodiments, the nucleic acid target sequence comprises RNA, such as mRNA. In further embodiments, the method comprises producing one or more single- or double-strand breaks in the target sequence.

[0023] In additional embodiments, the method is performed in a cell, such as a prokaryotic or eukaryotic cell. In certain embodiments, the cell constitutively expresses the CasM protein. In other embodiments, e.g., when the cell does not constitutively express the CasM protein, the cognate nucleic acid guide is complexed to the CasM protein prior to delivery to the nucleic acid target sequence. In other embodiments, the cell constitutively expresses the CasM protein and the cognate nucleic acid guide. In additional embodiments, the complex modifies the transcription or translation of a selected nucleic acid sequence in a host cell, such as a RNA sequence.

[0024] In further embodiments of the methods, the CasM protein comprises an amino acid sequence selected from the group consisting of the amino acid sequence of SEQ ID NOS:37-44 or 45; an ortholog of the amino acid sequence of

SEQ ID NOS:37-44 or 45; and a variant of the amino acid sequence of SEQ ID NOS:37-44 or 45.

[0025] In other embodiments of the methods, a donor polynucleotide is delivered to the region of the selected nucleic acid target sequence.

[0026] In additional embodiments, the CasM protein is capable of processing the CRISPR repeat-spacer array into individual repeat-spacer elements. The CasM protein cleaves the array within the 5' region of each repeat sequence, giving rise to a processed crRNA comprising, in a 5' to 3' direction, a repeat sequence and a spacer element.

[0027] In some embodiments the repeat sequence comprises a secondary structure that is recognized by the CasM protein. The secondary structure of the repeat may comprise a stem, a stem-loop duplex, a pseudoknot, or a tripartite duplex. CasM protein homologs may only recognize the repeat sequence or secondary structure of their cognate repeat elements. Alternatively, CasM protein homologs may recognize the repeat sequence or secondary structure of non-cognate repeat elements.

[0028] In some embodiments the crRNA/CasM complex is capable of sequence-specific single-stranded RNA activity. Recognition and cleavage of an initial ssRNA complementary to the crRNA target sequence activates the CasM protein to carry out endonuclease activity toward any single-stranded RNA in a sequence-independent manner. The sequence-specific recognition of RNA of the crRNA/CasM complex facilitates the target knockdown of gene transcripts perturbing translation of a specific protein. The non-specific endonuclease activity of an activated crRNA/CasM complex in a cellular environment can result in cell death due to depletion of RNA encoding for essential gene transcripts. The specific RNA targeting and collateral endonuclease activity of an activated crRNA/CasM complex enables the sequence-specific selection of cells expressing a RNA transcript.

[0029] In a further aspect, the present invention relates to a method of screening and killing cells that have not been modified by a DNA targeting nuclease (e.g., a Type II Cas9 nuclease). This method comprises contacting a crRNA/Cas9 complex to a locus of interest in a population of cells. Contacting the NATNA/Cas9 complex results in DNA cleavage and subsequent repair of the break by the endogenous cellular repair machine and the introduction of insertion and deletions ("indels") at the break site. The targeting of the NATNA/Cas9 to a targeted locus that encodes an RNA transcript results in indels in an RNA transcript sequence. This modified RNA transcript sequence is different compared to a transcript from an unmodified cell (a wild-type cell). A cognate nucleic acid guide/CasM complex can then be targeted to the unmodified transcript, wherein recognition of the unmodified transcript by the complex results in activation of the sequence independent, single-stranded RNA targeting activity of the CasM protein and subsequent cell death. Alternatively, this method can be adapted to screen for the incorporation of a donor-polynucleotide into NATNA/Cas9 break site.

[0030] These aspects and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

INCORPORATION BY REFERENCE

[0031] All publications, patents, and patent applications mentioned in this specification are herein incorporated by

reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

SEQUENCE LISTING

[0032] The sequences referred to herein are listed in the Sequence Listing submitted as an ASCII text file entitled “CBI025 10_ST25.txt”-300 KB and was created on Mar. 22, 2018. The Sequence Listing entitled “CBI025 10_ST25.txt” is incorporated herein by reference in its entirety.

BRIEF DESCRIPTION OF THE FIGURES

[0033] FIG. 1 depicts a representative CasM operon from *Eubacterium siraeum* (NCBI Accession No. NZ_DS499551.1).

[0034] FIG. 2 shows a structure of a CasM repeat sequence (SEQ ID NO:51).

[0035] FIG. 3 shows the results of the in vitro CRISPR array cleavage assay described in the Examples.

[0036] FIG. 4 shows a depiction of a synthetic CasM CRISPR array.

[0037] FIG. 5 shows a representation of the results of the in vitro CRISPR array cleavage assay results shown in FIG. 3 and described in the Examples.

[0038] FIG. 6 shows the results of the CasM ssRNA cleavage assay described in the Examples.

DETAILED DESCRIPTION OF THE INVENTION

[0039] It is to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms “a,” “an” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to a “guide/Cas complex” includes one or more such complexes, reference to “a polynucleotide” includes one or more polynucleotides, etc.

[0040] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although other methods and materials similar, or equivalent, to those described herein can be used in the practice of the present invention, preferred materials and methods are described herein.

[0041] In view of the teachings of the present specification, one of ordinary skill in the art can apply conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics, and recombinant polynucleotides, as taught, for example, by the following standard texts: Antibodies: A Laboratory Manual, Second edition, E. A. Greenfield, 2014, Cold Spring Harbor Laboratory Press, ISBN 978-1-936113-81-1; Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition, R. I. Freshney, 2010, Wiley-Blackwell, ISBN 978-0-470-52812-9; Transgenic Animal Technology, Third Edition: A Laboratory Handbook, 2014, C. A. Pinkert, Elsevier, ISBN 978-0124104907; The Laboratory Mouse, Second Edition, 2012, H. Hedrich, Academic Press, ISBN 978-0123820082; Manipulating the Mouse Embryo: A Laboratory Manual, 2013, R. Behringer, et al., Cold Spring Harbor Laboratory Press, ISBN 978-1936113019; PCR 2: A Practical Approach, 1995, M. J.

McPherson, et al., IRL Press, ISBN 978-0199634248; Methods in Molecular Biology (Series), J. M. Walker, ISSN 1064-3745, Humana Press; RNA: A Laboratory Manual, 2010, D. C. Rio, et al., Cold Spring Harbor Laboratory Press, ISBN 978-0879698911; Methods in Enzymology (Series), Academic Press; Molecular Cloning: A Laboratory Manual (Fourth Edition), 2012, M. R. Green, et al., Cold Spring Harbor Laboratory Press, ISBN 978-1605500560; Bioconjugate Techniques, Third Edition, 2013, G. T. Hermanson, Academic Press, ISBN 978-0123822390; Methods in Plant Biochemistry and Molecular Biology, 1997, W. V. Dashek, CRC Press, ISBN 978-0849394805; Plant Cell Culture Protocols (Methods in Molecular Biology), 2012, V. M. Loyola-Vargas, et al., Humana Press, ISBN 978-1617798177; Plant Transformation Technologies, 2011, C. N. Stewart, et al., Wiley-Blackwell, ISBN 978-0813821955; Recombinant Proteins from Plants (Methods in Biotechnology), 2010, C. Cunningham, et al., Humana Press, ISBN 978-1617370212; Plant Genomics: Methods and Protocols (Methods in Molecular Biology), 2009, D. J. Somers, et al., Humana Press, ISBN 978-1588299970; Plant Biotechnology: Methods in Tissue Culture and Gene Transfer, 2008, R. Keshavachandran, et al., Orient Blackswan, ISBN 978-8173716164.

[0042] As used herein, “a CasM protein” refers to a CRISPR protein capable of targeting RNA and causing single-strand RNA breaks when guided to a target site by a crRNA, without the necessity of association with a tracrRNA. CasM proteins typically include two or more higher eukaryotic and prokaryotic nucleotide-binding (HEPN) domains found in protein family PF05168 in the C-terminal region of the CasM sequence. CasM proteins show synteny with one or more WYL domain-containing proteins and sometimes with RtcB (RNA 3'-terminal phosphate cyclase, group B) domain-containing proteins. Based on the foregoing characteristics, CasM may be classified as a Class 2 Type VI CRISPR-Cas system because it is a single effector protein containing two HEPN domains used for targeted ssRNA interference. However, CasM has a very low degree of sequence similarity to other Type VI subtypes. Exemplary CasM proteins are shown in SEQ ID NOS:37-45, and are encoded by polynucleotides shown in SEQ ID NOS:28-36, respectively. These proteins display approximately 13.59% to 99.82% sequence identity to each other and show less than 8% sequence identity with other known CRISPR-Cas proteins. As used herein, the term “CasM protein” refers to a CasM protein derived from any species, subspecies, or strain of bacteria that encodes the CasM protein, as well as an ortholog of the CasM protein, i.e., a CasM protein from a species other than the species producing the reference CasM protein. For example, CasM orthologs of *Eubacterium siraeum* CasM, shown in Table 1, display approximately 13.59% to 99.82% sequence identity to each other. Thus, CasM orthologs are identified based on the CasM characteristics detailed herein. Reference to a CasM protein also encompasses a variant of the reference CasM protein, e.g., an active homolog of the reference amino acid sequence. Thus, CasM proteins include, but are not limited to, those proteins depicted in SEQ ID NOS:37-45, orthologs thereof, or variants thereof. Non-limiting examples of such proteins include CasM proteins from *Eubacterium siraeum*; *Ruminococcus* sp., such as from *Ruminococcus bicirculans*; *Ruminococcus flavefaciens*, such as, but not limited to, FD-1 and strain XPD3002; *Ruminococcus albus* such as, but not

limited to, strain KH2T6; *Ruminococcus* sp. isolates, such as but not limited to, isolates 2789STDY5834971, 2789STDY5608892 and 2789STDY5834894.

[0043] By “dCasM protein” is meant a deactivated CasM protein lacking activity, such as catalytic and/or binding activity, also termed “dead CasM.” Such molecules lack all or a portion of biological activity, such as nuclease and/or binding activity, and are therefore unable to bind and/or cleave a target nucleic acid of interest, respectively. In some embodiments, these deactivated CasM proteins can be used to regulate genes in a nucleic acid-guided manner. This is accomplished by introducing mutations that inactivate CasM nuclease function and typically involves mutating catalytic residues of the gene encoding CasM. dCasM can be used alone or in fusions to synthetically repress (CRISPR interference or CRISPRi) or activate (CRISPR activation or CRISPRa) gene expression. CRISPRi can work independently of host cellular machineries. In some embodiments a dCasM protein and a customized nucleic acid-targeting nucleic acid, i.e., a cognate nucleic acid guide designed with a complementary region to any gene of interest, are used to direct dCasM to a chosen genomic location. In other embodiments, dCasM can be fused to a transcription factor, such as a repressor, and the fused dCasM-transcription factor can then work in concert with cellular machineries. CRISPRa is carried out by dCasM-transcription factor (activator) fusions.

[0044] A “nucleic acid-targeting nucleic acid” (NATNA), as used herein, refers to one or more polynucleotides that guide a protein, such as a CasM protein, to preferentially target a nucleic acid target sequence present in a polynucleotide (relative to a polynucleotide that does not comprise the nucleic acid target sequence). Such NATNAs are also known herein as “cognate nucleic acid guides,” or “cognate guides.” NATNAs can comprise ribonucleotide bases (e.g., RNA), deoxyribonucleotide bases (e.g., DNA), combinations of ribonucleotide bases and deoxyribonucleotide bases (e.g., RNA/DNA), nucleotides, nucleotide analogs, modified nucleotides, and the like, as well as synthetic, naturally occurring, and non-naturally occurring modified backbone residues or linkages. Thus, a NATNA as used herein site-specifically guides a CasM, or a deactivated CasM, to a target nucleic acid. Many such NATNAs are known, such as but not limited to sgRNA (including miniature and truncated single-guide RNAs), crRNA, dual-guide RNA, including but not limited to, crRNA/tracrRNA molecules, as described herein, and the like, the use of which depends on the particular Cas protein. For a non-limiting description of exemplary NATNAs, see, e.g., PCT Publication No. WO 2014/150624 to May et al., published Sep. 29, 2014; PCT Publication No. WO 2015/200555 to May et al., published Mar. 10, 2016; PCT Publication No. WO 2016/201155 to Donohoue et al., published Dec. 15, 2016; PCT Publication No. WO 2017/027423 to Donohoue et al., published Feb. 16, 2017; and PCT Publication No. WO 2016/123230 to May et al., published Aug. 4, 2016; each of which is incorporated herein by reference in its entirety.

[0045] With reference to a NATNA or a cognate nucleic acid guide, a “spacer,” “spacer sequence,” or “spacer element,” as used herein, refers to the polynucleotide sequence that can specifically hybridize to a target nucleic acid sequence. The spacer element interacts with the target nucleic acid sequence through hydrogen bonding between complementary base pairs (i.e., paired bases). A spacer

element binds to a selected nucleic acid target sequence. Accordingly, the spacer element is the nucleic acid target-binding sequence. The spacer element determines the location of a Cas protein’s site-specific binding and nucleolytic cleavage. Spacer elements range from approximately 17 to approximately 84 nucleotides in length and have an average length of 36 nucleotides (see, e.g., Marraffini, et al., “CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea,” *Nature reviews Genetics* (2010) 11:181-190). Variability of the functional length for a spacer element is known in the art (e.g., U.S. Patent Publication 2014/0315985 to May et al., published Oct. 23, 2014, incorporated herein by reference in its entirety). The terms “nucleic acid target binding sequence” and “spacer sequence” are used interchangeably herein.

[0046] The term “sgRNA” typically refers to a single-guide RNA (i.e., a single, contiguous polynucleotide sequence) that essentially comprises a crRNA connected at its 3' end to the 5' end of a tracrRNA through a “loop” sequence (see, e.g., U.S. Published Patent Application No. 2014/0068797 to Duudna et al., published 6 Mar. 2014, incorporated herein by reference in its entirety). sgRNA interacts with a cognate Cas protein essentially as described for tracrRNA/crRNA polynucleotides. Similar to crRNA, sgRNA has a spacer, a region of complementarity to a potential DNA or RNA target sequence, adjacent a second region that forms base-pair hydrogen bonds that form a secondary structure, typically a stem structure. The term includes truncated single-guide RNAs (tru-sgRNAs) of approximately 17-18 nucleotides (nt) (see, e.g., Fu et. al., *Nat Biotechnol.* (2014) 32:279-284). The term also encompasses functional miniature sgRNAs with expendable features removed, but that retain an essential and conserved module termed the “nexus” located in the portion of sgRNA that corresponds to tracrRNA (not crRNA). See, e.g., U.S. Patent Publication 2014/0315985 to May et al., published Oct. 23, 2014, incorporated herein by reference in its entirety; Briner et al., “Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality,” *Molecular Cell* (2014) 56:333-339.

[0047] As used herein, “dual-guide RNA” refers to a two-component RNA system for a polynucleotide component capable of associating with a cognate Cas protein. A representative CRISPR Class 2 Type II CRISPR-Cas-associated dual-guide RNA includes a Cas-crRNA and Cas-tracrRNA, paired by hydrogen bonds to form secondary structure (see, e.g., U.S. Published Patent Application No. 2014/0068797 to Doudna et al., published 6 Mar. 2014, incorporated herein by reference in its entirety; see also Jinck M., et al., *Science* 337:816-21 (2012)). A Cas-dual-guide RNA is capable of forming a nucleoprotein complex with a cognate Cas protein, wherein the complex is capable of targeting a nucleic acid target sequence complementary to the spacer sequence.

[0048] As used herein, the term “cognate” typically refers to a Cas protein (e.g., CasM protein) and one or more polynucleotides (e.g., a CRISPR-CasM-associated cognate nucleic acid guide) capable of forming a nucleoprotein complex for site-directed binding to a nucleic acid target sequence complementary to the nucleic acid target binding sequence present in one of the one or more polynucleotides.

[0049] The terms “wild-type,” “naturally-occurring,” “native,” and “unmodified” are used herein to mean the typical (or most common) form, appearance, phenotype, or

strain existing in nature; for example, the typical form of cells, organisms, characteristics, polynucleotides, proteins, macromolecular complexes, genes, RNAs, DNAs, or genomes as they occur in and can be isolated from a source in nature. The wild-type form, appearance, phenotype, or strain serve as the original parent before an intentional modification. Thus, mutant, variant, engineered, recombinant, and modified forms are not wild-type forms.

[0050] As used herein, the terms “engineered,” “genetically engineered,” “recombinant,” “modified,” and “non-naturally occurring” are interchangeable and indicate intentional human manipulation.

[0051] “Covalent bond,” “covalently attached,” “covalently bound,” “covalently linked,” “covalently connected,” and “molecular bond” are used interchangeably herein, and refer to a chemical bond that involves the sharing of electron pairs between atoms. Examples of covalent bonds include, but are not limited to, phosphodiester bonds and phosphorothioate bonds.

[0052] “Non-covalent bond,” “non-covalently attached,” “non-covalently bound,” “non-covalently linked,” “non-covalent interaction,” and “non-covalently connected” are used interchangeably herein, and refer to any relatively weak chemical bond that does not involve sharing of a pair of electrons. Multiple non-covalent bonds often stabilize the conformation of macromolecules and mediate specific interactions between molecules. Examples of non-covalent bonds include, but are not limited to hydrogen bonding, ionic interactions (e.g., Na^+Cl^-), van der Waals interactions, and hydrophobic bonds.

[0053] As used herein, “hydrogen bonding,” “hydrogen base pairing,” and “hydrogen bonded” are used interchangeably and refer to canonical hydrogen bonding and non-canonical hydrogen bonding including, but not limited to, “Watson-Crick-hydrogen-bonded base pairs” (W-C-hydrogen-bonded base pairs or W-C hydrogen bonding); “Hoogsteen-hydrogen-bonded base pairs” (Hoogsteen hydrogen bonding); and “wobble-hydrogen-bonded base pairs” (wobble hydrogen bonding). W-C hydrogen bonding, including reverse W-C hydrogen bonding, refers to purine-pyrimidine base pairing, that is, adenine:thymine, guanine:cytosine, and uracil:adenine. Hoogsteen hydrogen bonding, including reverse Hoogsteen hydrogen bonding, refers to a variation of base pairing in nucleic acids wherein two nucleobases, one on each strand, are held together by hydrogen bonds in the major groove. This non-W-C hydrogen bonding can allow a third strand to wind around a duplex and form triple-stranded helices. Wobble hydrogen bonding, including reverse wobble hydrogen bonding, refers to a pairing between two nucleotides in RNA molecules that does not follow Watson-Crick base pair rules. There are four major wobble base pairs: guanine:uracil, inosine(hypoxanthine):uracil, inosine-adenine, and inosine-cytosine. Rules for canonical hydrogen bonding and non-canonical hydrogen bonding are known to those of ordinary skill in the art (see, e.g., *The RNA World*, Third Edition (Cold Spring Harbor Monograph Series), R. F. Gesteland, Cold Spring Harbor Laboratory Press, ISBN 978-0879697396 (2005); *The RNA World*, Second Edition (Cold Spring Harbor Monograph Series), R. F. Gesteland, et al., Cold Spring Harbor Laboratory Press, ISBN 978-0879695613 (1999); *The RNA World* (Cold Spring Harbor Monograph Series), R. F. Gesteland, et al., Cold Spring Harbor Laboratory Press, ISBN 978-0879694562 (1993) (see, e.g., Appendix 1: Struc-

tures of Base Pairs Involving at Least Two Hydrogen Bonds, I. Tinoco); *Principles of Nucleic Acid Structure*, W. Saenger, Springer International Publishing AG, ISBN 978-0-387-90761-1 (1988); *Principles of Nucleic Acid Structure*, First Edition, S. Neidle, Academic Press, ISBN 978-01236950791 (2007)).

[0054] “Connect,” “connected,” and “connecting” are used interchangeably herein, and refer to a covalent bond or a non-covalent bond between two macromolecules (e.g., polynucleotides, proteins, and the like). Thus, CasM and a cognate nucleic acid guide are “connected” in a cognate nucleic acid guide/CasM complex.

[0055] As used herein, the terms “nucleic acid,” “nucleic acid sequence,” “nucleotide sequence,” “oligonucleotide,” and “polynucleotide” are interchangeable and refer to a polymeric form of nucleotides. The nucleotides may be deoxyribonucleotides (DNA), ribonucleotides (RNA), analogs thereof, or combinations thereof, and may be of any length. Polynucleotides may perform any function and may have any secondary and tertiary structures. The terms encompass known analogs of natural nucleotides and nucleotides that are modified in the base, sugar and/or phosphate moieties. Analogs of a particular nucleotide have the same base-pairing specificity (e.g., an analog of A base pairs with T). A polynucleotide may comprise one modified nucleotide or multiple modified nucleotides. Examples of modified nucleotides include fluorinated nucleotides, methylated nucleotides, and nucleotide analogs. Nucleotide structure may be modified before or after a polymer is assembled. Following polymerization, polynucleotides may be additionally modified via, for example, conjugation with a labeling component or target binding component. A nucleotide sequence may incorporate non-nucleotide components. The terms also encompass nucleic acids comprising modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, and have similar binding properties as a reference polynucleotide (e.g., DNA or RNA). Examples of such analogs include, but are not limited to, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs), Locked Nucleic Acid (LNATM) (Exiqon, Inc., Woburn, Mass.) nucleosides, glycol nucleic acid, bridged nucleic acids, and morpholino structures.

[0056] Peptide-nucleic acids (PNAs) are synthetic homologs of nucleic acids wherein the polynucleotide phosphate-sugar backbone is replaced by a flexible pseudo-peptide polymer. Nucleobases are linked to the polymer. PNAs have the capacity to hybridize with high affinity and specificity to complementary sequences of RNA and DNA.

[0057] In phosphorothioate nucleic acids, the phosphorothioate (PS) bond substitutes a sulfur atom for a non-bridging oxygen in the polynucleotide phosphate backbone. This modification makes the internucleotide linkage resistant to nuclelease degradation. In some embodiments, phosphorothioate bonds are introduced between the last 3 to 5 nucleotides at the 5'-end or 3'-end sequences of a polynucleotide sequence to inhibit exonuclease degradation. Placement of phosphorothioate bonds throughout an entire oligonucleotide helps reduce degradation by nucleases as well.

[0058] Threose nucleic acid (TNA) is an artificial genetic polymer. The backbone structure of TNA comprises repeating threose sugars linked by phosphodiester bonds. TNA

polymers are resistant to nuclease degradation. TNA can self-assemble by base-pair hydrogen bonding into duplex structures.

[0059] Linkage inversions can be introduced into polynucleotides through use of “reversed phosphoramidites” (see, e.g., ucalgary.ca/dnalab/synthesis/-modifications/linkages). A 3'-3' linkage at a terminus of a polynucleotide stabilizes the polynucleotide to exonuclease degradation by creating an oligonucleotide having two 5'-OH termini but lacking a 3'-OH terminus. Typically, such polynucleotides have phosphoramidite groups on the 5'-OH position and a dimethoxytrityl (DMT) protecting group on the 3'-OH position. Normally, the DMT protecting group is on the 5'-OH and the phosphoramidite is on the 3'-OH.

[0060] Polynucleotide sequences are displayed herein in the conventional 5' to 3' orientation unless otherwise indicated.

[0061] As used herein, the term “complementarity” refers to the ability of a nucleic acid sequence to form hydrogen bond(s) with another nucleic acid sequence (e.g., through traditional Watson-Crick base pairing). A percent complementarity indicates the percentage of residues in a nucleic acid molecule that can form hydrogen bonds with a second nucleic acid sequence. When two polynucleotide sequences have 100% complementarity, the two sequences are perfectly complementary, i.e., all of a first polynucleotide's contiguous residues hydrogen bond with the same number of contiguous residues in a second polynucleotide.

[0062] As used herein, “binding” refers to a non-covalent interaction between macromolecules (e.g., between a protein and a polynucleotide, between a polynucleotide and a polynucleotide, or between a protein and a protein, and the like). Such non-covalent interaction is also referred to as “associating” or “interacting” (e.g., if a first macromolecule interacts with a second macromolecule, the first macromolecule binds to second macromolecule in a non-covalent manner). Some portions of a binding interaction may be sequence-specific (the terms “sequence-specific binding,” “sequence-specifically bind,” “site-specific binding,” and “site specifically binds” are used interchangeably herein). Sequence-specific binding, as used herein, typically refers to one or more cognate nucleic acid guides (i.e., NATNAs) capable of forming a complex with a protein (e.g., a CasM protein) to cause the protein to bind a nucleic acid sequence (e.g., a RNA or DNA sequence) comprising a nucleic acid target sequence (e.g., a RNA or DNA target sequence) preferentially relative to a second nucleic acid sequence (e.g., a second RNA or DNA sequence) without the nucleic acid target binding sequence (e.g., the RNA or DNA target binding sequence). All components of a binding interaction do not need to be sequence-specific, such as contacts of a protein with phosphate residues in a DNA backbone. Binding interactions can be characterized by a dissociation constant (K_d). “Binding affinity” refers to the strength of the binding interaction. An increased binding affinity is correlated with a lower K_d .

[0063] As used herein, a Cas protein (e.g., a CasM protein) is said to “target” a polynucleotide if a cognate nucleic acid/Cas protein nucleoprotein complex associates with, binds and/or cleaves a polynucleotide at the nucleic acid target sequence within the polynucleotide.

[0064] As used herein, “single-strand break” (SSB) refers to cleavage of a single strand of RNA or DNA. A “double-strand break” (DSB) refers to both strands of a double-

stranded segment of nucleic acid being severed. In some instances, if such a break occurs, one strand can be said to have a “sticky end” wherein nucleotides are exposed and not hydrogen bonded to nucleotides on the other strand. In other instances, a “blunt end” can occur wherein both strands remain fully base paired with each other.

[0065] As used herein, the term “recombination” refers to a process of exchange of genetic information between two polynucleotides.

[0066] As used herein, “nucleic acid repair,” such as but not limited to DNA repair, encompasses any process whereby cellular machinery repairs damage to a nucleic acid molecule contained in the cell. The damage repaired can include single-strand breaks or double-strand breaks (DSBs). At least three mechanisms exist to repair DSBs: homology-directed repair (HDR), classical non-homologous end joining (c-NHEJ), and microhomology-mediated end joining (MMEJ), all defined below. “Nucleic acid repair” is also used herein to refer to nucleic acid repair resulting from human manipulation, wherein a target locus is modified, e.g., by inserting, deleting, or substituting nucleotides, all of which represent forms of genome editing.

[0067] As used herein, the term “homology-directed repair” or “HDR” refers to nucleic acid repair that takes place in cells, for example, during repair of double-strand and single-strand breaks in a nucleic acid molecule, such as DNA. HDR requires nucleotide sequence homology and uses a “donor template” (donor template nucleic acid, such as DNA, polynucleotide donor, or oligonucleotide (used interchangably herein) to repair the sequence where the double-strand break occurred (e.g., DNA target sequence). This results in the transfer of genetic information from, for example, the donor template DNA to the DNA target sequence. HDR may result in alteration of the nucleic acid target sequence (e.g., insertion, deletion, mutation) if the donor template sequence or oligonucleotide sequence differs from the target sequence and part or all of the donor template polynucleotide or oligonucleotide is incorporated into the target sequence. In some embodiments, an entire donor template polynucleotide, a portion of the donor template polynucleotide, or a copy of the donor polynucleotide is copied or integrated at the site of the target sequence.

[0068] By “donor polynucleotide” is meant a polynucleotide that can be directed to, and inserted into a target site of interest, such as an integration locus, to modify the target nucleic acid. All or a portion of the donor polynucleotide can be inserted into the target nucleic acid. The donor polynucleotide can be used for repair of the break in the target nucleic acid sequence resulting in the transfer of genetic information (i.e., polynucleotide sequences) from the donor at the site or in close proximity of the break. Accordingly, new genetic information (i.e., polynucleotide sequences) may be inserted or copied at a target site. The donor polynucleotide can be double- or single-stranded RNA, DNA, a vector, plasmid, or the like. Thus, a donor polynucleotide can be an insertion cassette, a recombinase expression vector, and the like. Non-symmetrical polynucleotide donors can also be used that are composed of two oligonucleotides. They are partially complementary, and each can include a flanking region of homology. The donor can be used to insert or replace polynucleotide sequences in a target sequence, for example, to introduce a polynucleotide that encodes a protein or functional RNA (e.g., siRNA), to introduce a protein tag, to modify a regulatory sequence of

a gene, or to introduce a regulatory sequence to a gene (e.g., a promoter, an enhancer, an internal ribosome entry sequence, a start codon, a stop codon; a localization signal, or polyadenylation signal), to modify a nucleic acid sequence (e.g., introduce a mutation), and the like.

[0069] Targeted nucleic acid modifications using donor polynucleotides for large changes (e.g., more than 100 base pair (bp) insertions or deletions) traditionally use plasmid-based donor templates that contain homology arms flanking the site of alteration. Each arm can vary in length, but is typically longer than about 100 bp, such as 100-1500 bp, e.g., 100 . . . 200 . . . 300 . . . 400 . . . 500 . . . 600 . . . 700 . . . 800 . . . 900 . . . 1000 . . . 1500 bp or any integer between these values. However, these numbers can vary, depending on the size of the donor polynucleotide and the target polynucleotide. This method can be used to generate large modifications, including insertion of reporter genes such as fluorescent proteins or antibiotic resistance markers. For transfection in cells, such as HEK cells, approximately 100-1000 nanograms (ng), e.g., 100 . . . 200 . . . 300 . . . 400 . . . 500 . . . 600 . . . 700 . . . 800 . . . 900 . . . 1000 ng or any integer between these values, of a typical size donor plasmid (e.g., approximately 5 kb) containing a NATNA/Cas vector, can be used for one well in 24-well plate. (See, e.g., Yang et al., "One Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering" *Cell* (2013) 154:1370-1379).

[0070] Single-stranded and partially double-stranded oligonucleotides, such as DNA oligonucleotides, have been used in place of targeting plasmids for short modifications (e.g., less than 50 bp) within a defined locus without cloning. To achieve high HDR efficiencies, single-stranded oligonucleotides containing flanking sequences on each side that are homologous to the target region can be used, and can be oriented in either the sense or antisense direction relative to the target locus. The length of each arm can vary, but the length of at least one arm is typically longer than about 10 bases, such as from 10-150 bases, e.g., 10 . . . 20 . . . 30 . . . 40 . . . 50 . . . 60 . . . 70 . . . 80 . . . 90 . . . 100 . . . 110 . . . 120 . . . 130 . . . 140 . . . 150, or any integer within these ranges. However, these numbers can vary, depending on the size of the donor polynucleotide and the target polynucleotide. In some embodiments, the length of at least one arm is 10 bases or more. In other embodiments, the length of at least one arm is 20 bases or more. In yet other embodiments, the length of at least one arm is 30 bases or more. In some embodiments, the length of at least one arm is less than 100 bases. In further embodiments, the length of at least one arm is greater than 100 bases. In some embodiments, the length of at least one arm is zero bases. For single-stranded oligonucleotide design, typically an oligonucleotide with around 100-150 bp total homology is used. The mutation is introduced in the middle, giving 50-75 bp homology arms for a donor designed to be symmetrical about the target site. In other cases, no homology arms are required, and the donor polynucleotide is inserted using non-homologous repair mechanisms.

[0071] A "genomic region" is a segment of a chromosome in the genome of a host cell that is present on either side of the nucleic acid target sequence site or, alternatively, also includes a portion of the nucleic acid target sequence site. The homology arms of the donor polynucleotide have sufficient homology to undergo homologous recombination with the corresponding genomic regions. In some embodi-

ments, the homology arms of the donor polynucleotide share significant sequence homology to the genomic region immediately flanking the nucleic acid target sequence site; it is recognized that the homology arms can be designed to have sufficient homology to genomic regions farther from the nucleic acid target sequence site.

[0072] As used herein the terms "classical non-homologous end joining" or "c-NHEJ" refer to the repair of double-strand breaks in DNA by direct ligation of one end of the break to the other end of the break without a requirement for a donor template DNA. NHEJ in the absence of a donor template DNA often results in small insertions or deletions of nucleotides at the site of the double-strand break, also referred to as "indels." This DNA repair pathway is genetically defined and requires the activity of Ligase IV, DNA-PKcs, Polμ, Polλ, and the Ku70/80 heterodimer, among other proteins (see, e.g., Sfeir and Symington, *Trends Biochem Sci* (2015) 40:701-714).

[0073] "Microhomology-mediated end joining (MMEJ)," a form of alternative nonhomologous end-joining (alt-NHEJ), is another pathway for repairing double-strand breaks in DNA. MMEJ is associated with deletions flanking a DSB and involves alignment of microhomologous sequences internal to the broken ends before joining. The proposed mechanism entails 5'-3' resection of the DNA ends at a DSB, annealing of the microhomologies (1-16 nucleotides of homology), removal of heterologous flaps, gap filling DNA synthesis, and ligation. MMEJ is genetically defined and requires the activity of CtIP, PARP1, Polθ, Lig1 and Lig3, among other proteins (see, e.g., Sfeir and Symington, "Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?" *Trends Biochem Sci* (2015) 40:701-714).

[0074] Alternative mechanisms of nucleic acid insertion that do not require sequence homology between the donor and the target sequence can also be used for nucleic acid insertion. These mechanisms involve various components of the cellular repair machinery and it is to be understood that the scope of the invention is not bound by the use of any particular mechanism for insertion of nucleic acid after target nucleic acid is cut or nicked by a site-specific polynucleotide.

[0075] "Gene," as used herein, refers to a polynucleotide sequence comprising exon(s) and related regulatory sequences. A gene may further comprise intron(s) and/or untranslated region(s) (UTR(s)).

[0076] As used herein, "expression" refers to transcription of a polynucleotide from a DNA template, resulting in, for example, a messenger RNA (mRNA) or other RNA transcript (e.g., non-coding, such as structural or scaffolding RNAs). The term further refers to the process through which transcribed mRNA is translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be referred to collectively as "gene product(s)." Expression may include splicing the mRNA in a eukaryotic cell, if the polynucleotide is derived from genomic DNA.

[0077] As used herein, the term "modulate" refers to a change in the quantity, degree or amount of a function. For example, a cognate nucleic acid guide/CasM protein complex, as disclosed herein, may modulate the activity of a promoter sequence by binding to a nucleic acid target sequence at or near the promoter. Depending on the action occurring after binding, the cognate nucleic acid guide/CasM protein complex can induce, enhance, suppress, or

inhibit transcription of a gene operatively linked to the promoter sequence. Thus, “modulation” of gene expression includes both gene activation and gene repression.

[0078] Modulation can be assayed by determining any characteristic directly or indirectly affected by the expression of the target gene. Such characteristics include, e.g., changes in RNA or protein levels, protein activity, product levels, expression of the gene, or activity level of reporter genes. Accordingly, the terms “modulating expression,” “inhibiting expression,” and “activating expression” of a gene can refer to the ability of a cognate guide/CasM protein complex to change, activate, or inhibit transcription of a gene.

[0079] The terms “vector” and “plasmid” are used interchangeably and as used herein refer to a polynucleotide vehicle to introduce genetic material into a cell. Vectors can be linear or circular. Vectors can integrate into a target genome of a host cell or replicate independently in a host cell. Vectors can comprise, for example, an origin of replication, a multicloning site, and/or a selectable marker. An expression vector typically comprises an expression cassette. Vectors and plasmids include, but are not limited to, integrating vectors, prokaryotic plasmids, eukaryotic plasmids, plant synthetic chromosomes, episomes, viral vectors, cosmids, and artificial chromosomes. An expression vector typically comprises an expression cassette.

[0080] As used herein the term “expression cassette” is a polynucleotide construct, generated recombinantly or synthetically, comprising regulatory sequences operably linked to a selected polynucleotide to facilitate expression of the selected polynucleotide in a host cell. For example, the regulatory sequences can facilitate transcription of the selected polynucleotide in a host cell, or transcription and translation of the selected polynucleotide in a host cell. An expression cassette can, for example, be integrated in the genome of a host cell or be present in a vector to form an expression vector.

[0081] As used herein, a “targeting vector” is a recombinant DNA or RNA construct typically comprising tailored DNA or RNA arms, homologous to genomic DNA or RNA derived therefrom, that flank elements of a target gene or nucleic acid target sequence (e.g., a SSB or DSB). A targeting vector comprises a donor polynucleotide. Elements of the target sequence can be modified in a number of ways including deletions and/or insertions. A defective target gene can be replaced by a functional target gene, or in the alternative a functional gene can be knocked out. Optionally, the donor polynucleotide of a targeting vector comprises a selection cassette comprising a selectable marker that is introduced into the target gene. Targeting regions (i.e., nucleic acid target sequences) adjacent or within a target gene or region can be used to affect regulation of gene expression.

[0082] As used herein, the terms “regulatory sequences,” “regulatory elements,” and “control elements” are interchangeable and refer to polynucleotide sequences that are upstream (5' non-coding sequences), within, or downstream (3' non-translated sequences) of a polynucleotide target to be expressed. Regulatory sequences influence, for example, the timing of transcription, amount or level of transcription, RNA processing or stability, and/or translation of the related structural nucleotide sequence. Regulatory sequences may include activator binding sequences, enhancers, introns, polyadenylation recognition sequences, promoters, tran-

scription start sites, repressor binding sequences, stem-loop structures, translational initiation sequences, internal ribosome entry sites (IRES), translation leader sequences, transcription termination sequences (e.g., polyadenylation signals and poly-U sequences), translation termination sequences, primer binding sites, and the like.

[0083] Regulatory elements include those that direct constitutive, inducible, and repressible expression of a nucleotide sequence in many types of host cells and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). In some embodiments, a vector comprises one or more pol III promoters, one or more pol II promoters, one or more pol I promoters, or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer; see, e.g., Boshart, M., et al., *Cell* 41:521-530 (1985)), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. It will be appreciated by those skilled in the art that the design of an expression vector may depend on such factors as the choice of the host cell to be transformed, the level of expression desired, and the like. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

[0084] As used herein the term “operably linked” refers to polynucleotide sequences or amino acid sequences placed into a functional relationship with one another. For instance, a promoter or enhancer is operably linked to a coding sequence if it regulates, or contributes to the modulation of, the transcription of the coding sequence. Operably linked DNA sequences encoding regulatory sequences are typically contiguous to the coding sequence. However, enhancers can function when separated from a promoter by up to several kilobases or more. Accordingly, some polynucleotide elements may be operably linked but not contiguous.

[0085] As used herein, the term “expression” refers to transcription of a polynucleotide from a DNA template, resulting in, for example, an mRNA or other RNA transcript (e.g., non-coding, such as structural or scaffolding RNAs). The term further refers to the process through which transcribed mRNA is translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be referred to collectively as “gene product.” Expression may include splicing the mRNA in a eukaryotic cell, if the polynucleotide is derived from genomic DNA.

[0086] As used herein, the term “sequence identity” generally refers to the percent identity of bases or amino acids determined by comparing a first polynucleotide or polypeptide to a second polynucleotide or polypeptide using algorithms having various weighting parameters. Sequence identity between two polypeptides or two polynucleotides can be determined using sequence alignment by various methods and computer programs (e.g., BLAST, CS-BLAST, FASTA, HMMER, L-ALIGN, etc.), available through the worldwide web at sites including GENBANK (ncbi.nlm.nih.gov/genbank/) and EMBL-EBI (ebi.ac.uk/). Sequence identity between two polynucleotides or two polypeptide sequences is generally calculated using the standard default parameters

of the various methods or computer programs. Generally, Cas proteins, such as CasM homologs, for use herein will have at least about 75% or more sequence identity to the wild-type or naturally occurring sequence of the Cas protein of interest, such as about 80%, such as about 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or complete identity. CasM orthologs can vary widely from the reference sequence. For example, CasM orthologs shown in Table 1 display approximately 13.59% to 99.82% sequence identity to each other. Thus, CasM orthologs are identified based on the CasM characteristics detailed herein.

[0087] As used herein, “hybridization,” “hybridize,” or “hybridizing” is the process of combining two complementary single-stranded DNA or RNA molecules so as to form a single double-stranded molecule (DNA/DNA, DNA/RNA, RNA/RNA) through hydrogen base pairing. Hybridization stringency is typically determined by the hybridization temperature and the salt concentration of the hybridization buffer; e.g., high temperature and low salt provide high stringency hybridization conditions. Examples of salt concentration ranges and temperature ranges for different hybridization conditions are as follows: high stringency, approximately 0.01M to approximately 0.05M salt, hybridization temperature 5° C. to 10° C. below T_m; moderate stringency, approximately 0.16M to approximately 0.33M salt, hybridization temperature 20° C. to 29° C. below T_m; and low stringency, approximately 0.33M to approximately 0.82M salt, hybridization temperature 40° C. to 48° C. below T_m. T_m of duplex nucleic acids is calculated by standard methods well-known in the art (see, e.g., Maniatis, T., et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press: New York (1982); Casey, J., et al., *Nucleic Acids Research* 4:1539-1552 (1977); Bodkin, D. K., et al., *Journal of Virological Methods* 10(1): 45-52 (1985); Wallace, R. B., et al., *Nucleic Acids Research* 9(4):879-894 (1981)). Algorithm prediction tools to estimate T_m are also widely available. High stringency conditions for hybridization typically refer to conditions under which a polynucleotide complementary to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Typically, hybridization conditions are of moderate stringency, preferably high stringency.

[0088] As used herein, the term “amino acid” refers to natural and synthetic (unnatural) amino acids, including amino acid analogs, modified amino acids, peptidomimetics, glycine, and D or L optical isomers.

[0089] As used herein, the terms “peptide,” “polypeptide,” and “protein” are interchangeable and refer to polymers of amino acids. A polypeptide may be of any length. It may be branched or linear, it may be interrupted by non-amino acids, and it may comprise modified amino acids. The terms may be used to refer to an amino acid polymer that has been modified through, for example, acetylation, disulfide bond formation, glycosylation, lipidation, phosphorylation, cross-linking, and/or conjugation (e.g., with a labeling component or ligand). Polypeptide sequences are displayed herein in the conventional N-terminal to C-terminal orientation.

[0090] Polypeptides and polynucleotides can be made using routine techniques in the field of molecular biology (see, e.g., standard texts set forth above). Further, essentially any polypeptide or polynucleotide can be custom ordered from commercial sources.

[0091] The terms “fusion protein” and “chimeric protein,” as used herein, refer to a single protein created by joining two or more proteins, protein domains, or protein fragments that do not naturally occur together in a single protein. For example, a fusion protein can contain a first domain from a CasM protein and a second domain from a different Cas protein. The modification to include such domains in fusion proteins may confer additional activity on the modified site-directed polypeptides. Such activities can include nucleic acid activity, methyltransferase activity, demethylase activity, DNA or RNA repair activity, DNA or RNA damage activity, deamination activity, diisomerase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity, glycosylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristylation activity, or demyristylation activity) that modifies a polypeptide associated with nucleic acid target sequence (e.g., a histone). A fusion protein can also comprise epitope tags (e.g., histidine tags, FLAG® (Sigma Aldrich, St. Louis, Mo.) tags, Myc tags), reporter protein sequences (e.g., glutathione-S-transferase, beta-galactosidase, luciferase, green fluorescent protein, cyan fluorescent protein, yellow fluorescent protein), and/or nucleic acid binding domains (e.g., a DNA binding domain, an RNA binding domain). A fusion protein can also comprise activator domains (e.g., heat shock transcription factors, NFkB activators) or repressor domains (e.g., a KRAB domain). As described by Lupo, A., et al., *Current Genomics* 14(4): 268-278 (2013), the KRAB domain is a potent transcriptional repression module and is located in the amino-terminal sequence of most C2H2 zinc finger proteins (see, e.g., Margolin, J., et al., *Proceedings of the National Academy of Sciences of the United States of America* 91:4509-4513 (1994); Witzgall, R., et al., *Proceedings of the National Academy of Sciences of the United States of America* 91:4514-4518 (1994)). The KRAB domain typically binds to co-repressor proteins and/or transcription factors via protein-protein interactions, causing transcriptional repression of genes to which KRAB zinc finger proteins (KRAB-ZFPs) bind (see, e.g., Friedman J. R., et al., *Genes & Development* 10:2067-2078 (1996)). In some embodiments, linker nucleic acid sequences are used to join the two or more proteins, protein domains, or protein fragments.

[0092] A “moiety,” as used herein, refers to a portion of a molecule. A moiety can be a functional group or describe a portion of a molecule with multiple functional groups (e.g., that share common structural aspects). The terms “moiety” and “functional group” are typically used interchangeably; however, a “functional group” can more specifically refer to a portion of a molecule that comprises some common chemical behavior. “Moiety” is often used as a structural description. In some embodiments, a 5' terminus, a 3' terminus, or a 5' terminus and a 3' terminus (e.g., a non-native 5' terminus and/or a non-native 3' terminus in a first stem element) can comprise one or more moieties.

[0093] As used herein, the term “isolated” can refer to a nucleic acid or polypeptide that, by the hand of a human, exists apart from its native environment and is therefore not

a product of nature. Isolated means substantially pure. An isolated nucleic acid or polypeptide can exist in a purified form and/or can exist in a non-native environment such as, for example, in a recombinant cell.

[0094] As used herein, a “host cell” generally refers to a biological cell. A cell is the basic structural, functional and/or biological unit of an organism. A cell can originate from any organism having one or more cells. Examples of host cells include, but are not limited to: a prokaryotic cell, eukaryotic cell, a bacterial cell, an archaean cell, a cell of a single-cell eukaryotic organism, a protozoal cell, a cell from a plant (e.g., cells from plant crops (such as soy, tomatoes, sugar beets, pumpkin, hay, *cannabis*, tobacco, plantains, yams, sweet potatoes, cassava, potatoes, wheat, sorghum, soybean, rice, corn, maize, oil-producing *Brassica* (e.g., oil-producing rapeseed and canola), cotton, sugar cane, sunflower, millet, and alfalfa), fruits, vegetables, grains, seeds, flowering plants, conifers, gymnosperms, ferns, club-mosses, hornworts, liverworts, mosses), an algal cell, (e.g., *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens* C. agardh, and the like), seaweeds (e.g., kelp), a fungal cell (e.g., a yeast cell or a cell from a mushroom), an animal cell, a cell from an invertebrate animal (e.g., fruit fly, cnidarian, echinoderm, nematode, and the like), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, or mammal), a cell from a mammal (e.g., a pig, a cow, a goat, a sheep, a rodent, a rat, a mouse, a non-human primate, a human, and the like). Furthermore, a cell can be a stem cell or a progenitor cell.

[0095] As used herein, “stem cell” refers to a cell that has the capacity for self-renewal, i.e., the ability to go through numerous cycles of cell division while maintaining the undifferentiated state. Stem cells can be totipotent, pluripotent, multipotent, oligopotent, or unipotent. Stem cells can be embryonic, fetal, amniotic, adult, or induced pluripotent stem cells.

[0096] As used herein, “induced pluripotent stem cells” refers to a type of pluripotent stem cell that is artificially derived from a non-pluripotent cell, typically an adult somatic cell, by inducing expression of specific genes.

[0097] “Plant,” as used herein, refers to whole plants, plant organs, plant tissues, germplasm, seeds, plant cells, and progeny of the same. Plant cells include, without limitation, cells from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Plant parts include differentiated and undifferentiated tissues including, but not limited to roots, stems, shoots, leaves, pollens, seeds, tumor tissue, and various forms of cells and culture (e.g., single cells, protoplasts, embryos, and callus tissue). The plant tissue may be in plant or in a plant organ, tissue or cell culture. “Plant organ” refers to plant tissue or a group of tissues that constitute a morphologically and functionally distinct part of a plant.

[0098] “Subject,” as used herein, refers to any member of the phylum Chordata, including, without limitation, humans and other primates, including non-human primates such as rhesus macaques, chimpanzees and other monkey and ape species; farm animals, such as cattle, sheep, pigs, goats, and horses; domestic mammals, such as dogs and cats; laboratory animals, including rabbits, mice, rats, and guinea pigs; birds, including domestic, wild, and game birds, such as chickens, turkeys, and other gallinaceous birds, ducks, and

geese; and the like. The term does not denote a particular age or gender. Thus, the term includes adult, young, and newborn individuals as well as male and female. In some embodiments, a host cell is derived from a subject (e.g., stem cells, progenitor cells, or tissue-specific cells). In some embodiments, the subject is a non-human subject.

[0099] As used herein, “transgenic organism” refers to an organism whose genome is genetically modified. The term includes the progeny (any generation) of a transgenic organism, provided that the progeny has the genetic modification.

CRISPR Systems

[0100] The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) genomic locus is found in the genomes of many prokaryotes (e.g., bacteria and archaea). CRISPR loci provide resistance to foreign invaders (e.g., virus, phage) in prokaryotes. In this way, the CRISPR system functions as a type of immune system to help defend prokaryotes against foreign invaders. There are three main stages in CRISPR-Cas immune systems: (1) acquisition, (2) expression, and (3) interference. Acquisition involves cleaving the genome of invading viruses and plasmids and integrating segments (termed protospacers) of the genomic DNA into the CRISPR locus of the host organism. The segments that are integrated into the host genome are known as spacers, which mediate protection from subsequent attack by the same (or sufficiently related) virus or plasmid. Expression involves transcription of the CRISPR locus and subsequent enzymatic processing to produce short mature CRISPR RNAs, each containing a single spacer sequence. Interference is induced after the CRISPR RNAs associate with Cas proteins to form effector complexes, which are then targeted to complementary protospacers in foreign genetic elements to induce nucleic acid degradation.

[0101] Currently, two classes of CRISPR systems have been described, Class 1 and Class 2, based upon the genes encoding the effector module, i.e., the proteins involved in the interference stage. Class 1 systems have a multi-subunit crRNA-effector complex, whereas Class 2 systems have a single protein, such as Cas9, Cpf1, C2c1, C2c2, C2c3, or a crRNA-effector complex. Class 1 systems comprise Type I, Type III and Type IV systems. Class 2 systems comprise Type II, Type V and Type VI systems.

[0102] To date, there are six types (Types I-VI) and 19 subtypes of CRISPR systems categorized within these classes (Makarova et al., *Nature Reviews Microbiology* (2015) 13:1-15; Shmakov et al., *Nature Reviews Microbiology* (2017) 15:169-182).

[0103] CRISPR loci are currently characterized as including a number of short repeating sequences referred to as “repeats.” Repeats can form hairpin structures and/or repeats can be unstructured single-stranded sequences. The repeats occur in clusters. Repeats frequently diverge between species. Repeats are regularly interspersed with unique intervening sequences, referred to as “spacers,” resulting in a repeat-spacer-repeat locus architecture. Spacers are identical to or are homologous with known foreign invader sequences. In some instances, a spacer-repeat unit encodes a crisprRNA (crRNA). A crRNA refers to the mature form of the spacer-repeat unit. A crRNA contains a spacer sequence that is involved in targeting a target nucleic acid (e.g., possibly as a surveillance mechanism against foreign nucleic acid). Thus, crRNA has a region of complementarity to a potential DNA or RNA target sequence and in some

cases, e.g., in currently characterized Type II systems, a second region that forms base-pair hydrogen bonds with a transactivating CRISPR RNA (tracrRNA) to form a secondary structure, typically to form at least a stem structure. In this context, the tracrRNA and a crRNA interact through a number of base-pair hydrogen bonds to form secondary RNA structures. Complex formation between tracrRNA/crRNA and a Cas protein results in conformational change of the Cas protein that facilitates binding to DNA, nuclease activities of the Cas protein, and crRNA-guided site-specific DNA cleavage by the nuclease. For a Cas protein/tracrRNA/crRNA complex to cleave a DNA target sequence, the DNA target sequence is adjacent to a cognate protospacer adjacent motif (PAM).

[0104] A CRISPR locus comprises polynucleotide sequences encoding for CRISPR Associated Genes (cas) genes. Cas genes are involved in the biogenesis and/or the interference stages of crRNA function. Cas genes display extreme sequence (e.g., primary sequence) divergence between species and homologs. Some Cas genes comprise homologous secondary and/or tertiary structures. Cas genes are typically named according to the organism from which they are derived. For example, Cas genes in *Staphylococcus epidermidis* can be referred to as Csm-type, Cas genes in *Streptococcus thermophilus* can be referred to as Csn-type, and Cas genes in *Pyrococcus furiosus* can be referred to as Cmr-type.

[0105] The integration stage of a CRISPR system refers to the ability of the CRISPR locus to integrate new spacers into the crRNA array upon being infected by a foreign invader. Acquisition of the foreign invader spacers can help confer immunity to subsequent attacks by the same foreign invader. Integration typically occurs at the leader end of the CRISPR locus. Cas proteins are involved in integration of new spacer sequences. Integration proceeds similarly for some types of CRISPR systems (e.g., Types I-III).

[0106] Mature crRNAs are processed from a longer polycistronic CRISPR locus transcript (i.e., pre-crRNA array). A pre-crRNA array comprises a plurality of crRNAs. The repeats in the pre-crRNA array are recognized by cas genes. Cas genes bind to the repeats and cleave the repeats. This action can liberate the plurality of crRNAs. crRNAs can be subjected to further events to produce the mature crRNA form such as trimming (e.g., with an exonuclease). A crRNA may comprise all, some, or none of the CRISPR repeat sequence.

[0107] Interference refers to the stage in the CRISPR system that is functionally responsible for combating infection by a foreign invader. CRISPR interference follows a similar mechanism to RNA interference (RNAi: e.g., wherein a target RNA is targeted (e.g., hybridized) by a short interfering RNA (siRNA)), which results in target RNA degradation and/or destabilization. Currently characterized CRISPR systems perform interference of a target nucleic acid by coupling crRNAs and Cas genes, thereby forming CRISPR ribonucleoproteins (RNPs). crRNA of the RNP guides the RNP to foreign invader nucleic acid, (e.g., by recognizing the foreign invader nucleic acid through hybridization). Hybridized target foreign invader nucleic acid-crRNA units are subjected to cleavage by Cas proteins. Target nucleic acid interference typically requires a protospacer adjacent motif (PAM) in a target nucleic acid.

[0108] By a "CRISPR-Cas system" as used herein, is meant any of the various CRISPR-Cas classes, types, and

subtypes. Class 1 systems comprise Type I, Type III, and Type IV systems. Type I systems are currently characterized as having a Cas3 protein that has helicase activity and cleavage activity. Type I systems are further divided into several subtypes that have a defined combination of signature genes and distinct features of operon organization.

[0109] To date, it appears that all Type III systems possess a cas10 gene, which encodes a multidomain protein containing a Palm domain (a variant of the RNA recognition motif (RRM)) that is homologous to the core domain of numerous nucleic acid polymerases and cyclases and that is the largest subunit of Type III crRNA-effector complexes. All Type III loci also encode the small subunit protein, one Cas5 protein and typically several Cas7 proteins. Type III is also further divided into several subtypes.

[0110] Type IV systems encode a minimal multisubunit crRNA-effector complex comprising a partially degraded large subunit, Csfl, Cas5, Cas7, and in some cases, a putative small subunit. Type IV systems lack cas1 and cas2 genes. Type IV systems do not have subtypes, but there are two distinct variants. One Type IV variant has a DinG family helicase, whereas a second Type IV variant lacks a DinG family helicase, but has a gene encoding a small α -helical protein. An example of an organism with a Type IV system is *Acidithiobacillus ferrooxidans*.

[0111] Class 2 systems comprise Type II, Type V, and Type VI systems. Type II systems include cas1, cas2 and cas9 genes. There are two strands of RNA in Type II systems, a crRNA and a tracrRNA, that hybridizes to a complementary region of pre-crRNA causing maturation of the pre-crRNA to crRNA. The duplex formed by the tracrRNA and crRNA is recognized by, and associates with a multidomain protein, Cas9, encoded by the cas9 gene, which combines the functions of the crRNA-effector complex with target DNA cleavage. Cas9 is directed to a target nucleic acid by a sequence of the crRNA that is complementary to, and hybridizes with, a sequence in the target nucleic acid.

[0112] In Type V systems, nucleic acid target sequence binding involves a Cas12a protein and the crRNA, as does the nucleic acid target sequence cleavage. In Type V systems, the RuvC-like nuclease domain of Cas12a protein cleaves both strands of the nucleic acid target sequence in a sequential fashion (Swarts, et al., *Mol. Cell* (2017) 66:221-233.e4), producing 5' overhangs, which contrasts with the blunt ends generated by Cas9 protein cleavage.

[0113] The Cas12a protein cleavage activity of Type V systems does not require hybridization of crRNA to tracrRNA to form a duplex; rather Type V systems use a single crRNA that has a stem-loop structure forming an internal duplex. Cas12a protein binds the crRNA in a sequence- and structure-specific manner by recognizing the stem loop and sequences adjacent to the stem loop, most notably the nucleotides 5' of the spacer sequence, which hybridizes to the nucleic acid target sequence. This stem-loop structure is typically in the range of 15 to 19 nucleotides in length. Substitutions that disrupt this stem-loop duplex abolish cleavage activity, whereas other substitutions that do not disrupt the stem-loop duplex do not abolish cleavage activity.

[0114] Type VI systems include the Cas13a protein (also known as Class 2 candidate 2 protein, or C2c2) which does not share sequence similarity with other CRISPR effector proteins (see Abudayyeh, et al., *Science* (2016) 353: aaf5573). Cas13a proteins have two HEPN domains and

possess single-stranded RNA cleavage activity. Cas13a proteins are similar to Cas12a proteins in requiring a crRNA for nucleic acid target sequence binding and cleavage, but not requiring tracrRNA. Also, similar to Cas12a protein, the crRNA for Cas13a proteins forms a stable hairpin, or stem-loop structure, that aids in association with the Cas13a protein. Type VI systems have a single polypeptide RNA endonuclease that utilizes a single crRNA to direct RNA cleavage in a target-dependent fashion. Additionally, after hybridizing to the target RNA complementary to the spacer, Cas13a protein becomes a promiscuous RNA endonuclease exhibiting non-specific endonuclease activity toward any single-stranded RNA in a sequence independent manner (see East-Seletsky, et al., *Nature* (2016) 538:270-273).

[0115] As is readily apparent, the discovery and characterization of CRISPR systems is currently evolving.

Production of CRISPR Components

[0116] In all of the embodiments described herein, the various components can be produced by synthesis, or for example, using expression cassettes encoding CasM, a cognate guide, etc. The various components can be provided to a cell or used in vitro. These components can be present on a single cassette or multiple cassettes, in the same or different constructs. Expression cassettes typically comprise regulatory sequences functional in host cells into which they are introduced. Regulatory sequences are involved in one or more of the following: regulation of transcription, post-transcriptional regulation, and regulation of translation. Expression cassettes can be present in expression vectors and introduced into a wide variety of host cells including bacterial cells, yeast cells, plant cells, and mammalian cells.

[0117] In one aspect, all or a portion of the various components for use herein are produced in vectors, including expression vectors, comprising polynucleotides encoding therefor. Vectors useful for producing components for use in the present methods include plasmids, viruses (including phage), and integratable nucleic acid fragments (i.e., fragments integratable into the host genome by homologous recombination). A vector replicates and functions independently of the host genome, or may, in some instances, integrate into the genome itself. Suitable replicating vectors will contain a replicon and control sequences derived from species compatible with the intended expression host cell. In some embodiments, polynucleotides encoding one or more of the various components are operably linked to an inducible promoter, a repressible promoter, or a constitutive promoter. Expression vectors can also include polynucleotides encoding protein tags (e.g., poly-His tags, hemagglutinin tags, fluorescent protein tags, bioluminescent tags, nuclear localization tags). The coding sequences for such protein tags can be fused to the coding sequences or can be included in an expression cassette, for example, in a targeting vector.

[0118] General methods for construction of expression vectors are known in the art. Expression vectors for most host cells are commercially available. There are several commercial software products designed to facilitate selection of appropriate vectors and construction thereof, such as insect cell vectors for insect cell transformation and gene expression in insect cells, bacterial plasmids for bacterial transformation and gene expression in bacterial cells, yeast plasmids for cell transformation and gene expression in yeast and other fungi, mammalian vectors for mammalian

cell transformation and gene expression in mammalian cells or mammals, viral vectors (including retroviral, lentiviral, and adenoviral vectors) for cell transformation and gene expression and methods to easily enable cloning of such polynucleotides. SnapGene™ (GSL Biotech LLC, Chicago, Ill.; snapgene.com/resources/plasmid_files/your_time_is_valuable/), for example, provides an extensive list of vectors, individual vector sequences, and vector maps, as well as commercial sources for many of the vectors.

[0119] Several expression vectors have been designed for expressing guide polynucleotides. See, e.g., Shen et al. *Nat. Methods* (2014) 11:399-402. Additionally, vectors and expression systems are commercially available, such as from New England Biolabs (Ipswich, Mass.) and Clontech Laboratories (Mountain View, Calif.). Vectors can be designed to simultaneously express a target-specific NATNA using a U2 or U6 promoter, a CasM and/or dCasM, and if desired, a marker protein, for monitoring transfection efficiency and/or for further enriching/isolating transfected cells by flow cytometry.

[0120] For example, the various components can be incorporated into mammalian vectors for use in mammalian cells. A large number of mammalian vectors suitable for use with the systems of the present invention are commercially available (e.g., from Life Technologies, Grand Island, N.Y.; NeoBiolab, Cambridge, Mass.; Promega, Madison, Wis.; DNA2.0, Menlo Park, Calif.; Addgene, Cambridge, Mass.).

[0121] Vectors derived from mammalian viruses can also be used for expressing the various components of the present methods in mammalian cells. These include vectors derived from viruses such as adenovirus, papavirus, herpesvirus, polyomavirus, cytomegalovirus, lentivirus, retrovirus, vaccinia and Simian Virus 40 (SV40) (see, e.g., Kaufman, R. J., *Molec. Biotech.* (2000) 16:151-160; Cooray et al., *Methods Enzymol.* (2012) 507:29-57). Regulatory sequences operably linked to the components can include activator binding sequences, enhancers, introns, polyadenylation recognition sequences, promoters, repressor binding sequences, stem-loop structures, translational initiation sequences, translation leader sequences, transcription termination sequences, translation termination sequences, primer binding sites, and the like. Commonly used promoters are constitutive mammalian promoters CMV, EF1a, SV40, PGK1 (mouse or human), Ubc, CAG, CaMKIIa, and beta-Act, and others known in the art (Khan, K. H. *Advanced Pharmaceutical Bulletin* (2013) 3:257-263). Furthermore, mammalian RNA polymerase III promoters, including H1 and U6, can be used.

[0122] Numerous mammalian cell lines have been utilized for expression of gene products including HEK 293 (Human embryonic kidney) and CHO (Chinese hamster ovary). These cell lines can be transfected by standard methods (e.g., using calcium phosphate or polyethyleneimine (PEI), or electroporation). Other typical mammalian cell lines include, but are not limited to: HeLa, U2OS, 549, HT1080, CAD, P19, NIH 3T3, L929, N2a, Human embryonic kidney 293 cells, MCF-7, Y79, SO-Rb50, Hep G2, DUKX-X11, J558L, and Baby hamster kidney (BHK) cells.

[0123] Vectors can be introduced into and propagated in a prokaryote. Prokaryotic vectors are well known in the art. Typically a prokaryotic vector comprises an origin of replication suitable for the target host cell (e.g., oriC derived from *E. coli*, pUC derived from pBR322, pSC101 derived from *Salmonella*), 15A origin (derived from p15A) and bacterial artificial chromosomes). Vectors can include a

selectable marker (e.g., genes encoding resistance for ampicillin, chloramphenicol, gentamicin, and kanamycin). Zeocin™ (Life Technologies, Grand Island, N.Y.) can be used as a selection in bacteria, fungi (including yeast), plants and mammalian cell lines. Accordingly, vectors can be designed that carry only one drug resistance gene for Zeocin for selection work in a number of organisms. Useful promoters are known for expression of proteins in prokaryotes, for example, T5, T7, Rhamnose (inducible), Arabinose (inducible), and PhoA (inducible). Furthermore, T7 promoters are widely used in vectors that also encode the T7 RNA polymerase. Prokaryotic vectors can also include ribosome binding sites of varying strength, and secretion signals (e.g., mal, sec, tat, ompC, and pelB). In addition, vectors can comprise RNA polymerase promoters for the expression of NATNAs. Prokaryotic RNA polymerase transcription termination sequences are also well known (e.g., transcription termination sequences from *Streptococcus pyogenes*).

[0124] Expression of proteins in prokaryotes is typically carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins.

[0125] In some embodiments, a vector is a yeast expression vector comprising one or more components of the above-described methods. Examples of vectors for expression in *Saccharomyces cerevisiae* include, but are not limited to, the following: pYEpSecl, pMFa, pJRY88, pYES2, and picZ. Methods for gene expression in yeast cells are known in the art (see, e.g., Methods in Enzymology, Volume 194, "Guide to Yeast Genetics and Molecular and Cell Biology, Part A," (2004) Christine Guthrie and Gerald R. Fink (eds.), Elsevier Academic Press, San Diego, Calif.). Typically, expression of protein-encoding genes in yeast requires a promoter operably linked to a coding region of interest plus a transcriptional terminator. Various yeast promoters can be used to construct expression cassettes for expression of genes in yeast.

CasM Proteins

[0126] CasM, a new CRISPR-Cas protein, is described herein. CasM displays nucleic acid binding activity and produces breaks, such as single-strand breaks (SSBs) or DSBs, when brought into proximity with a nucleic acid target sequence, e.g., by association with a cognate nucleic acid guide, such as a cognate crRNA. As shown in the Examples herein, CasM targets RNA and is capable of cleaving ssRNA, such as when delivered to a genomic target when complexed with a crRNA, without the necessity of association with a tracrRNA. CasM proteins typically include two or more higher eukaryotic and prokaryotic nucleotide-binding (HEPN) domains found in protein family PF05168, in the C-terminal region of the CasM sequence. CasM proteins show synteny with one or more WYL domain-containing proteins and sometimes with RtcB (RNA 3'-terminal phosphate cyclase, group B) domain-containing proteins. Based on the foregoing characteristics, CasM may be classified as a Class 2 Type VI CRISPR-Cas system. However, CasM has a very low degree of sequence similarity to other Type VI subtypes.

[0127] Exemplary CasM proteins are shown in SEQ ID NOS:37-45, and are encoded by polynucleotides shown in SEQ ID NOS:28-36, respectively. These proteins display approximately 13.59% to 99.82% sequence identity to each other and show less than 8% sequence identity with other

known CRISPR-Cas proteins. CasM has been found in several species and isolates including, without limitation, *Eubacterium siraeum*; *Ruminococcus* sp., such as from *Ruminococcus bicirculans*; *Ruminococcus flavefaciens*, such as, but not limited to, FD-1 and strain XPD3002; *Ruminococcus albus* such as, but not limited to, strain KH2T6; *Ruminococcus* sp. isolates, such as but not limited to, isolates 2789STDY5834971, 2789STDY5608892 and 2789STDY5834894. However, it is to be understood that the term "CasM" refers to a protein derived from any species, subspecies or strain of bacteria that encodes a CasM protein, as well as orthologs thereof, or variants thereof. Representative CasM proteins include, but are not limited to, those proteins depicted as SEQ ID NOS:37-45 (see Table 1), orthologs thereof, or variants thereof. CasM proteins are approximately 800 to approximately 1000 amino acids in length.

TABLE 1

Representative CasM Proteins	
Species/Isolate	SEQ ID NO
<i>Eubacterium siraeum</i>	SEQ ID NO: 37
<i>Ruminococcus</i> sp., isolate 2789STDY5834971	SEQ ID NO: 38
<i>Ruminococcus bicirculans</i>	SEQ ID NO: 39
<i>Ruminococcus</i> sp., isolate 2789STDY5608892	SEQ ID NO: 40
<i>Ruminococcus</i> sp. CAG:57	SEQ ID NO: 41
<i>Ruminococcus flavefaciens</i> FD-1	SEQ ID NO: 42
<i>Ruminococcus albus</i> strain KH2T6	SEQ ID NO: 43
<i>Ruminococcus flavefaciens</i> strain XPD3002	SEQ ID NO: 44
<i>Ruminococcus</i> sp., isolate 2789STDY5834894	SEQ ID NO: 45

[0128] Analysis of these CasM protein sequences indicates the presence of two HEPN domains in the C-terminal region of the sequences. The HEPN domain is often involved in nucleic acid binding and can function as a metal-independent RNase in certain instances.

[0129] CasM systems display strong synteny with an open reading frame in WYL domain-(protein family PF13280) containing proteins. The sequences for WYL domains in various species that encode CasM proteins are shown as SEQ ID NOS:52-59 (see Table 2). WYL domains share similarities with CRISPR-associated Rossman fold (CARF) domains and are thought to bind ligands derived from host-virus conflict and regulate CRISPR-Cas systems. A WYL domain protein (SI17009) has been shown to be a negative regulator of the I-D CRISPR-Cas system in *Synechocystis* sp. (Hein et al., RNA Biol. (2013) 10: 852-864. In some instances, the WYL-containing protein contains at least two WYL domains. These duplications are consistent with the hypothesized multimeric assembly of these ligand-binding domains (Schumacher et al., EMBO J. (2002) 21:1210-1218). The N-termini of these WYL domains contain homology to transcriptional repressor CopG and the ParD anti-toxin domain. For use in eukaryotes, the WYL domain-containing proteins can be modified with a N- or C-terminal nuclear localization signal sequence (NLS). SEQ ID NOS:61-68 present exemplary WYL domain-containing proteins modified with a seven amino acid C-terminal NLS tag derived from the SV40 Large T-antigen.

TABLE 2

WYL domain sequences in various CasM-containing species	
Species/Isolate	SEQ ID NO
<i>Eubacterium siraeum</i>	SEQ ID NO: 52
<i>Ruminococcus</i> sp., isolate 2789STDY5834971	SEQ ID NO: 53
<i>Ruminococcus bicirculans</i>	SEQ ID NO: 54
<i>Ruminococcus</i> sp., isolate 2789STDY5608892	SEQ ID NO: 55
<i>Ruminococcus</i> sp. CAG:57	SEQ ID NO: 56
<i>Ruminococcus flavefaciens</i> FD-1	SEQ ID NO: 57
<i>Ruminococcus albus</i> strain KH2T6	SEQ ID NO: 58
<i>Ruminococcus flavefaciens</i> strain XPD3002	SEQ ID NO: 59

[0130] RtcB (RNA 3'-terminal phosphate cyclase, group B) is a protein domain superfamily and a RtcB homolog (SEQ ID NO. 60) proximal to the CasM loci has been identified. It has previously been reported that the CARF domain has sequence similarity with the N-terminal domain of the RtcR protein, which acts as the regulator of the Rtc RNA repair system. The Rtc system is comprised of the 3'-terminal phosphate cyclase RtcA and the RNA ligase RtcB. The RtcB domain-containing proteins can be modified with a N- or C-terminal NLS for use in eukaryotes. A RtcB domain with an associated NLS derived from the SV40 Large T-antigen is shown in SEQ ID NO:69.

[0131] A modified CasM protein can have a low degree of sequence identity, a moderate degree of sequence identity, or a high degree of sequence identity over its length to a reference CasM protein, depending on the intended function of the CasM in question. By a “high degree of sequence identity” is meant approximately 90% sequence identity to 100% sequence identity, for example, about 90% . . . 95% . . . 98% sequence identity or higher. A “moderate degree of sequence identity” is typically between about 80% sequence identity to about 85% sequence identity, for example, about 80% identity or higher, such as about 85% sequence identity. A “low degree of sequence identity” is typically between about 50% identity and 75% identity, for example, about 50% identity, preferably about 60% identity to about 75% identity.

[0132] In some embodiments, the amino acid sequence of the reference CasM protein may be modified by deletion, insertion, or substitution of one or more amino acid residues (either conservative or non-conservative in nature), such that the activity of the CasM protein is either largely retained, enhanced, or reduced. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts that produce the proteins or errors due to PCR amplification.

[0133] Conservative substitutions are generally those substitutions that take place within a family of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic—aspartate and glutamate; (2) basic—lysine, arginine, histidine; (3) non-polar—alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar—glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids. It is reasonably predictable that an isolated replacement of leucine with isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid, will not have a major effect on the desired biological activity.

For example, the CasM protein may include up to about 5-10 conservative or non-conservative amino acid substitutions, or even up to about 15-100 or more, e.g., 50 or more, conservative or non-conservative amino acid substitutions, or any number between 5-100, so long as the desired function of the molecule remains intact.

[0134] In other embodiments, it may be desirable to modify one or more catalytic domains in order to render a nuclease-deactivated CasM protein, also termed “catalytically inactive,” “catalytically dead CasM,” “dead CasM,” or “dCasM,” such that the protein either fails to produce nucleic acid breaks, and/or binds a target sequence but does not cleave it. Such molecules lack all or a portion of nuclease activity and are unable to cleave a nucleic acid of interest and can therefore be used to regulate genes in a nucleic acid-guided manner. These dCasM proteins can be used alone or in fusions to synthetically repress (CRISPRi) or activate (CRISPRa) gene expression.

[0135] The CasM proteins can either be directly isolated and purified from bacteria, or synthetically or recombinantly produced using polynucleotides encoding the same.

CasM Polynucleotides

[0136] Nucleic acid sequences encoding representative CasM proteins are shown in SEQ ID NOS:28-36 (see Table 3) and these polynucleotides can be used to produce CasM proteins as described herein.

TABLE 3

Representative CasM DNA Sequences	
Species/Isolate	SEQ ID NO
<i>Eubacterium siraeum</i>	SEQ ID NO: 28
<i>Ruminococcus</i> sp., isolate 2789STDY5834971	SEQ ID NO: 29
<i>Ruminococcus bicirculans</i>	SEQ ID NO: 30
<i>Ruminococcus</i> sp., isolate 2789STDY5608892	SEQ ID NO: 31
<i>Ruminococcus</i> sp. CAG:57	SEQ ID NO: 32
<i>Ruminococcus flavefaciens</i> FD-1	SEQ ID NO: 33
<i>Ruminococcus albus</i> strain KH2T6	SEQ ID NO: 34
<i>Ruminococcus flavefaciens</i> strain XPD3002	SEQ ID NO: 35
<i>Ruminococcus</i> sp., isolate 2789STDY5834894	SEQ ID NO: 36

[0137] These polynucleotides can be designed to encode native CasM proteins, such as the proteins depicted in SEQ ID NOS:37-45 (see Table 1); homologs thereof, including orthologs found in other species; or other variants thereof. Moreover, a casM polynucleotide sequence can be modified to have a low degree of sequence identity, a moderate degree of sequence identity, or a high degree of sequence identity, over its length to a reference casM polynucleotide, depending on the intended function of the encoded CasM in question. By “a high degree of sequence identity” is meant approximately 90% sequence identity to 100% sequence identity, for example, about 90% . . . 95% . . . 98% sequence identity or higher. A “moderate degree of sequence identity” is typically between about 80% sequence identity to about 85% sequence identity, for example, about 80% identity or higher, such as about 85% sequence identity. A “low degree of sequence identity” is typically between about 50% identity and 75% identity, for example, about 50% identity, preferably about 60% identity to about 75% identity.

[0138] In some embodiments, the polynucleotide sequences are modified to enhance expression in a selected host cell. Codon usage bias refers to differences in the

frequency of occurrence of synonymous codons in coding DNA. For example, for the 20 standard amino acids in the genetic code, there are 64 different codons (61 codons encoding for amino acids, and 3 stop codons). The over-abundance in the number of codons allows several amino acids to be encoded by more than one codon. The genetic codes of different organisms are often biased towards the usage of one of the several codons that encode a particular amino acid. Thus, a greater frequency of one codon will be found than expected by chance in particular organisms. Accordingly, in order to enhance expression in a particular host cell, it is often desirable to manipulate polynucleotides to include codons that are biased for expression in the selected host cell. Several software packages are available online for this purpose. For example, a database from Integrated DNA Technologies, Coralville, Iowa (idtdna.com/CodonOpt), is a tool for producing modified sequences for expression in dozens of organisms. GeneScript, Piscataway, N.J., also provides modification tools through the OptimumGene™ algorithm (genscript.com/codon_opt.html?src=google&gclid=CIX3uoqexdICFRSUfgodu3sAlQ). See also, U.S. Pat. No. 8,326,547, incorporated herein by reference in its entirety.

[0139] Typically, polynucleotide sequences modified for expression in particular host cells will display from about 50%-99% sequence identity to the native sequences, such as 60%-95%, e.g. 65% . . . 70% . . . 75% . . . 80% . . . 85% . . . 90% . . . 95% or more sequence identity, or any integer between these ranges, to the native sequences.

[0140] Using these tools, polynucleotide sequences can be modified for expression in any commonly used host cell, such as but not limited to, bacterial cells and eukaryotic cells, including without limitation, bacterial cells such as *E. coli*, *Lactococcus lactis*, *Pseudomonas* systems, *Streptomyces* systems, *Bacillus subtilis* systems, *Brevibacillus* systems, coryneform bacteria, and halophilic bacteria; algal cells; yeast and other fungal cells; plant cells; mammalian cells such as human cells; insect cells, and the like.

[0141] SEQ ID NOS:1-9 show representative CasM-encoding polynucleotide sequences modified for expression in *E. coli* cells (see Table 4). SEQ ID NOS:10-18 show representative CasM-encoding polynucleotide sequences modified for expression in human cells (see Table 5). SEQ ID NOS:19-27 show CasM-encoding polynucleotide sequences modified for expression in *Zea mays* cells (see Table 6).

TABLE 4

Representative casM DNA Sequences Modified for Expression in <i>E. coli</i>	
Species/Isolate	SEQ ID NO
<i>Eubacterium siraeum</i>	SEQ ID NO: 1
<i>Ruminococcus</i> sp., isolate 2789STDY5834971	SEQ ID NO: 2
<i>Ruminococcus bicirculans</i>	SEQ ID NO: 3
<i>Ruminococcus</i> sp., isolate 2789STDY5608892	SEQ ID NO: 4
<i>Ruminococcus</i> sp. CAG:57	SEQ ID NO: 5
<i>Ruminococcus flavefaciens</i> FD-1	SEQ ID NO: 6
<i>Ruminococcus albus</i> strain KH2T6	SEQ ID NO: 7
<i>Ruminococcus flavefaciens</i> strain XPD3002	SEQ ID NO: 8
<i>Ruminococcus</i> sp., isolate 2789STDY5834894	SEQ ID NO: 9

TABLE 5

Representative casM DNA Sequences Modified for Expression in Human Cells	
Species/Isolate	SEQ ID NO
<i>Eubacterium siraeum</i>	SEQ ID NO: 10
<i>Ruminococcus</i> sp., isolate 2789STDY5834971	SEQ ID NO: 11
<i>Ruminococcus bicirculans</i>	SEQ ID NO: 12
<i>Ruminococcus</i> sp., isolate 2789STDY5608892	SEQ ID NO: 13
<i>Ruminococcus</i> sp. CAG:57	SEQ ID NO: 14
<i>Ruminococcus flavefaciens</i> FD-1	SEQ ID NO: 15
<i>Ruminococcus albus</i> strain KH2T6	SEQ ID NO: 16
<i>Ruminococcus flavefaciens</i> strain XPD3002	SEQ ID NO: 17
<i>Ruminococcus</i> sp., isolate 2789STDY5834894	SEQ ID NO: 18

TABLE 6

Representative casM DNA Sequences Modified for Expression in <i>Zea mays</i>	
Species/Isolate	SEQ ID NO
<i>Eubacterium siraeum</i>	SEQ ID NO: 19
<i>Ruminococcus</i> sp., isolate 2789STDY5834971	SEQ ID NO: 20
<i>Ruminococcus bicirculans</i>	SEQ ID NO: 21
<i>Ruminococcus</i> sp., isolate 2789STDY5608892	SEQ ID NO: 22
<i>Ruminococcus</i> sp. CAG:57	SEQ ID NO: 23
<i>Ruminococcus flavefaciens</i> FD-1	SEQ ID NO: 24
<i>Ruminococcus albus</i> strain KH2T6	SEQ ID NO: 25
<i>Ruminococcus flavefaciens</i> strain XPD3002	SEQ ID NO: 26
<i>Ruminococcus</i> sp., isolate 2789STDY5834894	SEQ ID NO: 27

[0142] The casM polynucleotides can also be modified to include sequences encoding N- or C-terminal nuclear localization signal sequences (NLS), such as for expression in eukaryotic cells. Such sequences are known, and include, without limitation, an NLS tag derived from the SV40 Large T-antigen. Such tag is present at the C-terminus of the proteins shown in SEQ ID NOS:61-69 (i.e., the last seven amino acids in these sequences).

[0143] The casM polynucleotides can be used to recombinantly produce CasM proteins using methods well known in the art.

CasM Complexes

[0144] CasM proteins can be complexed to a cognate nucleic acid guide (cognate guide/CasM complex) in order to deliver CasM in proximity with a target nucleic acid sequence. A cognate guide, such as a crRNA, is a polynucleotide that site-specifically guides a CasM nuclease, or a deactivated CasM nuclease, to a target nucleic acid region. The binding specificity is determined jointly by the complementary region on the cognate guide and a short DNA motif (protospacer adjacent motif or PAM) juxtaposed to the complementary region. The spacer present in the guide specifically hybridizes to a target nucleic acid sequence and determines the location of a Cas protein's site-specific binding and nucleolytic cleavage.

[0145] Cognate guide/CasM complexes can be produced using methods well known in the art. For example, the guide components of the complexes can be produced in vitro and CasM components can be recombinantly produced and then the guides and CasM proteins can be complexed together using methods known in the art. Additionally, cell lines constitutively expressing CasM proteins can be developed and can be transfected with the guide components, and

complexes can be purified from the cells using standard purification techniques, such as but not limited to affinity, ion exchange and size exclusion chromatography. See, e.g., Jinek M., et al., "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science* (2012) 337:816-821.

[0146] Alternatively, the components, i.e., the cognate guides and casM polynucleotides may be provided separately to a cell, e.g., using separate constructs, or together, in a single construct, or in any combination, and complexes can be purified as above.

[0147] Methods of designing particular guides, such as for use in the complexes, are known. See, e.g., Briner et al., "Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality," *Molecular Cell* (2014) 56:333-339. To do so, the genomic sequence for the gene to be targeted is first identified. The exact region of the selected gene to target will depend on the specific application. For example, in order to activate or repress a target gene using, for example, Cas activators or repressors, cognate guide/CasM complexes can be targeted to the promoter driving expression of the gene of interest. For genetic knockouts, guides are commonly designed to target 5' constitutively expressed exons which reduces the chances or removal of the targeted region from mRNA due to alternative splicing. Exons near the N-terminus can be targeted because frameshift mutations here will increase the likelihood of the production of a nonfunctional protein product. Alternatively, cognate guides can be designed to target exons that code for known essential protein domains. In this regard, non-frameshift mutations such as insertions or deletions are more likely to alter protein function when they occur in protein domains that are essential for protein function. For gene editing using HDR, the target sequence should be close to the location of the desired edit. In this case, the location where the edit is desired is identified and a target sequence is selected nearby.

[0148] The guides can be delivered to a cell. If the cell constitutively expresses a CasM nuclease, the CasM nuclease will then be recruited to the target site to cleave the target nucleic acid. If the cell does not express a CasM nuclease, complexes of cognate guide/CasM can be delivered to the cells to make breaks in the genome, thereby triggering the repair pathways in the cells.

[0149] Treated cells are then screened using methods well known in the art, such as using high-throughput screening techniques including, but not limited to, fluorescence-activated cell sorting (FACS)-based screening platforms, microfluidics-based screening platforms, and the like. These techniques are well known in the art. See, e.g., Wojcik et al., *Int. J. Molec. Sci.* (2015) 16:24918-24945. The cells can then be expanded and re-transfected with additional cognate guide/CasM complexes to introduce further diversity and this process can be repeated iteratively until a population with the desired properties is obtained. Single cell clones are sorted from the population, expanded and sequenced to recover the mutations that resulted in the desired function.

Applications of CasM

[0150] Due to its RNA-targeting abilities, CasM can be used to edit RNA and in some embodiments, to treat diseases caused by toxic RNA or improperly spliced RNA.

[0151] In some embodiments cognate guide/CasM complexes, such as, but not limited to crRNA/CasM complexes, are capable of sequence-specific ssRNA activity. Recogni-

tion and cleavage of an initial ssRNA complementary to the crRNA target sequence activates the CasM protein to carry out endonuclease activity toward any single-stranded RNA in a sequence-independent manner. The sequence-specific recognition of RNA of the crRNA/CasM complex facilitates the target knockdown of gene transcripts perturbing translation of a specific protein. The non-specific endonuclease activity of an activated crRNA/CasM complex in a cellular environment can result in cell death due to depletion of RNA encoding for essential gene transcripts. Thus, the specific RNA targeting and collateral endonuclease activity of an activated crRNA/CasM complex enables the sequence-specific selection of cells expressing a RNA transcript.

[0152] Thus, in further aspects, CasM complexes, such as, but not limited to crRNA/CasM complexes, can be used in methods of screening and killing cells, such as bacterial cells, that have not been modified by a DNA targeting nuclease (i.e., a Type II Cas9 nuclease). This method comprises contacting a NATNA/Cas9 complex to a locus of interest in a population of cells. Contacting the NATNA/Cas9 complex with the locus results in DNA cleavage and subsequent repair of the break by the endogenous cellular repair machine and the introduction of indels at the break site. The targeting of the NATNA/Cas9 complex to a targeted locus that encodes an RNA transcript results in indels in an RNA transcript sequence. This modified RNA transcript sequence is different compared to a transcript from an unmodified cell (a wild-type cell). A crRNA/CasM complex can then be targeted to the unmodified transcript, wherein crRNA/CasM recognition of the unmodified transcript results in activation of the sequence independent, single-stranded RNA targeting activity of the CasM protein and subsequent cell death. Alternatively, this method can be adapted to screen for the incorporation of a donor-poly-nucleotide into NATNA/Cas9 break site.

[0153] In another aspect, CasM complexes can be targeted to a eukaryotic exon coding region to cause exon skipping. This method comprises contacting a crRNA/deactivated CasM complex, such as, but not limited to a crRNA/dCasM complex, with either a donor site (5' end of an intron), a branch site (proximal to the 3' end of an intron), or an acceptor site (5' of an exon) of a pre-mRNA. Contacting the crRNA/dCasM complex to the various regions involved in exon splice events prevents the proper splicing of one of more exons together and causes the target exon to be "skipped", and thus is not included in the mature mRNA and therefore omitted from the translated polypeptide sequence.

[0154] In yet another aspect, CasM complexes are used for the detection of one or more target molecules in vitro. This method comprises contacting a cognate guide/CasM complex, such as a crRNA complex, with a ssRNA target of interest within a pool of nucleic acids. The crRNA/CasM complex can be added to a sample potential containing the ssRNA target of interest, in combination with a quenched fluorescent RNA reporter, for example a RNA hexamer with a 6-Carboxyfluorescein at the 5' end and a Iowa Black® FQ quencher (Integrated DNA Technologies, Coralville, Iowa) at the 3'end. Contacting of the crRNA/CasM complex with the ssRNA target, activates the CasM protein to carry out collateral cleavage of the quenched fluorescent RNA reporter where cleavage of the reporter and resulting in an increase fluorescence that can be read out using a spectrophotometer. The gain in fluorescence is used as a measure of the presence of a ssRNA target of interest.

[0155] In another aspect, CasM can be used for the targeted cleavage of an endogenous mRNA transcript while simultaneously delivering an exogenous mRNA transcript in cells. This method comprises contacting a cognate guide/CasM complex, with a disease-associated endogenous mRNA transcript, while simultaneously delivering of a mRNA coding for the non-disease exogenous polypeptide into a cell. Thus, the disease-associated phenotype is repressed while the non-disease phenotype is restored.

[0156] The CasM proteins described herein can also be used with associated cognate guides in order to activate or repress a target gene, to knockout a gene, to produce a nonfunctional protein product, or to alter protein function. The present invention includes methods of modulating in vitro or in vivo transcription using the various components and complexes described herein. In one embodiment, a cognate guide/CasM protein complex can repress gene expression by interfering with transcription when the cognate guide directs nucleic acid target binding of the complex to the promoter region of the gene. Use of the complexes to reduce transcription also includes complexes wherein the CasM protein is fused to a known down-regulator of a target gene (e.g., a repressor polypeptide). For example, expression of a gene is under the control of regulatory sequences to which a repressor polypeptide can bind. A cognate guide can direct nucleic acid target-binding of a repressor protein complex to the sequences encoding the regulatory sequences or adjacent the regulatory sequences such that binding of the repressor protein complex brings the repressor protein into operable contact with the regulatory sequences. Similarly, CasM can be fused to an activator polypeptide to activate or increase expression of a gene under the control of regulatory sequences to which an activator polypeptide can bind.

[0157] In one embodiment, CasM can be fused with a nuclease, or a mutant or an active portion thereof, as well as a cognate guide, in order to bring the nuclease into proximity with a target nucleic acid sequence, wherein the nuclease can produce a single-strand or double-strand break. In this way, a locus-specific cut in a target nucleic acid can be achieved using a cognate guide in combination with CasM, and the nuclease of interest. For example, it may be desirable to associate CasM with a restriction endonuclease in order to cleave at a particular restriction site in a target nucleic acid sequence. The restriction endonuclease can be selected from any of the various types of restriction endonucleases, such as, but not limited to, type I, II, III or IV. See, e.g., PCT Publication No. WO 2013/098244 to Brouns et al., published 4 Jul. 2013, incorporated herein by reference in its entirety, for methods of producing complexes between a Cas protein and a restriction endonuclease.

[0158] Using the methods described herein, any desired nucleic acid sequence, and in particular RNA sequences, for modification can be targeted, including without limitation, protein coding mRNA sequences, in order to reduce or restore the function of the gene product; regions that have a propensity to incorporate nucleotide sequences from a donor template, termed "HDR hotspots" herein; safe harbor regions, i.e., regions where nucleotide sequences can be inserted without disrupting neighboring gene function; non-coding regulatory regions in nucleic acid sequences; and the like.

[0159] Protein coding sequences, including RNA such as mRNA, for targeting by the methods described herein include, but are not limited to, mammalian antibodies (ABs)

(IgG, IgA, IgM, IgE), antibody fragments such as Fc regions, antibody Fab regions, antibody heavy chains, antibody light chains, antibody CDRs, nanobodies, chimeric antibodies and other IgG domains; T cell receptors (TCR); endonucleases and exonucleases, such as TALENS, CRISPR nucleases such as Cas9, Cas3, Cpf1, ZnFN, meganucleases, nuclease domains such as HNH domain, RuvC domain; recombinases such as Cre, Tre, Brec1, Flp, γ -integrase, Int14 integrase, XerD recombinase, HP1 integrase; DNA topoisomerase; transposons such as the Tc1/mariner family, Tol2, piggyBac, Sleeping beauty; RAG proteins; retrotransposons such as LTR-retrotransposons and non-LTR retrotransposons (Alu, SINE, LINE); enzymes including but not limited to arginases, glycosyldases, proteases, kinases, and glycosylation enzymes such as glycosyltransferase; anticoagulants such as protein C, Protein S and antithrombin; coagulants such as thrombin; nucleases such as DNases, RNAses, helicases, GTPases; DNA or RNA binding proteins; reporter molecules, such as Green Fluorescent Protein (GFP); cell penetrating peptides and their fusions with cargo proteins; membrane proteins such as GPCRs, pain receptors such as TRP channels and ion channels; cell surface receptors including but not limited to EGFR, FGFR, VEGFR, IGFR and ephrin receptor; cell adhesion molecules like integrins and cadherins; ion channels; rhodopsins; immunoreceptors such as CD28, CD80, PD-1, PD-L1, CTLA-4, CXCR4, CXCR5, B2M, TRACA, TRBC; proteins known to be involved with genetic defects; secreted proteins including but not limited to hormones, cytokines, growth factors; vaccine antigens such as viral proteins from human immunodeficiency virus (HIV), Dengue, cytomegalovirus (CMV), Ebola, Zika and oncolytic viruses; snake toxin proteins and peptides including but not limited to phospholipases and metalloproteases; ribosomal cyclic peptides.

[0160] The present invention also encompasses genome engineering methods for preventing or treating diseases, disorders, and conditions using the various methods described herein. In one embodiment, a genome engineering method uses the introduction of nucleic acid sequences into an organism or cells of an organism (e.g., patient) to achieve expression of components of the present invention to provide modification of a target function. For example, cells from an organism may be engineered, *ex vivo*, by (i) introduction of vectors comprising expression cassettes expressing the various components, (ii) direct introduction of a NATNA and/or donor polynucleotides and CasM proteins, or (iii) introduction of combinations of these components. The engineered cells are provided to an organism (e.g., patient) to be treated.

[0161] Examples of genome engineering and techniques for therapy are known in the art (see, e.g., Kay, M. A., *Nature Reviews Genetics* (2011) 12:316-328; Wang et al., *Discov. Med.* (2014) 18:67-77; Wang et al., *Discov. Med.* (2014) 18:151-61; "The Clinibook: Clinical Gene Transfer State of the Art," Odile Cohen-Haguenauer (Editor), EDP Sciences (Oct. 31, 2012), ISBN-10: 2842541715).

[0162] In some aspects, components of the present invention are delivered using nanoscale delivery systems, such as nanoparticles. Additionally, liposomes and other particulate delivery systems can be used. For example, vectors including the components of the present methods can be packaged in liposomes prior to delivery to the subject or to cells derived therefrom, such as described in U.S. Pat. Nos.

5,580,859; 5,264,618; 5,703,055, each of which is incorporated herein by reference in its entirety. Lipid encapsulation is generally accomplished using liposomes that are able to stably bind or entrap and retain nucleic acid.

[0163] The methods described herein can also be used to generate non-human genetically modified organisms, such as mice, plants, and the like.

[0164] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. From the above description and the following Examples, one skilled in the art can ascertain essential characteristics of this invention, and without departing from the spirit and scope thereof, can make changes, substitutions, variations, and modifications of the invention to adapt it to various usages and conditions. Such changes, substitutions, variations, and modifications are also intended to fall within the scope of the present disclosure.

EXPERIMENTAL

[0165] Aspects of the present invention are further illustrated in the following Examples. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, concentrations, percent changes, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, temperature is in degrees Centigrade and pressure is at or near atmospheric. It should be understood that these Examples, while indicating some embodiments of the invention, are given by way of illustration only.

[0166] The following Examples are not intended to limit the scope of what the inventors regard as various aspects of the present invention.

Example 1

Discovery of a New CRISPR-Associated (Cas) Protein in Silico

[0167] This Example describes the in silico discovery of a new Cas protein, termed “CasM,” from genomic sequencing data. The overall approach used was similar to methods described in Shmakov et al., “Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems” *Molecular Cell* (2015) 60:385-397. In particular a computational pipeline was used to search sequencing data for CRISPR arrays in whole genomes and metagenomic contigs.

[0168] Every contig or genome in the data set was inspected to determine if it contained a CRISPR array using Minced (github.com/ctSkennerton/minced) and PILERCR (drive5.com/pilercr/).

[0169] Any time a CRISPR array was found in a contig or genome, the surrounding DNA sequence (up to 10 kilobases on either side of the CRISPR array) was further inspected for open reading frames (ORFs) using the tool getorf ([emboss.sourceforge.net/apps/cvs/emboss/apps/getorf.html](https://sourceforge.net/apps/cvs/emboss/apps/getorf.html)).

[0170] The primary amino acid sequence of each predicted ORF was analyzed for potential functional domain annotations using the tool HHpred (homology detection & structure prediction by HMM-HMM comparison; toolkit.tuebingen.mpg.de/hhpred). HHpred allows the user to specify which databases to compare the amino acid sequence against to find similar protein domains. The databases searched

included PFAM (which includes a large collection of protein families; <http://pfam.xfam.org/>), PDB (protein databank; www.pdb.org), CDD (conserved domain database; ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), and KEGG (Kyoto Encyclopedia of Genes and Genomes; genome.jp/kegg/).

[0171] Using these tools, ORFs encoding a new putative CRISPR-associated protein, termed “CasM,” was found in several species in the Clostridia family. The ORFs are proximal to a predicted HTH DNA binding protein with homology to a CRISPR-associated WYL domain. See FIG. 1 for a representative map of a CRISPR operon found in *Eubacterium siraeum* (Genome Accession No. NZ_DS499551.1, coordinates 211.800-220.497). The various CRISPR locus features are described in Table 7.

TABLE 7

CasM CRISPR locus for <i>Eubacterium siraeum</i> (FIG. 1)
101 corresponds to the CasM open reading frame (ORF) (Protein accession: WP_005358205.1)
102 corresponds to a CRISPR Array
103 corresponds to an ORF containing a RctB RNA ligase domain (Protein accession: WP_005358214.1)
104 corresponds to an ORF containing a WYL DNA binding domain (Protein accession: WP_005358216.1)

[0172] The results of HHpred analysis were analyzed to determine if the CasM-encoding ORFs had predicted domains commonly found in CRISPR-associated proteins. No annotations were found for CasM, thus indicating that the protein was novel.

[0173] The sequences for the various native CasM proteins are shown in SEQ ID NOS:37-45 (see Table 1) and the native polynucleotides encoding therefor are shown in SEQ ID NOS:28-36 (see Table 3). This protein has no significant homology to any known protein families or to any Class 2 Cas effectors.

Example 2

Codon Modification of Native casM Sequences

[0174] This Example describes the process of codon optimizing CasM coding sequences to improve expression in selected host cells.

[0175] Native casM nucleotide sequences were retrieved from the reference genomes or metagenomic contigs of the host microbes as described in Example 1. The amino acid sequences of the coding regions were generated with the ExPASy DNA translation tool (web.expasy.org/translate/). Next, these amino acid sequences were entered into the Integrated DNA Technologies (Coralville, Iowa) Codon Optimization tool (idtdna.com/CodonOpt). “Amino acid” was chosen for the “Sequence Type” option and “Gene” was chosen for the “Product Type” option. For each native casM sequence, codon modifications were performed to increase expression in *E. coli*, human, and *Zea mays* cells.

[0176] SEQ ID NOS:1-9 show the modified sequences for use in *E. coli* (see Table 4). SEQ ID NOS:10-19 show the modified sequences for use in human cells (see Table 5). SEQ ID NOS:20-27 show the modified sequences for use in *Z. mays* cells (see Table 6). Table 8 shows the percent identity of the modified sequences to the native sequences.

TABLE 8

casM bacterial strain	Percent Sequence Identity to Native casM Sequences		
	modified for <i>E. coli</i> cells	modified for human cells	modified for <i>Z. mays</i> cells
<i>Eubacterium siraeum</i>	75%	77%	77%
<i>Ruminococcus</i> sp., isolate 2789STDY5834971	77%	76%	76%
<i>Ruminococcus bicirculans</i>	76%	76%	77%
<i>Ruminococcus</i> sp., isolate 2789STDY5608892	76%	75%	77%
<i>Ruminococcus</i> sp. CAG:57	76%	77%	76%
<i>Ruminococcus flavefaciens</i>	76%	77%	76%
FD-1			
<i>Ruminococcus albus</i> strain KH2T6	76%	77%	77%
<i>Ruminococcus flavefaciens</i> strain XPD3002	76%	77%	77%
<i>Ruminococcus</i> sp., isolate 2789STDY5834894	75%	76%	78%

Example 3

Production of CasM Expression Plasmids for DNA Interference Assays

[0177] This Example describes the production of plasmids that express the CasM protein.

[0178] The modified casM nucleotide sequences set forth in Example 2 were synthesized in vitro. The DNA sequences were cloned into an appropriate plasmid for expression in *E. coli*.

[0179] For *E. coli* expression, the *E. coli*-modified sequences were cloned into a p14A plasmid backbone using appropriate restriction nucleases. The plasmid backbone contained a T7 promoter upstream of the CasM coding sequence to facilitate transcription in cells.

[0180] The p14A plasmid backbone also contained a cloning site enabling the insertion of a minimal CRISPR array. The minimal CRISPR array contained one repeat sequence, followed by one spacer sequence, followed by one repeat sequence. The plasmid backbone also contained a T7 promoter upstream of the CRISPR array site, a kanamycin resistance gene, and a Cole1 origin of replication.

[0181] Similar techniques are used for preparing plasmids for expression in human and *Zea mays* cells. Once the plasmids are produced, they are transfected into the selected cell, e.g., *E. coli*, human, or plant cells (e.g., *Zea mays* cells).

Example 4

Plasmid Interference Assay

[0182] This Example describes the use of CasM in an assay to evaluate its ability to cleave double-stranded DNA in the form of a target plasmid. The overall approach is similar to methods used in Burnstein et al., *Nature* (2016) 542:237-241.

[0183] The CasM expression plasmid in Example 3 is transformed into *E. coli* cells. The cells are grown in a medium containing kanamycin to select only for cells that contain the CasM expression plasmid.

[0184] A target plasmid is constructed that contains the spacer sequence contained in the CRISPR array of the CasM expression plasmid. Adjacent to the spacer sequence is a

randomized PAM sequence of 7 nucleotides. Plasmid libraries containing randomized PAM sequences are assembled by annealing a DNA oligonucleotide containing a target with a 7 nt randomized PAM region with a primer and extended with Klenow Fragment (New England Biolabs, Ipswich, Mass.). The double-stranded DNA is digested with EcoRI and NcoI and ligated into a pUC19 backbone. The ligated library is transformed into *E. coli* DH5α and cells are harvested, the plasmids extracted and purified. 200 ng of the pooled library is transformed into electro-competent *E. coli* harboring a CRISPR locus or a control plasmid with no locus. The transformed cells are plated on selective media containing carbenicillin (100 mg L⁻¹) and chloramphenicol (30 mg L⁻¹) for 30 hours at 25° C. Plasmid DNA is extracted and the PAM sequence is amplified with adapters for Illumina sequencing. The 7 nt PAM region is extracted and PAM frequencies calculated for each 7 nt sequence. PAM sequences depleted above the specified threshold are used to generate a sequence logo with WebLogo (weblogo.berkeley.edu). If depleted PAMs are present, this shows that the nuclease is a double-stranded DNA nuclease.

Example 5

Targeted Modification of HEK293 Cells Using CasM

[0185] This Example illustrates the use of CasM to modify human embryonic kidney (HEK293) cells at specific target locations.

[0186] casM polynucleotides are transfected into HEK293 cells constitutively expressing a CasM-GFP fusion (HEK293-CasM-GFP), using the Nucleofector™ 96-well Shuttle System (Lonza, Allendale, N.J.) and the following protocol. The casM polynucleotides are designed to target the FUT8 gene. Equal molar amounts of casM polynucleotide components are prepared in an annealing buffer (1.25 mM HEPES, 0.625 mM MgCl₂, 9.375 mM KCl at pH 7.5), incubated for 2 minutes at 95° C., removed from the thermocycler, allowed to equilibrate to room temperature, and dispensed in a 10 µL final volume in a 96-well plate. Culture medium is aspirated from HEK293-CasM-GFP cells, and the cells are washed once with calcium and magnesium-free PBS and then trypsinized by the addition of TrypLE (Life Technologies, Grand Island, N.Y.) followed by incubation at 37° C. for 3-5 minutes. Trypsinized cells are gently pipetted up and down to form a single cell suspension and added to DMEM complete culture medium composed of DMEM culture medium (Life Technologies, Grand Island, N.Y.) containing 10% FBS (Fisher Scientific, Pittsburgh, Pa.) and supplemented with penicillin and streptomycin (Life Technologies, Grand Island, N.Y.).

[0187] The cells are then pelleted by centrifugation for 3 minutes at 200xg, the culture medium aspirated and cells resuspended in PBS. The cells are counted using the Countess™ II Automated Cell Counter (Life Technologies, Grand Island, N.Y.). 2.2×10⁷ cells are transferred to a 50 ml tube and pelleted. The PBS is aspirated and the cells resuspended in Nucleofector™ SF (Lonza, Allendale, N.J.) solution to a density of 1×10⁷ cells/mL. 20 µL of the cell suspension are then added to individual wells containing 10 µL of casM polynucleotide components and the entire volume is transferred to the wells of a 96-well Nucleocuvette™ Plate (Lonza, Allendale, N.J.). The plate is loaded onto the Nucleofector™ 96-well Shuttle™ (Lonza, Allendale, N.J.)

and cells are nucleofected using the 96-CM-130 Nucleofector™ program (Lonza, Allendale, N.J.). Post-nucleofection, 70 µL DMEM complete culture medium is added to each well and 50 µL of the cell suspension are transferred to a collagen coated 96-well cell culture plate containing 150 µL pre-warmed DMEM complete culture medium. The plate is then transferred to a tissue culture incubator and maintained at 37° C. in 5% CO₂ for 48 hours.

[0188] Genomic DNA (gDNA) is prepped using the QuickExtract DNA extraction solution (Illumina, San Diego, Calif.) pursuant to the manufacturer instructions. Sequencing amplicons of between 150 bp to 200 bp are designed to span the CasM RNP FUT8 target site. Using previously isolated gDNA, a first PCR is performed using Herculase II Fusion DNA Polymerase™ (Agilent, Santa Clara, Calif.) with primers comprising an adapter sequences and a sequence specific to the region flanking the FUT8 target site. A second PCR is performed using the amplicons of the first round of PCR as template at $\frac{1}{20}^{\text{th}}$ the volume of the PCR reaction volume. The second PCR uses a second set of primers comprising a sequence complementary to the adapter sequence of the first primer pair, a barcode index sequence unique to each sample, and a flow cell adapter sequence. Amplicons are pooled and analyzed on a 2% TBE gel and bands of expected amplicon sizes are gel purified using the QIAEX II Gel extraction Kit™ (Qiagen, Venlo, Luxembourg). The concentrations of purified amplicons are evaluated using the double-stranded DNA BR Assay Kit and Qubit System™ (Life Technologies, South San Francisco, Calif.) and library quality determined using the Agilent DNA100Chip and Agilent Bioanalyzer 2100 System™ (Agilent, Santa Clara, Calif.). After validation of library quality, the library is sequenced on a MiSeq Benchtop Sequencer™ (Illumina, San Diego, Calif.) with the MiSeq Reagent Kit v2™ (300 cycles, Illumina, San Diego, Calif.) per manufacturer instructions for 151 bp paired end reads.

[0189] The identity of products in the sequencing data is analyzed based upon the index barcode sequence adapted onto the amplicon in the second round of PCR. A computational script is used to process the MiSeq data by executing the following tasks:

[0190] 1. Joining paired end reads with the aid of fastqjoin (Aronesty 2011: code.google.com/p/ea-utils);

[0191] 2. Validating the sequence reads for appropriate primer sequences being present at both 5' and 3' ends of the read sequence using fastx_barcode_splitter (hannonlab.cshl.edu/fastx_toolkit/index.html); reads lacking correct primer sequences at both ends are discarded.

[0192] 3. Comparing Read sequences to expected wild type FUT8 sequence; identical read sequences are classified as having the same indel modification.

[0193] Other chromosomal loci within HEK293 cells are similarly modified by selection of an appropriate spacer sequence for the CasM RNP. Selection is specific to a specific gene target and the procedure outlined in this

Example is readily modifiable by one of ordinary skill in the art for other gene targets.

[0194] This procedure can provide data to verify the CasM RNP and to detect nucleic acid-guided nuclease activity at targeted loci in HEK293 cells.

Example 6

CasM CRISPR Array Processing Assay

[0195] This Example describes the CRISPR array processing activity of a CasM protein. The following method may be practiced with other CasM protein homologs to characterize their CRISPR array processing capabilities.

[0196] A. Identification of the CRISPR Array Repeat Sequence

[0197] The CRISPR array of the CasM protein homolog (SEQ ID NO:39) was analyzed in silico and the repeat sequence identified. The in silico structure of a CRISPR repeat sequence (SEQ ID NO:51) associated with the CasM protein (SEQ ID NO:39) as predicted using an RNA folding algorithm (rna.urmc.rochester.edu/RNA structureWeb/Servers/Predict1.html) is shown in FIG. 2 (SEQ ID NO:51). The various CRISPR repeat sequence structural components represented in FIG. 2 are described in Table 9.

TABLE 9

Numerical Indicators Used to Illustrate CasM CRISPR Repeat Sequence Structural Components (FIG. 2)

470 corresponds to a 5' repeat handle sequence
471 corresponds to a stem-duplex formed by a first stem duplex strand hybridized to a second stem duplex strand
472 corresponds to a loop sequences
473 corresponds to a 5' repeat handle sequence
474 corresponds to the 3' attachment point of a spacer sequence
475 corresponds to the CRISPR repeat processing positions performed by CasM upon guide binding
10 indicates the tenth nucleotide position
20 indicates the twentieth nucleotide position
30 indicates the thirtieth nucleotide position

The repeat sequence was used to design a CRISPR array by incorporating spacer sequences 5', 3', or both 5' and 3' of the repeat sequence. The sequences were used for synthesis as RNA.

The synthetic CasM CRISPR array is represented in FIG. 4 and structural components 470-473 are detailed in Table 9. Additional components 490 and 491 correspond to a first 5' and a first 3' spacer sequence, respectively. The CRISPR array components are shown in Table 10.

TABLE 10

CRISPR Array Components			
SEQ ID NO:	CRISPR Array Configuration	Sequence	Size (nt)
SEQ ID NO: 46	spacer-repeat-spacer	UGAUACUGCUUUUGAUGUCAGCAUUGC AUAU <u>CUACUAUACUGGUGCGAAUUG</u> <u>CACUAGCUAAAAAUCAUAACCAUAU</u> GUUCUUCUGCGUUCAUAU	96
SEQ ID NO: 47	spacer-repeat	UGAUACUGCUUUUGAUGUCAGCAUUGC AUAU <u>CUACUAUACUGGUGCGAAUUG</u> <u>CACUAGCUAAAAAU</u>	66
SEQ ID NO: 48	repeat-spacer	<u>CUACUAUACUGGUGCGAAUUGCACU</u> AGUCUAAAUGAUACUGCUUUGAUG UCAGCAUUGCAUAU	66

*CRISPR repeat sequence is underlined

[0198] SEQ ID NO:46 comprises, in a 5' to 3'orientation, CRISPR array structural components 490, 470-473, and 491. SEQ ID NO:47 comprises, in a 5' to 3'orientation, CRISPR array structural components 490 and 470-473. SEQ ID NO:48 comprises, in a 5' to 3'orientation, CRISPR array structural components 470-473 and 491.

[0199] Alternative to synthesis, CRISPR arrays may be made via PCR using 3' overlapping primers containing DNA sequences corresponding to CRISPR array components and incorporation of a T7 promoter sequence 5' of the CRISPR arrays, followed by in vitro transcription.

[0200] B. CasM Purification

[0201] The CasM protein coding sequence was codon-optimized for expression in *E. coli* and incorporated into a modified pET plasmid backbone downstream of a maltose binding protein (MBP) using appropriate restriction nucleases. The plasmid backbone contained a T7-Lac promoter upstream of the MBP-CasM coding sequence to facilitate transcription in cells. Additionally, the plasmid backbone contained an kanamycin resistance gene and a ColE1 origin of replication.

[0202] The CasM expression plasmid was transformed into Rosetta2 (DE3) cells, and cells were grown in two 1L shake flasks at 37° C. until cells reached an optical density of 0.6, after which protein expression was induced by addition of 0.5 mM IPTG. Cells were then incubated at 16° C. overnight.

[0203] Cells were collected via centrifugation and lysed via sonication. Cell debris was pelleted, and the clarified lysate was purified using a combination of HisTrap column chromatography, followed by cleavage of the MBP tag, and finally cation exchange column chromatography. Final purified protein was quantified using a NanoDrop™ 2000 spectrophotometer (ThermoFisher, Waltham, Mass.), and stored at -80° C.

[0204] C. In Vitro CRISPR Array Processing

[0205] Synthetic CRISPR array reagents were resuspended in water to a final concentration of 250 μM and diluted to a working concentration of 250 nM. CRISPR arrays were incubated at 95° C. for two minutes and cooled by 0.5° C./sec in a thermocycler to a final temperature of 25° C.

[0206] CasM was diluted to a final concentration of 500 nM in 1× cleavage buffer (20 mM HEPES, 100 mM KCl, 5 mM MgCl₂, and 5% glycerol at pH 7.4). The reaction was

initiated by addition of CasM protein to denatured CRISPR arrays in a final reaction volume of 12 μL, followed by incubation at 37° C. for 15 minutes. The reaction was terminated by heat inactivation at 95° C. for 2 minutes, and 6 μL of the reaction was mixed with 6 μL of 2×RNA loading buffer (New England Biolabs, Ipswich, Mass.). Low Range ssRNA Ladder™ (New England Biolabs, Ipswich, Mass.) was diluted 125-fold in water and 7 μL were mixed with 7 μL of 2×RNA Loading Dye™ (New England Biolabs, Ipswich, Mass.) and incubated at 90° C. for 4 minutes and then incubated on ice for 5 minutes. CRISPR array processing reactions and ssRNA ladder were analyzed on a Mini-PROTEAN 15% TBE-Urea™ (Bio-RAD, Hercules, Calif.) run at 200 V for 1 hour in 1×TBE running buffer. The gel was stained using 2×SYBR Gold™ (MilliporeSigma, St. Louis, Mich.) for 15 minutes and visualized using a Gel Doc™ EZ System™ (Bio-RAD, Hercules, Calif.). The results of the CRISPR array processing reactions are shown in FIG. 3 and lane order is presented in Table 11.

TABLE 11

CRISPR Array Cleavage Gel Lane Order			
Lane	CRISPR Array Configuration	SEQ ID NO:	CasM
1		Low Range ssRNA Ladder	
2	spacer-repeat-spacer	SEQ ID NO: 46	-
3	spacer-repeat	SEQ ID NO: 47	-
4	repeat-spacer	SEQ ID NO: 48	-
5	spacer-repeat-spacer	SEQ ID NO: 46	+
6	spacer-repeat	SEQ ID NO: 47	+
7	repeat-spacer	SEQ ID NO: 48	+

[0207] The results of the CRISPR array cleavage assays (FIG. 3) demonstrated that the CasM protein is capable of processing a cognate CRISPR array. The three bands shown in Lane 1 correspond to 150, 80 and 50 nucleotide standards of the Low Range ssRNA Ladder™ (New England Biolabs, Ipswich, Mass.), respectively). Indicator 480 in FIG. 3 corresponds to a processed CasM crRNA comprising a portion of the CRISPR repeat sequence and a spacer sequence. Indicator 481 corresponds to RNA species cleaved from the 5' end of the CRISPR array following addition of CasM.

[0208] The CasM cleaved nucleotides in the 5' region of the repeat element (FIG. 3, comparing Lane 2 to 5; comparing Lane 3 to Lane 6), and exhibited no cleavage 3' of the repeat element (FIG. 3, comparing Lane 4 to Lane 7). crRNA proceeded from the CasM CRISPR array and therefore had a 5' repeat element and a spacer element 3' of the repeat. In the absence of CasM, no cleavage of the crRNA was observed (FIG. 3, Lanes 2, 3, and 4).

[0209] Schematics of the crRNA processing regimes are depicted in FIG. 5. In FIG. 5, panel I corresponds to the reaction in FIG. 3, Lane 5; FIG. 5 panel II corresponds to the reaction in FIG. 3, Lane 6; and FIG. 5 panel III corresponds to the reaction in FIG. 3, Lane 7. The various components represented in FIG. 5 are described in Table 12.

TABLE 12

Numerical Indicators Used to Illustrate the Results of the in vitro CRISPR Array Cleavage Assay (FIG. 5)
492 corresponds to a spacer-repeat-spacer CRISPR array (SEQ ID. NO: 46)
493 corresponds to a spacer-repeat CRISPR array (SEQ ID. NO: 47)
494 corresponds to a repeat-spacer CRISPR array (SEQ ID. NO: 48)
495 corresponds to a CasM protein
496 corresponds to a processed crRNA
497 corresponds to a RNA species cleaved from the 5' end of the CRISPR array
498 corresponds to a processed CRISPR repeat sequence
499 corresponds to a RNA species cleaved from the 5' end of the CRISPR repeat sequence

Example 7

CasM ssRNA Cleavage Assay

[0210] This Example illustrates the use of a crRNA/CasM protein complex to carry out ssRNA cleavage. The following method may be practiced with other CasM protein and crRNA to cleave ssRNA targets.

[0211] A. Generation of ssRNA Target

[0212] A ssRNA target was generated via PCR amplification of a 224 nucleotide target sequence from a plasmid. A T7 promoter sequence was incorporated into the 5' end of the reverse PCR primer (SEQ ID NO:50) for transcription. The primers used for ssRNA target DNA template are presented in Table 13,

TABLE 13

ssRNA Target DNA Template Primers		
SEQ ID NO:	Name	Sequence
SEQ ID NO: 49	Forward primer	CGAAATTAAATACGACTCACTATAGGTTTCGAT TATGCGGCCGTGT
SEQ ID NO: 50	Reverse primer	AGGAGATATAACCATGGGCAGCA

*T7 Promoter sequence underlined.

[0213] The primers were present at a concentration of 400 nM each. PCR reactions were performed using Q5 Hot Start High-Fidelity 2× Master Mix™ (New England Biolabs, Ipswich, Mass.) following the manufacturer's instructions with 10 ng of plasmid template. PCR assembly reactions were carried out using the following thermal cycling conditions: 98° C. for 2 minutes; 20 cycles of 10 seconds at 98° C.; 15 seconds at 60° C.; 30 seconds at 72° C.; and a final extension at 72° C. for 2 minutes. DNA product quality was evaluated after the PCR reaction by agarose gel electrophoresis (1.5%, SYBR® Safe; Life Technologies, Grand Island, N.Y.).

[0214] Between 0.1-0.5 µg of the amplified ssRNA target DNA template was used as a template for transcription using T7 High Yield RNA Synthesis Kit™ (New England Biolabs, Ipswich, Mass.) for approximately 16 hours at 37° C. Transcription reactions were treated with DNase I (New England Biolabs, Ipswich, Mass.) and purified using Gene-Jet RNA Cleanup and Concentration Kit™ (Life Technologies, Grand Island, N.Y.). The quality of the transcribed RNA was checked by agarose gel electrophoresis (2%, SYBR® Safe; Life Technologies, Grand Island, N.Y.) and quantified using the Quant-iT™ RNA Assay Kit™ (ThermoFisher, Waltham, Mass.).

[0215] B. Designing CasM crRNA

[0216] The 224 nucleotide ssRNA target sequence was probed in silico for a 30 nucleotide target sequence. The target sequence was appended in silico to the 3' end of the CasM crRNA repeat sequence and the crRNA sequence was provided to a commercial manufacturer for synthesis.

[0217] C. ssRNA Cleavage Assay

[0218] Synthetic crRNA reagents were resuspended in water to a final concentration of 250 µM and diluted to a suitable working concentration of 250 nM. In vitro transcribed ssRNA target was diluted to 43 ng/µL in water. Both the crRNA and the ssRNA target reagents were separately incubated at 95° C. for two minutes and cooled by 0.5° C./sec in a thermocycler to a final temperature of 25° C. The CasM protein was diluted to various concentrations in water and 1x cleavage buffer. Denatured crRNA was added at various concentrations to the CasM protein and incubated in a thermocycler for 10 minutes at 37° C. The cleavage reactions were initiated by the addition of the ssRNA target to a final concentration of 56.4 nM in a final reaction volume of 12 µL. The concentration of each component in the various reactions is shown in Table 14.

TABLE 14

ssRNA Targeting Reaction Component Concentrations				
Reaction	nM CasM	nM crRNA	nM ssRNA target	Molar ratio CasM:crRNA:ssRNA target
1	11.3	33.8	56.4	0.2:0.6:1
2	22.5	67.6	56.4	0.4:1.2:1
3	33.8	101.5	56.4	0.6:1.8:1
4	45.1	135.3	56.4	0.8:2.4:1
5	56.4	169.1	56.4	1:3:1
6	112.7	338.2	56.4	2:6:1
7	225.5	676.4	56.4	4:12:1
8	338.2	1014.6	56.4	6:18:1
9	450.9	1352.8	56.4	8:24:1
10	563.7	1691.0	56.4	10:30:1
11	0.0	0.0	56.4	0:3:1
12	56.4	0.0	56.4	1:0:1
13	0.0	169.1	56.4	0:3:1
14	0.0	169.1	0.0	1:3:0

[0219] Samples were mixed and centrifuged briefly before being incubated for 1 hour at 37° C. Reactions were terminated by incubating the reaction at 95° C. for 2 minutes followed by the addition of 100 U/μL of Proteinase K (New England Biolabs, Ipswich, Mass.), 4 M urea, 5 μM DTT, 50 μM EDTA and incubation at 37° C. for 15 minutes. 7 μL of each reaction was mixed with 6 μL of 2×RNA Loading Dye (New England Biolabs, Ipswich, Mass.) and incubated at 90° C. for two minutes. Low Range ssRNA Ladder™ (New England Biolabs, Ipswich, Mass.) was diluted 125-fold in water and 7 μL were mixed with 7 μL of 2×RNA Loading Dye™ (New England Biolabs, Ipswich, Mass.) and incubated at 90° C. for 4 minutes and then incubated on ice for 5 minutes. Cleavage reactions and ssRNA ladder were analyzed on a Mini-PROTEAN 15% TBE-Urea™ (Bio-RAD, Hercules, Calif.), run at 200 V for 1 hour in 1×TBE running buffer. Gel was stained using 2×SYBR Gold™ (Life Technologies, Grand Island, N.Y.) for 15 minutes and visualized with using a Gel Doc EZ System™ (BioRAD, Hercules, Calif.). The results of the crRNA cleavage assay are shown in FIG. 6 and the components of each lane shown in Table 15. Numerical indicator 500 corresponds to the ssRNA target. Numerical indicator 501 corresponds to a ssRNA target hybridized to the spacer sequences of the CasM crRNA. Numerical indicator 502 corresponds to an unprocessed CasM crRNA. Numerical indicator 503 corresponds to the processed CasM crRNA species.

TABLE 15

ssRNA Cleavage Gel Lane Order		
Lane	Molar ratio CasM:crRNA:ssRNA target	
1	0.2:0.6:1	
2	0.4:1.2:1	
3	0.6:1.8:1	
4	0.8:2.4:1	
5	1:3:1	
6	2:6:1	
7	4:12:1	
8	6:18:1	
9	8:24:1	
10	10:30:1	
11	ssRNA Ladder	
12	0:3:1	

TABLE 15-continued

ssRNA Cleavage Gel Lane Order		
Lane	Molar ratio CasM:crRNA:ssRNA target	
13	1:0:1	
14	0:3:1	
15	1:3:0	

[0220] The results of the ssRNA cleavage assay shown in FIG. 6 demonstrated that a CasM:crRNA protein complex was capable of ssRNA target cleavage. The results of this procedure demonstrate that increasing the amount of CasM: crRNA complexes resulted in decreased amounts of ssRNA target (FIG. 6, indicator 500) visualized on the gel (FIG. 6, Lanes 1-10).

Example 8

Production of CasM and RtcB Expression Plasmids for MS2 Phage Drop Plaque Assays

[0221] This Example describes the production of plasmids for the expression of CasM, RtcB (RNA 3'-terminal phosphate cyclase, group B), and a corresponding CRISPR array in *E. coli* for use in a MS2 phage drop plaque assay. The following method can be practiced with other CasM, RtcB, and CRISPR array homologs.

[0222] The casM and rtcB nucleotide sequence from *Eubacterium siraeum* (SEQ ID NO:37 and SEQ ID NO:60, respectively) are selected and codon optimized for expression in *E. coli*. The *E. coli*-modified sequences are cloned into a p14A plasmid backbone using appropriate restriction nucleases. The plasmid backbone contains a T7 promoter upstream of each protein coding sequence to facilitate transcription in cells. Two control plasmids, one containing only the casM gene sequence under the control of a T7 promoter and the other plasmid only containing the rtcB gene sequence under the control of a T7 promoter, can also be constructed.

[0223] A spacer sequence that has homology with the MS2 phage genome is engineered in silico flanked 5' and 3' by the *Eubacterium siraeum* CasM CRISPR repeat sequence. A non-targeting spacer with no homology to the MS2 phage or *E. coli* genome, is similarly engineered as a control. Both sequences are subcloned into separate plasmids between an upstream T7 promoter sequence and a downstream transcription terminator sequence.

Example 9

MS2 Phage Drop Plaque Assay

[0224] This Example describes the use of CasM and RtcB in an assay to evaluate the ability of the RtcB protein to modulate CasM's sequence-specific and collateral nuclease activity in *E. coli*. The method set forth herein is adapted from Smargon et al., *Molec. Cell* (2017) 65:618-630. Not all of the following steps are required for screening, nor must the order of the steps be as presented.

[0225] The expression plasmids constructed in Example 8 are individually and in combination transformed into BL21 (AI) *E. coli* cells from a commercial provider, such as Invitrogen (Carlsbad, Calif.). Transformed cells are grown

overnight at 37° C., with shaking, in lysogeny broth (LB) supplemented with 100 µg/mL carbenicillin, to select for cells that contain the CasM expression plasmid.

[0226] The following day, cells are diluted 1:100 and then grown at 37° C., with shaking, to an OD₆₀₀ of 2.0. The cells are then mixed with 4 mL of carbenicillin-containing top Agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride, 5 g/L agar) and poured onto LB-antibiotic base plates. The top agar also contains 0.2% arabinose to induce expression of the cash, rtcB and CRISPR array coding sequences. 10-fold serial dilutions of MS2 phage (ATCC 15597-B1, Manassas Va.) are made in LB and then spotted onto hardened top agar with a multi-channel pipette. Plaque formation is assessed after overnight incubation of the spotted plates at 37° C.

[0227] To assess whether the RtcB protein modulates CasM cleavage activity, the relative plaque formation is determined by comparing cells expressing CasM, RtcB, and the CRISPR array targeting MS2 phage; cells expressing CasM and the CRISPR array targeting MS2 phage; and cells expressing CasM and the CRISPR array not targeting MS2 phage; cells expressing CasM and RtcB only.

Example 10

Introduction of CasM RNP Complexes into Target Cells

[0228] This Example illustrates the design and delivery of CasM and crRNA ribonucleoprotein (RNP) complexes into human cells to enable mRNA cleavage of the human epidermal growth factor receptor (EGFR) gene and subsequent knockdown of EGFR gene expression.

[0229] A. Production of CasM Complexes and Transformation into Cells

[0230] Mature crRNAs (SEQ ID NOS:70-165) were designed to target the EGFR locus in the human genome. Each crRNA contained a 5' 36 nt repeat (SEQ ID NO:51) followed by a 30 nt spacer. crRNAs were designed to target 72 unique sequences complementary to the egfr mRNA within exons 1-3. Sequences were designed such that flanking sequences within 1 bp were not biased by any nucleotide. As negative controls, not predicted to induce cleavage, crRNAs were also designed to target (1) eight genomic sequences upstream of the predicted egfr mRNA; (2) eight sequences complementary to the vega mRNA exon 1; and (3) eight sequences identical to the egfr mRNA.

[0231] Double-stranded DNA (dsDNA) guide templates containing upstream T7 promoter sequences were created by annealing complementary oligonucleotides (Integrated DNA Technologies, Coralville, Iowa) at a final concentration of 10 µM in annealing buffer (30 mM HEPES, 300 mM KCL), then incubating at 95° C. for two minutes, and then slowly cooled to approximately 25° C., and incubated for an additional 20 minutes. Following annealing, guides were transcribed with T7 RNA polymerase HiScribe™ T7 High Yield RNA Synthesis Kit™ (New England Biolabs, Ipswich, Mass.) according to manufacturer's instructions. Next, samples were digested with RNase-free DNase-I (New England Biolabs, Ipswich, Mass.) according to manufacturer's instructions, then purified using RNAClean XP™ beads (Beckman Coulter, Indianapolis, Ind.).

[0232] For RNAClean XP™ bead purification, 30 µL of sample was combined with 155 µL of 100% isopropanol and 10 µL of 3 M sodium acetate and then mixed thoroughly.

Next, 50 µL of RNAClean XP™ beads were incubated on a magnet for three minutes to allow separation of the liquid and beads, and the supernatant was removed. Subsequently, the samples containing crRNA were added to the beads, mixed, incubated at approximately 25° C. for five minutes, then incubated on a magnet for three minutes. Finally, the supernatant was removed, the beads were washed once with 85% ethanol, dried, and then the crRNA was eluted in 20 µL of molecular biology grade water. crRNAs were quantified using ribogreen and then normalized to 1 µg/µL.

[0233] To assemble CasM RNPs, 120 pmols of each unique crRNA were added to a well then incubated at 95° C. for two minutes followed by 25° C. for approximately 10 minutes. Next, the denatured crRNA guides were combined with 20 pmol of CasM (SEQ ID NO:39) in RNP assembly buffer (20 mM HEPES; pH 7.4, 10 mM MgCl₂, 150 mM KCl, 5% glycerol) and then incubated at 37° C. for 10 minutes.

[0234] B. Transfection of CasM RNP Complexes into Eukaryotic Cells

[0235] HeLa cells (ATCC, Manassas, Va.) were cultured in suspension in DMEM medium supplemented with 10% FBS and 1x Antibiotic-Antimycotic Solution (Mediatech, Inc., Manassas, Va.) at 37° C., 5% CO₂ and 100% humidity. HeLa cells were transfected using the Nucleofector® 96-well Shuttle System (Lonza, Allendale, N.J.). Prior to nucleofection, 5 µL of the CasM:crRNA RNPs were assembled in individual wells of a 96-well plate. HeLa cells were transferred to a 50 ml conical centrifuge tube and centrifuged at 200×G for five minutes. The media was aspirated and the cell pellet was washed in calcium and magnesium-free PBS. The cells were centrifuged once more and resuspended in Nucleofector SF™ buffer (Lonza, Allendale, N.J.) at a concentration of 5×10⁶ cells/ml. 20 µL of this cell suspension was added to the CasM:crRNA RNPs in the 96 well plate, mixed, and then the entire volume was transferred to a 96-well Nucleocuvette™ Plate. The plate was then loaded into the Nucleofector 96-well Shuttle™ and cells were nucleofected using the 96-CN-114 Nucleofector™ program (Lonza, Allendale, N.J.). Immediately following nucleofection, 75 µL of complete DMEM medium was added to each well of the 96-well Nucleocuvette™ Plate. Half of the contents of each well were then transferred to a 96-well tissue culture plate containing 150 µL of complete DMEM medium. This procedure was then repeated in order to plate a duplicate for each well, one which would be used for lysis and genomic DNA analysis, and one for FACS analysis. The cells were cultured at 37° C., 5% CO₂ and 100% humidity for approximately 5 days.

[0236] C. FACS Analysis of CasM Mediated EGFR Knockdown

[0237] Fluorescence activated cell sorting (FACS) analysis was performed 5 days after nucleofection of HeLa cells with EGFR-targeting CasM2 RNPs. In brief, 2×10⁵-4×10⁵ cells/well were detached with TrypLE Express (Gibco), stained with 2 µL APC anti-human EGFR (Clone Y13, Sony Biotechnology) in 100 µL total volume and then analyzed using Intercity Flow Cytometer (Intercity, Albuquerque, N. Mex.). Results from these experiments are shown in Table 16.

TABLE 16

CasM Mediated EGFR knockdown			
Name	% EGFR negative cells	transcription	crRNA SEQ ID NO.
Untransfected reference cell	2%	n/a	—
Intergenic target-1	5%	Intergenic	SEQ ID NO: 70
Intergenic target-2	6%	Intergenic	SEQ ID NO: 71
Intergenic target-3	9%	Intergenic	SEQ ID NO: 72
Intergenic target-4	8%	Intergenic	SEQ ID NO: 73
Intergenic target-5	5%	Intergenic	SEQ ID NO: 74
Intergenic target-6	6%	Intergenic	SEQ ID NO: 75
Intergenic target-7	6%	Intergenic	SEQ ID NO: 76
Intergenic target-8	4%	Intergenic	SEQ ID NO: 77
Exon 1 target-1	6%	Exon 1/28	SEQ ID NO: 78
Exon 1 target-2	5%	Exon 1/28	SEQ ID NO: 79
Exon 1 target-3	5%	Exon 1/28	SEQ ID NO: 80
Exon 1 target-4	4%	Exon 1/28	SEQ ID NO: 81
Exon 1 target-5	6%	Exon 1/28	SEQ ID NO: 82
Exon 1 target-6	5%	Exon 1/28	SEQ ID NO: 83
Exon 1 target-7	6%	Exon 1/28	SEQ ID NO: 84
Exon 1 target-8	6%	Exon 1/28	SEQ ID NO: 85
Exon 1 target-9	4%	Exon 1/28	SEQ ID NO: 86
Exon 1 target-10	3%	Exon 1/28	SEQ ID NO: 87
Exon 1 target-11	3%	Exon 1/28	SEQ ID NO: 88
Exon 1 target-12	2%	Exon 1/28	SEQ ID NO: 89
Exon 1 target-13	4%	Exon 1/28	SEQ ID NO: 90
Exon 1 target-14	4%	Exon 1/28	SEQ ID NO: 91
Exon 1 target-15	5%	Exon 1/28	SEQ ID NO: 92
Exon 1 target-16	4%	Exon 1/28	SEQ ID NO: 93
Exon 1 target-17	6%	Exon 1/28	SEQ ID NO: 94
Exon 1 target-18	6%	Exon 1/28	SEQ ID NO: 95
Exon 1 target-19	6%	Exon 1/28	SEQ ID NO: 96
Exon 1 target-20	5%	Exon 1/28	SEQ ID NO: 97
Exon 1 target-21	5%	Exon 1/28	SEQ ID NO: 98
Exon 1 target-22	5%	Exon 1/28	SEQ ID NO: 99
Exon 1 target-23	6%	Exon 1/28	SEQ ID NO: 100
Exon 1 target-24	5%	Exon 1/28	SEQ ID NO: 101
Exon 2 target-1	6%	Exon 2/28	SEQ ID NO: 102
Exon 2 target-2	7%	Exon 2/28	SEQ ID NO: 103
Exon 2 target-3	11%	Exon 2/28	SEQ ID NO: 104
Exon 2 target-4	5%	Exon 2/28	SEQ ID NO: 105
Exon 2 target-5	6%	Exon 2/28	SEQ ID NO: 106
Exon 2 target-6	8%	Exon 2/28	SEQ ID NO: 107
Exon 2 target-7	11%	Exon 2/28	SEQ ID NO: 108
Exon 2 target-8	10%	Exon 2/28	SEQ ID NO: 109
Exon 2 target-9	13%	Exon 2/28	SEQ ID NO: 110
Exon 2 target-10	8%	Exon 2/28	SEQ ID NO: 111
Exon 2 target-11	10%	Exon 2/28	SEQ ID NO: 112
Exon 2 target-12	8%	Exon 2/28	SEQ ID NO: 113
Exon 2 target-13	13%	Exon 2/28	SEQ ID NO: 114
Exon 2 target-14	16%	Exon 2/28	SEQ ID NO: 115
Exon 2 target-15	19%	Exon 2/28	SEQ ID NO: 116
Exon2 target-16	11%	Exon 2/28	SEQ ID NO: 117
Exon 2 target-17	10%	Exon 2/28	SEQ ID NO: 118
Exon 2 target-18	19%	Exon 2/28	SEQ ID NO: 119
Exon 2 target-19	20%	Exon 2/28	SEQ ID NO: 120
Exon 2 target-20	25%	Exon 2/28	SEQ ID NO: 121
Exon 2 target-21	15%	Exon 2/28	SEQ ID NO: 122
Exon 2 target-22	17%	Exon 2/28	SEQ ID NO: 123
Exon 2 target-23	14%	Exon 2/28	SEQ ID NO: 124
Exon 2 target-24	12%	Exon 2/28	SEQ ID NO: 125
Exon 3 target-1	7%	Exon 3/28	SEQ ID NO: 126
Exon 3 target-2	7%	Exon 3/28	SEQ ID NO: 127

TABLE 16-continued

CasM Mediated EGFR knockdown			
Name	% EGFR negative cells	transcription	crRNA SEQ ID NO.
Exon 3 target-3	9%	Exon 3/28	SEQ ID NO: 128
Exon 3 target-4	9%	Exon 3/28	SEQ ID NO: 129
Exon 3 target-5	8%	Exon 3/28	SEQ ID NO: 130
Exon 3 target-6	11%	Exon 3/28	SEQ ID NO: 131
Exon 3 target-7	12%	Exon 3/28	SEQ ID NO: 132
Exon 3 target-8	12%	Exon 3/28	SEQ ID NO: 133
Exon 3 target-9	10%	Exon 3/28	SEQ ID NO: 134
Exon 3 target-10	9%	Exon 3/28	SEQ ID NO: 135
Exon 3 target-11	11%	Exon 3/28	SEQ ID NO: 136
Exon 3 target-12	40%	Exon 3/28	SEQ ID NO: 137
Exon 3 target-13	17%	Exon 3/28	SEQ ID NO: 138
Exon 3 target-14	15%	Exon 3/28	SEQ ID NO: 139
Exon 3 target-15	12%	Exon 3/28	SEQ ID NO: 140
Exon 3 target-16	21%	Exon 3/28	SEQ ID NO: 141
Exon 3 target-17	48%	Exon 3/28	SEQ ID NO: 142
Exon 3 target-18	41%	Exon 3/28	SEQ ID NO: 143
Exon 3 target-19	19%	Exon 3/28	SEQ ID NO: 144
Exon 3 target-20	9%	Exon 3/28	SEQ ID NO: 145
Exon 3 target-21	19%	Exon 3/28	SEQ ID NO: 146
Exon 3 target-22	8%	Exon 3/28	SEQ ID NO: 147
Exon 3 target-23	8%	Exon 3/28	SEQ ID NO: 148
Exon 3 target-24	6%	Exon 3/28	SEQ ID NO: 149
VEGFA target-1	6%	Exon 1/8	SEQ ID NO: 150
VEGFA target-2	8%	Exon 1/8	SEQ ID NO: 151
VEGFA target-3	8%	Exon 1/8	SEQ ID NO: 152
VEGFA target-4	7%	Exon 1/8	SEQ ID NO: 153
VEGFA target-5	8%	Exon 1/8	SEQ ID NO: 154
VEGFA target-6	7%	Exon 1/8	SEQ ID NO: 155
VEGFA target-7	6%	Exon 1/8	SEQ ID NO: 156
VEGFA target-8	8%	Exon 1/8	SEQ ID NO: 157
Nontargeting target-1	5%	Exon 1/28	SEQ ID NO: 158
Nontargeting target-2	7%	Exon 1/28	SEQ ID NO: 159
Nontargeting target-3	6%	Exon 1/28	SEQ ID NO: 160
Nontargeting target-4	6%	Exon 1/28	SEQ ID NO: 161
Nontargeting target-5	6%	Exon 1/28	SEQ ID NO: 162
Nontargeting target-6	6%	Exon 1/28	SEQ ID NO: 163
Nontargeting target-7	7%	Exon 1/28	SEQ ID NO: 164
Nontargeting target-8	4%	Exon 1/28	SEQ ID NO: 165

[0238] The data presented in Table 16 shows that CasM did not produce egfr knockdown when targeted to (1) sequences upstream of the predicted exon 1 start site using SEQ ID NOS:70-77; (2) an unrelated vascular endothelial growth factor A (vegfa) gene using SEQ ID NOS:150-157; or (3) the reverse complement of sequences contained in egfr exon 1 mRNA using SEQ ID NOS:158-165. Conversely, CasM enabled approximately 40% egfr knockdown when targeted to mRNA sequences contained in exon 2 using SEQ ID NOS:102-125, and exon 3 of egfr using SEQ ID NOS: 126-149.

[0239] Although preferred embodiments of the subject methods have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 165

<210> SEQ ID NO 1
<211> LENGTH: 2862
<212> TYPE: DNA

-continued

```

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Eubacterium
      siraeum, modified for expression in Escherichia coli

<400> SEQUENCE: 1

atgggaaaga agattcatgc gcgcgattta cgcgaaacac gcaaaacgga tcgcactgag      60
aaatttgcgg atcaaaacaa aaagcgcgag gccgagcgcg ctgttctaa aaaggacgcc      120
gcagtctcg ttaagagtgt atcgtccgtg tcttcaaaaa aggacaacgt cactaaaagc      180
atggcgaagg ccgctgggt aaagtctgtt tttgccgtg gtaacacggt atacatgaca      240
tcgttcggcc gcggcaacga cgctgtactg gagcaaaaga tcgtggatac atccatgaa      300
ccacttaaca tcgacgatcc agcatatcaa ttgaacgttg ttacaatgaa cggttatcc      360
gtcaccggcc accgcggaga gaccgttct gcagtaacgg acaaccctt acggcgttc      420
aatggccgca aaaaggacga acctgagca tcgggttcca ctgacatgct ttgtttaaa      480
cctacgttag agaagaagtt cttcggcaag gagtttgacg acaacatcca catccagttg      540
atttataaca ttttagatat tgagaagatc ttagcagttt attcaaccaa tgcaatttac      600
gctttaaca acatgagcgc cgacgaaaac atcgaaaatt cggattttt catgaaacgt      660
accacagacg aaaccttga cgacttggaa aagaaaaaaag aatctactaa ctcacgcgaa      720
aaggcagact tcgacgcgtt tgaaaaattt atggaaaact accgttgc gtacttcgc      780
gtatgtttct atgtcaataa aaaaaaccct aagggaaagg ctaagaatgt tctgcgtgaa      840
gataaggagc tttactcggt cttaactctt atcggtaaac tgcgcattg gtgcgtacat      900
agcgaggagg gacgtgcaga gttctggctg tataagttt acgagttttt agacgatccc      960
aaaaatgtat tggacgtcgt gtacaaccgt cccgtggaaag aatcaacaa ccgtttatt      1020
gagaataaca aagttaatat ccaattctg gggagcgtgt aaaaaaacac agacatcgct      1080
gaacttgtc gctcgttta cgaattctt attacaaaa aataaaaaa tatggctt      1140
tctattaaga aactcgtga atcaatgtt gaaaggtaaag gttacgcaga caaggatata      1200
gactccgtcc gtaataagtt gtaccaaattt acagacttca ttctgtatac gggatacatc      1260
aacgaagact cagatcgtgc agacgtatctg gtcaataacc tgcgccttc tctgaaggag      1320
gatgataaga cgactgtata ctgtaaagag gccgactatt tggaaagaa gtatcgcgaa      1380
tcgatccgtg aggttgcggta tgcactggat ggtgataaca tcaagaagtt gagtaagtcg      1440
aacatcgaga tccaaagagga taaacttcgt aagtgcattca ttagttatgc agactccgtt      1500
tcagagttca caaaactgtat ctactcggtt acccgcttcc tgagcggaaa ggaaattat      1560
gacctggtaa ctactttat caataaaattt gataacatcc gctttttct tgagattatg      1620
gacgagctgg gattagatcg tacgtttacc gccgaatatt cgttcttga aggctcaacg      1680
aaataacttgg cggagcttgtt agagttaaat tctttgtaa aatcttgcctc ttttgatatt      1740
aacgccaagg gcacaatgtt tgcgcacgcc tttagacattt tggggattga atcggacaag      1800
actgaagagg atattgaaaa gatgattgtt aatatccttc agattgtatc gaatggcgac      1860
aagaaaactta agaaaaataa tggcctgcgtt aacttcattt caagtaacgt tattgacagt      1920
aaccgtttca aatacttagt acgctacggg aaccctaaaaa aaatccgcga aacagctaag      1980
tgcaaaaccgg ctgttcgtt cgtgttgaac gagatccccg acgcacagat cgagcgctat      2040
tacgaggcat gctgtccaaa gaacacagcc ctttgctcag cgaacaagcg tcgcgagaag      2100

```

-continued

ttagctgaca	tgattgccga	gattaagtgc	gagaacttct	ctgacgctgg	aaattatcaa	2160
aaagctaacg	ttacctcgcg	cacatcagag	gcccggaaatca	aacgtaaaaa	ccaggcgatt	2220
attcgcgttgc	atttgacgggt	catgtacatt	atgctgaaga	acttagtcaa	cgtgaacgct	2280
cgttaacgtga	tcgcatttca	ctgtgtggag	cgtgatacta	agttgtatgc	cgaatctggaa	2340
ttggagggttg	ggaacattga	aaagaataaa	actaatctta	ccatggccgt	aatgggagtt	2400
aagcttggaga	atggtatcat	caagactgag	tttgataat	cttttgcgga	aaacgcagca	2460
aatcgttacc	ttcgtaacgc	acgctggtat	aaacttatct	tagacaattt	aaaaaagtca	2520
gaacgcgcgg	tagtaaacga	atttcgttaac	acagttatgtc	atttaaacgc	catccgcaac	2580
attaacatta	acatcaagga	gattaaggag	gtagaaaattt	attttgcctt	gtaccactat	2640
ttgatccaaa	aacatttggaa	gaaccgtttc	gccgacaaaaa	aagttgaacgc	cgatacgggt	2700
gactttttttt	ccaaatttggaa	agagcataag	acgtactgtt	aggactttgt	aaaagcatac	2760
tgtacgcccgt	ttggatataa	tttagtacgt	tataagaact	tgactattga	cggaactttc	2820
gataaaaaact	accctggggaa	ggatgattct	gatgaacaga	aa		2862

<210> SEQ ID NO 2
<211> LENGTH: 2757
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus sp., isolate 2789STDY5834971, modified for expression in Escherichia coli

<400> SEQUENCE: 2

atggcaaaga	aaaataaaaat	gaagccgcgc	gagttacgcg	aggcccagaa	gaaagctcg	60
caattaaaag	cggccgagat	caacaataac	gcagccccag	caattgcgc	aatgcgcgc	120
gccgaagtga	ttgcgcggc	tgccagagaag	aagaagagct	cagtcaaggc	agcaggatg	180
aagagcatcc	ttgttagcga	gaacaagatg	tacattacat	cttttggaa	aggaaactca	240
gggttattgg	aatacgaggt	tgataacaac	gattacaatc	agacgcagg	atcatccaag	300
gacaacagca	acatccaact	gggtggcgtc	aatgaggctc	acattacttt	ttcaagcaag	360
cacggctttg	aaagtggcgt	ggaaatttaac	acttctaattc	cgacacacccg	ttcaggagaa	420
agttcccttg	ttcggtggcga	tatgttaggg	cttaagtca	aaactggaaa	gcgccttc	480
ggtaagacct	tcgatgataa	cattcacatt	caacttatct	acaacatcct	tgatattgaa	540
aagatccttg	cagtgtacgt	tacgaacatc	gtctacgctc	tgaataat	tttaggtgtc	600
aaggggtctg	aatccatga	tgacttcatt	ggttacttgt	cgacaaataa	tatctacgt	660
gtcttcattt	atccagataa	tagttcccttg	agcgacgaca	agaaagcaaa	cgtacgtaaa	720
agtcttagta	aatttatgc	gttgtaaaa	actaaacgtc	tgggtatatt	cggtttagag	780
gaaccaaaga	ccaaagacaa	ccgtgtaaac	caggcgtata	agaagcgtgt	gtatcacatg	840
cttgcatttgc	tcgggcaaat	tcgtcaatgc	gtatccatg	acaaaagcgg	tgccaaacgt	900
tttgatcttt	attctttcat	taacaatatt	gatccagat	accgtgacac	gcctgattat	960
tttgtttagaa	agcgcctgaa	gtcaattaac	aaagacttta	ttgaagacaa	caaagtaaac	1020
atcagccttt	taattgat	gatgaagggt	tacgaggcgg	acgatatacat	tcgcctgtac	1080
tacgacttca	ttgttattaa	atctcagaaa	aacctggggt	tctctattaa	gaagttacgt	1140

-continued

gagaagatgc tggacgagta tggtttccgt ttcaaagata aacaatacga ttctgttcgt	1200
tccaaatgt ataaattgtat ggatttttg ctttttgtat actattaccg caatgtatatt	1260
gctgcggggg aatctctggt acgtaaactg cggtttcgta tgacagacga tgaaaaggag	1320
ggcatttatg cggacaaagc cgcttaatttggggaaat ttctgtatgtat ctttggaaat	1380
atcgccggacc acatgaatgg cgatgttatt aaggagttgg gaaaagctga catggatttc	1440
gacgaaaaga tcttggattc tgagaagaaa aacgcttccg acctgtgtat tttttcaaaa	1500
atgatttata tgctgacata tttcttagat gggaaagaga ttaacgactt gctgacgact	1560
ctgatttcaaa aatttgcacaa tatcaaagag tttttgaaaa ttatgttgc ttctgtcagtc	1620
gatgttagt gtgaacttac agctgggtac aagctgtca atgacagtca acgtatcacc	1680
aacgaattat ttatcgtaa aaatattgcc tccatgtgtat agccagccgc aagtgc当地	1740
ctgacaatgt tccgcgatgc actgacgatt ctgggaatttgc acgataagat tacggatgac	1800
cgtatccatcg gaatcttgc gcttaaagag aaggcaagg gcattcatgg acttcgttaac	1860
ttccatccatca acaacgtgtat cgagatgtac cggtttgtt accttatcaa atatgc当地	1920
gcacaaaaga tccgcgaaat ggcgaaaaac gagaaggctg taatgttgc attaggttgg	1980
attccagata cgcaatttgc ggcgttatttgc aagtcgttgc tagatgtccc ggatatgc当地	2040
agctcatttgc gatgttgc ttccatgtgtat ggcgcgtatgtat ttaagaatata cagtttgc当地	2100
gatttcaaga acgttgc当地 acaagcgaaa ggacgc当地 acgtcgcaaa agacgc当地	2160
aaggccgtca ttgggttgc cttaacgtgtat atgtacttac ttgtcaaaaa cctggatata	2220
gttaacgc当地 gctatgtcat cgccatccat tgcgttgc gtttttgc ttcttataag	2280
gagatttttgc ctgaacttgc gtc当地 aacgc当地 attaccgc当地 ttatctcag	2340
actctgtgtatc aactgtgtatc taatgttccat aattttgttgc tgaagaagaa tgagcgc当地	2400
cgttaatgttgc ttgaagtc当地 catcaataat gcagacatgc cgtatgtatc taaatatcgc	2460
aactgtatgc ctcaacttgc tgcgtccgt gaatttttgc agtacatgttgc tgc当地	2520
accgttgc当地 ttatatttgc tatttaccat tatgtatgc aacgc当地 taaaaggc当地	2580
gaaaacgc当地 ccaaggc当地 ggaaaaatc aaatgc当地 acgttgc当地 taagaatc当地	2640
ggctatacaa aagacttgc当地 aaaaggc当地 aactcacctt tcggatataa catccgc当地	2700
tttaaaaatc ttcaatttgc gcaacttttgc gatgtatgc当地 agtacatgttgc当地	2757

<210> SEQ ID NO 3
 <211> LENGTH: 2754
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic: CasM DNA sequence
 from Ruminococcus bicirculans,
 modified for expression in Escherichia coli

<400> SEQUENCE: 3

atggcgaaaa agaataaaaat gaaacctcgc gaattgc当地 aggcacaaaaa gaaagcgc当地	60
caattgttgc当地 cagcggagat caacaataac gcagttcccg cc当地tgc当地tgc tatgc当地	120
gctgaggccg ctgccccccgc agcggagaaa aagaagtcat cggtaaagc ggc当地ggatg	180
aagtcaatct tagtctccgc当地 gaacaagatc tacatccatca gttttggaaa aggttaactcg	240
gcgggttgc当地 agtacgaggtt agacaataat gactataaca aaactcagtt atcctcg当地	300

-continued

gataatagca atattgagtt gtgtgatgtg gggaggta atatcacgtt cagctctcg	360
cgtggcttg aatcggagt cgagattaat acgagtaacc caacccaccc ctccggagag	420
tgcgtcgtag tccgtggga tatgtggc ttgaaaacg agttggaaaa acgtttttt	480
ggcaagaatt tcgacgataa tatccatatt caacttattt acaacatctt ggacatcgag	540
aagatccttg ctgtgtatgt tacgaacatt gtttacgccc tgaataatat gcttggcgaa	600
ggggatgaat ctaactacga ctttatgggg tattttagac cattcaacac atataaagtc	660
tttacgaatc cgaatggttc aacgctgtct gatgacaaga aagagaacat tcgcaaatca	720
ttatcgaaat ttaatgcttt gttgaaaacg aagcgcttag gttatttccg gtttagaggag	780
cctaaaacaa aggacacgcg cgcatcgag gcttacaaga aacgcgtata tcacatgctg	840
gctatcgttg ggcaaattccg tcagtgcgta tttcatgata agagcggggc caagcgttc	900
gacccttattt cattttatcaa taacattgtt ccagaatatc gtgaaaactct ggattactg	960
gtcgacgaac gctttgacag tattaataaa ggatttatcc aaggtataaa agtaaacatc	1020
agcttactga tcgatatgtat gaagggttac gaggcggatg acatcatccg tctttactac	1080
gatttcatttgc tccttaatc gcagaaaaac ctgggcttca gtatcaaaaa gttacgcgaa	1140
aagatgttgg atgagttatgg ctttcgttcc aaagataacg aatacgatag cgttcgcagc	1200
aagatgtata aattatggaa ttctttatca ttctgcaattt actaccgcaaa cgacatttgc	1260
gcggggcgaat ctcttgcgtcc caagctgcgc ttttagtgcg ccgtatgtgcg gaaggaggg	1320
atctacgcag atgaggctgc aaaactgtgg ggcaaatttc gtaacgactt tgagaacatc	1380
gcccggaccaca tgaacggtgcg cgtcattaa gagttggggaa aagcagatgg gactttgtat	1440
gaaaagatcc ttgattccga aaagaaaaat gcgtcgatgc tggtgtatgg tagtaaaatg	1500
atttacatgc ttacgtatgg tctggacggaa aaagaaatca acgacttact tactacattt	1560
atttcaagt ttgataacat taaggagttt ttaaaaatca tgaaaagcag tgcaaggatc	1620
gttgaatgtg aacttacagc aggttataaa ttatgtatgc acagccaaacg catcacaat	1680
gaattgttca tcgtgaagaa tatcgcttct atgcgcaaac ccgtcgatcc ggcgaagctg	1740
acaatgtttc gcgacgcttt aacaatccgtt gggatcgacg ataagatcac tgatgtatgc	1800
atttccgaaa tcttaaaattt aaaggagaaa ggaaaaggta tccatggctt acgcaattt	1860
atcactaata atgtatgtt aagtagccgc tttgtgtacc ttatcaagtg cgcacacgc	1920
caaaaaatcc gcgtggatgc caaaaacgag aaagtcgttgc tggtgtatgg ggggtggatt	1980
cccgacacac aaatcgacgc ctactacaaa agttgtgtgg aattccggaa catgaactcg	2040
agtctgggtt ttaagcgatgc tgaattggcc cgtatgtatca agaatatcag ttttgcgtat	2100
ttcaagaatgtg tgaaacagca ggccaaagggg cgtgagaacgc tgccaaaggaa acgcgttac	2160
gctgtatcg gtttatatct gaccgtgtatc tacttgcgttgc tgaagaattt ggtgaacgtt	2220
aacgcgcgtt acgttattgc cattcattgc tttagaacgcg acctttggact gtataaggag	2280
attattccctg aatttagccatc caaaaacgtt aaaaacgtt atcgatctt gagccaaacc	2340
ctttgcgttac tttgtgtatca aagccaaac ttgtttttaa aaaaaatgcg gctgttacgc	2400
aaatgcgtgg aggttgcgtat taataatgtt gattcctcgat tgacccgcaaa ataccgttac	2460
tgttattgcgttcc atttgcgtatc agtccgcgcgat ttgttgcgttgc acattggaga tatttgcact	2520
gtggacagttt acttcgtat ttaccattat gtaatgcac gctgcattac aaagcgccgag	2580

-continued

aacgacacta	agcaggagga	aaaaatcaag	tacgaggatg	atctgctgaa	aatcatggc	2640
tacaccaagg	actttgttaa	ggccttgaac	tctccgttc	ggtataacat	tccccgctc	2700
aaaaatctga	gtattgagca	gttgtttgc	cgtaatgagt	atcttacaga	gaag	2754
<210>	SEQ ID NO 4					
<211>	LENGTH:	2766				
<212>	TYPE:	DNA				
<213>	ORGANISM:	Artificial Sequence				
<220>	FEATURE:					
<223>	OTHER INFORMATION:	Synthetic: CasM DNA sequence from Ruminococcus sp., isolate 2789STDY5608892, modified for expression in Escherichia coli				
<400>	SEQUENCE:	4				
atggccaaaa	agaacaaaat	gaagccccgc	gaacttcgtg	aggcccaaaa	gaaagctgc	60
caattaaaag	cagccgagat	caacaacaac	gcagctccgg	ccattgcagc	aatgcctgct	120
gcagaagtga	ttgcgccagt	cgccgaaaag	aagaaatcca	gtgttaaagc	tgcaggtatg	180
aagtctattt	tggtttcgg	gaacaagatg	tatatcacaa	gttcgggaa	aggtaatagt	240
gctgttcttg	agtatgaagt	agataacaac	gactataata	aaacccaact	tagttctaag	300
gataactcta	atattgaatt	gggggacgtt	aatgaggtaa	atatcacgtt	ctcatcgaag	360
catggctttg	gttccggggt	ggaatcaat	acctctaatac	ccactcatcg	ttcgggtgaa	420
tcctcccccag	tccgtggta	tatgttgggg	cttaaatcgg	agtttagagaa	acgcttctt	480
ggtaaaacct	ttgatgataa	tattcatatt	caattgattt	ataacatttt	ggatatcgag	540
aaagattttgg	ctgtatacgt	tacaaatatac	gtgtatgcac	ttaataat	gttgggtatt	600
aaagattctg	aatcgatgat	tgatttcatg	ggctatttga	gcmcacgcaa	tacctatgaa	660
gtcttcactc	atccgtataa	aagcaactta	agtgataagg	ttaaagggaa	cattaagaag	720
agtttatcaa	agttcaatga	cttgtttaag	accaagcgcc	ttgggtactt	cggtctttag	780
gaaccgaaga	ccaaagatac	ccgcgttct	gaggcgtata	agaagcgcgt	ctaccacatg	840
cttgcataatc	taggtcaaat	ccgtcaagtgt	gtgtttcagc	acaaatcagg	agcgaacacgt	900
ttcgatttgt	actccattcat	taataacatc	gaccagagt	atcgcgacac	tcttgactac	960
ttagttgagg	aacgtttgaa	gtcaattaat	aaggattca	ttgagggaaa	taaagtaaac	1020
attagccctc	ttatcgacat	gatgaaggga	tacgaggccg	acgatattat	tcgcctgtat	1080
tatgatttta	ttgtgttgaa	atcacaaaag	aatttgggg	ttagcattaa	aaaattgcgc	1140
gagaagatgt	tggaggagta	tgggttcgc	ttaaggata	aacagtatga	ctcagtcgc	1200
tcaaaaatgt	ataagttaat	ggacttcctg	ctttttgtat	attattacgg	taatgacgtc	1260
gcgcgggtg	aaggccctgg	tcgttaatttgc	cgcttctcaa	tgactgacga	tgagaaggag	1320
ggaatttatg	ctgtatgaggc	tgcaagtttgc	tggggaaagt	ttcgttaacga	cttcgaaaat	1380
atcgccgacc	acatgaatgg	agatgttac	aaggagcttg	gcaaggcgga	tatggattt	1440
gatgaaaaga	tccttgacag	cgaaaagaag	aatgcctccg	atttgcgtt	cttttcgaaa	1500
atgatctaca	tgcttaccta	tttcctggac	ggcaaagaga	tcaacgatct	tttgaccacc	1560
cttatttcta	agtgcataa	tatcaaagag	ttttgttttt	tcatgaagag	ttcggcggtc	1620
gatgttgaat	gtgaattaac	ggccgggtat	aaatttatttta	acgactccca	acgttattacg	1680
aatgaattat	ttatcgtaaa	aaacatcgct	tctatgcgc	aaccagcagc	gtccgcacaa	1740

-continued

cttacgatgt ttctgtacgc cttaccatt ttggaaatcg acgataacat cacagatgtat	1800
cgcatattctg agatcttggaa gcttaaggaa aaggccaagg gcatccatgg tttacgtaat	1860
tttatcacaa acaacgtgtat cgagtcgtact cggtttgtct atctgtcaa gtatgcaaacc	1920
gcccggaaaa ttctgtgaat ggccaaaaat gagaaagttag taatgtttgt tttgggtgg	1980
atccctgaca cccagattga gcgttactac aagtctgttg tagaattccc tgacatgaat	2040
agcagcttag aagctaaacgc ctctgaactt gcgcgtatg taaaaatata tcgcgtcgat	2100
gacttcaaga acgtttaaca acaggccaaa ggccgtgaga atgttgctaa agaacgcgcg	2160
aaggctgtta ttggattata cttctactgtat atgtatctgt tagtggaaaaa cttgtgttgc	2220
gtcaacgcgc gctacgtcat tgcgatccat tggttgagc gtgactttgg gttatacaag	2280
gagatcatcc cagaactggc ctcaaaaaac ttaaaaaatgt actaccgtat tttgagtcgt	2340
acccctgtcg actgtgtcgta tgaccgttaac gaatccctgca acttgttctt gaagaagaat	2400
aaacgtttgc gcaaatgtgtat cgaggtagat atcaacaatg cagacagctc tatgacgcgt	2460
aagtaccgtta actgttattgc tcacttaacc gtatgttgc aacttaaaga atacattgg	2520
gacattcgtta cagttgtatcg ctacttcaactt atttatcaact atgttatgca ggcgtgtatc	2580
actaagcgtg gggatgatac gaagcaagaa gagaaaattttt agtacgaaga tgacctgttg	2640
aaaaaccacg ggtacactaa ggactttgtc aaagctctga attccccgtt cgggtacaat	2700
atccctcgat ttaagaatct gaggattgaa cagttatgtt accgcaacga atacattacg	2760
gagaag	2766

```

<210> SEQ_ID NO 5
<211> LENGTH: 2766
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
sp. CAG:57, modified for expression in Escherichia coli

```

<400> SEQUENCE: 5	
atggctaaaa agaataaaaat gaaacctcgca gagtttgcgc aagcccggaa aaaagctcgat	60
cagttaaagg cagccggaaat taataataat gcagcaccccg ccattcgccgc gatccccgc	120
gctgaagttt tgcggccctgt tgcgtggaaat aagaaatccca gctgtggaaat ggcgggtatg	180
aagtccatatt tggtcagcgat gaataaaaatg tacattacgt cggttggaa aggcaactcc	240
gctgtcccttgc atgtatgtat agacaacaat gactacaaca aaactcaact gtcaagcaaa	300
gacaacacgtt acatcgaaactt gggggacgtgtt aatggggatcgat atatcacgtt ttcatcaaaa	360
catggggatcgat gaaatcgatcgat acaagcaatcccgatccatcgatccatcgatccatcgat	420
tgcgtcccttgc tgcgtggaaat catgttgggtt cttaagtccatcgatccatcgatccatcgat	480
ggcaagacat tgcgtatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgat	540
aagatggggatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgat	600
aaggactcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgat	660
gtcttcactccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgat	720
tctttgtcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgat	780
gaaccgaaatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgatccatcgat	840

-continued

ctggcaatcg tgggccaaat ccgtcagtgt gttttcatg acaaagggtt agctaaacgc	900
tttgatttgt acagcttcat taataacatt gatcctgaat atcgcgacac tttggattat	960
ttagtagaaag aacgccttaa atctattaat aaagacttta ttgaaggaa taaggtaac	1020
atcagcttac tgatcgacat gatgaagggt tacgaggctg acgacattat ccgcttgat	1080
atgatttca ttgtattaaa atctcagaaa aacctggat tcagtattaa gaaattacgc	1140
gagaaaaatgc ttgaggagta cggattccgt ttcaggata aacaatatga ttctgtgcgt	1200
agtaaatgt acaaacttat ggactttta ttgttctgtt actattaccg taatgacgtt	1260
geccgcaggcg aagccttggt acgtaagttt cgttcagca tgacagatga cgaaaaggag	1320
ggcatttacg cggatgaaac agcgaagctg tggggtaaat tccgcaacga ttttggaaat	1380
attgctgacc acatgaatgg tcatgttatac aaagaactgg gaaaagccga tatggattt	1440
gacgagaaga tcttggacag tgaaaaaaag aatgccagcg atctttata tttctccaaa	1500
atgatctaca tgcttactta ttcttgc gggaaagaga ttaatgatct gctgaccacg	1560
ctgattagta agttcgacaa cattaaggag tttttaaga tcatgaaatc gtccgctgtt	1620
gacgtagaat gcgagtttgc ggcagggttac aaactgttca acgatagtca acgcattacc	1680
aatgaacttt tcategtcaa aaacatttgc tccatgcgcg agcccgcggc tagcgctaaa	1740
ttaacgatgt tccgtgacgc cttgacgatt ttaggcattcg acgacaacat cacggacat	1800
cgcatttcgg aaatccttaa acttaaggaa aaggggaaag gtatccatgg tcttgcgaaat	1860
tttatcacta acaatgtaat tgaatcatca cgcttcgtttt acttaatcaa atacgcaat	1920
gctcaaaaga ttctgtgaatg agccaaggat gaaaagggtt tcatgtttgt cctgggggg	1980
atccagaca cccaaattga acgttattac aagtcttgc tggaaattccc cgatatgaat	2040
agctccttgg aggccaaacg ctctgagttt gcccgcattt ttaagaacat ttccctcgac	2100
gattttaaa atgtcaaaaca acaggcaaaa ggcgcgaga atgttagccaa ggagcgtgcc	2160
aaggcgttac tcggattgtt tcttactgtc atgtattttgc ttgttaagaa tcttgcgttac	2220
gttaacgcgc gctatgtat cgttattcat tgcttagaac gcgactttgg cctttataag	2280
gagatttttc ccgagtttgc atccaaaat ctttggaaacg acttccgtat tttgtcaca	2340
accttatgcg agttatgcg tgaccgcac gacttccatc atctgtttct taaaaaaaac	2400
aaacgttcc gcaaatgcgtt ggaagtggac atcaacaacg ccgacagtag tatgactcgt	2460
aagtatcgta actgtattgc gcacttgcact gtagtgcgcg agttgaagga gtatattggg	2520
gatatccgc cctggatttcc atacttgcgtt atcttaccact acgtcatgc acgttgcattc	2580
acgaaacgtt gagacgcacac caaacaagag gaaaagatgtt aatgtatgcgatc cgaccttttgc	2640
aagaaccacg gctacaccaa agatttgtt aaggcttgc atagtcctt cgggtataac	2700
atccccgtt tcaaaaaactt gagcattgaa cagctgttgc accgcaatga atacttgaca	2760
gaaaaag	2766

```

<210> SEQ ID NO 6
<211> LENGTH: 2799
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
      flavefaciens FD-1, modified for expression in Escherichia coli

```

<400> SEQUENCE: 6

-continued

atgaaaaaaaaaa	aatgtctct	gcgtgaaaag	cgtgaggcgg	agaagcaagc	aaagaaagcc	60
gcttattccg	ctgcttagtaa	gaataactgac	agcaaaccgg	cagagaagaa	ggcgaaaca	120
ccaaagcccg	cagaatttat	ctcgataac	tcgcgcaata	aaactgctgt	taaagccgc	180
ggcttgaat	caactatcat	cagtgggat	aaattataca	tgacgtcatt	tggtaaggg	240
aatgccgccc	tgatcgaaca	gaagattgat	attaatgact	actcttttc	tgccatgaag	300
gatacccccta	gcttagaggt	tgataaggcc	gagagcaagg	agatctctt	ttccctctcac	360
catcccttcg	taaagaatga	caaattgacc	acttacaacc	ccctgtacgg	cgccaaggac	420
aatccggaaa	agccagtggg	acgtgacatg	ctgggggtga	aagacaattt	ggaggaacgt	480
tatTTggat	gcactttcaa	tgataatctg	cacatccaga	tcatctacaa	tatcttagac	540
atcgagaaaa	tcctggctgt	tcatagcgca	aatatcacca	ccgcactgga	tcacatggta	600
gacgaggatg	acgaaaaata	cttgaactct	gactacattt	gttacatgaa	caccattaat	660
acgtacgacg	tatTTatgg	cccgtaaaag	aactcttctt	tgtcgccgaa	agatcgcaag	720
aacatcgaca	actcccgcc	caagttttag	aagttattgt	caacgaacg	ttttaggatac	780
tttgggggg	actatgatgc	gaatggcaag	gataagaaga	agaacgagga	gattaagaag	840
cgtctgtacc	atcttaccgc	gtttcggggt	cagttcgtc	agtggccctt	tcacagcgct	900
ggcaattatc	cacgtacatg	gctgtacaaa	cttgatagtt	tggacaaga	ataccttgat	960
acacttgatc	actatTCGA	taaacgcctc	aatgacatta	atgacgattt	cgttacaaag	1020
aacgcgacga	atTTatata	tcttaaggaa	gttttccgg	aggcgaactt	taaagatatc	1080
gcagatcttt	attacgactt	catcgtaatc	aaatcccaca	aaaatatggg	tttctctatt	1140
aaaaaaattgc	gtgaaaaaat	gttagagtgt	gatggtgcgg	atcgcatcaa	agaacaagat	1200
atggacagcg	tacgttcaaa	gctgtataaa	cttattgact	tttgcatttt	caaatattac	1260
catgagttcc	cggaactgtc	tgagaagaat	gttgatatct	tacgtgctc	cgtctccgac	1320
acgaagaaag	ataatcttta	tagcgcacg	gccgcgcgtc	tgtggagtat	cttcaaggag	1380
aagttctgg	gtttctgtga	caaaaattgtc	gtatgggtga	ctggtaaca	tgaaaaagat	1440
atcaacttcgg	taatcgataa	agacgcgtat	cgcaaccgt	gcaatgtcag	ttatTTtcg	1500
aaactgtatgt	atgcgtatgt	cttttcctt	gatggtaagg	aaattaacga	tttattgaca	1560
accctgatta	ataaattcga	taatatcgca	aatcagatca	aaacggcaaa	ggaacttgg	1620
attaacacag	cttcgtaaa	gaattatgac	tttttaacc	actcggagaa	gtatgtcgac	1680
gaactgaata	ttgtgaaaaa	catecgctgc	atgaaaaaggc	ctagtagcaa	cgctaaaaaa	1740
gtatgttacc	acgatgcatt	gacgatctt	gggattcctg	aagatatgga	tgagaaagcc	1800
tttagatgagg	agctggactt	gattctggaa	aaaaagaccg	atccagtaac	cgggaaagcct	1860
ttgaaaaggga	aaaaccgc	tcgcaactt	atcgctaaca	atgtatcg	aaactctcgc	1920
ttcatctatt	tgattaagtt	ttgcaatccg	aaaaacgtac	gtaagattgt	taataacacc	1980
aaagttacag	agtttgcatt	gaagcgcac	ccagatgcgc	agatcgacg	ctattacaag	2040
tcttgactg	actcgaaat	gaacccccc	acggaaaaga	aaattacgga	gttagccgg	2100
aaacttaagg	acatgaattt	tggaaacttc	cgcaacgtgc	gtcaaagtgc	aaaggagaac	2160
atggaaaagg	agcgTTTaa	agcagtgatt	ggtttgcacc	ttaccgtat	ctatcgctt	2220
gtaaaaaaatc	tggttgcatt	taattccgc	tacatcatgg	cgtttcatc	gctggagcgc	2280

-continued

gacagtcagt tatataatgt ctcggtcac aacgactacc tggcctaac cgatacgta	2340
gtaaaaggagg gagataattc ccgtccgt tacttagcgg ggaataaacg cttgegtgac	2400
tgtgtgaaac aggataattga taatgctaag aaatggctcg tcagtgataa gtacaactct	2460
atcacaaaat accgtaataa cgtagcacat ttaactgcag tacgtaattg cgccgaattt	2520
atcggtgaca ttactaagat cgactcgtat tttgcattat atcactacct tattcagcgt	2580
caactggcta agggtttggg tcacgagcgt tcgggatttg accgcaacta tccgcagttat	2640
gttccacttt ttaagtggca tacttacgtg aaagacgtgg ttaaagcctt aaatgctccc	2700
ttcggatatac acatcccacg cttaagaat ttgtctattt atgctttt tgatcgcaat	2760
gagatcaaaa agaatgacgg agagaagaag tctgtatgt	2799

```

<210> SEQ_ID NO 7
<211> LENGTH: 2832
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
albus strain KH2T6, modified for expression in Escherichia coli

```

<400> SEQUENCE: 7	
atggcaaaga aatccaaggg gatgtcgta cgtgagaaac gcgaatttggg aaaacagaag	60
cgcattcaaa aggctgctgt taactccgtc aacgacactc ctgaaaagac agaaggaggct	120
aacgtggtat cagtgaatgt gcgcacttct gccgaaaaca agcactccaa aaagtcagcg	180
gccaaggctt tggggctgaa atctggcttg gtaatttggg atgagctgtt tctgacatcg	240
ttcggtcgca gcaacgaagc caagttggaa aagaaaatct caggtgatac ggttgagaaa	300
tttaggtatcg ggcgttttga ggttagcttag cgtgacgagt cgacgctgac gcttggaaagt	360
ggacgcatta aggacaagac ggcgcgtcca aaggacccac gtcacattac ggttgataca	420
caaggtaaat tcaaagagga tatgtgggt attcgacgatc tgtagaaaa aaagatttt	480
ggaaagacct ttgacgataa catccatgtt caactggcat acaacattct tgatgtcgag	540
aaaattatgg cacagtatgt cagtatatt gtttatgtc tgcacaacac ggacaagacg	600
gagcgtaatg ataacctgtat gggttacatg tcaatccgtc acacatacaa gacgttctgt	660
gatacttcaa acttgcctga tgatactaaa caaaaatgtt aaaacaaaaa acgtgaattt	720
gataaaatca ttaagagtgg ccgtctgggc tatttcgggg aagctttat ggtaaatagc	780
ggcaactcta caaaaactgcg cccggaaaaa gagatctatc atattttgc gctgtatggcg	840
tcgttacgcc aaagttactt tcatgggtat gtcaaagata ccgattacca agggaccact	900
tggcgtata cactggagga caaaactgaag gggccctctc acgagttccg cgagacgatt	960
gacaaaatct ttgacgaggg atttccaaa atctcgaaag atttcgccaa aatgaacaag	1020
gtgaacctgc aaattttggg gcaaatgtatc ggggagttgt acgggtccat tgagcgccaa	1080
aacttaactt gtgactacta cgatttcattc cagttaaaga aacataagta tcttggcttt	1140
agcattaaac gtttacgcga gacgtatgtt gagactactc ccgcagatgt ctataaggca	1200
gagtgttaca actctgagcg ccagaaactg tacaagtttgc tgcactttt aatctacgac	1260
ctttattaca atcgtaagcc cgacgtatc gaagagatcg tcgataagct gcgtaatct	1320
gtgaatgtatg aagaaaaaga gtctatattac tcagtagagg ctaagttgt ctatgaaagc	1380

-continued

cttccaaag	tccttgacaa	gagttgaag	aatagtgtt	ctggggaaac	cattaaagac	1440
cttcagaaac	gttatgtga	tgaacagct	aaccgtattt	gggacatctc	gcaacattca	1500
atcagtgcca	acgtcaattt	cttctgtaaa	ttaatttaca	tcatgactct	tatgtggac	1560
ggaaaagaaa	tcaatgtct	gttgacaacg	ctgggtaaca	aattcgataa	cattgccagt	1620
ttcattgtat	tcatggatga	gttaggatta	gagcaactcat	tcaactgataa	ctataagatg	1680
ttcgctgatt	otaaagctat	ttgtctggat	ttgcaattta	tcaatttatt	tgcccgatg	1740
tcgaagatcg	atgacgaaaa	gtcgaaacgt	caacttttc	gtgacgcgt	ggtttattta	1800
gatattggta	ataaggacga	gacatggatt	aataactact	tagattccga	tatcttaag	1860
ctggacaagg	aaggtaataa	gttaaaggga	gcccgcattt	attttcgcaa	ctttatcgca	1920
aataacgtga	ttaagtcttc	acgcttcaa	tatttatgt	agtattcgag	tgccggatggc	1980
atgatttaat	taaagacaaa	tgagaagctt	attgggttcg	ttctggataa	gttaccagag	2040
acgcaaatcg	accgttacta	cgagtcttgc	gggttagaca	atgcccgtgt	ggacaaaaaa	2100
gtccgtattt	agaagctgag	tggggtaatt	cgtgatgt	agttcgacga	ttttctggc	2160
gtaaaaacta	gtaacaaagc	tggogacaat	gacaaggcagg	acaaggccaa	atatcaggcc	2220
attatttcgt	tataccttat	ggtgctttac	cagatcgtaa	agaacatgtat	ttacgtcaac	2280
tcacgttacg	tcattgtttt	ccactgttta	gaacgcgtt	ttgggatgt	tggcaaggat	2340
tttggaaat	attaccaggg	gtgcccgaag	ctgactgtatc	acttcatcg	agagaataac	2400
atgaaggaag	gaaaattggg	atgcaacaaa	aaagttaggac	gttatcttaa	aaataatatt	2460
tcctgctgca	cggatggact	gattaacaca	taccgttaacc	aggtggatca	tttcgcagt	2520
gttcgcaaaa	ttggtaacta	tgcggcctat	atcaaattct	tccgaagctg	gttcgaactt	2580
taccattatg	tgattcaacg	tattgtgtt	gatgagtatc	gtttcgact	taacaacaca	2640
gagtccaaact	ataaaaaactc	cattatcaa	caccatacg	actgttaaga	tatgttaag	2700
gcattgaata	cgcctttgg	ctacgacgt	cctcgctaca	agaacttgc	gatccccgac	2760
ttgttcgacc	gtaacaattt	tttaaacaag	acgaaggaat	cgattgtgc	taattcaac	2820
attgattcac	ag					2832

```

<210> SEQ ID NO 8
<211> LENGTH: 2901
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
    flavefaciens strain XPD3002, modified for expression in
    Escherichia coli

```

```

<400> SEQUENCE: 8

atgatcgaga aaaaaaaaaatc ttttgctaag ggcatggcg ttaagtccac cttggttca 60
ggttctaagg tatatatgac cacttcgca gagggatccg acgcacgtct ggagaaaatt 120
gtcgaaggag attcgatccg ttccgtgaat gagggggagg cgttctccgc ggagatggcg 180
gacaaaaatg cgggttataa gattggaaac gctaaatttt cccacccgaa aggatacgca 240
gtggtagcca ataacccctt ttacacaggg cctgtgcaac aggacatgtt gggattgaag 300
gagactttgg aaaagcgcta ttttggtag tccgcagatg gaaacgataa tatctgtatc 360
caggtaattc acaatatctt ggatattgaa aagatcctt ctgagtcat taccaacgt 420

```

-continued

gcctacgccc tgaataatat ctccggctta gacaaggaca ttattggcctt tgsgaaagttc	480
agtaccgtct atacgtatga cgaatthaag gacccagaac accatcgatc cgccctcaat	540
aataatgata agttgatcaa tgcaattaaa gcccagtagc acgaatttga taacttcttg	600
gataatcccc gcttaggcta ctccggcaa gctttttca gtaaggaggg gcgttaactac	660
attattaatt acggcaatga gtgttacat atccttgcatt tactttcggg gcttcggcac	720
tgggttgtac acaataatga ggaagagtcg cgcattagcc gcacgtgggt gtataacctt	780
gataagaacc ttgacaatga atacatctct accctgaact acttatatga tcgcattacg	840
aatgagttaa ccaattcatt ctcaaagaat agtgcagecca acgtcaacta tatcgacag	900
acgctgggta tcaaccggc ggaattcgcc gagcagtatt tccgcttttc aatcatgaag	960
gaacaaaaga atctgggtt caatattacc aagttacgtg aagtaatgtt ggatcgtaag	1020
gatatgtctg agattcgaa aaaccataaa gtgttgaca gcatccgtac gaaggctac	1080
actatgatgg acttcgttat ctaccgctat tacatcgaaag aggatgocca agtggcagcg	1140
gogaacaaat cccttccaga caacgagaaa agtctttctg agaaagacat cttgtatac	1200
aacttgcgc gttccttaa tgatgaccag aaagatgcgt tgtaactatga tgaagctaat	1260
cgtatggc gtaagttgaa aaacatcatg cataacatta aggagttcg tgggaacaag	1320
acacgtgagt ataaaaaaaaa ggatgctcca cgtctccgc gcattttgcc tgcaggacgc	1380
gatgtcagtg cttagcgtt attaatgtat gcactgacaa tgtttctgga cgggaaggaa	1440
atcaatgatc ttctgactac acttattaaac aagtttgata atattcagtc cttcttaaag	1500
gttatgcctt tgattgggtt aaacgcgaaa ttgtcgaag agtatgcctt tttcaaggat	1560
agcgcgaaaa ttgcgcacga actgcgtttt attaagagtt tcgctcgat gggggagcca	1620
atcgctgacg cccgecgccgca tatgtacatc gatgtatcc gcatcttgg tacaaacttg	1680
tcatacgatg aacttaaagc tttagcagac acctttcgc tggatgaaaa cggaaacaag	1740
ttgaaaaagg ggaagcatgg aatgcgcaat ttattatca ataacgtgtat ctcaaataag	1800
cgtttccact atcttatccg ttatggagat cggcacacc tgcataatgc tggcaagaat	1860
gaggccgtgg tgaaattcgt tttaggcgc attgctgata ttcagaagaa acagggcag	1920
aatggaaaga atcaaatcga ccgttactat gagacgtgtt ttggcaaga caaggggaaa	1980
tcggtttccgg aaaaagttga cgccttgcacg aagatcatca cgggcataaa ctacgacac	2040
tttgacaaaaa aacgcgttgtt aattgaagat accggacgtg agaatgcgga acgtgagaaa	2100
tttaaaaaga tcatacgatgtat gatctgacc gtaatttatac atattttaaa aaatatcgta	2160
aacatcaacg cacgttatgt gatcggttc cactgtgttag aacgcgcacgc tcaactttat	2220
aaagaaaaagg ggtatgatata taacttgaaa aagtttagagg agaagggatt ctcatcgatc	2280
accaagttgtt ggcgggttat tgacgaaacg gcacccggaca agcgcataa cgttggaaa	2340
gagatggccg aacgcgcacaa ggaaagtatc gactcattag aaagcgcaaa tcccaagctg	2400
tatgccttattt atatcaagta tagcgatgatc aagaaggccg aggagtttac ggcgcacatc	2460
aaccgtgaaa aggccaaaaac tgcatgtatc gcctacttgc gcaatcgaa atggatgt	2520
atcatccgtg aggacctgtgc gctgtatcgat aacaaaacat gtacttttatt tcgcaataaa	2580
gcgggtacatc ttgaagttggc gctgtacgtt cacgcgtata tcaatgcacat tgcagggat	2640
aattccattt tccagctgtatc tcaactacatt atgcacacgca ttattatgaa cgagcgatc	2700

-continued

gagaaaagca gcggcaaagt atccgaatac tttgacgcag ttaacgtga gaagaatata	2760
aacgaccgct tactgaaatt gctgtgtga cctttgggt attgcatccc ccgtttaaa	2820
aacctgagta tcgaagctct gtttgaccgc aacgaggccc ccaaatttga taaggaaaag	2880
aaaaagggttt cggaaatag t	2901

<210> SEQ ID NO 9
<211> LENGTH: 2388
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus sp., isolate 2789STDY5834894, modified for expression in Escherichia coli

<400> SEQUENCE: 9

atggaaatca acacttcgaa ccccacccat cgacgcggtg aaagtagcag tgttcggtgg	60
gacatgcttg gactgaagtc agagctggag aaacgccttt ttggaaagac ctgcacat	120
aacattcata ttcaattgtat ctacaatatac ttggacattt aaaaatcct ggccgtgtac	180
gtcactaata ttgttatatgc actgaacaat atgctggag tgaaggcag tgagagctac	240
gatgacttca tgggetatct gtcagcgcag aatacatattt acatcttac tcateccagat	300
aagtcaaacc tgagtgacaa agtgaaaaggc aacattaaaa agagtctgtc caaatat	360
gatctgctga aaacaaaacg tttgggttat tttggactgg aggagccaa aactaaggac	420
aagcgegtga gcgaagccta caagaaacgt gtttatata tgctggcaat tgtgggtcag	480
atccgtcaaa gcgtttcca tgacaagtct aatgaattgg atgagtatct gtactcggtt	540
atcgacatta tcgacagcga atatcggtac acgctggatt atttgggtga tgaacgttcc	600
gatagcatca ataagggtt cgtccagggg aataaggtaa acatctcggtt actgattgtac	660
atgatgaagg ggtatgaggc cgatgacatt atccgcattat actatgactt catcggttg	720
aaatccaaa agaaccttgg ctttccatt aaaaacttc gtgagaagat gcttgatgag	780
tacggtttcc gttcaagga taaaacaatac gattcagtgc ttagcaaat gtacaaggat	840
atggattttt tattattctg caactattat cgtaacgcac tggttagcggg cgaggcttt	900
gtccgtaaac tgcgttctc gatgacagat gacgaaaaag aaggcatcta tgccgacaa	960
gcccggaaaat tggggccaa gttccgtaat gactttgaga atatcgctga tcatatgaat	1020
ggagacgtta tcaaggaact tggcaaagcc gacatggatt tcgacgagaa gatcctggat	1080
tctgaaaaga agaacgcgtc ggacttgcgt tattttcgta agatgtatcta tatgttact	1140
tatttcttgg atggcaaaga aattaacgcac ctgttgcacca cactgatttag caaatttgc	1200
aacattaagg agttccctaa aattatgaag tctagcgcac ttgacgtgga gtgcgagctg	1260
actgcgggat acaaattgtt taacgacagt caacgtatca cgaatgaact tttcattgt	1320
aagaacattt cgatgcgcg caagccggct gccagtgcaa agttgaccat gtttcgtgtat	1380
gtctgtacca tcttaggcat tgatgacaag attaccgtatcc accgcatttc cgaaatttctt	1440
aagttaaaag aaaaaggaa aggaatccat ggtttcgta actttatcac caacaatgt	1500
atcgagtccct cgcgtttgt ctacttgatt aaatatgtca acgcacaaaa gattcgcgaa	1560
ttagctaaaa acgaaaaagt tggatgttt gtttaggtt gcattcccgat tacccagatt	1620
gaacgctact ataaaagctg tgcgaaattc cggacatga actcatctt agaggcaaaa	1680

-continued

tgttcagagt tagctcgat gatcaagaat attagttcg atgactcaa gaatgtaaaa	1740
cagcaagcaa agggccgcga aaatgttagcc aaagagcgcg ctaaggctgt catcgattg	1800
tatctgacag tcatgtaccc tcttgtcaag aatttggtca acgtaaatgc tcgctatgtt	1860
attgctatcc attgtttaga acgcgacttc ggcttatata aagaaattat tccggagttg	1920
gacctaaaaa acttgaagaa cgattaccgt attttgagtc agaccctgtg cgaactgtgc	1980
gacgaccgcg acgagtcacc taacctgttc ttgaagaaaa acaagcgtt acgtaagtgt	2040
gtggaggtgg acatcaacaa tgccgatagc tccatgaccg gttaataccg taattgcatt	2100
geccatctta ccgtggttcg cgaattaaaa gagtatattg gcatatccg tactgtcgat	2160
tcttatttca gcatctacca ctacgtttagc cagcgttgc tcacgaaacg tgaggacgat	2220
accaaacaag agaaaaagat taagtacgaa gacgatctgc tgaaaaacca tgggtatacg	2280
aaggacttcg taaaagcgtt gaactcccc ttcggctata acattcctcg cttcaagaac	2340
ttatctatcg agcaactttt tgaccgtaac gagtatttaa cggagaaaa	2388

<210> SEQ ID NO 10
<211> LENGTH: 2862
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Eubacterium
siraenum, modified for expression in human cells

<400> SEQUENCE: 10

atggcaaaaa aaatccacgc ccgggacttg agggagcaga gaaaaactga tcgcacagaa	60
aaattcgcgc atcaaaacaa aaaaaggaa gctgagagag ccgtccctaa gaaagatgca	120
gcgggtctcg tgaaaagcgt gагtagcgtt tccagtaaaa aagacaatgt aaccaagagt	180
atggccaagg cagccggcgt aaagtcaagg ttcgcgggtgg gtaacactgt ttacatgaca	240
agtttggtc gaggaaacga cgctgtattt gacgagaaga ttgtggatac aagccatgaa	300
cccctgaaca ttgacgatcc agcctatcaa ctgaatgtgg taaccatgaa cggatactca	360
gttacaggcc ataggggtga gactgtttct gccgttaccg acaaccgtt gagggcgttt	420
aatggacgaa aaaaagacga gcctgagcag tccgtaccaa ccgatgtct ttgcctgaag	480
cccacccctcg agaaaaaaatt ttttggaaag gagttcgatg ataatattca catccagctt	540
atatacaaca ttctcgacat agaaaagatt ctgtgtct actcaacaaa tgcgattac	600
geactcaata acatgagcgc cgacgagaat atcgaaaata gcgatcccc catgaaaagg	660
actacggacg agacattcga tgactttgaa aaaaaaaaaa agtccacaaa cagtagggag	720
aaggcggatt ttgacgcctt cgagaaattt atcggttaact acaggcttgc ctatggcg	780
gacgcgttct atgtgaataa aaaaaatccc aaaggaaaaag caaagaatgt gctcagagag	840
gataaaagAAC ttttactcgat ttttgcgttc atcggttaacg tccgccactg gtgtgtacat	900
tctgaagagg ggagagcggaa gttctggctc tataaaattgg acgagcttaa ggacgacttc	960
aagaacgttc tcgacgttagt gtacaaccga cctgtggaaag agataaataa cagatttac	1020
gaaaacaata agttaaacat ccaaataattt ggctccgtct aaaaaacac agatattgcc	1080
gaacttgcgtca gaagctacta cgagtttttgg attaccaaga agtataaaaaa catggattt	1140
tcaattaaga agttgagaga aagcatgctc gagggaaaag gttacgcggaa taaagagttat	1200
gacagcgtga ggaacaaact ttaccaaatac acggacttca ttctctacac aggttacata	1260

-continued

aatgaggaca	gcgacagagc	agacgatctt	gtaaatacgc	ttcgctcttc	cctgaaggaa	1320
gacgacaaga	ccactgtgta	ctgcaaggag	gctgattacc	tctggaagaa	gtaccgagaa	1380
tccattcggg	aagttagccga	cgcacttgcac	ggcgacaata	ttaaaaagg	tgatggaa	1440
aacattgaga	ttcagaga	taagcttgc	aagtgc	tctttatgc	ggattctgtc	1500
agtgaattca	caaagctgat	ctacttgctt	actagattct	tgagtggtaa	ggaaattaat	1560
gaccttgtt	caacttgat	caataagttc	gacaatatta	gatccttct	cgaaattatg	1620
gtgagctt	gtctggacc	aacttca	gctgagttact	catttttga	aggttcaaca	1680
aaatatctgg	ctgaatttgg	tgagctcaac	tcctttgtca	agagttgttag	ctttgacatc	1740
aatgcaa	gcacgatgt	tcgagatgt	ttggatatcc	tggaaatcga	gtctgacaaa	1800
acggaagagg	acatcgaaaa	aatgatagac	aatatcttc	agattgacgc	aatggggat	1860
aaaaaaactca	aaaagaataa	cggcttgcga	aattttat	catcta	acgt catagacagc	1920
aaccgggttca	aatacctcg	gcttatggc	aatccaaaaa	agattagaga	gaccgcaag	1980
tgcaaaaccag	cggtccgg	tgtgctgaac	gaaatttcc	acgcacagat	tgaacggtat	2040
atgaa	gcat	gctgcctaa	aaacacgg	ctgtgc	cgaataaaag	2100
ttggcggata	tgatcg	cgat	aaattt	ttt	catatc	2160
aaagcgaacg	ttacctc	acg	gac	ctc	agatgat	2220
ataagactgt	atcttactgt	tatgt	atc	atc	tgta	2280
cggta	cgt	ttgcgt	ttgcgt	cgag	agctgtatgc	2340
ctggaggt	gaaat	atc	aaaga	aca	acg	2400
aaactcgaaa	acgg	tatt	caag	actg	aaat	2460
aacagg	tatc	tgag	gac	atgc	tgat	2520
gagcgg	gg	ttgt	aaac	gca	atgc	2580
attaacatta	acat	taa	aaat	ggaa	tttgc	2640
cttata	acat	ttt	aaat	cgatt	tttgc	2700
gat	ttt	ttt	ctaa	actt	gttgc	2760
tgcac	ccgt	tcgg	ctataa	tttgc	tat	2820
gacaaaact	acc	ccgg	ggaa	agac	gatgt	2862

<210> SEQ ID NO 11
<211> LENGTH: 2757
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus sp., isolate 2789STDY5834971, modified for expression in human cells

<400> SEQUENCE: 11

atggcaaaaa	agaataaaat	gaagccgcgg	gaacttaggg	aagctcagaa	aaaggcccga	60
caacttaaag	ctgccgagat	aaacaacaac	gctgcacccgg	cgatagccgc	catgcctgca	120
gctgaggtga	ttgcacctgc	tgccgaaaaa	aagaaatcaa	gctgaaagc	agccggcatg	180
aaatctatcc	tcgtgtccga	aaataagat	tatattacgt	cttttgaaa	agggatagt	240
gcgg	ttctcg	agtacga	agataataat	gattataatc	aaactcaact	300

-continued

gacaatagca atatacaact tggcgggtt aacgaggta acattacctt ttcaagcaag	360
cacggcttg agtcaggtgt agaaataat acaagtaacc ccactcatcg ctcagggaa	420
tcatcacctg tacgcggga catgctcggg cttaagttag aactggagaa acgcttctt	480
ggtaaaaat ttgacgacaa tattcatata cagctgatct ataatatct tgatataag	540
aaaatcttgg ctgtatacgt cacaacatc gtatacgcac ttaataatat gctcgggtt	600
aaaggcagcg aaagccatga cgacttcatt ggatacctta gcaccaataa catctacgc	660
gtattcatcg acccagacaa tagcagtcg agcgatgaca agaaggctaa cgtgagaaag	720
tcactctcca aatttatgc cttgtttaaa acaaagagat tgggtactt tgggttgaa	780
gagcctaaga cgaaggataa tcgcgtatca caagcctata agaagcgggt ctatcacatg	840
ctggcgatcg tgggtcaaat tcgccaatgt gtttccacg acaagtctgg cgctaagaga	900
ttcgatcttt acagttcat caacaacatc gaccccgagt accgggacac cctggactac	960
ctcgtggagg aaagactcaa gtcaatcaat aaggattttt ttaagataaa caaggtaat	1020
atatccctcc tcataagatat gatgaaaggt tacgaggccg atgatatcat tcgactgtat	1080
tacgatttca ttgtactgaa gagtcaaaaa aatctggct tctcaatcaa aaaactgcgg	1140
gagaaaaatgc tggacgagta tggttttagg ttcaaggata agcaatacga cagtgtccgc	1200
agcaagatgt acaagctcat ggatttttg ctctttgtt attactacg aaatgacata	1260
gctgcaggcg agtctttggt gcgaaaattt cgctttcca tgacagacga tgaaaaggag	1320
ggcatatatg ccgtatgaaatc tgctaaattt tggggaaaat ttccgaacga tttcgaaac	1380
atccggacc acatgaatgg agatgtcato aaggagctt gtaaagctga tatggacttt	1440
gacgaaaaga tattggacag tgaaaaaaaaa aacgctagcg atcttctta ttttccaag	1500
atgatataata tgctgacgta tttcttgac ggtaaagaaa taaacgacct gctgactaca	1560
ttgatttcaa aatttgacaa catcaaggaa tttctgaaaa taatgaagag ttccgcggta	1620
gatgttagat gtgagggtac agccggatac aaattgttca atgatagtcg gaggatcacc	1680
aatgagttgt tcattgttaa gaatattgcg tctatgagga aaccagcggc aagtgcata	1740
ttgacgatgt ttcgagacgc gcttacaatt cttggatcg atgacaaaat cactgacgac	1800
cggatttcag ggatactgaa gctcaaggaa aaggaaaag gcattcatgg gcttaggaac	1860
tttatcaact acaatgtaat tgaatctgcg cggttcgtct acttgcataa gtacgccaat	1920
gcccggaaa ttagagaagt tgccaagaat gaaaaggctg tgatgttcgtt attgggggt	1980
attccagata cacagatcga acgctactac aagtcttgcg ttgagttccc ggacatgaac	2040
tccctctgg gggtaagcg ctccgaactg gctcgatgt ttaagaacat tagcttcac	2100
gatttcaaaa acgtcaagca acaagcgaag gggcgcaaa acgttgcacaa ggagaggct	2160
aaagcgtga tcggcttta tctcacatgt atgtatctt ttgttaagaa tcttgtcaat	2220
gtcaatgcac ggtatgttat agtatacacc tgcgtcgaac gagacttcgg tctctacaaa	2280
gaaattatttc cagagcttgc aagtaaaaac ctgaaaaatg attatgcac tttgtcaca	2340
acgttgtgtg agctgtgcga taagtctcca aaccttcc ttaagaaaaa cgaacgattt	2400
cgaaagtgtg tcgagggtgaa tatcaataat gcccgttcc ccatgacccg aaaatata	2460
aactgtatttgcgacttgcgacttgcgacttgcgacttgcgacttgcgacttgcgacttgcg	2520
acgttgtgtg acgttgtgtg acgttgtgtg acgttgtgtg acgttgtgtg acgttgtgtg	2580

-continued

gagaatgata cgaagcaaga agaaaagata aagtatgaag atgaccttt gaaaaaccac	2640
ggttatacga aggacttcgt aaaagcttt aactcaccat ttggttacaa tatcccaaga	2700
ttcaagaacc tctcaatcga gcaattgttc gatcgaaatg agtatctgac ggagaaa	2757

```

<210> SEQ ID NO 12
<211> LENGTH: 2754
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
bicirculans, modified for expression in human cells

```

<400> SEQUENCE: 12

atggcaaaga agaacaaaat gaagccgcgc gagttgcggg aggccaaaa gaaagctgc	60
cagctgaagg ccgcggaaat caataacaac gcagtcctt ccatagctgc catgccagca	120
gcgcgaagccg ccgcaccggc tgccggaaag aagaagtctt cagtaaaagc tgccggcatg	180
aaaagtatac ttgtgtcaga gaacaagatg tatatcacca gttttggaaa aggcaactcc	240
gcagtgottt agtatgaggt agataacaat gattacaaca agacgcgtt gtccagcaaa	300
gataactcaa acattgaact gtgcgacgtt ggcaaggtaataatactt cagtagtcgc	360
cgccggattt aatcagggggt ggaatcaat acttctaacc caactcatcg gtctgggag	420
agctcttcag tacgcgggta tatgttggga cttaaatctg agctcgaaaa gagattttt	480
ggtaagaact tcgatgataa catccacatc caattgattt ataatatctt ggatataag	540
aagatactcg cagttatgt gactaacatc gtctacgcgc ttaacaatat gctcggtag	600
ggagatgagt ctaactacga ctttatggc tatctgagca catttaacac ctataaagt	660
ttcactaatac ccaatggaa tactttgagc gatgacaaga aagaaaacat tcgcaagtca	720
ctctctaagt tcaacgcctt cctcaagacc aaacgcgtt ggtatttgg tctggagaa	780
cccaaaaacgaa aagacactag agcttcagag gcataacaaga aacgagata ccatatgct	840
gccattgtcg ggcagatccg ccagtggtg tttcatgata agtctggagc aaaacgattc	900
gacctgtata gttttatcaa caatatacg cccgagttt gggaaacttt ggactacatt	960
gtagatgagc gttttgactc cataaacaag ggctttatac aaggaaataa agtcaatatc	1020
agtctgtca tagatatgt gaaagggtat gaagctgacg acattattcg cctgtactat	1080
gactttatcg ttcttaagt tcagaaaaat ctggcttca gtataaaaaa gctccgcag	1140
aagatgtgg atgagttatgg atttagattc aaggataagc agtacgacag tgtaagatct	1200
aaaatgtata aacttatgga tttctgttg ttctgcaact actaccggaa cgacatcgcc	1260
gcgggtgaga gtttggtaga aaagcttcgg ttctccatg ccgacgacga aaaggaagg	1320
atatatgcag atgaagccgc taaactctgg ggcaagtttc gaaatgactt cgaaaacatt	1380
gcggatata tgaacggta tgtgataaaa gaacttggaa aagccgatat ggactttgt	1440
aaaaagatac tggactcaga aaagaaaaac gccagtgacc ttctttactt cagcaagat	1500
atctacatgc tcacctactt tctggatgg aaagaaaatca atgatttgc tacaacctt	1560
atctctaagt tcgataatataa aagaaattt ttgaagatc tgaaatctag tgctgtggac	1620
gtagagtgta aactcacagc aggtatataag ctcttaatg atagccaacg aataacaaac	1680
gagctttca tagtggaaaaa cattgccagc atgcggaagc cggcggcgtc agcaaaattt	1740

-continued

accatgttcc gcgatgcact gactattctt gggatcgatg ataaaataac ggatgatcgc	1800
ataaggcaga ttctgaaatt gaaggaaaag ggttaagggtt tacacggttt gcggaaacttc	1860
attacgaaca acgtcattga atccagtcga ttttgttatac tgataaaagta cgcgaatgcg	1920
cagaaaataa gggaggttgc taaaaatgag aaggtcgta tgttcgtaact tggcggcatt	1980
cccgacacac aaatcgaaag gtattacaaa agttgttagt agttcccaga tatgaacagt	2040
tccctggag taaaacggtc tgaactggcg agaatgataa agaatatac attcgacgac	2100
ttcaaaaatg taaagcaaca ggcgaaagga agagagaacg tggctaagga acggggccaaa	2160
gecgttattg gactttacct tacggttatg tacttgttgg taaaaaacct tgttaatgta	2220
aacgcacgct atgttatagc aatacattgc ctggagagag acttcgggct ctacaaggaa	2280
ataattcccc aactcgcttc aaagaacctt aaaaacgatt accgcattct tagtcaaacg	2340
ctctgcgagc tctgcgacaa atccccatac ctgttctca aaaaaaatga gagactcagg	2400
aagtgcgtcg aggttgacat caataatgca gattctagta tgactcgaaa gtatcgaaac	2460
tgtatcgccg acttgacagt tgcgcgaa ctgaaagaat acataggcga tatctgtacc	2520
gttagactcat atttctcaat ttaccactat gtgtatgc当地 gatgcataac caagagggag	2580
aacgcacgca aacaggagga aaagattaag tacgaggatg acttgttggaa aaaccacggt	2640
tatacaaaaag attttgc当地 ggcactgaat agtcctttt ggtataatat cccgaggttc	2700
aaaaaccttt caattgaaca actcttcgat aggaacgagt acctgacgga gaag	2754

<210> SEQ ID NO 13
<211> LENGTH: 2766
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus sp., isolate 2789STDY5608892, modified for expression in human cells

<400> SEQUENCE: 13

atggcaaaaa agaacaagat gaagccccga gagttgcggg aagcgcagaa aaaagcgagg	60
cagcttaagg ccgctgaaat caacaacaat gcccgtcccg caatagctgc gatgectgcc	120
gcggaggtga ttgcaccagt agcggagaag aagaaaagtt ctgtaaaagc tgcaggtatg	180
aaaagcatat tggtaagtga aaacaagatg tatataacta gtttcggcaa aggttaattct	240
gccgtgttgg aatatgaggt tgataataac gattacaata aaacccaact ctccctctaaa	300
gacaattcaa atatagagct cggcgacgta aatgaagtga acattacgat ctccagcaaa	360
cacggtttcg gctcagggggt ggaattaat acttctaaacc cgacacacccg gagttgttag	420
tcatctccag tgagaggaga tatgctcgaa ttgaaatccg aactcgagaa acggttttc	480
ggcaagacat tcgacgacaa catccatac cagttgattt ataacataact cgacatcgag	540
aaaattttgg ccgtgtatgt gacaaacatt gtttatgcat tgaacaacat gctgggtata	600
aaagattcg agagctatga cgactttatg gggtaacttgc gtgcacgcaa tacctacgag	660
gtgtttacgc acccagacaa gagaatttg tctgacaagg tgaaggtaa tattaagaag	720
tcccttcaa aatttaacga cttgtgaaa actaaacgcg tgggtactt tggactcgaa	780
gaaccaaaaa ccaaggatac aaggccatca gaagcctaca agaagagggt gtaccatgt	840
ctggctatag taggtcagat tcggcagtgc gtattccacg acaagtcagg tgcaaagaga	900

-continued

tttgatctt actcattcat aaacaacatt gatccggaaat accgggatac gctggactat	960
ctggtagaaag agcgattgaa gtcaatcaat aaagattta ttgaaggaaa caaagtgaat	1020
attagcctgc tgatcgacat gatgaaaggg tatgaagctg atgacatcat acggctctac	1080
tacgacttca tagtactcaa gagtcagaag aacctgggtt tttccatcaa aaaactgcga	1140
gaaaagatgt tggagaataa cggcttcgc ttcaaagaca aacagtatga ttccgtccga	1200
agcaaaatgt ataagttat ggatttcgt ctttctgca attattacag aaatgacgta	1260
gcccggggag aagccctggt acgaaagtgg agattctcta tgacggatga cgagaaggaa	1320
ggcatctatg ctgacgaggc agcgaagctg tggggaaaat tccgcaacga cttcgaaaac	1380
atagcggatc atatgaatgg ggacgttataa aaagaactcg gaaaagcggaa tatggacttt	1440
gatgagaaga tcctggattc tgagaaaaaa aacgctagtg atcttctcta tttctctaag	1500
atgatttaca tgctcacgta tttttggat ggcaaaagaaa ttaatgatct cctcaactacc	1560
ctcatttcta agttcgacaa tattaaggaa ttccctaaga tcatgaagag ttcatcggtc	1620
gacgtagaat gtgagttac tgccggatac aaattgttta acgatagcca gcgaatcacf	1680
aatgagctgt tcattgtcaa gaatatcgcc agtatgagga agcccgctgc gtctgcaaaa	1740
ttgactatgt tccgegeatgc tcttaccatt ctggccattt acgacaatata aactgacgac	1800
cgcacatcgatg agatcctgaa gctcaaggag aaggggaagg ggttccacgg attgeggaaat	1860
ttcatcacaa ataacgtaat tgagagttcc cgggtcgatg atcttattaa atatgc当地	1920
gctcaaaaga taagagaagt agcaaaaaac gagaagggtgg tcatgtttgt actggccgaa	1980
atacccgaca cccaaatcgaa acggatttat aaatcttgcg tagaattccc agacatgaaac	2040
agttcactcg aagcgaagag atcagaactc gcgcggatga taaaaacat ttccctcgac	2100
gacttcaaaa acgtcaaaaca gcaggcgaaa ggttagggaga atgttgcgaa agaaagagct	2160
aaagcgttaa ttggctgtt tctgaccgtc atgtacgtt tggtgaaaaa tcttgtcaac	2220
gttaaatgcgc gatacgatcat cgcgcattt tgtcttgacg gagacttcgg gctctataag	2280
gagattatcc ctgagttggc cagtaaaaat cttaaaaacg actacagaat ccttagccag	2340
acgctttgtg agctttgtga cgacagggaaac gagtcttcca atctgtttct caagaaaaat	2400
aagaggctca gaaaatgtgt agagggttgcg atcaataacg ctgatagctc tatgactcg	2460
aagtatcgaa attgtattgc acaccttacg gtatgttaggg agctgaaaga atatatcgcc	2520
gatatacgaa cagtagacag ctatccatg atataccatt atgtcatgca acgctgcatt	2580
accaagaggg gggacgatac caagcaggag gagaaaaatca aatacgaaga tgacttgctc	2640
aagaatcacg gttatactaa ggatgggtt aaagcgctca atagtcctt tggctacaac	2700
atccccccat tcaagaacctt gatgtattgaa caactttcg atagaaacga gtaccttact	2760
gagaaaa	2766

```

<210> SEQ ID NO 14
<211> LENGTH: 2766
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
      sp. CAG:57, modified for expression in human cells

<400> SEQUENCE: 14
atggccaaaaaaaataaagat gaaaccacgc gaattgcggg aagctcagaa aaaggctaga      60

```

-continued

cagttgaagg ccgcggagat aaacaacaat gcagcacctg ctatgccgc catgecagct	120
gccgagggtga ttgccccgt agcgaaaaag aagaaatcc cgtaaaaac ggccccggatg	180
aagagcatcc ttgtgagcga gaacaaaatg tacattacaa gcttttgtaa aggaaactca	240
gctgtgttgg agtacagaatg cgacaataac gactacaaca agacccagct gtcctctaaa	300
gacaatagca acatagaact gggcgacgta aacgaggtaa atataacgtt ctcttctaa	360
catggcttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgc	420
atgtagccccgtttaggggaga catgctcgcc ttgaaatcag agctggagaa gagatTTT	480
ggaaaaacat tcgacgataa tatacacatc cagctgatatacatttgcagtttgcagtttgc	540
aaaatacttg cagtgtacgt tacgaacatt gtctatgtt tgaacaatatacatttgcagtttgc	600
aggattcccg agtccctacgatgatgatgatgatgatgatgatgatgatgatgatgatgatgat	660
gtgttcaatc atccggacaa atccaatctc agtgataaag tgaaggccaa cataaagaaaa	720
tcccttcataatttacgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgat	780
gaaacctaaaaa cgaaagacac tagagccagc gaggctataaaaaaagagtcatcataatc	840
ctcgctatag ttggacaaat taggacatgt gtgttcatg acaaaaatgg tgcaaaacgg	900
ttcgatctgt actcatttat caacaacatttgcagtttgcagtttgcagtttgcagtttgc	960
ttgggttggagg aacgattgaa atctataaac aaggattca ttgaggggaa caaggtaat	1020
ataaggcatttc tcattgatatacatttgcagtttgcagtttgcagtttgcagtttgcagtttgc	1080
tatgattttatgttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgc	1140
gagaagatgc ttggaaataa ttggtttcgg tttaaagata aacaatatgc ctccgtgagg	1200
agtaaaatgttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgc	1260
gcagcaggcg aagcactcgttgcagtttgcagtttgcagtttgcagtttgcagtttgcagtttgc	1320
ggaatatacg ctgacgaagc ggcgaaactg ttggggaaat ttgcacgcgatgatgatgatgat	1380
atagctgacc atatgaatgg cgacgttacaaatgggggggggggggggggggggggggggggggg	1440
gacgagaaaa ttctcgacag tgagaaaaag aacgccatgttgcatgttgcatgttgcatgtt	1500
atgatataca tgctcacata cttttcgtatgttgcatgttgcatgttgcatgttgcatgtt	1560
cttattagca aatttgcataa catcaaagatgttgcatgttgcatgttgcatgttgcatgtt	1620
gatgtggagt gcgagtcac ggcagggttat aaacttttgcatgttgcatgttgcatgttgc	1680
aatgagctgt tcattgtcaaaatgg	1740
cttacgtgttgcggacgc cctcacgtatgttgcatgttgcatgttgcatgttgcatgttgc	1800
agaatcagtg agatacttaa gctcaaggaa aagggggaaag ggatacacgg tctgcgcac	1860
ttcataacgatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgtt	1920
gcccaaaaga taaggaaatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgtt	1980
attcccgacatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgtt	2040
agctccctcg aggctaagcgatgttgcatgttgcatgttgcatgttgcatgttgcatgttgc	2100
gattttaaaaa atgtaaagca acaagctaaatgggggggggggggggggggggggggggggg	2160
aaagcagtgatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgtt	2220
gtcaatgcaatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgttgcatgtt	2280
gaaatcatcc cggagttggc atctaaaaac cttaagaatgttgcatgttgcatgttgcatgtt	2340

-continued

accttgtcg	aactctgcg	tgaccgaaac	gaatcatcta	acctcttcct	aaaaaaaaac	2400
aagagactca	gaaagtgtgt	ggaggtggat	atcaataatg	ccgattccag	tatgactaga	2460
aaataccgca	actgcacgc	acacctgact	gtggtcagag	aacttaaggaa	gtacatttgg	2520
gatattagaa	cggtcgactc	atattttgc	atctatcatt	atgtcatgca	gaggtgtatc	2580
accaagagag	gagatgatac	aaagcaggaa	gagaagataa	agtacgagga	cgatcttctt	2640
aagaaccatg	gctacactaa	ggacttcgta	aaagcgttga	actccccgtt	cgggtataaac	2700
ataccttaggt	ttaagaatct	ttcaattttag	caattgttttgc	accgcaatga	gtaccttaca	2760
gagaag						2766

<210> SEQ ID NO 15
<211> LENGTH: 2799
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus flavefaciens FD-1, modified for expression in human cells

<400> SEQUENCE: 15						
atgaaaaaga	aatgtccctt	gcgagaaaaa	agggaaagctg	aaaaacaagc	aaagaaggcc	60
gcgtactcag	cagcttccaa	gaataccgac	tccaaaccag	cgaaaagaa	ggcagaaacc	120
ccgaagccgg	cagagataat	aagtgacaac	agtcggataa	aaacggctgt	gaaagctgcg	180
ggccttaaat	ctaccattat	atctggagat	aagctgtaca	tgacatcatt	tggtttaggg	240
aacgctgcgg	ttattgaaca	gaagatcgc	atcaatgact	atagcttctc	tgcttatgaaa	300
gatacacccat	cccttggaa	ggacaaggct	gaaagcaagg	aaatttcatt	tagcagccac	360
caccgcgtcg	tgaaaaatga	taaactgacc	acctacaacc	cattgtatgg	tggaaagat	420
atccggaaa	aaccaggtagg	aagagacatg	ctggactga	aggacaagct	tgaagaacgg	480
tatttcggat	gcaccccaa	tgataacttg	catattcaga	ttatatataa	catactcgat	540
atcgaaaaga	tacttgcagt	gcactccgca	aacatcacga	ccgcgttgg	tcacatgg	600
gacgaagatg	atgagaata	tcttaacagt	gattacatcg	ggtacatgaa	cacaattaac	660
acatacgcacg	tatttatgga	cccttctaaa	aattccagcc	tctcacctaa	ggaccgcaag	720
aatatcgaca	acagtcgagc	caagtttggaa	aaactgttga	gcacgaaaag	gcttggatata	780
ttcggattcg	attatgacgc	caatggtaag	gacaaaaaaa	agaatgaaga	gataaaaaaa	840
cggctgtatc	atttgactgc	attcgcttgc	caactgagac	agtggccctt	ccattctgct	900
gggaactacc	ctcgcacgtg	gctctacaaa	ttggacagct	tggacaaggaa	ataccatgc	960
acgctggacc	attactttga	taaacgggttc	aatgatattt	acgatgattt	tgttaccaaa	1020
aacgccacta	acttgcataat	actcaaggaa	gtatcccgg	aggcaatttt	caaagacata	1080
gccgacctt	actacgactt	tattgttatac	aagagccaca	agaacatggg	gtttccatt	1140
aaaaaaactcc	gcgagaagat	gctcgaaatgc	gatggtgctg	accgcatcaa	ggagcaggat	1200
atggactcag	taaggagtaa	gctttacaaa	ctgatcgact	tttgtatttt	taagtattac	1260
cacgaaatttc	ctgagggtgc	agagaagaac	gtcgacatac	ttcgagcagc	ggtttctgat	1320
acgaaaaagg	ataacccat	ttcagacgag	gctgctcggc	tgtggagcat	attcaaagaa	1380
aagttccctcg	gctttgtga	caaaatttgc	gtttgggtca	ccggagagca	cgaaaaggac	1440

-continued

atcacgtcag tgattgataa agacgcatac cgaatcgca gtaacgttc ttacttctcc	1500
aagcttatgt acgcaatgtg tttttttttt gatggtaagg agataaacga cctcctcact	1560
acccttatca ataaggctga caaatatagca aatcagatta agacggccaa agaactcgga	1620
ataaacactg catttgaaa gaactacgac ttcttcaatc atagcgagaa atacgttagac	1680
gagctgaata tcgtaaaaa tattcgctgg atgaaaaaac ccagttcaaa cgaaaaaaag	1740
geaatgttac atgacgcatt gacgatattt ggaatcccag aggacatgga tgagaaggct	1800
ctcgacgaag aattggacct cattttggag aaaaagactg atccgggtgac tggcaaacca	1860
ctgaaaggca aaaaccctct gcgaaatttc atagccaaca acgtaatcga aaacagttaga	1920
ttcatataacc ttattaagtt ctgcaccccc gagaatgtcc gcaagatagt caacaacaca	1980
aaggtcacgg aattcggttctt gaagcgcatt cctgatgccca aatcgagcg gtactacaag	2040
agttgtactg atagttagat gaacccccc acggaaaaaa agattacgga gctcgcttgt	2100
aagctgaaag atatgaattt tgggaacttc aggaacgtaa ggcaatctgc aaaggaaaaac	2160
atggaaaagg agcgcattcaa agcagtgtt ggcctgtatc tcaccgttgt gtaccgagtc	2220
gtcaagaatc ttgttagatgt gaacagtcga tacatcatgg cttttcacag tctggAACGG	2280
gatagtcacgc tgtacaacgt ctccgtggat aacgattacc tcgacttac ggacactctt	2340
gtcaaggaag gcgacaattc ccggtcacga tatctggccg gaaataaacg cctcgagat	2400
tgtgtaaagc aggatattga taacgcaag aagtggttt tgagcgacaa gtacaatagc	2460
ataactaaat accgaaacaa ttagtgcac cttaccgtt taaggatattt cgccggattt	2520
atcgggtata ttactaagat tgattccat ttgcactgt atcattatct gatacagagg	2580
caacttgcca agggccttggc ccatgaacgg agtggctttt atcgaaacta tccccaaatac	2640
gcaccattgt ttatggca tacttacgtt aaggacgtt gtaaggctct taatgctcct	2700
ttcggttaca atatacctag attcaaaaat ctgagcatcg atgcactttt cgaccgcaat	2760
gagattaaaa agaacgcacgg agagaaaaag tccgacgtat	2799

<210> SEQ ID NO 16
<211> LENGTH: 2832
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
albus strain KH2T6, modified for expression in human cells

<400> SEQUENCE: 16	
atggctaaaa aatcaaagg aatgagcctc cgcgaaaaac gggaaactcga gaagcaaaag	60
aggattcaaa aggcgcggtaa taattcgtt aacgatacac ccggaaaaac ggagggaaact	120
aatgtcgta gcgtcaacgt tcgacacctca gctgagaata agcactccaa aaaatctcg	180
gccaaggctc tggcccttaa gagtggctgt gttataggag acgaacttta cttgacgac	240
ttcgggtcggt gtaatgaagc gaagtggag aaaaagatggat gcgccgacac ggtcgagaag	300
ctggggatcg ggcgccttcga gggtgctgaa agggacgtat ctaacttcac gctcgagac	360
ggtcgcacca aagacaagac agccagacca aaagatccac ggcataattac tggtgatata	420
caaggaaaat tcaaaaagaat tatgtgggtt atccggagcg tactcgagaa aaaaatattt	480
ggcaaaaactt ttgtatgataa catccacgtt caactggcgat ataacattct tgacgttgaa	540
aaaatcatgg ctcagtgatgtt ctcagacata gtatacatgt tgcacaacac ggataagacc	600

-continued

gagcgcaatg ataacctgat gggatatacg tccatcgaa acacatataa gacatttc	660
gacactagca atctgcgtga cgacacaaag caaaaagttt aaaaccaaaa gagagagttc	720
gataagataa tcaagtccgg ccgactcgga tattttggag aagcatttat ggttaatca	780
ggcaataga cgaagtcgg acctgagaaa gaaatctacc atatttcgc gcttatggca	840
tccctgcgcc aaagctactt tcatggttac gtcaaggata cagattacca gggtaccacg	900
tggcgtata cgcttgaaga caaaactcaag ggtccatctc atgagttcg agaaacgatc	960
gataagattt ttgacgaggg gtttcaaaa atcagtaaag atttcggaaa gatgaacaag	1020
gttaatctcc agattttggaa acaaatacgata ggcgagctgt atggctccat cgagcgc当地	1080
aaccttacgt gtgactatta tgatTTATA cagctaaaaa aacacaaaata tctgggttc	1140
tccataaaac gcctcaggaa aacgatgctt gagacaacac ctgcggaaatg ttataaggca	1200
gaatgttata actctgagag gcaaaaactg tacaagctga tcgacttcct gatctacgat	1260
ctctactaca atcgcaagcc agcacgaattt gaagagatag tcgataagct gcgggagagc	1320
gtgaacgacg aggagaagga gtccatatac tcagttgagg caaatgtatgt ctatgagtc	1380
ttgtcaaaag tgctcgacaa gagtctcaaa aactctgtga gcggtgagac gatcaaagac	1440
cttcagaaac ggtatgacga tgagacggcc aaccggatct gggacatctc ccagcatcc	1500
atatccggta acgtgaactg ttctgttaag cttatctaca tcatgacact gatgtcgac	1560
ggcaaggaaa tcaatgatct cctgactaca cttgttaaca agttcgataa cattgttct	1620
ttcatagacg ttatggatga gcttggctg gagcacagtt ttaccgataa ctataagatg	1680
tttgagattt ccaaggccat atgcttggat ctgcaattta taaattccctt cgcttagatg	1740
tctaagattt atgacgaaaa atctaaacga cagctttca gggatgcgt cgtaattctt	1800
gacatcgaa ataaagatga gacctggata aacaactact tggattccga catattcaag	1860
ttggataagg aaggaaacaa actcaagggt gcccggcatg actttaggaa ctttatttgc	1920
aacaacgtca tcaagtccctc ccgggttaag tatctcgtaa agtactctag cgctgacggg	1980
atgataaaagc tgaaaaacgaa cgagaaactc atcggattcg tcctggacaa gctgcctgag	2040
acgcagatag atcgatattt tgaatcatgc ggccttgaca atgcggctgt cgacaagaaa	2100
gtgcgaatag agaagttgag cggacttatac agggacatga agtttgatga cttctccggc	2160
gtgaagactt ctaacaaggc cgagacaaat gataaaacaag ataaggcgaa gtaccaggct	2220
attattatgtt tggatctgtat ggtactgtac cagatagtaa aaaacatgtat ttacgtcaat	2280
tcccgctatg tcattgtttt ccactgcctt gaacgcgact ttggatgtta tggcaaaat	2340
tttggaaagt actaccaggg ctgtcgaaag ttgaccgacc acttcatacga agaaaagtac	2400
atgaagggaa gaaagttggg gtgcaacaaa aaggtcggcc ggtacctgaa aaacaatatt	2460
tcctgtgtt cggacggatt gataaaatact taccgaaatc aggtggacca ttttgcgtt	2520
gtccgaaaga taggaaacta cgcagcctac attaagtcaa taggctctg gtttgaactg	2580
taccactacg taattcagag gattgtctc gacgaataca gattcgctct taacaacacc	2640
gagtcaaattt ataagaattt catcatcaaa catcacacgt attgttaagga tatggtaag	2700
ggtgttgcata cggccgttgg ttatgttttgc ccacggatcaaaaatcttc cattggggat	2760
cttttcgacc gcaataacta tctcaacaaa actaaggaaa gcatcgacgc taatagtca	2820
atagattctc aa	2832

-continued

```
<210> SEQ_ID NO 17
<211> LENGTH: 2901
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
flavifaciens strain XPD3002, modified for expression in human
cells

<400> SEQUENCE: 17

atgatagaga aaaaaaaaaag ctttgcggaa gggatggcg taaaaagtac actggtatca      60
ggctctaagg tctacatgac aacgttcgca gaaggaagcg atgcacgcct cgaaaagatt     120
gttgagggag atagcattag gtccgtcaat gaaggagaag ccttttagtgc agaaatggca    180
gacaagaacg ctggatacaa gattggaaac gcaaaatttt cccatccaaa gggatacgca    240
gttgttagcta acaatccccct ctataccggg cccgtccagc aagacatgct tggcctcaaa   300
gagacgcttg agaagaggta ttttggagag agtgctgtatc gtaatgacaa tatctgtatc   360
caagttatttc ataacatcct cgacatagag aaaatccttgc cagaatatata caccaacgcc  420
gcataatgcag tgaataatata atccggctcg gataaagaca taatcgatt cggcaagttt  480
agtacagtat atacctatga cgagttcaaa gacccggagc atcatcgagc cgcttcaat    540
aacaacgaca aacttatcaa tgccattaag gctcaatatg acgagttcga taatttttg   600
gacaatccca gacttgggta ttccggccag gccttctttt ctaaggaaagg caggaattac  660
atcattaatt acggaaacgaa atgttacgat atcctcgctt tgctctctgg cctgcgccac  720
tgggttgtac acaacaacgaa ggaggaaatct cgaatttcac gaacttggct gtacaatttg  780
gataaaaaact tggataatga atacatcagt actctgaact atctctacga taggatcacc  840
aacgaaactta cgaattcatt ttcaaaaaat tccggccgcaa acgttaattt catcgctgag  900
acggtggca taaatccggc cgagttcgcc gagcaatattt ttaggttcag tatcatgaaag 960
gagcaaaaga atttggggtt caacatcagc aaactccgag aagtcatgct cgaccggaaaa 1020
gatatgtccg aaattcggaa gaaccataag gtattcgaca gcatccgcac aaaagtgtac 1080
acaatgatgg atttcgttat atacaggtat tatatacgagg aagatgcggaa agttggccgc 1140
gcaaaacaaa gtcttccaga taatgaaaag agcttgatgt aaaaagatata ttttggataa 1200
aaccttcgctt gttcccttcaa ttagtgcacca aaggatgtct tgcgtactacgaa cgaggcaaac 1260
cgaatctggc gaaaacttggaa aaacatcatg cataatataa aggaatttcg cgggaaacaaa 1320
acgagggagt ataagaagaa ggatgtctt cgcctccccc ggataactccc tgcggggcaga 1380
gacgtctccg catttagcaa actgtatgtat gctctcaacta tggtttggaa tggggaggaa 1440
ataaacgatc ttctgactac gttgatgttac aaatttgcata acattcagag ttttctcaag 1500
gtcatgccac ttatcggtt aatgcggaaat tttgttgggg aataacgcctt cttaaagac 1560
tccgctaaaa tagcggttgc gctccgcctt attaaatccct tcggccgaaat ggggttgcac 1620
atagcggtatc cccggcgagc tatgtatcgtt gatgtatca ggatccctgg aactaacttg 1680
agctacgacg aacttaaggc tctggcgac actttcgtt gggacgagaa tggggacaa 1740
ctgaaaaagg gaaaggacgg gatgagaaac ttctataataa ataatgttcat ttccaacaag 1800
aggttccatt atttggatcg gttatgtatc gctgcgcacc ttcatgaaat tgcgttgc 1860
gaagctgtgg tttaaatttgc tcttggcaga attgcccaca tccaaaaaaa acagggggcaa 1920
```

-continued

aatggtaaga accaaattga tagatactac gaaacttgca taggtaaaga caaaggtaaa	1980
agtgtctctg aaaagggttga tgccctgacg aaaatcatca caggtatgaa ctatgaccaa	2040
ttcgacaaaa agagaagtgt aattgaggat actggcggg aaaacgctga aagagagaag	2100
tttaagaaga ttattagtct ctatcttacc gttatattatc acatttc当地 aaacatagtc	2160
aacatcaatg ccagatatgt catcgattc cactgcgttg aacgagatgc tcagttgtac	2220
aaggagaaag gctacgacat caacctcaaa aaactggagg aaaagggggt tagttccgtt	2280
acaaggatgtt gcgcggaat tgacgagacg gccccagata aacgaaagga cgttgagaaa	2340
gaaatggcgg aacgagcga agagtccatc gactcttgc agtcagctaa tcctaaatttgc	2400
tatgcaaact atattaaata ctctgatgag aagaaagcgg aggaattcac acgacagatc	2460
aatcgggaga aagcaaaaac ggcactgaat gcatacttgc ggaacacgaa gtggAACGTG	2520
attatccatcagg aggacctgtt gaggatcgac aataaaacgt gtaccctgtt tagaaataaa	2580
gcgcgttcatc tcgaggatggc cccgtacgtg cacgcctata ttaatgacat tgccggagtt	2640
aattcttatt ttcaactgtt ccattacatc atgcagagaa ttatcatgaa tgaacgatac	2700
gaaaagagca gcggcaagt gtctgagtt tttgatgccg tcaatgatga gaaaaataac	2760
aatgacaggc tggtaagct gctgtgcgtt ccatttgcgtt attgtattcc tcgggtttaaa	2820
aatcttagta ttgaggctct ttttgcgtt aatgaagccc caaagtttgc taaggagaag	2880
aaaaaggtat ccggtaacag c	2901

<210> SEQ_ID NO 18
<211> LENGTH: 2388
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus sp., isolate 2789STDY5834894, modified for expression in human cells

<400> SEQUENCE: 18

atggaaatta atactagtaa tcccactcat aggtccgggt aatcttctag cgtacgagga	60
gacatgttg gtctcaatc agagctcgag aagagatttt tcggggaaac atttgatgat	120
aatatccaca ttcaacttat atataatatc cttgatatcg agaagatcct tgcgtctat	180
gtgactaata ttgtctacgc acttaacaat atgctcggtt taaaaggctc agagtccat	240
gacgacttta tggctatct ttcagcacag aatacgtact acatatttac acatcccac	300
aagagcaact tgagcataa agtgaaggc aatattaaga aatctttag taaattcaat	360
gaccttctga agacgaagcg acttggctat tttggctgg aggagccaa aaccaagat	420
aagcgagtgt ctgaagctt aaaaaacga gtgtatcaca tgctggctat agtgggtcaa	480
attcgccagt cagtcttca cgacaagtc aacgaattgg atgagttactt gtattccctt	540
atagacatca tcgatagcga gtatcgagac acatggact acctgggttgc tgaacgatt	600
gattccatta acaaaggatt cggtcagggg aataaggtaa acatctccctt gcttatcgac	660
atgatgaagg gctacgaggc tggatgatata ataagattgt actatgactt tattgtcctc	720
aagtctcaaa aagaatctggg tttcgtata aaaaaattgc gggagaagat gctcgacgag	780
tatggattta ggtttaagga caagcgtat gatagcgttc gctctaagat gtataaactt	840
atggactttc ttctgttctg taactactat cgaaacgacg tagtcgcagg ggaggcactg	900

-continued

gttaggaaac tgaggtttag catgaccgac gacgagaaaag aaggtattta tgccggacaa	960
gccccggaaa tttaggaat gacttgcgaa acatcgccga tcacatgaaac	1020
ggatgtgtga taaaggagct cggggaggcg gatatggact ttgacgagaa aatactggat	1080
tctgaaaaga agaatgcaag tgacccctt tacttcagaa aaatgtatcta catgttgacg	1140
tatTTTGG atggtaaaga gatcaacgat ctgcttacaa cgcttatttc taaatttgc	1200
aacataaaagg agttttgaa gatcatgaaa tccctccgcg tggatgtaga gtgtgagctg	1260
accgcgggct ataaaactgtt taacgattct caacggataa cgaacgagct cttcatagtg	1320
aagaacatcg cttccatgcg caagccggcg gcttcagccaa aattgactat gttccgcgt	1380
gcgcgtacaa tactcgggat tgacgataaa attacggacg accgaatatc agaaatttctt	1440
aaattgaagg aaaaggccaa gggcatccat ggcctgcggaa acttcatcac gaacaacgat	1500
atccgagtcttata gtcgggtttgt ttatcttata aaatacgcgaa atgcgcgaaa aattcgggag	1560
gtcgcaaaaa atgaaaagggt ggtaatgttt gtgcgcgggg ggatcttgcgac cacacagatt	1620
gagcggtaact ataaaaggttt cgttgagttc cctgacatgcg attcttcaact cgaagccaa	1680
tgcagtgcg tggcacggat gatcaagaat atctcccttc atgatTTAA gaacgtaaaa	1740
caacaaagctt aaggacgcgaa aatgtggcg aaagagaggg ccaaggcagt catcggtctc	1800
taccttacag ttatgtaccc ctttgtaaa aaccttgcgaa acgtcaatgc tcggatgt	1860
atagcaatcc actgtttggaa gagagatttcc ggcctctata aggagatcat cccggagctc	1920
gcttcacaaaa acttgaaaaaa tgattatcgat attcttctc aaactcttgcg tgaactttgt	1980
gatgacacggg acgagagtcc taacctgttc ttgaagaaga acaaaggact gcggaaatgt	2040
gtggaggctcg atataaacaa tgccgattctt agcatgaccc ggaaataccg gaattgcatt	2100
gcacacacca cagtggatcg cgagctcaag gaatacatcg gtgatatacg caccgtcgac	2160
tcctactttt ctatctacca ctatgttatgc caacgggtgtaa tcacccaaag ggaggatgt	2220
actaagcaag aagaaaaat caagtatgaa gatgacatgc ttaagaacca tggatacacg	2280
aaagatTTG tgaaaggccct taatagtcca ttccggatca atattccgcg attcaaaaaac	2340
cttccatcg aacaactctt cgatcgaaat ggttacccat ccggagaaa	2388

<210> SEQ ID NO 19
<211> LENGTH: 2862
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Eubacterium
siraeum, modified for expression in Zea mays cells

<400> SEQUENCE: 19

atgggtaaaa agatacatgc acgggacttg cgcgagcaga ggaagacgga ccggaccgaa	60
aaattcgcag accaaaaacaa aaaaagagag gctgaacggg cagtcggccaa gaaagatgc	120
gcgcgtgtcg tcaaatcggt ctcacgcgc tcacccaaaa aggacaatgt taccaatct	180
atggcgaaag ccgcggaggt caagtctgtt ttgcgtgttgc gcaatcggt ctacatgaca	240
tccttcgggc gcggaaatga tgccgttctt gaacagaaaa ttgttgcata ttccacacgaa	300
ccactcaaca ttgtatgaccc agcttataa ctcaatgtgg ttacgtgaa tgggtattca	360
gtgaccgggc ataggggaga aacggtctcg gcagtcacag acaatccctt gagaagattc	420

-continued

aacggcagaa	aaaaggacga	gccggaacaa	tcagtgcgca	ctgacatgtt	gtgtctcaaa	480
ccaaccctgg	aaaagaaaatt	ttttggcaaa	gagttcgacg	acaatatcca	cattcagttg	540
atataataaca	tcctggatat	tgagaaaatt	ttggccgtct	actcgaccaa	cggcatataac	600
getctcaaca	acatgtcagc	agatgagaac	attgagaact	cagactttt	tatgaaacgc	660
accacggatg	agaccttcga	tgacttcgag	aaaaagaaaag	agtccacgaa	cagcagagag	720
aaagctgatt	tcgatgcgtt	cgaaaagttc	atcggcaact	acaggctggc	gtatccgca	780
gatgcatttt	atgtcaacaa	gaaaaatccc	aagggttaagg	ccaaaaatgt	cctccgcaaa	840
gacaaggAAC	tctactcagt	gctcacattt	atcggaaagt	tgccgcatgg	gtgcgttcat	900
tccgaggagg	gtcgggcaga	gttctggctt	tataaactgg	acgaattgaa	ggacgatTTT	960
aagaacgtgc	ttgatgtcgt	ctacaataga	ccagtcgaag	aaattaataa	ccgctttatt	1020
aaaaacaata	aggtcaacat	acaatcttgc	ggatcggatct	ataaaaaacac	cgacatcgca	1080
gagctggatca	gaagctacta	cgagttctg	ataactaaaa	agtacaagaa	catgggcttc	1140
tcaataaaaa	aactgcgcga	atcaatgttt	gaaggtaagg	gatatgcggg	taaagaataac	1200
gattctgtta	gaaacaagct	ctaccagatg	actgacttca	ttctctatac	cggttatata	1260
aacgaagata	gcgacagggc	tgatgacctg	gtcaacacac	tgccggagctc	cctgaaagag	1320
gacgataaga	ccacagtgtt	ctgttaaggag	gccgattacc	tgtgaaagaa	ataccgcgag	1380
tctattaggg	aggtegcgga	cgcctggac	ggtgacaata	ttaaaaaact	ctctaaaagc	1440
aatategaga	tacaagaaga	caaactgcgc	aagtgtttt	tatcttatgc	ggattcagtc	1500
tcggagttca	cgaaactgtat	atatctctg	acacgccttc	tgagcgggaa	ggagattaat	1560
gacttggatga	caactttgat	taacaagttc	gacaacataa	ggagctttct	tgaaatcatg	1620
gatgagctgg	gcctegatag	aacgttcacc	gccccgatct	cgttctcga	gggttcaaca	1680
aaatatcttgc	cggaactcg	tgaattgaat	tcgttcgtt	aaagctgttc	ttttgtatata	1740
aatgccccaa	gaacaatgtt	ccgggacgcg	cttgatatcc	tgggcataga	atcggataaa	1800
accgaggaag	atatcgaaaa	gatgatagac	aatatcttgc	aaatcgacgc	aatgggtac	1860
aagaagcttca	aaaagaataa	cggttgcgc	aattttatcg	cttcgaatgt	catcgatcg	1920
aacaggttca	aatatcttgc	tcggtacgg	aacccgaa	agattagaga	aacagctaag	1980
tgtaagccat	cggtcagatt	tgtcttgcac	gaaataccgg	atgcgcagat	cgaaagatata	2040
tacgaaggcct	gctgccctaa	gaacaccgca	tttgttagcg	cgaataagcg	gcgggagaaa	2100
ctcgctgata	tgatagcgga	gattaaattc	gaaaatttct	cgacgcggg	caactaccaa	2160
aaagctaacc	ttacttcccg	cacttcggag	gcccggatata	aacggaaagaa	tcaaggcgata	2220
attagacttt	atctgaccgt	catgtacatt	atgcttaaga	atctcgatca	cgttaatgct	2280
agatatgtca	tcgccttca	ctgcgtggaa	cgcgataacta	aactgtatgc	cgaatcggt	2340
cttgaagtcg	ggaacataga	aaaaaataag	accaacctta	ctatggccgt	gatgggtgtc	2400
aaactggaga	acggcattat	caaaaactgaa	tttgataaaa	gottcgccga	aaacgcacgc	2460
aatcgctatc	tgcggAACG	aagatggat	aagcttatac	tcgataatct	taagaagtcg	2520
gaaaggccgc	tggtaacga	gttccggat	accgtttgc	acttgaacgc	gatccggat	2580
attaacatca	atataaaga	aattaaagaa	gtcgaaaact	actttgcgt	ctatcattac	2640
ttgatacaga	agcatctcga	gaatcgcttc	gccgataaaa	aggtggagag	ggacacaggt	2700

-continued

gactttat tccaagctcg a agacgataaa acctattgc a aggatttgt taaagcatat	2760
tgtacgccat tcggttataa tcttggtagg tacaagaatc tgacaatcg a cggcttgtc	2820
gataaaaatt atccgggca a ggacgatgc gatgagcaga ag	2862
<210> SEQ ID NO 20	
<211> LENGTH: 2757	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus sp., isolate 2789STDY5834971, modified for expression in Zea mays cells	
<400> SEQUENCE: 20	
atggccaaga agaataaaat gaagccacgc gagctgaggg aggctcaaaa aaaagccgg	60
cagcttaagg ctgcggagat caataataat gctgcccccc ctatcgacgc aatgcccgc	120
gcagaggctca ttgcgcggc cgccgaaaag aaaaaaagct cagtgaaggc tgcaggaatg	180
aagtcaattt tggtagcga gaataagatg tatattacct cgtttggcaa gggaaacagc	240
gcgcgtgtgg aatacgaagt tgataacaat gactataacc agacacagct ttcatcgaa	300
gataattcca acatccaatt ggggggcgtg aacgaagtta atataacgtt ttcttcaaaa	360
catggtttcg aatctggagt cgaaataaat acgtctaattc cgactcatag gtccggtgag	420
tccageccctg tccggggggc catgctcggt ctcaagtccg aactcgaaaa acgggttttc	480
ggtaagactt tcgatgataa tattcatatt cagtttatata acaatatctt ggatatacg	540
aaaattctgg cggtgtatgt cacaatata gtgtatgctc tgaataatata gctcggtgt	600
aaagggttcgg agagccatga tgatttcattt ggatatctt ctacaatataa catctacgat	660
gtgtttatag accccgataa ctcttcctcg agcgatgaca aaaaagccaa tgtgagaaag	720
agccttcga agtttaacgc cctgctcaaa acaaaacgc tggctattt tggattggaa	780
gaaccgaaga caaaagacaa tcgggttcg caggcctaca aaaagcgct gtatcacatg	840
cttgcaatcg tcggcaaat caggcaatgt gtcttcacg acaaaagcgg ggcaaaacgc	900
ttcgacctgt actctttat taataacata gatccggat ataggataac acttgattac	960
ctggcgaag aacgccttaa atccataaac aaagactta tagaagacaa taaagtgaat	1020
atttcttcgc tgatcgacat gatgaaggc tacgaagcgg acgacataat aagggttat	1080
tatgacttta tcgttcttaa gtcccagaaa aatctgggt tttcaattaa aaagcttagg	1140
aaaaaaatgt tggatgagta tggttccgg ttcaaagata agcaatacga ttcagtcaga	1200
tccaaatgt acaagctcat ggacttttt ctgttctgtat attactacgc caatgacata	1260
gcagctggtg a aagcctcggt gaggaaatgg agatttccca tgaccgacga tgagaaagag	1320
ggtattttatc cagatgaggc agccaagctc tggggaaatgt ttagaaatga cttcgagaat	1380
atcgccgacc atatgaacgg ggtatgtcatc aaagagctgg gaaaggcgg tatggacttc	1440
gacgagaaaa tactggattc tgaaaaaaaaa aatgcgagc acctccctta cttctccaag	1500
atgatctata tgcttactta ttcttcgtat gggaaaggaga taaacgacat gctgactaca	1560
cttataatcg aattcgacaa tatcaaaagaa ttccctaaaaa taatgaagtc ttcaagcggtt	1620
gatgtggagt gcgaattgac cgctggttac aagctgtta acgattcgca gcggatcacc	1680
aatgaattgt ttattgtcaa aaatatcgcc tctatgagaa aacctgctgc atctgcgaag	1740

-continued

ctcaccatgt tcagggatgc actcaccata ttgggcattg acgataagat caccgatgac	1800
aggagttctg gtatattgaa gctaaggaa aagggttaagg gaatacatgg tctcagaaac	1860
tttatcacta acaaactcat cgaatcccg cgctttgtct acctgataaa atatgctaac	1920
gctcagaaga tccgggaggt tgcgaaagaaat gaaaaagtgc tcatgttcgt tttggggggg	1980
attccccata cgcaaattga gaggtattat aagtcgtgtg tgcatttcc tgacatgaac	2040
tcatcaacttgc gcgtcaaacgc ctccgaatttgc acggatgtac tcaaaaacat ttcattcgac	2100
gacttcaaaa acgtcaaca gcaagctaa ggccgcgaga acgttgcaaa ggaaaggcca	2160
aaggcagtca taggacttta ctttactgtt atgtacccgc tgcgtttagaa cctggtcaat	2220
gtcaacgcgc ggtatgtcat tgccattcat tgcttggaaac gggacttcgg actttacaaa	2280
gagattatcc ctgaactggc gtcgaagaaac ttgaaaaacgc actaccggat tctgagccag	2340
acgctctgtt aactttgcga caaagggcct aaccttttc ttaaaaaaaaa cgagccgctt	2400
aggaaaatgtt tggaggttggaa tattaacaac gctgatacgct cgatgactcg gaagtaccgg	2460
aattgtatttgcgc accatccat tacgttatgc aaagatgc aacgaaaaga	2520
acgggttact catacttttgcgc accatccat tacgttatgc aaagatgc aacgaaaaga	2580
gagaacgata ctaaacagggaa ggaaaagata aagtatgaag atgacttgct taaaaatcac	2640
ggctacacta aagactttgtt taaagcactc aatagccctt ttggctacaa catacctaga	2700
ttcaaaaatc tqtcaattq aqcaqttttt qacaqaaacq aatatcqac aqaaaaaq	2757

<210> SEQ ID NO 21
<211> LENGTH: 2754
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
bicirculans, modified for expression in Zea mays cells

<400> SEQUENCE: 21

atggcaaaaa agaataagat gaagccgcgg gagcttcgag aggcccaga aaaggcgcgg	60
cagcttaaag cggctgaat taataataat gctgtcccg cgatagccgc aatgcctcg	120
gctgaagcgg cggctccgc ggccgagaag aaaaaatcat ctgttaaagc cgccggatg	180
aaaagcatcc tcgtgtcgga gaataagat tacattacgt cgttcggtaa gggaaattcg	240
gcggtccttg aatacgaagt tgataacaat gattataaca aactcagct ttccagcaaa	300
gacaattcga atattgagct ctgtgacgct gggaaagtga atataacggtt ttcttcccg	360
aggggtttcg agagcggtgtt gggaaatcaat acaagcaatc caactcatcg gtccccggag	420
tcctcctctg tgccccggcga catgttgggg cttaagtccg aacttgaaaa gcggttttt	480
ggaaaaattt tcgacgacaa tatacacatc caacttatct acaacatact ggacatagag	540
aagattttgg cagtgtatgt gaccaatata gtctacgccc tcaacaacat gctgggttag	600
ggcgacgaat caaattacga ctttatgggt tatctgtcaa cttttacac atataaggctc	660
tttacaaacc cgaatgggtc tacattgtcc gacgataaga aagaaaatat aaggaagtcc	720
ctttctaaat tcaacgcgtc ccttaaaaca aagagattgg gctacttcgg ccttgaagag	780
cccaagacaa aggacactcg ggcctcgaaa gcttataaga agagagtcta ccacatgctc	840
gccatagtgg gccaaattag gcagtgcgtc ttccacgaca agtctggtag aaagagattt	900
qatctqtact cattcattaa taatatcgat ccaqaqtaacc qcqaqacatt qqattatctt	960

-continued

gtcgacgaaa ggttcgattc tatcaataag ggtttatcc aaggtataa agtcaacatc	1020
tccctctgta ttgacatgtat gaaaggctat gaagccgatg acatcattag gctgtactac	1080
gactttatag ttctcaaatac acagaaaaac ctggggttct ctattaagaa gcttagagag	1140
aaaatgttgg acgaatacgg tttccgcctc aaagataagc aatacgactc agtgaggct	1200
aaaatgtaca aactcatgga ttttcttctg ttctgtactt actatcgaa tgatatcgca	1260
gccggtaaat ctctcgtagt aaaactcagg ttttcgtatc cggacgacga gaaagaaggg	1320
atatacgcgg acgaagccgc taagttgtgg ggaaaatttc gcaacgattt tgaaaatata	1380
gctgatcaca tgaatgggaa cgttataaaa gagcttggaa aagccgacat ggatttgac	1440
gagaagatat tggactctga gaagaagaat gcgtcagact tgctttatc ttcaaaaatg	1500
atatatatgc tcacgtactt cttggacggg aaggagataa acgtatgtt gacgacgctg	1560
attagcaaat tcgacaataat caaagagtcc ctgaaaataa tgaagagctc agctgtcgat	1620
gtcgagtgtg aactgacggc tggctacaaa ttgtttaacg attcgcaacg cattacgaat	1680
gagctgttta tagtggaaaaa cattgcataat atgcgcaaac cagctgcccag cgctaagctt	1740
acaatgttcc gggacgctct gacgattttgc ggcacatcgac ataaaattac tgacgatagg	1800
atcagcgaga tactgaaattt gaaagagaaa gggaaaggaa ttcacggcct cagaaacttt	1860
attactaata atgtcatcga atcgtcaagg tttgtgtact tgattaaata tgcaaatgca	1920
caaaaagattc gggaaagtcgc taaaaatgaa aagggttta tgtttgcct cggggggata	1980
cccgataaccc aaattgagcg gtattacaag agctgcgtgg agttccaga catgaactcg	2040
tctctggggg tgaaacggtc cgaactcgct cgcacatgatc cttcgacgac	2100
tttaagaacg tgaagcaaca agctaagggg cgccgagaacg tcgcgaaaga aaggccaaa	2160
gcgggttatcg gtctgtaccc tacggatcg tttgtgtgg tgaaaaacct tgtgaatgt	2220
aacgctcggt acgtgatcgc gatccactgt ctggagcgcg attttggct gtataaagag	2280
atcatccccg agctggcttc caaaaacctg aaaaatgact accgcataact gtcccagaca	2340
ctttgcgagt tttgcgacaa gagccgaat ctgtttctgaa aaaaaaaaaa ggcgcctgcgg	2400
aagtgcgttgg aggttgcataat aaacaacgcg gactcctcaat tgacgagaaa gtacagaaat	2460
tgcatacgctc atttgaccgt cgtcaggagat ctcaaaagaat acataggaa catttgact	2520
gtggactcgat attttccat ctaccactac gtgtatcaca ggtgttatcata taagcggaa	2580
aacgatacca aacaagagga gaagatcaag tacgaggatg acctttgaa aaatcacgg	2640
tatacgaagg acttcgtgaa ggcattgaaatc tctccgttc gttataatat cccttaggttc	2700
aagaatttgttccatgataaca gctttcgat cgcaatgagt atcttacaga aaaa	2754

```

<210> SEQ ID NO 22
<211> LENGTH: 2766
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
sp., isolate 2789STDY5608892, modified for expression in Zea
mays cells

```

```

<400> SEQUENCE: 22
atggcaaga agaacaaaat gaagccacgc gaactgagag aagctaaaaa gaaggcgaga      60
cagcttaaag ctgcggagat caataataac gcagctccgg ccattgcccgc aatgcccggc      120

```

-continued

gctgaagtga tagctccagt tgccggagaag aagaaatctt cagttaaagc agctggaatg	180
aaatccattc tcgtctcgga gaataaaaatg tatattacgt ctttcggaaa aggaaatcc	240
gcgggttctcg agtatgaggt ggacaacaac gactacaaca agactcaact gtgcgacaaa	300
gacaactcaa atattgaact cggggacgtt aacgaagtca atataaacatt ttccctcaaag	360
catggattcg gcagcggtgt cgaaattaat acttcaaatac cgacacatag gtctggagaa	420
tcgtcgcctc tcaggggcga tatgcttggt ttgaagtccg aactggagaa gcgggttctt	480
gggaagactt ttgacataa cattcatata caactgatct acaacatact ggatatcgag	540
aaaatcctcg cagtgtatgt cactaatatt gtttacgcct tgaacaacat gctggcatt	600
aaagactctg aatcatatga tgacttcatg gggtatctca gcccaggaa cacatatgaa	660
gtgtttacgc accccggacaa gtctaattcg tctgataagg tcaaggtaa tattaagaag	720
tcaactcagca agttcaacga ctggcttaag acgaaggcgc tcggctactt tgggcttgag	780
gaaccaaaaa cgaaggacac cagagcctct gaggcttata agaaaagagt gtatcatatg	840
ctcgcgatag tcggtcaaatt tagacagtgt gtttccacg ataaatctgg agcaaagagg	900
ttcgacctt actcatttat aaacaatatc gaccctgaat atagagacac gctggattac	960
cttgtggagg agcggctgaa gtcgattaat aaggacttta tagaaggcaa taaagtcaat	1020
atctctctcc tcatagacat gatgaaaggt tatgaagccg acgacataat aaggcttta	1080
tacgattttc tcgttcttaa gtcacagaaa aatttgggtt ttgcgtcaa aaaacttcgg	1140
gaaaagatgt tggagaata cgggttcaga ttcaaagaca agcagtagca tagcgtgagg	1200
tcaaaaatgt acaagctgat ggacttcctg ctgtttgcatttactacag aatgtatgtc	1260
gccgcggggg aggcgttggt tcgcaagctt cgctttcaa tgacagatga tgaaaaagag	1320
gggattttatg cggatgaggc cgccaagctc tggggcaaat ttaggaatga ttttggaaac	1380
attgctgatc atatgaatgg cgtatgtgatt aaggaactgg gcaaaaggaga catggattt	1440
gatgaaaaga tcctcgactc agaaaagaag aatgccagcg atttgttgta ttctcaag	1500
atgatctaca tgctgacgta tttttggac ggtaaagaga taaacgatct gctcacgacg	1560
ttgatttcta aattcgacaa tattaaggag ttctttaaga ttatgaagtc ttccggcagtt	1620
gacgttgaat gcgaactgac tgctggctac aaactcttca acgactcaca acgcatcacc	1680
aatgaacttt ttatcgtaa aaatatagcc agcatgcggaa agccggcagc ttctgccaag	1740
ctcaccatgt ttcgegatgc tttgaccatc ttgggcattt atgacaatat tacagatgt	1800
cggatatctg agatactcaa acttaaggag aaaggcaagg gcatacatgg ctttcggaaat	1860
ttcattacta ataacgtgat agaaagcagc cgctttgtt acctcattaa atacgcaaatt	1920
gcaccaaaaa taagggaaatgt tgctaaaaac gaaaaagtgg tgatgttgcgt gcttggagga	1980
atacctgaca cacaatcgaa ggcgttattac aagtcgtgt tcgaattccc cgatatgaaat	2040
tcttccttgg aggcttaacgc gtcagagctc gccagaatgtca aacaaatcttgcatttgcatt	2100
gacttcaaaa atgtgaaaca gcaagctaa ggtcgcaaaa acgtcgctaa agagaggggcc	2160
aaggctgtta tcggcctcta tcttacgggtt atgtatgttgcgttggaaatcctcgtaat	2220
gtcaacgcac ggtatgttat agcaatacat tgcctcgaaac gggatggggatggatggat	2280
gagattatcc cagaattggc gtcacaaacac ctcaagaacac actatcgat attgtctcag	2340
acgctttgtt aattgtgcga tgaccgcaat gagtcttcca acttggatggaaatgaaat	2400

-continued

aagcggttgc gcaagtgcgt tgaagtggac ataaataacg ccgactcttc aatgactcgc	2460
aagtacagaa attgtatagc gcacccact gtcgtgcggg aattgaaaga atacatcgga	2520
gacataagga cccgtcgatag ctattttagc atttaccact atgtcatgca aaggtgtata	2580
actaaacgcg gtgatgatac caaacaggaa gaaaagatca aatacgaaga cgatctgctc	2640
aagaatcatg gctacaccaa agatttcgtt aaagcattga atagccctt cgggtataat	2700
attcccagat taaaaaacct cagcattgaa caactgttcg accgcaacga atacctcagc	2760
gaaaag	2766

```

<210> SEQ_ID NO 23
<211> LENGTH: 2766
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
sp. CAG:57, modified for expression in Zea mays cells

```

<400> SEQUENCE: 23	
atggcgaaga agaacaaaaat gaaaccacgc gaactcagag aggacaaaaa gaaagcccg	60
cagttgaagg cccgcggat aaacaacaac gcccccccg caattgcggc aatgcggat	120
gccccggatca tcgctccgt cggcgagaag aagaagagct cggtaaggc agccggatg	180
aaatcttac ttgtgtcaga gaataagat tacattacgt ctttcggcaa gggaaatagc	240
gcagtcttgg agtatgaagt tgacaacaac gactataaca aaacacaact ttctagcaaa	300
gacaactcga atatagaatt gggagatgc aatggatc acataaccc ttgtccaaag	360
catggctttg gtcgggtgt ggaaattaac acgtccaatc ctacccatcg gtcggcgag	420
tgcgtcccgat ttagggggaa catgctgggt ctcaagagcg agttggagaa aagattttc	480
ggtaagacct tcgatgataa cattcatc caacttatct ataacatctt ggacatagaa	540
aaaatacttg cagtgtacgt cactaatatc gtttatgcgt tgaataatat gttggaaatt	600
aaggactctg aatcctatga cgatttatg ggctatctga ggcgtcgaa tacctacaa	660
gtgtttatctc atccagataa aagcaaccc ttgtcgataa aagcgataagg tcaaggccaa cataaaaaag	720
tccctgtcaa agtttaacga tcttctcaa accaaacggc tgggtactt tggactcgag	780
gagcctaaga cgaaagacac gccccatct gaggcataca agaaaagggt ttatcatatg	840
ctggcaatag tcggtaaat caggcagtgc gtcttcacg acaagagccg agcgaagccg	900
tttgacctt attcttcat caataacatc gatccggat accgcgacac attggattac	960
ctggtcgagg aagggttgaa gtcataaac aaggactca tcgaggggaa caaggtaac	1020
atttcacttc tgattgacat gatgaaaggc tacgggtctg acgatatcat aagactttat	1080
tatgacttta tcgtgctgaa atcgcagaaa aatttggat tttctatcaa aaagctcaga	1140
gagaagatgc ttgaggagta tggattttaga tttaggaca agcagtacga ttctgtcgc	1200
tctaaaatgt acaagctcat ggatttctc ctctttgcataattacatc gacgtatgtt	1260
ggccgcaggcg aggctttgt ccggaaagctc cgcttctccat gacggacgca cgaaaaggaa	1320
ggcatataacg cggatgaggc agcggaaattg tgggttaagt tcaggaatga ttttggaaat	1380
atagctgatc acatgaacgg tgacgtcatac aaggagctgg ggaaagccga tatggattt	1440
gatgagaaaa ttctggattc ggaaaagaaa aatgcgagcg acttgcgtcta cttagcaaa	1500

-continued

atgatttata tggaccta ttccctcgat ggaaagaga tcaacgatt gcttacgact	1560
ctgataagca aattcgataa tataaaagag ttttgaaaa taatgaagtc ctcagcgtt	1620
gatgttgaat gcgaactgac agccggctat aagctttca atgattcaca gaggattacc	1680
aacgaacttt ttatagtcaa aaacatcgcc tcaatgagga aacccgcgc gagcgcgaag	1740
ttgacaatgt ttagggacgc tctgacgatt ttggaaatcg acgataatat cactgacgac	1800
aggatttcgg agatcctcaa attgaaagag aaggcggaaag ggatccacgg gttgagaaat	1860
tttataacca ataacgttat agaatcatcg aggtttgtt atctgatcaa atacgcgaat	1920
getcaaaaaga tcagggaaatg ggaaaggac gagaaggttt tcatgttgc cctgggtgg	1980
atccctgaca occagataga aagataactat aagtccctgc tgaaattccc tgatatgaat	2040
tttccctcg aggctaaaag atctgagtttgc acggatgttgc tcaagaatat ttctgttgc	2100
gatttcaaaa acgtgaagca acaagctaaa gggcgggaaa acgttgcac ggaacgggct	2160
aaagctgtca ttggcctta cctcaactgttgc atgtatgttgc tggtaagaa ttctgttgc	2220
gttaacgcata gatacgttgc cgctatccac tgcttggagc gcttgcggacttgc actgtacaag	2280
gagattatac cagagcttgc ttccaagaat cttaagaatg actatcgat attgtccaa	2340
actctttgcg agttgtgcga cgatcggaaac ggttgccttca atctgttgc taagaaaaat	2400
aaaaggctgc ggaaatgcgt cgaagtcgac attaacaatg cggattcttc tatgacgaga	2460
aagtaccgca actgtatcgcc ccatctcactgcg gttgtcaggag agctcaagga atacatagga	2520
gacattagaa cgggtggactc atattttca atataccatt atgttatgca aaggtgttatt	2580
acaaaacggg gggatgacac aaaacaagag gaaaagatta aatatgaaga cgatttgc	2640
aagaaccatg gttacacgaa agatttcgtt aaagcgttta attcgccatt tggtaat	2700
attcccgat taaaaattt gggatagag cagctttcg atagaaatga atacttgacc	2760
gagaag	2766

<210> SEQ ID NO 24
<211> LENGTH: 2799
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus flavefaciens FD-1, modified for expression in Zea mays cells

<400> SEQUENCE: 24

atgaaaaaaaaa agatgagctt gcggggaaaaa agagaggcag aaaagcaggc caagaaagct	60
gcatacagcg ctgcgtctaa gaacactgtat tccaaaccac cggagaaaaa agcggagact	120
ccaaaacctg ccggaaattat atctgataac tcacgcaata agacggcggt caaggcagcg	180
ggactcaagt cgacgatcat atcaggcgtt aaattgtata tgaccagctt tggcaaggc	240
aatgcagctg tgatagaaca aaagatagac atcaatgact attcttttag cgcaatgaag	300
gacaccccaa gccttgaatg cgacaaggca gaatctaagg aaatatcctt ttctgtccat	360
catccctttg tgaagaacga caagttgacg acatataatc ctctttacgg tggaaaggat	420
aacccagaga agccgggtgg gcgcgatatg ttgggggttga aagataaact tgaggaacgg	480
tactttgggtt gtacattcaa tgacaacctc cacattcaga tcatttacaa tattttggat	540
attgagaaga tcctcgctgt tcattccgca aatattacga cagctttga tcataatgg	600
gatgaggacg atgagaaata ccttaactct gactatatcg gctacatgaa cacgtcaac	660

-continued

acctaegacg tcttcatgga tccctctaag aattccttctt tgtcgc当地 agacaggaaa	720
aacatcgaca attcgagggc gaagtttgag aagctctct ctacaaaaag gttgggtac	780
tttgggttcg actatgacgc gaacggaaa gacaaaaaga agaatgagga aattaaaaag	840
cggcttacc acttgacgac atttgcaggc cagctgaggc agtggccctt ccactcagca	900
ggaaactatc ccagaacctg gttgtataaa ttggactccc tggataaaga gtatctggac	960
acgctcgacc actatccga taagggttt aatgatataa atgacgat ttgtcactaaa	1020
aacgcaacga acctgtat atctgaggag gtttccctg aggctaactt taaagatatt	1080
gcccacttgt attatgactt tattgtcatc aagtccacaca agaacatggg attctcgatc	1140
aagaaacttc gggaaaaat gctcgagtgc gatggagctg accgcatcaa agaacaggat	1200
atggattctg tccgctccaa gctctacaag ctcattgatt ttgcataatt caagtattac	1260
catgagttcc cagagctcag cgagaagaac gtcgacatcc tgagggctgc cgtgagcgat	1320
actaagaagg acaatctcta ctcagatgaa gctgctcggt tgggtcaat ttcaaggaa	1380
aaatttctcg gatTTTGTGA caaaatttggt gtttgggtga ccggagagca tgagaaagat	1440
atcacgtctg tcattgataa agacgcctac aggaacagaaa gcaatgtctc gtatTTTCA	1500
aagctcatgt acgcaatgtg ttttttctt gatggaaagg agataaaacga ccttctgact	1560
accttgatta acaagttga caaatcgcc aaccagat agacagcaaa ggaattgggg	1620
atcaacacgg cgttcgtaa aaactatgac ttctcaacc attctgagaa atatgtcgac	1680
gaattgaaca tagtggaaaa tatcgctcg gatggggaaa ccttcaaaa cgcggaaaa	1740
gtatgttacc atgacgcctt tactattttt ggcatttcctg aagatgtgaa cgaaaaggct	1800
ttggatgaag aactcgacct tatactcgaa aaaaagaccg atcccgtcac aggtaaaccg	1860
ctgaagggtt acaatccctt ggcataatcc atagctaaaca acgttataaga gaactctcg	1920
ttcatctacc ttataaaatt ctgtaatccg gaaaacgtga gaaaatttgt gaataacact	1980
aagggtgacag agttcggtct gaaacgcata ccagatgccca aaatttggagag gtattacaaa	2040
tcttgcgttcc atagcgatgat gaaaccctccg actggggggggg aaatttccgaa gttggcttgt	2100
aaactttaaag acatgaactt cggcaacttc cggaaatgtcc ggcagtctgc aaaagagaat	2160
atggagaaag agagggtttaa agccgtcatt ggactgtacc ttaccgttgt gtacagggtt	2220
gttaagaatc tcgtcgacgt gaaactcaaga tacattatgg cattccatc actcgagaga	2280
gactccaaat tgtataacgt ctcagtcgac aacgattatc tggcactgac cgatacactg	2340
gtcaaaagggtt gtcgacactc acgctcacgg tacttggccg ggaataaaag attgcgggat	2400
tgtgtcaaac aggttatttga taacgcaaaa aagtggtttgg ttagcgataa atataattcc	2460
ataaccaagt ataggaacaa tggcgccac ctgaccggcc ttcggaaactg tgccgaaattt	2520
ataggcgaca taacgaaagat tgactccctac ttccctctt accactactt tatccagcg	2580
caactcgcca aagggtctcgta tcatgagagg tcagggtttt accgcaat tccacagtg	2640
gcaccactgt tcaagtggca tacttatgtt aaagatgttgg tggaaagcgct gaatgcaccc	2700
ttcgggtata atattccaaat gttcaagaat cttccatttgc acgactctt cgaccggaaat	2760
gagatcaaga agaatgtatgg agaaaagaaa tctgacgac	2799

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
albus strain KH2T6, modified for expression in Zea mays cells

<400> SEQUENCE: 25

atggccaaa aatctaaagg catgtccctg agggaaaaac gcgagctgga gaagcaaaag      60
cggatccaga aagctgcagt gaactctgtc aacgacactc ccgaaaagac cgaggaagca     120
aacgttgtt ctgtcaatgt gagaacgtct gcggaaaaca agcacagcaa gaagagcgct     180
getaaagctc ttggacttaa atcggggttg gttattgggg acgaattgta cctcacatca    240
tttggcagag gaaatgagggc gaaactcgaa aagaaaataa gcggggatac cgtggaaaaa    300
ttgggcattt gtgcatttcga agtggcggaa agggatgagt ctacactcac acttgaatct   360
gggcgcattt aagataaaaac tgccagacccg aaagatcccc gacatattac agtggacaca  420
caagggaaat ttaaggaaga tatgctcgga atacgctctg tgcttgagaa aaagatattt   480
ggtaagacct tcgatgacaa catccatgtc caacttgcgt acaatatcct cgatgtcgag  540
aaagatcatgg cacagtaacgt ctctgacatt gtttacatgc tccacaaacac cgataagacg 600
gaacgcataatg acaaccgtat ggggtatatg tccatcgagaa atacttacaa aacctttgt 660
gataacttcca accttccggaa cgatacaaaa caaaaggctcg agaatcaaaa acgggaattc 720
gacaagataaa ttaagtctgg ggcgttggaa tactttggcg aggcatatg ggtcaactcc 780
ggcaactcta caaaaattgcg gcctgagaaa gaaatctatc atatttcgc tctcatggcc 840
tcacttaggc agtcctactt ccacgggtat gtgaaggaca cggactacca aggaacaacg 900
tggcgtaca cattggagga caagttgaag ggcccggtcac acgagttcag agaaacaatt 960
gataagatattt ttgtatggg attctctaag atatcaaagg acttcgggaa aatgaacaaa 1020
gttaatctgc aaattctggaa gcagatgata gggagctgt acggatctat tgagcgcag 1080
aatctcacat gtgattacta cgacttcata caatttgaaga aacataagta cttggggttc 1140
tctataaagc ggttggagaa aacgatgtt gaaacgacac cggcggaaatg ttacaaggca 1200
gaatgttaca atagegagcg gcagaagtt tacaaacttta tagattttct gatctatgat 1260
ttgttactata accgcaagcc ggcgcggatc gaggaaatttgcgataagct tagggagttc 1320
gtgaacgtatc agggaaaatg atcgattttgcgatgtt acggtcgaa ctaagtatgt ctatgttcc 1380
ctctccaaag tgctggataa gtcctcaag aactccgtt cggggagac catcaaagat 1440
ctccagaaaaa ggtatgttgcgatgtt aatagaatat gggacatctc gcaacactcg 1500
atttctggaa acgtcaacttgcgatgtt ttgtatctaca taatgaccctt catgtggac 1560
ggggaaagaaa ttaacgaccc ctttacaacg ctcgtgaaca aattcgataa tattgttca 1620
ttcattgtatc ttatggacga attggggtttgcgatgtt aacactcat ttactgttataa 1680
tttgcgatgtt aatggcttat ctcgttgcgtt cttcaatttta ttaattcggtt tgacggatgtt 1740
agcaaaaatcg acgtaaaaatcg atctaagcgca caattgttta gggacgttctt ggttcatctc 1800
gacataggca ataaggacga gacctggata aataactact tggactccga tattttcaaa 1860
ttggatataaag agggaaataa gttgaaggcgc gcaaggcatg actttcgaa ctttattgtt 1920
aacaacgttgcgatgtt aacgttgcgtt aacactgttgcgatgtt aacactgttgcgatgtt 1980
atgataaaaac tgaaaactaa cggaaaagctt ataggctttgcgtt ctttggacaa gctccctgag 2040

```

-continued

acacagatag atagatacta cgaatcgtgt ggacttgate atgcgttgtt cgacaaaaaa	2100
gtcaggatcg agaagctgtc agggcttata cgcgacatga aatttgatga tttctccggt	2160
gtcaaaaacat caaataaggc gggcgataac gataagcaaa acaaagcaaa gtatcaggca	2220
attatcagct tgtaccttat ggttctgtac caaattgtga aaaacatgtat ctatgtcaat	2280
tcacggtacg tgatcgcgtt ccattgcctt gagagggatt tcggcatgtc cgaaaaagac	2340
ttcggaaat attaccaggg atgttagaaaa ttgactgacc atttcataga agagaaatat	2400
atgaagaaag ggaaacttgg ttgcaataag aagggtggaa ggtatctcaa aaataatatt	2460
tcatgctgtc cggatggctt gatcaataacc tataggaacc aagtggacca tttcgctgtt	2520
gttcggaaga tagggaatta tgcagcataat atcaaatactc tccggctcatg gtttgaactg	2580
tatcaactacg tcattcagag gatcgtgttt gatgagtaca gatttgcact gaataatacg	2640
gagagcaact acaagaatc aatcattaag caccatactt attgcaaaaga catggtaag	2700
gtctcaata ogcctttgg gtatgacctc cccagatata agaatctctc catcgggat	2760
ctttcgata gaaacaatta tcttaataag acgaaggaaat cgatagatgc taattccagc	2820
attgactcac ag	2832

<210> SEQ ID NO 26
<211> LENGTH: 2901
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus flavefaciens strain XPD3002, modified for expression in Zea mays cells

<400> SEQUENCE: 26

atgatecgaga agaaaaagtc ttgcgaaaaa ggaatggag tcaagtctac attggttct	60
ggttcgaagg ttatatatgac gacgttcgcc gagggctctg acgcgcgcctt ggagaagata	120
gtggaggggg attcaatacg gtcgtgtac gaaaggcgaag cttttcgcc cgagatggcg	180
gacaagaatg cagggtataa aattggaaat gcaaagtttt cgcaccccaa aggttacgca	240
gtcggtgcga ataaccgcct ctatactggt ccagtcacg aagatgtct cgggctgaaa	300
gagaccctcg agaaacgcta ttttggag agcgccgatg ggaatgacaa tatatgtatc	360
caagttatac ataattatct ggatatcgaa aagatcctt ctgaatacat taccaacgct	420
gcttatgcgg tcaacaatat ttcccggactt gataaagata taatcgccctt cggtaaattc	480
agcactgtct atacatacga tgagttcaag gatccagacg atcatagacg ggcgttcaat	540
aataacgaca aactgattaa cgcaattaaa gcgcaatatg acgagttcga caatttctc	600
gacaacccac ggcttgctt ctttggccag gcattttctt cgaaggaggg taggaactac	660
ataatcaatt atggcaatga atgctatgac atacttgctc tgctttcagg tctcagacat	720
tgggtcggtt acaataacga agaagaatct cggatctctc ggacttggct ctataacatt	780
gacaagaacc ttgataacga gtacatctct acgctgaact acctttacga cagaatcact	840
aacgagctca ccaattcatt ctccaaaaat tctgcccac acgtcaacta catcgccgaa	900
acccttggaa tcaacccacg agagttgtctt gaacagtattt ttccgttctc aatcatgaaa	960
gaacagaaaa atctgggctt caatataacg aaactgcgcg aggtcatgtt ggatagaaaa	1020
gatatgtccg aaatcaggaa aaaccataaa gtcttcgact caataaggac caaagtgtat	1080

-continued

accatgatgg atttgtcat ctaccgctat tacatagagg aggtatgcaaa agtcgctgcc	1140
gctaacaaga gccttccaga taatgaaaag tctctgtcgaa aaaaggatataatttgatt	1200
aatctccggg gaagctttaa cgacgatcaa aaggatgccg tgtactacga tgaggcaaac	1260
agaatttggaa ggaagctgga aaacattatg cataacatta aggagttccg cgaaaataaa	1320
acgaggaaataa agatgctccg aggttgcctc ggattcttc tgctggtagg	1380
gatgtttcggtt catttcgaa gctgtatgtac gcaactcacca tgttccttga cggttaagag	1440
atcaacgatc tcttgacaac gcttattaaat aagtttgata atatacagtc tttcccttaag	1500
gttatgcggg ttattggagt taatgctaaa ttctgttggaaag agtatgtttt cttcaaggac	1560
agcgcgaaaa ttgtctgacga actcgccctt atcaagtctc tcgcgcggat gggagagcct	1620
atagctgacg ctcgcagggc aatgtatatac gacgcccattcc gcatccttgg caccaatctg	1680
agctatgtatg agcttaaagc ctcgcggac accttcagcc tggacgaaaaa cggcaacaaa	1740
ctcaagaagg gcaaggcacgg catcgcaat ttctattatca ataacgtgat ctcgaataag	1800
agatttact atctgatacg gtatggcgcac ccggccccacc tccatgagat tgcaaaaaaac	1860
gaagctgttg tggaaatttgt gcttggtaga attgcggaca tacaaaaaaa acaaggccaa	1920
aatggccaaa atcaaattga cagatattac gaaacatgc ttggaaagga taagggaaag	1980
tctgtgagcg agaagggttga tgctgttgcacc aaaataatca caggaatgaa ttacgatcg	2040
ttcgataaaa agaggtcagt gatagaagac acggggggggg aaaacgctga acgcgaaaaa	2100
tttaagaaaa taatttcgct ctatcttacg gtcattttatc acatcttgcgaa gaatatagtc	2160
aatatcaacg ctagatacgt gattggtttc cattgtgtgg aaagagacgc tcaactgtac	2220
aaggaaaagg gttatgatat aaacctcaag aagctggagg aaaagggttt tagctcggtg	2280
actaaattgt gcgctgaat cgtaaacc gcccggatataaaagga tggtgagaag	2340
gagatggccg agagagcgaa ggaatctatc gacagcctgg aaagcgccaa tcccaactt	2400
tatgccaact acatcaagta ctctgacgag aaaaaagcggg aagagtttac tagacaatc	2460
aatcgggaga aagctaagac cgcctcaat gcttacttgc gcaataccaa atggAACGTT	2520
atcattcgccg aagaccttgcgcatatgat aataaaacat gtacattttt tagaaataaa	2580
gcagtcgacc tcgaggctgc cagatacgtt cagcatata taaatgacat cgctgaggtg	2640
aactctgtact ttcaagttgtt ccattacatt atgcaaaaggatc tcaatgttgcgaa cgaaaggatc	2700
gagaaatcggtt caggtaaagt ttccgaatat ttgcacgcag tcaatgttgcgaa aaagaaggatc	2760
aacgaccggc ttgttggaaat ttgcacgcag ctttcgggtt actgttatccc tcgggttcaaa	2820
aacctgttcca tagaggcatt ttgcacgcag aacgaccggc caaagtttgcgaa caaggaaaaag	2880
aaaaagggtt cgggttactc g	2901

```

<210> SEQ_ID NO 27
<211> LENGTH: 2388
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM DNA sequence from Ruminococcus
sp., isolate 2789STDY5834894, modified for expression in Zea
mays cells

```

```
<400> SEQUENCE: 27
```

```
atggaaatatacgttccaatccgcacacac agatcaggcg aatcttcctc agtttagaggt
```

60

-continued

gatatgttgg gacttaaatc cgaattggaa aagagggttt ttggcaagac attcgatgtat	120
aacattcaca tacaacttat atataacatc ctgtatatag aaaagatact tgctgttat	180
gtgacaaaca tagtttatgc actgaacaac atgcttggcg tgaagggatc agaaaagctac	240
gatgattca tgggttacct ctccgcttag aacacctatt acatattcac gcacccagat	300
aaatctaacc tgcggataa agttaagggg aatattaaga agtcgtttc taaatttaac	360
gaccttctta agacaaaaag actgggctac tttgggttg aggagccaaa gacgaaagac	420
aaacgggtta gcgaggcata taaaaagagg gtttatata tgcttgcatt agtggccag	480
atacgccagt ccgtcttc tgataaatct aacgagttgg acgagtatct ttactcttc	540
atcgacatca tcgactccga atatagagac acgctcgact atcttgtcga cgaacggttt	600
gattcgataa ataagggttt tgtccaaggc aacaaagtca atatatact cctcatagat	660
atgatgaaag gatacgaagc agacgatata atcagacttt attacgactt tattgttctt	720
aagagccaga aaaatcttgg attctcaata aagaaactga gggagaaaat gttggacgag	780
tatgggttgc ggtttaaaga taaacaatat gactcggcga ggttcaagat gtacaagctt	840
atggactttc ttttgttctg taattactat aggaatgacg ttgttgcgg ggaggcctt	900
gttagaaaat tgagattcag catgaccat gacgaaaaag aaggcatcta tgccggatgag	960
gcagagaagt tggggggaa atttaggaat gacttgaaa acatagccga tcatatgaat	1020
ggcgatgtca taaaggagtt gggaaagct gacatggatt ttgacgaaaa aatctggat	1080
agcgaaaaaa agaatgcttc cgatctgtt tatttctcta agatgatcta tatgetcact	1140
tactttctgg acggtaaaga gatcaacgc cttcttacta cccttatttc aaagttcgat	1200
aacattaagg aatttctgaa aataatgaaa tcctcggtcg tcgacgttga atgcgaactt	1260
actgcagggt acaagctgtt taacgactcg caaaggatttta ctaatgaact gttcattgtc	1320
aagaacatag cgtccatgag aaacgctgca gcaagcgaa agctgacgat gttccgcgt	1380
gctctcacca ttctggaaat tgatgacaag attaccgatg accgcatttc ggagatcctt	1440
aagcttaagg aaaaggggaa ggggattcac ggactgagaa attttatcac caataacgtg	1500
atcgaatcgt ctaggtttgt ctatgtata aagtatgcca atgcgaaaa aattcgcgaa	1560
gtcgccaaaga atgagaaggt cgttatgttc gtgctcgag gaattcccgat tacacagatt	1620
gaacggtaact ataaatcctg tggaaattc ccggatatac actcatccct cgaggccaaa	1680
tgctctgagc ttgcgaggat gatcaagaat atctccttgc atgatttaa aaacgtgaag	1740
cagcaggcga agggccggga gaatgtggcg aaggagcggg ctaaagctgt gataggcctt	1800
tatcttaactg ttatgtaccc tctcggtgaa aacctcggtgaa atgtgaacgc caggtacgtt	1860
atagcgatcc attgtcttgc ggcgcacttc ggtttgcata aggagataat tccagagctg	1920
gcatcgaaga acctgaaaaa cgattacaga attctgtcac aaactctctg tgaactctgc	1980
gatgaccgcg atgagtcacc gaatctcttc ctcaaaaaaa acaagaggct gagaaatgt	2040
gtgaaagttg acatcaataa cgcggattcg agcatgacac gcaagtacccg gaattgtatt	2100
gctcatctca cagtcgtccg cgagctcaaag gatgtatatac gtgatgtccg gaccgttgc	2160
tcttattttt ctatctatca ttacgttgcg cagcggtgc ttacaaaaag ggaagatgtat	2220
acccaaacaag aagaaaaat aaagtatgag gatgacttgt tgaaaaatca tggatataact	2280
aaagactttg tcaaggctct caactcacccg ttccggatcaca acatacccg attaaaaaac	2340

-continued

ttgtcaattt aacagggttt tgaccggaaac gaataacctga cagaaaaaa	2388
<210> SEQ_ID NO 28	
<211> LENGTH: 2865	
<212> TYPE: DNA	
<213> ORGANISM: Eubacterium siraeum	
<220> FEATURE:	
<221> NAME/KEY: misc_feature	
<222> LOCATION: (1)..(2865)	
<223> OTHER INFORMATION: native CasM DNA sequence from Eubacterium siraeum	
<400> SEQUENCE: 28	
atgggtaaga aaatacacgc acgagatctc agagaacaaa gaaagaccga tagaacggaa	60
aaatttgcag atcagaacaa aaaacgtcaa gcagagaggg cagttccgaa aaaagacgca	120
gccgttctg taaaatcagt ttcttctgtt tcataaaaaa aagacaatgt aacaaaatct	180
atggctaaag ccgcaggcgt gaagtcggtt tttgctgttag gaaataactgtt ttatatgact	240
tcatcggca gaggaaacga tgctgtactt gagcagaaaa tagtcgatac atcgcacgaa	300
ccgctgaata ttgacgatcc tgcataatcg ttgacgttg tcacaatgaa cggttatcc	360
gttaccggtc acagaggtga aacggtatct gccgtaacgg ataatccgt gcccgtttt	420
aacggaagaa agaaagatga accggaacag tctgtgccta cggatatgtt gtgcctgaaa	480
ccgactctt aaaaaggatt ctccggaaaaa gaattcgtatc ataataataca tatccagtt	540
atttacaata ttcttgacat tgaaaaataa ctggcggtt attcgaccaaa cgcttattac	600
gcattgaata atatgagtgc tgacgaaaat atcgaaaaca gcgatttctt catgaaacgt	660
accaccgatg aaaccttga cgattttgaa aagaaaaagg agagtacaaa cagtcgagag	720
aaagccgatt ttgacgcatt tgaaaaattc atcggcaattt acaggctggc ttatggcc	780
gatgcatttt atgtaataaa aaagaatccc aaaggttaag caaaaaatgt tctgcgttag	840
gataaaagaac ttactccgt gctactctg atcggtaaac tgcgtcattt gtgtgttac	900
agtggagggg gcagagcaga attctggctg tataagctcg atgaacttaa agatgatttc	960
aaaaatgtac tcgacgttgt ttataaccgt cctgttgaag aaataaaacaa ccgtttata	1020
gaaaacaata aggttaacat acagatactg ggctcggtat acaagaacac cgatattgcc	1080
gaacttgtaa ggtcatatta cgaatttctt atcacaaaga agtataaaaa tatggcctt	1140
tcaataaaaga agctccgtga ggttatctg gaaggttaag gttacgcccga taaaagatata	1200
gattctgtaa ggaataagct gtatcagatg acggatttca tcttatacac aggatatac	1260
aacgaagaca gcgatagagc cgacgatctt gtgaaacactt tgagaagttc gctcaaagag	1320
gatgataaga caaccgtata ttgcaaggaa gctggattatc tggggaaaaa ataccgtaa	1380
tccataagag aggttgcgca tgcgttcat ggcgataaca taaaaaagct gagcaaatcg	1440
aatattgaaa ttccggaaaga caagctgaga aaatgtttt tcaagctatgc cgacacgtta	1500
tccggatattt ccaagcttata ttatctgtcg acaagatttt taagcggtaa ggagatcaac	1560
gatcttgcata caacgctgtat aaacaagttt gacaatataca gaagcttccct tggaaataatg	1620
gacgagcttg ggcttgacag gaccttcacc gcccgttaca gcttcttgc aggcgttaca	1680
aagtatcttcccgacatggatgttgcgat gatgttgcgat aatatccttc agatcgacgc aaacgggtat	1740
aacgcggaaaaa gaacaatgttca tcgcgttgcgat ctggatattt tcggcatttgc atcgatggata	1800
accggaaagatatttgcgat gatgttgcgat aatatccttc agatcgacgc aaacgggtat	1860

-continued

aaaaagctca agaaaaacaa cggctcgaga aattcattg caagtaacgt tataaatatca	1920
aaccgattca agtaccttgt gccggtaacgga aatccaaaga agatccgtga aacggcaaaa	1980
tgcaggcccg ctgttaagggtt tgcgtgtatc gagatcccg acgcacagat cgaaagatat	2040
tatgaggcctt gttccccaaa aaatacagct ttatgctctg caaataagag acgtgagaaaa	2100
ctggctgata tgatagctga aataaagttt gagaattttt cggatgccgg caattatcg	2160
aaagcaaataatc tacatcaag aacgtctgaa gctgaaatca agcggaaagaa tcaggctata	2220
atccgtcttt atcttaccgt tatgtacatt atgctgaaatc accttgtaaa tgtgaacgcc	2280
agatacgtta tcgcttcca ttgcgttcaa agggatacga agctgtatgc ggaaagcggt	2340
ctggaaagtgc gtaatataga aaaaaacaag acaaataatc ctatggctgt aatgggagtc	2400
aagctcgaaa acggaatcat aaaaacggaa tttgacaaga gctttgcaga aatgccgca	2460
aacagatatac tcaggaatgc acgctggatc aagctgatac tggataattt aaagaagtgc	2520
gaaagagcggtt ttgtcaatga gttcagaaat actgtctgccc atctgaaatgc gataaggaat	2580
atcaatataatca atatcaagga aataaaagag gtcgagaact actttgctct gtaccactac	2640
ctcattcaga aacatctcgaa aatcgaaaaa gccgataaaaaa aagtagaaaaa agacaccggc	2700
gattttataa gcaagctcgaa agaacacaag acttactgca aggacttgtt aaaagcatat	2760
tgtacgcctt tcggatataa ctttgtgaga tataaaaacc ttacgataga cgggctgttt	2820
gataagaatt accccggaaaa agacgattct gatgaacaga aataaa	2865

```

<210> SEQ_ID NO 29
<211> LENGTH: 2760
<212> TYPE: DNA
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1) .. (2760)
<223> OTHER INFORMATION: native CasM DNA sequence from Ruminococcus
sp., isolate 2789STDY5834971

```

<400> SEQUENCE: 29	
atggcaaaaa agaataaaaat gaagcctaga gagctgegtg aggctcagaa aaaagccaga	60
cagctaaag cggctgagat aaataataac gctgctctg caatcgctgc catgcctgct	120
gcagaggctca ttgcacctgc ggcagagaag aaaaaatccct ccgtaaaggc ggcaggaatg	180
aagtctattt ttgtcagcga aaataaaatg tacataacccct ctttcggcaa gggcaattct	240
gctgtgtttt aatatgagggt ggataataat gactacaacc aaactcagct ttcttcaaag	300
gacaacacgca atatccagct tgggtggta aacgaagtaa acatcacttt ttcaagcaag	360
catggctttt agagcggagt ggaataaaac acttcaaacc ctactcacag aagcgggtgaa	420
agctcgccctt taagaggggaa tatgtgtgggg cttaaatcgg agcttggaaaa ggcgttttc	480
ggcaaaaactt ttgatgataa tatacatatc cagtttattt acaacattct ggatatcgaa	540
aagataacttg cgggttatgt aacgaatatac gtttatgcgc tgaacaatata gctcggtgta	600
aagggttcag aaagtcatga cgattttattt gggttatctt ccacaatataa tatttatgat	660
gtttttattt accctgataa cagcgttta tctgtatgata agaaagcgaa tgtcagaaaa	720
agccttagca agttcaatgc cctgctgaaa actaaggcgcc ttggcttattt cggcttgc	780
gagccaaaga cgaaagataa tagagttcg caagcttaca aaaagcgtgt ttatcatatg	840

-continued

```
<210> SEQ ID NO 30
<211> LENGTH: 2757
<212> TYPE: DNA
<213> ORGANISM: Ruminococcus bicirculans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(2757)
<223> OTHER INFORMATION: native CasM DNA sequence from Ruminococcus
      bicirculans

<400> SEQUENCE: 30
```

-continued

atggcaaaaa agaataaaat gaagcctaga gagctgegtg aggctcagaa aaaagccaga	60
cagctcaaag cggctgagat aaataataac gctgttctg caatcgctgc catgectgct	120
gcagaggctg ctgcacctgc ggcagagaag aaaaaatccc ccgtaaaggc ggcaggaaatg	180
aagtctattc ttgtcagtga aaataaaaatg tacataaacct ctttcggcaa gggcaattct	240
gcgggtcttg aatatgaggt ggataataat gactacaaca aaactcagct ttccctcaaag	300
gacaacagta atatcgagct ctgtgatgtaa ggcaaaagtaa acatcacttt ttcgagcaga	360
cgtggctttg agagcggtgt ggagataaac acttcaaacc ctactcacag aagcggtgaa	420
agctcgtctg taagagggga tatgctgggg cttaaatcgg agcttggaaa ggccttttc	480
ggcaagaatt ttgatgataa tatacatatc cagcttattt acaacattct gnatatcgaa	540
aagataacttg cagtgtatgt gacgaatatc gtttatgcac tgaacaatat gcttggggaa	600
ggcgcgatgaga gcaattacga tttcatgggg tatctttcca catttaacac ttataaagtt	660
tttactaatac ctaatggcag cactttatcc gacgataaga aagagaatat cagaaaaagt	720
cttagcaaatac tcaatgccct gctgaaaact aagcgttttgc gctatttcgg ccttgaagag	780
ccaaagacaa aggatacaag agcttcggaa gcatacaaaa agcgtgttta tcataatgctt	840
gcaatttgtgg ggcagataag acagtgtgtt tttcatgata aatcgggtgc aaaaagattt	900
gacccttatac gttttattaa caatattgtat cccgaataaca gagaaaccct tgactatctt	960
gtagatgaga gatttgattc tataaataag ggctttatcc agggcaacaa ggtcaatatc	1020
agcttgccta ttgatgatgtat gaaaggctat gaggctgatg atatcatacg cctttattac	1080
gatttcatttgc tgcttaatac tcagaaaaat ctcgggtttt ctatcaaaaa gcttcgtgag	1140
aaaatgctgg acgaatacgg cttcagattt aaggacaagc aatatgactc tgtgcgtca	1200
aagatgtaca agcttatggat tttctgtttt ttctgcaact actacagaaa tgacattgcc	1260
gcaggcgaat ctcttgcgc caaactgcgt tttcaatga ccgatgatga aaaagagggg	1320
atatatgctg atgaagcggc aaagctttgg ggcaaaattca ggaatgattt tgaaaatatc	1380
gccgaccaca tgaacgggtga cggttatcaag gagcttggca aggctgacat ggattttgat	1440
gagaaaatttc ttgacagcga aaagaagaat ggcgtctgacc ttttgtattt ctccaaaatg	1500
atatatatgc tcacatattt tcttgacggc aaggagataa acgacccct tacaacgcctt	1560
atcagcaagt ttgataacat caaggagttt ttgaagataa tgaaaagctc tgctgttgat	1620
gttgagtgatg aacttacggc gggctacaag ctgttcaatg acagccagag gataaccaac	1680
gagcttttta tcgttaaagaa cattgttcc atgagaaagc ctgcggcttc ggcgaagctt	1740
acgatgttcc gtgacgcact gactatactc ggtatagacg acaagatcac ggacgatagg	1800
ataagcggaga ttctaaaact taaagaaaaa ggcaaggcga tacatggcct gagaatttc	1860
ataacaaaca atgttatcga gtcctctcg tttgtatacc ttatcaagta tgcaacgc	1920
cagaagataa gagaagtggc taagaatgag aaagttgtca tgtttggatc tgggggtatc	1980
cctgacacgc agatagagcg ttattacaag agttgtgtgg aatttcctga catgaacagt	2040
tctttggag taaagcgcag tgagcttgcg agaatgataa agaacatcag ctttgcgtat	2100
ttcaaaaatg tgaaacagca ggcaaaaggc agagaaaaacg tggctaagga gaggcggaaag	2160
gctgttatcg ggctttatct tacggctatg tatctgtgg tgaaaaatct tgcgtatgc	2220
aatgcggat atgttattgc gatacactgc cttgaacgtg attttgggct gtataaggag	2280

-continued

ataattcctg agttggcttc aaagaacttg aaaaatgact acaggatact ttcacagacg	2340
cttgcgtgaac ttgtgataa gtcgccaaat ttgttcttga aaaagaacga gcggctgcgc	2400
aagtgcgttg aagtgtatcaataatgca gacagcagca tgacaagaaa ataccgcaac	2460
tgtattgctc atcttactgt agttcgtgaa ctgaaagaat acataggaga tatttgtaca	2520
gtggattctt acttctccat ttatcattat gttatgcagc gctgtatcac gaaaaggaa	2580
aatgacacaa agcaagaaga gaaaataaaag tatgaggacg atctttaaa aaatcacggc	2640
tatacgaaag actttgtaaa ggctctcaac tcgcccgttg gataacaacat tccgaggttt	2700
aaaaatctt caattgagca gttgtttgac agaaatgaat atcttactga aaagtag	2757

<210> SEQ_ID NO 31
<211> LENGTH: 2769
<212> TYPE: DNA
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(2769)
<223> OTHER INFORMATION: native CasM DNA sequence from Ruminococcus
sp., isolate 2789STDY5608892

<400> SEQUENCE: 31	
atggcaaaaa agaataaaat gaagcctaga gagctgcgtg aggctcagaa aaaagccaga	60
cagctcaaag cggctgagat aaataataac gctgctctg cgatcgctgc catgectgct	120
gcagaggtca ttgcacctgt ggcagagaag aaaaatcct ccgtaaaggc ggcaggaatg	180
aagtctattc ttgtcagcga aaataaaaatg tacataaccc ctttcggcaa gggcaattct	240
gctgtgcttg aatatgaggt ggacaataat gactacaaca aaactcagct ttcttcaaag	300
gacaacagca atategagct tggtgatgta aacgaggtaa acatcactt ttcaagcaag	360
catggctttg ggagcggagt ggagataat acttcaaacc ctactcacag aagcggtaa	420
agctcgctg taagagggga tatgctgggg cttaaatcgg agcttggaaa ggccttttc	480
ggcaaaacctt ttgtgataa tatacatatc cagtttattt acaacattct ggatatcgaa	540
aagataacttg cggtgtatgt aacgaatatc gtttatgcgc tgaacaatat gcttggata	600
aaggattctg aaagttatga tgatttatg gggtatctt ctgcaagaaa tacttatgaa	660
gtttttactc accctgacaa aagcaatctt tccgataagg taaaggtaa tatcaagaaa	720
agccttagca agttaatga cttgctgaaa actaagcgcc ttggcttattt cggccttcaa	780
gagccaaaga caaaagacac aagagcttcg gaagcataca aaaagcgtgt ttatcatatg	840
cttgcaatttggggcagat aagacagtgt gttttcatg ataaatcggg tgcaaaaaga	900
tttgacattt acagttttat taacaatatt gatcccgaat acagagatac tcttgactat	960
cttggaggagg agcgtttaaa gtccataaac aaggacttta tcgagggtaa caaggtaat	1020
atcagcctgc ttattgataat gatgaaaggc tatgaggctg atgatatcat acgcctttat	1080
tacgattca ttgtgcttaa atctcagaaa aatctcggt tttctatcaa aaagcttgcgt	1140
gagaaaaatgc tggaggaata cggtttcaga tttaaggaca agcaatatga ctctgtgcgc	1200
tcaaagatgt acaagcttgc ttgttgcata actactacag aatgacgtt	1260
gccgcaggcg aagctttgt gctgaaactg cgttttcaa tgaccgatga tgaaaaagag	1320
gggatatatg ctgatgaagc ggcaagctt tggggcaaat tcaggaatga tttgaaaat	1380

-continued

atcgccgacc acatgaacgg tgacgttatac aaggagcttg gcaaggctga catggatttt	1440
gatgagaaaa ttcttgacag taaaagaag aatgcgtctg acctttgtat tttctccaaa	1500
atgatatata tgctcacata ttttcttgcg ggcaaggaga taaacgatct tcttacaacg	1560
cattatcagca agtttgataa catcaaggag ttttgaaga taatgaaaag ctctgctgtt	1620
gatgtttagt gtgagcttac ggcgggctac aagctgttca atgacagcga gaggataacc	1680
aacgagcttt ttatcgtaaa gaacattgtt tccatgagaa agcctgcggc ttcagcgaag	1740
cattacgatgt tccgtgacgc actgactata ctcggtagt acgacaatata cacggacgat	1800
aggataagcg agattctaaa acttaaagaa aaaggcaagg gcatacatgg tctgagaaat	1860
tttataaaca acaatgttat cgagtccctct cggtttgtat accttataa gatgcgaac	1920
gtcagaaga taagagaagt ggctaagaat gagaaagttt tcatgtttgt tcttgggggt	1980
atccctgaca cgcagataga gcttattac aagagttgtg tggagtttcc tgacatgaat	2040
agttcttgg aagcaaagcg cagttagctt gcgagaatga taaagaacat cagctttgt	2100
gatttcaaaa atgtgaaaca gcaggcaag ggcagagaaa acgtggctaa ggagaggcga	2160
aaggctgtta tcgggttta tcttacggtc atgtatctgc tggtaaaaaa tcttgtgaat	2220
gtcaatgcaa ggtatgttat tgcgatacac tgccttgaac gtgattttgg gctgtataag	2280
gagataattc ctgagttggc ttcaaagaac ttgaaaaatg actacaggat actttcacag	2340
acgctttgtg aactttgtga tgatcgtaat gagtcgtcga attttgttctt gaaaaagaac	2400
aagcggctgc gcaagtgcgt tgaagttgtat atcaataatg cagacagcag catgacaaga	2460
aaataccgcg actgttattgc tcatcttact gtatcgatg aactgaaaga atacatagga	2520
gatattcgta cagtggattt ttacttctcc atttatcatt atgttatgca gcgttgcata	2580
acgaaaaggg gagatgacac aaagcaagaa gagaaaataa agttaggaa cgatcttta	2640
aaaaatcagc gctatacgaa agactttgtat aaggctctca actcgccgtt tggataaac	2700
attcccgaggt taaaaatct ttcaatttgag cagttgtttt acagaaatga atatcttact	2760
gaaaagtag	2769

<210> SEQ ID NO 32
<211> LENGTH: 2769
<212> TYPE: DNA
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(2769)
<223> OTHER INFORMATION: native CasM DNA sequence from Ruminococcus
sp. CAG:57

<400> SEQUENCE: 32	
atggcaaaaa agaataaaaat gaagcctaga gagctgcgtg aggctcagaa aaaagccaga	60
cagctcaaag cggctgagat aaataataac gctgctctg cgatcgctgc catgcctgt	120
gcagaggatca ttgcacctgt ggcagagaag aaaaaatccct ccgtaaaggc ggcaggaatg	180
aagtctatttcttgcgaa aaataaaaatg tacataacctt ctttcggcaaa gggcaattct	240
gctgtgttttgcgaa aatatgaggt ggacaataat gactacaaca aaactcagct ttcttcaag	300
gacaacagca atatcgagct tgggtatgtat aacgaggtaa acatcaactt ttcaagcag	360
catggcttttgcgaa ggagcggagt ggagataat acttcaaaacc ctactcacag aagcggtgaa	420
agctcgccctg taagagggga tatgtgtggg cttaaatcgagtttgcgaa gcgcttttc	480

-continued

ggcaaaacctt ttgatataa tatacatatc cagttattt acaacattct ggatatcgaa	540
aagataacttg cggtgtatgt aacgaatatc gtttatgcgc tgaacaatat gcttggata	600
aaggattctg aaagtatgt tgatttatg gggtatctt ctgcaagaaa tacttatgaa	660
gttttactc accctgacaa aagcaatctt tccgataagg taaaggtaa tatcaagaaa	720
agccttagca agttaatgt cttgctgaaa actaagcgcc ttggctattt cggcattgaa	780
gagccaaaga caaaagacac aagagcttcg gaagcataca aaaagcgtgt ttatcatatg	840
cttgcaattt ggggcagat aagacagtgt gttttcatg ataaatcggg tgcaaaaaga	900
tttgacccctt acagtttat taacaatatt gatcccgaat acagagatac tcttgactat	960
cttggagg agcgtttaaa gtccataaac aaggactta tcgaggtaa caaggtaat	1020
atcagectgc ttattgatat gatgaaaggc tatgaggctg atgatatcat acgccttatt	1080
tacgattca ttgtgcttaa atctcagaaa aatctcggt tttctatcaa aaagcttcgt	1140
gagaaaaatgc tggaggataa cggtttcaga ttaaggaca agcaatatga ctctgtgcgc	1200
tcaaagatgt acaagcttat ggattccctg ctttctgca actactacag aaatgacgtt	1260
gccgcaggcg aagctttgt gcgttaactg cgttttcaa tgaccgatga tgaaaaagag	1320
gggatataatgc ctgatgaaagc ggcaagctt tggggcaaat tcaggaatga ttttggaaat	1380
atcgccgacc acatgaaacgg tgacgttatac aaggagctt gcaaggctga catggattt	1440
gtgagaaaa ttcttgacag tgaaaagaag aatgcgtctg acctttgtt tttctccaaa	1500
atgatataata tgctcacata ttttcttgcg ggcaaggaga taaacgatct tcttacaacg	1560
cttacatcgca agtttgatata catcaaggag ttttgaaga taatgaaaag ctctgtgtt	1620
gtgttgagt gtgagcttac ggccggctac aagctttca atgacagccca gaggataacc	1680
aacgagctt ttatcgtaaa gaacattgtc tccatgagaa agcctgcggc ttcaagcgaag	1740
cttacgtgt tccgtacgc actgactata ctcggatatac acgacaatatac cacggacat	1800
aggataagcg agattctaaa acttaaagaa aaaggcaagg gcatacatgg tctgagaaaat	1860
tttataacaa acaatgttat cgagtctct cggttgtat accttataa gtagcgaac	1920
gtcagaaga taagagaagt ggctaaggat gagaaaggttt tcatgtttgt tcttgggggt	1980
atccctgaca cgcaagataga gcgttattac aagagttgtg tggagttcc tgacatgaat	2040
agttcttgg aagcaagcg cagttagttt gcgagaatga taaagaacat cagctttgtat	2100
gatttcaaaa atgtgaaaca gcaggcaag ggcagagaaa acgtggctaa ggagaggca	2160
aaggctgtta tcgggttta tcttacggtc atgtatctgc tggtaaaaaa tcttgtgaat	2220
gtcaatgca ggtatgttac tgcgatacac tgccttgcac gtgatttgg gctgtataag	2280
gagataattc ctgagttgc ttcaaagaac ttgaaaaatg actacaggat actttcacag	2340
acgctttgtt aactttgttga tgatcgtaat gagtcgtcg attttgttctt gaaaaagaaac	2400
aagcggctgc gcaagtgcgt tgaagttgtat atcaataatg cagacagcag catgacaaga	2460
aaataccgca actgtattgc tcatcttact gtatgttgc aactgaaaga atacatagga	2520
gatattcgta cagtggattt ttacttctcc atttatcatt atgttatgcg gctgtatc	2580
acgaaaagggg gagatgacac aaagcaagaa gagaaaataa agttaggaa cgtatttta	2640
aaaaatcacg gctatacgaa agactttgtt aaggctctca actcgccgtt tggataaac	2700
attccgaggt taaaaatct ttcaatttgcg cagttgttttgc acagaaatga atatcttact	2760

-continued

gaaaagtag	2769
<hr/>	
<210> SEQ ID NO 33	
<211> LENGTH: 2802	
<212> TYPE: DNA	
<213> ORGANISM: Ruminococcus flavefaciens	
<220> FEATURE:	
<221> NAME/KEY: misc_feature	
<222> LOCATION: (1) ..(2802)	
<223> OTHER INFORMATION: native CasM DNA sequence from Ruminococcus flavefaciens FD-1	
<400> SEQUENCE: 33	
ataaaaaaga aaatgtctct ccgtgaaaag cgtgaagccg agaaacaggc taaaaaaagct	60
gcataattcag cagcttcaaa aaatacagat tctaagcctg cggaaaagaa agcagaaaact	120
ccaaagcctg cggagattat ttccgataat tccagaaata agaccgctgt aaaggcggct	180
ggtctgaaat caacaattat cagcggcgat aagctgtata tgacatctt cggcaagggt	240
aacgctgctg ttattgagca gaaaatagat atcaatgattt attcttttc agctatgaaa	300
gatactocgt cgcttgaagt tgataaaagca gaatcaaaag agatctt ttcaagtac	360
catccttttg taaagaatga taagctgaca acatataacc ctttatacgg cggcaaggat	420
aaccccgaaa agcctgtcgg cagggatatg ctccggcttaa aagataagct tgaagaacgc	480
tatttggat gtacattcaa tgataatctt cacatccaga ttatctataa catacttgac	540
atcgagaaga ttttagctgt tcattctgca aatatcacaa ctgcgttga ccacatggtt	600
gatgaagacg atgaaaaata tcttaacacg gattatatcg gctacatgaa taccataaat	660
acatatgacg tgtttatgga tccttcaaag aattcttcat taagccctaa agatagaaaag	720
aatattgaca acageccgtgc aaaattttag aactctttt caactaagcg ccttggctat	780
tttggatttg actatgatgc aaacggtaag gacaagaaaa agaacgagga aataaaaaaag	840
cgtttatatac atctcacacg ttttgcaggc cagctccgtc agtggagttt tcatatgtct	900
ggcaattatac cgagaacatg gctttacaag ctgcattcac tggataagga atatcttgat	960
actcttgacc attacttcga taaacgtttt aacgatataa acgtatgtt cgtaactaag	1020
aatgctacca atctctatata tctgaaagaa gtatttcccg aagcaaaactt caaggatatt	1080
gccgatctttt attacgattt catagttata aagtcgcaca aaaatatggg attctccata	1140
aaaaagctga gggagaagat gcttgaatgt gatggcgcag acaggataaa agaacaggat	1200
atggactctg ttgcgtcaaa gctgtataag ctcatagact ttgcattttt caagtattat	1260
cacgaatttc ctgaacttgc tgaaaagaat gtggatatac tcagagcgc tggatccgt	1320
acaaaaaaaaa ataaccttta ttctgtatgag gctgcacgt tatggagcat atttaagaa	1380
aaattccctcg gcttctgtga taagatagtt gtatggtaa caggagagca tgagaaagat	1440
atcacatccg ttattgataa ggatgcttac aggaacagga gcaatgtttc atattctca	1500
aagctgtatgt atgcaatgtg cttttcctt gacggaaaag agataaaatga ctttctcaact	1560
actcttatca acaaattcga taatatcgat aaccagataa aaacagccaa agaacttggc	1620
attaatactg cttttgtaaa gaattacgt ttcttcaatc acagcgagaa atatgtcgat	1680
gaaactgaaca tcgtcaagaa tattgcaaga atgaagaacg cttcaagtaa tgccaaaaaa	1740
gctatgtatc atgatgcgt tactattctc ggaatacctg aggatatgga tgaaaaagct	1800

-continued

cttgatgagg aactggattt aattcttcaa aaaaagacag acccagtaac tggcaagcca	1860
ctgaaaggta agaatccctt acgttaatttt atcgcaaaca atgtgataga gaattcaaga	1920
tccatataatc ttatcaagtt ctgcaatctt gagaatgtac gtaaaatcgt gaataataca	1980
aagggtcactg agtttgttta aaagcgtatt cccgatgetc agatcgaacg ctattataag	2040
tctgttacag attctgaaat gaatccgcct actgaaaaga agatcaccga acttgtcggt	2100
aagttaaagg atatgaactt tggcaacttc cgaaatgtga gacagtctgc taaagagaat	2160
atggagaagg agcgcattcaa agctgttata gggctttatc tcacggtagt atatcggtt	2220
gtcaagaatc ttgttcatgt aaactcacga tatatcatgg cttttcattc gcttgaacgt	2280
gattcacaac tgtataacgt atctgttcatgt aatgattatc ttgcacttac cgatacttt	2340
gttaaggagg gagataattc cagaaggcaga tatcttgcag gcaacaagcg tctgagagat	2400
tgtgtgaagc aggataatcga taatgcaaaa aagtggttt ttagtgataa gtacaatagc	2460
ataaccaagt acaggaataa cgttgccat cttaccgttg tacgttaactg cgctgaattc	2520
atcggagata taacgaagat agactcctat tttgcattgt atcattatct catcagaga	2580
cagcttgcga aagggtttga ccatgagcga agtggcttg acagaaacta tccacagtat	2640
gcaccgcgtgt ttaagtggca tacgtatgt aaggatgttgc tcaaggctct gaatgctcca	2700
tttggctaca atatccctcg tttcaagaat ctcagcatag atgcacttt tgacggcaac	2760
gaaataaaga agaatacggtt ccggatgtt ga	2802

```

<210> SEQ ID NO 34
<211> LENGTH: 2835
<212> TYPE: DNA
<213> ORGANISM: Ruminococcus albus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(2835)
<223> OTHER INFORMATION: native CasM DNA sequence from Ruminococcus
albus strain KH2T6

```

<400> SEQUENCE: 34

atggcaaaaa aatcgaaagg tatgagcctt agagaaaaac gtgaacttga aaagcagaaaa	60
aggataaaaa aggcagctgt gaattcgtt aatgatacac ctgaaaaaac agaagaagca	120
aatgtcgtat ctgttaatgt caggacatcg gctgagaata agcatagtaa aaaatctgt	180
gecaaaagctt tggactgaa atccggctgt gttatcggtg atgagctgtt ccttacttca	240
ttcggcagag gtaacgaagc aaagcttcaa aagaagatat ccggtgacac tgcgaaaaaa	300
cttggcatttgc tgcgtttgttgc agtgcggaa cgtgacgat caacgcgttac cctcgaaagt	360
ggcaggataa aggacaagac cgccagaccc aaagacccccaa gacatataac cgtcgatata	420
caaggtaataa tcaaggaaaga tatgcttggg atacgcgttgc tactggagaa aaagatattt	480
ggcaaaaat ttgtatgtaa tatccatgtt cagcttgcgt acaatatctt ggatgtcgaa	540
aaagataatgg cacagtatgt cagcgatatac gtatatatgc tgcataatac tgataaaaca	600
gaaagaaaacg ataataatctt ggggtatgttgc agcatcgatataa gacattttgtt	660
gatacgtcaa atcttcccgaa tgatacaaaa caaaaatgttgcgaa gagatgtcgaa	720
gacaagatca taaaaggcgg cagacttggg tatttcggcgttgc aagctttatc ggtaaacagc	780
ggcaatagta ccaagcttagt acccgagaaa gagatatac atatcttgc gcttgcgttgc	840
agcctgaggc agagttactt tcacggatataa gtaaaagata ccgattatca gggaaaccaca	900

-continued

tgggcataata	ctcttgagga	caagctgaaa	ggtccgagcc	atgagttcag	ggaaaccatt	960
gataagatata	ttgtatgggg	attcagcaag	atcagcaagg	actttggcaa	gatgaacaag	1020
gtcaaccttc	agataacttga	acagatgatt	ggtgaactgt	atggcgat	agaacgacaa	1080
aacctcaactt	gcgattacta	tgacttcatt	caactgaaaa	agcataagta	tcttggattt	1140
tctataaagc	gtcttagaga	gaccatgctt	gaaacaacac	cggctgaatg	ttataaaagct	1200
gaatgctata	acagecgagcg	tcaaaagctg	tataagctga	tagattcct	gatatatgat	1260
ctttactata	accgttaagcc	tgcacgcattc	gaagaaaatcg	tggacaagct	gaggaaatct	1320
gtgaacgacg	aagagaaaaga	atccatataat	tcagttgagg	cgaagtatgt	ctatgaatca	1380
cttagcaaag	ttctggataa	atcgctgaaa	aacagtgtgt	ctgggtaaac	gataaaggat	1440
ctccaaaaga	gatatgatga	cgaaacacgca	aacaggatct	gggatatctc	acagcacagt	1500
ataagtggaa	atgtcaactg	tttctgcaag	ctaatttata	ttatgaccct	gatgttgac	1560
ggcaaggaga	taaatgatct	gctgacaacg	ctggtaaaca	agttcgataa	catagcatca	1620
tttatacatgt	ttatggacga	acttggctt	gagcatagtt	ttacagataa	ctataaaatg	1680
tttgcgcaca	gcaaggctat	atgccttgc	ctgcagttca	taaacagttt	tgcacgtatg	1740
tcaaagatcg	atgatgagaa	gtcaaaaaga	cagctttcc	gtgatgcgt	tgtcataactg	1800
gatatcggtt	ataaaagatga	gacttggata	aataattatc	tggattctga	tatttcaaa	1860
ctggacaaag	aaggtaacaa	gtttaaggc	gcaaggcatg	atttcaggaa	ctttatagcc	1920
aataatgtta	taaagtcatc	acgtttcaa	tacctagtaa	aatacagcag	tgccgatggt	1980
atgataaaagc	tgaaaacgaa	tgaaaagctg	ataggcttg	ttctggataa	gcttccagaa	2040
acgcagatag	accgctacta	tgaatcatgc	ggacttgaca	atgcggtagt	agataagaaa	2100
gtcaggatag	aaaagctatc	ggggcttatac	agagatatga	agttcgatga	tttcagcggt	2160
gtcaaaacct	caaacaacgc	aggagataat	gacaaacagg	ataaggcgaa	atatcaggcg	2220
ataataagcc	tgtacccat	ggtgctgtat	cagatagtca	agaacatgtat	atatgtcaac	2280
tcacgttatg	ttatcgctt	ccattgtctt	gaacgtgact	ttggatgtta	tggaaaagat	2340
tttggaaagt	attatcaagg	ctgccgaaaa	cttacagatc	attttattga	agaaaagatc	2400
atgaaagagg	gtaaacttgg	ctgcaataaa	aaagtccggca	gatatctgaa	aaataatatt	2460
tctctgtca	ctgatggact	gataaatacc	taccgtatc	aggttgatca	ctttcgatg	2520
gtaaggaaga	taggcaacta	tgccggcatat	atcaagagta	tcggttcg	gtttgaactt	2580
tatcactatg	taatacagag	gatagtttt	gacgaaataca	gatttgact	taacaacact	2640
gaaagcaact	ataagaacag	catcatcaag	caccataacct	actgtaaagg	tatggtcaag	2700
gcactgaaca	caccccccgg	ttatgacatg	ccgagataca	agaatcttc	tatcggtat	2760
ctgtttgatc	gcaataatta	tctgaataaa	acaaaagagt	caatagatgc	aaatagctct	2820
attgacagtc	agtga					2835

```

<210> SEQ_ID NO 35
<211> LENGTH: 2904
<212> TYPE: DNA
<213> ORGANISM: Ruminococcus flavefaciens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)...(2904)
<223> OTHER INFORMATION: native CasM DNA sequence from Ruminococcus

```

-continued

flavefaciens strain XPD3002	
<400> SEQUENCE: 35	
atgatcgaaa agaagaagtc atttgcaaag ggcattggag taaaatcaac acttgtatcc	60
ggttcaaagg tatacatgac gacgttcgca gaaggaagcg atgccagact tgaaaagatc	120
gttgaaggcg attctatcag atctgtcaac gaaggagaag cgttctcagc tgaaatggct	180
gataagaatg caggctacaa gatcgtaac gcaaagttca gccacccaaa gggctatgct	240
gtatggcaa acaaccctt atacaccgga ccggtagcagg aggatatgct cggctgaag	300
gaaacgcttg aaaagagata ttttgagag tctgcccacg gaaatgataa tatctgtatt	360
caggtcattcc ataataatcct cgatatcgaa aagatcctcg ctgaatataat aaccaatgct	420
gtttatgcgg taaacaatat ttccgggttt gataaggata tcattcggtt tggttaagttc	480
agtacggctt atacttatga tgatgtcaag gatcctgaac atcacagagc agctttcaac	540
aataacgata agttaattaa tgccatcaag gcacagtatg atgaatttga caatttcctt	600
gataatcctc gtctcggcta ctttggacag gctttttca gtaaggaaagg cagaaattac	660
attatcaatt acggcaacga gtgttatgtt attttgtt tactcagcgg attgcgtcac	720
tgggttagtac ataataatga ggaagaatca aggatttccc gtacatggct ttataatctc	780
gacaagaatc ttgacaacga atatatctct actctcaatt atctgtatga tagaattaca	840
aacgaattaa caaattcctt ctcaaagaat agtgcagcca acgtaaacta taticgtgaa	900
acccttggta ttaatcctgc tgaatttgc gaggcattt tcagattcag taticgtgaa	960
gaacagaaga atctcggtt caatattact aagctgagag aagtaatgct tgacagaaag	1020
gatatgtctg agatccgtaa aaatcataag gtctttgatt caatccgtac taaggcttat	1080
actatgtatgg atttcgtt atctacatgat tacatttgcg aggatgcataa ggttgctgct	1140
gccaacaagt ctctgcggta taacgaaaaa agcctcgtg aaaaggatata ctgtttata	1200
aatctcagag gaagctttaa cgatgtatgc aaggatgcc tttattatga tgaggccat	1260
cgtatttggaa gaaagctcgaa aacattatgc cacaatatac aggaattcag aggcaataag	1320
acacgtgaat acaagaagaa ggatgctcca agactccccaa gaattttcc tgccggaaagg	1380
gatgtttccg cgttctcaaa gttgtatgtc gctcttacca tgttcttgc tggttggag	1440
atcaatgtatgc ttctcaccac gctcatcaat aagttcgtata acatcccgat tttctcaag	1500
gtaatgcctc ttatcggtt gaatgcaag tttgttggagg aatatgcctt cttcaaggac	1560
agcgcaaaga ttgctgacga actcaggctg attaagagct ttgcccataat gggagaacct	1620
atcgcagatg caagacgtgc tatgtatatac gatgtatca ggatttcgg aacaacatctc	1680
agctatgtatgc agcttaaggc ctttgcgtatc actttttccg ttgtatggaaa cggcaacaag	1740
cttaagaagg gcaaggcacgg catgagaaac ttcatcatta ataatgtataat cagtaacaag	1800
cgcttccatt atctcattcg ttacgggtat cctgcacatc tccatgagat cgccaagaat	1860
gaagctgtt gaaatgttgcgtt ccttgcgtatc atagctgataat tcacatggaaa gcaaggcacag	1920
aacggaaaga atcagatcgaa cagggtactat gagacgttgc tcggcaaggaa caaggccaaag	1980
tctgtctccg aaaaggttgcgtt tgccctcaca aagattatca ccggatgttgcgatc ttcacatcgat	2040
ttcgataaga agagaagcgat tatttggaggat actggaaaggaa aaaacgctgaa gagagaaaaag	2100
ttcaagaaga tcattcgtatc ctatcttact gtcatttatac acatcattaa gaatattgtt	2160

-continued

aatatcaatg	cgcgttacgt	tatcggttc	cattgcgttg	agcgtatgc	acagcttat	2220
aaggaaaagg	gctatgatat	caacctcaag	aagctcgaag	aaaaggggtt	ttcatcagtc	2280
acaaggctgt	gtgcaggtat	tgtgagact	gctcctgaca	agcgtaaagga	tgttggaaaag	2340
gaaatggctg	agcgtgcaaa	ggaatctatc	gatagccttg	aatctgcaaa	tcctaagctt	2400
tacgcaact	ataatcaagta	ttctgacgag	aagaaggctg	aggaatttac	tagacagatc	2460
aaccgtgaga	aggcaaagac	cgcgtctgaat	gcatatctca	gaaatactaa	gtggaaatgtg	2520
ataatcaggg	aagatcttct	tagaatcgat	aataagacat	gtacgcttt	tagaaataag	2580
gcccgttcatc	ttgaagttgc	aagatatgtt	catgcatata	tcaacgatat	tgccgaaagta	2640
aacagctatt	tccagcttta	tcattacatc	atgcagagaa	tcatcatgaa	cgaaagatata	2700
gaaaagtctt	ctggaaaggt	aagcgaatac	ttcgatgttg	tgaacgtatg	aaagaagtac	2760
aacgacagggc	ttctgaaagct	gttgtgcgtt	ccatTTGGT	actgcataccc	gagattcaag	2820
aatctctcca	ttgaagcttt	gttcgacagg	aacgaagcag	ctaagtttga	caaggaaaag	2880
aagaaaatgt	caggtaattc	atag				2904

```

<210> SEQ ID NO 36
<211> LENGTH: 2391
<212> TYPE: DNA
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(2391)
<223> OTHER INFORMATION: native CasM DNA sequence from Ruminococcus
    sp., isolate 2789STDY5834894

```

<400> SEQUENCE: 36

gtggagataa	acacttcaaa	ccctactcac	agaagcggtg	aaagctcgtc	tgttaagaggg	60
gatatgctgg	ggcttaatc	ggagcttcaa	aagcgctttt	tcggcaagac	ttttgatgat	120
aatatacata	tccagcttat	ttacaacatt	ctggatatcg	aaaagatact	tgcagtgtat	180
gtgacgaata	tcgtttatgc	actgaacaat	atgcttgggt	taaagggttc	tgaaagttat	240
gatgatttta	tgggtatct	ttctgccccaa	aatacttatt	atattttac	tcaccctgac	300
aaaagtaatc	tttccgataa	ggtaaagggt	aatatcaaga	aaagccttag	caagtttaat	360
gacctgctga	aaactaagcg	tcttggctat	tttggtcttg	aagagectaa	gacgaaagat	420
aaaagagttt	cggaggcata	caaaaagcgt	gtttatcata	tgcttgcaat	tgtggggcag	480
ataaggcaga	gtgtttcca	tgataagtca	aatgagcttg	atgagttact	ttacagcttt	540
attgacatta	ttgattccga	atacagagac	actcttgcact	atcttgcata	tgagagattt	600
gattctataa	ataagggctt	tgtccaggc	aacaaggctca	atatcagctt	gcttattgtat	660
atgatgaaag	gctatgaggc	tgtatgatc	atacgcctt	attatgattt	cattgtgctt	720
aaatctcaga	aaaatctcg	ttttctatc	aaaaagcttc	gtgagaaaat	gctggacgaa	780
tacggcttca	gatttaagga	caagcaatat	gactctgtgc	gctcaaagat	gtacaagctt	840
atggatttgc	tgctttctg	caactattac	agaaatgacg	ttgtcgcagg	cgaagcttt	900
gtgcgcaaac	tgcgttttc	aatgaccgat	gatgaaaaag	aggggatata	tgctgtatgaa	960
ggtggaaaagc	tttggggcaa	attcaggaat	gatTTGAA	atatcgccga	ccacatgaac	1020
ggtgacgttca	tcaaggagct	tggcaaggct	gacatggatt	ttgatgagaa	aattcttgac	1080
agcggaaaaga	agaatgcgtc	tgacccttttgc	tatttctcca	aaatgatata	tatgtcaca	1140

-continued

```
tattttcttg acggcaagga gataaacat cttcttacaa cgcttatcag caagtttcat 1200
aacatcaagg agttttgaa gataatgaaa agctctgctg ttgtatgttga gtgtgagctt 1260
acggcggttaca acaagctgtt caatgacacg cagaggataa ccaacgactt ttttatcgta 1320
aagaacatttgcgttccatgag aaacgcctcgccgttccggcga agcttacat gttccgtgac 1380
gcactgacta tactcggtt agacgacaag atcacggacg ataggataag cgagattta 1440
aaacttaaaggaaaaaggccaa gggcatacat ggtctgagaa attttataac aaacaatgtt 1500
atcgagtccttctcggttgtt ataccttac aagtatgcga acgctcagaa gataagagaa 1560
gtggctaaga atgaaaaagt tgtcatgttt gttcttgggg gtatccctga cacgcagata 1620
gagcgttatt acaagagttt tggttgcattt cctgacatga acagttctt ggaagcaaag 1680
tgcagtgagc ttgcgagaat gataaagaac atcagctttg atgatttcaa aaatgtgaaa 1740
cagcaggcaaaaggccagaga aaacgtggct aaggagagg caaaggctgt tatcgggctt 1800
tatcttacgg tcatgtatct gctgggttgcattt aatcttgcga atgtcaatgc aaggtatgtt 1860
attgcgatac actgccttgc acgtgattttt gggctgtata aggagataat tcctgagtt 1920
gcttcaaaga acttggaaaaaa tgactacagg atactttcac agacgctttg tgaactttgt 1980
gatgatcgttgc atgagtcgccc gaatttgcattt ttggaaaaaga acaagcggct ggcgaatgtc 2040
gttgaagtttgc atatcaataa tgcagacacg agcatgacaa gaaaataccg caactgtatt 2100
gctcatcttgc tctgtatgc tgaactgaaa gaatacatag gagatattcg tacagtggat 2160
tcttacttcttccatattca ttatgttgc tgcgtgttgc tcacgaaaag ggaagatgac 2220
acaaggaaag aagagaaaaat aaagtatgcg gacgatctt taaaaaatca cggctatacg 2280
aaagacttttgc taaaggctctt caactcgccg tttggataca acattccgag gttaaaaat 2340
cttcaatttgc acgttgcattt tgacgaaaat gaatatcttgc tggaaaagtg 2391
```

```
<210> SEQ_ID NO 37
<211> LENGTH: 954
<212> TYPE: PRT
<213> ORGANISM: Eubacterium siraeum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(954)
<223> OTHER INFORMATION: native CasM protein sequence from
Eubacterium siraeum
```

```
<400> SEQUENCE: 37
```

Met	Gly	Lys	Ile	His	Ala	Arg	Asp	Leu	Arg	Glu	Gln	Arg	Lys	Thr	
1								5		10		15			
Asp	Arg	Thr	Glu	Lys	Phe	Ala	Asp	Gln	Asn	Lys	Lys	Arg	Glu	Ala	Glu
								20		25		30			
Arg	Ala	Val	Pro	Lys	Lys	Asp	Ala	Ala	Val	Ser	Val	Lys	Ser	Val	Ser
								35		40		45			
Ser	Val	Ser	Ser	Lys	Lys	Asp	Asn	Val	Thr	Lys	Ser	Met	Ala	Lys	Ala
								50		55		60			
Ala	Gly	Val	Lys	Ser	Val	Phe	Ala	Val	Gly	Asn	Thr	Val	Tyr	Met	Thr
								65		70		75		80	
Ser	Phe	Gly	Arg	Gly	Asn	Asp	Ala	Val	Leu	Glu	Gln	Lys	Ile	Val	Asp
								85		90		95			
Thr	Ser	His	Glu	Pro	Leu	Asn	Ile	Asp	Asp	Pro	Ala	Tyr	Gln	Leu	Asn
								100		105		110			

-continued

Val Val Thr Met Asn Gly Tyr Ser Val Thr Gly His Arg Gly Glu Thr
115 120 125

Val Ser Ala Val Thr Asp Asn Pro Leu Arg Arg Phe Asn Gly Arg Lys
130 135 140

Lys Asp Glu Pro Glu Gln Ser Val Pro Thr Asp Met Leu Cys Leu Lys
145 150 155 160

Pro Thr Leu Glu Lys Lys Phe Phe Gly Lys Glu Phe Asp Asp Asn Ile
165 170 175

His Ile Gln Leu Ile Tyr Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala
180 185 190

Val Tyr Ser Thr Asn Ala Ile Tyr Ala Leu Asn Asn Met Ser Ala Asp
195 200 205

Glu Asn Ile Glu Asn Ser Asp Phe Phe Met Lys Arg Thr Thr Asp Glu
210 215 220

Thr Phe Asp Asp Phe Glu Lys Lys Glu Ser Thr Asn Ser Arg Glu
225 230 235 240

Lys Ala Asp Phe Asp Ala Phe Glu Lys Phe Ile Gly Asn Tyr Arg Leu
245 250 255

Ala Tyr Phe Ala Asp Ala Phe Tyr Val Asn Lys Lys Asn Pro Lys Gly
260 265 270

Lys Ala Lys Asn Val Leu Arg Glu Asp Lys Glu Leu Tyr Ser Val Leu
275 280 285

Thr Leu Ile Gly Lys Leu Arg His Trp Cys Val His Ser Glu Glu Gly
290 295 300

Arg Ala Glu Phe Trp Leu Tyr Lys Leu Asp Glu Leu Lys Asp Asp Phe
305 310 315 320

Lys Asn Val Leu Asp Val Val Tyr Asn Arg Pro Val Glu Glu Ile Asn
325 330 335

Asn Arg Phe Ile Glu Asn Asn Lys Val Asn Ile Gln Ile Leu Gly Ser
340 345 350

Val Tyr Lys Asn Thr Asp Ile Ala Glu Leu Val Arg Ser Tyr Tyr Glu
355 360 365

Phe Leu Ile Thr Lys Tyr Lys Asn Met Gly Phe Ser Ile Lys Lys
370 375 380

Leu Arg Glu Ser Met Leu Glu Gly Lys Gly Tyr Ala Asp Lys Glu Tyr
385 390 395 400

Asp Ser Val Arg Asn Lys Leu Tyr Gln Met Thr Asp Phe Ile Leu Tyr
405 410 415

Thr Gly Tyr Ile Asn Glu Asp Ser Asp Arg Ala Asp Asp Leu Val Asn
420 425 430

Thr Leu Arg Ser Ser Leu Lys Glu Asp Asp Lys Thr Thr Val Tyr Cys
435 440 445

Lys Glu Ala Asp Tyr Leu Trp Lys Lys Tyr Arg Glu Ser Ile Arg Glu
450 455 460

Val Ala Asp Ala Leu Asp Gly Asp Asn Ile Lys Lys Leu Ser Lys Ser
465 470 475 480

Asn Ile Glu Ile Gln Glu Asp Lys Leu Arg Lys Cys Phe Ile Ser Tyr
485 490 495

Ala Asp Ser Val Ser Glu Phe Thr Lys Leu Ile Tyr Leu Leu Thr Arg
500 505 510

-continued

Phe	Leu	Ser	Gly	Lys	Glu	Ile	Asn	Asp	Leu	Val	Thr	Thr	Leu	Ile	Asn
515															525
Lys	Phe	Asp	Asn	Ile	Arg	Ser	Phe	Leu	Glu	Ile	Met	Asp	Glu	Leu	Gly
530															540
Leu	Asp	Arg	Thr	Phe	Thr	Ala	Glu	Tyr	Ser	Phe	Phe	Glu	Gly	Ser	Thr
545															560
Lys	Tyr	Leu	Ala	Glu	Leu	Val	Glu	Leu	Asn	Ser	Phe	Val	Lys	Ser	Cys
565															575
Ser	Phe	Asp	Ile	Asn	Ala	Lys	Arg	Thr	Met	Tyr	Arg	Asp	Ala	Leu	Asp
580															590
Ile	Leu	Gly	Ile	Glu	Ser	Asp	Lys	Thr	Glu	Glu	Asp	Ile	Glu	Lys	Met
595															605
Ile	Asp	Asn	Ile	Leu	Gln	Ile	Asp	Ala	Asn	Gly	Asp	Lys	Lys	Leu	Lys
610															620
Lys	Asn	Asn	Gly	Leu	Arg	Asn	Phe	Ile	Ala	Ser	Asn	Val	Ile	Asp	Ser
625															640
Asn	Arg	Phe	Lys	Tyr	Leu	Val	Arg	Tyr	Gly	Asn	Pro	Lys	Lys	Ile	Arg
645															655
Glu	Thr	Ala	Lys	Cys	Lys	Pro	Ala	Val	Arg	Phe	Val	Leu	Asn	Glu	Ile
660															670
Pro	Asp	Ala	Gln	Ile	Glu	Arg	Tyr	Tyr	Glu	Ala	Cys	Cys	Pro	Lys	Asn
675															685
Thr	Ala	Leu	Cys	Ser	Ala	Asn	Lys	Arg	Arg	Glu	Lys	Leu	Ala	Asp	Met
690															700
Ile	Ala	Glu	Ile	Lys	Phe	Glu	Asn	Phe	Ser	Asp	Ala	Gly	Asn	Tyr	Gln
705															720
Lys	Ala	Asn	Val	Thr	Ser	Arg	Thr	Ser	Glu	Ala	Glu	Ile	Lys	Arg	Lys
725															735
Asn	Gln	Ala	Ile	Ile	Arg	Leu	Tyr	Leu	Thr	Val	Met	Tyr	Ile	Met	Leu
740															750
Lys	Asn	Leu	Val	Asn	Val	Asn	Ala	Arg	Tyr	Val	Ile	Ala	Phe	His	Cys
755															765
Val	Glu	Arg	Asp	Thr	Lys	Leu	Tyr	Ala	Glu	Ser	Gly	Leu	Glu	Val	Gly
770															780
Asn	Ile	Glu	Lys	Asn	Lys	Thr	Asn	Leu	Thr	Met	Ala	Val	Met	Gly	Val
785															800
Lys	Leu	Glu	Asn	Gly	Ile	Ile	Lys	Thr	Glu	Phe	Asp	Lys	Ser	Phe	Ala
805															815
Glu	Asn	Ala	Ala	Asn	Arg	Tyr	Leu	Arg	Asn	Ala	Arg	Trp	Tyr	Lys	Leu
820															830
Ile	Leu	Asp	Asn	Leu	Lys	Ser	Glu	Arg	Ala	Val	Val	Asn	Glu	Phe	
835															845
Arg	Asn	Thr	Val	Cys	His	Leu	Asn	Ala	Ile	Arg	Asn	Ile	Asn	Ile	Asn
850															860
Ile	Lys	Glu	Ile	Lys	Glu	Val	Glu	Asn	Tyr	Phe	Ala	Leu	Tyr	His	Tyr
865															880
Leu	Ile	Gln	Lys	His	Leu	Glu	Asn	Arg	Phe	Ala	Asp	Lys	Lys	Val	Glu
885															895
Arg	Asp	Thr	Gly	Asp	Phe	Ile	Ser	Lys	Leu	Glu	Glu	His	Lys	Thr	Tyr
900															910
Cys	Lys	Asp	Phe	Val	Lys	Ala	Tyr	Cys	Thr	Pro	Phe	Gly	Tyr	Asn	Leu

-continued

915	920	925
Val Arg Tyr Lys Asn Leu Thr Ile Asp Gly Leu Phe Asp Lys Asn Tyr		
930	935	940
Pro Gly Lys Asp Asp Ser Asp Glu Gln Lys		
945	950	

```

<210> SEQ ID NO 38
<211> LENGTH: 919
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(919)
<223> OTHER INFORMATION: native CasM protein sequence from Ruminococcus
sp., isolate 2789STDY5834971

```

<400> SEQUENCE: 38

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln	15		
1	5	10	
Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala			
20	25	30	
Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Ala Ala			
35	40	45	
Glu Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu			
50	55	60	
Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser			
65	70	75	80
Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Gln Thr Gln			
85	90	95	
Leu Ser Ser Lys Asp Asn Ser Asn Ile Gln Leu Gly Gly Val Asn Glu			
100	105	110	
Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Glu Ser Gly Val Glu			
115	120	125	
Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val			
130	135	140	
Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe			
145	150	155	160
Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile			
165	170	175	
Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr			
180	185	190	
Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser His Asp Asp			
195	200	205	
Phe Ile Gly Tyr Leu Ser Thr Asn Asn Ile Tyr Asp Val Phe Ile Asp			
210	215	220	
Pro Asp Asn Ser Ser Leu Ser Asp Asp Lys Lys Ala Asn Val Arg Lys			
225	230	235	240
Ser Leu Ser Lys Phe Asn Ala Leu Leu Lys Thr Lys Arg Leu Gly Tyr			
245	250	255	
Phe Gly Leu Glu Pro Lys Thr Lys Asp Asn Arg Val Ser Gln Ala			
260	265	270	
Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg			
275	280	285	
Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr			

-continued

290	295	300
Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp Tyr		
305	310	315
Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Glu Asp		
325	330	335
Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu		
340	345	350
Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser		
355	360	365
Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu		
370	375	380
Asp Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg		
385	390	395
Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr		
405	410	415
Arg Asn Asp Ile Ala Ala Gly Glu Ser Leu Val Arg Lys Leu Arg Phe		
420	425	430
Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala		
435	440	445
Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His		
450	455	460
Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe		
465	470	475
Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu		
485	490	495
Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys		
500	505	510
Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile		
515	520	525
Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys		
530	535	540
Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr		
545	550	555
Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala		
565	570	575
Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly		
580	585	590
Ile Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Gly Ile Leu Lys Leu		
595	600	605
Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn		
610	615	620
Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn		
625	630	635
Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe		
645	650	655
Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser		
660	665	670
Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Gly Val Lys Arg Ser		
675	680	685
Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn		
690	695	700

-continued

Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala
705 710 715 720

Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys
725 730 735

Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
740 745 750

Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
755 760 765

Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
770 775 780

Leu Cys Asp Lys Ser Pro Asn Leu Phe Leu Lys Lys Asn Glu Arg Leu
785 790 795 800

Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser Ser Met Thr
805 810 815

Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val Arg Glu Leu
820 825 830

Lys Glu Tyr Ile Gly Asp Ile Cys Thr Val Asp Ser Tyr Phe Ser Ile
835 840 845

Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu Asn Asp Thr
850 855 860

Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys Asn His
865 870 875 880

Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr
885 890 895

Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe Asp Arg
900 905 910

Asn Glu Tyr Leu Thr Glu Lys
915

```
<210> SEQ_ID NO 39
<211> LENGTH: 918
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus bicirculans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(918)
<223> OTHER INFORMATION: native CasM protein sequence from Ruminococcus
bicirculans

<400> SEQUENCE: 39

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1 5 10 15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Val
20 25 30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Ala Ala Ala Pro Ala Ala
35 40 45

Glu Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
50 55 60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65 70 75 80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
85 90 95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Cys Asp Val Gly Lys
100 105 110
```

-continued

Val Asn Ile Thr Phe Ser Ser Arg Arg Gly Phe Glu Ser Gly Val Glu
115 120 125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Ser Val
130 135 140

Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145 150 155 160

Gly Lys Asn Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
165 170 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
180 185 190

Ala Leu Asn Asn Met Leu Gly Glu Gly Asp Glu Ser Asn Tyr Asp Phe
195 200 205

Met Gly Tyr Leu Ser Thr Phe Asn Thr Tyr Lys Val Phe Thr Asn Pro
210 215 220

Asn Gly Ser Thr Leu Ser Asp Asp Lys Lys Glu Asn Ile Arg Lys Ser
225 230 235 240

Leu Ser Lys Phe Asn Ala Leu Leu Lys Thr Lys Arg Leu Gly Tyr Phe
245 250 255

Gly Leu Glu Pro Lys Thr Lys Asp Thr Arg Ala Ser Glu Ala Tyr
260 265 270

Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg Gln
275 280 285

Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr Ser
290 295 300

Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Glu Thr Leu Asp Tyr Leu
305 310 315 320

Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Ile Gln Gly Asn
325 330 335

Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu Ala
340 345 350

Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser Gln
355 360 365

Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu Asp
370 375 380

Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg Ser
385 390 395 400

Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr Arg
405 410 415

Asn Asp Ile Ala Ala Gly Glu Ser Leu Val Arg Lys Leu Arg Phe Ser
420 425 430

Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala Lys
435 440 445

Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His Met
450 455 460

Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe Asp
465 470 475 480

Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu Tyr
485 490 495

Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys Glu
500 505 510

-continued

Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile Lys
 515 520 525
 Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys Glu
 530 535 540
 Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr Asn
 545 550 555 560
 Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala Ala
 565 570 575
 Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly Ile
 580 585 590
 Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu Lys
 595 600 605
 Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn Asn
 610 615 620
 Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn Ala
 625 630 635 640
 Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe Val
 645 650 655
 Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser Cys
 660 665 670
 Val Glu Phe Pro Asp Met Asn Ser Ser Leu Gly Val Lys Arg Ser Glu
 675 680 685
 Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn Val
 690 695 700
 Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala Lys
 705 710 715 720
 Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys Asn
 725 730 735
 Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu Glu
 740 745 750
 Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser Lys
 755 760 765
 Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu Leu
 770 775 780
 Cys Asp Lys Ser Pro Asn Leu Phe Leu Lys Lys Asn Glu Arg Leu Arg
 785 790 795 800
 Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser Ser Met Thr Arg
 805 810 815
 Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val Arg Glu Leu Lys
 820 825 830
 Glu Tyr Ile Gly Asp Ile Cys Thr Val Asp Ser Tyr Phe Ser Ile Tyr
 835 840 845
 His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Glu Asn Asp Thr Lys
 850 855 860
 Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu Lys Asn His Gly
 865 870 875 880
 Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro Phe Gly Tyr Asn
 885 890 895
 Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu Phe Asp Arg Asn
 900 905 910
 Glu Tyr Leu Thr Glu Lys

-continued

915

```

<210> SEQ_ID NO 40
<211> LENGTH: 922
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(922)
<223> OTHER INFORMATION: native CasM protein sequence from Ruminococcus
sp., isolate 2789STDY5608892

<400> SEQUENCE: 40

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1 5 10 15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Asn Ala Ala
20 25 30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
35 40 45

Glu Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
50 55 60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Gly Asn Ser
65 70 75 80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
85 90 95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asp Val Asn Glu
100 105 110

Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Gly Ser Gly Val Glu
115 120 125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
130 135 140

Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
145 150 155 160

Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
165 170 175

Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
180 185 190

Ala Leu Asn Asn Met Leu Gly Ile Lys Asp Ser Glu Ser Tyr Asp Asp
195 200 205

Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Glu Val Phe Thr His
210 215 220

Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn Ile Lys Lys
225 230 235 240

Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg Leu Gly Tyr
245 250 255

Phe Gly Leu Glu Pro Lys Thr Lys Asp Thr Arg Ala Ser Glu Ala
260 265 270

Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
275 280 285

Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
290 295 300

Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp Tyr
305 310 315 320

Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Glu Gly

```

-continued

325	330	335
Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu		
340	345	350
Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser		
355	360	365
Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu		
370	375	380
Glu Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg		
385	390	395
Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr		
405	410	415
Arg Asn Asp Val Ala Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe		
420	425	430
Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala		
435	440	445
Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His		
450	455	460
Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe		
465	470	475
Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp Leu Leu		
485	490	495
Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys		
500	505	510
Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile		
515	520	525
Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys		
530	535	540
Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg Ile Thr		
545	550	555
Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys Pro Ala		
565	570	575
Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile Leu Gly		
580	585	590
Ile Asp Asp Asn Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu Lys Leu		
595	600	605
Lys Glu Lys Gly Lys Gly Ile His Gly Leu Arg Asn Phe Ile Thr Asn		
610	615	620
Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr Ala Asn		
625	630	635
Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val Met Phe		
645	650	655
Val Leu Gly Gly Ile Pro Asp Thr Gln Ile Glu Arg Tyr Tyr Lys Ser		
660	665	670
Cys Val Glu Phe Pro Asp Met Asn Ser Ser Leu Glu Ala Lys Arg Ser		
675	680	685
Glu Leu Ala Arg Met Ile Lys Asn Ile Ser Phe Asp Asp Phe Lys Asn		
690	695	700
Val Lys Gln Gln Ala Lys Gly Arg Glu Asn Val Ala Lys Glu Arg Ala		
705	710	715
Lys Ala Val Ile Gly Leu Tyr Leu Thr Val Met Tyr Leu Leu Val Lys		
725	730	735

-continued

Asn Leu Val Asn Val Asn Ala Arg Tyr Val Ile Ala Ile His Cys Leu
740 745 750

Glu Arg Asp Phe Gly Leu Tyr Lys Glu Ile Ile Pro Glu Leu Ala Ser
755 760 765

Lys Asn Leu Lys Asn Asp Tyr Arg Ile Leu Ser Gln Thr Leu Cys Glu
770 775 780

Leu Cys Asp Asp Arg Asn Glu Ser Ser Asn Leu Phe Leu Lys Lys Asn
785 790 795 800

Lys Arg Leu Arg Lys Cys Val Glu Val Asp Ile Asn Asn Ala Asp Ser
805 810 815

Ser Met Thr Arg Lys Tyr Arg Asn Cys Ile Ala His Leu Thr Val Val
820 825 830

Arg Glu Leu Lys Glu Tyr Ile Gly Asp Ile Arg Thr Val Asp Ser Tyr
835 840 845

Phe Ser Ile Tyr His Tyr Val Met Gln Arg Cys Ile Thr Lys Arg Gly
850 855 860

Asp Asp Thr Lys Gln Glu Glu Lys Ile Lys Tyr Glu Asp Asp Leu Leu
865 870 875 880

Lys Asn His Gly Tyr Thr Lys Asp Phe Val Lys Ala Leu Asn Ser Pro
885 890 895

Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser Ile Glu Gln Leu
900 905 910

Phe Asp Arg Asn Glu Tyr Leu Thr Glu Lys
915 920

<210> SEQ ID NO 41
<211> LENGTH: 922
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(922)
<223> OTHER INFORMATION: native CasM protein sequence from Ruminococcus
sp. CAG:57

<400> SEQUENCE: 41

Met Ala Lys Lys Asn Lys Met Lys Pro Arg Glu Leu Arg Glu Ala Gln
1 5 10 15

Lys Lys Ala Arg Gln Leu Lys Ala Ala Glu Ile Asn Asn Ala Ala
20 25 30

Pro Ala Ile Ala Ala Met Pro Ala Ala Glu Val Ile Ala Pro Val Ala
35 40 45

Glu Lys Lys Ser Ser Val Lys Ala Ala Gly Met Lys Ser Ile Leu
50 55 60

Val Ser Glu Asn Lys Met Tyr Ile Thr Ser Phe Gly Lys Asn Ser
65 70 75 80

Ala Val Leu Glu Tyr Glu Val Asp Asn Asn Asp Tyr Asn Lys Thr Gln
85 90 95

Leu Ser Ser Lys Asp Asn Ser Asn Ile Glu Leu Gly Asp Val Asn Glu
100 105 110

Val Asn Ile Thr Phe Ser Ser Lys His Gly Phe Gly Ser Gly Val Glu
115 120 125

Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser Pro Val
130 135 140

-continued

Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg Phe Phe
 145 150 155 160
 Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr Asn Ile
 165 170 175
 Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile Val Tyr
 180 185 190
 Ala Leu Asn Asn Met Leu Gly Ile Lys Asp Ser Glu Ser Tyr Asp Asp
 195 200 205
 Phe Met Gly Tyr Leu Ser Ala Arg Asn Thr Tyr Glu Val Phe Thr His
 210 215 220
 Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn Ile Lys Lys
 225 230 235 240
 Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg Leu Gly Tyr
 245 250 255
 Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Thr Arg Ala Ser Glu Ala
 260 265 270
 Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln Ile Arg
 275 280 285
 Gln Cys Val Phe His Asp Lys Ser Gly Ala Lys Arg Phe Asp Leu Tyr
 290 295 300
 Ser Phe Ile Asn Asn Ile Asp Pro Glu Tyr Arg Asp Thr Leu Asp Tyr
 305 310 315 320
 Leu Val Glu Glu Arg Leu Lys Ser Ile Asn Lys Asp Phe Ile Glu Gly
 325 330 335
 Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly Tyr Glu
 340 345 350
 Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu Lys Ser
 355 360 365
 Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys Met Leu
 370 375 380
 Glu Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser Val Arg
 385 390 395 400
 Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn Tyr Tyr
 405 410 415
 Arg Asn Asp Val Ala Ala Gly Glu Ala Leu Val Arg Lys Leu Arg Phe
 420 425 430
 Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu Ala Ala
 435 440 445
 Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala Asp His
 450 455 460
 Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met Asp Phe
 465 470 475 480
 Asp Glu Lys Ile Leu Asp Ser Glu Lys Asn Ala Ser Asp Leu Leu
 485 490 495
 Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp Gly Lys
 500 505 510
 Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp Asn Ile
 515 520 525
 Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val Glu Cys
 530 535 540

-continued

Glu	Leu	Thr	Ala	Gly	Tyr	Lys	Leu	Phe	Asn	Asp	Ser	Gln	Arg	Ile	Thr	
545																560
Asn	Glu	Leu	Phe	Ile	Val	Lys	Asn	Ile	Ala	Ser	Met	Arg	Lys	Pro	Ala	
565																575
Ala	Ser	Ala	Lys	Leu	Thr	Met	Phe	Arg	Asp	Ala	Leu	Thr	Ile	Leu	Gly	
580																590
Ile	Asp	Asp	Asn	Ile	Thr	Asp	Asp	Arg	Ile	Ser	Glu	Ile	Leu	Lys	Leu	
595																605
Lys	Glu	Lys	Gly	Lys	Gly	Ile	His	Gly	Leu	Arg	Asn	Phe	Ile	Thr	Asn	
610																620
Asn	Val	Ile	Glu	Ser	Ser	Arg	Phe	Val	Tyr	Leu	Ile	Lys	Tyr	Ala	Asn	
625																640
Ala	Gln	Lys	Ile	Arg	Glu	Val	Ala	Lys	Asp	Glu	Lys	Val	Val	Met	Phe	
645																655
Val	Leu	Gly	Gly	Ile	Pro	Asp	Thr	Gln	Ile	Glu	Arg	Tyr	Tyr	Lys	Ser	
660																670
Cys	Val	Glu	Phe	Pro	Asp	Met	Asn	Ser	Ser	Leu	Glu	Ala	Lys	Arg	Ser	
675																685
Glu	Leu	Ala	Arg	Met	Ile	Lys	Asn	Ile	Ser	Phe	Asp	Asp	Phe	Lys	Asn	
690																700
Val	Lys	Gln	Gln	Ala	Lys	Gly	Arg	Glu	Asn	Val	Ala	Lys	Glu	Arg	Ala	
705																720
Lys	Ala	Val	Ile	Gly	Leu	Tyr	Leu	Thr	Val	Met	Tyr	Leu	Leu	Val	Lys	
725																735
Asn	Leu	Val	Asn	Val	Asn	Ala	Arg	Tyr	Val	Ile	Ala	Ile	His	Cys	Leu	
740																750
Glu	Arg	Asp	Phe	Gly	Leu	Tyr	Lys	Glu	Ile	Ile	Pro	Glu	Leu	Ala	Ser	
755																765
Lys	Asn	Leu	Lys	Asn	Asp	Tyr	Arg	Ile	Leu	Ser	Gln	Thr	Leu	Cys	Glu	
770																780
Leu	Cys	Asp	Asp	Arg	Asn	Glu	Ser	Ser	Asn	Leu	Phe	Leu	Lys	Lys	Asn	
785																800
Lys	Arg	Leu	Arg	Lys	Cys	Val	Glu	Val	Asp	Ile	Asn	Asn	Ala	Asp	Ser	
805																815
Ser	Met	Thr	Arg	Lys	Tyr	Arg	Asn	Cys	Ile	Ala	His	Leu	Thr	Val	Val	
820																830
Arg	Glu	Leu	Lys	Glu	Tyr	Ile	Gly	Asp	Ile	Arg	Thr	Val	Asp	Ser	Tyr	
835																845
Phe	Ser	Ile	Tyr	His	Tyr	Val	Met	Gln	Arg	Cys	Ile	Thr	Lys	Arg	Gly	
850																860
Asp	Asp	Thr	Lys	Gln	Glu	Lys	Ile	Lys	Tyr	Glu	Asp	Asp	Leu	Leu		
865																880
Lys	Asn	His	Gly	Tyr	Thr	Lys	Asp	Phe	Val	Lys	Ala	Leu	Asn	Ser	Pro	
885																895
Phe	Gly	Tyr	Asn	Ile	Pro	Arg	Phe	Lys	Asn	Leu	Ser	Ile	Glu	Gln	Leu	
900																910
Phe	Asp	Arg	Asn	Glu	Tyr	Leu	Thr	Glu	Lys							
915																920

<210> SEQ_ID NO 42
<211> LENGTH: 933
<212> TYPE: PRT

-continued

```

<213> ORGANISM: Ruminococcus flavefaciens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(933)
<223> OTHER INFORMATION: native CasM protein sequence from Ruminococcus
flavefaciens FD-1

<400> SEQUENCE: 42

Met Lys Lys Met Ser Leu Arg Glu Lys Arg Glu Ala Glu Lys Gln
1           5          10          15

Ala Lys Lys Ala Ala Tyr Ser Ala Ala Ser Lys Asn Thr Asp Ser Lys
20          25          30

Pro Ala Glu Lys Lys Ala Glu Thr Pro Lys Pro Ala Glu Ile Ile Ser
35          40          45

Asp Asn Ser Arg Asn Lys Thr Ala Val Lys Ala Ala Gly Leu Lys Ser
50          55          60

Thr Ile Ile Ser Gly Asp Lys Leu Tyr Met Thr Ser Phe Gly Lys Gly
65          70          75          80

Asn Ala Ala Val Ile Glu Gln Lys Ile Asp Ile Asn Asp Tyr Ser Phe
85          90          95

Ser Ala Met Lys Asp Thr Pro Ser Leu Glu Val Asp Lys Ala Glu Ser
100         105         110

Lys Glu Ile Ser Phe Ser Ser His His Pro Phe Val Lys Asn Asp Lys
115         120         125

Leu Thr Thr Tyr Asn Pro Leu Tyr Gly Lys Asp Asn Pro Glu Lys
130         135         140

Pro Val Gly Arg Asp Met Leu Gly Leu Lys Asp Lys Leu Glu Glu Arg
145         150         155         160

Tyr Phe Gly Cys Thr Phe Asn Asp Asn Leu His Ile Gln Ile Ile Tyr
165         170         175

Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val His Ser Ala Asn Ile
180         185         190

Thr Thr Ala Leu Asp His Met Val Asp Glu Asp Asp Glu Lys Tyr Leu
195         200         205

Asn Ser Asp Tyr Ile Gly Tyr Met Asn Thr Ile Asn Thr Tyr Asp Val
210         215         220

Phe Met Asp Pro Ser Lys Asn Ser Ser Leu Ser Pro Lys Asp Arg Lys
225         230         235         240

Asn Ile Asp Asn Ser Arg Ala Lys Phe Glu Lys Leu Leu Ser Thr Lys
245         250         255

Arg Leu Gly Tyr Phe Gly Phe Asp Tyr Asp Ala Asn Gly Lys Asp Lys
260         265         270

Lys Lys Asn Glu Glu Ile Lys Lys Arg Leu Tyr His Leu Thr Ala Phe
275         280         285

Ala Gly Gln Leu Arg Gln Trp Ser Phe His Ser Ala Gly Asn Tyr Pro
290         295         300

Arg Thr Trp Leu Tyr Lys Leu Asp Ser Leu Asp Lys Glu Tyr Leu Asp
305         310         315         320

Thr Leu Asp His Tyr Phe Asp Lys Arg Phe Asn Asp Ile Asn Asp Asp
325         330         335

Phe Val Thr Lys Asn Ala Thr Asn Leu Tyr Ile Leu Lys Glu Val Phe
340         345         350

Pro Glu Ala Asn Phe Lys Asp Ile Ala Asp Leu Tyr Tyr Asp Phe Ile

```

-continued

355	360	365
Val Ile Lys Ser His Lys Asn Met Gly Phe Ser Ile Lys Lys Leu Arg		
370	375	380
Glu Lys Met Leu Glu Cys Asp Gly Ala Asp Arg Ile Lys Glu Gln Asp		
385	390	395
400		
Met Asp Ser Val Arg Ser Lys Leu Tyr Lys Leu Ile Asp Phe Cys Ile		
405	410	415
Phe Lys Tyr Tyr His Glu Phe Pro Glu Leu Ser Glu Lys Asn Val Asp		
420	425	430
Ile Leu Arg Ala Ala Val Ser Asp Thr Lys Lys Asp Asn Leu Tyr Ser		
435	440	445
Asp Glu Ala Ala Arg Leu Trp Ser Ile Phe Lys Glu Lys Phe Leu Gly		
450	455	460
Phe Cys Asp Lys Ile Val Val Trp Val Thr Gly Glu His Glu Lys Asp		
465	470	475
480		
Ile Thr Ser Val Ile Asp Lys Asp Ala Tyr Arg Asn Arg Ser Asn Val		
485	490	495
Ser Tyr Phe Ser Lys Leu Met Tyr Ala Met Cys Phe Phe Leu Asp Gly		
500	505	510
Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn		
515	520	525
Ile Ala Asn Gln Ile Lys Thr Ala Lys Glu Leu Gly Ile Asn Thr Ala		
530	535	540
Phe Val Lys Asn Tyr Asp Phe Phe Asn His Ser Glu Lys Tyr Val Asp		
545	550	555
560		
Glu Leu Asn Ile Val Lys Asn Ile Ala Arg Met Lys Lys Pro Ser Ser		
565	570	575
Asn Ala Lys Lys Ala Met Tyr His Asp Ala Leu Thr Ile Leu Gly Ile		
580	585	590
Pro Glu Asp Met Asp Glu Lys Ala Leu Asp Glu Glu Leu Asp Leu Ile		
595	600	605
Leu Glu Lys Lys Thr Asp Pro Val Thr Gly Lys Pro Leu Lys Gly Lys		
610	615	620
Asn Pro Leu Arg Asn Phe Ile Ala Asn Asn Val Ile Glu Asn Ser Arg		
625	630	635
640		
Phe Ile Tyr Leu Ile Lys Phe Cys Asn Pro Glu Asn Val Arg Lys Ile		
645	650	655
Val Asn Asn Thr Lys Val Thr Glu Phe Val Leu Lys Arg Ile Pro Asp		
660	665	670
Ala Gln Ile Glu Arg Tyr Tyr Lys Ser Cys Thr Asp Ser Glu Met Asn		
675	680	685
Pro Pro Thr Glu Lys Lys Ile Thr Glu Leu Ala Gly Lys Leu Lys Asp		
690	695	700
Met Asn Phe Gly Asn Phe Arg Asn Val Arg Gln Ser Ala Lys Glu Asn		
705	710	715
720		
Met Glu Lys Glu Arg Phe Lys Ala Val Ile Gly Leu Tyr Leu Thr Val		
725	730	735
Val Tyr Arg Val Val Lys Asn Leu Val Asp Val Asn Ser Arg Tyr Ile		
740	745	750
Met Ala Phe His Ser Leu Glu Arg Asp Ser Gln Leu Tyr Asn Val Ser		
755	760	765

-continued

Val Asp Asn Asp Tyr Leu Ala Leu Thr Asp Thr Leu Val Lys Glu Gly
770 775 780

Asp Asn Ser Arg Ser Arg Tyr Leu Ala Gly Asn Lys Arg Leu Arg Asp
785 790 795 800

Cys Val Lys Gln Asp Ile Asp Asn Ala Lys Lys Trp Phe Val Ser Asp
805 810 815

Lys Tyr Asn Ser Ile Thr Lys Tyr Arg Asn Asn Val Ala His Leu Thr
820 825 830

Ala Val Arg Asn Cys Ala Glu Phe Ile Gly Asp Ile Thr Lys Ile Asp
835 840 845

Ser Tyr Phe Ala Leu Tyr His Tyr Leu Ile Gln Arg Gln Leu Ala Lys
850 855 860

Gly Leu Asp His Glu Arg Ser Gly Phe Asp Arg Asn Tyr Pro Gln Tyr
865 870 875 880

Ala Pro Leu Phe Lys Trp His Thr Tyr Val Lys Asp Val Val Lys Ala
885 890 895

Leu Asn Ala Pro Phe Gly Tyr Asn Ile Pro Arg Phe Lys Asn Leu Ser
900 905 910

Ile Asp Ala Leu Phe Asp Arg Asn Glu Ile Lys Lys Asn Asp Gly Glu
915 920 925

Lys Lys Ser Asp Asp
930

<210> SEQ ID NO 43
<211> LENGTH: 944
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus albus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(944)
<223> OTHER INFORMATION: native CasM protein sequence from Ruminococcus
albus strain KH2T6

<400> SEQUENCE: 43

Met Ala Lys Lys Ser Lys Gly Met Ser Leu Arg Glu Lys Arg Glu Leu
1 5 10 15

Glu Lys Gln Lys Arg Ile Gln Lys Ala Ala Val Asn Ser Val Asn Asp
20 25 30

Thr Pro Glu Lys Thr Glu Glu Ala Asn Val Val Ser Val Asn Val Arg
35 40 45

Thr Ser Ala Glu Asn Lys His Ser Lys Lys Ser Ala Ala Lys Ala Leu
50 55 60

Gly Leu Lys Ser Gly Leu Val Ile Gly Asp Glu Leu Tyr Leu Thr Ser
65 70 75 80

Phe Gly Arg Gly Asn Glu Ala Lys Leu Glu Lys Lys Ile Ser Gly Asp
85 90 95

Thr Val Glu Lys Leu Gly Ile Gly Ala Phe Glu Val Ala Glu Arg Asp
100 105 110

Glu Ser Thr Leu Thr Leu Glu Ser Gly Arg Ile Lys Asp Lys Thr Ala
115 120 125

Arg Pro Lys Asp Pro Arg His Ile Thr Val Asp Thr Gln Gly Lys Phe
130 135 140

Lys Glu Asp Met Leu Gly Ile Arg Ser Val Leu Glu Lys Lys Ile Phe
145 150 155 160

-continued

Gly Lys Thr Phe Asp Asp Asn Ile His Val Gln Leu Ala Tyr Asn Ile
165 170 175

Leu Asp Val Glu Lys Ile Met Ala Gln Tyr Val Ser Asp Ile Val Tyr
180 185 190

Met Leu His Asn Thr Asp Lys Thr Glu Arg Asn Asp Asn Leu Met Gly
195 200 205

Tyr Met Ser Ile Arg Asn Thr Tyr Lys Thr Phe Cys Asp Thr Ser Asn
210 215 220

Leu Pro Asp Asp Thr Lys Gln Lys Val Glu Asn Gln Lys Arg Glu Phe
225 230 235 240

Asp Lys Ile Ile Lys Ser Gly Arg Leu Gly Tyr Phe Gly Glu Ala Phe
245 250 255

Met Val Asn Ser Gly Asn Ser Thr Lys Leu Arg Pro Glu Lys Glu Ile
260 265 270

Tyr His Ile Phe Ala Leu Met Ala Ser Leu Arg Gln Ser Tyr Phe His
275 280 285

Gly Tyr Val Lys Asp Thr Asp Tyr Gln Gly Thr Thr Trp Ala Tyr Thr
290 295 300

Leu Glu Asp Lys Leu Lys Gly Pro Ser His Glu Phe Arg Glu Thr Ile
305 310 315 320

Asp Lys Ile Phe Asp Glu Gly Phe Ser Lys Ile Ser Lys Asp Phe Gly
325 330 335

Lys Met Asn Lys Val Asn Leu Gln Ile Leu Glu Gln Met Ile Gly Glu
340 345 350

Leu Tyr Gly Ser Ile Glu Arg Gln Asn Leu Thr Cys Asp Tyr Tyr Asp
355 360 365

Phe Ile Gln Leu Lys Lys His Lys Tyr Leu Gly Phe Ser Ile Lys Arg
370 375 380

Leu Arg Glu Thr Met Leu Glu Thr Thr Pro Ala Glu Cys Tyr Lys Ala
385 390 395 400

Glu Cys Tyr Asn Ser Glu Arg Gln Lys Leu Tyr Lys Leu Ile Asp Phe
405 410 415

Leu Ile Tyr Asp Leu Tyr Tyr Asn Arg Lys Pro Ala Arg Ile Glu Glu
420 425 430

Ile Val Asp Lys Leu Arg Glu Ser Val Asn Asp Glu Glu Lys Glu Ser
435 440 445

Ile Tyr Ser Val Glu Ala Lys Tyr Val Tyr Glu Ser Leu Ser Lys Val
450 455 460

Leu Asp Lys Ser Leu Lys Asn Ser Val Ser Gly Glu Thr Ile Lys Asp
465 470 475 480

Leu Gln Lys Arg Tyr Asp Asp Glu Thr Ala Asn Arg Ile Trp Asp Ile
485 490 495

Ser Gln His Ser Ile Ser Gly Asn Val Asn Cys Phe Cys Lys Leu Ile
500 505 510

Tyr Ile Met Thr Leu Met Leu Asp Gly Lys Glu Ile Asn Asp Leu Leu
515 520 525

Thr Thr Leu Val Asn Lys Phe Asp Asn Ile Ala Ser Phe Ile Asp Val
530 535 540

Met Asp Glu Leu Gly Leu Glu His Ser Phe Thr Asp Asn Tyr Lys Met
545 550 555 560

-continued

Phe	Ala	Asp	Ser	Lys	Ala	Ile	Cys	Leu	Asp	Leu	Gln	Phe	Ile	Asn	Ser
565								570					575		
<hr/>															
Phe	Ala	Arg	Met	Ser	Lys	Ile	Asp	Asp	Glu	Lys	Ser	Lys	Arg	Gln	Leu
580							585				590				
<hr/>															
Phe	Arg	Asp	Ala	Leu	Val	Ile	Leu	Asp	Ile	Gly	Asn	Lys	Asp	Glu	Thr
595							600			605					
<hr/>															
Trp	Ile	Asn	Asn	Tyr	Leu	Asp	Ser	Asp	Ile	Phe	Lys	Leu	Asp	Lys	Glu
610					615				620						
<hr/>															
Gly	Asn	Lys	Leu	Lys	Gly	Ala	Arg	His	Asp	Phe	Arg	Asn	Phe	Ile	Ala
625					630			635				640			
<hr/>															
Asn	Asn	Val	Ile	Lys	Ser	Ser	Arg	Phe	Lys	Tyr	Leu	Val	Lys	Tyr	Ser
645					650			655							
<hr/>															
Ser	Ala	Asp	Gly	Met	Ile	Lys	Leu	Lys	Thr	Asn	Glu	Lys	Leu	Ile	Gly
660					665			670							
<hr/>															
Phe	Val	Leu	Asp	Lys	Leu	Pro	Glu	Thr	Gln	Ile	Asp	Arg	Tyr	Tyr	Glu
675					680			685							
<hr/>															
Ser	Cys	Gly	Leu	Asp	Asn	Ala	Val	Val	Asp	Lys	Lys	Val	Arg	Ile	Glu
690					695			700							
<hr/>															
Lys	Leu	Ser	Gly	Leu	Ile	Arg	Asp	Met	Lys	Phe	Asp	Asp	Phe	Ser	Gly
705					710			715				720			
<hr/>															
Val	Lys	Thr	Ser	Asn	Lys	Ala	Gly	Asp	Asn	Asp	Lys	Gln	Asp	Lys	Ala
725					730			735							
<hr/>															
Lys	Tyr	Gln	Ala	Ile	Ile	Ser	Leu	Tyr	Leu	Met	Val	Leu	Tyr	Gln	Ile
740					745			750							
<hr/>															
Val	Lys	Asn	Met	Ile	Tyr	Val	Asn	Ser	Arg	Tyr	Val	Ile	Ala	Phe	His
755					760			765							
<hr/>															
Cys	Leu	Glu	Arg	Asp	Phe	Gly	Met	Tyr	Gly	Lys	Asp	Phe	Gly	Lys	Tyr
770					775			780							
<hr/>															
Tyr	Gln	Gly	Cys	Arg	Lys	Leu	Thr	Asp	His	Phe	Ile	Glu	Glu	Lys	Tyr
785					790			795				800			
<hr/>															
Met	Lys	Glu	Lys	Leu	Gly	Cys	Asn	Lys	Lys	Val	Gly	Arg	Tyr	Leu	
805					810			815							
<hr/>															
Lys	Asn	Asn	Ile	Ser	Cys	Cys	Thr	Asp	Gly	Leu	Ile	Asn	Thr	Tyr	Arg
820					825			830							
<hr/>															
Asn	Gln	Val	Asp	His	Phe	Ala	Val	Val	Arg	Lys	Ile	Gly	Asn	Tyr	Ala
835					840			845							
<hr/>															
Ala	Tyr	Ile	Lys	Ser	Ile	Gly	Ser	Trp	Phe	Glu	Leu	Tyr	His	Tyr	Val
850					855			860							
<hr/>															
Ile	Gln	Arg	Ile	Val	Phe	Asp	Glu	Tyr	Arg	Phe	Ala	Leu	Asn	Asn	Thr
865					870			875				880			
<hr/>															
Glu	Ser	Asn	Tyr	Lys	Asn	Ser	Ile	Ile	Lys	His	His	Thr	Tyr	Cys	Lys
885					890			895							
<hr/>															
Asp	Met	Val	Lys	Ala	Leu	Asn	Thr	Pro	Phe	Gly	Tyr	Asp	Leu	Pro	Arg
900					905			910							
<hr/>															
Tyr	Lys	Asn	Leu	Ser	Ile	Gly	Asp	Leu	Phe	Asp	Arg	Asn	Asn	Tyr	Leu
915					920			925							
<hr/>															
Asn	Lys	Thr	Lys	Glu	Ser	Ile	Asp	Ala	Asn	Ser	Ser	Ile	Asp	Ser	Gln
930					935			940							

<210> SEQ_ID NO 44

<211> LENGTH: 967

<212> TYPE: PRT

-continued

<213> ORGANISM: Ruminococcus flavefaciens
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1) .. (967)
 <223> OTHER INFORMATION: native CasM protein sequence from Ruminococcus
 flavefaciens strain XPD3002

<400> SEQUENCE: 44

Met Ile Glu Lys Lys Ser Phe Ala Lys Gly Met Gly Val Lys Ser
 1 5 10 15

Thr Leu Val Ser Gly Ser Lys Val Tyr Met Thr Thr Phe Ala Glu Gly
 20 25 30

Ser Asp Ala Arg Leu Glu Lys Ile Val Glu Gly Asp Ser Ile Arg Ser
 35 40 45

Val Asn Glu Gly Glu Ala Phe Ser Ala Glu Met Ala Asp Lys Asn Ala
 50 55 60

Gly Tyr Lys Ile Gly Asn Ala Lys Phe Ser His Pro Lys Gly Tyr Ala
 65 70 75 80

Val Val Ala Asn Asn Pro Leu Tyr Thr Gly Pro Val Gln Gln Asp Met
 85 90 95

Leu Gly Leu Lys Glu Thr Leu Glu Lys Arg Tyr Phe Gly Glu Ser Ala
 100 105 110

Asp Gly Asn Asp Asn Ile Cys Ile Gln Val Ile His Asn Ile Leu Asp
 115 120 125

Ile Glu Lys Ile Leu Ala Glu Tyr Ile Thr Asn Ala Ala Tyr Ala Val
 130 135 140

Asn Asn Ile Ser Gly Leu Asp Lys Asp Ile Ile Gly Phe Gly Lys Phe
 145 150 155 160

Ser Thr Val Tyr Thr Tyr Asp Glu Phe Lys Asp Pro Glu His His Arg
 165 170 175

Ala Ala Phe Asn Asn Asp Lys Leu Ile Asn Ala Ile Lys Ala Gln
 180 185 190

Tyr Asp Glu Phe Asp Asn Phe Leu Asp Asn Pro Arg Leu Gly Tyr Phe
 195 200 205

Gly Gln Ala Phe Phe Ser Lys Glu Gly Arg Asn Tyr Ile Ile Asn Tyr
 210 215 220

Gly Asn Glu Cys Tyr Asp Ile Leu Ala Leu Ser Gly Leu Arg His
 225 230 235 240

Trp Val Val His Asn Asn Glu Glu Ser Arg Ile Ser Arg Thr Trp
 245 250 255

Leu Tyr Asn Leu Asp Lys Asn Leu Asp Asn Glu Tyr Ile Ser Thr Leu
 260 265 270

Asn Tyr Leu Tyr Asp Arg Ile Thr Asn Glu Leu Thr Asn Ser Phe Ser
 275 280 285

Lys Asn Ser Ala Ala Asn Val Asn Tyr Ile Ala Glu Thr Leu Gly Ile
 290 295 300

Asn Pro Ala Glu Phe Ala Glu Gln Tyr Phe Arg Phe Ser Ile Met Lys
 305 310 315 320

Glu Gln Lys Asn Leu Gly Phe Asn Ile Thr Lys Leu Arg Glu Val Met
 325 330 335

Leu Asp Arg Lys Asp Met Ser Glu Ile Arg Lys Asn His Lys Val Phe
 340 345 350

Asp Ser Ile Arg Thr Lys Val Tyr Thr Met Met Asp Phe Val Ile Tyr

-continued

355	360	365	
Arg Tyr Tyr Ile Glu Glu Asp Ala Lys Val Ala Ala Ala Asn Lys Ser			
370	375	380	
Leu Pro Asp Asn Glu Lys Ser Leu Ser Glu Lys Asp Ile Phe Val Ile			
385	390	395	400
Asn Leu Arg Gly Ser Phe Asn Asp Asp Gln Lys Asp Ala Leu Tyr Tyr			
405	410	415	
Asp Glu Ala Asn Arg Ile Trp Arg Lys Leu Glu Asn Ile Met His Asn			
420	425	430	
Ile Lys Glu Phe Arg Gly Asn Lys Thr Arg Glu Tyr Lys Lys Lys Asp			
435	440	445	
Ala Pro Arg Leu Pro Arg Ile Leu Pro Ala Gly Arg Asp Val Ser Ala			
450	455	460	
Phe Ser Lys Leu Met Tyr Ala Leu Thr Met Phe Leu Asp Gly Lys Glu			
465	470	475	480
Ile Asn Asp Leu Leu Thr Thr Leu Ile Asn Lys Phe Asp Asn Ile Gln			
485	490	495	
Ser Phe Leu Lys Val Met Pro Leu Ile Gly Val Asn Ala Lys Phe Val			
500	505	510	
Glu Glu Tyr Ala Phe Phe Lys Asp Ser Ala Lys Ile Ala Asp Glu Leu			
515	520	525	
Arg Leu Ile Lys Ser Phe Ala Arg Met Gly Glu Pro Ile Ala Asp Ala			
530	535	540	
Arg Arg Ala Met Tyr Ile Asp Ala Ile Arg Ile Leu Gly Thr Asn Leu			
545	550	555	560
Ser Tyr Asp Glu Leu Lys Ala Leu Ala Asp Thr Phe Ser Leu Asp Glu			
565	570	575	
Asn Gly Asn Lys Leu Lys Lys Gly Lys His Gly Met Arg Asn Phe Ile			
580	585	590	
Ile Asn Asn Val Ile Ser Asn Lys Arg Phe His Tyr Leu Ile Arg Tyr			
595	600	605	
Gly Asp Pro Ala His Leu His Glu Ile Ala Lys Asn Glu Ala Val Val			
610	615	620	
Lys Phe Val Leu Gly Arg Ile Ala Asp Ile Gln Lys Lys Gln Gly Gln			
625	630	635	640
Asn Gly Lys Asn Gln Ile Asp Arg Tyr Tyr Glu Thr Cys Ile Gly Lys			
645	650	655	
Asp Lys Gly Lys Ser Val Ser Glu Lys Val Asp Ala Leu Thr Lys Ile			
660	665	670	
Ile Thr Gly Met Asn Tyr Asp Gln Phe Asp Lys Lys Arg Ser Val Ile			
675	680	685	
Glu Asp Thr Gly Arg Glu Asn Ala Glu Arg Glu Lys Phe Lys Lys Ile			
690	695	700	
Ile Ser Leu Tyr Leu Thr Val Ile Tyr His Ile Leu Lys Asn Ile Val			
705	710	715	720
Asn Ile Asn Ala Arg Tyr Val Ile Gly Phe His Cys Val Glu Arg Asp			
725	730	735	
Ala Gln Leu Tyr Lys Glu Lys Gly Tyr Asp Ile Asn Leu Lys Lys Leu			
740	745	750	
Glu Glu Lys Gly Phe Ser Ser Val Thr Lys Leu Cys Ala Gly Ile Asp			
755	760	765	

-continued

Glu Thr Ala Pro Asp Lys Arg Lys Asp Val Glu Lys Glu Met Ala Glu
770 775 780

Arg Ala Lys Glu Ser Ile Asp Ser Leu Glu Ser Ala Asn Pro Lys Leu
785 790 795 800

Tyr Ala Asn Tyr Ile Lys Tyr Ser Asp Glu Lys Lys Ala Glu Glu Phe
805 810 815

Thr Arg Gln Ile Asn Arg Glu Lys Ala Lys Thr Ala Leu Asn Ala Tyr
820 825 830

Leu Arg Asn Thr Lys Trp Asn Val Ile Ile Arg Glu Asp Leu Leu Arg
835 840 845

Ile Asp Asn Lys Thr Cys Thr Leu Phe Arg Asn Lys Ala Val His Leu
850 855 860

Glu Val Ala Arg Tyr Val His Ala Tyr Ile Asn Asp Ile Ala Glu Val
865 870 875 880

Asn Ser Tyr Phe Gln Leu Tyr His Tyr Ile Met Gln Arg Ile Ile Met
885 890 895

Asn Glu Arg Tyr Glu Lys Ser Ser Gly Lys Val Ser Glu Tyr Phe Asp
900 905 910

Ala Val Asn Asp Glu Lys Lys Tyr Asn Asp Arg Leu Leu Lys Leu Leu
915 920 925

Cys Val Pro Phe Gly Tyr Cys Ile Pro Arg Phe Lys Asn Leu Ser Ile
930 935 940

Glu Ala Leu Phe Asp Arg Asn Glu Ala Ala Lys Phe Asp Lys Glu Lys
945 950 955 960

Lys Lys Val Ser Gly Asn Ser
965

<210> SEQ ID NO 45
<211> LENGTH: 796
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(796)
<223> OTHER INFORMATION: native CasM protein sequence from Ruminococcus
sp., isolate 2789STDY5834894

<400> SEQUENCE: 45

Met Glu Ile Asn Thr Ser Asn Pro Thr His Arg Ser Gly Glu Ser Ser
1 5 10 15

Ser Val Arg Gly Asp Met Leu Gly Leu Lys Ser Glu Leu Glu Lys Arg
20 25 30

Phe Phe Gly Lys Thr Phe Asp Asp Asn Ile His Ile Gln Leu Ile Tyr
35 40 45

Asn Ile Leu Asp Ile Glu Lys Ile Leu Ala Val Tyr Val Thr Asn Ile
50 55 60

Val Tyr Ala Leu Asn Asn Met Leu Gly Val Lys Gly Ser Glu Ser Tyr
65 70 75 80

Asp Asp Phe Met Gly Tyr Leu Ser Ala Gln Asn Thr Tyr Tyr Ile Phe
85 90 95

Thr His Pro Asp Lys Ser Asn Leu Ser Asp Lys Val Lys Gly Asn Ile
100 105 110

Lys Lys Ser Leu Ser Lys Phe Asn Asp Leu Leu Lys Thr Lys Arg Leu
115 120 125

-continued

Gly Tyr Phe Gly Leu Glu Glu Pro Lys Thr Lys Asp Lys Arg Val Ser
130 135 140

Glu Ala Tyr Lys Lys Arg Val Tyr His Met Leu Ala Ile Val Gly Gln
145 150 155 160

Ile Arg Gln Ser Val Phe His Asp Lys Ser Asn Glu Leu Asp Glu Tyr
165 170 175

Leu Tyr Ser Phe Ile Asp Ile Ile Asp Ser Glu Tyr Arg Asp Thr Leu
180 185 190

Asp Tyr Leu Val Asp Glu Arg Phe Asp Ser Ile Asn Lys Gly Phe Val
195 200 205

Gln Gly Asn Lys Val Asn Ile Ser Leu Leu Ile Asp Met Met Lys Gly
210 215 220

Tyr Glu Ala Asp Asp Ile Ile Arg Leu Tyr Tyr Asp Phe Ile Val Leu
225 230 235 240

Lys Ser Gln Lys Asn Leu Gly Phe Ser Ile Lys Lys Leu Arg Glu Lys
245 250 255

Met Leu Asp Glu Tyr Gly Phe Arg Phe Lys Asp Lys Gln Tyr Asp Ser
260 265 270

Val Arg Ser Lys Met Tyr Lys Leu Met Asp Phe Leu Leu Phe Cys Asn
275 280 285

Tyr Tyr Arg Asn Asp Val Val Ala Gly Glu Ala Leu Val Arg Lys Leu
290 295 300

Arg Phe Ser Met Thr Asp Asp Glu Lys Glu Gly Ile Tyr Ala Asp Glu
305 310 315 320

Ala Glu Lys Leu Trp Gly Lys Phe Arg Asn Asp Phe Glu Asn Ile Ala
325 330 335

Asp His Met Asn Gly Asp Val Ile Lys Glu Leu Gly Lys Ala Asp Met
340 345 350

Asp Phe Asp Glu Lys Ile Leu Asp Ser Glu Lys Lys Asn Ala Ser Asp
355 360 365

Leu Leu Tyr Phe Ser Lys Met Ile Tyr Met Leu Thr Tyr Phe Leu Asp
370 375 380

Gly Lys Glu Ile Asn Asp Leu Leu Thr Thr Leu Ile Ser Lys Phe Asp
385 390 395 400

Asn Ile Lys Glu Phe Leu Lys Ile Met Lys Ser Ser Ala Val Asp Val
405 410 415

Glu Cys Glu Leu Thr Ala Gly Tyr Lys Leu Phe Asn Asp Ser Gln Arg
420 425 430

Ile Thr Asn Glu Leu Phe Ile Val Lys Asn Ile Ala Ser Met Arg Lys
435 440 445

Pro Ala Ala Ser Ala Lys Leu Thr Met Phe Arg Asp Ala Leu Thr Ile
450 455 460

Leu Gly Ile Asp Asp Lys Ile Thr Asp Asp Arg Ile Ser Glu Ile Leu
465 470 475 480

Lys Leu Lys Glu Lys Gly Lys Ile His Gly Leu Arg Asn Phe Ile
485 490 495

Thr Asn Asn Val Ile Glu Ser Ser Arg Phe Val Tyr Leu Ile Lys Tyr
500 505 510

Ala Asn Ala Gln Lys Ile Arg Glu Val Ala Lys Asn Glu Lys Val Val
515 520 525

-continued

Met	Phe	Val	Leu	Gly	Gly	Ile	Pro	Asp	Thr	Gln	Ile	Glu	Arg	Tyr	Tyr
530						535					540				
Lys	Ser	Cys	Val	Glu	Phe	Pro	Asp	Met	Asn	Ser	Ser	Leu	Glu	Ala	Lys
545						550					555				560
Cys	Ser	Glu	Leu	Ala	Arg	Met	Ile	Lys	Asn	Ile	Ser	Phe	Asp	Asp	Phe
						565				570					575
Lys	Asn	Val	Lys	Gln	Gln	Ala	Lys	Gly	Arg	Glu	Asn	Val	Ala	Lys	Glu
						580				585					590
Arg	Ala	Lys	Ala	Val	Ile	Gly	Leu	Tyr	Leu	Thr	Val	Met	Tyr	Leu	Leu
						595			600			605			
Val	Lys	Asn	Leu	Val	Asn	Val	Ala	Arg	Tyr	Val	Ile	Ala	Ile	His	
						610			615			620			
Cys	Leu	Glu	Arg	Asp	Phe	Gly	Leu	Tyr	Lys	Glu	Ile	Ile	Pro	Glu	Leu
625						630				635				640	
Ala	Ser	Lys	Asn	Leu	Lys	Asn	Asp	Tyr	Arg	Ile	Leu	Ser	Gln	Thr	Leu
						645			650			655			
Cys	Glu	Leu	Cys	Asp	Asp	Arg	Asp	Glu	Ser	Pro	Asn	Leu	Phe	Leu	Lys
						660			665			670			
Lys	Asn	Lys	Arg	Leu	Arg	Lys	Cys	Val	Glu	Val	Asp	Ile	Asn	Asn	Ala
						675			680			685			
Asp	Ser	Ser	Met	Thr	Arg	Lys	Tyr	Arg	Asn	Cys	Ile	Ala	His	Leu	Thr
						690			695			700			
Val	Val	Arg	Glu	Leu	Lys	Glu	Tyr	Ile	Gly	Asp	Ile	Arg	Thr	Val	Asp
						705			710			715			720
Ser	Tyr	Phe	Ser	Ile	Tyr	His	Tyr	Val	Met	Gln	Arg	Cys	Ile	Thr	Lys
						725			730			735			
Arg	Glu	Asp	Asp	Thr	Lys	Gln	Glu	Lys	Ile	Lys	Tyr	Glu	Asp	Asp	
						740			745			750			
Leu	Leu	Lys	Asn	His	Gly	Tyr	Thr	Lys	Asp	Phe	Val	Lys	Ala	Leu	Asn
						755			760			765			
Ser	Pro	Phe	Gly	Tyr	Asn	Ile	Pro	Arg	Phe	Lys	Asn	Leu	Ser	Ile	Glu
						770			775			780			
Gln	Leu	Phe	Asp	Arg	Asn	Glu	Tyr	Leu	Thr	Glu	Lys				
						785			790			795			

```

<210> SEQ_ID NO 46
<211> LENGTH: 96
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CRISPR Arrays Component

<400> SEQUENCE: 46

ugauacugcu uugaugucag cauugcauau cuacuaucu ggugcgaauu ugcacuaguc      60
uaaaaucuau aaccuaagu ucuucugcgu ucauau                                96

```

```

<210> SEQ_ID NO 47
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CRISPR Arrays Component

<400> SEQUENCE: 47

ugauacugcu uugaugucag cauugcauau cuacuaucu ggugcgaauu ugcacuaguc      60

```

-continued

aaaaau	66
<210> SEQ ID NO 48	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CRISPR Arrays Component	
<400> SEQUENCE: 48	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaauugau acugcuuuga ugucagcauu	60
gcauau	66
<210> SEQ ID NO 49	
<211> LENGTH: 45	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: Forward primer	
<400> SEQUENCE: 49	
cgaatataat acgactcact ataggtttcg attatgcggc cgtgt	45
<210> SEQ ID NO 50	
<211> LENGTH: 22	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: Reverse primer	
<400> SEQUENCE: 50	
aggagatata ccatgggcag ca	22
<210> SEQ ID NO 51	
<211> LENGTH: 36	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM repeat sequence	
<400> SEQUENCE: 51	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaau	36
<210> SEQ ID NO 52	
<211> LENGTH: 824	
<212> TYPE: PRT	
<213> ORGANISM: Eubacterium siraeum	
<220> FEATURE:	
<221> NAME/KEY: misc_feature	
<222> LOCATION: (1)..(824)	
<223> OTHER INFORMATION: WYL Eubacterium siraeum	
<400> SEQUENCE: 52	
Met Lys Lys Thr Glu Lys Phe Asp Asp Val Gln Ser Gly Tyr Glu Tyr 1 5 10 15	
Lys Tyr Phe Leu Glu Ser Ile Asp Lys Tyr Arg Ala Ala Val Gln Asn 20 25 30	
Ile Tyr Thr Tyr Gly Cys Phe Asn Gln Lys Gln Leu Ser Glu Gln Cys 35 40 45	
Asn Cys Ser Asp Gln Thr Ile Lys Lys Ala Phe Asn Phe Tyr Asn Leu 50 55 60	

-continued

Cys Leu Ala Asn Tyr Ile Lys Lys Lys Gly Thr Leu Ser Lys Lys
65 70 75 80

Ala Lys Gly Arg Pro Thr Glu Ala Lys Tyr Leu Glu Tyr Asp Arg Phe
85 90 95

Thr Leu Asn Glu Asn Tyr Leu Tyr Asn Ile Tyr Leu Trp Ala Arg Ile
100 105 110

Thr Lys Lys Gln Met Trp Ala Phe Ser Tyr Phe Arg Arg His Thr Ser
115 120 125

Leu Leu Ile Asn Ala Ser Arg Thr Glu Ile Lys Asn Gln Leu Ser Asp
130 135 140

Phe Phe Leu Tyr Phe Ser Glu Tyr Met Asp Arg Ser Lys Lys Ala Glu
145 150 155 160

Asn Ser Gln Asp Leu Gly Tyr Ile Ile Asp Met Thr Ala Pro Thr Glu
165 170 175

Lys Asn Met Leu Ile Ser Ser Met Cys Asp Ala Leu Ala Val Phe Gly
180 185 190

Arg Lys Ala Pro Tyr Ser Val Pro Ala Tyr Ser Ile Ser His Lys Leu
195 200 205

Lys Lys Leu Cys Gly Asn Asp Ser Lys Ser Leu Trp Ser Phe Met Tyr
210 215 220

Asp Asn Tyr Asp Arg Ile Leu Tyr Asp Glu Ala Val Tyr Thr Ile Arg
225 230 235 240

Gln Ala Ile Arg Asp Arg Lys Leu Ile Gly Tyr Gln Thr Val Gly Thr
245 250 255

Glu Lys Gln Lys Ser Val Asn Tyr Val Val Pro Leu Lys Ile Met Tyr
260 265 270

Glu Tyr Asn Leu Gly Arg Cys Tyr Leu Leu Tyr Ser Pro Leu Asn Ser
275 280 285

Asp Ser Ile Ile Lys Ser Ile Arg Leu Asp Lys Leu Tyr Lys Val Ala
290 295 300

Ala Tyr Glu Pro Asp Ser Ile Ile Asn Tyr Glu Lys Leu Tyr Asp Val
305 310 315 320

Leu Ala Val Ala Glu Asn Glu Ile Trp Leu Ser Gly Asp Tyr Thr Lys
325 330 335

Lys Asp Cys Leu Ser Arg Ile Val Leu Lys Asn Val Lys Pro Gln Ala
340 345 350

Phe Ser Leu Ile Glu Lys Tyr Gly Val Cys Tyr Thr Glu Asp Arg Glu
355 360 365

Ala Lys Thr Val Thr Phe Asn Ile Arg Lys Ala Asp Asp Ile Lys Pro
370 375 380

Phe Ile Arg Thr Leu Gly Gly Asp Ala Val Ile Ser Glu Glu Asp Asn
385 390 395 400

Pro Gly Leu Phe Arg Glu Phe Ala Tyr Asp Ala Arg Ile Gly Arg Gln
405 410 415

Met Tyr Tyr Asp Asp Ser Phe Ala Asp Cys Pro Ala Glu Lys Asp Ser
420 425 430

Gln Pro Ala Lys Asp Ser Lys Thr Ala Ser Gly Asn Asp Asn Ile Lys
435 440 445

Lys Tyr Ala Ser Tyr Pro Thr Leu Arg Leu Phe Asn Lys Tyr Gly Ser
450 455 460

-continued

Phe	Met	Asn	Ile	Leu	Ala	Glu	Glu	Leu	Ala	Glu	His	Ile	Phe	Ser	Glu
465				470				475							480
Ile Ile Arg Met Pro Val Glu Lys Arg Ala Gly Gln Ile Glu Tyr Ser															
485					490				495						
Ser Asn Arg Leu Glu Arg Val Leu Asn Ser Tyr Phe Lys Ile Tyr Gly															
500					505				510						
Phe Asp Glu Leu Arg Thr Glu Ala Ser Asn Ile Thr Glu Trp Phe Thr															
515					520				525						
Lys Ala Thr Glu Glu Leu Ser Asp Ser Asp Tyr Ser Ser Trp Phe Ser															
530					535				540						
Val Asn Gly Gly Lys Phe Glu Ala Val Ala Asp Leu Asn Glu Tyr Glu															
545					550				555				560		
His Lys Gln Leu Leu Thr Asn Ile Glu Tyr Glu Tyr Leu Arg Leu Met															
565					570				575						
Leu Gly Asp Pro Asp Ala Arg Ala Ile Ile Gly Asn Glu Tyr Cys Glu															
580					585				590						
Lys Leu Ser Glu Tyr Val Gly Ser Ala Asp Thr Thr Leu Asp Glu Phe															
595					600				605						
Phe Thr Val Arg Tyr Ala Asn Arg Asn Glu Lys Thr Ile Glu Asn Lys															
610					615				620						
His Ser Val Leu Arg Thr Ile Met Arg Ala Met Asn Asn Glu Lys Lys															
625					630				635				640		
Ala Asp Ile Glu Tyr Lys Gly Lys His Tyr Ile Cys Ser Ala Tyr Arg															
645					650				655						
Phe Thr Tyr Ser Leu Arg Glu Arg Lys His Arg Leu Met Val Phe Asp															
660					665				670						
Gly Asn Tyr Ile Met Gln Ile Asn Leu Cys Asp Ile Lys Asp Ala Gln															
675					680				685						
Met Thr Lys Glu Pro Ser Leu Ser Asp Glu Glu Met Asn Lys Leu Leu															
690					695				700						
Thr Glu Arg Lys Tyr Ile Glu Ile Ala Ile Pro Gln Asn Ala Asp															
705					710				715				720		
Ala Gln Gln Arg Asn Val Phe Glu Arg Ala Leu Arg Leu Phe Gly Gly															
725					730				735						
Phe Glu Arg Tyr Ser Trp Asn Asp Ala Lys Asn Gly Glu Tyr Val Ile															
740					745				750						
Ala Val Ala Tyr Tyr Glu Pro Asp Ile Ser Val Ser Ser Ala Asp															
755					760				765						
Arg Arg Ile Tyr Arg Arg Asp Thr Val Ala Ala Asp Ile Met Ser Leu															
770					775				780						
Gly Arg Tyr Ala Arg Val Met Lys Gln Pro Gly Phe Glu Leu Asp Gly															
785					790				795				800		
Val Arg Tyr Asp Ser Ser Leu Tyr Asp Tyr Ile Ser Lys Asn Tyr Ser															
805					810				815						
Gly Thr Ala Ala Arg Tyr Glu Lys															
820															

```

<210> SEQ ID NO 53
<211> LENGTH: 389
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature

```

-continued

<222> LOCATION: (1)...(389)

<223> OTHER INFORMATION: WYL Ruminococcus sp.isolate 2789STDY5834971

<400> SEQUENCE: 53

Met	Leu	Ile	Leu	Pro	Ser	Thr	Phe	Leu	Pro	Lys	Arg	Asp	Lys	Asn	Val
1				5				10			15				

Pro	Tyr	Ile	Ala	Glu	Val	Gln	Ser	Ile	Pro	Leu	Ser	Pro	Ser	Ala	Tyr
		20					25				30				

Ser	Val	Ile	Ile	Lys	Asp	Lys	Ser	Ile	Phe	Glu	Thr	Ser	Leu	Ser	Pro
		35				40				45					

Asn	Gly	Ser	Val	Ser	Met	Ser	Ser	Phe	Leu	Thr	Ser	Ile	Phe	Asp	Ser
		50			55				60						

Ala	Tyr	Ile	Ala	Ser	Leu	Lys	Tyr	Lys	Ser	Glu	Lys	Tyr	Asn	Gly	Ile
65				70			75			80					

Pro	Leu	Leu	Asn	Ala	Phe	Val	Lys	Trp	Gln	Ile	Glu	Glu	Ile	Asn	Asp
			85				90				95				

Gly	Leu	Asp	Asp	Lys	Ser	Lys	Glu	Ile	Ile	Lys	Ser	Tyr	Leu	Ile	Ser
			100			105				110					

Lys	Leu	Ser	Ala	Lys	Tyr	Glu	Lys	Thr	Lys	Thr	Glu	Asn	Ala	Val	Arg
		115			120			125							

Val	Arg	Leu	Ser	Ile	Cys	Arg	Asp	Leu	Tyr	Asp	Thr	Leu	Ser	Ser	Asp
		130			135			140							

Asp	Leu	Tyr	Tyr	Glu	Asn	Lys	Val	Tyr	Ser	Ser	Thr	Leu	Arg	Arg	Phe
145				150			155				160				

Leu	Lys	Ala	Val	Tyr	Glu	Asp	Tyr	Ala	Leu	Leu	Ser	Asp	Cys	Glu	Arg
		165			170			175							

Glu	Arg	Leu	Ile	Phe	Ala	Asp	Asn	Ile	Ile	Lys	Ile	Asn	Glu	Val	Ile
		180			185			190							

Lys	Gln	Asn	Gly	Ser	Arg	Tyr	Tyr	Ser	Phe	Ile	Tyr	Ala	Tyr	Ser	Asn
		195			200			205							

Met	Tyr	Ser	Arg	Glu	Lys	Arg	Arg	Ile	Arg	Leu	Ile	Pro	Tyr	Arg	Ile
		210			215			220							

Val	Ser	Asp	Glu	Tyr	Lys	Met	Tyr	Asn	Tyr	Leu	Val	Cys	Leu	Ser	Asp
225			230			235			240						

Glu	Lys	Ser	Ala	Gly	Lys	Glu	Phe	Lys	Ala	Asp	Ser	Tyr	Arg	Ile	Ser
		245			250			255							

Arg	Leu	Ser	Gly	Leu	Ser	Ile	Ala	Glu	Lys	Leu	Ser	Gln	Lys	Glu	Tyr
		260			265			270							

Ser	Ser	Val	Thr	Glu	Tyr	Glu	Arg	Leu	Lys	Glu	Gly	His	Val	Lys	Ser
		275			280			285							

Val	Lys	His	Leu	Leu	Ser	Asp	Pro	Arg	Phe	Gly	Ser	Asp	Glu	Ser	Asp
		290			295			300							

Ile	Ser	Lys	Val	Tyr	Leu	Thr	Glu	Lys	Val	Glu	Met	Phe	Gly	Lys	
305				310			315			320					

Ile	Leu	Tyr	Gln	Arg	Pro	Ile	Leu	Lys	Gly	Asn	Glu	Lys	Pro	Lys	Pro
		325			330			335							

Asn	Ala	Val	Asn	Glu	Phe	Ile	Ser	Pro	Pro	Ile	Gln	Val	Lys	Tyr	Tyr
		340			345			350							

Phe	Asn	Lys	Phe	Gly	Lys	Asp	Gly	Val	Ile	Leu	Ser	Pro	Ser	Asp	Ser
		355			360			365							

Phe	Glu	Glu	Met	Arg	Thr	Leu	Tyr	Val	Glu	Gly	Ala	Glu	Ala	Tyr	Asn
		370			375			380							

-continued

Arg Glu Val Glu Met
385

<210> SEQ_ID NO 54
<211> LENGTH: 392
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus bicirculans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(392)
<223> OTHER INFORMATION: WYL Ruminococcus bicirculans

<400> SEQUENCE: 54

Met Ser Met Thr Pro Ser Thr Phe Leu Pro Lys Arg Glu Asp Gly Val
1 5 10 15

Pro Tyr Ile Ala Glu Val Gln Ser Ile Pro Leu Ser Pro Ser Ala Tyr
20 25 30

Ser Val Ile Ile Lys Asp Lys Ser Ile Phe Glu Thr Ser Leu Ser Pro
35 40 45

Asn Gly Ser Val Ser Met Ser Ser Phe Leu Thr Ser Ile Phe Asp Ser
50 55 60

Ala Tyr Ile Ala Ser Leu Lys Tyr Lys Ser Asp Asp Asn Tyr Lys Tyr
65 70 75 80

Ile Gly Ile Pro Leu Leu Asn Ala Phe Val Lys Trp Gln Ile Glu Glu
85 90 95

Ile Asp Asp Ser Leu Asp Asp Lys Ser Lys Glu Ile Ile Lys Ser Tyr
100 105 110

Leu Ile Ser Lys Leu Ser Ala Lys Tyr Glu Lys Thr Lys Thr Glu Asn
115 120 125

Ala Val Arg Val Arg Leu Ser Ile Cys Arg Asp Leu Tyr Asp Thr Leu
130 135 140

Ser Ser Asp Asp Leu Tyr Tyr Glu Asn Lys Val Tyr Ser Ser Thr Leu
145 150 155 160

Arg Arg Phe Leu Lys Ala Val Tyr Glu Asp Tyr Ala Leu Leu Ser Asp
165 170 175

Cys Glu Arg Glu Arg Leu Ile Phe Ala Asp Asn Ile Ile Lys Ile Asn
180 185 190

Glu Val Ile Lys Gln Asn Gly Ser Arg Tyr Tyr Ser Phe Ile Tyr Ala
195 200 205

Tyr Ser Asn Met Tyr Ser Arg Glu Lys Arg Arg Ile Arg Leu Ile Pro
210 215 220

Tyr Arg Ile Val Ser Asp Glu Tyr Lys Met Tyr Asn Tyr Leu Val Cys
225 230 235 240

Leu Ser Asp Glu Lys Ser Ala Gly Lys Glu Phe Lys Ala Asp Ser Tyr
245 250 255

Arg Ile Ser Arg Leu Ser Gly Leu Ser Ile Ala Glu Lys Leu Ser Gln
260 265 270

Lys Glu Tyr Ser Ser Val Thr Glu Tyr Glu Arg Leu Lys Glu Gly His
275 280 285

Val Lys Ser Val Lys His Leu Leu Ser Asp Pro Arg Phe Gly Ser Asp
290 295 300

Glu Ser Asp Ile Ser Lys Val Tyr Leu Thr Glu Lys Gly Val Glu Met
305 310 315 320

-continued

Phe	Gly	Lys	Ile	Leu	Tyr	Gln	Arg	Pro	Ile	Leu	Lys	Gly	Asn	Glu	Lys
325															335
Pro	Lys	Pro	Asn	Ala	Val	Asn	Glu	Phe	Ile	Ser	Pro	Pro	Ile	Gln	Val
340															350
Lys	Tyr	Tyr	Phe	Asn	Lys	Phe	Gly	Lys	Asp	Gly	Val	Ile	Leu	Ser	Pro
355															365
Ser	Asp	Ser	Phe	Glu	Glu	Met	Arg	Thr	Leu	Tyr	Val	Glu	Gly	Ala	Glu
370															380
Ala	Tyr	Asn	Arg	Glu	Val	Glu	Met								
385															390
<210> SEQ_ID NO 55															
<211> LENGTH: 392															
<212> TYPE: PRT															
<213> ORGANISM: Ruminococcus sp.															
<220> FEATURE:															
<221> NAME/KEY: misc_feature															
<222> LOCATION: (1)..(392)															
<223> OTHER INFORMATION: WYL Ruminococcus sp. isolate 2789STDY5608892															
<400> SEQUENCE: 55															
Met	Leu	Ile	Pro	Pro	Ser	Thr	Phe	Leu	Pro	Lys	Arg	Asp	Lys	Asn	Val
1															15
Pro	Tyr	Ile	Ala	Glu	Val	Gln	Ser	Ile	Pro	Leu	Ser	Pro	Ser	Ala	Tyr
	20														30
Ser	Val	Ile	Ile	Lys	Asp	Lys	Ser	Ile	Phe	Glu	Thr	Ser	Leu	Ser	Pro
	35														45
Asn	Gly	Ser	Val	Ser	Met	Ser	Ser	Phe	Leu	Thr	Ser	Ile	Phe	Asp	Ser
	50														60
Ala	Tyr	Ile	Ala	Ser	Leu	Lys	Tyr	Lys	Ser	Asp	Asp	Asn	Tyr	Lys	Tyr
	65														80
Ile	Gly	Ile	Pro	Leu	Leu	Asn	Ala	Phe	Val	Glu	Trp	Gln	Ile	Glu	Glu
															95
Ile	Asp	Asp	Ser	Leu	Asp	Asp	Lys	Ser	Lys	Glu	Ile	Ile	Lys	Ser	Tyr
	100														110
Leu	Ile	Ser	Lys	Leu	Ser	Ala	Lys	Tyr	Glu	Lys	Thr	Lys	Thr	Glu	Asn
	115														125
Ala	Val	Arg	Val	Leu	Ser	Ile	Cys	Arg	Asp	Leu	Tyr	Asp	Thr	Leu	
	130														140
Ser	Ser	Asp	Asp	Leu	Tyr	Tyr	Glu	Asn	Lys	Tyr	Ser	Leu	Thr	Leu	
	145														160
Arg	Arg	Phe	Leu	Lys	Ala	Val	Tyr	Glu	Asp	Tyr	Ala	Leu	Leu	Ser	Asp
	165														175
Cys	Glu	Arg	Glu	Arg	Leu	Ile	Phe	Ala	Asp	Asn	Ile	Ile	Lys	Ile	Asn
	180														190
Glu	Val	Ile	Lys	Gln	Asn	Gly	Ser	Arg	Tyr	Tyr	Ser	Phe	Ile	Tyr	Ala
	195														205
Tyr	Ser	Asn	Met	Tyr	Ser	Arg	Glu	Lys	Arg	Arg	Ile	Arg	Leu	Ile	Pro
	210														220
Tyr	Arg	Ile	Val	Ser	Asp	Glu	Tyr	Lys	Met	Tyr	Asn	Tyr	Leu	Val	Cys
	225														240
Leu	Ser	Asp	Glu	Lys	Ser	Ala	Gly	Lys	Glu	Phe	Lys	Ala	Asp	Ser	Tyr
	245														255
Arg	Ile	Ser	Arg	Leu	Ser	Gly	Leu	Ser	Ile	Ala	Glu	Lys	Leu	Ser	Gln

-continued

260	265	270	
Lys Glu Tyr Ser Ser Val Thr Glu Tyr Glu Arg Leu Lys Glu Gly His			
275	280	285	
Val Lys Ser Val Lys His Leu Leu Ser Asp Pro Arg Phe Gly Ser Asp			
290	295	300	
Glu Ser Asp Ile Ser Lys Val Tyr Leu Thr Glu Lys Gly Val Glu Met			
305	310	315	320
Phe Gly Lys Ile Leu Tyr Gln Arg Pro Ile Leu Lys Gly Asn Glu Lys			
325	330	335	
Pro Lys Pro Asn Thr Val Asn Glu Phe Ile Ser Pro Pro Ile Gln Val			
340	345	350	
Lys Tyr Tyr Phe Asn Lys Phe Gly Lys Asp Gly Val Ile Leu Ser Pro			
355	360	365	
Ser Asp Ser Phe Glu Glu Met Arg Thr Leu Tyr Val Glu Gly Ala Glu			
370	375	380	
Ala Tyr Asn Arg Glu Val Glu Met			
385	390		

<210> SEQ ID NO 56
<211> LENGTH: 392
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus sp.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(392)
<223> OTHER INFORMATION: WYL Ruminococcus sp. CAG:57

<400> SEQUENCE: 56			
Met Leu Ile Pro Pro Ser Thr Phe Leu Pro Lys Arg Asp Lys Asn Val			
1	5	10	15
Pro Tyr Ile Ala Glu Val Gln Ser Ile Pro Leu Ser Pro Ser Ala Tyr			
20	25	30	
Ser Val Ile Ile Lys Asp Lys Ser Ile Phe Glu Thr Ser Leu Ser Pro			
35	40	45	
Asn Gly Ser Val Ser Met Ser Ser Phe Leu Thr Ser Ile Phe Asp Ser			
50	55	60	
Ala Tyr Ile Ala Ser Leu Lys Tyr Lys Ser Asp Asp Asn Tyr Lys Tyr			
65	70	75	80
Ile Gly Ile Pro Leu Leu Asn Ala Phe Val Glu Trp Gln Ile Glu Glu			
85	90	95	
Ile Asp Asp Ser Leu Asp Asp Lys Ser Lys Glu Ile Ile Lys Ser Tyr			
100	105	110	
Leu Ile Ser Lys Leu Ser Ala Lys Tyr Glu Lys Thr Lys Thr Glu Asn			
115	120	125	
Ala Val Arg Val Arg Leu Ser Ile Cys Arg Asp Leu Tyr Asp Thr Leu			
130	135	140	
Ser Ser Asp Asp Leu Tyr Tyr Glu Asn Lys Val Tyr Ser Leu Thr Leu			
145	150	155	160
Arg Arg Phe Leu Lys Ala Val Tyr Glu Asp Tyr Ala Leu Leu Ser Asp			
165	170	175	
Cys Glu Arg Glu Arg Leu Ile Phe Ala Asp Asn Ile Ile Lys Ile Asn			
180	185	190	
Glu Val Ile Lys Gln Asn Gly Ser Arg Tyr Tyr Ser Phe Ile Tyr Ala			
195	200	205	

-continued

Tyr Ser Asn Met Tyr Ser Arg Glu Lys Arg Arg Ile Arg Leu Ile Pro
210 215 220

Tyr Arg Ile Val Ser Asp Glu Tyr Lys Met Tyr Asn Tyr Leu Val Cys
225 230 235 240

Leu Ser Asp Glu Lys Ser Ala Gly Lys Glu Phe Lys Ala Asp Ser Tyr
245 250 255

Arg Ile Ser Arg Leu Ser Gly Leu Ser Ile Ala Glu Lys Leu Ser Gln
260 265 270

Lys Glu Tyr Ser Ser Val Thr Glu Tyr Glu Arg Leu Lys Glu Gly His
275 280 285

Val Lys Ser Val Lys His Leu Leu Ser Asp Pro Arg Phe Gly Ser Asp
290 295 300

Glu Ser Asp Ile Ser Lys Val Tyr Leu Thr Glu Lys Gly Val Glu Met
305 310 315 320

Phe Gly Lys Ile Leu Tyr Gln Arg Pro Ile Leu Lys Gly Asn Glu Lys
325 330 335

Pro Lys Pro Asn Thr Val Asn Glu Phe Ile Ser Pro Pro Ile Gln Val
340 345 350

Lys Tyr Tyr Phe Asn Lys Phe Gly Lys Asp Gly Val Ile Leu Ser Pro
355 360 365

Ser Asp Ser Phe Glu Glu Met Arg Thr Leu Tyr Val Glu Gly Ala Glu
370 375 380

Ala Tyr Asn Arg Glu Val Glu Met
385 390

<210> SEQ ID NO 57
<211> LENGTH: 280
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus flavefaciens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(280)
<223> OTHER INFORMATION: WYL Ruminococcus flavefaciens FD-1

<400> SEQUENCE: 57

Met Ile Ile Ala Ile Asn Gln Trp Lys Arg Arg Phe Ser Leu Val Ile
1 5 10 15

Tyr Gly Lys Ser Glu Gly Glu Thr Ile Val Lys Ile Lys Leu Leu Leu
20 25 30

Ile Ser Leu Ala Tyr Leu Ile Ser Ile Tyr Leu Leu Cys Ser Pro Gly
35 40 45

Cys Ile Gly Ile Phe Thr His Gly Met Leu Thr Thr Val Ile Gly Val
50 55 60

Val Thr Met Leu Ala Ala Thr Gly Thr Tyr Gly Met Tyr Leu Tyr Ser
65 70 75 80

Ser Ala Ile Gly Glu Arg Ser Leu Pro Glu Ile Pro Met Asn Lys Glu
85 90 95

Thr Glu Tyr Ser Arg Tyr Lys Glu Leu Glu Asn Trp Phe Arg Ala Phe
100 105 110

Arg Tyr Leu Asp Arg Asn Asn Phe Ala Met Leu Ser Ser Asp Leu
115 120 125

Ala Thr Ser Tyr His Asp Gly Leu Ile Arg Asp Asn Pro Phe Arg Asn
130 135 140

-continued

Thr	Glu	Leu	Gly	Asp	Arg	Leu	Gln	Thr	Thr	Ser	Ser	Asp	Ile	Ser	Ile
145						150			155						160
Lys	Tyr	Asp	Gln	Thr	Leu	Lys	Ile	Leu	Ser	Glu	Ser	Phe	Glu	Lys	Asn
							165		170						175
Asp	Ile	Thr	Tyr	Gln	Asn	Tyr	Leu	Ser	Val	Leu	Asp	Asn	Val	Leu	Lys
							180		185						190
Leu	Ser	Ser	His	Leu	Lys	Ala	Ile	Lys	Lys	Arg	Val	Cys	Val	Phe	
							195		200						205
Asp	Tyr	Arg	Thr	Trp	Ala	Asp	Asn	Lys	Asn	Asp	Glu	Met	Cys	Arg	Lys
						210		215							220
Tyr	Ile	Glu	Glu	Val	Lys	Ser	Ser	Val	Ile	Arg	Leu	Glu	Glu	Ile	Glu
						225		230		235					240
Gly	Lys	Phe	Asp	Asn	Leu	Leu	His	Glu	Leu	Ile	Cys	Leu	Ser	Glu	Ile
						245		250		255					255
Ser	Glu	Asp	Pro	Leu	Leu	Glu	Met	Gln	Asp	Leu	Ile	Glu	Thr	Thr	Ser
						260		265							270
Asp	Tyr	Lys	Ser	Ile	Glu	Asp	Gln								
				275		280									

<210> SEQ_ID NO 58
<211> LENGTH: 226
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus albus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(226)
<223> OTHER INFORMATION: WYL Ruminococcus albus strain KH2T6

<400> SEQUENCE: 58

Met	Cys	Thr	Trp	Tyr	Tyr	Ala	Glu	Ala	Lys	Ser	Leu	Ser	Phe	Phe	Ile
1						5			10						15
Asp	Lys	Ala	Ser	Gln	Leu	Pro	Leu	Ser	Asp	Ile	Ile	Met	Asn	Thr	Met
						20		25							30
Ser	Lys	Ser	Lys	Ala	Met	Ser	Gly	Asn	Ile	Arg	Pro	Thr	Asp	Met	Ala
						35		40		45					
Ala	Val	Leu	Ala	Pro	Asn	Lys	Gln	Gly	Asn	Val	Ala	Val	Phe	Pro	Met
						50		55		60					
Ile	Trp	Gly	Phe	Thr	His	Glu	Ser	Thr	Ser	Lys	Pro	Val	Ile	Asn	Cys
						65		70		75					80
Arg	Ile	Glu	Ser	Ala	Asp	Thr	Lys	Pro	Leu	Trp	Lys	Asp	Ser	Trp	Tyr
						85		90							95
Arg	Arg	Arg	Cys	Val	Ile	Pro	Ala	Ser	Trp	Tyr	Tyr	Glu	Trp	Gly	Val
						100		105							110
Pro	Pro	Ser	Glu	Gly	Glu	Leu	Tyr	His	Lys	Asn	Glu	Tyr	Asn	Lys	Ile
						115		120		125					
Gln	Lys	Glu	Lys	Tyr	Ala	Ile	Gln	Pro	Glu	Gly	Ala	Glu	Ile	Thr	Tyr
						130		135		140					
Leu	Ala	Gly	Leu	Tyr	Arg	Phe	Glu	Glu	His	Arg	Gly	Val	Gln	Val	Pro
						145		150		155					160
Met	Phe	Ala	Val	Ile	Thr	Arg	Glu	Ser	Val	Glu	Pro	Val	Ser	Ser	Ile
						165		170		175					175
His	Asp	Arg	Met	Pro	Leu	Ile	Leu	Gly	Lys	Asp	Ser	Leu	Ser	Glu	Trp
						180		185							190
Ile	His	Pro	Asn	Gly	Asp	Pro	Asn	Lys	Ile	Ala	Lys	Thr	Ala	Leu	Thr

-continued

195

200

205

Lys Met Val Met Glu Lys Ala Ile Asp Tyr Pro Glu Pro Glu Pro Ser
210 215 220

Phe Met
225

<210> SEQ ID NO 59
<211> LENGTH: 314
<212> TYPE: PRT
<213> ORGANISM: Ruminococcus flavefaciens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(314)
<223> OTHER INFORMATION: WYL Ruminococcus flavefaciens strain XPD3002

<400> SEQUENCE: 59

Met Glu Leu Phe Asn Glu Tyr Arg Asn Lys Ser Leu Arg Ala Phe Leu
1 5 10 15

Lys Leu Ala Glu Arg Ile Ser Tyr Gly Glu Glu Leu Ser Ile Asp Glu
20 25 30

Phe Glu Ala Glu Tyr Tyr Arg Leu Ser Gly Asp Asn Lys Lys Ile Thr
35 40 45

Ser Val Phe Tyr Lys Asn Thr Leu Tyr Asn Asp Lys Leu Pro Ile Phe
50 55 60

Asp Thr Arg Glu Gly Lys Val Arg Leu Phe Gly Glu Pro Asp Lys Cys
65 70 75 80

Ser Asn Lys His Ile Ser Asp Thr Leu Leu Lys Ser Glu Ile Thr Trp
85 90 95

Leu His Asn Ala Leu Asn Asp Lys Leu Ser Lys Leu Phe Leu Ser Asp
100 105 110

Glu Glu Arg Ile Ser Ile Asp Ala Lys Leu Ser Asp Tyr Thr Glu Tyr
115 120 125

Tyr Lys Asn Ile Asp Asp Met Trp Arg Ser Asn Glu Asp Ile Ser Glu
130 135 140

Glu Val Glu Lys Asn Phe Lys Ile Ile Leu Lys Ala Ile Asn Glu Lys
145 150 155 160

Gln Ala Leu Ser Tyr Thr Phe Lys Asn Lys Asn Cys Glu Gly Phe Pro
165 170 175

Val Arg Ile Glu Tyr Asp Glu Arg Thr Cys Arg Ile Tyr Met Ile Ile
180 185 190

Tyr Asp Gly Asn Arg Phe Val Lys Ser Asp Ile Ser Lys Leu Ser Asp
195 200 205

Ile Tyr Ile Thr Glu Asn Ser Ile Asp Thr Ile Pro Glu Ile Lys Asp
210 215 220

Asp Met Leu Asn Lys Lys Ala Tyr Leu Pro Val Val Phe Thr Val Thr
225 230 235 240

Asp Asp Lys Asn Arg Lys Ala Ile Asp Arg Ala Leu Leu Ala Phe Ser
245 250 255

Val Tyr Asp His Val Val Glu Pro Ile Asp Glu Lys Thr Ala Arg Phe
260 265 270

Thr Ile Gln Tyr Tyr Thr Met Asp Leu Asp Leu Leu Ile Lys Asp Ile
275 280 285

Leu Ala Phe Gly Ser Asp Ile Lys Val Glu Ser Pro Arg Tyr Val Val
290 295 300

-continued

Lys Arg Ile Thr Asp Ile Leu Arg Lys Val
305 310

<210> SEQ_ID NO 60
<211> LENGTH: 412
<212> TYPE: PRT
<213> ORGANISM: Eubacterium siraeum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(412)
<223> OTHER INFORMATION: RtcB Eubacterium siraeum

<400> SEQUENCE: 60

Met Ile Val Leu Glu Ile Ile Gly Glu Arg Asn Thr Ala Val Val Tyr
1 5 10 15

Gly Glu Ile Ile Asp Glu Cys Ala Val Ser Gln Ile Glu Glu Ile Cys
20 25 30

Asn His Pro Ala Phe Glu Asn Ser Arg Ile Arg Ile Met Pro Asp Cys
35 40 45

His Ala Gly Lys Gly Cys Val Ile Gly Phe Thr Cys Val Thr Ser Asn
50 55 60

Arg Met Ile Val Pro Asn Ile Val Gly Val Asp Ile Gly Cys Gly Ile
65 70 75 80

Leu Thr Thr Val Phe Thr Ala Asp Arg Glu Ile Asp Tyr Arg Ala Leu
85 90 95

Asp Thr Phe Ile Arg Ser Asn Ile Pro Ser Gly Met Glu Ile His Asp
100 105 110

Ser Val Ser Asp Thr Val Ala Glu Asn Thr Ala Leu Ile Ala Lys Val
115 120 125

Asn Gly Ile Cys Asp Ala Ile Gly Glu Ser Ala Asp Val Asp Tyr His
130 135 140

Leu Arg Ser Ile Gly Thr Leu Gly Gly Asn His Phe Ile Glu Ile
145 150 155 160

Asp Arg Leu Asn Asn Gly Asn Tyr Ala Leu Thr Val His Thr Gly Ser
165 170 175

Arg Asn Leu Gly Lys Arg Ile Cys Gly Tyr Phe Gln Ser Asn Ala Ser
180 185 190

Val Ile Asp Thr Glu Leu Arg Arg Ser Ile Leu Leu Arg His Arg Ser
195 200 205

Ala Thr Thr Ser Glu Glu His Glu Glu Ile Asp Arg Arg Ala Ala Gln
210 215 220

Ile Ala Pro Val Ser Lys Glu Leu Ala Phe Ile Thr Gly Glu Arg Tyr
225 230 235 240

Asp Ser Tyr Ile Gly Cys Met Leu Asp Ala Lys Ala Leu Ala Ala Phe
245 250 255

Asn Arg Thr Val Ile Ser Asp Arg Ile Met Ser Phe Leu Ala Asp Glu
260 265 270

Tyr Gly Val Glu Ile Lys Asp Arg Phe Asp Thr Val His Asn Tyr Ile
275 280 285

Asp Trp Tyr Asp Asp Thr His Thr Ser Val Val Ile Arg Lys Gly Ala
290 295 300

Ile Ser Ala Arg Lys Gly Glu Arg Ile Val Ile Pro Leu Asn Met Arg
305 310 315 320

-continued

Asp Gly Ile Ile Ile Ala His Gly Arg Gly Asn Glu Glu Trp Asn Cys
325 330 335

Ser Ala Pro His Gly Ser Gly Arg Ala Tyr Ser Arg Ser Asp Ala Arg
340 345 350

Arg Thr Phe Thr Leu Glu Glu Tyr Val Glu Glu Met Asp Gly Val Asn
355 360 365

Thr Trp Ser Val Ser Glu Ser Thr Ile Asp Glu Cys Pro Met Ala Tyr
370 375 380

Lys Pro Ser Glu Met Ile Ile Gly Ser Ile Gly Asp Thr Val Glu Ile
385 390 395 400

Glu Ser Ile Ala His Thr Val Tyr Asn Phe Lys Ala
405 410

<210> SEQ ID NO 61

<211> LENGTH: 831

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic: WYL Eubacterium siraeum + C-term NLS

<400> SEQUENCE: 61

Met Lys Lys Thr Glu Lys Phe Asp Asp Val Gln Ser Gly Tyr Glu Tyr
1 5 10 15

Lys Tyr Phe Leu Glu Ser Ile Asp Lys Tyr Arg Ala Ala Val Gln Asn
20 25 30

Ile Tyr Thr Tyr Gly Cys Phe Asn Gln Lys Gln Leu Ser Glu Gln Cys
35 40 45

Asn Cys Ser Asp Gln Thr Ile Lys Lys Ala Phe Asn Phe Tyr Asn Leu
50 55 60

Cys Leu Ala Asn Tyr Ile Lys Lys Lys Gly Thr Leu Ser Lys Lys
65 70 75 80

Ala Lys Gly Arg Pro Thr Glu Ala Lys Tyr Leu Glu Tyr Asp Arg Phe
85 90 95

Thr Leu Asn Glu Asn Tyr Leu Tyr Asn Ile Tyr Leu Trp Ala Arg Ile
100 105 110

Thr Lys Lys Gln Met Trp Ala Phe Ser Tyr Phe Arg Arg His Thr Ser
115 120 125

Leu Leu Ile Asn Ala Ser Arg Thr Glu Ile Lys Asn Gln Leu Ser Asp
130 135 140

Phe Phe Leu Tyr Phe Ser Glu Tyr Met Asp Arg Ser Lys Lys Ala Glu
145 150 155 160

Asn Ser Gln Asp Leu Gly Tyr Ile Ile Asp Met Thr Ala Pro Thr Glu
165 170 175

Lys Asn Met Leu Ile Ser Ser Met Cys Asp Ala Leu Ala Val Phe Gly
180 185 190

Arg Lys Ala Pro Tyr Ser Val Pro Ala Tyr Ser Ile Ser His Lys Leu
195 200 205

Lys Lys Leu Cys Gly Asn Asp Ser Lys Ser Leu Trp Ser Phe Met Tyr
210 215 220

Asp Asn Tyr Asp Arg Ile Leu Tyr Asp Glu Ala Val Tyr Thr Ile Arg
225 230 235 240

Gln Ala Ile Arg Asp Arg Lys Leu Ile Gly Tyr Gln Thr Val Gly Thr
245 250 255

-continued

Glu	Lys	Gln	Lys	Ser	Val	Asn	Tyr	Val	Val	Pro	Leu	Lys	Ile	Met	Tyr
260					265						270				
Glu	Tyr	Asn	Leu	Gly	Arg	Cys	Tyr	Leu	Leu	Tyr	Ser	Pro	Leu	Asn	Ser
275					280						285				
Asp	Ser	Ile	Ile	Lys	Ser	Ile	Arg	Leu	Asp	Lys	Leu	Tyr	Lys	Val	Ala
290					295					300					
Ala	Tyr	Glu	Pro	Asp	Ser	Ile	Ile	Asn	Tyr	Glu	Lys	Leu	Tyr	Asp	Val
305					310					315				320	
Leu	Ala	Val	Ala	Glu	Asn	Glu	Ile	Trp	Leu	Ser	Gly	Asp	Tyr	Thr	Lys
325					330					335					
Lys	Asp	Cys	Leu	Ser	Arg	Ile	Val	Leu	Lys	Asn	Val	Lys	Pro	Gln	Ala
340					345					350					
Phe	Ser	Leu	Ile	Glu	Lys	Tyr	Gly	Val	Cys	Tyr	Thr	Glu	Asp	Arg	Glu
355					360					365					
Ala	Lys	Thr	Val	Thr	Phe	Asn	Ile	Arg	Lys	Ala	Asp	Asp	Ile	Lys	Pro
370					375					380					
Phe	Ile	Arg	Thr	Leu	Gly	Gly	Asp	Ala	Val	Ile	Ser	Glu	Asp	Asn	
385					390					395				400	
Pro	Gly	Leu	Phe	Arg	Glu	Phe	Ala	Tyr	Asp	Ala	Arg	Ile	Gly	Arg	Gln
405					410					415					
Met	Tyr	Tyr	Asp	Asp	Ser	Phe	Ala	Asp	Cys	Pro	Ala	Glu	Lys	Asp	Ser
420					425					430					
Gln	Pro	Ala	Lys	Asp	Ser	Lys	Thr	Ala	Ser	Gly	Asn	Asp	Asn	Ile	Lys
435					440					445					
Lys	Tyr	Ala	Ser	Tyr	Pro	Thr	Leu	Arg	Leu	Phe	Asn	Lys	Tyr	Gly	Ser
450					455					460					
Phe	Met	Asn	Ile	Leu	Ala	Glu	Glu	Leu	Ala	Glu	His	Ile	Phe	Ser	Glu
465					470					475				480	
Ile	Ile	Arg	Met	Pro	Val	Glu	Lys	Arg	Ala	Gly	Gln	Ile	Glu	Tyr	Ser
485					490					495					
Ser	Asn	Arg	Leu	Glu	Arg	Val	Leu	Asn	Ser	Tyr	Phe	Lys	Ile	Tyr	Gly
500					505					510					
Phe	Asp	Glu	Leu	Arg	Thr	Glu	Ala	Ser	Asn	Ile	Thr	Glu	Trp	Phe	Thr
515					520					525					
Lys	Ala	Thr	Glu	Glu	Leu	Ser	Asp	Ser	Asp	Tyr	Ser	Ser	Trp	Phe	Ser
530					535					540					
Val	Asn	Gly	Gly	Lys	Phe	Glu	Ala	Val	Ala	Asp	Leu	Asn	Glu	Tyr	Glu
545					550					555				560	
His	Lys	Gln	Leu	Leu	Thr	Asn	Ile	Glu	Tyr	Glu	Tyr	Leu	Arg	Leu	Met
565					570					575					
Leu	Gly	Asp	Pro	Asp	Ala	Arg	Ala	Ile	Ile	Gly	Asn	Glu	Tyr	Cys	Glu
580					585					590					
Lys	Leu	Ser	Glu	Tyr	Val	Gly	Ser	Ala	Asp	Thr	Thr	Leu	Asp	Glu	Phe
595					600					605					
Phe	Thr	Val	Arg	Tyr	Ala	Asn	Arg	Asn	Glu	Lys	Thr	Ile	Glu	Asn	Lys
610					615					620					
His	Ser	Val	Leu	Arg	Thr	Ile	Met	Arg	Ala	Met	Asn	Asn	Glu	Lys	Lys
625					630					635				640	
Ala	Asp	Ile	Glu	Tyr	Lys	Gly	Lys	His	Tyr	Ile	Cys	Ser	Ala	Tyr	Arg
645					650					655					
Phe	Thr	Tyr	Ser	Leu	Arg	Glu	Arg	Lys	His	Arg	Leu	Met	Val	Phe	Asp

-continued

660	665	670	
Gly Asn Tyr Ile Met Gln Ile Asn Leu Cys Asp Ile Lys Asp Ala Gln			
675	680	685	
Met Thr Lys Glu Pro Ser Leu Ser Asp Glu Glu Met Asn Lys Leu Leu			
690	695	700	
Thr Glu Arg Lys Lys Tyr Ile Glu Ile Ala Ile Pro Gln Asn Ala Asp			
705	710	715	720
Ala Gln Gln Arg Asn Val Phe Glu Arg Ala Leu Arg Leu Phe Gly Gly			
725	730	735	
Phe Glu Arg Tyr Ser Trp Asn Asp Ala Lys Asn Gly Glu Tyr Val Ile			
740	745	750	
Ala Val Ala Tyr Tyr Glu Pro Asp Ile Ser Val Ser Ser Ser Ala Asp			
755	760	765	
Arg Arg Ile Tyr Arg Arg Asp Thr Val Ala Ala Asp Ile Met Ser Leu			
770	775	780	
Gly Arg Tyr Ala Arg Val Met Lys Gln Pro Gly Phe Glu Leu Asp Gly			
785	790	795	800
Val Arg Tyr Asp Ser Ser Leu Tyr Asp Tyr Ile Ser Lys Asn Tyr Ser			
805	810	815	
Gly Thr Ala Ala Arg Tyr Glu Lys Pro Lys Lys Lys Arg Lys Val			
820	825	830	

```

<210> SEQ ID NO: 62
<211> LENGTH: 396
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: WYL Ruminococcus sp. isolate
2789STDY5834971 + C-term NLS

```

<400> SEQUENCE: 62

Met Leu Ile Leu Pro Ser Thr Phe Leu Pro Lys Arg Asp Lys Asn Val			
1	5	10	15
Pro Tyr Ile Ala Glu Val Gln Ser Ile Pro Leu Ser Pro Ser Ala Tyr			
20	25	30	
Ser Val Ile Ile Lys Asp Lys Ser Ile Phe Glu Thr Ser Leu Ser Pro			
35	40	45	
Asn Gly Ser Val Ser Met Ser Ser Phe Leu Thr Ser Ile Phe Asp Ser			
50	55	60	
Ala Tyr Ile Ala Ser Leu Lys Tyr Lys Ser Glu Lys Tyr Asn Gly Ile			
65	70	75	80
Pro Leu Leu Asn Ala Phe Val Lys Trp Gln Ile Glu Glu Ile Asn Asp			
85	90	95	
Gly Leu Asp Asp Lys Ser Lys Glu Ile Ile Lys Ser Tyr Leu Ile Ser			
100	105	110	
Lys Leu Ser Ala Lys Tyr Glu Lys Thr Lys Thr Glu Asn Ala Val Arg			
115	120	125	
Val Arg Leu Ser Ile Cys Arg Asp Leu Tyr Asp Thr Leu Ser Ser Asp			
130	135	140	
Asp Leu Tyr Tyr Glu Asn Lys Val Tyr Ser Ser Thr Leu Arg Arg Phe			
145	150	155	160
Leu Lys Ala Val Tyr Glu Asp Tyr Ala Leu Leu Ser Asp Cys Glu Arg			
165	170	175	

-continued

Glu Arg Leu Ile Phe Ala Asn Ile Ile Lys Ile Asn Glu Val Ile
 180 185 190
 Lys Gln Asn Gly Ser Arg Tyr Tyr Ser Phe Ile Tyr Ala Tyr Ser Asn
 195 200 205
 Met Tyr Ser Arg Glu Lys Arg Arg Ile Arg Leu Ile Pro Tyr Arg Ile
 210 215 220
 Val Ser Asp Glu Tyr Lys Met Tyr Asn Tyr Leu Val Cys Leu Ser Asp
 225 230 235 240
 Glu Lys Ser Ala Gly Lys Glu Phe Lys Ala Asp Ser Tyr Arg Ile Ser
 245 250 255
 Arg Leu Ser Gly Leu Ser Ile Ala Glu Lys Leu Ser Gln Lys Glu Tyr
 260 265 270
 Ser Ser Val Thr Glu Tyr Glu Arg Leu Lys Glu Gly His Val Lys Ser
 275 280 285
 Val Lys His Leu Leu Ser Asp Pro Arg Phe Gly Ser Asp Glu Ser Asp
 290 295 300
 Ile Ser Lys Val Tyr Leu Thr Glu Lys Gly Val Glu Met Phe Gly Lys
 305 310 315 320
 Ile Leu Tyr Gln Arg Pro Ile Leu Lys Gly Asn Glu Lys Pro Lys Pro
 325 330 335
 Asn Ala Val Asn Glu Phe Ile Ser Pro Pro Ile Gln Val Lys Tyr Tyr
 340 345 350
 Phe Asn Lys Phe Gly Lys Asp Gly Val Ile Leu Ser Pro Ser Asp Ser
 355 360 365
 Phe Glu Glu Met Arg Thr Leu Tyr Val Glu Gly Ala Glu Ala Tyr Asn
 370 375 380
 Arg Glu Val Glu Met Pro Lys Lys Lys Arg Lys Val
 385 390 395

```
<210> SEQ ID NO 63
<211> LENGTH: 399
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: WYL Ruminococcus bicirculans +
C-term NLS
```

<400> SEQUENCE: 63

Met	Ser	Met	Thr	Pro	Ser	Thr	Phe	Leu	Pro	Lys	Arg	Glu	Asp	Gly	Val
1						5				10					15

Pro Tyr Ile Ala Glu Val Gln Ser Ile Pro Leu Ser Pro Ser Ala Tyr
20 25 30

Ser Val Ile Ile Lys Asp Lys Ser Ile Phe Glu Thr Ser Leu Ser Pro
35 40 45

Asn	Gly	Ser	Val	Ser	Met	Ser	Ser	Phe	Leu	Thr	Ser	Ile	Phe	Asp	Ser
50						55					60				

Ala Tyr Ile Ala Ser Leu Lys Tyr Lys Ser Asp Asp Asn Tyr Lys Tyr
65 70 75 80

Ile Gly Ile Pro Leu Leu Asn Ala Phe Val Lys Trp Gln Ile Glu Glu
85 90 95

Ile Asp Asp Ser Leu Asp Asp Lys Ser Lys Glu Ile Ile Lys Ser Tyr
 100 105 110

Leu Ile Ser Lys Leu Ser Ala Lys Tyr Glu Lys Thr Lys Thr Glu Asn
115 120 125

-continued

Ala Val Arg Val Arg Leu Ser Ile Cys Arg Asp Leu Tyr Asp Thr Leu
 130 135 140

Ser Ser Asp Asp Leu Tyr Tyr Glu Asn Lys Val Tyr Ser Ser Thr Leu
 145 150 155 160

Arg Arg Phe Leu Lys Ala Val Tyr Glu Asp Tyr Ala Leu Leu Ser Asp
 165 170 175

Cys Glu Arg Glu Arg Leu Ile Phe Ala Asp Asn Ile Ile Lys Ile Asn
 180 185 190

Glu Val Ile Lys Gln Asn Gly Ser Arg Tyr Tyr Ser Phe Ile Tyr Ala
 195 200 205

Tyr Ser Asn Met Tyr Ser Arg Glu Lys Arg Arg Ile Arg Leu Ile Pro
 210 215 220

Tyr Arg Ile Val Ser Asp Glu Tyr Lys Met Tyr Asn Tyr Leu Val Cys
 225 230 235 240

Leu Ser Asp Glu Lys Ser Ala Gly Lys Glu Phe Lys Ala Asp Ser Tyr
 245 250 255

Arg Ile Ser Arg Leu Ser Gly Leu Ser Ile Ala Glu Lys Leu Ser Gln
 260 265 270

Lys Glu Tyr Ser Ser Val Thr Glu Tyr Glu Arg Leu Lys Glu Gly His
 275 280 285

Val Lys Ser Val Lys His Leu Leu Ser Asp Pro Arg Phe Gly Ser Asp
 290 295 300

Glu Ser Asp Ile Ser Lys Val Tyr Leu Thr Glu Lys Gly Val Glu Met
 305 310 315 320

Phe Gly Lys Ile Leu Tyr Gln Arg Pro Ile Leu Lys Gly Asn Glu Lys
 325 330 335

Pro Lys Pro Asn Ala Val Asn Glu Phe Ile Ser Pro Pro Ile Gln Val
 340 345 350

Lys Tyr Tyr Phe Asn Lys Phe Gly Lys Asp Gly Val Ile Leu Ser Pro
 355 360 365

Ser Asp Ser Phe Glu Glu Met Arg Thr Leu Tyr Val Glu Gly Ala Glu
 370 375 380

Ala Tyr Asn Arg Glu Val Glu Met Pro Lys Lys Lys Arg Lys Val
 385 390 395

<210> SEQ ID NO 64
 <211> LENGTH: 399
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Synthetic: WYL Ruminococcus sp. isolate
 2789STDY5608892 + C-term NLS

<400> SEQUENCE: 64

Met Leu Ile Pro Pro Ser Thr Phe Leu Pro Lys Arg Asp Lys Asn Val
 1 5 10 15

Pro Tyr Ile Ala Glu Val Gln Ser Ile Pro Leu Ser Pro Ser Ala Tyr
 20 25 30

Ser Val Ile Ile Lys Asp Lys Ser Ile Phe Glu Thr Ser Leu Ser Pro
 35 40 45

Asn Gly Ser Val Ser Met Ser Ser Phe Leu Thr Ser Ile Phe Asp Ser
 50 55 60

Ala Tyr Ile Ala Ser Leu Lys Tyr Lys Ser Asp Asp Asn Tyr Lys Tyr

-continued

65	70	75	80
Ile Gly Ile Pro Leu Leu Asn Ala Phe Val Glu Trp Gln Ile Glu Glu			
85	90	95	
Ile Asp Asp Ser Leu Asp Asp Lys Ser Lys Glu Ile Ile Lys Ser Tyr			
100	105	110	
Leu Ile Ser Lys Leu Ser Ala Lys Tyr Glu Lys Thr Lys Thr Glu Asn			
115	120	125	
Ala Val Arg Val Arg Leu Ser Ile Cys Arg Asp Leu Tyr Asp Thr Leu			
130	135	140	
Ser Ser Asp Asp Leu Tyr Tyr Glu Asn Lys Val Tyr Ser Leu Thr Leu			
145	150	155	160
Arg Arg Phe Leu Lys Ala Val Tyr Glu Asp Tyr Ala Leu Leu Ser Asp			
165	170	175	
Cys Glu Arg Glu Arg Leu Ile Phe Ala Asp Asn Ile Ile Lys Ile Asn			
180	185	190	
Glu Val Ile Lys Gln Asn Gly Ser Arg Tyr Tyr Ser Phe Ile Tyr Ala			
195	200	205	
Tyr Ser Asn Met Tyr Ser Arg Glu Lys Arg Arg Ile Arg Leu Ile Pro			
210	215	220	
Tyr Arg Ile Val Ser Asp Glu Tyr Lys Met Tyr Asn Tyr Leu Val Cys			
225	230	235	240
Leu Ser Asp Glu Lys Ser Ala Gly Lys Glu Phe Lys Ala Asp Ser Tyr			
245	250	255	
Arg Ile Ser Arg Leu Ser Gly Leu Ser Ile Ala Glu Lys Leu Ser Gln			
260	265	270	
Lys Glu Tyr Ser Ser Val Thr Glu Tyr Glu Arg Leu Lys Glu Gly His			
275	280	285	
Val Lys Ser Val Lys His Leu Leu Ser Asp Pro Arg Phe Gly Ser Asp			
290	295	300	
Glu Ser Asp Ile Ser Lys Val Tyr Leu Thr Glu Lys Gly Val Glu Met			
305	310	315	320
Phe Gly Lys Ile Leu Tyr Gln Arg Pro Ile Leu Lys Gly Asn Glu Lys			
325	330	335	
Pro Lys Pro Asn Thr Val Asn Glu Phe Ile Ser Pro Pro Ile Gln Val			
340	345	350	
Lys Tyr Tyr Phe Asn Lys Phe Gly Lys Asp Gly Val Ile Leu Ser Pro			
355	360	365	
Ser Asp Ser Phe Glu Glu Met Arg Thr Leu Tyr Val Glu Gly Ala Glu			
370	375	380	
Ala Tyr Asn Arg Glu Val Glu Met Pro Lys Lys Arg Lys Val			
385	390	395	

```

<210> SEQ_ID NO 65
<211> LENGTH: 399
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: WYL Ruminococcus sp.
          CAG:57 + C-term NLS

```

<400> SEQUENCE: 65

Met Leu Ile Pro Pro Ser Thr Phe Leu Pro Lys Arg Asp Lys Asn Val			
1	5	10	15

-continued

Pro	Tyr	Ile	Ala	Glu	Val	Gln	Ser	Ile	Pro	Leu	Ser	Pro	Ser	Ala	Tyr
20															30
Ser	Val	Ile	Ile	Lys	Asp	Lys	Ser	Ile	Phe	Glu	Thr	Ser	Leu	Ser	Pro
35															45
Asn	Gly	Ser	Val	Ser	Met	Ser	Ser	Phe	Leu	Thr	Ser	Ile	Phe	Asp	Ser
50															60
Ala	Tyr	Ile	Ala	Ser	Leu	Lys	Tyr	Lys	Ser	Asp	Asp	Asn	Tyr	Lys	Tyr
65															80
Ile	Gly	Ile	Pro	Leu	Leu	Asn	Ala	Phe	Val	Glu	Trp	Gln	Ile	Glu	Glu
85															95
Ile	Asp	Asp	Ser	Leu	Asp	Asp	Lys	Ser	Lys	Glu	Ile	Ile	Lys	Ser	Tyr
100															110
Leu	Ile	Ser	Lys	Leu	Ser	Ala	Lys	Tyr	Glu	Lys	Thr	Lys	Thr	Glu	Asn
115															125
Ala	Val	Arg	Val	Arg	Leu	Ser	Ile	Cys	Arg	Asp	Leu	Tyr	Asp	Thr	Leu
130															140
Ser	Ser	Asp	Asp	Leu	Tyr	Tyr	Glu	Asn	Lys	Val	Tyr	Ser	Leu	Thr	Leu
145															160
Arg	Arg	Phe	Leu	Lys	Ala	Val	Tyr	Glu	Asp	Tyr	Ala	Leu	Ser	Asp	
165															175
Cys	Glu	Arg	Glu	Arg	Leu	Ile	Phe	Ala	Asp	Asn	Ile	Ile	Lys	Ile	Asn
180															190
Glu	Val	Ile	Lys	Gln	Asn	Gly	Ser	Arg	Tyr	Tyr	Ser	Phe	Ile	Tyr	Ala
195															205
Tyr	Ser	Asn	Met	Tyr	Ser	Arg	Glu	Lys	Arg	Arg	Ile	Arg	Leu	Ile	Pro
210															220
Tyr	Arg	Ile	Val	Ser	Asp	Glu	Tyr	Lys	Met	Tyr	Asn	Tyr	Leu	Val	Cys
225															240
Leu	Ser	Asp	Glu	Lys	Ser	Ala	Gly	Lys	Glu	Phe	Lys	Ala	Asp	Ser	Tyr
245															255
Arg	Ile	Ser	Arg	Leu	Ser	Gly	Leu	Ser	Ile	Ala	Glu	Lys	Leu	Ser	Gln
260															270
Lys	Glu	Tyr	Ser	Ser	Val	Thr	Glu	Tyr	Glu	Arg	Leu	Lys	Glu	Gly	His
275															285
Val	Lys	Ser	Val	Lys	His	Leu	Leu	Ser	Asp	Pro	Arg	Phe	Gly	Ser	Asp
290															300
Glu	Ser	Asp	Ile	Ser	Lys	Val	Tyr	Leu	Thr	Glu	Lys	Gly	Val	Glu	Met
305															320
Phe	Gly	Lys	Ile	Leu	Tyr	Gln	Arg	Pro	Ile	Leu	Lys	Gly	Asn	Glu	Lys
325															335
Pro	Lys	Pro	Asn	Thr	Val	Asn	Glu	Phe	Ile	Ser	Pro	Pro	Ile	Gln	Val
340															350
Lys	Tyr	Tyr	Phe	Asn	Lys	Phe	Gly	Lys	Asp	Gly	Val	Ile	Leu	Ser	Pro
355															365
Ser	Asp	Ser	Phe	Glu	Glu	Met	Arg	Thr	Leu	Tyr	Val	Glu	Gly	Ala	Glu
370															380
Ala	Tyr	Asn	Arg	Glu	Val	Glu	Met	Pro	Lys	Lys	Arg	Lys	Val		
385															395

<210> SEQ_ID NO 66

<211> LENGTH: 287

<212> TYPE: PRT

-continued

```

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: WYL Ruminococcus flavefaciens FD-1 +
C-term NLS

<400> SEQUENCE: 66

Met Ile Ile Ala Ile Asn Gln Trp Lys Arg Arg Phe Ser Leu Val Ile
1 5 10 15

Tyr Gly Lys Ser Glu Gly Glu Thr Ile Val Lys Ile Lys Leu Leu Leu
20 25 30

Ile Ser Leu Ala Tyr Leu Ile Ser Ile Tyr Leu Leu Cys Ser Pro Gly
35 40 45

Cys Ile Gly Ile Phe Thr His Gly Met Leu Thr Thr Val Ile Gly Val
50 55 60

Val Thr Met Leu Ala Ala Thr Gly Thr Tyr Gly Met Tyr Leu Tyr Ser
65 70 75 80

Ser Ala Ile Gly Glu Arg Ser Leu Pro Glu Ile Pro Met Asn Lys Glu
85 90 95

Thr Glu Tyr Ser Arg Tyr Lys Glu Leu Glu Asn Trp Phe Arg Ala Phe
100 105 110

Arg Tyr Leu Asp Arg Asn Asn Phe Ala Met Leu Ser Ser Asp Leu
115 120 125

Ala Thr Ser Tyr His Asp Gly Leu Ile Arg Asp Asn Pro Phe Arg Asn
130 135 140

Thr Glu Leu Gly Asp Arg Leu Gln Thr Thr Ser Ser Asp Ile Ser Ile
145 150 155 160

Lys Tyr Asp Gln Thr Leu Lys Ile Leu Ser Glu Ser Phe Glu Lys Asn
165 170 175

Asp Ile Thr Tyr Gln Asn Tyr Leu Ser Val Leu Asp Asn Val Leu Lys
180 185 190

Leu Ser Ser Ser His Leu Lys Ala Ile Lys Lys Arg Val Cys Val Phe
195 200 205

Asp Tyr Arg Thr Trp Ala Asp Asn Lys Asn Asp Glu Met Cys Arg Lys
210 215 220

Tyr Ile Glu Glu Val Lys Ser Ser Val Ile Arg Leu Glu Glu Ile Glu
225 230 235 240

Gly Lys Phe Asp Asn Leu Leu His Glu Leu Ile Cys Leu Ser Glu Ile
245 250 255

Ser Glu Asp Pro Leu Leu Glu Met Gln Asp Leu Ile Glu Thr Thr Ser
260 265 270

Asp Tyr Lys Ser Ile Glu Asp Gln Pro Lys Lys Lys Arg Lys Val
275 280 285

```

```

<210> SEQ ID NO 67
<211> LENGTH: 233
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: WYL Ruminococcus albus strain
KH2T6 + C-term NLS

```

<400> SEQUENCE: 67

```

Met Cys Thr Trp Tyr Tyr Ala Glu Ala Lys Ser Leu Ser Phe Phe Ile
1 5 10 15

Asp Lys Ala Ser Gln Leu Pro Leu Ser Asp Ile Ile Met Asn Thr Met

```

-continued

20	25	30	
Ser Lys Ser Lys Ala Met Ser Gly Asn Ile Arg Pro Thr Asp Met Ala			
35	40	45	
Ala Val Leu Ala Pro Asn Lys Gln Gly Asn Val Ala Val Phe Pro Met			
50	55	60	
Ile Trp Gly Phe Thr His Glu Ser Thr Ser Lys Pro Val Ile Asn Cys			
65	70	75	80
Arg Ile Glu Ser Ala Asp Thr Lys Pro Leu Trp Lys Asp Ser Trp Tyr			
85	90	95	
Arg Arg Arg Cys Val Ile Pro Ala Ser Trp Tyr Tyr Glu Trp Gly Val			
100	105	110	
Pro Pro Ser Glu Gly Glu Leu Tyr His Lys Asn Glu Tyr Asn Lys Ile			
115	120	125	
Gln Lys Glu Lys Tyr Ala Ile Gln Pro Glu Gly Ala Glu Ile Thr Tyr			
130	135	140	
Leu Ala Gly Leu Tyr Arg Phe Glu Glu His Arg Gly Val Gln Val Pro			
145	150	155	160
Met Phe Ala Val Ile Thr Arg Glu Ser Val Glu Pro Val Ser Ser Ile			
165	170	175	
His Asp Arg Met Pro Leu Ile Leu Gly Lys Asp Ser Leu Ser Glu Trp			
180	185	190	
Ile His Pro Asn Gly Asp Pro Asn Lys Ile Ala Lys Thr Ala Leu Thr			
195	200	205	
Lys Met Val Met Glu Lys Ala Ile Asp Tyr Pro Glu Pro Glu Pro Ser			
210	215	220	
Phe Met Pro Lys Lys Arg Lys Val			
225	230		

```

<210> SEQ_ID NO 68
<211> LENGTH: 321
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER_INFORMATION: Synthetic: WYL Ruminococcus flavefaciens strain
 XPD3002 + C-term NLS

```

210	211	212	
<210> SEQ_ID NO 68			
<211> LENGTH: 321			
<212> TYPE: PRT			
<213> ORGANISM: Artificial Sequence			
<220> FEATURE:			
<223> OTHER_INFORMATION: Synthetic: WYL Ruminococcus flavefaciens strain			
XPD3002 + C-term NLS			
<400> SEQUENCE: 68			
Met Glu Leu Phe Asn Glu Tyr Arg Asn Lys Ser Leu Arg Ala Phe Leu			
1	5	10	15
Lys Leu Ala Glu Arg Ile Ser Tyr Gly Glu Glu Leu Ser Ile Asp Glu			
20	25	30	
Phe Glu Ala Glu Tyr Tyr Arg Leu Ser Gly Asp Asn Lys Ile Thr			
35	40	45	
Ser Val Phe Tyr Lys Asn Thr Leu Tyr Asn Asp Lys Leu Pro Ile Phe			
50	55	60	
Asp Thr Arg Glu Gly Lys Val Arg Leu Phe Gly Glu Pro Asp Lys Cys			
65	70	75	80
Ser Asn Lys His Ile Ser Asp Thr Leu Leu Lys Ser Glu Ile Thr Trp			
85	90	95	
Leu His Asn Ala Leu Asn Asp Lys Leu Ser Lys Leu Phe Leu Ser Asp			
100	105	110	
Glu Glu Arg Ile Ser Ile Asp Ala Lys Leu Ser Asp Tyr Thr Glu Tyr			
115	120	125	

-continued

Tyr	Lys	Asn	Ile	Asp	Asp	Met	Trp	Arg	Ser	Asn	Glu	Asp	Ile	Ser	Glu
130						135					140				
<hr/>															
Glu	Val	Glu	Lys	Asn	Phe	Lys	Ile	Ile	Leu	Lys	Ala	Ile	Asn	Glu	Lys
145						150				155					160
<hr/>															
Gln	Ala	Leu	Ser	Tyr	Thr	Phe	Lys	Asn	Lys	Asn	Cys	Glu	Gly	Phe	Pro
											165	170		175	
<hr/>															
Val	Arg	Ile	Glu	Tyr	Asp	Glu	Arg	Thr	Cys	Arg	Ile	Tyr	Met	Ile	Ile
											180	185		190	
<hr/>															
Tyr	Asp	Gly	Asn	Arg	Phe	Val	Lys	Ser	Asp	Ile	Ser	Lys	Leu	Ser	Asp
						195			200			205			
<hr/>															
Ile	Tyr	Ile	Thr	Glu	Asn	Ser	Ile	Asp	Thr	Ile	Pro	Glu	Ile	Lys	Asp
						210			215			220			
<hr/>															
Asp	Met	Leu	Asn	Lys	Lys	Ala	Tyr	Leu	Pro	Val	Val	Phe	Thr	Val	Thr
						225			230			235			240
<hr/>															
Asp	Asp	Lys	Asn	Arg	Lys	Ala	Ile	Asp	Arg	Ala	Leu	Leu	Ala	Phe	Ser
						245			250			255			
<hr/>															
Val	Tyr	Asp	His	Val	Val	Glu	Pro	Ile	Asp	Glu	Lys	Thr	Ala	Arg	Phe
						260			265			270			
<hr/>															
Thr	Ile	Gln	Tyr	Tyr	Thr	Met	Asp	Leu	Asp	Leu	Ile	Lys	Asp	Ile	
						275			280			285			
<hr/>															
Leu	Ala	Phe	Gly	Ser	Asp	Ile	Lys	Val	Glu	Ser	Pro	Arg	Tyr	Val	Val
						290			295			300			
<hr/>															
Lys	Arg	Ile	Thr	Asp	Ile	Leu	Arg	Lys	Val	Pro	Lys	Lys	Lys	Arg	Lys
						305			310			315			320

Val

```
<210> SEQ ID NO 69
<211> LENGTH: 419
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: RtcB Eubacterium siraeum +
C-term NLS

<400> SEQUENCE: 69

Met Ile Val Leu Glu Ile Ile Gly Glu Arg Asn Thr Ala Val Val Tyr
1 5 10 15

Gly Glu Ile Ile Asp Glu Cys Ala Val Ser Gln Ile Glu Glu Ile Cys
20 25 30

Asn His Pro Ala Phe Glu Asn Ser Arg Ile Arg Ile Met Pro Asp Cys
35 40 45

His Ala Gly Lys Gly Cys Val Ile Gly Phe Thr Cys Val Thr Ser Asn
50 55 60

Arg Met Ile Val Pro Asn Ile Val Gly Val Asp Ile Gly Cys Gly Ile
65 70 75 80

Leu Thr Thr Val Phe Thr Ala Asp Arg Glu Ile Asp Tyr Arg Ala Leu
85 90 95

Asp Thr Phe Ile Arg Ser Asn Ile Pro Ser Gly Met Glu Ile His Asp
100 105 110

Ser Val Ser Asp Thr Val Ala Glu Asn Thr Ala Leu Ile Ala Lys Val
115 120 125

Asn Gly Ile Cys Asp Ala Ile Gly Glu Ser Ala Asp Val Asp Tyr His
130 135 140
```

-continued

Leu	Arg	Ser	Ile	Gly	Thr	Leu	Gly	Gly	Asn	His	Phe	Ile	Glu	Ile
145				150			155				160			
Asp Arg Leu Asn Asn Gly Asn Tyr Ala Leu Thr Val His Thr Gly Ser														
	165				170			175						
Arg Asn Leu Gly Lys Arg Ile Cys Gly Tyr Phe Gln Ser Asn Ala Ser														
	180				185			190						
Val Ile Asp Thr Glu Leu Arg Arg Ser Ile Leu Leu Arg His Arg Ser														
	195				200			205						
Ala Thr Thr Ser Glu Glu His Glu Glu Ile Asp Arg Arg Ala Ala Gln														
	210				215			220						
Ile Ala Pro Val Ser Lys Glu Leu Ala Phe Ile Thr Gly Glu Arg Tyr														
	225				230			235			240			
Asp Ser Tyr Ile Gly Cys Met Leu Asp Ala Lys Ala Leu Ala Ala Phe														
	245				250			255						
Asn Arg Thr Val Ile Ser Asp Arg Ile Met Ser Phe Leu Ala Asp Glu														
	260				265			270						
Tyr Gly Val Glu Ile Lys Asp Arg Phe Asp Thr Val His Asn Tyr Ile														
	275				280			285						
Asp Trp Tyr Asp Asp Thr His Thr Ser Val Val Ile Arg Lys Gly Ala														
	290				295			300						
Ile Ser Ala Arg Lys Gly Glu Arg Ile Val Ile Pro Leu Asn Met Arg														
	305				310			315			320			
Asp Gly Ile Ile Ala His Gly Arg Gly Asn Glu Glu Trp Asn Cys														
	325				330			335						
Ser Ala Pro His Gly Ser Gly Arg Ala Tyr Ser Arg Ser Asp Ala Arg														
	340				345			350						
Arg Thr Phe Thr Leu Glu Glu Tyr Val Glu Glu Met Asp Gly Val Asn														
	355				360			365						
Thr Trp Ser Val Ser Glu Ser Thr Ile Asp Glu Cys Pro Met Ala Tyr														
	370				375			380						
Lys Pro Ser Glu Met Ile Ile Gly Ser Ile Gly Asp Thr Val Glu Ile														
	385				390			395			400			
Glu Ser Ile Ala His Thr Val Tyr Asn Phe Lys Ala Pro Lys Lys Lys														
	405				410			415						
Arg Lys Val														

```

<210> SEQ ID NO 70
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 1

<400> SEQUENCE: 70
cuacauauacu ggugcgaauu ugcacauaguc uaaaaaucuua caucuuuccu ccucauccag      60
caaaauu                                         66

```

```

<210> SEQ ID NO 71
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown
experiment target 2

```

-continued

<400> SEQUENCE: 71
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaaucaca auccugaagu aagugaagcu 60
acagac 66

<210> SEQ ID NO 72
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 3

<400> SEQUENCE: 72
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaucugu caaaaaucac aauccugaag 60
uaagug 66

<210> SEQ ID NO 73
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 4

<400> SEQUENCE: 73
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaucucu guaaaaaac acaauccuga 60
aguaag 66

<210> SEQ ID NO 74
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 5

<400> SEQUENCE: 74
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauucuc uucacgagau ucacuaggac 60
cuucag 66

<210> SEQ ID NO 75
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 6

<400> SEQUENCE: 75
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaaucucc ucucuuacg agauucacua 60
ggaccu 66

<210> SEQ ID NO 76
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 7

-continued

<400> SEQUENCE: 76
cuacuaucu ggugcgaaau ugcacuaguc uaaaaaugucc ucuaggucca uguuacagcc 60
agaccc 66

<210> SEQ ID NO 77
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 8

<400> SEQUENCE: 77
cuacuaucu ggugcgaaau ugcacuaguc uaaaaauagu ccucuagguc cauguuacag 60
ccagac 66

<210> SEQ ID NO 78
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 9

<400> SEQUENCE: 78
cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucgcc aggagcgug ccccgccgu 60
cccgga 66

<210> SEQ ID NO 79
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 10

<400> SEQUENCE: 79
cuacuaucu ggugcgaaau ugcacuaguc uaaaaugcgc caggagcgcu gccccggccg 60
ucccg 66

<210> SEQ ID NO 80
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 11

<400> SEQUENCE: 80
cuacuaucu ggugcgaaau ugcacuaguc uaaaaugcag cgccaggagc gcugccccgg 60
ccgucc 66

<210> SEQ ID NO 81
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 12

<400> SEQUENCE: 81

-continued

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauagca ggcgcaggag cgccugcccg 60

gccccuc 66

<210> SEQ ID NO 82
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 13

<400> SEQUENCE: 82

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaucagc agcgccaggaa ggcgcugccccc 60

ggccgu 66

<210> SEQ ID NO 83
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 14

<400> SEQUENCE: 83

cuacuaauacu ggugcgaaau ugcacuaguc uaaaauccag cagcgccagg agcgccugccc 60

cggccg 66

<210> SEQ ID NO 84
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 15

<400> SEQUENCE: 84

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaugcca gcagcgccag gagcgccugcc 60

cgggcc 66

<210> SEQ ID NO 85
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 16

<400> SEQUENCE: 85

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaucagc cagcgccgc aggagcgug 60

ccccgg 66

<210> SEQ ID NO 86
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 17

<400> SEQUENCE: 86

-continued

cuacuaucu ggugcgaaau ugcacuaguc uaaaaugcag ccagcagcgc caggagcgcu	60
gccccg	66
<210> SEQ ID NO 87	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 18	
<400> SEQUENCE: 87	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaucgca gccagcagcg ccaggagcgc	60
ugcccc	66
<210> SEQ ID NO 88	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 19	
<400> SEQUENCE: 88	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaugcgc agccagcagc gccaggagcg	60
cugccc	66
<210> SEQ ID NO 89	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 20	
<400> SEQUENCE: 89	
cuacuaucu ggugcgaaau ugcacuaguc uaaaauagcg cagccagcag cgccaggagc	60
gcugcc	66
<210> SEQ ID NO 90	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 21	
<400> SEQUENCE: 90	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaugagc gcagccagca gcccaggag	60
cgcugc	66
<210> SEQ ID NO 91	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 22	
<400> SEQUENCE: 91	
cuacuaucu ggugcgaaau ugcacuaguc uaaaauagag cgccagccagc agccaggaga	60

-continued

gcgcug	66
 <pre><210> SEQ ID NO 92 <211> LENGTH: 66 <212> TYPE: RNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 23</pre>	
<400> SEQUENCE: 92	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaucaga ggcggccagg	60
agcgcu	66
 <pre><210> SEQ ID NO 93 <211> LENGTH: 66 <212> TYPE: RNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 24</pre>	
<400> SEQUENCE: 93	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaugcag agcgccgcgc	60
gagcgc	66
 <pre><210> SEQ ID NO 94 <211> LENGTH: 66 <212> TYPE: RNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 25</pre>	
<400> SEQUENCE: 94	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaugggc agagcgccgc cagcagcgcc	60
aggagc	66
 <pre><210> SEQ ID NO 95 <211> LENGTH: 66 <212> TYPE: RNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 26</pre>	
<400> SEQUENCE: 95	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaucggc cagagcgccag ccagcagcgc	60
caggag	66
 <pre><210> SEQ ID NO 96 <211> LENGTH: 66 <212> TYPE: RNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 27</pre>	
<400> SEQUENCE: 96	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaugccg ggcagagcgc agccagcagc	60

-continued

gccagg	66
<210> SEQ ID NO 97	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 28	
<400> SEQUENCE: 97	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucgc gggcagagcg cagccagcag	60
cggccag	66
<210> SEQ ID NO 98	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 29	
<400> SEQUENCE: 98	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucgc cgggcagagc gcagccagca	60
gccccca	66
<210> SEQ ID NO 99	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 30	
<400> SEQUENCE: 99	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucgc cggggcagag cgccagccagc	60
agcgcc	66
<210> SEQ ID NO 100	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 31	
<400> SEQUENCE: 100	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucgc gccgggcaga ggcgcagccag	60
cagcgc	66
<210> SEQ ID NO 101	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 32	
<400> SEQUENCE: 101	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaugacu cgccgggcag agcgcagcca	60
gcagcg	66

-continued

<210> SEQ ID NO 102
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 33

<400> SEQUENCE: 102

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauaaa gugcccaacu gcgugagcuu	60
guuacu	66

<210> SEQ ID NO 103
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 34

<400> SEQUENCE: 103

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauaucu ucaaaagugc ccaacugcgu	60
gagcuu	66

<210> SEQ ID NO 104
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 35

<400> SEQUENCE: 104

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauuga ucuucaaag ugcccacug	60
cgugag	66

<210> SEQ ID NO 105
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 36

<400> SEQUENCE: 105

cuacuaucu ggugcgaaau ugcacuaguc uaaaauccuc uggaggcuga gaaaaugauc	60
uucaaa	66

<210> SEQ ID NO 106
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 37

<400> SEQUENCE: 106

cuacuaucu ggugcgaaau ugcacuaguc uaaaauacau ccucuggagg cugagaaaaau	60
gaucuu	66

-continued

<210> SEQ ID NO 107
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 38

<400> SEQUENCE: 107

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauaguu auugaacauc cucuggaggc	60
ugagaa	66

<210> SEQ ID NO 108
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 39

<400> SEQUENCE: 108

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauacag uuauugaaca uccucuggag	60
gcugag	66

<210> SEQ ID NO 109
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 40

<400> SEQUENCE: 109

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaucaca guuauugaac auccucugga	60
ggcuga	66

<210> SEQ ID NO 110
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 41

<400> SEQUENCE: 110

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauaccu cacaguauu gaacaucuc	60
uggagg	66

<210> SEQ ID NO 111
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 42

<400> SEQUENCE: 111

cuacuaauacu ggugcgaaau ugcacuaguc uaaaauagga ccaccucaca guuauugaac	60
auccuc	66

-continued

<210> SEQ ID NO 112
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 43

<400> SEQUENCE: 112

cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucaag gaccaccuca caguuaug	60
acaucc	66

<210> SEQ ID NO 113
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 44

<400> SEQUENCE: 113

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauuucc caaggaccac cucacaguua	60
uugaac	66

<210> SEQ ID NO 114
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 45

<400> SEQUENCE: 114

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauaaau ucctaaggac caccucacag	60
uuauuug	66

<210> SEQ ID NO 115
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 46

<400> SEQUENCE: 115

cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucaa uucctaagg ccaccucaca	60
guuauu	66

<210> SEQ ID NO 116
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 47

<400> SEQUENCE: 116

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauccaa auucctaagg accaccucac	60
aguuaau	66

<210> SEQ ID NO 117

-continued

<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 48

<400> SEQUENCE: 117

cuacuaucuacu ggugcgaaau ugcacuaguc uaaaauuuuuc caauuccca aggaccacu 60
cacagu 66

<210> SEQ ID NO 118
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 49

<400> SEQUENCE: 118

cuacuaucuacu ggugcgaaau ugcacuaguc uaaaauuuaau uuccaaaauuc ccaaggacca 60
ccucac 66

<210> SEQ ID NO 119
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 50

<400> SEQUENCE: 119

cuacuaucuacu ggugcgaaau ugcacuaguc uaaaaauguaa uuuccaaauu cccaaggacc 60
accuca 66

<210> SEQ ID NO 120
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 51

<400> SEQUENCE: 120

cuacuaucuacu ggugcgaaau ugcacuaguc uaaaaauaggu aauuuuccaaa uucccaagga 60
ccaccu 66

<210> SEQ ID NO 121
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 52

<400> SEQUENCE: 121

cuacuaucuacu ggugcgaaau ugcacuaguc uaaaaucuaa gguauuuucc aaauucccaa 60
ggacca 66

<210> SEQ ID NO 122
<211> LENGTH: 66

-continued

<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 53

<400> SEQUENCE: 122

cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucugc acauagguaa uuucccaaau	60
cccaag	66

<210> SEQ ID NO 123
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 54

<400> SEQUENCE: 123

cuacuaucu ggugcgaaau ugcacuaguc uaaaaucucu gcacauaggu aauuuuccaaa	60
uuuccca	66

<210> SEQ ID NO 124
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 55

<400> SEQUENCE: 124

cuacuaucu ggugcgaaau ugcacuaguc uaaaauaauu ccucugcaca uagguauuu	60
ccaaau	66

<210> SEQ ID NO 125
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 56

<400> SEQUENCE: 125

cuacuaucu ggugcgaaau ugcacuaguc uaaaauagau cauaauuccu cugcacauag	60
guaauu	66

<210> SEQ ID NO 126
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 57

<400> SEQUENCE: 126

cuacuaucu ggugcgaaau ugcacuaguc uaaaauauga ggacauaacc agccaccucc	60
uggaug	66

<210> SEQ ID NO 127
<211> LENGTH: 66
<212> TYPE: RNA

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 58

<400> SEQUENCE: 127

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaugcaa ugaggacaua accagccacc	60
uccugg	66

<210> SEQ ID NO 128
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 59

<400> SEQUENCE: 128

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauauu cgcuccacug uguugaggc	60
aauagag	66

<210> SEQ ID NO 129
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 60

<400> SEQUENCE: 129

cuacuaauacu ggugcgaaau ugcacuaguc uaaaauguuu uccaaaggaa uucgcuccac	60
uguguu	66

<210> SEQ ID NO 130
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 61

<400> SEQUENCE: 130

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauucug cagguuuucc aaaggaauuc	60
gcucca	66

<210> SEQ ID NO 131
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 62

<400> SEQUENCE: 131

cuacuaauacu ggugcgaaau ugcacuaguc uaaaaugauc ugcagguuuu ccaaaggaaau	60
ucgcuc	66

<210> SEQ ID NO 132
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 63

<400> SEQUENCE: 132

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaaugaug	aucugcagg	uuuccaaagg	60
aauucg						66

<210> SEQ ID NO 133
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 64

<400> SEQUENCE: 133

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaauugau	gaucugcagg	uuuuccaaag	60
gaauuc						66

<210> SEQ ID NO 134
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 65

<400> SEQUENCE: 134

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaauucug	augaucugca	gguuuuccaa	60
aggaaau						66

<210> SEQ ID NO 135
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 66

<400> SEQUENCE: 135

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaauauuu	ccucugauga	ucugcagg	60
uuccaa						66

<210> SEQ ID NO 136
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 67

<400> SEQUENCE: 136

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaauauau	uuccucugau	gaucugcagg	60
uuuucc						66

<210> SEQ ID NO 137
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 68

<400> SEQUENCE: 137

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaauuuuc	guaguacaua	uuuccucuga	60
ugaucu						66

<210> SEQ ID NO 138

<211> LENGTH: 66

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 69

<400> SEQUENCE: 138

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaauuagg	aauuuucgua	guacauauu	60
ccucug						66

<210> SEQ ID NO 139

<211> LENGTH: 66

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 70

<400> SEQUENCE: 139

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaaucua	ggaauuuucg	uaguacauau	60
uuccuc						66

<210> SEQ ID NO 140

<211> LENGTH: 66

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 71

<400> SEQUENCE: 140

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaaucugc	uaaggcauag	gaaauuuucgu	60
aguaca						66

<210> SEQ ID NO 141

<211> LENGTH: 66

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 72

<400> SEQUENCE: 141

cuacuaucu	ggugcgaaau	ugcacuaguc	aaaaaugaua	agacugcuaa	ggcauaggaa	60
uuuucg						66

<210> SEQ ID NO 142

<211> LENGTH: 66

<212> TYPE: RNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment

-continued

target 73

<400> SEQUENCE: 142

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauauag uuagauaaga cugcuaaggc 60
auagga 66

<210> SEQ ID NO 143
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 74

<400> SEQUENCE: 143

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauauca uaguuagaua agacugcuaa 60
ggcaua 66

<210> SEQ ID NO 144
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 75

<400> SEQUENCE: 144

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauuu ugcaucauag uuagauaaga 60
cugcua 66

<210> SEQ ID NO 145
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 76

<400> SEQUENCE: 145

cuacuaucu ggugcgaaau ugcacuaguc uaaaaauaguc cgguuuuauu ugcaucauag 60
uuagau 66

<210> SEQ ID NO 146
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 77

<400> SEQUENCE: 146

cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucuuc aguccgguuu uauuugcauc 60
auaguu 66

<210> SEQ ID NO 147
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment
target 78

-continued

<400> SEQUENCE: 147
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauuggg cagcuccuuuc aguccgguuu 60
uuauuug 66

<210> SEQ ID NO 148
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 79

<400> SEQUENCE: 148
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaaucaug ggcagcuccu ucaguccgg 60
uuuauu 66

<210> SEQ ID NO 149
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 80

<400> SEQUENCE: 149
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauuaaa uuucucaugg gcagcuccuu 60
cagucc 66

<210> SEQ ID NO 150
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 81

<400> SEQUENCE: 150
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaauuagc ccccagcgcc acgaccuccg 60
agcuac 66

<210> SEQ ID NO 151
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 82

<400> SEQUENCE: 151
cuacuaauacu ggugcgaaau ugcacuaguc uaaaaaugccu cccgacagag cgugugcu 60
agcccc 66

<210> SEQ ID NO 152
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 83

-continued

<400> SEQUENCE: 152
cuacuaaucu ggugcgaaau ugcacuaguc uaaaauuucc agcacccgagc gcccuggccg 60
gugagu 66

<210> SEQ ID NO 153
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 84

<400> SEQUENCE: 153
cuacuaaucu ggugcgaaau ugcacuaguc uaaaauagaa aaaagaagag ggauaaaacc 60
cggauc 66

<210> SEQ ID NO 154
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 85

<400> SEQUENCE: 154
cuacuaaucu ggugcgaaau ugcacuaguc uaaaaugggg aguagagcaa ucuccccaag 60
ccgucg 66

<210> SEQ ID NO 155
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 86

<400> SEQUENCE: 155
cuacuaaucu ggugcgaaau ugcacuaguc uaaaaugggg aggagguggu agcuggggcu 60
gggggc 66

<210> SEQ ID NO 156
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 87

<400> SEQUENCE: 156
cuacuaaucu ggugcgaaau ugcacuaguc uaaaaucacc ccgccuccgg gcgcgggcuc 60
cggccc 66

<210> SEQ ID NO 157
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 88

<400> SEQUENCE: 157

-continued

cuacuaauacu ggugcgaauu ugcacuaguc uaaaaaucacg gcuuccuccga agcgagaaca 60
gccccag 66

<210> SEQ ID NO 158
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 89

<400> SEQUENCE: 158

cuacuaauacu ggugcgaauu ugcacuaguc uaaaaaucccg ggacggccgg ggcagcgcuc 60
cuggcgc 66

<210> SEQ ID NO 159
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 90

<400> SEQUENCE: 159

cuacuaauacu ggugcgaauu ugcacuaguc uaaaaauccgg gacggccggg gcagcgcucc 60
uggcgc 66

<210> SEQ ID NO 160
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 91

<400> SEQUENCE: 160

cuacuaauacu ggugcgaauu ugcacuaguc uaaaauggac ggccggggca ggcuccugg 60
cgcugc 66

<210> SEQ ID NO 161
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 92

<400> SEQUENCE: 161

cuacuaauacu ggugcgaauu ugcacuaguc uaaaaugacg gcccggggcag cgcuccuggc 60
gcugcu 66

<210> SEQ ID NO 162
<211> LENGTH: 66
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 93

<400> SEQUENCE: 162

-continued

cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucgg cggggcagc gcuccuggcg	60
cugcug	66
<210> SEQ ID NO 163	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 94	
<400> SEQUENCE: 163	
cuacuaucu ggugcgaaau ugcacuaguc uaaaaaucggc cggggcagcg cuccuggcg	60
ugcugg	66
<210> SEQ ID NO 164	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 95	
<400> SEQUENCE: 164	
cuacuaucu ggugcgaaau ugcacuaguc uaaaauggcc ggggcagcgc uccuggcg	60
gcuggc	66
<210> SEQ ID NO 165	
<211> LENGTH: 66	
<212> TYPE: RNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic: CasM EGFR mRNA knockdown experiment target 96	
<400> SEQUENCE: 165	
cuacuaucu ggugcgaaau ugcacuaguc uaaaauccgg ggcagcgcuc cuggcgcug	60
uggcug	66

1. An isolated CasM protein capable of producing a single-strand break at an RNA target site when guided to the RNA target site by a cognate nucleic acid guide.

2. The CasM protein of claim 1, wherein the cognate nucleic acid guide is crRNA.

3. The CasM protein of claim 1, wherein the protein comprises an amino acid sequence selected from the group consisting of the amino acid sequence of SEQ ID NOS:37-44 or 45; an ortholog of the amino acid sequence of SEQ ID NOS:37-44 and 45; and a variant of the amino acid sequence of SEQ ID NOS:37-44 or 45.

4. The CasM protein of claim 3, wherein the protein comprises the amino acid sequence of SEQ ID NOS:37-44 or 45.

5. A complex comprising:

a CasM protein according to claim 1; and
a cognate nucleic acid guide.

6. The complex of claim 5, wherein the cognate nucleic acid guide comprises RNA.

7. The complex of claim 6, wherein the cognate nucleic acid guide comprises a crRNA.

8. The complex of claim 5, wherein the cognate nucleic acid guide comprises a repeat sequence and a spacer sequence, wherein the repeat sequence and the spacer sequence do not naturally occur together.

9. The complex of claim 5, wherein the cognate nucleic acid guide comprises a modified base analog.

10. An isolated polynucleotide encoding the CasM protein of claim 1.

11. The polynucleotide of claim 10, wherein the polynucleotide comprises the sequence of SEQ ID NOS:28-35 or 36.

12. A modified polynucleotide encoding a CasM protein capable of producing a single strand break at a RNA target site when guided to the RNA target site by a cognate nucleic acid guide, wherein the polynucleotide is modified relative to its native sequence.

13. The modified polynucleotide of claim 12, wherein the polynucleotide is modified for expression in a selected host cell.

- 14.** The modified polynucleotide of claim **13**, wherein the polynucleotide is modified for expression in a bacterial cell.
- 15.** The modified polynucleotide of claim **14**, wherein the bacterial cell is an *Escherichia coli* cell.
- 16.** The modified polynucleotide of claim **15**, wherein the polynucleotide comprises the sequence of SEQ ID NOS:1-8 or 9.
- 17.** The modified polynucleotide of claim **13**, wherein the polynucleotide is modified for expression in a eukaryotic cell.
- 18.** The modified polynucleotide of claim **17**, wherein the polynucleotide is modified for expression in a mammalian cell.
- 19.** The modified polynucleotide of claim **18**, wherein the polynucleotide is modified for expression in a human cell.
- 20.** The modified polynucleotide of claim **19**, wherein the polynucleotide comprises the sequence of SEQ ID NOS:10-17 or 18.
- 21.** The modified polynucleotide of claim **18**, wherein the polynucleotide comprises a sequence encoding a N- or C-terminal nuclear localization signal sequence (NLS).
- 22.** The modified polynucleotide of claim **17**, wherein the polynucleotide is modified for expression in a plant cell.
- 23.** The modified polynucleotide of claim **22**, wherein the plant cell is a *Zea mays* cell.
- 24.** The modified polynucleotide of claim **23**, wherein the polynucleotide comprises the sequence of SEQ ID NOS:20-26 or 27.
- 25.** A recombinant vector comprising:
a polynucleotide comprising a coding sequence for a CasM protein according to claim 1; and
at least one control element operably linked to said polynucleotide, whereby the CasM coding sequence in said polynucleotide is capable of being transcribed and translated in a host cell.
- 26.** The recombinant vector of claim **25**, wherein the polynucleotide encoding the CasM protein comprises the nucleic acid sequence of SEQ ID NOS:1-35 or 36.
- 27.** The recombinant vector of claim **25**, wherein at least one of the control elements is heterologous to the coding sequence.
- 28.** A host cell comprising the recombinant vector of claim **25**.
- 29.** The host cell of claim **28**, wherein the host cell is a eukaryotic cell.
- 30.** A method of producing the CasM protein of claim 1, the method comprising:
providing a population of host cells according to claim **28**; and
culturing said population of cells under conditions whereby the CasM protein encoded by the polynucleotide present in said recombinant vector is expressed.
- 31.** A eukaryotic host cell comprising the CasM protein of claim 1.
- 32.** A eukaryotic host cell comprising the complex of claim 5.
- 33.** A method of directing a CasM protein to a selected nucleic acid target sequence, the method comprising:
contacting the selected nucleic acid target sequence with a complex according to claim 5, whereby the CasM protein is delivered to the nucleic acid target sequence.
- 34.** The method of claim **33**, wherein the nucleic acid target sequence comprises RNA.
- 35.** The method of claim **34**, wherein the nucleic target sequence is in an mRNA sequence.
- 36.** The method of claim **34**, further comprising
producing one or more single-strand breaks in the RNA.
- 37.** The method of claim **33**, wherein the method is performed in a cell.
- 38.** The method of claim **37**, wherein the cell is a eukaryotic cell.
- 39.** The method of claim **37**, wherein the cell constitutively expresses the CasM protein.
- 40.** The method of claim **39**, wherein the cell constitutively expresses the CasM protein and the cognate nucleic acid guide.

* * * * *