新型 CRISPR/Cas12f 酶和系统

技术领域

本发明涉及核酸编辑领域，特别是规律成簇的间隔短回文重复(CRISPR)技术领域。具体而言，本发明涉及 Cas 效应蛋白，包含此类蛋白的融合蛋白，以及编码它们的核酸分子。本发明还涉及用于核酸编辑（例如，基因或基因组编辑）的复合物和组合物，其包含本发明的蛋白或融合蛋白，或编码它们的核酸分子。本发明还涉及用于核酸编辑（例如，基因或基因组编辑）的方法，其使用包含本发明的蛋白或融合蛋白。

背景技术

CRISPR/Cas 技术是一种被广泛使用的基因编辑技术，它通过 **RNA** 引导对基因组上的靶序列进行特异性结合并切割 **DNA** 产生双链断裂，利用生物非同源末端连接或同源重组进行定点基因编辑。

CRISPR/Cas9 系统是最常用的 **II** 型 **CRISPR** 系统，它识别 **3'-NGG** 的 **PAM** 基序，对靶标序列进行平末端切割。**CRISPR/Cas Type V** 系统是一类近两年新发现的 **CRISPR** 系统，它具有 **5'-TTN** 的基序，对靶标序列进行粘性末端切割，例如 **Cpf1, C2c1, CasX, CasY**。然而目前存在的不同的 **CRISPR/Cas** 各有不同的优点和缺陷。例如 **Cas9, C2c1** 和 **CasX** 均需要两条 **RNA** 进行导向 **RNA**，而 **Cpf1** 只需要一条导向 **RNA** 而且可以用来进行多重基因编辑。**CasX** 具有 **980** 个氨基酸的大小，而常见的 **Cas9**，**C2c1,CasY** 和 **Cpf1** 通常大小在 **1300** 个氨基酸左右。此外，**Cas9**，**Cpf1**，**CasX**，**CasY** 的 **PAM** 序列都比较复杂多样，而 **C2c1** 识别严谨的 **5'-TTN**，因此它的靶标位点比其他系统容易被预测从而降低了潜在的脱靶效应。

总之，鉴于目前可获得的 **CRISPR/Cas** 系统都受限于一些缺陷，开发一种更稳健的、具有多方面良好性能的新型 **CRISPR/Cas** 系统对生物技术的发展具有重要意义。

发明内容

本申请的发明人经过大量实验和反复摸索，出人意料地发现了一种新型 **RNA** 指导的核酸内切酶。基于这一发现，本发明人开发了新的 **CRISPR/Cas** 系统以及基于该系统的基因编辑方法。

Cas 效应蛋白

因此，在第一方面，本发明提供了一种蛋白，其具有 SEQ ID NO：1 所示的氨基酸序列，或具有与 SEQ ID NO: 1 相比具有至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、或至少 99%的序列同一性的氨基酸序列，所述氨基酸序列基本保留了 SEQ ID NO: 1 的生物学功能。

在某些实施方案中，本发明提供了一种蛋白，其具有 SEQ ID NO：1 所示的氨基酸序列或其直系同源物（ortholog）、同源物或变体；其中，所述直系同源物、同源物或变体与 SEQ ID NO：1 相比具有至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、或至少 99%的序列同一性，并且基本保留了 SEQ ID NO: 1 的生物学功能。

在本发明中，上述序列的生物学功能包括但不限于，与导向 RNA 结合的活性、核酸内切酶活性、在导向 RNA 引导下与靶序列特定位点结合并切割的活性。

在某些实施方案中，所述蛋白是 CRISPR/Cas 系统中的效应蛋白。

在某些实施方案中，本发明的蛋白具有 SEQ ID NO：1 所示的氨基酸序列。

衍生的蛋白

本发明的蛋白可进行衍生化，例如被连接至另一个分子（例如另一个多肽或蛋白）。通常，蛋白的衍生化（例如，标记）不会不利影响该蛋白的期望活性（例如，与导向 RNA 结合的活性、核酸内切酶活性、在导向 RNA 引导下与靶序列特定位点结合并切割的活性）。因此，本发明的蛋白还意欲包括此类衍生化的形式。例如，可以将本发明的蛋白功能性连接（通过化学偶合、基因融合、非共价连接或其它方式）于一个或多个其它分子基团，例如另一个蛋白或多肽，检测试剂，药用试剂等。

特别地，可以将本发明的蛋白连接其他功能性单元。例如，可以将其与核定位信号（NLS）序列连接，以提高本发明的蛋白进入细胞核的能力。例如，可以将其与靶向部分连接，以使得本发明的蛋白具有靶向性。例如，可以将其与可检测的标记连接，以便于对本发明的蛋白进行检测。例如，可以将其与表位标签连接，以便于本发明的蛋白的表达、检测、示踪和/或纯化。

缀合物

因此，在第二方面，本发明提供了一种缀合物，其包含如上所述的蛋白和修饰部

分。

在某些实施方案中，所述修饰部分选自另外的蛋白或多肽、可检测的标记或其任意组合。

在某些实施方案中，所述另外的蛋白或多肽选自表位标签、报告基因序列、核定位信号（NLS）序列、靶向部分、转录激活结构域（例如，VP64）、转录抑制结构域（例如，KRAB 结构域或 SID 结构域）、核酸酶结构域（例如，Fok1），具有选自下列的活性的结构域：核苷酸脱氨酶、甲基化酶活性，去甲基化酶，转录激活活性，转录抑制活性，转录释放因子活性，组蛋白修饰活性，核酸酶活性，单链 RNA 切割活性，双链 RNA 切割活性，单链 DNA 切割活性，双链 DNA 切割活性和核酸结合活性；以及其任意组合。

在某些实施方案中，本发明的缀合物包含一个或多个 NLS 序列，例如 SV40 病毒大 T 抗原的 NLS。在某些示例性实施方案中，所述 NLS 序列如 SEQ ID NO: 19 所示。在某些实施方案中，所述 NLS 序列位于、靠近或接近本发明的蛋白的末端（例如，N 端或 C 端）。在某些示例性实施方案中，所述 NLS 序列位于、靠近或接近本发明的蛋白的 C 端。

在某些实施方案中，本发明的缀合物包含表位标签(epitope tag)。这类表位标签是本领域技术人员熟知的，其实例包括但不限于 His、V5、FLAG、HA、Myc、VSV-G、Trx 等，并且本领域技术人员已知如何根据期望目的（例如，纯化、检测或示踪）选择合适的表位标签。

在某些实施方案中，本发明的缀合物包含报告基因序列。这类报告基因是本领域技术人员熟知的，其实例包括但不限于 GST、HRP、CAT、GFP、HcRed、DsRed、CFP、YFP、BFP 等。

在某些实施方案中，本发明的缀合物包含能够与 DNA 分子或细胞内分子结合的结构域，例如麦芽糖结合蛋白(MBP)、Lex A 的 DNA 结合结构域(DBD)、GAL4 的 DBD 等。

在某些实施方案中，本发明的缀合物包含可检测的标记，例如荧光染料，例如 FITC 或 DAPI。

在某些实施方案中，本发明的蛋白任选地通过接头与所述修饰部分偶联、缀合或融合。

在某些实施方案中，所述修饰部分直接连接至本发明的蛋白的 N 端或 C 端。

在某些实施方案中，所述修饰部分通过接头连接至本发明的蛋白的 N 端或 C 端。这类接头是本领域熟知的，其实例包括但不限于包含一个或多个（例如，1 个，2 个，3

个，4 个或 5 个）氨基酸（如，Glu 或 Ser）或氨基酸衍生物（如，Ahx、β-Ala、GABA 或 Ava）的接头，或 PEG 等。

融合蛋白

在第三方面，本发明提供了一种融合蛋白，其包含本发明的蛋白以及另外的蛋白或多肽。

在某些实施方案中，所述另外的蛋白或多肽选自表位标签、报告基因序列、核定位信号（NLS）序列、靶向部分、转录激活结构域（例如，VP64）、转录抑制结构域（例如，KRAB 结构域或 SID 结构域）、核酸酶结构域（例如，Fok1），具有选自下列的活性的结构域：核苷酸脱氨酶、甲基化酶活性，去甲基化酶，转录激活活性，转录抑制活性，转录释放因子活性，组蛋白修饰活性，核酸酶活性，单链 RNA 切割活性，双链 RNA 切割活性，单链 DNA 切割活性，双链 DNA 切割活性和核酸结合活性；以及其任意组合。

在某些实施方案中，本发明的融合蛋白包含一个或多个 NLS 序列，例如 SV40 病毒大 T 抗原的 NLS。在某些实施方案中，所述 NLS 序列位于、靠近或接近本发明的蛋白的末端（例如，N 端或 C 端）。在某些示例性实施方案中，所述 NLS 序列位于、靠近或接近本发明的蛋白的 C 端。

在某些实施方案中，本发明的融合蛋白包含表位标签。

在某些实施方案中，本发明的融合蛋白包含报告基因序列。

在某些实施方案中，本发明的融合蛋白包含能够与 DNA 分子或细胞内分子结合的结构域。

在某些实施方案中，本发明的蛋白任选地通过接头与所述另外的蛋白或多肽融合。

在某些实施方案中，所述另外的蛋白或多肽直接连接至本发明的蛋白的 N 端或 C 端。

在某些实施方案中，所述另外的蛋白或多肽通过接头连接至本发明的蛋白的 N 端或 C 端。

在某些示例性实施方案中，本发明的融合蛋白具有 SEQ ID NO: 20 所示的氨基酸序列。

本发明的蛋白、本发明的缀合物或本发明的融合蛋白不受其产生方式的限定，例如，其可以通过基因工程方法（重组技术）产生，也可以通过化学合成方法产生。

同向重复序列

在第四方面，本发明提供了一种分离的核酸分子，其包含选自下列的序列，或由选自下列的序列组成：

(i) SEQ ID NO：7 或 13 所示的序列；

(ii) 与 **SEQ ID NO：7 或 13** 所示的序列相比具有一个或多个碱基的置换、缺失或添加（例如 **1** 个，**2** 个，**3** 个，**4** 个，**5** 个，**6** 个，**7** 个，**8** 个，**9** 个或 **10** 个碱基的置换、缺失或添加）的序列；

(iii) 与 **SEQ ID NO：7 或 13** 所示的序列具有至少 **20%**、至少 **30%**、至少 **40%**、至少 **50%**、至少 **60%**、至少 **70%**、至少 **80%**、至少 **90%**、至少 **95%**的序列同一性的序列；

(iv) 在严格条件下与**(i)-(iii)**任一项中所述的序列杂交的序列；或

(v) (i)-(iii)任一项中所述的序列的互补序列；

并且，**(ii)-(v)**中任一项所述的序列基本保留了其所源自的序列的生物学功能，所述序列的生物学功能是指，作为 **CRISPR-Cas** 系统中的同向重复序列的活性。

在某些实施方案中，所述分离的核酸分子是 **CRISPR-Cas** 系统中的同向重复序列。

在某些实施方案中，所述核酸分子包含选自下列的序列，或由选自下列的序列组成：

(a) SEQ ID NO：7 或 13 所示的核苷酸序列；

(b) 在严格条件下与**(a)**中所述的序列杂交的序列；或

(c) SEQ ID NO：7 或 13 所示的核苷酸序列的互补序列。

在某些实施方案中，所述分离的核酸分子是 **RNA**。

CRISPR/Cas 复合物

在第五方面，本发明提供了一种复合物，其包含：

(i) 蛋白组分，其选自：本发明的蛋白、缀合物或融合蛋白，及其任意组合；和

(ii) 核酸组分，其从 **5’**至 **3’**方向包含第四方面所述的分离的核酸分子和能够与靶序列杂交的导向序列，

其中，所述蛋白组分与核酸组分相互结合形成复合物。

在某些实施方案中，所述导向序列连接于所述核酸分子的 **3’**端。

在某些实施方案中，所述导向序列包含所述靶序列的互补序列。

在某些实施方案中，所述核酸组分是 CRISPR-Cas 系统中的导向 RNA。

在某些实施方案中，所述核酸分子是 RNA。

在某些实施方案中，所述复合物不包含反式作用 crRNA(tracrRNA)。

在某些实施方案中，所述导向序列在长度上为至少 5 个、至少 10 个、在某些实施方案中，所述导向序列在长度上为 10-30 个、或 15-25 个、或 15-22 个、或 19-25 个或 19-22 个核苷酸。

在某些实施方案中，所述分离的核酸分子在长度上为 55-70 个核苷酸，例如 55-65 个核苷酸，例如 60-65 个核苷酸，例如 62-65 个核苷酸，例如 63-64 个核苷酸。在某些实施方案中，所述分离的核酸分子在长度上为 15-30 个核苷酸，例如 15-25 个核苷酸，例如 20-25 个核苷酸，例如 22-24 个核苷酸，例如 23 个核苷酸。

在具体的实施方式中，本发明提供了一种 CRISPR-Cas 系统，其包含：

a) 导向 RNA，其从 5'至 3'方向包含同向重复序列和能够与靶序列杂交的导向序列，和

b) Cas12f 效应蛋白；

所述导向 RNA 与所述 Cas12f 效应蛋白形成复合物；

其中所述 Cas12f 蛋白大小为 900-1200 个氨基酸，在靠近其 C 端存在一个 RuvC 结构域，由 RuvC-I, RuvC-II 和 RuvC-III 基序构成；

其中所述 Cas12f 位于细菌基因组中 CRISPR 座 500bp 以内；

优选地，所述同向重复序列的长度为 21nt-36nt，所述导向序列的长度为 1-80nt，并且所述同向重复序列的末尾 16 或 17 个碱基可以形成一个茎环，环大小为 8 或 9nt，茎由五对互补碱基构成。

<u>编码核酸、载体及宿主细胞</u>

在第六方面，本发明提供了一种分离的核酸分子，其包含：

(i) 编码本发明的蛋白或融合蛋白的核苷酸序列；

(ii) 编码如第四方面所述的分离的核酸分子；或

(iii) 包含(i)和(ii)的核苷酸序列。

在某些实施方案中，(i)-(iii)任一项中所述的核苷酸序列经密码子优化用于在原核细胞

中进行表达。在某些实施方案中，**(i)-(iii)**任一项中所述的核苷酸序列经密码子优化用于在真核细胞中进行表达。

在第七方面，本发明还提供了一种载体，其包含如第六方面所述的分离的核酸分子。本发明的载体可以是克隆载体，也可以是表达载体。在某些实施方案中，本发明的载体是例如质粒，粘粒，噬菌体，柯斯质粒等等。在某些选实施方案中，所述载体能够在受试者（例如哺乳动物，例如人）体内表达本发明的蛋白、融合蛋白、如第四方面所述的分离的核酸分子或如第五方面所述的复合物。

在第八方面，本发明还提供了包含如上所述的分离的核酸分子或载体的宿主细胞。此类宿主细胞包括但不限于，原核细胞例如大肠杆菌细胞，以及真核细胞例如酵母细胞，昆虫细胞，植物细胞和动物细胞（如哺乳动物细胞，例如小鼠细胞、人细胞等）。本发明的细胞还可以是细胞系，例如 **293T** 细胞。

组合物及载体组合物

在第九方面，本发明还提供了一种组合物，其包含：

(i) 第一组分，其选自：本发明的蛋白、缀合物、融合蛋白、编码所述蛋白或融合蛋白的核苷酸序列，以及其任意组合；和

(ii) 第二组分，其为包含导向 **RNA** 的核苷酸序列，或者编码所述包含导向 **RNA** 的核苷酸序列的核苷酸序列；

其中，所述导向 **RNA** 从 **5’**至 **3’**方向包含同向重复序列和导向序列，所述导向序列能够与靶序列杂交；

所述导向 **RNA** 能够与**(i)**中所述的蛋白、缀合物或融合蛋白形成复合物。

在某些实施方案中，所述同向重复序列是如第四方面所定义的分离的核酸分子。

在某些实施方案中，所述导向序列连接至所述同向重复序列的 **3’**端。在某些实施方案中，所述导向序列包含所述靶序列的互补序列。

在某些实施方案中，所述组合物不包含 **tracrRNA**。

在某些实施方案中，所述组合物是非天然存在的或经修饰的。在某些实施方案中，所述组合物中的至少一个组分是非天然存在的或经修饰的。在某些实施方案中，所述第一组分是非天然存在的或经修饰的；和/或，所述第二组分是非天然存在的或经修饰的。

在某些实施方案中，当所述靶序列为 **DNA** 时，所述靶序列位于原间隔序列临近基序**(PAM)**的 **3’**端，并且所述 **PAM** 具有 **5’-TTN** 所示的序列，其中，**N** 选自 **A**、**G**、**T**、**C**。

在某些实施方案中，N 选自 A、T、C。

在某些实施方案中，当所述靶序列为 RNA 时，所述靶序列不具有 PAM 结构域限制。

在某些实施方案中，所述靶序列是来自原核细胞或真核细胞的 DNA 或 RNA 序列。在某些实施方案中，所述靶序列是非天然存在的 DNA 或 RNA 序列。

在某些实施方案中，所述靶序列存在于细胞内。在某些实施方案中，所述靶序列存在于细胞核内或细胞质（例如，细胞器）内。在某些实施方案中，所述细胞是真核细胞。在某些实施方案中，所述细胞是原核细胞。

在某些实施方案中，所述蛋白连接有一个或多个 NLS 序列。在某些实施方案中，所述缀合物或融合蛋白包含一个或多个 NLS 序列。在某些实施方案中，所述 NLS 序列连接至所述蛋白的 N 端或 C 端。在某些实施方案中，所述 NLS 序列融合至所述蛋白的 N 端或 C 端。

在第十方面，本发明还提供了一种组合物，其包含一种或多种载体，所述一种或多种载体包含：

(i) 第一核酸，其为编码本发明的蛋白或融合蛋白的核苷酸序列；任选地所述第一核酸可操作地连接至第一调节元件；以及

(ii) 第二核酸，其编码包含导向 RNA 的核苷酸序列；任选地所述第二核酸可操作地连接至第二调节元件；

其中：

所述第一核酸与第二核酸存在于相同或不同的载体上；

所述导向 RNA 从 5’至 3’方向包含同向重复序列和导向序列，所述导向序列能够与靶序列杂交；

所述导向 RNA 能够与(i)中所述的效应蛋白或融合蛋白形成复合物。

在某些实施方案中，所述同向重复序列是如第四方面所定义的分离的核酸分子。

在某些实施方案中，所述导向序列连接至所述同向重复序列的 3’端。在某些实施方案中，所述导向序列包含所述靶序列的互补序列。

在某些实施方案中，所述组合物不包含 tracrRNA。

在某些实施方案中，所述组合物是非天然存在的或经修饰的。在某些实施方案中，所述组合物中的至少一个组分是非天然存在的或经修饰的。

在某些实施方案中，所述第一调节元件是启动子，例如诱导型启动子。

在某些实施方案中，所述第二调节元件是启动子，例如诱导型启动子。

在某些实施方案中，当所述靶序列为 DNA 时，所述靶序列位于原间隔序列临近基序(PAM)的 3’端，并且所述 PAM 具有 5’-TTN 所示的序列，其中，N 选自 A、G、T、C。在某些实施方案中，N 选自 A、T、C。

在某些实施方案中，当所述靶序列为 RNA 时，所述靶序列不具有 PAM 结构域限制。

在某些实施方案中，所述靶序列是来自原核细胞或真核细胞的 DNA 或 RNA 序列。在某些实施方案中，所述靶序列是非天然存在的 DNA 或 RNA 序列。

在某些实施方案中，所述靶序列存在于细胞内。在某些实施方案中，所述靶序列存在于细胞核内或细胞质（例如，细胞器）内。在某些实施方案中，所述细胞是真核细胞。在某些实施方案中，所述细胞是原核细胞。

在某些实施方案中，所述蛋白连接有一个或多个 NLS 序列。在某些实施方案中，所述缀合物或融合蛋白包含一个或多个 NLS 序列。在某些实施方案中，所述 NLS 序列连接至所述蛋白的 N 端或 C 端。在某些实施方案中，所述 NLS 序列融合至所述蛋白的 N 端或 C 端。

在某些实施方案中，一种类型的载体是质粒，其是指其中可以例如通过标准分子克隆技术插入另外的 DNA 片段的环状双链 DNA 环。另一种类型的载体是病毒载体，其中病毒衍生的 DNA 或 RNA 序列存在于用于包装病毒(例如，逆转录病毒、复制缺陷型逆转录病毒、腺病毒、复制缺陷型腺病毒、以及腺相关病毒)的载体中。病毒载体还包含由用于转染到一种宿主细胞中的病毒携带的多核苷酸。某些载体(例如，具有细菌复制起点的细菌载体和附加型哺乳动物载体)能够在它们被导入的宿主细胞中自主复制。其他载体(例如，非附加型哺乳动物载体)在引入宿主细胞后整合到该宿主细胞的基因组中，并且由此与该宿主基因组一起复制。而且，某些载体能够指导它们可操作连接的基因的表达。这样的载体在此被称为“表达载体”。在重组 DNA 技术中使用的普通表达载体通常是质粒形式。

重组表达载体可包含处于适合于在宿主细胞中的核酸表达的形式的本发明的核酸分子，这意味着这些重组表达载体包含基于待用于表达的宿主细胞而选择的一种或多种调节元件，所述调节元件可操作地连接至待表达的核酸序列。

递送及递送组合物

本发明的蛋白、缀合物、融合蛋白、如第四方面所述的分离的核酸分子、本发明的复合物、如第六方面所述的分离的核酸分子、如第七方面所述的载体、如第九方面及第十方面所述的组合物，可以通过本领域已知的任何方法进行递送。此类方法包括但不限于，电穿孔、脂转染、核转染、显微注射、声孔效应、基因枪、磷酸钙介导的转染、阳离子转染、脂质体转染、树枝状转染、热激转染、核转染、磁转染、脂转染、穿刺转染、光学转染、试剂增强性核酸摄取、以及经由脂质体、免疫脂质体、病毒颗粒、人工病毒体等的递送。

因此，在另一个方面，本发明提供了一种递送组合物，其包含递送载体，以及选自下列的一种或多种：本发明的蛋白、缀合物、融合蛋白、如第四方面所述的分离的核酸分子、本发明的复合物、如第六方面所述的分离的核酸分子、如第七方面所述的载体、如第九方面及第十方面所述的组合物。

在某些实施方案中，所述递送载体是粒子。

在某些实施方案中，所述递送载体选自脂质颗粒、糖颗粒、金属颗粒、蛋白颗粒、脂质体、外泌体、微泡、基因枪或病毒载体（例如，复制缺陷型逆转录病毒、慢病毒、腺病毒或腺相关病毒）。

试剂盒

在另一个方面，本发明提供了一种试剂盒，其包含如上所述的组分中的一种或多种。在某些实施方案中，所述试剂盒包含一种或多种选自下列的组分：本发明的蛋白、缀合物、融合蛋白、如第四方面所述的分离的核酸分子、本发明的复合物、如第六方面所述的分离的核酸分子、如第七方面所述的载体、如第九方面及第十方面所述的组合物。

在某些实施方案中，本发明的试剂盒包含如第九方面所述的组合物。在某些实施方案中，所述试剂盒还包含使用所述组合物的说明书。

在某些实施方案中，本发明的试剂盒包含如第十方面所述的组合物。在某些实施方案中，所述试剂盒还包含使用所述组合物的说明书。

在某些实施方案中，本发明的试剂盒中包含的组分可以被提供于任何适合的容器中。

在某些实施方案中，所述试剂盒还包含一种或多种缓冲液。缓冲液可以是任何缓冲

液，包括但不限于碳酸钠缓冲液、碳酸氢钠缓冲液、硼酸盐缓冲液、Tris 缓冲液、MOPS 缓冲液、HEPES 缓冲液及其组合。在某些实施方案中，该缓冲液是碱性的。在某些实施方案中，该缓冲液具有从约 7 至约 10 的 pH。

在某些实施方案中，该试剂盒还包括一个或多个寡核苷酸，该一个或多个寡核苷酸对应于一个用于插入进载体中的导向序列，以便可操作地连接该导向序列和调节元件。在某些实施方案中，该试剂盒包括同源重组模板多核苷酸。

方法及用途

在另一个方面，本发明提供了一种修饰靶基因的方法，其包括：将如第五方面所述的复合物、如第九方面所述的组合物或如第十方面所述的组合物与所述靶基因接触，或者递送至包含所述靶基因的细胞中；所述靶序列存在于所述靶基因中。

在某些实施方案中，所述靶基因存在于细胞内。在某些实施方案中，所述细胞是原核细胞。在某些实施方案中，所述细胞是真核细胞。在某些实施方案中，所述细胞是哺乳动物细胞。在某些实施方案中，所述细胞是人类细胞。在某些实施方案中，所述细胞选自非人灵长类动物、牛、猪或啮齿类动物细胞。在某些实施方案中，所述细胞是非哺乳动物真核细胞，例如家禽或鱼等。在某些实施方案中，所述细胞是植物细胞，例如栽培植物（如木薯、玉米、高粱、小麦或水稻）、藻类、树或蔬菜具有的细胞。

在某些实施方案中，所述靶基因存在于体外的核酸分子（例如，质粒）中。在某些实施方案中，所述靶基因存在于质粒中。

在某些实施方案中，所述修饰是指所述靶序列的断裂，如 DNA 的双链断裂或 RNA 的单链断裂。

在某些实施方案中，所述断裂导致靶基因的转录降低。

在某些实施方案中，所述方法还包括：将编辑模板与所述靶基因接触，或者递送至包含所述靶基因的细胞中。在此类实施方案中，所述方法通过与外源模板多核苷酸同源重组修复所述断裂的靶基因，其中所述修复导致一种突变，包括所述靶基因的一个或多个核苷酸的插入、缺失、或取代。在某些实施方案中，所述突变导致在从包含该靶序列的基因表达的蛋白质中的一个或多个氨基酸改变。

因此，在某些实施方案中，所述修饰还包括将编辑模板（例如外源核酸）插入所述断裂中。

在某些实施方案中，所述的蛋白、缀合物、融合蛋白、分离的核酸分子、复合物、

载体或组合物包含于递送载体中。

在某些实施方案中，所述递送载体选自脂质颗粒、糖颗粒、金属颗粒、蛋白颗粒、脂质体、外泌体、病毒载体（如复制缺陷型逆转录病毒、慢病毒、腺病毒或腺相关病毒）。

在某些实施方案中，所述方法其用于改变靶基因或编码靶基因产物的核酸分子中的一个或多个靶序列来修饰细胞、细胞系或生物体。

在另一个方面，本发明提供了一种改变基因产物的表达的方法，其包括：将如第五方面所述的复合物、如第九方面所述的组合物或如第十方面所述的组合物与编码所述基因产物的核酸分子接触，或者递送至包含所述核酸分子的细胞中，所述靶序列存在于所述核酸分子中。

在某些实施方案中，所述核酸分子存在于细胞内。在某些实施方案中，所述细胞是原核细胞。在某些实施方案中，所述细胞是真核细胞。在某些实施方案中，所述细胞是哺乳动物细胞。在某些实施方案中，所述细胞是人类细胞。在某些实施方案中，所述细胞选自非人灵长类动物、牛、猪或啮齿类动物细胞。在某些实施方案中，所述细胞是非哺乳动物真核细胞，例如家禽或鱼等。在某些实施方案中，所述细胞是植物细胞，例如栽培植物（如木薯、玉米、高粱、小麦或水稻）、藻类、树或蔬菜具有的细胞。

在某些实施方案中，所述核酸分子存在于体外的核酸分子（例如，质粒）中。在某些实施方案中，所述核酸分子存在于质粒中。

在某些实施方案中，所述基因产物的表达被改变（例如，增强或降低）。在某些实施方案中，所述基因产物的表达被增强。在某些实施方案中，所述基因产物的表达被降低。

在某些实施方案中，所述基因产物是蛋白。

在某些实施方案中，所述的蛋白、缀合物、融合蛋白、分离的核酸分子、复合物、载体或组合物包含于递送载体中。

在某些实施方案中，所述递送载体选自脂质颗粒、糖颗粒、金属颗粒、蛋白颗粒、脂质体、外泌体、病毒载体（如复制缺陷型逆转录病毒、慢病毒、腺病毒或腺相关病毒）。

在某些实施方案中，所述方法其用于改变靶基因或编码靶基因产物的核酸分子中的一个或多个靶序列来修饰细胞、细胞系或生物体。

在另一个方面，本发明涉及如第一方面所述的蛋白、如第二方面所述的缀合物、如第三方面所述的融合蛋白、如第四方面所述的分离的核酸分子、如第五方面所述的复合物、如第六方面所述的分离的核酸分子、如第七方面所述的载体、如第九方面所述的组合物、如第十方面所述的组合物、本发明的试剂盒或递送组合物，用于核酸编辑的用途。

在某些实施方案中，所述核酸编辑包括基因或基因组编辑，例如修饰基因、敲除基因、改变基因产物的表达、修复突变、和/或插入多核苷酸。

在另一个方面，本发明涉及如第一方面所述的蛋白、如第二方面所述的缀合物、如第三方面所述的融合蛋白、如第四方面所述的分离的核酸分子、如第五方面所述的复合物、如第六方面所述的分离的核酸分子、如第七方面所述的载体、如第九方面所述的组合物、如第十方面所述的组合物、本发明的试剂盒或递送组合物，在制备制剂中的用途，所述制剂用于：

(i) 离体基因或基因组编辑；

(ii) 离体单链 **DNA** 的检测；

(iii) 编辑靶基因座中的靶序列来修饰生物或非人类生物；

(iv) 治疗由靶基因座中的靶序列的缺陷引起的病症。

细胞及细胞子代

在某些情况下，由本发明的方法引入到细胞的修饰可以使得细胞和其子代被改变以改进其生物产物（如抗体、淀粉、乙醇或其他期望的细胞输出物）的产生。在某些情况下，由本发明的方法引入到细胞的修饰可以使得细胞和其子代包括使所生产生物产物发生变化的改变。

因此，在另一方面，本发明还涉及如上所述的方法获得的细胞或其子代，其中所述细胞含有在其野生型中不存在的修饰。

本发明还涉及如上所述的细胞或其子代的细胞产物。

本发明还涉及一种体外的、离体的或体内的细胞或细胞系或它们的子代，所述细胞或细胞系或它们的子代包含：如第一方面所述的蛋白、如第二方面所述的缀合物、如第三方面所述的融合蛋白、如第四方面所述的分离的核酸分子、如第五方面所述的复合

物、如第六方面所述的分离的核酸分子、如第七方面所述的载体、如第九方面所述的组合物、如第十方面所述的组合物、本发明的试剂盒或递送组合物。

在某些实施方案中，所述细胞是原核细胞。

在某些实施方案中，所述细胞是真核细胞。在某些实施方案中，所述细胞是哺乳动物细胞。在某些实施方案中，所述细胞是人类细胞。某些实施方案中，所述细胞是非人哺乳动物细胞，例如非人灵长类动物、牛、羊、猪、犬、猴、兔、啮齿类（如大鼠或小鼠）的细胞。在某些实施方案中，所述细胞是非哺乳动物真核细胞，例如家禽鸟类（如鸡）、鱼类或甲壳动物（如蛤蜊、虾）的细胞。在某些实施方案中，所述细胞是植物细胞，例如单子叶植物或双子叶植物具有的细胞或栽培植物或粮食作物如木薯、玉米、高粱、大豆、小麦、燕麦或水稻具有的细胞，例如藻类、树或生产植物、果实或蔬菜（例如，树类如柑橘树、坚果树；茄属植物、棉花、烟草、番茄、葡萄、咖啡、可可等）。

在某些实施方案中，所述细胞是干细胞或干细胞系。

术语定义

在本发明中，除非另有说明，否则本文中使用的科学和技术名词具有本领域技术人员所通常理解的含义。并且，本文中所用的分子遗传学、核酸化学、化学、分子生物学、生物化学、细胞培养、微生物学、细胞生物学、基因组学和重组**DNA**等操作步骤均为相应领域内广泛使用的常规步骤。同时，为了更好地理解本发明，下面提供相关术语的定义和解释。

在本发明中，表述**“Cas12f”**是指，本发明人首次发现并鉴定的一种**Cas**效应蛋白，其具有选自下列的氨基酸序列：

(i) SEQ ID NOs：1、2、3任一项所示的序列；

(ii) 与**SEQ ID NOs：1、2、3**任一项所示的序列相比具有一个或多个氨基酸的置换、缺失或添加（例如**1**个，**2**个，**3**个，**4**个，**5**个，**6**个，**7**个，**8**个，**9**个或**10**个氨基酸的置换、缺失或添加）的序列；或

(iii) 与**SEQ ID NOs：1、2、3**任一项所示的序列具有至少**20%**、至少**30%**、至少**40%**、至少**50%**、至少**60%**、至少**70%**、至少**80%**、至少**80%**、至少**85%**、至少**90%**、至少**91%**、至少**92%**、至少**93%**、至少**94%**、至少**95%**、至少**96%**、至少**97%**、至少**98%**、或至少**99%**的序列同一性的序列。

本发明的**Cas12f**是一种在导向**RNA**引导下与靶序列特定位点结合并切割的核酸

内切酶，同时具有DNA和RNA内切酶活性。

如本文中所使用的，术语“规律成簇的间隔短回文重复(CRISPR)-CRISPR-相关(Cas)(CRISPR-Cas)系统”或“CRISPR 系统”可互换地使用并且具有本领域技术人员通常理解的含义，其通常包含与 CRISPR 相关(“Cas”)基因的表达有关的转录产物或其他元件，或者能够指导所述 Cas 基因活性的转录产物或其他元件。此类转录产物或其他元件可以包含编码 Cas 效应蛋白的序列和包含 CRISPR RNA(crRNA)的导向 RNA，以及在 CRISPR-Cas9 系统中所含有的反式作用 crRNA(tracrRNA)序列，或来自 CRISPR 基因座的其他序列或转录产物。在本发明所述的基于 Cas12f 的 CRISPR 系统中，不需要 tracrRNA 序列。

如本文中所使用的，术语“Cas 效应蛋白”、“Cas 效应酶”可互换地使用并且是指，CRISPR-Cas 系统中呈现的任一种大于长度 900 个氨基酸的蛋白质。在某些情况下，这类蛋白是指从 Cas 基因座中鉴定的蛋白。

如本文中所使用的，术语“导向 RNA(guide RNA)”、“成熟 crRNA”可互换地使用并且具有本领域技术人员通常理解的含义。一般而言，导向 RNA 可以包含同向(direct)重复序列和导向序列(guide sequence)，或者基本上由或由同向重复序列和导向序列（在内源性 CRISPR 系统背景下也称为间隔序列(spacer)）组成。在某些情况下，导向序列是与靶序列具有足够互补性从而与所述靶序列杂交并引导 CRISPR/Cas 复合物与所述靶序列的特异性结合的任何多核苷酸序列。在某些实施方案中，当最佳比对时，导向序列与其相应靶序列之间的互补程度为至少 50%、至少 60%、至少 70%、至少 80%、至少 90%、至少 95%、或至少 99%。确定最佳比对在本领域的普通技术人员的能力范围内。例如，存在公开和可商购的比对算法和程序，诸如但不限于 ClustalW、matlab 中的史密斯-沃特曼算法(Smith-Waterman)、Bowtie、Geneious、Biopython 以及 SeqMan。

在某些情况下，所述导向序列在长度上为至少 5 个、至少 10 个、至少 15 个、至少 16 个、至少 17 个、至少 18 个、至少 19 个、至少 20 个、至少 21 个、至少 22 个、至少 23 个、至少 24 个、至少 25 个、至少 26 个、至少 27 个、至少 28 个、至少 29 个、至少 30 个、至少 35 个、至少 40 个、至少 45 个或至少 50 个核苷酸。在某些情况下，所述导向序列在长度上为不超过 50 个、45 个、40 个、35 个、30 个、25 个、24 个、23 个、22 个、21 个、20 个、15 个、10 个或更少个核苷酸。在某些实施方案中，所述导向序列在长度上为 10-30 个、或 15-25 个、或 15-22 个、或 19-25 个或 19-22 个核苷酸。

在某些情况下，所述同向重复序列在长度上为至少 10 个、至少 15 个、至少 16 个、

至少 17 个、至少 18 个、至少 19 个、至少 20 个、至少 21 个、至少 22 个、至少 23 个、至少 24 个、至少 25 个、至少 26 个、至少 27 个、至少 28 个、至少 29 个、至少 30 个、至少 35 个、至少 40 个、至少 45 个、至少 50 个、至少 55 个、至少 56 个、至少 57 个、至少 58 个、至少 59 个、至少 60 个、至少 61 个、至少 62 个、至少 63 个、至少 64 个、至少 65 个或至少 70 个核苷酸。在某些情况下，所述同向重复序列在长度上为不超过 70 个、65 个、64 个、63 个、62 个、61 个、60 个、59 个、58 个、57 个、56 个、55 个、50 个、45 个、40 个、35 个、30 个、29 个、28 个、27 个、26 个、25 个、24 个、23 个、22 个、21 个、20 个、15 个、10 个或更少个核苷酸。在某些实施方案中，所述同向重复序列在长度上为 55-70 个核苷酸，例如 55-65 个核苷酸，例如 60-65 个核苷酸，例如 62-65 个核苷酸，例如 63-64 个核苷酸。在某些实施方案中，所述同向重复序列在长度上为 15-30 个核苷酸，例如 15-25 个核苷酸，例如 20-25 个核苷酸，例如 22-24 个核苷酸，例如 23 个核苷酸。

如本文中所使用的，术语“CRISPR/Cas 复合物”是指，导向 RNA(guide RNA)或成熟 crRNA 与 Cas 蛋白结合所形成的核糖核蛋白复合体，其包含杂交到靶序列上并且与 Cas 蛋白结合的导向序列。该核糖核蛋白复合体能够识别并切割能与该导向 RNA 或成熟 crRNA 杂交的多核苷酸。

因此，在形成 CRISPR/Cas 复合物的情况下，“靶序列”是指被设计为具有靶向性的导向序列所靶向的多核苷酸，例如与该导向序列具有互补性的序列，其中靶序列与导向序列之间的杂交将促进 CRISPR/Cas 复合物的形成。完全互补性不是必需的，只要存在足够互补性以引起杂交并且促进一种 CRISPR/Cas 复合物的形成即可。靶序列可以包含任何多核苷酸，如 DNA 或 RNA。在某些情况下，所述靶序列位于细胞的细胞核或细胞质中。在某些情况下，该靶序列可位于真核细胞的一个细胞器例如线粒体或叶绿体内。可被用于重组到包含该靶序列的靶基因座中的序列或模板被称为“编辑模板”或“编辑多核苷酸”或“编辑序列”。在某些实施方案中，所述编辑模板为外源核酸。在某些实施方案中，该重组是同源重组。

在本发明中，表述“靶序列”或“靶多核苷酸”可以是对细胞（例如，真核细胞）而言任何内源或外源的多核苷酸。例如，该靶多核苷酸可以是一种存在于真核细胞的细胞核中的多核苷酸。该靶多核苷酸可以是一个编码基因产物(例如，蛋白质)的序列或一个非编码序列(例如，调节多核苷酸或无用 DNA)。在某些情况下，据信该靶序列应该与原间隔序列临近基序(PAM)相关。对 PAM 的精确序列和长度要求取决于使用的 Cas 效应酶而不

同，但是 PAM 典型地是临近原间隔序列(也即，靶序列)的 2-5 个碱基对序列。本领域技术人员能够鉴定与给定的 Cas 效应蛋白一起使用的 PAM 序列。

在某些情况下，靶序列或靶多核苷酸可以包括多个疾病相关基因和多核苷酸以及信号传导生化途径相关基因和多核苷酸。此类靶序列或靶多核苷酸的非限制性实例，包括分别提交于 2012 年 12 月 12 日和 2013 年 1 月 2 日的美国临时专利申请 61/736,527 和 61/748,427、提交于 2013 年 12 月 12 日的国际申请 PCT/US2013/074667 中所列举的那些，其全部通过引用并入本文。

在某些情况下，靶序列或靶多核苷酸的实例包括与信号传导生化途径相关的序列，例如信号传导生化途径相关基因或多核苷酸。靶多核苷酸的实例包括疾病相关基因或多核苷酸。“疾病相关”基因或多核苷酸是指与非疾病对照的组织或细胞相比，在来源于疾病影响的组织的细胞中以异常水平或以异常形式产生转录或翻译产物的任何基因或多核苷酸。在改变的表达与疾病的出现和/或进展相关的情况下，它可以是一个以异常高的水平被表达的基因；或者，它可以是一个以异常低的水平被表达的基因。疾病相关基因还指具有一个或多个突变或直接负责或与一个或多个负责疾病的病因学的基因连锁不平衡的遗传变异的基因。转录的或翻译的产物可以是已知的或未知的，并且可以处于正常或异常水平。

如本文中所使用的，术语“野生型”具有本领域技术人员通常理解的含义，其表示生物、菌株、基因的典型形式或者当它在自然界存在时区别于突变体或变体形式的特征，其可从自然中的来源分离并且没有被人为有意地修饰。

如本文中所使用的，术语“非天然存在的”或“工程化的”可互换地使用并且表示人工的参与。当这些术语用于描述核酸分子或多肽时，其表示该核酸分子或多肽至少基本上从它们在自然界中或如发现于自然界中的与其结合的至少另一种组分游离出来。

如本文中所使用的，术语“直系同源物（orthologue, ortholog）”具有本领域技术人员通常理解的含义。作为进一步指导，如本文中所述的蛋白质的“直系同源物”是指属于不同物种的蛋白质，该蛋白质执行与作为其直系同源物的蛋白相同或相似的功能。

如本文中所使用的，术语“同一性”用于指两个多肽之间或两个核酸之间序列的匹配情况。当两个进行比较的序列中的某个位置都被相同的碱基或氨基酸单体亚单元占据时(例如，两个 DNA 分子的每一个中的某个位置都被腺嘌呤占据，或两个多肽的每一个中的某个位置都被赖氨酸占据)，那么各分子在该位置上是同一的。两个序列之间的“百分数同一性”是由这两个序列共有的匹配位置数目除以进行比较的位置

数目×100 的函数。例如，如果两个序列的 10 个位置中有 6 个匹配，那么这两个序列具有 60%的同一性。例如，DNA 序列 CTGACT 和 CAGGTT 共有 50%的同一性(总共 6 个位置中有 3 个位置匹配)。通常，在将两个序列比对以产生最大同一性时进行比较。这样的比对可通过使用，例如，可通过计算机程序例如 Align 程序(DNAstar, Inc.)方便地进行的 Needleman 等人(1970)*J. Mol. Biol.* 48：443-453 的方法来实现。还可使用已整合入 ALIGN 程序(版本 2.0)的 E. Meyers 和 W. Miller (Comput. Appl Biosci.，4:11-17 (1988))的算法，使用 PAM120 权重残基表(weight residue table)、12 的缺口长度罚分和 4 的缺口罚分来测定两个氨基酸序列之间的百分数同一性。此外，可使用已整合入 GCG 软件包(可在 www.gcg.com 上获得)的 GAP 程序中的 Needleman 和 Wunsch (J MoI Biol. 48:444-453 (1970))算法，使用 Blossum 62 矩阵或 PAM250 矩阵以及 16、14、12、10、8、6 或 4 的缺口权重(gap weight)和 1、2、3、4、5 或 6 的长度权重来测定两个氨基酸序列之间的百分数同一性。

如本文中所使用的，术语“载体”是指，可将多聚核苷酸插入其中的一种核酸运载工具。当载体能使插入的多核苷酸编码的蛋白获得表达时，载体称为表达载体。载体可以通过转化，转导或者转染导入宿主细胞，使其携带的遗传物质元件在宿主细胞中获得表达。载体是本领域技术人员公知的，包括但不限于：质粒；噬菌粒；柯斯质粒；人工染色体，例如酵母人工染色体(YAC)、细菌人工染色体(BAC)或 P1 来源的人工染色体(PAC)；噬菌体如 λ 噬菌体或 M13 噬菌体及动物病毒等。可用作载体的动物病毒包括但不限于，逆转录酶病毒(包括慢病毒)、腺病毒、腺相关病毒、疱疹病毒(如单纯疱疹病毒)、痘病毒、杆状病毒、乳头瘤病毒、乳头多瘤空泡病毒(如 SV40)。一种载体可以含有多种控制表达的元件，包括但不限于，启动子序列、转录起始序列、增强子序列、选择元件及报告基因。另外，载体还可含有复制起始位点。

如本文中所使用的，术语“宿主细胞”是指，可用于导入载体的细胞，其包括但不限于，如大肠杆菌或枯草菌等的原核细胞，如酵母细胞或曲霉菌等的真菌细胞，如 S2 果蝇细胞或 Sf9 等的昆虫细胞，或者如纤维原细胞，CHO 细胞，COS 细胞，NSO 细胞，HeLa 细胞，BHK 细胞，HEK 293 细胞或人细胞等的动物细胞。

本领域技术人员将理解，表达载体的设计可取决于诸如待转化的宿主细胞的选择、所希望的表达水平等因素。一种载体可以被引入到宿主细胞中而由此产生转录物、蛋白质、或肽，包括由如本文所述的蛋白、融合蛋白、分离的核酸分子等(例如，CRISPR 转

录物，如核酸转录物、蛋白质、或酶)。

如本文中所使用的，术语“调节元件”旨在包括启动子、增强子、内部核糖体进入位点(IRES)、和其他表达控制元件(例如转录终止信号，如多聚腺苷酸化信号和多聚 U 序列)，其详细描述可参考戈德尔(Goeddel)，《基因表达技术：酶学方法》(GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY)185，学术出版社(Academic Press)，圣地亚哥(San Diego)，加利福尼亚州(1990)。在某些情况下，调节元件包括指导一个核苷酸序列在许多类型的宿主细胞中的组成型表达的那些序列以及指导该核苷酸序列只在某些宿主细胞中表达的那些序列(例如，组织特异型调节序列)。组织特异型启动子可主要指导在感兴趣的期望组织中的表达，所述组织例如肌肉、神经元、骨、皮肤、血液、特定的器官(例如肝脏、胰腺)、或特殊的细胞类型(例如淋巴细胞)。在某些情况下，调节元件还可以时序依赖性方式(如以细胞周期依赖性或发育阶段依赖性方式)指导表达，该方式可以是或者可以不是组织或细胞类型特异性的。在某些情况下，术语“调节元件”涵盖的是增强子元件，如 WPRE；CMV 增强子；在 HTLV-I 的 LTR 中的 R-U5’片段((Mol.Cell.Biol.，第 8(1)卷，第 466-472 页，1988)；SV40 增强子；以及在兔 β-珠蛋白的外显子 2 与 3 之间的内含子序列(Proc.Natl.Acad.Sci.USA.，第 78(3)卷，第 1527-31 页，1981)。

如本文中所使用的，术语“启动子”具有本领域技术人员公知的含义，其是指一段位于基因的上游能启动下游基因表达的非编码核苷酸序列。组成型(constitutive)启动子是这样的核苷酸序列：当其与编码或者限定基因产物的多核苷酸可操作地相连时，在细胞的大多数或者所有生理条件下，其导致细胞中基因产物的产生。诱导型启动子是这样的核苷酸序列，当可操作地与编码或者限定基因产物的多核苷酸相连时，基本上只有当对应于所述启动子的诱导物在细胞中存在时，其导致所述基因产物在细胞内产生。组织特异性启动子是这样的核苷酸序列：当可操作地与编码或者限定基因产物的多核苷酸相连时，基本上只有当细胞是该启动子对应的组织类型的细胞时，其才导致在细胞中产生基因产物。

如本文中所使用的，术语“可操作地连接”旨在表示感兴趣的核苷酸序列以一种允许该核苷酸序列的表达的方式被连接至该一种或多种调节元件(例如，处于一种体外转录/翻译系统中或当该载体被引入到宿主细胞中时，处于该宿主细胞中)。

如本文中所使用的，术语“互补性”是指核酸与另一个核酸序列借助于传统的沃森-克里克或其他非传统类型形成一个或多个氢键的能力。互补百分比表示一个核酸分子中可

与一个第二核酸序列形成氢键(例如，沃森-克里克碱基配对)的残基的百分比(例如，10 个之中有 5、6、7、8、9、10 个即为 50%、60%、70%、80%、90%、和 100%互补)。“完全互补”表示一个核酸序列的所有连续残基与一个第二核酸序列中的相同数目的连续残基形成氢键。如本文使用的“基本上互补”是指在一个具有 8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50 个或更多个核苷酸的区域上至少为 60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%、或 100%的互补程度，或者是指在严格条件下杂交的两个核酸。

如本文中所使用的，对于杂交的“严格条件”是指与靶序列具有互补性的一个核酸主要地与该靶序列杂交并且基本上不杂交到非靶序列上的条件。严格条件通常是序列依赖性的，并且取决于许多因素而变化。一般而言，该序列越长，则该序列特异性地杂交到其靶序列上的温度就越高。严格条件的非限制性实例描述于蒂森(Tijssen)(1993)的《生物化学和分子生物学中的实验室技术-核酸探针杂交》(Laboratory Techniques In BiochemistryAnd Molecular Biology-Hybridization With Nucleic Acid Probes)，第 I 部分，第二章，“杂交原理概述和核酸探针分析策略”(“Overview of principles of hybridization andthe strategy of nucleic acid probe assay”)，爱思唯尔(Elsevier)，纽约。

如本文中所使用的，术语“杂交”是指其中一个或多个多核苷酸反应形成一种复合物的反应，该复合物经由这些核苷酸残基之间的碱基的氢键键合而稳定化。氢键键合可以借助于沃森-克里克碱基配对、Hoogstein 结合或以任何其他序列特异性方式而发生。该复合物可包含形成一个双链体的两条链、形成多链复合物的三条或多条链、单个自我杂交链、或这些的任何组合。杂交反应可以构成一个更广泛的过程(如 PCR 的开始、或经由一种酶的多核苷酸的切割)中的一个步骤。能够与一个给定序列杂交的序列被称为该给定序列的“互补物”。

如本文中所使用的，术语“表达”是指，藉此从 DNA 模板转录成多核苷酸(如转录成 mRNA 或其他 RNA 转录物)的过程和/或转录的 mRNA 随后藉此翻译成肽、多肽或蛋白质的过程。转录物和编码的多肽可以总称为“基因产物”。如果多核苷酸来源于基因组 DNA，表达可以包括真核细胞中 mRNA 的剪接。

如本文中所使用的，术语“接头”是指，由多个氨基酸残基通过肽键连接形成的线性多肽。本发明的接头可以为人工合成的氨基酸序列，或天然存在的多肽序列，例如具有铰链区功能的多肽。此类接头多肽是本领域众所周知的(参见例如，Holliger, P.等人（1993)Proc. Natl. Acad. Sci. USA 90 :6444-6448 ； Poljak, R. J.等人（1994) Structure

2:1121-1123)。

如本文中所使用的，术语“治疗”是指，治疗或治愈病症，延缓病症的症状的发作，和/或延缓病症的发展。

如本文中所使用的，术语“受试者”包括但不限于各种动物，例如哺乳动物，例如牛科动物、马科动物、羊科动物、猪科动物、犬科动物、猫科动物、兔科动物、啮齿类动物（例如，小鼠或大鼠）、非人灵长类动物（例如，猕猴或食蟹猴）或人。在某些实施方式中，所述受试者（例如人）患有病症（例如，疾病相关基因缺陷所导致的病症）。

发明的有益效果

与现有技术相比，本发明的 Cas 蛋白及系统具有显著的有利方面。例如，本发明的 Cas 效应蛋白的 PAM 结构域为严格的 5’-TTN 结构，而且靶标序列前面的第二位和第三位碱基近 100%为 T，而其他位置可以为任意的序列，比目前已经报道的最严谨 PAM 识别的 C2c1 拥有更加严谨的 PAM 识别方式，从而显著降低脱靶效应。例如，本发明的 Cas 效应蛋白能在真核生物体内进行 DNA 切割，在分子大小上比 Cpf1 和 Cas9 蛋白小约 200-300 个氨基酸，因此转染效率上明显优于 Cpf1 和 Cas9。

下面将结合附图和实施例对本发明的实施方案进行详细描述，但是本领域技术人员将理解，下列附图和实施例仅用于说明本发明，而不是对本发明的范围的限定。根据附图和优选实施方案的下列详细描述，本发明的各种目的和有利方面对于本领域技术人员来说将变得显然。

附图说明

图 1 为实施例 2 中 Cas12f.4，Cas12f.5 和 Cas12f.6 的 crRNA 结构分析结果，显示了 Repeat 序列的二级结构。

图 2 为实施例 3 中 PAM 结构域分析结果。

图 3a-图 3c 为实施例 4 中 Cas12f.4 在人类细胞系中的切割活性检测结果。

图 4a-图 4c 为实施例 5 中 Cas12f.4 在玉米原生质体细胞中的切割活性检测结果。

序列信息

本发明涉及的部分序列的信息提供于下面的表 1 中。

表 1：序列的描述

SEQ ID NO:	描述
1	Cas12f.4 的氨基酸序列
2	Cas12f.5 的氨基酸序列
3	Cas12f.6 的氨基酸序列
4	Cas12f.4 的编码核酸序列
5	Cas12f.5 的编码核酸序列
6	Cas12f.6 的编码核酸序列
7	Cas12f.4/原型同向重复序列
8	Cas12f.5/原型同向重复序列
9	Cas12f.6/原型同向重复序列
10	Cas12f.4/原型同向重复序列的编码核酸序列
11	Cas12f.5/原型同向重复序列的编码核酸序列
12	Cas12f.6/原型同向重复序列的编码核酸序列
13	Cas12f.4/成熟同向重复序列
14	Cas12f.5/成熟同向重复序列
15	Cas12f.6/成熟同向重复序列
16	Cas12f.4/成熟同向重复序列的编码核酸序列
17	Cas12f.5/成熟同向重复序列的编码核酸序列
18	Cas12f.6/成熟同向重复序列的编码核酸序列
19	NLS 序列
20	Cas12f.4-NLS 融合蛋白的氨基酸序列
21	Cas12f.5-NLS 融合蛋白的氨基酸序列
22	Cas12f.6-NLS 融合蛋白的氨基酸序列
23	表达 Cas12f.4 系统的质粒
24	PAM 文库序列
25	Cas12f.4 系统的导向 RNA-VEGFA
26	Cas12f.5 系统的导向 RNA-VEGFA
27	Cas12f.6 系统的导向 RNA-VEGFA
28	Cas12f.4 系统的导向 RNA-PDI1
29	Cas12f.4 系统的导向 RNA-SBE2.2

具体实施方式

现参照下列意在举例说明本发明（而非限定本发明）的实施例来描述本发明。

除非特别指明，否则基本上按照本领域内熟知的以及在各种参考文献中描述的常规方法进行实施例中描述的实验和方法。例如，本发明中所使用的免疫学、生物化学、化学、分子生物学、微生物学、细胞生物学、基因组学和重组 DNA 等常规技术，可参见参见萨姆布鲁克(Sambrook)、弗里奇(Fritsch)和马尼亚蒂斯(Maniatis)，《分子克隆：实验室手册》(MOLECULAR CLONING:A LABORATORY MANUAL)，第 2 次编辑(1989)；《当代分子生物学实验手册》(CURRENT PROTOCOLS IN MOLECULAR BIOLOGY)(F.M.奥苏贝尔(F.M.Ausubel)等人编辑，(1987))；《酶学方法》(METHODS IN ENZYMOLOGY)系列(学术出版公司)：《PCR 2：实用方法》(PCR 2:A PRACTICAL APPROACH)(M.J.麦克弗森(M.J.MacPherson)、B.D.黑姆斯(B.D.Hames)和 G.R.泰勒(G.R.Taylor)编辑(1995))、哈洛(Harlow)和拉内(Lane)编辑(1988)《抗体：实验室手册》(ANTIBODIES,A LABORATORY MANUAL)，以及《动物细胞培养》(ANIMAL CELL CULTURE)(R.I.弗雷谢尼(R.I.Freshney)编辑(1987))。

另外，实施例中未注明具体条件者，按照常规条件或制造商建议的条件进行。所用试剂或仪器未注明生产厂商者，均为可以通过市购获得的常规产品。本领域技术人员知晓，实施例以举例方式描述本发明，且不意欲限制本发明所要求保护的范围。本文中提及的全部公开案和其他参考资料以其全文通过引用合并入本文。

以下实施例涉及的部分试剂的来源如下：

LB 液体培养基：10g 胰蛋白胨（Tryptone），5g 酵母提取物（Yeast Extract），10g NaCl，定容至 1L，灭菌。若需加抗生素，则待培养基冷却后加，50μg/ml 的终浓度。

氯仿/异戊醇：240ml 的氯仿加 10ml 的异戊醇，混匀。

RNP 缓冲液：100 mM 氯化钠, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 μg/ml BSA, pH 7.9。

原核表达载体 pACYC-Duet-1 和 pUC19 购自北京全式金生物技术有限公司。

大肠杆菌感受态 EC100 购自 Epicentre 公司。

实施例 1. Cas12f 基因和 Cas12f 导向 RNA 的获得

1、 CRISPR 和基因的注释：使用 Prodigal 对将 NCBI 和 JGI 数据库的微生物基因

组和宏基因组数据进行基因注释得到所有蛋白，同时用 Piler-CR 进行 CRISPR 座的注释，参数均为默认参数。

2、蛋白质的过滤：通过序列一致性对注释蛋白去冗余，去除序列完全一致的蛋白，同时将长度大于 800 个氨基酸的蛋白划分为大分子蛋白。由于目前发现的所有第二类 CRISPR/Cas 系统的效应蛋白长度多大于 900 个氨基酸，所以为了降低计算复杂度，我们在挖掘 CRISPR 效应蛋白的时候只对大分子蛋白进行考虑。

3、CRISPR 相关大分子蛋白的获得：将每一个 CRISPR 座上下游延伸 10 Kb，将对 CRISPR 邻近区间内的非冗余大分子蛋白进行鉴定。

4、CRISPR 相关大分子蛋白质的聚类：使用 BLASTP 对非冗余大分子 CRISPR 相关蛋白进行内部的两两比对，输出 Evalue<1E-10 的比对结果。使用 MCL 对 BLASTP 的输出结果进行聚类分析，CRISPR 相关蛋白质家族。

5、CRISPR 富集大分子蛋白质家族的鉴定：使用 BLASTP 对 CRISPR 相关蛋白质家族的蛋白比对到去除去 CRISPR 相关蛋白的非冗余大分子蛋白数据库，输出 Evalue<1E-10 的比对结果。如果一个非 CRISPR 相关蛋白数据库发现的同源蛋白小于 100%，那么则说明这个家族的蛋白在 CRISPR 区域是富集的，通过这种方法我们对 CRISPR 富集大分子蛋白质家族进行鉴定。

6、蛋白功能和结构域的注释：利用 Pfam 数据库，NR 数据库以及从 NCBI 收集的 Cas 蛋白对 CRISPR 富集大分子蛋白质家族进行注释，得到新的 CRISPR/Cas 蛋白质家族。利用 Mafft 对每个 CRISPR/Cas 家族蛋白进行多重序列比对，然后用 JPred 和 HHpred 进行保守结构域分析，鉴定含有 RuvC 结构域的蛋白质家族。

在此基础上，本发明人获得了一种全新的 Cas 效应蛋白，即 Cas12f，以其三种活性同源物序列，分别命名为 Cas12f.4（SEQ ID NO: 1）、Cas12f.5（SEQ ID NO: 2）、Cas12f.6（SEQ ID NO: 3），三种同源物的编码 DNA 分别如 SEQ ID NOs: 4、5、6 所示。Cas12f.4、Cas12f.5、Cas12f.6 所对应的原型同向重复序列（pre-crRNA 中所含有的 repeat 序列）分别如 SEQ ID NOs: 7、8、9 所示。Cas12f.4、Cas12f.5、Cas12f.6 所对应的成熟同向重复序列（成熟 crRNA 中所含有的 repeat 序列）分别如 SEQ ID NOs: 13、14、15 所示。

实施例 2. Cas12f 基因对成熟 crRNA 的加工

1、人工合成 SEQ ID NO: 4 所示的双链 DNA 分子，同时人工合成 SEQ ID NO: 10

所示双链 DNA 分子。

2、将步骤 1 合成的双链 DNA 分子与原核表达载体 pACYC-Duet-1 连接，得到重组质粒 pACYC-Duet-1+CRISPR/Cas12f。

对重组质粒 pACYC-Duet-1+CRISPR/Cas12f 进行测序。测序结果表明，重组质粒 pACYC-Duet-1+CRISPR/Cas12f 中含有 SEQ ID NO: 4 和 SEQ ID NO: 10 所示的序列，并且表达 SEQ ID NO: 1 所示的 Cas12f.4 蛋白和 SEQ ID NO: 7 所示的 Cas12f.4 原型同向重复序列。将重组质粒 pACYC-Duet-1+CRISPR/Cas12f 导入大肠杆菌 EC100，得到重组菌，将该重组菌命名为 EC100-CRISPR/Cas12f。

3、取 EC100-CRISPR/Cas12f 的单克隆，接种至 100mL LB 液体培养基（含 50μg/mL 氨苄霉素），37℃、200rpm 振荡培养 12h，得到培养菌液。

4、提取细菌 RNA：转移 1.5 mL 细菌培养物到预冷微量离心管中，在 4℃，6000×g，离心 5 分钟。离心后，弃上清液，将细胞沉淀重新悬浮于预热至 95°C 的 200μL Max Bacterial Enhancement Reagent 中，吹吸混匀混合。95℃孵育 4 分钟。向溶解产物中加入 1 mL TRIzol®Reagent 并吹吸混匀，室温下孵育 5 分钟。加入 0.2mL 冷氯仿，用手摇动管混合 15 秒，室温下孵育 2-3 分钟。4℃，12,000×g 离心 15 分钟。取 600μL 上清于新管中，加入 0.5mL 冷异丙醇沉淀 RNA，颠倒混匀，室温下孵育 10 分钟。4℃下以 15,000×g 离心 10 分钟，弃上清，加入 1 mL 75％乙醇，涡旋混匀。4℃，7500×g 离心 5 分钟，弃上清，空气风干。将 RNA 沉淀溶解到 50μL RNase-free water 中，在 60℃下孵育 10 分钟。

5、DNA 的消化：20ugRNA 溶解到 39.5μL dH_2O，65℃，5min。冰上 5min，加入 0.5μL RNAI，5μL buffer, 5μL DNaseI,37℃ 45min(50μL 体系)。加 50μL dH_2O，调整体积到 100μL。2mL Phase-Lock tube 16000g 离心 30s 后，加 100μL 酚：氯仿：异戊醇（25:24:1）、100μL 消化的 RNA，摇 15s，15℃，16000g 离心 12min。取上清于一个新的 1.5mL 离心管中，加入与上清等体积的异丙醇 1/10 NaoAC，反应 1h 或-20℃过夜。4℃,16000g 离心 30min，弃上清。加 350μL 75% 乙醇洗涤沉淀，4℃,16000g 离心 10min，弃上清。晾干，加入 20μL RNase-free water，65℃，5min 溶解沉淀。NanoDrop 测浓度，跑胶。

6、3’脱磷酸化及 5’磷酸化：将消化的 RNA~20ug，各加水至 42.5μL，90℃ 2min。冰上冷却 5min。加 5μL 10×T4 PNK buffer；0.5 μL RNaI，2 μL T4 PNK（50μL），37℃ 6h。加 1μL T4 PNK，1.25μL(100mM)ATP，37℃ 1h。加 47.75μL dH_2O，调整体积

到 100μL。2mL Phase-Lock tube 16000g 离心 30s 后，加 100μL 酚：氯仿：异戊醇（25:24:1）、100μL 消化的 RNA，摇 15s，15℃，16000g 离心 12min。取上清于一个新的 1.5mL 离心管中，加与上清等体积的异丙醇，总体积 1/10 NaoAC，反应 1h 或-20℃过夜。4℃,16000g 离心 30min，弃上清。加 350μL 75%乙醇洗涤沉淀，4℃,16000g 离心 10min，弃上清。晾干，加入 21μL RNase-free water，65℃，5min 溶解沉淀，NanoDrop 测浓度。

7、RNA 单磷酸化：20μL RNA ，90℃ 1min，冰上冷却 5min。加入 2μL RNA 5' Polphosphatase 10×Reaction buffer，0.5μL Inhibitor，1μL RNA 5' Polphosphatase（20Units），加 RNase-free water 至 20μL，37℃ 60min。加 80μL dH_2O，调整体积到 100μL。2mL Phase-Lock tube16000g 离心 30s 后，加 100μL 酚：氯仿：异戊醇（25:24:1）、100μL 消化的 RNA，摇 15s，15℃，16000g 离心 12min。取上清于一个新的 1.5mL 离心管中，加与上清等体积的异丙醇，总体积 1/10 NaoAC，反应 1h 或-20℃过夜。4℃,16000g 离心 30min，弃上清，加 350μL 75%乙醇洗涤沉淀，4℃,16000g 离心 10min，弃上清。晾干，加入 21μL RNase-free water，65℃，5min 溶解沉淀，NanoDrop 测浓度。

8、cDNA 文库的准备：16.5μL RNase-free water。5μL Poly（A）Polymerase 10×Reaction buffer。5μL 10mM ATP。1.5μL RiboGuard RNase Inhibitor。20μL RNA Substrate。2μL Poly（A）Polymerase（4Units）。50μL 总体积。37℃ 20min。加 50μL dH_2O，调整体积到 100μL。2mL Phase-Lock tube16000g 离心 30s 后，加 100μL 酚：氯仿：异戊醇（25:24:1）、100μL 消化的 RNA，摇 15s，15℃，16000g 离心 12min。取上清于一个新的 1.5mL 离心管中，加与上清等体积的异丙醇，总体积 1/10 NaoAC，反应 1h 或-20℃过夜。4℃,16000g 离心 30min，弃上清，晾干，加入 11μL RNase-free water，65℃，5min 溶解沉淀，NanoDrop 测浓度。

9、将 cDNA 文库加上测序接头后送至北京贝瑞合康进行测序。

10、对原始数据进行质量过滤，去除碱基平均质量值低于 30 的序列。对序列去除接头后，保留 25 nt 到 50 nt 的 RNA 序列，用 bowtie 将其比对到 CRISPR 阵列的参考序列上。

11、通过比对我们发现 Cas12f.4 的 pre-crRNA 在大肠杆菌体内能成功的被加工成 45nt 的成熟 crRNA，其中由 23nt 的 Repeat 序列和 19-22nt 的导向序列构成。

12、用 ViennaRNA 和 VARNA 对成熟的 crRNA 进行结构预测和可视化分析我们发

现，crRNA 的 Repeat 序列的 3’端可以形成一个 8 个碱基大小的颈环（图 1）。

13、我们对 Cas12f.5 和 Cas12f.6 的 crRNA 3’端 23nt 的序列进行预测后发现了相似的二级结构（图 1）。

实施例 3. Cas12f 基因的 PAM 结构域鉴定

1.构建重组质粒 pACYC-Duet-1+CRISPR/Cas12f 并测序。根据测序结果，对重组质粒 pACYC-Duet-1+CRISPR/Cas12f 进行结构描述如下：将载体 pACYC-Duet-1 的限制性内切酶 Pml I 和 Kpn I 识别序列间的小片段替换为 SEQ ID NO: 4 所示的序列中自 5’末端起第 1 至 3713 位所示的双链 DNA 分子。重组质粒 pACYC-Duet-1+CRISPR/Cas12f 表达 SEQ ID NO: 1 所示的 Cas12f.4 蛋白和 SEQ ID NO: 25 所示的 Cas12f 导向 RNA。

2.重组质粒 pACYC-Duet-1+CRISPR/Cas12f 中含有表达盒，该表达盒的核苷酸序列如 SEQ ID NO: 23 所示。SEQ ID NO: 23 所示的序列中，自 5’末端起第 1 至 44 位为 pLacZ 启动子的核苷酸序列，第 45 至 3326 位为 Cas12f.4 基因的核苷酸序列，第 3327 至 3412 位为终止子的核苷酸序列（用于终止转录）。自 5’末端起第 3413 至 3452 位为 J23119 启动子的核苷酸序列，第 3453 至 3,628 位为 CRISPR 阵列的核苷酸序列，第 3627 至 3713 位为 rrnB-T1 终止子的核苷酸序列（用于终止转录）。

3.重组大肠杆菌的获得：将重组质粒 pACYC-Duet-1+CRISPR/Cas12f 导入大肠杆菌 EC100 中，得到重组大肠杆菌，命名为 EC100/pACYC-Duet-1+CRISPR/Cas12f。将重组质粒 pACYC-Duet-1 导入大肠杆菌 EC100 中，得到重组大肠杆菌，命名为 EC100/pACYC-Duet-1。

4.PAM 文库的构建：人工合成 SEQ ID NO: 24 所示的序列，并连接到 pUC19 载体，其中 SEQ ID NO: 24 所示的序列包括 5’端八个随机碱基和靶序列。对 PAM 文库的靶标序列 5’端前面设计了 8 个随机碱基构建质粒文库。将质粒分别转入到含有 Cas12f.4 基因座的大肠杆菌中和不含有 Cas.12f.4 基因座的大肠杆菌中。在 37°C 下处理 1 小时后，我们对质粒进行提取，并对 PAM 区域序列进行 PCR 扩增和测序。

5.PAM 文库结构域的获得：分别统计实验组和对照组中 65,536 种组合的 PAM 序列出现次数，并用各自组所有的 PAM 序列数目进行标准化。对于任意一条 PAM 序列，当 log2(对照组标准化值/实验组标准化值)大于 3.5 时，我们认为这条 PAM 被显著消耗，我们总共得到了 3,548 条显著被消耗的 PAM 序列，所有占比例 5.41%。我们用 Weblogo 对显著消耗的 PAM 序列进行预测，发现 Cas12f.4 的 PAM 结构域为严格的 5’-TTN 结构

（图 2），而且靶标序列前面的第二位和第三位碱基几乎 100%都是 T，而其他位置可以为任意的序列，这是比已经报道的最严谨 PAM 识别的 C2c1 拥有更加严谨的 PAM 识别方式。

6.PAM 文库结构域的验证：通过 PAM 文库消耗实验，我们获得了 Cas12f.4 的 PAM 结构域，为了验证这一结构域的严谨性，我们设置了 10 组 PAM 进行了体内实验，测序 Cas12f 对这些 PAM 的编辑活性。首先，我们将 30nt 的靶标和 PAM 序列整合到质粒的耐卡那基因的非保守位置上，然后用 CRSPR/Cas12f 和导向 RNA 形成的复合物与之进行混合培养 8 小时。通过涂板并统计菌落数目，我们便能判断 Cas12f 对不同 PAM 序列的消耗活性。通过实验结果，我们可以看到 CRISPR/Cas12f.4 系统只能对带有 5'-TTA、5'-TTT、 5'-TTC 和 5'-TTG PAM 的靶标序列进行有效地编辑，而对带有 5'-TAT、5'-TCT、5'-TCG、5'-ATT、5'-CTT 和 5'-GTT PAM 的靶标序列则没有编辑活性，从而验证了 Cas12f.4 的 PAM 结构域识别的验证性。通过对不同 PAM 的菌落进行统计，我们发现 CRISPR/Cas12f.4 系统对 5'-TTA、5'-TTT 和 5'-TTC 的编辑活性高于 5'-TTG。

实施例 4. Cas12f.4，Cas12f.5，Cas12f.6 在人类细胞系中的切割

将含有 Cas12f.4 基因的真核表达载体和含有 U6 启动子与 crRNA（SEQ ID NO: 25）序列的 PCR 产物通过脂质体转染的方法转入到人类 HEK293T 细胞中（图 3a），在 37 摄氏度 5%二氧化碳浓度下培养 72h。提取全部细胞的 DNA，并对包含靶位点 700bp 的序列进行扩增，将 PCR 产物通过 Tn5 进行二代测序文库构建，测序由北京安诺优达基因科技有限公司完成，将测序结果比对到人类基因组的 VEGFA 基因上，鉴定到 Cas12f.4 对目标靶位点的切割方式（图 3b），CRISPR/Cas12f.4 系统对 VEGFA 的编辑效率能达到 4.2%，原始测序数据如图 3c 所示(图 3c)。

使用相同方法检测 Cas12f.5、Cas12f.6 对 VEGFA 的切割活性，其 crRNA 分别如 SEQ ID NO: 26、SEQ ID NO: 27 所示。图 3c 的结果显示，CRISPR/Cas12f.5、CRISPR/Cas12f.6 系统对 VEGFA 的编辑效率分别为 0.31%和 0.19%。

实施例 5. Cas12f.4 在玉米原生质体中的切割

将纯化的 Cas12f.4 蛋白（60μg）和 SEQ ID NO: 28 或 29 所示导向 RNA（120μg）在 37 摄氏度下混合成核糖核蛋白复合体(RNP)，再利用 PEG4000 介导的原生质体转化将 CRISPR/Cas12f.4 RNP 转入玉米原生质体细胞中，在 37 摄氏度下暗培养 24 小时（图

4a）。培养结束后，离心去上清收集原生质体，提取原生质体 DNA，扩增靶位点上下游约 600bp 的 DNA 片段。将包含靶位点的 DNA 片段进行 T7 核酸内切酶酶切检测，结果如图 4b 所示，CRISPR/Cas12f.4 系统对 PDI1、SEB2.2 具有高效的切割活性。将含有靶位点的 DNA 片段连接 Blunt Simple 载体，涂板，通过赛默飞世尔科技（中国）有限公司对单克隆进行 Sanger 测序，将测序结果比对到玉米组的 PDI1、SEB2.2 基因上，结果如图 4b-4c 所示，鉴定得到 Cas12f.4 对目标靶位点的切割效率分别为 33.5%和 16.7%。

尽管本发明的具体实施方式已经得到详细的描述，但本领域技术人员将理解：根据已经公布的所有教导，可以对细节进行各种修改和变动，并且这些改变均在本发明的保护范围之内。本发明的全部分为由所附权利要求及其任何等同物给出。

权 利 要 求

1. 一种蛋白，其具有 SEQ ID NO：1 所示的氨基酸序列，或具有与 SEQ ID NO: 1 相比具有至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、或至少 99%的序列同一性的氨基酸序列；

例如，所述蛋白是 CRISPR/Cas 系统中的效应蛋白。

2. 一种缀合物，其包含权利要求 1 所述的蛋白以及修饰部分。

3. 权利要求 2 所述的缀合物，其中，所述修饰部分选自另外的蛋白或多肽、可检测的标记，及其任意组合。

4. 权利要求 2 或 3 所述的缀合物，其中，所述修饰部分任选地通过接头连接至所述蛋白的 N 端或 C 端。

5. 权利要求 2-4 任一项所述的缀合物，其中，所述另外的蛋白或多肽选自表位标签、报告基因序列、核定位信号（NLS）序列、靶向部分、转录激活结构域（例如，VP64）、转录抑制结构域（例如，KRAB 结构域或 SID 结构域）、核酸酶结构域（例如，Fok1），具有选自下列的活性的结构域：核苷酸脱氨酶、甲基化酶活性，去甲基化酶，转录激活活性，转录抑制活性，转录释放因子活性，组蛋白修饰活性，核酸酶活性，单链 RNA 切割活性，双链 RNA 切割活性，单链 DNA 切割活性，双链 DNA 切割活性和核酸结合活性；以及其任意组合。

6. 权利要求 2-5 任一项所述的缀合物，其中，所述缀合物包含表位标签。

7. 权利要求 2-6 任一项所述的缀合物，其中，所述缀合物包含 NLS 序列；

例如，所述 NLS 序列如 SEQ ID NO: 19 所示；

例如，所述 NLS 序列位于、靠近或接近所述蛋白的末端（例如，N 端或 C 端）。

8. 一种融合蛋白，其包含权利要求 1 所述的蛋白以及另外的蛋白或多肽。

9. 权利要求 8 所述的融合蛋白，其中，所述另外的蛋白或多肽任选地通过接头连接至所述蛋白的 N 端或 C 端。

10. 权利要求 8 或 9 所述的融合蛋白，其中，所述另外的蛋白或多肽选自表位标签、报告基因序列、核定位信号（NLS）序列、靶向部分、转录激活结构域（例如，VP64）、转录抑制结构域（例如，KRAB 结构域或 SID 结构域）、核酸酶结构域（例如，Fok1），具有选自下列的活性的结构域：核苷酸脱氨酶、甲基化酶活性，去甲基化酶，转录激活活性，转录抑制活性，转录释放因子活性，组蛋白修饰活性，核酸酶活性，单链 RNA 切割活性，双链 RNA 切割活性，单链 DNA 切割活性，双链 DNA 切割活性和核酸结合活性；以及其任意组合。

11. 权利要求 8-10 任一项所述的融合蛋白，其中，所述融合蛋白包含表位标签。

12. 权利要求 8-11 任一项所述的融合蛋白，其中，所述融合蛋白包含 NLS 序列；

例如，所述 NLS 序列如 SEQ ID NO: 19 所示；

例如，所述 NLS 序列位于、靠近或接近所述蛋白的末端（例如，N 端或 C 端）。

13. 权利要求 8-12 任一项所述的融合蛋白，其中，所述融合蛋白具有 SEQ ID NO: 20 所示的氨基酸序列。

14. 一种分离的核酸分子，其包含选自下列的序列，或由选自下列的序列组成：

(i) SEQ ID NO：7 或 13 所示的序列；

(ii) 与 SEQ ID NO：7 或 13 所示的序列相比具有一个或多个碱基的置换、缺失或添加（例如 1 个，2 个，3 个，4 个，5 个，6 个，7 个，8 个，9 个或 10 个碱基的置换、缺失或添加）的序列；

(iii) 与 SEQ ID NO：7 或 13 所示的序列具有至少 20%、至少 30%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 90%、至少 95%的序列同一性的序列；

(iv) 在严格条件下与(i)-(iii)任一项中所述的序列杂交的序列；或

(v) (i)-(iii)任一项中所述的序列的互补序列；

并且，(ii)-(v)中任一项所述的序列基本保留了其所源自的序列的生物学功能；

例如，所述分离的核酸分子是 RNA；

例如，所述分离的核酸分子是 CRISPR/Cas 系统中的同向重复序列。

15. 权利要求 14 所述的分离的核酸分子，其中，所述核酸分子包含一个或多个茎环或优化的二级结构；

例如，(ii)-(v)中任一项所述的序列保留了其所源自的序列的二级结构。

16. 权利要求 14 或 15 所述的分离的核酸分子，其中，所述核酸分子包含选自下列的序列，或由选自下列的序列组成：

(a) SEQ ID NO：7 或 13 所示的核苷酸序列；

(b) 在严格条件下与(a)中所述的序列杂交的序列；或

(c) SEQ ID NO：7 或 13 所示的核苷酸序列的互补序列。

17. 一种复合物，其包含：

(i) 蛋白组分，其选自：权利要求 1 所述的蛋白、权利要求 2-7 任一项所述的缀合物、权利要求 8-13 任一项所述的融合蛋白，及其任意组合；和

(ii) 核酸组分，其从 5’至 3’方向包含权利要求 14-16 任一项所述的分离的核酸分子和能够与靶序列杂交的导向序列，

其中，所述蛋白组分与核酸组分相互结合形成复合物；

例如，所述核酸组分是 CRISPR/Cas 系统中的导向 RNA；

例如，所述核酸分子是 RNA；

例如，所述复合物不包含反式作用 crRNA(tracrRNA)。

18. 权利要求 17 所述的复合物，其中，所述导向序列连接于所述核酸分子的 3’端。

19. 权利要求 17 或 18 所述的复合物，其中，所述导向序列包含所述靶序列的互补序列。

20. 一种分离的核酸分子，其包含：

(i) 编码权利要求 **1** 所述的蛋白，或权利要求 **8-13** 任一项所述的融合蛋白的核苷酸序列；

(ii) 编码权利要求 **14-16** 任一项所述的分离的核酸分子的核苷酸序列；和/或，

(iii) 包含**(i)**和**(ii)**的核苷酸序列；

例如，**(i)-(iii)**任一项中所述的核苷酸序列经密码子优化用于在原核细胞或真核细胞中进行表达。

21. 一种载体，其包含权利要求 **20** 所述的分离的核酸分子。

22. 一种宿主细胞，其包含权利要求 **20** 所述的分离的核酸分子或权利要求 **21** 所述的载体。

23. 一种组合物，其包含：

(i) 第一组分，其选自：权利要求 **1** 所述的蛋白、权利要求 **2-7** 任一项所述的缀合物、权利要求 **8-13** 任一项所述的融合蛋白、编码所述蛋白或融合蛋白的核苷酸序列，以及其任意组合；和

(ii) 第二组分，其为包含导向 **RNA** 的核苷酸序列，或者编码所述包含导向 **RNA** 的核苷酸序列的核苷酸序列；

其中，所述导向 **RNA** 从 **5'**至 **3'**方向包含同向重复序列和导向序列，所述导向序列能够与靶序列杂交；

所述导向 **RNA** 能够与**(i)**中所述的蛋白、缀合物或融合蛋白形成复合物；

所述同向重复序列是权利要求 **14-16** 任一项中所定义的分离的核酸分子；

例如，所述组合物不包含反式作用 **crRNA(tracrRNA)**。

24. 一种组合物，其包含一种或多种载体，所述一种或多种载体包含：

(i) 第一核酸，其为编码权利要求 **1** 所述的蛋白或权利要求 **8-13** 任一项所述的融合蛋白的核苷酸序列；任选地所述第一核酸可操作地连接至第一调节元件；以及

(ii) 第二核酸，其编码包含导向 **RNA** 的核苷酸序列；任选地所述第二核酸可操作地连接至第二调节元件；

其中：

所述第一核酸与第二核酸存在于相同或不同的载体上；

所述导向 RNA 从 5’至 3’方向包含同向重复序列和导向序列，所述导向序列能够与靶序列杂交；

所述导向 RNA 能够与(i)中所述的效应蛋白或融合蛋白形成复合物；

所述同向重复序列是权利要求 14-16 任一项中所定义的分离的核酸分子；

例如，所述组合物不包含反式作用 crRNA(tracrRNA)。

25. 权利要求 24 所述的组合物，其中，所述第一调节元件和/或第二调节元件是启动子，例如诱导型启动子。

26. 权利要求 23-25 任一项所述的组合物，其中，所述组合物中的至少一个组分是非天然存在的或经修饰的。

27. 权利要求 23-26 任一项所述的组合物，其中，所述导向序列连接至所述同向重复序列的 3’端。

28. 权利要求 23-27 任一项所述的组合物，其中，所述导向序列包含所述靶序列的互补序列。

29. 权利要求 23-28 任一项所述的组合物，其中，当所述靶序列为 DNA 时，所述靶序列位于原间隔序列临近基序(PAM)的 3’端，并且所述 PAM 具有 5’-TTN 所示的序列，其中，N 选自 A、G、T、C；当所述靶序列为 RNA 时，所述靶序列不具有 PAM 结构域限制。

30. 权利要求 23-29 任一项所述的组合物，其中，所述靶序列是来自原核细胞或真核细胞的 DNA 或 RNA 序列；或者，所述靶序列是非天然存在的 DNA 或 RNA 序列。

31. 权利要求 23-30 任一项所述的组合物，其中，所述靶序列存在于细胞内；

例如，所述靶序列存在于细胞核内或细胞质（例如，细胞器）内；

例如，所述细胞是真核细胞；

例如，所述细胞是原核细胞。

32. 权利要求 **23-31** 任一项所述的组合物，其中，所述蛋白连接有一个或多个 **NLS** 序列，或者，所述缀合物或融合蛋白包含一个或多个 **NLS** 序列；

例如，所述 **NLS** 序列连接至所述蛋白的 **N** 端或 **C** 端；

例如，所述 **NLS** 序列融合至所述蛋白的 **N** 端或 **C** 端。

33. 一种试剂盒，其包括一种或多种选自下列的组分：权利要求 **1** 所述的蛋白、权利要求 **2-7** 任一项所述的缀合物、权利要求 **8-13** 任一项所述的融合蛋白、权利要求 **14-16** 任一项所述的分离的核酸分子、权利要求 **17-19** 任一项所述的复合物、权利要求 **20** 所述的分离的核酸分子、权利要求 **21** 所述的载体、权利要求 **23-32** 任一项所述的组合物；

例如，所述试剂盒包含权利要求 **23**、**26-32** 任一项所述的组合物，以及使用所述组合物的说明书；

例如，所述试剂盒包含权利要求 **24**、**25-32** 任一项所述的组合物，以及使用所述组合物的说明书。

34. 一种递送组合物，其包含递送载体，以及选自下列的一种或多种：权利要求 **1** 所述的蛋白、权利要求 **2-7** 任一项所述的缀合物、权利要求 **8-13** 任一项所述的融合蛋白、权利要求 **14-16** 任一项所述的分离的核酸分子、权利要求 **17-19** 任一项所述的复合物、权利要求 **20** 所述的分离的核酸分子、权利要求 **21** 所述的载体、权利要求 **23-32** 任一项所述的组合物；

例如，所述递送载体是粒子；

例如，所述递送载体选自脂质颗粒、糖颗粒、金属颗粒、蛋白颗粒、脂质体、外泌体、微泡、基因枪或病毒载体（例如，复制缺陷型逆转录病毒、慢病毒、腺病毒或腺相关病毒）。

35. 一种修饰靶基因的方法，其包括：将权利要求 **17-19** 任一项所述的复合物或权利要求 **23-32** 任一项所述的组合物与所述靶基因接触，或者递送至包含所述靶基因的细胞中；所述靶序列存在于所述靶基因中。

36. 权利要求 **35** 所述的方法，其中，所述靶基因存在于细胞内；

例如，所述细胞是原核细胞；

例如，所述细胞是真核细胞，例如哺乳动物细胞（例如人类细胞）或植物细胞。

37. 权利要求 **35** 所述的方法，其中，所述靶基因存在于体外的核酸分子（例如，质粒）中。

38. 权利要求 **35-37** 任一项所述的方法，其中，所述修饰是指所述靶序列的断裂，如 **DNA** 的双链断裂或 **RNA** 的单链断裂；

例如，所述修饰还包括将外源核酸插入所述断裂中。

39. 一种改变基因产物的表达的方法，其包括：将权利要求 **17-19** 任一项所述的复合物或权利要求 **23-32** 任一项所述的组合物与编码所述基因产物的核酸分子接触，或者递送至包含所述核酸分子的细胞中，所述靶序列存在于所述核酸分子中。

40. 权利要求 **39** 所述的方法，其中，所述核酸分子存在于细胞内；

例如，所述细胞是原核细胞；

例如，所述细胞是真核细胞，例如哺乳动物细胞（例如人类细胞）或植物细胞。

41. 权利要求 **39** 所述的方法，其中，所述核酸分子存在于体外的核酸分子（例如，质粒）中。

42. 权利要求 **39-41** 任一项所述的方法，其中，所述基因产物的表达被改变（例如，增强或降低）。

43. 权利要求 **39-42** 任一项所述的方法，其中，所述基因产物是蛋白。

44. 权利要求 **35-43** 任一项所述的方法，其中所述的蛋白、缀合物、融合蛋白、分离的核酸分子、复合物、载体或组合物包含于递送载体中；

例如，所述递送载体选自脂质颗粒、糖颗粒、金属颗粒、蛋白颗粒、脂质体、外泌

体、病毒载体（如复制缺陷型逆转录病毒、慢病毒、腺病毒或腺相关病毒）。

45. 权利要求 **35-44** 任一项所述的方法，其用于改变靶基因或编码靶基因产物的核酸分子中的一个或多个靶序列来修饰细胞、细胞系或生物体。

46. 一种由权利要求 **35-45** 任一项所述的方法获得的细胞或其子代，其中所述细胞包含在其野生型中不存在的修饰。

47. 权利要求 **46** 所述的细胞或其子代的细胞产物。

48. 一种体外的、离体的或体内的细胞或细胞系或它们的子代，所述细胞或细胞系或它们的子代包含：权利要求 **1** 所述的蛋白、权利要求 **2-7** 任一项所述的缀合物、权利要求 **8-13** 任一项所述的融合蛋白、权利要求 **14-16** 任一项所述的分离的核酸分子、权利要求 **17-19** 任一项所述的复合物、权利要求 **20** 所述的分离的核酸分子、权利要求 **21** 所述的载体、权利要求 **23-32** 任一项所述的组合物；

例如，所述细胞是真核细胞；

例如，所述细胞是动物细胞（例如，哺乳动物细胞，例如人类细胞）或植物细胞；

例如，所述细胞是干细胞或干细胞系。

49. 权利要求 **1** 所述的蛋白、权利要求 **2-7** 任一项所述的缀合物、权利要求 **8-13** 任一项所述的融合蛋白、权利要求 **14-16** 任一项所述的分离的核酸分子、权利要求 **17-19** 任一项所述的复合物、权利要求 **20** 所述的分离的核酸分子、权利要求 **21** 所述的载体、权利要求 **23-32** 任一项所述的组合物或权利要求 **33** 所述的试剂盒，用于核酸编辑（例如，基因或基因组编辑）的用途；

例如，所述基因或基因组编辑包括修饰基因、敲除基因、改变基因产物的表达、修复突变、和/或插入多核苷酸。

50. 权利要求 **1** 所述的蛋白、权利要求 **2-7** 任一项所述的缀合物、权利要求 **8-13** 任一项所述的融合蛋白、权利要求 **14-16** 任一项所述的分离的核酸分子、权利要求 **17-19** 任一项所述的复合物、权利要求 **20** 所述的分离的核酸分子、权利要求 **21** 所述的载体、权

利要求 **23-32** 任一项所述的组合物或权利要求 **33** 所述的试剂盒，在制备制剂中的用途，所述制剂用于：

(i) 离体基因或基因组编辑；

(ii) 离体单链 **DNA** 的检测；

(iii) 编辑靶基因座中的靶序列来修饰生物或非人类生物；

(iv) 治疗由靶基因座中的靶序列的缺陷引起的病症。