



HAL
open science

Cas9 off-target binding to the promoter of bacterial genes leads to silencing and toxicity

William Rostain, Theophile Grebert, Danylo Vyhovskyi, Paula Thiel Pizarro, Gatwa Tshinsele-van Bellinghen, Lun Cui, David Bikard

► To cite this version:

William Rostain, Theophile Grebert, Danylo Vyhovskyi, Paula Thiel Pizarro, Gatwa Tshinsele-van Bellinghen, et al.. Cas9 off-target binding to the promoter of bacterial genes leads to silencing and toxicity. *Nucleic Acids Research*, 2023, 51 (7), pp.3485-3496. 10.1093/nar/gkad170 . pasteur-04130634

HAL Id: pasteur-04130634

<https://pasteur.hal.science/pasteur-04130634v1>

Submitted on 16 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Cas9 off-target binding to the promoter of bacterial genes leads to silencing and toxicity

William Rostain^{1,†}, Theophile Grebert^{1,†}, Danylo Vyhovskyi^{1,2}, Paula Thiel Pizarro¹, Gatwa Tshinsele-Van Bellinghen¹, Lun Cui¹ and David Bikard^{1,*}

¹Synthetic Biology, Institut Pasteur, Université Paris Cité, CNRS UMR 6047, 75015 Paris, France and ²Collège Doctoral, Sorbonne Université, F-75005 Paris, France

Received December 15, 2022; Revised February 21, 2023; Editorial Decision February 22, 2023; Accepted February 26, 2023

ABSTRACT

Genetic tools derived from the Cas9 RNA-guided nuclease are providing essential capabilities to study and engineer bacteria. While the importance of off-target effects was noted early in Cas9's application to mammalian cells, off-target cleavage by Cas9 in bacterial genomes is easily avoided due to their smaller size. Despite this, several studies have reported experimental setups in which Cas9 expression was toxic, even when using the catalytic dead variant of Cas9 (dCas9). Specifically, dCas9 was shown to be toxic when in complex with guide RNAs sharing specific PAM (protospacer adjacent motif)-proximal sequence motifs. Here, we demonstrate that this toxicity is caused by off-target binding of Cas9 to the promoter of essential genes, with silencing of off-target genes occurring with as little as 4 nt of identity in the PAM-proximal sequence. Screens performed in various strains of *Escherichia coli* and other enterobacteria show that the nature of toxic guide RNAs changes together with the evolution of sequences at off-target positions. These results highlight the potential for Cas9 to bind to hundreds of off-target positions in bacterial genomes, leading to undesired effects. This phenomenon must be considered in the design and interpretation of CRISPR–Cas experiments in bacteria.

INTRODUCTION

Over the last decade, the Cas9 protein has been harnessed as a powerful genetic tool in bacteria (1–3). Guide RNAs can be programmed to target any gene of interest with the simple requirement of a protospacer adjacent motif (PAM) next to the target position. In addition to genome editing applications that utilize the introduction of double-strand

breaks (4–6), the catalytic dead variant of Cas9 (dCas9) has been used in several powerful methods. dCas9 can efficiently block transcription initiation when binding within promoter sequences or transcription elongation when binding downstream of promoters (7,8). This strategy, also termed CRISPR interference (CRISPRi), has been developed in a wide range of bacteria to control the expression of individual genes, as well as in high-throughput screens to study gene function or optimize pathways of interest (9,10). Furthermore, protein domains can be fused to dCas9 to act at the target locus in various ways, such as using transcriptional activation domains to induce gene expression (8,11–14) or fusing cytosine or adenine deaminases to chemically modify bases in the target sequence (15,16). The list of applications continues to grow as researchers find ever more clever ways to combine Cas9 with other protein domains (17).

While Cas9-derived tools have demonstrated effectiveness in bacteria, several studies have reported that Cas9 expression can be toxic. For example, the commonly used dCas9 protein from *Streptococcus pyogenes* (SpydCas9) is lethal in *Mycobacterium smegmatis* when expressed with a nontargeting control guide (18). To address this issue, Rock *et al.* screened natural Cas9 variants from various origins and identified the dCas9 protein derived from the *Streptococcus thermophilus* CRISPR locus 1 as a suitable alternative. Similarly, SpydCas9 was seen to be toxic in *Synechococcus* (19) and in *Corynebacterium glutamicum* (20).

In *Escherichia coli*, the SpydCas9 protein could readily be used to control gene expression for a variety of tasks, but some early reports also noted toxicity and morphological defects when dCas9 was overexpressed (21,22). We have previously demonstrated that dCas9 can inhibit the growth of *E. coli* when combined with guide RNAs that have specific 5-nt motifs at their PAM-proximal end, also known as the seed sequence, in what we refer to as the 'bad-seed' effect (23). Over 100 of the 1024 possible combinations of 5 nt caused a significant decrease in fitness, ranging from mild reduction of growth rate to complete inhibition

*To whom correspondence should be addressed. Tel: +33 140613924; Email: david.bikard@pasteur.fr

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors. Present address: Lun Cui, CCZU-JITRI Joint Bio-X Lab, Changzhou University, Changzhou, Jiangsu 213100, China.

of colony formation. We have found that this ‘bad-seed’ effect depends on the concentration of dCas9 in bacteria and that the expression level of dCas9 can be adjusted to reduce toxicity while still maintaining good on-target activity (23). However, the mechanism responsible for this toxicity phenomenon has so far remained mysterious.

Cas9 searches for target DNA sequences by randomly probing DNA at positions carrying a PAM (24). The double helix is slightly unwound to enable pairing between the guide RNA and the target DNA at the PAM-proximal region. Single-molecule approaches have shown how dCas9 pauses at PAMs (24,25), while chromatin immunoprecipitation experiments (ChIP-seq) have detected dCas9 binding in the chromosome of eukaryotic cells at positions requiring only 5 nt of identity to the seed (26,27). However, this level of identity is thought to be too small to mediate strong binding of dCas9 and to have substantial impact on gene expression at the target position (23).

Here, we investigate the mechanism behind the sequence-specific toxicity of dCas9 when in complex with guide RNAs carrying bad-seed sequences. We use RNA-seq assays to study the transcriptional response to bad-seed toxicity and ChIP-seq assays to identify positions bound by dCas9 when in complex with toxic guide RNAs. In combination with a series of genetic experiments, we demonstrate how dCas9 can bind to off-target positions with as little as 4 or 5 nt of identity between the guide RNA and the target. When such off-target binding occurs in the promoter sequence of an essential gene, it can block expression and inhibit cell growth. This phenomenon was also observed in a panel of various *E. coli* isolates and other Enterobacteriaceae. The nature of the toxic sequences varies between strains and species, reflecting differences in DNA sequences at the off-target positions and changes in how gene expression affects bacterial growth.

MATERIALS AND METHODS

Strains and plasmids

Plasmid pTG34 is a p15A plasmid carrying dCas9 under the control of the 2,4-diacetylphloroglucinol (DAPG)-inducible pPhIF promoter, a guide RNA with BsaI cloning sites framing a *ccdB* counterselection marker, an RP4 origin of transfer and a chloramphenicol resistance marker. It is a derivative of plasmid pFR56 (28) (GenBank: MT412099.1) in which the 5′ UTR of dCas9 (5′-GGAGGCATATCAAA GGACGTGTGCAGGTGGCAAAA-3′) was replaced by a sequence with a stronger RBS (5′-TACTGGAGCTTAAAAGGAGAAAGTTTCT-3′).

MFDpir* cells were used as conjugation donor to transfer pTG34 to recipient strains (see Supplementary Table S1). MFDpir* was grown in LB medium supplemented with 300 μM diaminopimelic acid (DAP), and where appropriate supplemented with 25 μg/ml chloramphenicol. Recipient strains were grown in LB, supplemented where appropriate with 25 μg/ml chloramphenicol.

RNA-seq experiments were performed using strain *E. coli* LC-E18, an MG1655 derivative carrying dCas9 un-

der the control of a p_{tet} promoter integrated at the HK022 attB site (23). Strain AV01 is an MG1655 derivative carrying dCas9 under the control of a p_{tet} promoter integrated at the 186 primary attB site. Guide RNAs were cloned into psgRNACos using Golden Gate assembly (23).

ChIP-seq experiments were performed using *E. coli* strain MG1655. To enable immunoprecipitation, a C-terminal 3× FLAG tag was added to dCas9 and the gene cloned under the control of a p_{tet} promoter on plasmid pLC133 (pSC101, with chloramphenicol resistance gene, Addgene plasmid # 198639). Strains and plasmids are described in Supplementary Tables S5 and S6. These include various strains of *E. coli* as well as strains of *E. albertii*, *E. fergusonii*, *K. pneumoniae* and *C. freundii* used in the screening experiment. All strains were grown in LB with the appropriate antibiotics and inducers as indicated below.

Plasmid pTG131 (Addgene plasmid # 198640) was used for plasmid curing during the generation of variants of the glyQS promoter (see below). pTG131 is a thermosensitive pSC101-based plasmid expressing a single guide RNA (sgRNA) under the control of the J23119 promoter and Cas9 under the control of the DAPG-inducible PhIF promoter.

The sequence of guide RNAs used in this study is provided in Supplementary Table S7. Oligonucleotides used in plasmid construction are listed Supplementary Table S8.

RNA-seq

Samples were prepared from *E. coli* LC-E18 harboring psgRNACos plasmids with three different guides (control guide, AGGAA_n1 and ACCCA_n1). Samples were harvested 3 h after aTc induction, centrifuged at 10 000 × *g* for 2 min and resuspended with 200 μl of lysozyme buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 1% Triton X-100 and 20 mg/ml lysozyme). One milliliter of TRIzol was added to the samples and RNA was extracted following the manufacturer’s protocol. RNA samples were treated with DNase (TURBO DNA-free kit) and then treated with Ribo-Zero rRNA Removal Kit. Sequencing libraries were prepared using the TruSeq Stranded mRNA Sample Preparation LS protocol. The experiment was performed in three biological replicates. Samples were sequenced on an Illumina NextSeq 550 sequencer with 2 × 75 paired-end sequencing. Reads were aligned to the LC-E18 and psgRNACos sequences using Bowtie2 (29). Read counts were computed using featureCounts (30). Differential expression analysis was conducted using DESeq2 tools in R (31).

ChIP-seq

Samples were prepared from *E. coli* MG1655 harboring pLC133 and psgRNACos plasmids with three different guide RNAs, a control guide targeting *lacZ*, the AGGAA_n1 guide or the ACCCA_n1 guide (Supplementary Table S7). Cells were grown at 30°C and induced with aTc when OD₆₀₀ reached 0.2. When OD₆₀₀ reached 1.0, formaldehyde (final concentration: 1%) was added to fix the cells. After that, glycine was added to block formaldehyde. Samples were further washed with PBS and pelleted.

Samples were then resuspended in 1 ml of IP lysis buffer (50 mM HEPES, pH 7.5, 0.4 M NaCl, 1 mM EDTA, 1 mM DTT, 0.5% Triton X-100, 10% glycerol), 20 μ l of 0.1 M phenylmethylsulfonyl fluoride, 20 μ l complete protease inhibitor (50 \times) and 5 μ l RNase inhibitor (40 U/ μ l). Cell lysis was performed using a Covaris sonicator (S220), 10 \times (10 s on, 10 s off, amplitude 15 μ m) until the samples appeared clear. Samples were centrifuged and the supernatant mixed with Anti-Flag M2 Affinity Gel (Sigma–Aldrich) beads and incubated overnight at 4°C on a rotating wheel. After that, the beads were washed five times with IP lysis buffer and then eluted in ChIP-seq elution buffer (50 mM Tris–HCl, pH 8.0, 10 mM EDTA, 1.0% SDS). Bead pellets were incubated at 65°C in a water bath for 1 h and then spun at 14 000 rpm for 3 min. The supernatant, which contains the chipped DNA, was then collected.

Samples were column purified to remove RNA and then fragment ends were repaired and phosphorylated with T4 polymerase and T4 PNK. After that, an A was added to 3' end of repaired ends and then the DNA fragments with 3' A overhang were ligated with Illumina TruSeq adaptors. Samples were PCR amplified using primers (LC1504.TruSeq_P5, LC1505.TruSeq_P7). Samples were sequenced on an Illumina NextSeq 500 sequencer (150 cycles, 2 \times 75 paired end).

Fastp (v.0.20.1) was used to trim, filter low-quality bases and remove adapter sequences from raw reads. Filtered reads <30 nt were discarded. Only pairs for which both reads passed filtering were mapped to MG1655 genome using Bowtie2 (v.2.3.4.3). Aligned reads were sorted and indexed with SAMtools (v.1.9). The bamCoverage function of deepTools (v3.4.1) was used to compute coverage values and ChIP peaks were identified using the find_peaks function of the scipy.signal (v1.10) Python package with the following parameters: distance = 25 and prominence = 30.

RT-qPCR

Overnight cultures were diluted 1:100 in 3 ml of LB and induced with 1 nM aTc when cultures reached an OD of 0.2 (~1 h after dilution). After 2 h of induction, cells were treated with 2 \times volume of RNAProtect Bacteria Reagent (Qiagen) for 10 min at room temperature and collected by centrifuging for 10 min at 3300 \times g. Cells were lysed with lysozyme and RNA extracted using TRIzol followed by isopropanol precipitation and two ethanol washes. TURBO DNA-free kit (Thermo Fisher Scientific) was used for DNase treatment, and RNAs were reverse transcribed using the Transcriptor First Strand cDNA Synthesis Kit (Roche) in 20 μ l using 100 ng RNA (concentration was measured with Agilent RNA ScreenTape). Dual-labeled probes (5' FAM, 3' BHQ1) for *glyQ* and *rrsA* genes were used to perform qPCR with the FastStart Essential DNA Probes Master Mix (Roche) in a LightCycler 96 (Roche) following the manufacturer's instructions. qPCR was performed in two technical replicates each time with 1 μ l of the nonpurified cDNA subjected for amplification and three biological replicates. Relative gene expression was computed using the $\Delta\Delta C_q$ method after normalization by 5S rRNA (*rrsA*). qPCR primers and probes are listed in Supplementary Table S9.

Validation of off-target sites

Mutations were introduced in the PAM sequence of candidate off-target positions by recombineering. Plasmid pORTMAGE-Ec1 (32) was introduced in *E. coli* AV01 cells using Transformation and Storage Solution (TSS) (33). Strains were grown in selective media and transformed by a second TSS transformation with a psgRNAc expressing a nontargeting sgRNA with the corresponding bad seed (see Supplementary Table S7). This second plasmid was used to select mutants at the off-target position, which are expected to grow better upon dCas9 induction than the wild type (WT) in the presence of the cognate guide RNA. Oligonucleotide recombineering was then performed as described in (32). In summary, overnight cultures were diluted 1:100 in LB and grown for 1 h until an OD₆₀₀ of around 0.3. Cells were then induced with 1 mM *m*-toluic acid and incubated with agitation at 37°C for 1 h. Then, cells were kept on ice for 10 min, washed with cold water three times and resuspended in 10% glycerol (1/100 of the culture volume). For the electroporation, 81 μ l of the prepared cells were mixed with 9 μ l of the oligo at 100 μ M (final concentration of 10 μ M) and electroporated at 1.8 kV, 200 Ω and 25 μ F in 0.1 cm electroporation cuvettes. Finally, cells were recovered in LB for 1 h and plated in media supplemented with chloramphenicol. Mutations were verified by Sanger sequencing. A list of oligonucleotides used in recombineering can be found in Supplementary Table S10.

Mutants obtained in this manner as well as strain AV01 (WT) expressing the same guide RNA on psgRNAc were grown overnight in LB + chloramphenicol. Cultures were then serially diluted and spotted on plates with or without aTc.

Off-targets in the promoter of *rpmH* were validated using a plasmid complementation assay. Plasmid pTG40 expressing *rpmH-rnpA* operon was introduced in strain AV01 together with psgRNAcos carrying toxic guides with candidate off-targets in the *rpmH* promoter (ACCCA-n1, CACTC-n1, GAGGC-n1).

glyQ promoter variants

Oligo recombineering was used to generate variants of the AGGAAGGG motif present in the promoter of *glyQ* in *E. coli* MG1655. Plasmid pORTMAGE-Ec1 (32), which harbors the recombineering machinery, was introduced in strain AV01, an MG1655 derivative with P_{tet}-dCas9 integrated at the 186 primary attB site. We relied on the AGGAA bad-seed effect for efficient counterselection of the WT. To this end, a plasmid expressing a nontargeting sgRNA with an AGGAA seed (psgRNAc-AGGAA-n1) was also introduced in strain AV01. Recombineering was achieved by electroporating oligo TG395 in cells AV01/pORTMAGE-Ec1/psgRNAc-AGGAA-n1. Counterselection was performed by plating on plates containing chloramphenicol and aTc. Individual colonies were picked and the *glyQ* promoter sequenced. Plasmid psgRNAc-AGGAA-n1 was then cured through targeting with Cas9 using plasmid pTG131 (Addgene plasmid # 198640). The plasmid was programmed with a guide targeting the *cat* gene in psgRNAc, giving pTG131-*cat*. This

plasmid was then introduced in each variant, and cells selected on plates containing ampicillin and DAPG at 30°C. Curing of plasmid pTG131 was performed by growing the cells at 42°C. Successful curing of plasmids psgRNAc and pTG131 was verified by checking susceptibility to chloramphenicol and ampicillin. For each variant, a psgRNAc plasmid expressing an sgRNA matching the new seed sequence was assembled. The sequence of all guide RNAs used can be found in Supplementary Table S7.

Library cloning and conjugation

The TG120 oligonucleotide containing a stretch of 20 random bases was PCR amplified with primers FR221/FR222 using Phusion polymerase (six cycles, T_m 63°C, 3% DMSO), PCR purified and cloned into BsaI-linearized pTG34 by Gibson assembly. This was followed by electroporation into fresh MFDpir* *E. coli* cells (34). An aliquot was serially diluted for measuring library size, while the rest of the library was plated on six 12 cm × 12 cm plates, grown at 30°C overnight. Each culture was resuspended in 10 ml LB before pooling all six. To reduce library diversity to manageable numbers for high-throughput sequencing and data analysis, we applied an artificial bottleneck. Serial dilutions of the mother library were prepared, and for each dilution, a CFU spot assay was carried out in parallel to a culture started with 1 ml of the dilution. A post-bottleneck donor library containing a maximum of 1.8×10^5 guides was selected.

The pTG34 containing the random library was then introduced into recipient strains by conjugation. Recipient strains were grown to OD ~ 1 from an overnight starter culture, while the donor strain was grown to OD 0.6 directly from a DMSO stock. Two milliliters of donor was centrifuged, rinsed in LB + DAP twice and then mixed with 2 ml of recipient cells. The mixture was centrifuged, resuspended in 50 µL LB + DAP, pipetted onto LB + DAP agar plates and incubated for 2.5 h at 37°C. After conjugation, the cells were resuspended in 2 ml of LB and then plated on two 12 cm × 12 cm LB plates and grown overnight at 30°C, while an aliquot was serially diluted and plated for library size measurement. Over 10^7 transconjugants were obtained for all strains. Trans-conjugants were resuspended in 10 ml LB and stored at -80°C after addition of 10% DMSO until screening.

Random guide library screening

Fifty microliters of each DMSO stock was used to inoculate 1450 µl LB + Cm in deep 96-well plates, with duplicates for each strain, and grown for 12 h. Plasmid DNA was extracted from these cultures, providing the reference samples. Three passages were then performed in LB + 50 µM DAPG, inoculating 1485 µl medium with 15 µl cells, representing ~20 generations [$\log_2(100^3) \approx 19.9$]. The cultures were incubated for 5 h after the first and second dilutions, and 12 h after the last dilution. Plasmid DNA was extracted from these cultures, providing the induced samples.

Library sequencing

Library sequencing was performed as described previously (23,28). Briefly, 50 ng of template was amplified by two consecutive PCRs with primers described in Supplementary Table S11 with KAPA HiFi polymerase (Roche). Cycling conditions were as follows: first PCR: 95°C for 3 min; $9 \times 98^\circ\text{C}$ for 20 s, 60°C for 15 s and 72°C for 20 s; 72°C for 10 min, using 8.6 pmol of each primer. A 20 µl mix containing 100 pmol of primers was then added to the reaction and a second PCR conducted (cycling conditions: 95°C for 3 min; $9 \times 98^\circ\text{C}$ for 20 s, 60°C for 15 s and 72°C for 20 s; 72°C for 10 min). The resulting 354-bp fragments were gel extracted, pooled (50 ng per sample) and the final library was quantified by qPCR (NEBNext Library Quant Kit for Illumina) and sequenced on a NextSeq500 sequencer (Illumina) using a custom protocol as described previously (23,35). We obtained an average of 2.7 million reads per sample, and an average of 5.3×10^5 unique guides with >20 reads in the reference sample of each strain.

Screening data analysis

Sequences were demultiplexed and count tables generated with a custom Python script. We obtained a total of 1.3×10^8 usable reads with a mean of 2.77×10^6 reads per sample. Guides with <20 reads in the reference sample were discarded. The $\log_2\text{FC}$ value was calculated for each guide g and strain s as follows (s_{initial} and s_{final} represent the normalized read counts of strain s at the initial and final time points, respectively):

$$\log_2 \text{FC}_{g,s} = \log_2 \frac{\text{Reads}_{g,s_{\text{final}}} + 1}{\text{Reads}_{g,s_{\text{initial}}} + 1}$$

We observed a good reproducibility between biological duplicates (Supplementary Figure S1) except for *C. freundii* + DAPG, where one replicate had a low read count. This sample was discarded. For other samples, the read counts of biological repeats were summed before computing $\log_2\text{FC}$. Read counts are provided in Supplementary Table S12 and $\log_2\text{FC}$ values are provided in Supplementary Table S13.

RESULTS

Validation of the bad-seed toxicity phenomenon using high-throughput screening of guide RNAs with random sequences

We first carried out a CRISPRi screen with a library of random guide RNAs to corroborate our previous findings and ascertain that the bad-seed phenotype does not require the presence of bona fide targets in the host genome. Oligonucleotides in which random bases were incorporated at each of the 20 positions of the guide RNA were synthesized and cloned in plasmid pTG34 (Figure 1A). This plasmid contains a constitutively expressed sgRNA as well as a DAPG-inducible *dcas9* gene. A bottleneck was applied to yield a library of $\sim 10^5$ guide RNAs. This library size ensures that on average each 5-nt combination will be present in ~ 100 different guides while enabling sufficient sequencing coverage to obtain robust estimates of fold changes for each

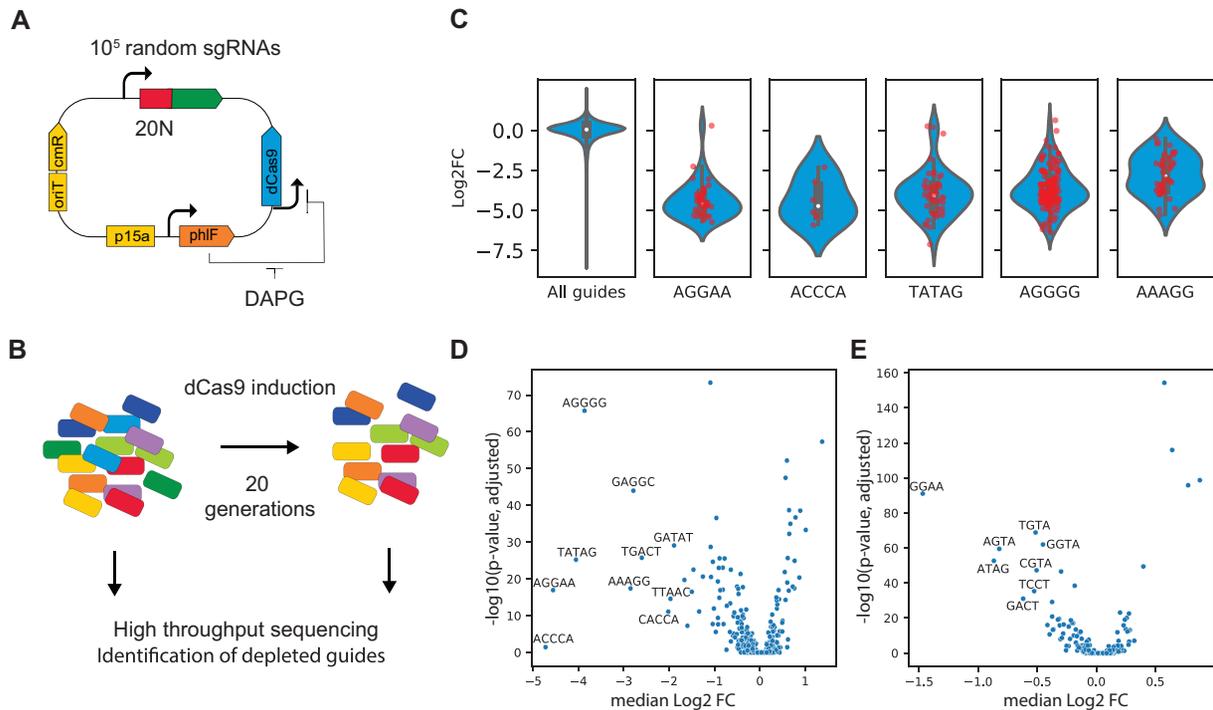


Figure 1. A random guide RNA library identifies toxic seed sequences. (A) Architecture of pTG34, which contains a DAPG-inducible dCas9 and a library of random guide RNAs. (B) Diagram of CRISPRi screen to identify unexpected guide toxicity. (C) Violin plots of \log_2FC for guide RNAs carrying different PAM-proximal sequences. The median \log_2FC of guide RNAs grouped by (D) their 5-nt PAM-proximal sequence or (E) their 4-nt PAM-proximal sequence is plotted against the $-\log_{10}(P\text{-values})$ of a Bonferroni adjusted Mann–Whitney U test.

guide. The expression of *dcas9* was then induced for ~ 20 generations and the library was sequenced before and after induction (Figure 1B). As expected, the distribution of \log_2FC values of guides was centered around 0 with a skew toward lower values, indicating that most guides have no effect while a few are depleted (Figure 1C and Supplementary Table S1). Guides were grouped based on their seed sequences, and the median for each group calculated (Figure 1D and Supplementary Table S2). When compared to our previous dataset obtained with *dcas9* expressed under the control of a pTet promoter from the chromosome (strain LC-E18) (23), we see that the \log_2FC values are smaller, indicating that the dCas9 levels when expressed from plasmid pTG34 are likely lower (Supplementary Figure S2). Nonetheless, the average \log_2FC values of guides sharing the same 5-nt seeds are well correlated in between the two experimental systems (Pearson $R = 0.65$). This confirms the toxicity of guide RNAs sharing specific 5-nt seed sequences regardless of the other 15 nt of the guide sequence or the existence of fully complementary targets in the chromosome. The screen also revealed a significant depletion of some 4-nt seed patterns, notably GGAA, a subsequence of the second most depleted 5-nt bad seed AGGAA (Figure 1E and Supplementary Table S3). We could also observe that some sets of guides sharing 6-nt seeds were depleted, while other guides sharing the same 5-nt seeds were not. Likewise, some 7-nt seeds were depleted, while the corresponding 6-nt seeds were not (Supplementary Figure S3 and Supplementary Table S4).

Bad-seed toxicity is due to off-target binding in the promoter of essential genes

We then investigated the transcriptional response to the two most depleted 5-nt seeds, AGGAA and ACCCA, by RNA-seq. Experiments were performed using strain LC-E18, an MG1655 derivative carrying dCas9 under the control of a pTet promoter in the chromosome (23). Guide RNAs were expressed constitutively from the psgRNA plasmid. The expression profile of *E. coli* was measured after 3 h of dCas9 induction in the presence of nontargeting guides containing the toxic ACCCA or AGGAA seed sequences (median \log_2FC of -4.6 and -4.7 , respectively) or the nontoxic TCTCG seed sequence (median \log_2FC of 0.2). Differential expression analysis was performed by computing fold changes in expression relative to the nontoxic guide RNA. The two most significantly repressed genes in the AGGAA bad-seed condition were *glyQ* and *glyS*, while *rpmH* was the most significantly repressed gene in the ACCCA samples (Figure 2A and B).

The *glyQS* operon encodes the two subunits of the essential glycine-tRNA ligase. To investigate whether repression of these genes could be responsible for the toxicity of guides containing the AGGAA seed sequence, we cloned the *glyQS* operon and its promoter (97-bp intergenic region between the upstream *ysaB* gene and *glyQ* ATG codon) on plasmid pTG36. We then introduced this plasmid in strain LC-E18, in which dCas9 is expressed at high levels, in the presence of a guide RNA with AGGAA seed sequence or a control guide RNA. This complementation rescued the for-

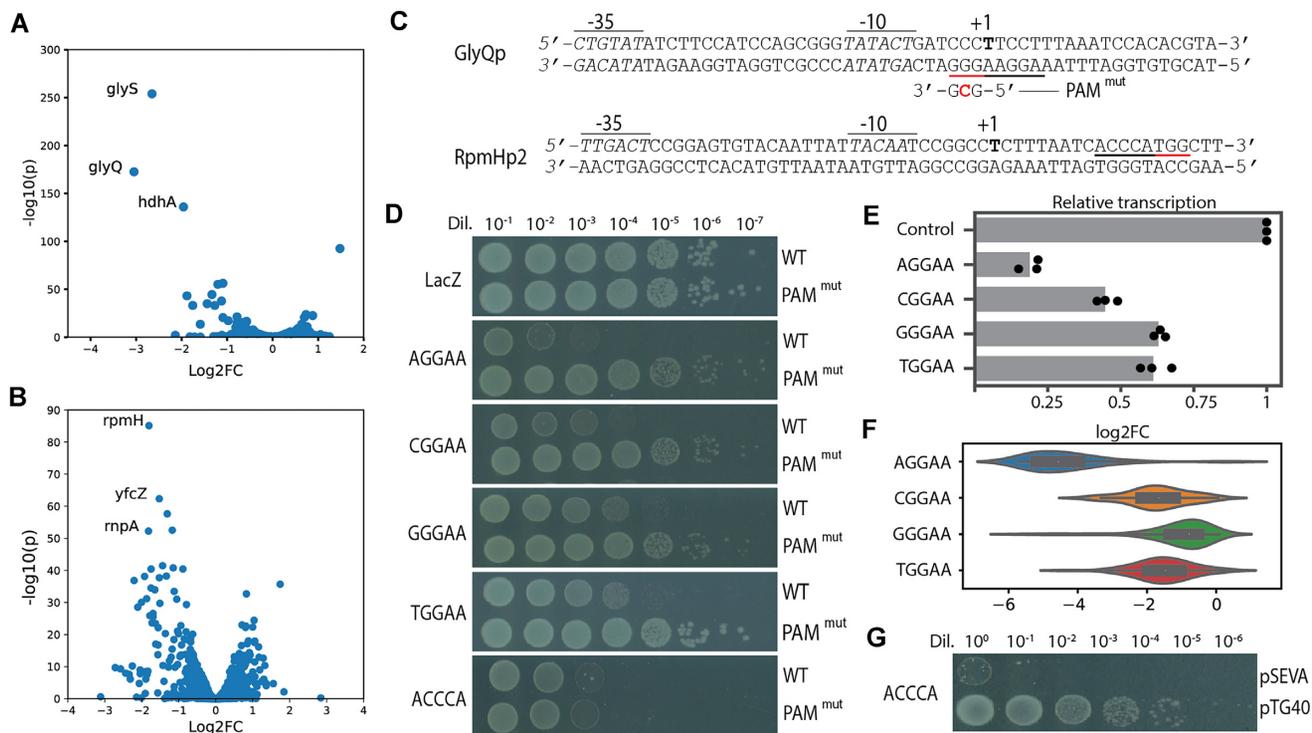


Figure 2. The bad-seed toxicity phenomenon is caused by off-target binding in the promoter of essential genes. Differential expression of genes after 3 h expression of dCas9 in the presence of a guide RNA with (A) an AGGAA seed or (B) an ACCCA seed sequence, measured by RNA-seq. (C) Diagram of the *glyQS* and *rpmH* promoters, indicating the location of potential 5-nt off-target sites. The genotype of the PAM^{mut} strain is indicated. (D) Serial dilution and spotting of strain AV01 (WT) or TG01 (PAM^{mut}) expressing dCas9 under the control of a p_{tet} promoter in the presence of various guide RNAs: LacZ (negative control), AGGAA, CGGAA, GGGAA, TGGAA and ACCCA guides with random sequences in the first 15 nt of the guide. (E) Expression of the *glyQ* gene measured by RT-qPCR in strain AV01 (WT) in the presence of all four NGGAA guide RNAs after 2 h of dCas9 induction. Points show three biological replicates. (F) Violin plot of log₂FC values for guides sharing each of the four NGGAA seed sequences. (G) Serial dilution and spotting of strain AV01 carrying a plasmid expressing *rpmH* (pTG40) or a control empty vector (pSEVA271), in the presence of the ACCCA guide RNA.

mation of colonies in the presence of the AGGAA bad seed, but colonies still appeared smaller than those in the presence of the control guide (Supplementary Figure S4). Upon examining the promoter of the *glyQS* operon, we noticed that it contained a possible off-target binding site for AGGAA guides overlapping the transcription start site (Figure 2C). Therefore, we mutated the PAM site of this off-target from NGG to NCG on plasmid pTG36, resulting in plasmid pTG37. Complementation with plasmid pTG37 fully abolished the AGGAA bad-seed effect (Supplementary Figure S4).

To unequivocally demonstrate that the AGGAA bad-seed effect is caused by off-target binding of dCas9 in the promoter of *glyQS*, we used oligonucleotide recombineering to introduce a mutation in the PAM of the candidate off-target position directly in the chromosome of *E. coli* (32). The resulting strain, TG01, did not exhibit any growth defects when dCas9 expression was induced in the presence of the AGGAA guide RNA (Figure 2D). As expected, a guide carrying the ACCCA toxic seed sequence remained toxic in the TG01 strain.

We then tested whether off-target binding to the promoter of *glyQS* could also explain the toxicity of the other seed sequences ending in GGAA. We measured *glyQS* repression by qPCR and observed moderate to strong silencing with all NGGAA seed sequences (Figure 2E). The

silencing levels obtained are in good agreement with the log₂FC values of guides sharing each of the four NGGAA seed sequences in the screen described above (Figure 2F). In addition, the fitness defect observed for these seed sequences was abolished in strain TG01 (Figure 2D). Altogether, these results demonstrate that the toxicity of all four NGGAA seeds comes from dCas9 binding to an off-target position in the promoter of the *glyQS* operon with as little as 4 nt of complementarity.

The ACCCA bad seed showed repression of the essential *rpmH-rnpA* operon by ~4-fold in the RNA-seq data (Figure 2B). As with AGGAA, we observed a potential off-target site in the promoter of this operon, this time located downstream of the transcriptional start (Figure 2C). The expression of the *rpmH-rnpA* operon from plasmid pTG40 rescued the negative effect of the ACCCA bad seed (Figure 2G). Despite multiple attempts, we were unable to introduce a point mutation in the PAM motif of this off-target on the *E. coli* chromosome, indicating that this position in the *rpmH-rnpA* promoter may be critical for function.

We then sought to determine whether other highly toxic seed sequences could be explained by off-target binding to the promoter of essential genes. To do this, we compiled a list of candidate off-targets for the most detrimental seeds (Supplementary Table S14). We then attempted to validate some candidate off-target using complementa-

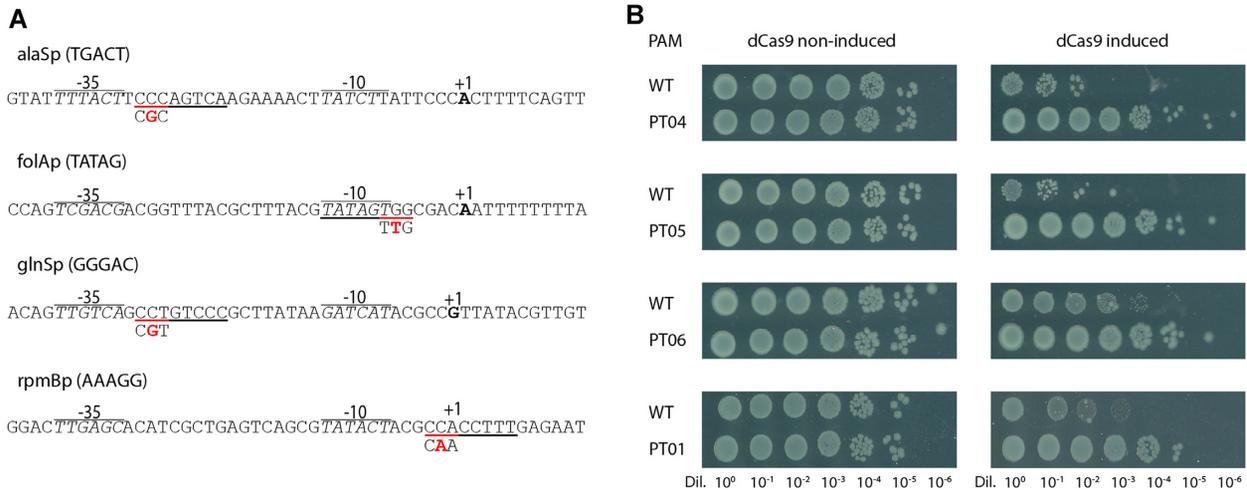


Figure 3. Identification of off-target sites responsible for the toxicity of a few bad-seed sequences. (A) The promoters of *alaS*, *folA*, *glnS* and *rpmB* are shown and the off-target positions identified for the TGACT, TATAG, GGGAC and AAAGG bad seeds are underlined. Mutations made to the PAM site are shown below the promoter sequence in red. (B) Serial dilution and spotting of strain AV01 (WT) or of the PAM site mutants depicted in panel (A), in the presence of the cognate guide RNA and in the absence or presence of aTc.

tion or targeted mutation of the PAM, as described above. This led to the identification of off-target positions responsible for the toxicity of six seeds: TGACT (*alaS*), AAAGG (*rpmB*), GAGGC (*rpmH*), CACTC (*rpmH*), TATAG (*folA*) and GGGAC (*glnS*) (see Figure 3 and Supplementary Figure S5).

Interestingly, all of the identified off-target positions are in the promoter of genes involved directly or indirectly in protein synthesis: RpmB and RpmH are ribosomal proteins, AlaS, GlnS and GlyQS are aminoacyl-tRNA synthetases, and FolA is involved in the synthesis of tetrahydrofolate, a metabolite required for the synthesis of amino acids and nucleic acids. Some of these genes (*glyQS*, *rpmB* and *alaS*) were previously identified in a CRISPRi screen as causing a fitness defect even when weakly repressed (35).

We then compared the off-target positions identified within these promoters to assess whether silencing with as little as 5 nt of identity requires specific binding positions and orientations. In five out of eight cases (*glyQS*, *folA*, *glnS*, *rpmB*, *rpmH-GAGGC*), the PAM was located between the -10 element and the transcriptional start. In the case of *alaS* and *rpmH-CACTC*, the off-target PAM is just downstream of the -35 element, while in the case of *rpmH-ACCCA* it is 16 nt downstream of the transcriptional start. The off-targets are oriented such that the guide will bind to the top strand of the promoter in six out of eight cases.

Bad seeds cannot not be predicted from genomic data

Our findings demonstrate that dCas9 binding to a gene promoter using only 4–5 nt can effectively silence gene expression. However, it is worth noting that all 5-nt sequences with a correct 5'-NGG-3' adjacent motif are commonly found in the *E. coli* genome, with a median of 451 occurrences. In fact, 86% of all possible seed sequences have at least one such occurrence within the -35 and +16 positions of an essential gene promoter. Yet, only ~7% of seed sequences

were found to cause a significant fitness defect in our dataset (XL-mHG test, P -value $< 10^{-14}$) (36). This suggests that 5 nt of identity can direct dCas9 to silence a promoter with sufficient strength for it to affect growth only under certain conditions.

To gain further insight into these circumstances, we examined other possible binding positions of dCas9 within the promoters identified above. All these promoters contained alternative binding positions for dCas9 with a 5'-NGG-3' PAM, either on one strand or on the other. While some of these positions are associated with toxic seed sequences such as 5'-CACTC-3' and 5'-GAGGC-3' in the promoter of *rpmH*, most are not associated with toxic seeds, as seen in our screening data (Supplementary Figure S6). For instance, the TTGAC seed has an off-target just 1 nt away from the toxic TGACT seed in the promoter of *alaS*, but is not toxic itself (Supplementary Figure S6).

Based on this information, we can conclude that while the toxicity of bad-seed sequences can be explained by off-target silencing of essential genes, dCas9 binding to a 5-nt sequence in the promoter of an essential gene is not sufficient to reduce its expression to a detrimental level in most cases. The bad-seed sequences must therefore have specific properties in terms of composition or position within the promoter of genes. Additionally, it is likely that only certain promoters are affected by dCas9 binding mediated by just 5 nt of homology.

ChIP-seq identifies pervasive binding of dCas9 to off-target positions

To investigate whether dCas9 binding to off-target positions responsible for toxicity was particularly strong, we performed ChIP-seq assays using FLAG-tagged dCas9 in the presence of the AGGAA or ACCCA guides. The results revealed pervasive binding of dCas9 at off-target positions throughout the genome. In the presence of the

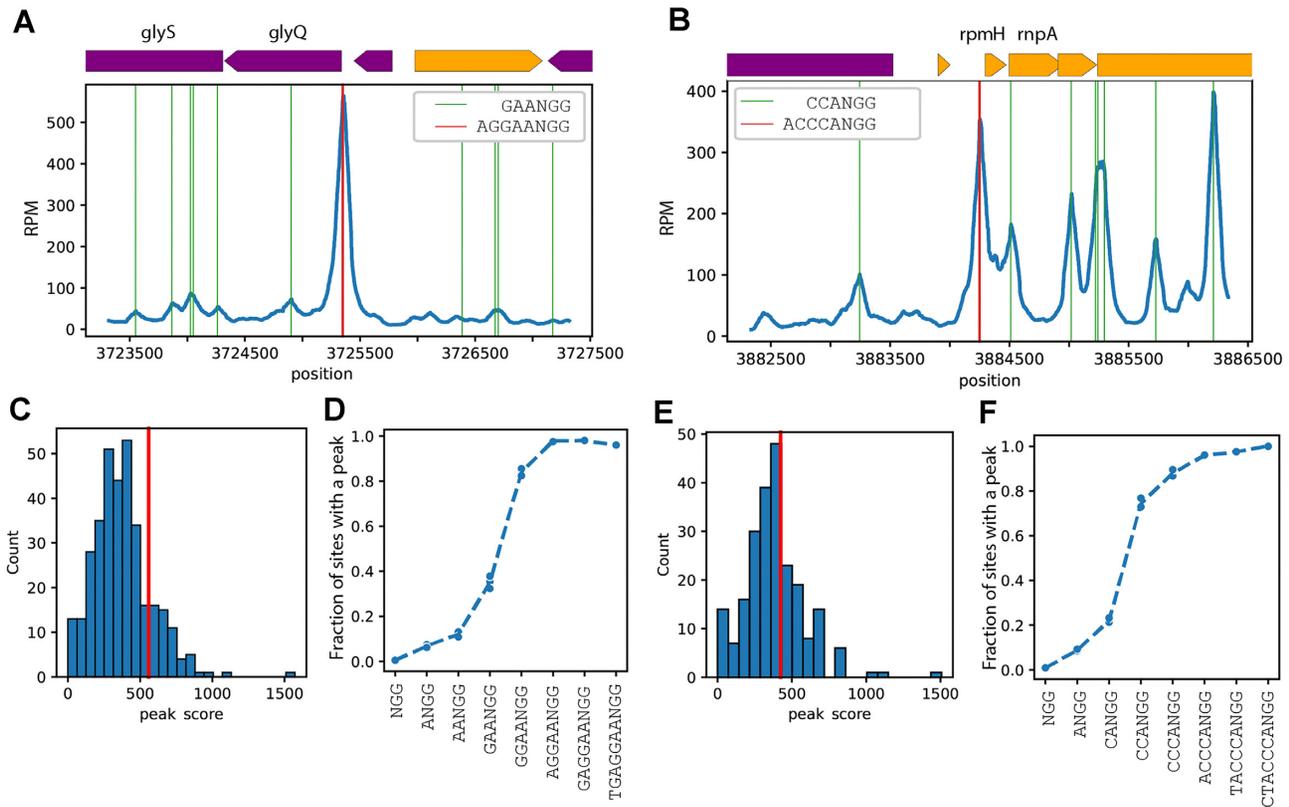


Figure 4. ChIP-seq analysis of dCas9 binding to the chromosome of *E. coli* in the presence of the toxic AGGAA or ACCCA guide RNAs. Normalized read counts showing the binding of dCas9 to (A) the off-target position in the promoter of *glyQS* in the presence of the AGGAA guide RNA or (B) the off-target position in the promoter of *rpmH* in the presence of the ACCCA guide RNA. (C) Distribution of peak scores for all peaks over AGGAANGG positions in the chromosome of *E. coli*. The score of the *glyQS* off-target is shown with a red bar. (D) Fraction of genome locations for which a ChIP peak is detected. The data are shown for locations with an increasing complementarity to the seed sequence of the AGGAA.n1 guide. Points show three biological replicates. (E) Distribution of peak scores for all peaks over ACCCANGG positions in the chromosome of *E. coli*. The score of the *rpmH* off-target is shown with a red bar. (F) Fraction of genome locations for which a ChIP peak is detected. The data are shown for locations with an increasing complementarity to the seed sequence of the ACCCA.n1 guide. Points show three biological replicates.

AGGAA guide, ChIP-seq peaks could be identified at 97% of all AGGAANGG positions including the off-target site in the *glyQS* promoter, which was among the 20% stronger peaks (Figure 4A, C and D). Likewise, in the presence of the ACCCA guide, ChIP-seq peaks could be identified at 96% of all AGGAANGG positions. However, the peak that mapped to the off-target site in the *rpmH* promoter was of average strength (Figure 4B, E and F). These results indicate that dCas9 binding to positions responsible for off-target effects is not particularly strong. In addition, off-target binding of dCas9 could be observed at most 4-nt off-target positions and large fraction of 3-nt positions (Figure 4A, B, D and F).

A corollary of these observations is that many genes may be affected by off-target effects, beyond the essential genes identified through their impact on growth. Our RNA-seq data indeed provide two examples of this. The nonessential *hdhA* gene was strongly repressed in the presence of the AGGAA guide RNA, and the *yfcZ* gene was strongly repressed in the presence of the ACCCA guide RNA (Figure 2B). Both genes carry an off-target binding position in their promoter that could explain this silencing (Supplementary Figure S7).

dCas9 silencing can occur with a wide range of seed sequences

To further investigate the sequence determinants of how dCas9 binding to 5-nt targets can mediate silencing of a promoter, we sought to replace the AGGAA sequence in the *glyQS* promoter with other sequences. Recombineering was used to replace this sequence in the chromosome of *E. coli* using a pool of 1024 oligonucleotides containing all possible bases at these five positions. Successful recombination events were selected by expressing the AGGAA guide RNA to counterselect WT clones. From this experiment, 29 different promoter sequences were isolated. The essential nature of the *glyQS* operon likely restricts the sequences obtained to sequences that preserved the expression of *glyQS*. We then designed guide RNAs targeting each of these sequences with just 5 nt of identity and evaluated their toxicity when expressed from the psgRNA in the presence of the cognate modified *glyQS* promoter or the WT promoter. In one case, the AGGAA sequence was replaced by another known toxic seed sequence (TATAG), and the cognate guide was thus toxic in both the modified and WT strains. In all other cases, the guide RNA became toxic only when the matching seed was carried by the *glyQS* promoter, with phenotypes ranging from a reduction in colony size to

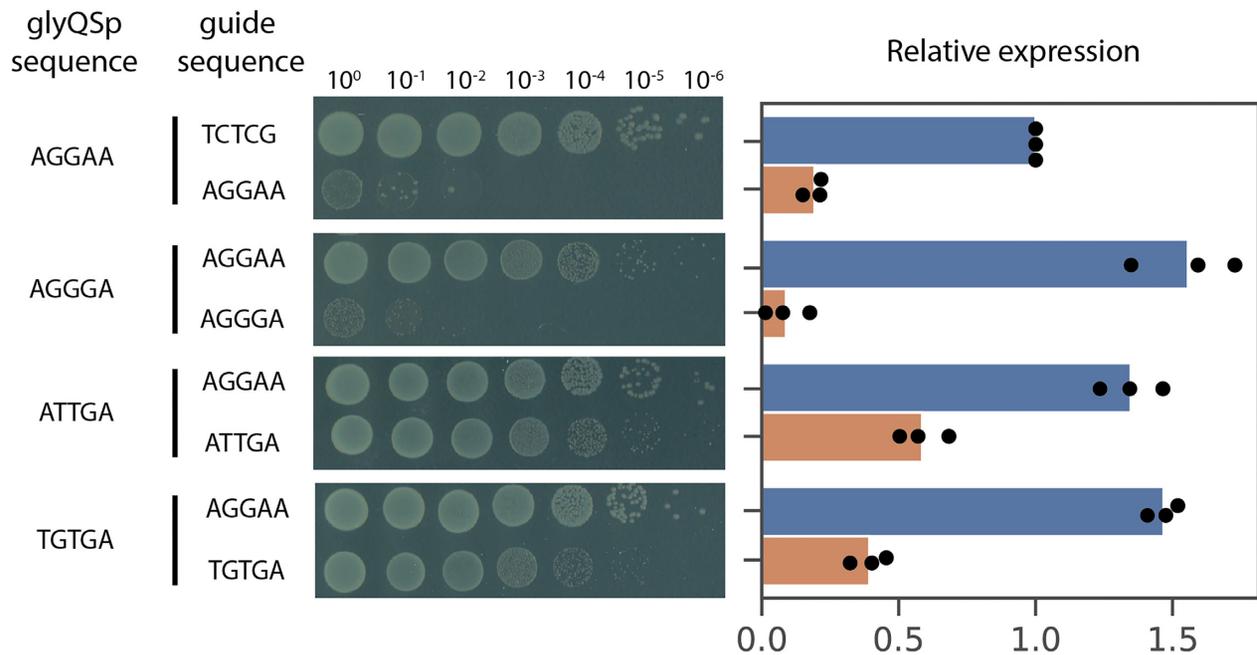


Figure 5. Off-target binding of dCas9 to multiple seed sequences in modified variants of the GlyQS promoter leads to silencing and toxicity. The AGGAA sequence in the promoter of *glyQS* was mutated to AGGGA, ATTGA or TGTGA in strain AV01. Plates show serial dilution and spotting in the presence of guide RNA matching the glyQSp sequence or not. Relative expression of *glyQ* was measured by RT-qPCR. Points show biological replicates ($n = 3$).

a strong reduction in the number of CFUs (Figure 5 and Supplementary Figure S8).

These results show that all the tested seed sequences can direct dCas9 binding with sufficient strength to reduce the expression of the *glyQS* promoter. We quantified the repression obtained with three seeds using RT-qPCR (Figure 5). One of the seed sequences that could be made toxic by the modification of the *glyQS* promoter was TAATA (Supplementary Figure S8), suggesting that the strength of the Watson–Crick interactions between the guide and the target is not a direct determinant of the ability of dCas9 to silence an off-target promoter.

Identification of toxic seed sequences in various enterobacteria

Our results suggest that toxic seed sequences may vary among different bacterial strains and species due to changes in gene essentiality and promoter sequences. To test this, we introduced our library of random guide RNA into eight *E. coli* strains as well as in other enterobacteria species: *E. albertii*, *E. fergusonii*, *K. pneumoniae* and *C. freundii*. We then induced dCas9 for ~20 generations and sequenced the library before and after induction (Figure 6A). Guides were grouped based on their 5-nt seed sequences, and the median for each group calculated. Toxic seed sequences could be identified across samples, but some strains displayed more bad seeds and with a more pronounced effect than others (Supplementary Figure S9). These differences are likely due to differences in the expression levels of dCas9 among strains. Interestingly, we found that some patterns, such as TATAG, were toxic in most tested strains except a few, while others (e.g. GATGG) were only depleted in one or a few strains (Figure 6A).

For several seed sequences, changes in toxicity could be attributed to genomic changes. The TGACT seed had previously been identified as targeting the *alaS* promoter, a sequence that is mutated in *K. pneumoniae* and *C. freundii*. In the former, the PAM is mutated and the target sequence is changed to TTACG, while in the latter the target sequence is changed to TTACT (Figure 6B). Guides with a TGACT seed are nontoxic in both strains, while the TTACT seed was toxic in *C. freundii* (Figure 6C). The TTACG seed is not toxic in any strain including *K. pneumoniae*, as expected given the PAM mutation. Other examples are the ACCCA target, where a mutation in the PAM abolishes bad-seed toxicity in *K. pneumoniae* and *C. freundii* (Figure 6D and E), and TATAG, which is nontoxic in *E. fergusonii*, the only strain with an alternative dihydrofolate reductase gene (*dfrAI*) (37), which has a different promoter and no target site for TATAG (Figure 6F and G). For some 5-nt patterns, candidate binding sites could be identified based on the presence of the 5-nt + PAM pattern in the promoter of essential genes and the disappearance of the bad-seed effect in strains where this sequence is mutated (Supplementary Figure S10).

DISCUSSION

In this study, we demonstrate that dCas9 can effectively silence genes at off-target positions in bacterial chromosomes with as little as 4 nt of identity to the seed sequence. When binding occurs in the promoter of an essential gene, this can result in toxicity. We measured a repression of up to ~26-fold when silencing was mediated by 5 nt, and a repression of ~2-fold with 4 nt of complementarity. Although the repression obtained with fully matched guides is typically much stronger, our findings show that weak silencing

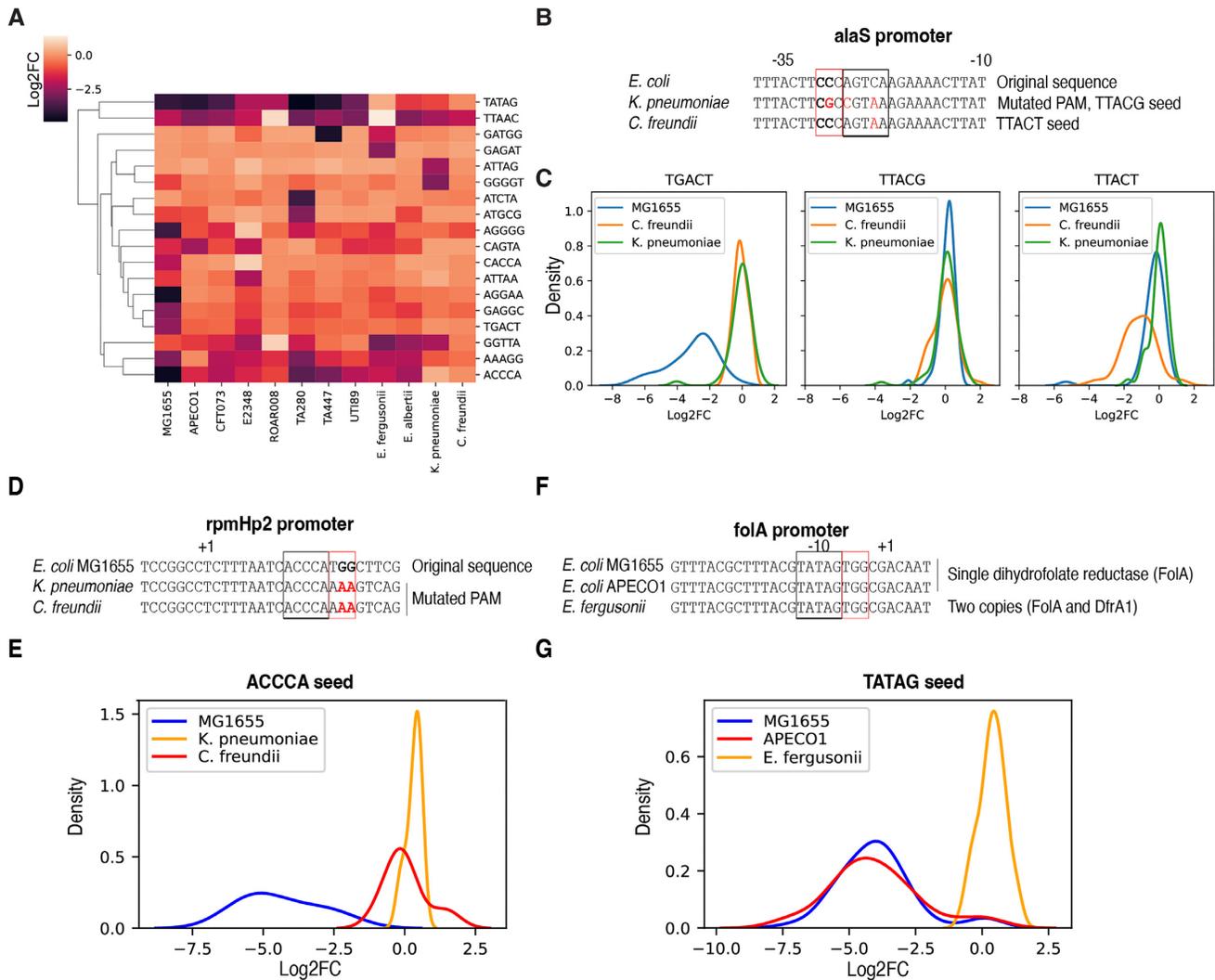


Figure 6. A screen performed in 12 Enterobacteriaceae identifies changes in toxic seed sequences. (A) Median \log_2FC of most toxic 5-nt seed sequences (dendrogram of hierarchical clustering shown on the left). (B) Sequence of *alaS* promoter in three strains. (C) Distribution of \log_2FC of all guide RNAs with a TGACT or TTACT seed. (D) Sequence of RpmHp2 in three strains. (E) Distribution of \log_2FC of guide RNAs with an ACCCA seed. (F) Sequence of *folA* promoter in three strains. (G) Distribution of \log_2FC of guide RNAs with a TATAG seed.

mediated by short complementarity to the seed sequence can have a significant impact on the growth of *E. coli*. This is particularly true for a subset of essential genes whose moderate silencing can have a large fitness effect. While we now provide a mechanism for the bad-seed toxicity phenomenon, it remains currently impossible to accurately predict which gene will be silenced by any given guide and to what level. In addition, off-target silencing mediated by this mechanism is not limited to essential genes, and we expect that many guides will perturb the expression of genes when Cas9 is expressed at high levels. This could of course be problematic in the many situations where silencing an off-target nonessential gene could impact the phenotype of interest. In a previous study, we showed how the bad-seed effect could be mitigated by optimizing the expression level of dCas9 while maintaining a good on-target activity (23). Our findings now further highlight the importance of such optimization step when designing novel Cas9 tools.

While optimizing the expression level of dCas9 greatly reduces off-target effects, they can still occur as we previously showed (23). Researchers should therefore also rely on consistent results obtained from multiple guide RNAs targeting the same gene before inferring its role in a phenotype of interest.

Beyond its relevance to CRISPRi experiments in bacteria, the off-target binding of Cas9 to positions with as little as 4 nt of complementarity to the guide could lead to undesired effects for other CRISPR–Cas9 technologies. In genome editing applications, off-target silencing might interfere with the use of Cas9 to edit some target positions. Most importantly, CRISPR–Cas9 techniques that only rely on Cas9 binding and not on its nuclease activity, such as some variations of the base-editing technology or transcriptional activation strategies, could be impacted by the extensive off-target binding of dCas9. In this context, it may be desirable to work with Cas9 variants that have stricter PAM

sequence requirements to limit the number of off-target positions. This is in contrast to the current efforts in the field of Cas9 engineering that have mostly focused on relaxing the PAM constraint (38).

Beyond the relevance of our findings for CRISPR–Cas9 technologies, it is interesting to hypothesize that off-target silencing could have implications for natural systems as well. We observed that the bad-seed toxicity phenomenon occurs with both Cas9 and dCas9 (Supplementary Figure S11). This is consistent with previous observations that mismatches in the PAM-distal region do not disrupt Cas9 binding but prevent cleavage (8,39). While our study was performed in an artificial setup in which Cas9 is overexpressed, recent work has identified natural cases of gene regulation mediated by Cas9. For example, Cas9 naturally downregulates its own expression through binding to its promoter mediated by a long form of the tracrRNA (tracr-L) that acts as a natural sgRNA (40). The interaction between the tracr-L and its target sequence is only mediated by 11 nt of identity (40). Additionally, degenerate forms of CRISPR RNAs, called scaRNAs, have been identified that guide Cas9 to silence endogenous genes through an interaction mediated by short homologies between the scaRNA and the promoter (41). It is tempting to hypothesize that scaRNAs evolved from the acquisition of novel spacers with serendipitous complementarity to endogenous genes of the bacterium, whose downregulation by Cas9 was then selected. However, we can assume that most spacer acquisition events leading to serendipitous downregulation of endogenous genes would be deleterious and counterselected, potentially representing a form of autoimmune reaction and cost of CRISPR–Cas9 immunity.

DATA AVAILABILITY

RNA-seq and ChIP-seq data have been deposited at GEO with the following number: GSE217006. CRISPRi screen read counts are provided in Supplementary Table S12 and log₂FC values are provided in Supplementary Table S13.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Erick Denamur for providing *E. coli* strains, and Alvaro San Milan and Jeronimo Rodríguez-Beltrán for providing *C. freundii* CF12.

FUNDING

European Research Council [677823]; European Research Council [101044479]; Agence Nationale de la Recherche [ANR-10-LABX-62-IBEID]. Funding for open access charge: Institut Pasteur.

Conflict of interest statement. None declared.

REFERENCES

- Vigouroux, A. and Bikard, D. (2020) CRISPR tools to control gene expression in bacteria. *Microbiol. Mol. Biol. Rev.*, **84**, e00077–19.
- Luo, M.L., Leenay, R.T. and Beisel, C.L. (2016) Current and future prospects for CRISPR-based tools in bacteria. *Biotechnol. Bioeng.*, **113**, 930–943.
- Adli, M. (2018) The CRISPR tool kit for genome editing and beyond. *Nat. Commun.*, **9**, 1911.
- Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR–Cas systems. *Nat. Biotechnol.*, **31**, 233–239.
- Wang, K., Fredens, J., Brunner, S.F., Kim, S.H., Chia, T. and Chin, J.W. (2016) Defining synonymous codon compression schemes by genome recoding. *Nature*, **539**, 59–64.
- Choudhury, A., Fenster, J.A., Fankhauser, R.G., Kaar, J.L., Tenaillon, O. and Gill, R.T. (2020) CRISPR/Cas9 recombining-mediated deep mutational scanning of essential genes in *Escherichia coli*. *Mol. Syst. Biol.*, **16**, e9265.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
- 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 Res.*, **41**, 7429–7437.
- Rousset, F. and Bikard, D. (2020) CRISPR screens in the era of microbiomes. *Curr. Opin. Microbiol.*, **57**, 70–77.
- Peters, J.M., Koo, B.-M., Patino, R., Heussler, G.E., Hearne, C.C., Qu, J., Inclan, Y.F., Hawkins, J.S., Lu, C.H.S., Silvis, M.R. *et al.* (2019) Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nat. Microbiol.*, **4**, 244.
- Ameruso, A., Villegas Kcam, M.C., Cohen, K.P. and Chappell, J. (2022) Activating natural product synthesis using CRISPR interference and activation systems in *Streptomyces*. *Nucleic Acids Res.*, **50**, 7751–7760.
- Ho, H.-I., Fang, J.R., Cheung, J. and Wang, H.H. (2020) Programmable CRISPR–Cas transcriptional activation in bacteria. *Mol. Syst. Biol.*, **16**, e9427.
- Liu, Y., Wan, X. and Wang, B. (2019) Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. *Nat. Commun.*, **10**, 3693.
- Klanschnig, M., Cserjan-Puschmann, M., Sriedner, G. and Grabherr, R. (2022) CRISPRactivation-SMS, a message for PAM sequence independent gene up-regulation in *Escherichia coli*. *Nucleic Acids Res.*, **50**, 10772–10784.
- Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I. and Liu, D.R. (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.
- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420–424.
- Anzalone, A.V., Koblan, L.W. and Liu, D.R. (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.*, **38**, 824–844.
- Rock, J.M., Hopkins, F.F., Chavez, A., Diallo, M., Chase, M.R., Gerrick, E.R., Pritchard, J.R., Church, G.M., Rubin, E.J., Sasseti, C.M. *et al.* (2017) Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat. Microbiol.*, **2**, 16274.
- Wendt, K.E., Ungerer, J., Cobb, R.E., Zhao, H. and Pakrasi, H.B. (2016) CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Microb. Cell Fact.*, **15**, 115.
- Jiang, Y., Qian, F., Yang, J., Liu, Y., Dong, F., Xu, C., Sun, B., Chen, B., Xu, X., Li, Y. *et al.* (2017) CRISPR–Cpf1 assisted genome editing of *Corynebacterium glutamicum*. *Nat. Commun.*, **8**, 15179.
- Lee, Y.J., Hoynes-O'Connor, A., Leong, M.C. and Moon, T.S. (2016) Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Res.*, **44**, 2462–2473.
- Cho, S., Choe, D., Lee, E., Kim, S.C., Palsson, B. and Cho, B.-K. (2018) High-level dCas9 expression induces abnormal cell morphology in *Escherichia coli*. *ACS Synth. Biol.*, **7**, 1085–1094.
- Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V. and Bikard, D. (2018) A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nat. Commun.*, **9**, 1912.

24. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62–67.
25. Jones, D.L., Leroy, P., Unoson, C., Fange, D., Ćurić, V., Lawson, M.J. and Elf, J. (2017) Kinetics of dCas9 target search in *Escherichia coli*. *Science*, **357**, 1420–1424.
26. Kuscü, C., Arslan, S., Singh, R., Thorpe, J. and Adli, M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.*, **32**, 677–683.
27. Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S. *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.*, **32**, 670–676.
28. Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., Rocha, E.P.C. and Bikard, D. (2021) The impact of genetic diversity on gene essentiality within the *Escherichia coli* species. *Nat. Microbiol.*, **6**, 301–312.
29. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
30. Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
31. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
32. Wannier, T.M., Nyerges, A., Kuchwara, H.M., Czikkely, M., Balogh, D., Filsinger, G.T., Borders, N.C., Gregg, C.J., Lajoie, M.J., Rios, X. *et al.* (2020) Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl Acad. Sci. U.S.A.*, **117**, 13689–13698.
33. Chung, C.T., Niemela, S.L. and Miller, R.H. (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl Acad. Sci. U.S.A.*, **86**, 2172–2175.
34. Ferrières, L., Hemery, G., Nham, T., Guerout, A.M., Mazel, D., Beloin, C. and Ghigo, J.M. (2010) Silent mischief: bacteriophage mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.*, **192**, 6418–6427.
35. Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F. and Bikard, D. (2018) Genome-wide CRISPR–dCas9 screens in *E. coli* identify essential genes and phage host factors. *PLoS Genet.*, **14**, e1007749.
36. Wagner, F. (2017) The XL-mHG test for gene set enrichment. PeerJ Preprints doi: <https://doi.org/10.7287/peerj.preprints.1962v3>, 12 February 2017, preprint: not peer reviewed.
37. Ambrose, S.J. and Hall, R.M. (2021) dfrA trimethoprim resistance genes found in Gram-negative bacteria: compilation and unambiguous numbering. *J. Antimicrob. Chemother.*, **76**, 2748–2756.
38. Collias, D. and Beisel, C.L. (2021) CRISPR technologies and the search for the PAM-free nuclease. *Nat. Commun.*, **12**, 555.
39. 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.
40. Workman, R.E., Pammi, T., Nguyen, B.T.K., Graeff, L.W., Smith, E., Sebald, S.M., Stoltzfus, M.J., Euler, C.W. and Modell, J.W. (2021) A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR–Cas expression. *Cell*, **184**, 675–688.
41. Ratner, H.K., Escalera-Maurer, A., Le Rhun, A., Jaggavarapu, S., Wozniak, J.E., Crispell, E.K., Charpentier, E. and Weiss, D.S. (2019) Catalytically active Cas9 mediates transcriptional interference to facilitate bacterial virulence. *Mol. Cell*, **75**, 498–510.