ACS Synthetic Biology

Subscriber access provided by SHANGHAI DOC & INFO CENTER

Letter

Efficient Transcriptional Gene Repression by Type V-A CRISPR-Cpf1 from *Eubacterium eligens*

Seong Keun Kim, Haseong Kim, Woo-Chan Ahn, Kwang-Hyun Park, Euijeon Woo, Dae-Hee Lee, and Seung-Goo Lee

ACS Synth. Biol., **Just Accepted Manuscript** • Publication Date (Web): 04 Apr 2017

Downloaded from http://pubs.acs.org on April 5, 2017

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications

ACS Synthetic Biology is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Efficient Transcriptional Gene Repression by Type V-A CRISPR-Cpf1 from *Eubacterium eligens*

Seong Keun Kim,[†,§,#] Haseong Kim,[†,§,#] Woo-Chan Ahn,[‡] Kwang-Hyun Park,[‡] Eui-Jeon Woo,[‡,¶] Dae-Hee Lee,[†,§,*] and Seung-Goo Lee,[†,§,*]

[†]Synthetic Biology and Bioengineering Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic of Korea

[§]Biosystems and Bioengineering Program, University of Science and Technology (UST), Daejeon 34113, Republic of Korea

[‡]Disease Target Structure Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Republic of Korea

[¶]Bio-Analytical Science Program, University of Science and Technology (UST), Daejeon 34113, Republic of Korea

[*]Corresponding Authors

[#]These authors equally contributed to this work.

ABSTRACT

Clustered regularly interspaced short palindromic repeats interference (CRISPRi) is an emerging technology for artificial gene regulation. Type II CRISPR-Cas endonuclease Cas9 is the most widely used protein for gene regulation with CRISPRi. Here, we present type V-A CRISPR-Cas endonuclease Cpf1-based CRISPRi. We constructed an L-rhamnose-inducible CRISPRi system with DNase-deactivated Cpf1 from *Eubacterium eligens* (EedCpf1) and compared its performance with catalytically deactivated Cas9 from *Streptococcus pyogenes* (SpdCas9). In contrast to SpdCas9, EedCpf1 showed stronger gene repression when it was targeted to the template strand than when it was targeted to the non-template strand of the 5′ untranslated region or coding DNA sequences. EedCpf1 exhibited no strand bias when targeted to the promoter, and preferentially used the 5′-TTTV-3′ (V= A, G, or C) protospacer adjacent motif. Multiplex repression of the EedCpf1-based CRISPRi system was demonstrated using episomal and chromosomal gene targets. Our findings will guide an efficient EedCpf1-mediated CRISPRi genetic control.

KEYWORDS: CRISPRi, deactivated Cpf1, deactivated Cas9, *Eubacterium eligens*, protospacer adjacent motif, *Streptococcus pyogenes*

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins form an adaptive immune system in eubacteria and archaea[1]. They have been repurposed for targeted genome editing in humans and other organisms[1-7]. To repurpose the CRISPR-Cas system for gene regulation instead of genome editing, CRISPR interference (CRISPRi) using a catalytically inactive Cas9 protein has been developed and used as an exceptionally efficient tool for sequence-specific regulation of gene expression in various organisms[8]. Catalytically deactivated Cas9 of *Streptococcus pyogenes* (SpdCas9) derived from a type II CRISPR system is the best studied and most widely used protein in CRISPRi[8-12]. The only requirements for this CRISPRi system are the SpdCas9 protein and a single guide RNA (sgRNA). The SpdCas9-sgRNA complex binds to a non-template strand of target DNA, which blocks transcription by RNA polymerase (RNAP). A multimeric CRISPRi system derived from a type I CRISPR system has been also reported; it requires a deletion of a Cas3 protein that is involved in the cleavage and degradation of target DNA[13, 14]. These CRISPRi systems derived from type I and II CRISPR systems allow efficient, reversible, and multiplexible repression of gene transcription.

Recently, type V-A CRISPR systems have been identified and introduced as targeted genome editing tools for human cells[15, 16]; they are composed of a single Cpf1 (CRISPR from *Prevotella* and *Francisella* 1) protein and its cognate CRISPR RNA (crRNA) and do not require an additional trans-activating crRNA (tracrRNA)[17, 18]. In contrast to Cas9, which identifies guanidine-rich protospacer adjacent motif (PAM) sequences downstream of the target region, Cpf1 recognizes thymidine-rich PAM sequences upstream of the target region and cleaves target DNA, generating staggered ends[17]. Even though type V-A CRISPR-Cpf1 is an attractive alternative for the Cas9-based genome engineering tool, the DNA-binding activity of Cpf1 in terms of PAM sequence diversity has been characterized only in catalytically deactivated Cpf1 of *Francisella novicida* U112 (FndCpf1)[19]. FndCpf1 has been

used for CRISPRi to measure the extent of gene repression by targeting the *lacZ* promoter upstream of a green fluorescent protein (*gfp*) gene, demonstrating that FndCpf1 can be readily repurposed for programmable gene regulation[19]. Recently, the diversity of Cpf1 family proteins was explored by searching public sequence databases. Among 46 non-redundant Cpf1 family proteins found, 16 Cpf1 candidate proteins were selected for PAM sequence determination and functional analysis. However, only eight new Cpf1 family members, from *F. novicida* U112, *Prevotella disiens*, *Acidaminococcus* sp. BV3L6, *Lachnospiraceae bacterium* ND2006, *Lachnospiraceae bacterium* MA2020, *Candidatus* Methanoplasma termitum, *Moraxella bovoculi* 237, and *Porphyromonas crevioricanis*, showed efficient cleavage of target DNA with identified PAM sequences[17]. Here, we report a tunable CRISPRi system for efficient gene regulation using a novel nuclease-deactivated Cpf1 from *Eubacterium eligens* (EedCpf1) and a designed crRNA.

To explore the feasibility of the dCpf1-based CRISPRi system for gene expression regulation, we first generated a deactivated EedCpf1 by introducing a mutation into wild-type (WT) EeCpf1 in a key amino acid involved in DNase activity. A recent study on FnCpf1 indicated that Cpf1 proteins have an RuvC-like endonuclease domain similar to that of Cas9 and harboring at least three essential catalytic residues (D917, E1006, and D1255 in FnCpf1)[17]. Based on amino acid sequence alignment of FnCpf1 and EeCpf1, we created a mutation in D880A, one of the three essential catalytic residues (D880, E965, and D1233) in EeCpf1, to produce an EedCpf1 (Figure S1). We found that the D880A mutation completely deactivated the DNA cleavage activity of EeCpf1. WT EeCpf1 can cleave supercoiled target DNA in the presence of Mn^{2+} in a crRNA-dependent manner, while the EedCpf1 has no nuclease activity, indicating that the RuvC-like domain of EeCpf1 cleaves both strands of the target DNA (Figure 1). This result is contrast to the mutation studies of SpCas9 because deactivation of each of the RuvC and HNH domains abolished its ability to cleave one of the

DNA double strands[1]. Furthermore, FnCpf1 was reported to process pre-crRNA into crRNA, a function independent from the DNA nuclease activity[18]. This previous report identified four residues (H843, K852, K869, and F873) of FnCpf1 essential to pre-crRNA processing, which correspond to H765, K774, K833, and Y837 in EeCpf1 (Figure S1).

After the EedCpf1 protein was created, we constructed a pSECRVi plasmid using a pSECRi plasmid that was previously generated for a SpdCas9-mediated CRISPRi system[11]. Plasmid pSECRVi encodes L-rhamnose-inducible EedCpf1 protein and constitutive BioBrick J23119 promoter-driven crRNA cassettes (Figure 2A). To design a specific crRNA cassette, we used a 5′-mature repeat sequence deduced from nine crRNA sequences of *E. eligens* (Figure S2A), a 20-nt spacer, a 3′-repeat sequence, and a terminator, resulting in crRNAR(T1) (Figure S2B). Following transcription, the crRNA is further processed by the RNase activity of EedCpf1, yielding native crRNA[18]. To examine the effect of redundant 334-bp sequences at the 3′-end of the spacer in crRNAR(T1) on repression efficiency, we synthesized another crRNA cassette of EedCpf1, i.e., crRNA(T1), which lacked the 3′-repeat sequence of crRNAR(T1) (Figure S2B). To compare the CRISPRi efficiencies of SpdCas9 and EedCpf1 in *Escherichia coli* DH5α, we used the pSECRi plasmid (Figure 2B) and a constructed reporter plasmid, pREGFP3(NT1) (Figure 2C). The latter harbors a *gfp* gene under the control of a constitutive BioBrick J23100 promoter, a 20-nt sequence complementary to the spacer sequence (5′-GCGTTGTGCCGATTCTGGTG-3′), and PAM sequences (5′-CCG-3′ for SpdCas9; 5′-GAAAA-3′ for EedCpf1) on the non-template strand. Previously, *in vitro* PAM identification assay using eight Cpf1 orthologs revealed that the PAM sequences of Cpf1 family proteins are predominantly T-rich and varied only in the number of thymidines constituting each PAM[17]. Among them, *Candidatus* Methanoplasma termitum Cpf1, which is the closest ortholog to EeCpf1, has a 5′-TTTTA-3′ PAM sequence.

Expression of SpdCas9 (from pSECRi) and EedCpf1 (from pSECRVi) was induced by 1 mM L-rhamnose. The expression ratios were calculated as $expression\ (\%) = \frac{RFU_{xv}/OD_{xv}}{RFU_{null}/OD_{null}} \times 100$, where RFU and OD are relative fluorescence units and optical density values at 600 nm, respectively. The subscript *xv* designates the tested cells harboring the pSECRi or pSECRVi plasmid (in the presence of L-rhamnose), whereas *null* indicates an empty-vector control (in the presence of L-rhamnose). Mean expression levels were compared using mainly two-tailed *t*-tests. CRISPR-SpdCas9 repressed *gfp* expression from pREGFP3(NT1) to approximately 2.8% of non-repressed levels (Figure S2C), which is comparable to the previously reported efficiency of SpdCas9-mediated CRISPRi[11]. Meanwhile, CRISPR-EedCpf1 induced no significant repression (*gfp* expression was approximately 80%). It was reported that targeting the template DNA strand with multimeric CRISPRi results in better repression than targeting the non-template strand[14], although contrasting results were observed in other studies[13, 20]. Therefore, we generated another reporter plasmid, pREGFP3(T1), by relocating the binding sequence of pREGFP3(NT1) on the template strand (Figure 2D). Indeed, when targeting the template strand, CRISPR-EedCpf1 reduced the expression of the GFP to 13.3%, a more pronounced repression than that observed for non-template-strand targeting (73.4%) (Figure 2E). This repression efficiency is comparable to that elicited by SpdCas9 targeting the template strand (13.8%). The strongest repression was achieved when SpdCas9 targeted the non-template strand at the 5′ untranslated region (UTR) (2.4%). Cassettes crRNA(T1) and crRNAR(T1) showed similar repression efficiencies (13.3% and 14% in crRNA(T1) and crRNAR(T1), respectively) (Figure 2E). In the above experiments exploring the ability of the CRISPR-EedCpf1 system for gene repression, we used the 20-nt guide sequence for the EedCpf1 crRNA. However, since the WT guide sequence for the Cpf1 family proteins is 25-nt long, we explored the length requirement of the guide sequence for gene repression with CRISPR-EedCpf1. We

found that the efficiency was not significantly different between 20 and 25-nt of guide sequences, but started to decrease with the guide sequences shorter than 20-nt (Figure 2F). Previous examination of the length requirement for the guide sequence of FnCpf1 revealed that it requires at least 16-nt of guide sequence to achieve detectable DNA-cleavage and a minimum of 18-nt of guide sequence to achieve DNA cleavage *in vitro*[17]. These requirements of EedCpf1 and FnCpf1 are similar to those demonstrated for SpdCas9, in which a minimum of 16–17-nt of spacer sequence is required for DNA cleavage[21, 22]. Overall, these results indicate that CRISPR-EedCpf1 can be employed as a highly specific tool for gene expression regulation. In addition, it is possible to generate chimeric crRNAs through fusion to the 3′ end of the spacer that can recruit other RNA-binding proteins to endow the system with a novel function, as the redundant 334-bp sequences at the 3′-end of spacer did not affect repression efficiency[22, 23]. Further, since oligonucleotide-mediated ligation cloning of spacer sequences lacking the 3′-repeat sequences using Type IIS restriction enzymes[12] is more cost-effective than cloning spacer sequences with the 3′-repeat sequences, we used the former method in subsequent experiments (Figure S3).

Next, we explored the effects of the binding strand and location bias of the EedCpf1-crRNA complex on the repression of gene expression using a single crRNA(T1) binding site. We constructed six additional reporter plasmids harboring a maltose-binding protein (MBP) with a C-terminal fusion with enhanced GFP (MBP-EGFP), and inserted the crRNA(T1) binding site in different coding regions of MBP-EGFP, either on the template (T1) (Figure 3A) or the non-template (NT1) DNA strand (Figure 3B). The in-frame 30-bp sequence, consisting of a 20-bp spacer (5′-CACCAGAATCGGCACAACGC-3′) sandwiched between 5′-ATTTTC-3′ for EedCpf1 and 5′-CGGT-3′ for SpdCas9, was inserted after three different codons [namely, methionine 1 (M1), alanine 206 (A206), and asparagine 372 (N372)] of the MBP-EGFP fusion protein. We chose A206, a known permissive site within MBP that allows

heterologous sequence insertions without adversely affecting protein function[24, 25]. In all cases, in agreement with the results of experiments targeting the 5′ UTR (Figure 2E), the repression of MBP-EGFP expression by EedCpf1-crRNA was more pronounced when targeting the template strand of the transcribed region than when targeting the non-template strand (Figure 3C). This was the exact opposite of repression by SpdCas9-crRNA (Figure 3D). These results indicate that targeting the 5′ UTR and coding regions yielded strong repression and showed consistent strand bias toward the template strand. Strand bias is also observed in type I and II CRISPR dCas9s in bacteria, which prefer the non-template strand for repression[8, 9]. In contrast to EedCpf1, SpdCas9 targeting the coding region on the non-template strand generally shows a stronger repression effect than that targeting the template strand. Stronger inhibition of RNAP when EedCpf1 is bound to the template strand can be explained by the fact that RNAP primarily needs access to the template strand for transcription elongation[13, 20]. Alternatively, this might be related to different conformations of the EedCpf1-crRNA-DNA and SpdCpf1-crRNA-DNA complexes[26]. The repression efficiency of EedCpf1-crRNA targeted to the template strand was slightly reduced when the binding location was further away from the transcription start site of the MBP-EGFP fusion (Figure 3C). The extremely low expression of MBP-EGFP encoded by the pMEGFP(NT1) plasmid did not allow for quantification. Overall, these results indicate that type V-A CRISPR-EedCpf1 can potentially block transcription elongation activity of RNAP by binding to the 5′ UTR or coding DNA sequence (CDS)[8].

Identification of PAM sequences that are preferentially used by EedCpf1 is indispensable for the design of guide crRNA sequences to enable versatile applications of this artificial repressor. Various Cpf1 family proteins recognize thymidine-rich PAM sequences[17]; however, EeCpf1 PAM sequences have not yet been identified. We designed and synthesized three different PAMs, 5′-CTTTC-3′, 5′-CCTTC-3′, and 5′-CCCTC-3′, by modifying the 5′-

TTTTC-3′ PAM sequence located upstream of the spacer in pREGFP3(T1). As shown in Figure 4A, the 5′-CTTTC-3′ PAM repressed *gfp* expression to a level (15.5%) similar to that induced by the 5′-TTTTC-3′ PAM (11.4%); the transcriptional repression activities of the 5′-CCTTC-3′ (66.1%) and 5′-CCCTC-3′ (88.3%) PAM sequences were lower. This indicated that minimum three thymidine nucleotides are required for efficient repression of gene transcription by the EedCpf1-CRISPRi system. Next, we further characterized the EedCpf1 PAM sequences using additionally designed 5′-NTTTC-3′, 5′-CNTTC-3′, and 5′-CTTTN-3′ PAM sequences (N, any nucleotide). As expected, the transcriptional repression activity of 5′-CNTTC-3′ PAM sequences was low, except for 5′-CTTTC-3′, which contains three thymidine nucleotides (Figure 4B). All 5′-NTTTC-3′ PAM sequences exhibited high repression efficiency, with less than 20% residual *gfp* expression (Figure 4C); 5′-CTTTT-3′ exhibited lower repression activity (40.9%) than other 5′-CTTTN-3′ PAM sequences (Figure 4D). From these results, we conclude that 5′-TTTV-3′ (V=A, G, or C) PAM sequences are preferred by EedCpf1-CRISPRi. However, care should be taken in using the 5′-TTTT-3′ PAM sequence as TTTV is the favorable PAM sequence used by Cpf1 proteins from *Acidaminococcus* sp. BV3L6 and *L. bacterium* ND2006 in mammalian cells[27], which is in line with *in vitro* results[17].

Since multimeric CRISPRi systems preferentially target the promoter region[13, 14], we compared the repression efficiencies of EedCpf1-CRISPRi targeted to the promoter, 5′ UTR, and CDS. Ten pSECRVi plasmids were constructed using the single-stranded DNA oligonucleotide-mediated DNA assembly method (Figure S3) to target the BioBrick J23100 promoter (P1 and P2) and *gfp* CDS (C1, C3, C4, C5 targeting the template strand and C2, C6, C7, C8 targeting the non-template strand) in addition to the predesigned pSECRVi plasmid targeting the 5′ UTR (T1) (Figure 5A). As anticipated, all EedCpf1 constructs targeting the

template strand effectively repressed *gfp* expression from pREGFP3(P2T1) (Figure 5B). A non-template strand targeting the promoter region was also effective in blocking transcription initiation (P2: 7.6%), on a par with a multimeric CRISPRi system derived from type I CRISPR[14]. Consistent with our previous results, EedCpf1 targeting the non-template strand in the *gfp* CDS hardly repressed expression (C2: 98.2%; C6: 96.5%; C7: 98.1%; C8: 92.9%). Further, repression of *gfp* expression was slightly reduced with increasing distance of the binding site from the promoter (C1, covering the ribosome-binding site and ATG start codon: 9.3% expression; C3, 138 bp downstream from the start codon: 12.2% expression; C4, 295 bp downstream from the start codon: 21.3% expression; C5, 618 bp downstream from the start codon, 18% expression). Because of the relatively short *gfp* sequence (714 bp), CRISPRi near the C-terminal region (C5) also considerably repressed *gfp* expression. Based on these results, and considering the difficulty of promoter identification along with low occurrence of the 5′-TTTV-3′ PAM sequence in promoter regions, we recommend that a target sequence for EedCpf1-mediated CRISPRi should be designed proximal to the translation start site in the CDS, targeting the template strand.

Next, we examined the tunability of EedCpf1-regulated gene expression by the L-rhamnose inducer. *E. coli* cells harboring the reporter plasmid pREGFP3(P2T1) and the pSECRVi(C1) plasmid targeting C1 in the *gfp* CDS were incubated in Luria-Bertani (LB) medium containing 1 mM L-rhamnose to induce EedCpf1 expression. The pre-induced cells were diluted (1:99) in LB medium containing various L-rhamnose concentrations (0–1000 μM). After 200 min of cultivation, cell fluorescence was inversely proportional to the final L-rhamnose concentration (Figure 5C). At the end of the experiment (after 750 min), the *gfp* expression levels were as follows: 15% (1000 μM L-rhamnose), 37% (250 μM L-rhamnose), 66% (64 μM L-rhamnose), 91% (16 μM L-rhamnose), and 99% (4 μM L-rhamnose), of fluorescence produced by *E. coli* cells in the absence of L-rhamnose, which indicated that

higher concentrations of L-rhamnose result in stronger inhibition of *gfp* expression. Cell growth under all these conditions was virtually identical (Figure 5D). These results indicated that EedCpf1 can be used to tune gene expression over a broad range, enabling the control of cell growth or metabolite yields by targeting essential or toxic genes. In the presented CRISPRi system, we used an L-rhamnose-inducible promoter with RhaS and RhaR regulators for orthogonal control of the transcription of the *EedCpf1* gene. The L-rhamnose-inducible promoter is capable of homogeneous and rheostatic transcriptional control of heterologous genes and shows undetectable background expression in the absence of L-rhamnose[28, 29]. However, it is important to note that the homogeneous expression of the L-rhamnose-inducible promoter must be confirmed on a case-by-case basis because this promoter yielded a bistable response to L-rhamnose in certain experimental conditions[30].

Finally, we evaluated the ability of the EedCpf1-mediated CRISPRi system to regulate the expression of a chromosomally integrated reporter gene. A chloramphenicol resistance gene was incorporated into a reporter cassette, pREGFP3(P2T1), to allow reporter strain selection. The reporter cassette was inserted into the *bglA* genomic locus of *E. coli* DH5α using λ Red-mediated homologous recombination[31]. Similarly to the results obtained with episomal plasmid reporters, targeting of the EedCpf1-crRNA complex to the promoter regions of *gfp* gene resulted in efficient gene repression (P1: 3.2%; P2: 4.3%) irrespective of the binding DNA strand (Figure 6A, right). When the template strand of *gfp* CDS was targeted, the C1 binding site yielded the highest *gfp* repression (2.0% expression). The repression efficiency gradually decreased with increasing distance from the translation start site (C3: 6.7%; C4: 16.5%; C5: 23%), which is comparable with the results obtained using the MBP-EGFP fusion protein. Non-template strand targeting the *gfp* CDS resulted in almost no repression (C2: 93.7%), as anticipated. Further, single-cell fluorescence analysis revealed that gene repression using EedCpf1-CRISPRi generated homogeneous single-cell populations

without the all-or-none expression phenotype (Figure 6A, left). Considering the low repression efficiency of the C4 and C5 binding sites, multiplex repression with a newly designed chimeric crRNA(C4C5) (Figure 6B) was tested and found to be more effective than single-site targeting of C4 or C5 (Figure 6C), proving the applicability of EedCpf1 in a multiplex repression approach. To test the general applicability of EedCpf1-based multiplex repression, we designed two crRNAs to repress different genes, *lacZ* and *gfp*. A crRNA(lacZ) is targeting endogenous *lacZ* gene of *E. coli* K-12 MG1655 whereas crRNA(C1lacZ) is designed for repressing exogenous *gfp* gene of pREGFP3(P2T1) plasmid and endogenous *lacZ* gene of MG1655. As expected, crRNA(C1) and crRNA(C1lacZ) effectively repressed *gfp* expression from pREGFP3(P2T1) whereas crRNA(lacZ) hardly repressed *gfp* expression (Figure 6D, left). In case of *lacZ* repression, crRNA(lacZ) and crRNA(C1lacZ) strongly repressed *lacZ* expression whereas crRNA(C1) did not repressed *lacZ* expression in *E. coli* K-12 MG1655 that was grown in LB solid medium containing 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 80 µg/ mL of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (Figure 6D, right). Therefore, crRNA(C1lacZ) simultaneously repressed the plasmid-borne *gfp* and chromosomal *lacZ* in *E. coli* K-12 MG1655, suggesting that CRISPR-EedCpf1 could enable simultaneous control of multiple genes. Very recently, CRISPR-Cpf1 system was used for multiplexed genome editing using a single crRNA array, which edited up to four genes in mammalian cells and three in the mouse brain, simultaneously[32].

In this study, the binding strand bias of the CRISPRi system employing the type V-A CRISPR EedCpf1 protein, PAM sequence preference, and the tunability of episomal and chromosomal target gene expression (multiplex targeting) were explored. To be effective, EedCpf1 requires a target binding site on the template strand, within the 5′ UTR or CDS; 5′-TTTV-3′ (V=A, G, or C) PAM sequences are preferred. This knowledge will be also useful

for genome editing with the EeCpf1-based CRISPR system. Functional expression of EedCpf1 in *E. coli* quantitatively repressed transcription of a plasmid- or chromosome-encoded target gene. In terms of genetic circuits in synthetic biology, the CRISPRi system can be used as an actuator, which, together with metabolite-responsive sensors[33, 34], might be integrated into a powerful intelligent genetic circuit. Such sensor-actuator circuits have already been harnessed as promising genetic tools for generating intelligent cells for biotechnological and medical applications[35-37]. We anticipate that our findings will inform the design of a CRISPR-EedCpf1 system as a ubiquitous genetic actuator.

METHODS

Bacterial Strains, Media, and Reagents. *E. coli* DH5α was used for cloning and plasmid maintenance. LB medium (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L sodium chloride) was used for bacterial cultivation. SOC medium (20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L sodium chloride, 2.4 g/L magnesium sulfate, 186 mg/L potassium chloride, and 4 g/L glucose) was used as a recovery medium after cell transformation. L-Rhamnose and antibiotics were purchased from Sigma-Aldrich (St. Louis, MO). Ampicillin, kanamycin, and chloramphenicol were used at final concentrations of 100 μg/mL, 25 μg/mL, and 10 μg/mL, respectively. For polymerase chain reaction (PCR), high fidelity KOD-Plus-Neo polymerase (Toyobo, Osaka, Japan) was used following a standard protocol. All restriction and modification enzymes were purchased from New England BioLabs (NEB; Ipswich, MA).

Plasmid Construction. Primers, plasmids, and crRNAs used in this study are listed in Tables S1, S2, and S3, respectively. The pSECRi(T1) plasmid was constructed using a previously reported inverse PCR method[38]. Briefly, CRI(T1)-F and CRI(T1)-R primers were used with pSECRi plasmid template to alter the sgRNA region of the SpdCas9-based CRISPRi system. After PCR amplification and agarose gel electrophoresis, a DNA fragment

of the correct size was purified from the gel using Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, WI) and treated with *Dpn*I. The digestion product was ligated using T4 DNA ligase and T4 polynucleotide kinase, as per manufacturer's instructions.

To construct the pSECRVi plasmid, first, the *SpdCas9* gene in pSECRi was replaced with the *EedCpf1* gene. To this end, *EedCpf1* was amplified using dCpf1-IF and dCpf1-IR primers from pET22b-EedCpf1 plasmid. The fragment comprised a DNase-inactive Cpf1(D880A) CDS from *E. eligens*. The vector backbone was amplified using dCpf1-VF and dCpf1-VR primers and pSECRi. The two fragments were assembled by the Gibson assembly method as per manufacturer's instructions (NEB), resulting in pSECRi-EedCpf1 plasmid. The crRNA expression cassettes were generated as depicted in Figure S2B using the inverse PCR method. The entire region was amplified from pG-sgRNA plasmid using crRNA-F and crRNA-R primers. After amplification and electrophoresis, a DNA fragment of the correct size was purified and treated with *Dpn*I. The digestion product was ligated using T4 DNA ligase and T4 polynucleotide kinase, resulting in pG-crRNA. Similarly, a pG-crRNA(T1) plasmid was constructed using crRNA(T1)-F and crRNA(T1)-R primers with pG-sgRNA, and pG-crRNAR(T1) was constructed using crRNAR(T1)-F and crRNAR(T1)-R primers with pG-crRNA(T1). Finally, the three crRNA cassettes were amplified using ST-F and ST-R primers from pG-crRNA, pG-crRNA(T1), or pG-crRNAR(T1). The amplified fragments were then individually assembled with a linear vector fragment generated by digestion of pSECRi-EedCpf1 with *Age*I/*Not*I, resulting in pSECRVi, pSECRVi(T1), and pSECRVRi(T1) plasmids.

To generate pREGFP3- and pMEGFP-based reporter plasmids, primer pairs listed in Table S1 were used in inverse PCR to replace the CRISPRi binding site in pREGFP3 or pMEGFP. To insert the various spacer sequences, single-stranded oligonucleotide-mediated assembly was used. pSECRVi was linearized with the type IIS restriction enzyme *Sap*I, and a

single-stranded oligonucleotide containing 20-bp spacer sequences was incorporated into the *Sap*I-digested plasmid using NEBuilder HiFi DNA Assembly Master Mix as per the manufacturer's instructions. To construct the pSECRVi(C4C5) plasmid, pSECRVi(C4) was linearized with *Sap*I and then, two Multi(C5)-F and Multi(C5)-R oligonucleotides were assembled into the digested pSECRVi(C4) plasmid using NEBuilder HiFi DNA Assembly Master Mix. The pSECRVi(C1LacZ) plasmid was also constructed by assembling *Sap*I-linearized pSECRVi(C1) with two Multi(LacZ)-F and Multi(LacZ)-R oligonucleotides using NEBuilder HiFi DNA Assembly Master Mix.

Reporter Strain Construction. For integration of the reporter cassette with *E. coli* DH5α chromosome, plasmid pREGFP3C(P2T1) was constructed, which contained a chloramphenicol resistance gene for reporter strain selection. To this end, the chloramphenicol resistance cassette was amplified from pKD3/I-*Sce*I plasmid using Cm-F and Cm-R primers. The amplified fragment was assembled using Gibson assembly with a linear fragment generated by *Hind*III digestion of pREGFP3(P2T1). Next, the *gfp* gene and chloramphenicol expression cassette were amplified with Int-F and Int-R primers from pREGFP3C(P2T1) and integrated into the *bglA* genomic locus of *E. coli* DH5α via λ Red-mediated homologous recombination[31].

Purification of EeCpf1 and EedCpf1. The gene encoding Cpf1 (WP_012739647.1) was amplified from the genomic DNA of *E. eligens* (ATCC 27750) by PCRs and ligated into a modified pET-22b(+) plasmid to produce the protein with a 6xHis-tag at the C-terminus. The resulting plasmid pET22b-EeCpf1 was transformed into *E. coli* BL21-CodonPlus(DE3)-RIL strain (Agilent Technologies). The transformant cells were cultured in LB medium containing ampicillin to an OD_{600} of 0.6 and induced by adding 1 mM IPTG and incubating at 18 °C for 20 h. The cells were collected by centrifugation (8,000 xg, 30 min), re-suspended in 400 mL of lysis buffer (30 mM Tris-HCl pH 7.5, 140 mM NaCl, 5 mM β-mercaptoethanol,

10% glycerol), and disrupted by sonication in an ice bath (VC-600 sonicator; Sonics & Materials, Newtown, CT). The supernatant was clarified by centrifugation (10,000 xg, 30 min, 4 °C) and the protein was purified using HisTrap HP, Heparin HP, and Superdex 200 pg columns (GE Healthcare Life Sciences) with an ÄKTA FPLC system (GE Healthcare Life Sciences, Chicago, IL) and the elution buffer (30 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol, 10% Glycerol). The EedCpf1 mutant containing the D880A substitution (pET22b-EedCpf1) was generated with the site-directed mutagenesis kit (Enzynomics) and purified in the same way as the wild-type protein.

***In Vitro* Nuclease Activity Assays**. Synthetic 37-mer crRNA; UAAUUUCUACUUUGUAGAUAAGUUCUGCUAUGUGGCG were synthesized (Integrated DNA Technologies). Target dsDNA of pUC19 was purchased (Enzynomics). The purified EeCpf1 or EedCpf1 (160 nM) and the crRNA (7.6 μM) were incubated at 37 °C for 5 min in reaction buffer (1x PBS) with 5 mM $MgSO_4$. The reaction was initiated by the addition of target dsDNA (10 nM) and incubated at 37 °C for 20 min and quenched by the addition of 6x DNA loading dye (Fermentas) before analysed on 1% agarose gel.

Fluorescence Assay. Single *E. coli* colonies harboring a reporter plasmid and a CRISPRi plasmid were individually inoculated into LB medium containing appropriate antibiotics and cultured at 37 °C, 200 rpm overnight. Then, the cultures were diluted (1:99) with fresh LB medium supplemented with appropriate antibiotics and cultured at 37 °C and 200 rpm for 8 h (for pREGFP3 reporter plasmid) or 12 h (for pMEGFP reporter plasmid and a chromosomal reporter strain). For the induction of EedCpf1 or SpdCas9 protein expression, the culture medium was supplemented with 1 mM L-rhamnose, unless specified otherwise. After cultivation, the cells were washed once with phosphate-buffered saline and resuspended in this buffer. Fluorescence and OD_{600} measurements were conducted with the Victor X multi-label plate reader (PerkinElmer, Waltham, MA) using black-walled 96-well

polystyrene plates; single-cell fluorescence analysis was performed using FACSCalibur (BD Bioscience, Franklin Lakes, NJ). For time-course monitoring of fluorescence, three single *E. coli* colonies harboring pREGFP3(P2T1) and pSECRVi(C1) were inoculated into LB medium containing appropriate antibiotics and 1 mM L-rhamnose; they were then cultured at 37 °C, 200 rpm, overnight. Then, the cultures were diluted (1:99) with fresh LB medium containing appropriate antibiotics and various concentrations of L-rhamnose in black-walled 96-well polystyrene plates. Cell growth and fluorescence were measured using an Infinite 200 PRO microplate reader (Tecan, Männedorf, Switzerland).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website. Additional tables and figures include primers, plasmids, sgRNA/crRNA binding sites, and maps of crRNA and pSECRVi.

ABBREVIATIONS

CDS, coding DNA sequence; CRISPRi, clustered regularly interspaced short palindromic repeats interference; crRNA, CRISPR RNA; EedCpf1, DNase-deactivated Cpf1 from *Eubacterium eligens*; GFP, green fluorescent protein; MBP, maltose-binding protein; PAM, protospacer adjacent motif; PCR, polymerase chain reaction; RNAP, RNA polymerase; sgRNA, single guide RNA; SpdCas9, nuclease deactivated Cas9 from *Streptococcus pyogenes*; 5′ UTR, 5′ untranslated region.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: dhlee@kribb.re.kr

*E-mail: sglee@kribb.re.kr

Author Contributions

SKK conducted most of the CRISPRi experiments, including plasmid construction and reporter assays. WA and KP constructed the plasmid expressing DNase-deactivated Cpf1 from *E. eligens* and conducted *in vitro* nuclease assay. SL and DL supervised the study, designed experiments, and analyzed and interpreted the results. SKK, HK, EW, SL, and DL wrote the manuscript. SKK and HK equally contributed to this work.

Conflict of Interest

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Victor D. Lorenzo for the kind donation of the pSEVA plasmids and members of the Synthetic Biology Laboratory in the Synthetic Biology and Bioengineering Center at KRIBB for their valuable comments and helpful discussions. This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) (Grant numbers: 20163030091540) funded by the Ministry of Trade, Industry and Energy (MOTIE) and C1 Gas Refinery Program through the National Research Foundation of Korea (NRF) (Grant number: NRF-2015M3D3A1A01064875) funded by the Ministry of Science, ICT & Future Planning (MSIP) of the Republic of Korea. This work is also supported by the KRIBB Research Initiative Program.

REFERENCES

1. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity, *Science 337*, 816-821.

2. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems, *Nature Biotechnology 31*, 233-239.

3. Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., Peterson, R. T., Yeh, J. R. J., and Joung, J. K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system, *Nature Biotechnology 31*, 227-229.

4. Cho, S. W., Kim, S., Kim, J. M., and Kim, J.-S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease, *Nature Biotechnology 31*, 230-232.

5. Kim, H., and Kim, J.-S. (2014) A guide to genome engineering with programmable nucleases, *Nature Reviews in Genetics 15*, 321-334.

6. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-guided human genome engineering via Cas9, *Science 339*, 823-826.

7. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems, *Science 339*, 819-823.

8. Qi, Lei S., Larson, Matthew H., Gilbert, Luke A., Doudna, Jennifer A., Weissman, Jonathan S., Arkin, Adam P., and Lim, Wendell A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression, *Cell 152*, 1173-1183.

9. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system, *Nucleic Acids Research 41*, 7429-7437.

10. Choudhary, E., Thakur, P., Pareek, M., and Agarwal, N. (2015) Gene silencing by CRISPR interference in mycobacteria, *Nature Communications 6*, 6267.

11. Kim, S. K., Han, G. H., Seong, W., Kim, H., Kim, S.-W., Lee, D.-H., and Lee, S.-G. (2016) CRISPR interference-guided balancing of a biosynthetic mevalonate pathway increases terpenoid production, *Metabolic Engineering 38*, 228-240.

12. Cress, B. F., Toparlak, Ö. D., Guleria, S., Lebovich, M., Stieglitz, J. T., Englaender, J. A., Jones, J. A., Linhardt, R. J., and Koffas, M. A. G. (2015) CRISPathBrick: Modular combinatorial assembly of type II-A CRISPR arrays for dCas9-mediated multiplex transcriptional repression in *E. coli*, *ACS Synthetic Biology 4*, 987-1000.

13. Luo, M. L., Mullis, A. S., Leenay, R. T., and Beisel, C. L. (2015) Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression, *Nucleic Acids Research 43*, 674-681.

14. Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P. R., and Lundgren, M. (2015) Efficient programmable gene silencing by Cascade, *Nucleic Acids Research 43*, 237-246.

15. Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S.-h., and Kim, J.-S. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells, *Nature Biotechnology 34*, 863-868.

16. Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., Welch, M. M., Lopez, J. M., McCaw, Z. R., Aryee, M. J., and Joung, J. K. (2016) Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells, *Nature Biotechnology 34*, 869-874.

17. Zetsche, B., Gootenberg, Jonathan S., Abudayyeh, Omar O., Slaymaker, Ian M., Makarova, Kira S., Essletzbichler, P., Volz, Sara E., Joung, J., van der Oost, J., Regev, A., Koonin, Eugene V., and Zhang, F. (2015) Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system, *Cell 163*, 759-771.

18. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA, *Nature 532*, 517-521.

19. Leenay, Ryan T., Maksimchuk, Kenneth R., Slotkowski, Rebecca A., Agrawal, Roma N., Gomaa, Ahmed A., Briner, Alexandra E., Barrangou, R., and Beisel, Chase L. (2016) Identifying and visualizing functional PAM diversity across CRISPR-Cas systems, *Molecular Cell 62*, 137-147.

20. Chang, Y., Su, T., Qi, Q., and Liang, Q. (2016) Easy regulation of metabolic flux in *Escherichia coli* using an endogenous type I-E CRISPR-Cas system, *Microbial Cell Factories 15*, 195.

21. Cencic, R., Miura, H., Malina, A., Robert, F., Ethier, S., Schmeing, T. M., Dostie, J., and Pelletier, J. (2014) Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage, *PLOS ONE 9*, e109213.

22. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., and Joung, J. K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs, *Nature Biotechnology 32*, 279-284.

23. Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., Nureki, O., and Zhang, F. (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, *Nature 517*, 583-588.

24. Betton, J.-M., Martineau, P., Saurin, W., and Hofnung, M. (1993) Location of tolerated insertions/deletions in the structure of the maltose binding protein, *FEBS Letters 325*, 34-38.

25. Betton, J. M., and Hofnung, M. (1994) In vivo assembly of active maltose binding protein from independently exported protein fragments, *The EMBO Journal 13*, 1226-1234.

26. Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., Fedorova, I., Nakane, T., Makarova, K. S., Koonin, E. V., Ishitani, R., Zhang, F., and Nureki, O. (2016) Crystal structure of Cpf1 in complex with guide RNA and target DNA, *Cell 165*, 949-962.

27. Kim, H. K., Song, M., Lee, J., Menon, A. V., Jung, S., Kang, Y.-M., Choi, J. W., Woo, E., Koh, H. C., Nam, J.-W., and Kim, H. (2017) In vivo high-throughput profiling of CRISPR-Cpf1 activity, *Nature Methods 14*, 153-159.

28. Giacalone, M. J., Gentile, A. M., Lovitt, B. T., Berkley, N. L., Gunderson, C. W., and Surber, M. W. (2006) Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system, *BioTechniques 40*, 355-364.

29. Wegerer, A., Sun, T., and Altenbuchner, J. (2008) Optimization of an *E. coli* L-rhamnose-inducible expression vector: test of various genetic module combinations, *BMC Biotechnology 8*, 2.

30. Afroz, T., Biliouris, K., Kaznessis, Y., and Beisel, C. L. (2014) Bacterial sugar utilization gives rise to distinct single-cell behaviours, *Molecular microbiology 93*, 1093-1103.

31. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proceedings of the National Academy of Sciences 97*, 6640-6645.

32. Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y.,

Scott, D. A., Severinov, K., van der Oost, J., and Zhang, F. (2017) Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array, *Nature Biotechnology 35*, 31-34.

33. Choi, S.-L., Rha, E., Lee, S. J., Kim, H., Kwon, K., Jeong, Y.-S., Rhee, Y. H., Song, J. J., Kim, H.-S., and Lee, S.-G. (2014) Toward a generalized and high-throughput enzyme screening system based on artificial genetic circuits, *ACS Synthetic Biology 3*, 163-171.

34. Kim, H., Rha, E., Seong, W., Yeom, S.-J., Lee, D.-H., and Lee, S.-G. (2016) A cell–cell communication-based screening system for novel microbes with target enzyme activities, *ACS Synthetic Biology 5*, 1231-1238.

35. Liu, D., Xiao, Y., Evans, B. S., and Zhang, F. (2015) Negative feedback regulation of fatty acid production based on a malonyl-coA sensor–actuator, *ACS Synthetic Biology 4*, 132-140.

36. Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., and Voigt, C. A. (2016) Genetic circuit design automation, *Science 352*, aac7341.

37. Rogers, J. K., Taylor, N. D., and Church, G. M. (2016) Biosensor-based engineering of biosynthetic pathways, *Current Opinion in Biotechnology 42*, 84-91.

38. Yoo, S. M., Na, D., and Lee, S. Y. (2013) Design and use of synthetic regulatory small RNAs to control gene expression in *Escherichia coli*, *Nature Protocols 8*, 1694-1707.

FIGURE LEGENDS

Figure 1. Generation of DNase-deactivated EeCpf1. Based on amino acid sequence alignment of FnCpf1, AsCpf1, and EeCpf1, we determined the three essential catalytic residues (D880, E965, and D1233) in EeCpf1 and created a mutation in D880A to produce an

EedCpf1 (Figure S1). Complexes of WT EeCpf1 or EedCpf1(D880A) with crRNA were assayed for DNase activity. The D880A mutation completely deactivated the DNA cleavage activity of EeCpf1. WT EeCpf1 can cleave supercoiled pUC19 plasmid DNA in the presence of Mn^{2+} in a crRNA-dependent manner, while the EedCpf1(D880A) has no nuclease activity.

Figure 2. Effect of strand bias on CRISPRi activity in EedCpf1 and SpdCas9 systems targeting the 5′ UTR region. Schematic representation of CRISPRi plasmids bearing EedCpf1 (**A**) and SpdCas9 (**B**). Reporter plasmids with binding sites targeting the non-template (NT1) (**C**) and template (T1) (**D**) strands in the 5′ UTR. CRISPRi targeting the template (T1, blue bar) or non-template (NT1, red bar) strand (**E**). To examine the effect of redundant 334-bp sequences at the 3′ end of the spacer in crRNAR(T1) on repression efficiency, we synthesized another crRNA cassette of EedCpf1, i.e., crRNA(T1), which lacked the 3′ repeat sequence of crRNAR(T1). Data are means from three biological replicates, error bars represent standard deviations. Template strand-targeting EedCpf1 significantly repressed *gfp* expression in both cases of crRNA(T1) and crRNAR(T1), with *p*-values of 0.0003 and 0.0001, respectively. The spacer length requirement of crRNA for gene repression with CRISPR-EedCpf1 (**F**). All cases were significantly repressed *gfp* expression, but the maximum repression efficiency was saturated from the case of 20-nt spacer. Data are means from four biological replicates, error bars represent standard deviations. The two tailed t-tests evaluating whether the average difference of fluorescence intensity between each spacer (16, 18, 20, 22, 24, 25) and empty vector (EV) is zero, yielding that the *p*-values were 0.00117, 0.00034, 0.00029, 0.00013, 0.00029, and 0.00026, respectively.

Figure 3. Effectiveness of EedCpf1 and SpdCas9 CRISPRi targeting the CDS region. Schematic representation of reporter plasmids bearing a crRNA(T1) binding site in various

CDS regions and strand bias. A crRNA(T1) binding site was inserted in-frame after the methionine 1 (M1), alanine 206 (A206), or asparagine 372 (N372) codon of the MBP-GFP fusion protein-coding sequence on the template (T) (**A**) or non-template (NT) (**B**) strand. CRISPRi activity assays of EedCpf1 (**C**) and SpdCas9 (**D**) with the six reporter plasmids. Data are means from three biological replicates, error bars represent standard deviations. The expression of MBP-EGFP encoded by pMEGFP(NT1) was too low to allow quantification. ND, not determined. Non-template- versus template-targeted expression levels with EedCpf1 were significantly different in the A206 and N372 cases ($p = 0.0098$ and $p = 0.0013$, respectively, where p indicates p-value of a t-test).

Figure 4. Characterization of the PAM domain of EedCpf1. Repression activities of 5′-CTTTC-3′, 5′-CCTTC-3′, and 5′-CCTC-3′ PAM sequences, examined after mutating the 5′-TTTTC-3′ PAM sequence of pREGFP(T1) (**A**). 5′-NTTTC-3′ (**B**), 5′-CNTTC-3′ (**C**), and 5′-CTTTN-3′ (**D**) PAM sequences were used to examine EedCpf1 binding affinity. Data are means from three biological replicates, error bars represent standard deviations. In statistical comparison of the repression levels of two groups (group1: 8 cases of 5′-NTTTV-3′ (V=A, G, or C), group2: the other 6 PAM sequences), the mean relative expression levels were 13.91% for group1 and 70.62% for group2, with the log of the p-value being approximately –10, which strongly supports that EedCpf1 with 5′-NTTTV-3′ PAM sequences has a more efficient repression effect than EedCpf1 with the other PAM sequences.

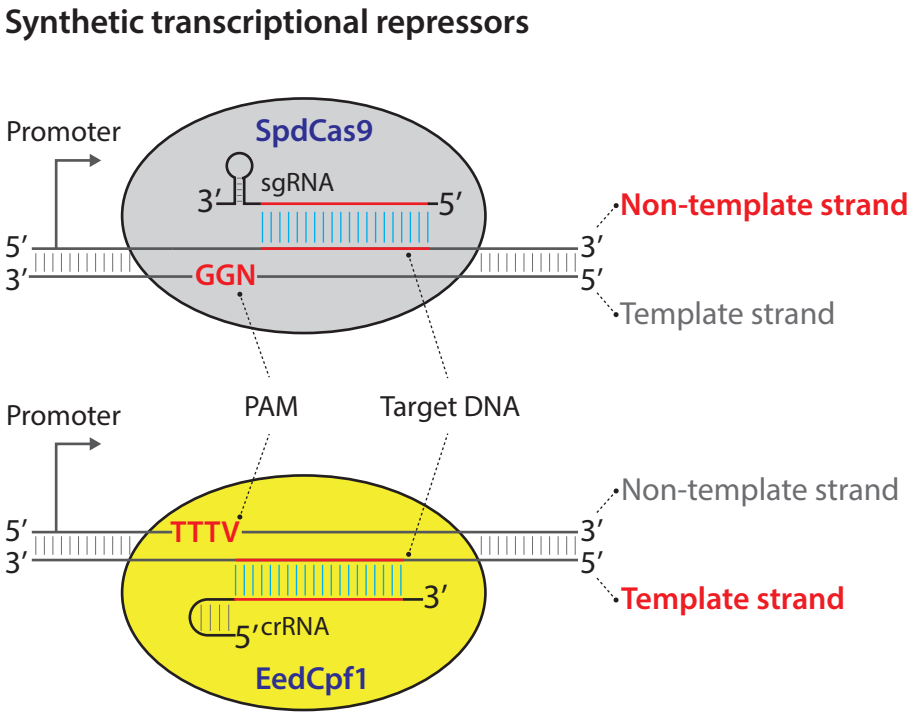
Figure 5. Comparison of the repression activity of the designed EedCpf1 system targeting the promoter, 5′ UTR, and CDS regions. Eleven sites were selected within the promoter (P1 and P2), 5′ UTR (T1), and CDS (C1, C2, C3, C4, C5, C6, C7 and C8) regions (**A**). P1, C2, C6, C7, and C8 sites reside on the non-template strand of the reporter plasmid, while the

remaining sites reside on the template strand. Repression of gene expression using the reporter plasmid (**B**). Data are means from three biological replicates, error bars represent standard deviations. Tunability of EedCpf1 with L-rhamnose as an inducer is shown (**C** and **D**). See text for details. In statistical comparison of the repression levels of two groups (template group: C1, C3, C4, C5, non-template group: C2, C6, C7, C8), the mean expression levels are 17.11% and 93.8% for template and non-template groups respectively, and the log of the *p*-value is approximately –9, which strongly supports that template targeting is much more efficient than non-template targeting in the use of EedCpf1. EV, empty vector.

Figure 6. Repression of chromosomal gene expression by the designed EedCpf1 system. A *gfp* expression cassette with eight designed CRISPRi-targeted sites within the promoter, 5′ UTR, and CDS regions was integrated into the chromosome of *E. coli* DH5α strain (reporter strain). The repression of gene expression by EedCpf1 targeting the eight binding sites was assayed (**A**, right) and single-cell flow cytometry fluorescence assays evaluating population homogeneity were conducted (**A**, left). To test multiplex repression of chromosomal targets, CRISPRi targeting C4 and/or C5 regions in the chromosomal reporter gene (*gfp*) was performed (**B**) and single-cell fluorescence assays were conducted (**C**). NC: *E. coli* DH5α strain, PC: reporter strain harboring empty vector (pSEVA221), ND: not detected. Data are means from three biological replicates, error bars represent standard deviations. C4 significantly repressed *gfp* expression in comparison to PC. C4C5 showed significantly stronger repression than C4 ($p = 0.009$, two-tailed *t*-test). Simultaneous repression of multiple genes was performed with endogenous *lacZ* gene of *E. coli* K-12 MG1655 and exogenous *gfp* gene of pREGFP3(P2T1) plasmid (**D**). Since *E. coli* DH5α is *lacZ* negative, we used K-12 MG166 strain that was grown in LB solid medium containing 0.5 mM IPTG and 80 μg/ mL

of X-Gal. The crRNA(C1lacZ) simultaneously repressed the plasmid-borne *gfp* and chromosomal *lacZ* in *E. coli* K-12 MG1655. EV, empty vector.

Graphic abstract



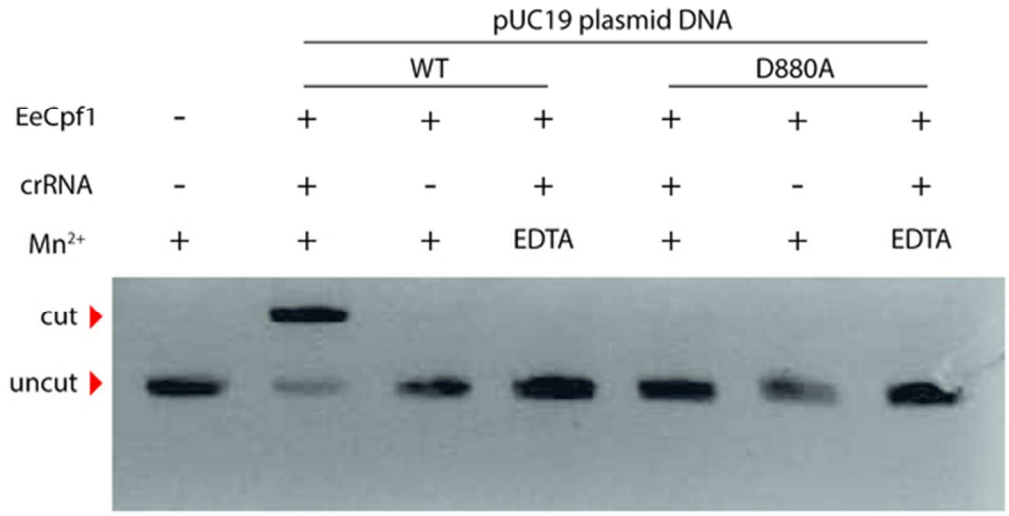


Figure 1

50x25mm (300 x 300 DPI)

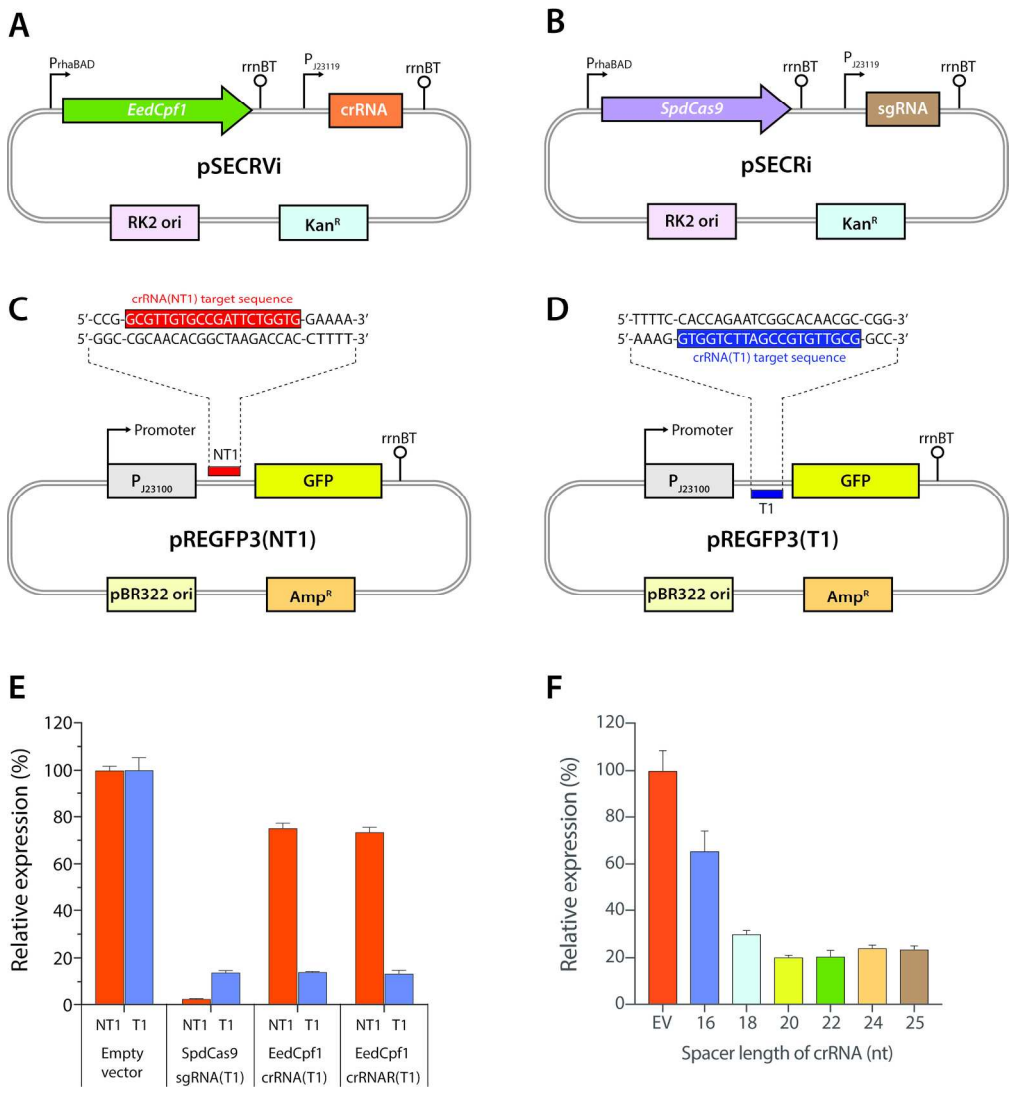


Figure 2

195x211mm (300 x 300 DPI)

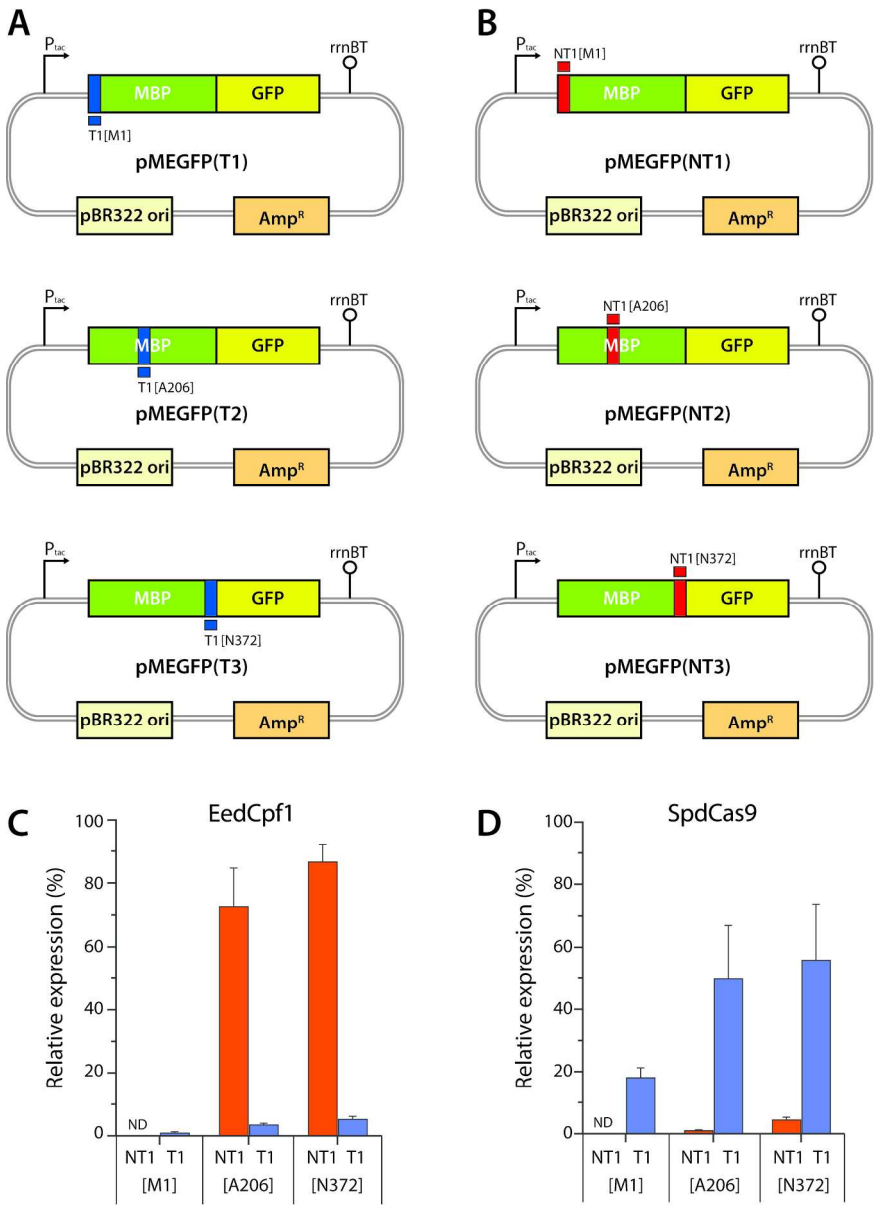


Figure 3

200x277mm (300 x 300 DPI)

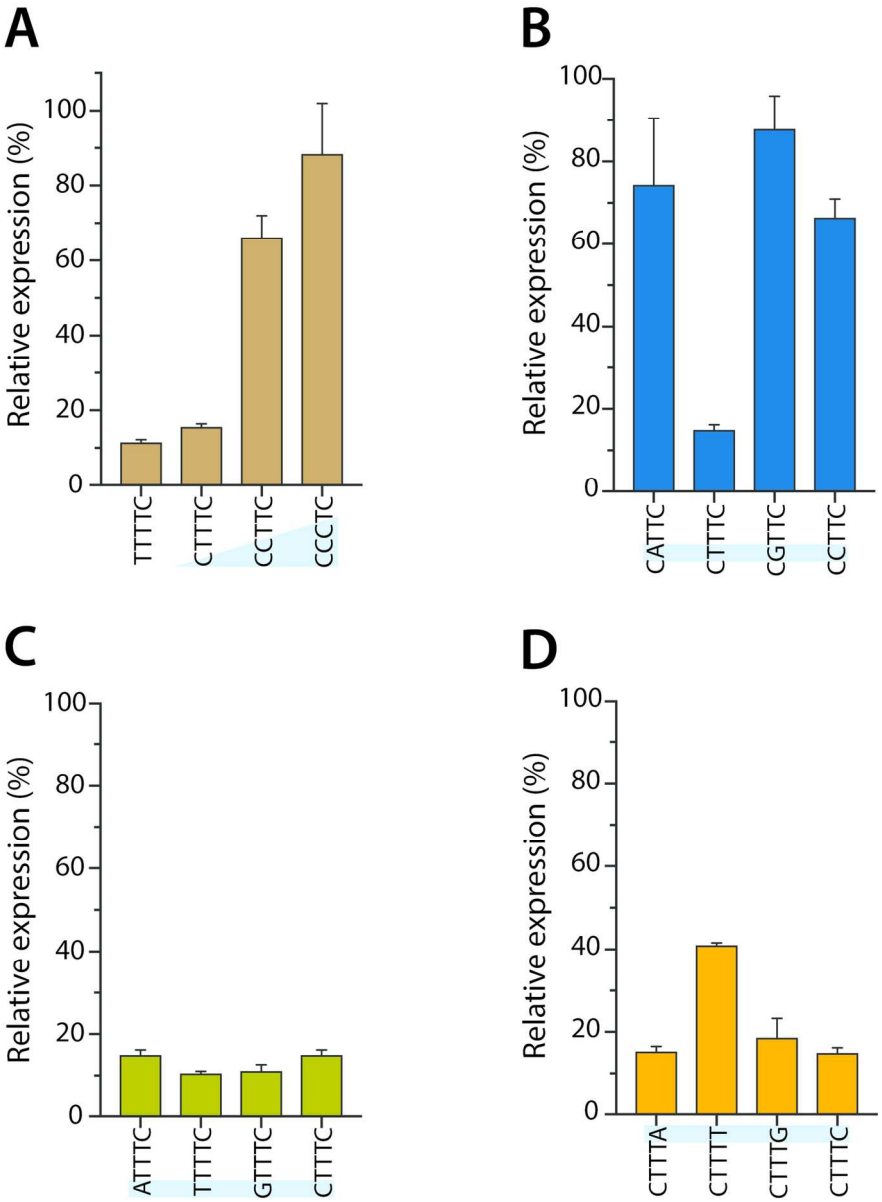


Figure 4

131x180mm (300 x 300 DPI)

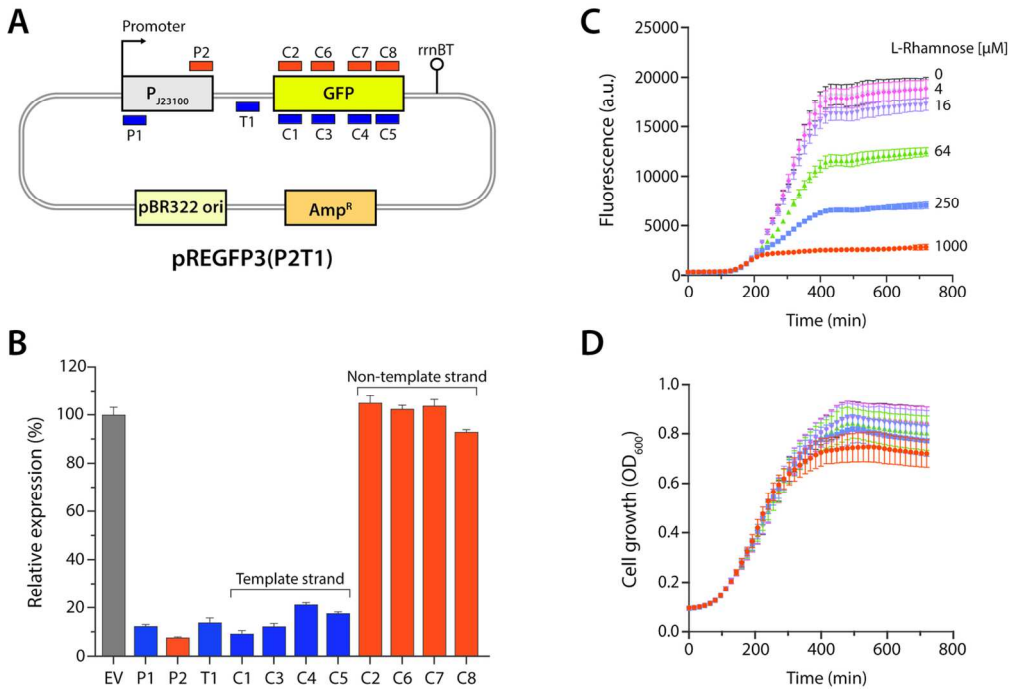


Figure 5

120x82mm (300 x 300 DPI)

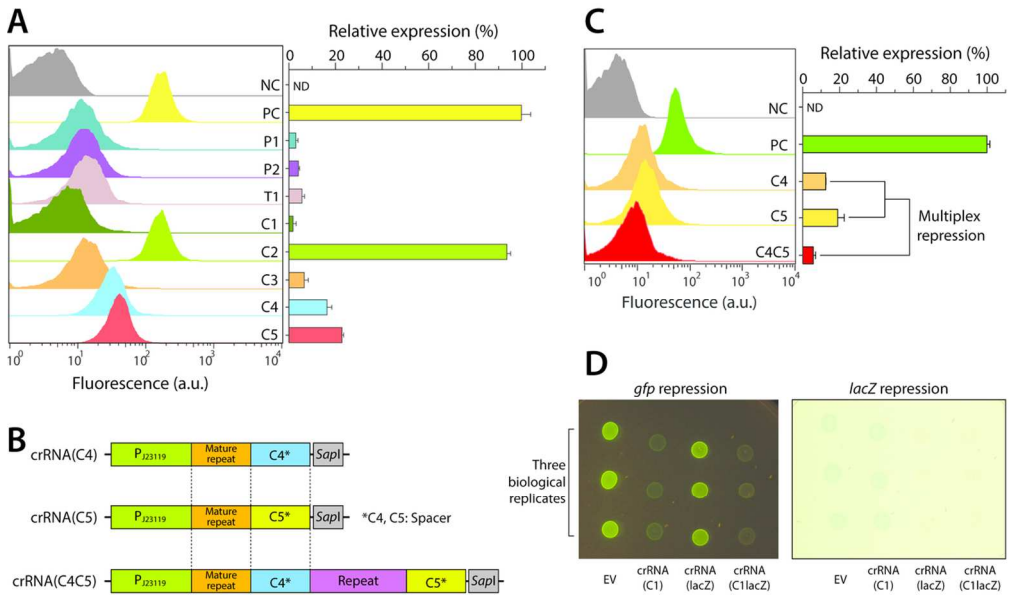


Figure 6

118x69mm (300 x 300 DPI)