Published online 13 October 2016 *Nucleic Acids Research, 2016, Vol. 44, No. 22* ***10849–10861***
doi: 10.1093/nar/gkw914

Altered stoichiometry *Escherichia coli* Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation

Konstantin Kuznedelov[1], Vladimir Mekler[1], Sofia Lemak[2], Monika Tokmina-Lukaszewska[3], Kirill A. Datsenko[4], Ishita Jain[1], Ekaterina Savitskaya[5,6], John Mallon[7], Sergey Shmakov[5,8], Brian Bothner[3], Scott Bailey[7], Alexander F. Yakunin[2], Konstantin Severinov[1,5,6,9,*] and Ekaterina Semenova[1,*]

[1]Waksman Institute of Microbiology, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA, [2]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario M5S 3E5, Canada, [3]Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT 59717, USA, [4]Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA, [5]Skolkovo Institute of Science and Technology, Skolkovo 143025, Russia, [6]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia, [7]Department of Biochemistry and Molecular Biology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21205, USA, [8]NCBI, NLM, NIH, Bethesda, MD 20894, USA and [9]Peter the Great St Petersburg State Polytechnic University, St. Petersburg, 195251, Russia

Received September 02, 2016; Revised September 29, 2016; Accepted October 03, 2016

ABSTRACT

The *Escherichia coli type* I-E CRISPR-Cas system Cascade effector is a multisubunit complex that binds CRISPR RNA (crRNA). Through its 32-nucleotide spacer sequence, Cascade-bound crRNA recognizes protospacers in foreign DNA, causing its destruction during CRISPR interference or acquisition of additional spacers in CRISPR array during primed CRISPR adaptation. Within Cascade, the crRNA spacer interacts with a hexamer of Cas7 subunits. We show that crRNAs with a spacer length reduced to 14 nucleotides cause primed adaptation, while crRNAs with spacer lengths of more than 20 nucleotides cause both primed adaptation and target interference *in vivo*. Shortened crRNAs assemble into altered-stoichiometry Cascade effector complexes containing less than the normal amount of Cas7 subunits. The results show that Cascade assembly is driven by crRNA and suggest that multisubunit type I CRISPR effectors may have evolved from much simpler ancestral complexes.

INTRODUCTION

The CRISPR-Cas systems comprise of Clustered Regular Interspaced Short Palindromic Repeats separated by spacers (CRISPR arrays) and *cas* (CRISPR Associated) genes (1). CRISPR-Cas systems provide prokaryotes with small RNA-based adaptive immunity against foreign DNA and RNA (2). Defensive CRISPR RNAs (crRNAs) are generated after processing of primary (pre-crRNA) CRISPR array transcripts (3,4). The pre-crRNA processing is catalyzed by either Cas6 in Type I and III systems (3,5,6); a small RNA and a cellular non-Cas RNAse in Type II systems (7), or the Cas effector nuclease in Type V systems (8). Processing generally occurs at the repeat segments of pre-crRNA, generating short RNAs containing a central guide segment with flanking fragments of repeats (4). In some systems, additional processing that removes repeat fragments at one side and even shortens the spacer part of crRNA has been documented (7,9–12). In Type I systems mature crRNA contains CRISPR repeat fragments on both sides and specific Cas proteins of the multisubunit effector complex Cascade bind to repeat sequences at both sides of the spacer. These proteins are localized on opposing sides of the elongated effector complex molecule (13). The spacer part remains exposed and extended on the surface of homooligomer of Cas proteins located in the central part of the effector complex. Each monomer of such proteins (Cas7 in the case of the *Escherichia coli* Type I-E Cascade complex) binds to an equally sized segment of spacer RNA (14–16). The stoichiometry of the Cas7 oligomer and the length of crRNA spacer must therefore correspond to each other to fulfill the architectural constraints of the effector

*To whom correspondence should be addressed. Tel: +1 848 445 6096; Fax: +1 848 445 5735; Email: severik@waksman.rutgers.edu
Correspondence may also be addressed to Ekaterina Semenova. Tel: +1 848 445 6096; Fax: +1 848 445 5735; Email: semenova@waksman.rutgers.edu

complex. While the length of the spacers in CRISPR arrays from the same organisms is kept largely constant (with only minor variation of 1–3 base pairs), the variability of spacer lengths in CRISPR arrays of systems belonging to the same type and even subtype but from different organisms can be substantial (17). In principle, the oligomeric structure of spacer-binding subunits of Type I (and Type III) effectors should allow to accommodate crRNAs with different length spacers. In fact the spacer length may dictate the oligomerization state of such subunits, leading to production of effector complexes with different subunit stoichiometries and functional properties. Indeed, a recent report showed that the *E. coli* crRNA spacer, normally 32-nt long, can be extended leading to formation of functional Cascade complexes with extra Cas7 monomers (18). Functional Cascades with longer crRNA spacers were also reported for a Type I-F system from *Shewanella putrefaciens*, however, shortening of spacer rendered effector complexes inactive (19). Here, we show that the crRNA spacer length can be significantly decreased for *E. coli* Cascade. *In vitro* experiments and direct measurements of subunit stoichiometry show that shortening of crRNA leads to progressive loss of Cas7 monomers from Cascade. Cascades containing crRNA shortened by as much as 12 nucleotides are capable of CRISPR interference. Removal of as many as 18 spacer nucleotides is tolerated during primed acquisition, which requires weak interaction of Cascade-crRNA with target DNA and appears to be a much more sensitive test for Cascade function. Our results shed light on the possible paths of evolution of multisubunit CRISPR-Cas effectors and have a potential to expand existing bioengineering tool set by creating Cascades recognizing shorter targets.

MATERIALS AND METHODS

Strains and plasmids

Strains and plasmids used in this work are listed in Supplementary Tables S1 and S2, respectively. The strains producing crRNA with altered spacer length are derivatives of KD263 (20) and were constructed using a technique based on the previously published Red recombinase protocol (21). Synthetic dsDNA fragments (gBlocks from IDT Inc.) containing CRISPR arrays with appropriate spacer lengths were used for recombination.

The target plasmids pG8 and pG8mut are pT7blue-based plasmids containing, respectively, a wild-type protospacer (5′-CTGTCTTTCGCTGCTGAGGGTGACGA TCCCGC) and preceding ATG PAM and a mutant variant carrying the C1T substitution at the first position of the protospacer (22). The protospacer is followed by a 5′-ATGTAT-3′ sequence matching the extension in the +6 spacer.

For *in vitro* studies of Cascade complexes assembled on crRNAs with altered-length spacer, the pCDF-based plasmids co-expressing *cse2, cas7, cas5 and cas6e* genes and appropriate CRISPR arrays were generated. DNA fragments containing CRISPR arrays with altered spacer length were PCR amplified using genomic DNA isolated form appropriate strains and the following primers: XhoI-forward (5′-ACCCCTCGAGATTTGG ATGGTTTAAGGTTGGTG-3′) and AvrII-reverse (5′-AGGCCTAGGCGAAGGCGTCTTGATGGG-3′) and cloned into XhoI and AvrII sites of the pCDF-casBCDE plasmid (23).

RNA extraction and Northern blotting

The procedure was performed as described previously (23). ^{32}P-end labeled probe (5′-GCAGCGAAAGACAGCGGT TTAT-3′) complementary to the 5′-part of all crRNA spacer-length variants was used for hybridization.

CRISPR interference and adaptation assays

E. coli strain KD263 (20) carrying a genomic CRISPR array with 32-nucleotide (wt) spacer, strain KD390 (23) harboring a single repeat in CRISPR array or derivative strains (KD692, KD696, KD698, KD700, KD702) harboring CRISPR arrays containing altered-length spacers were used to determine cell sensitivity to M13 phage infection by a spot test method as described (24). Efficiency of plaquing was calculated as a ratio of the number of plaques formed on a lawn of tested cells to the number of plaques on sensitive (non-targeting, KD390) cell lawn. For each strain, plaquing efficiency was determined in at least three independent experiments.

For plasmid transformation efficiency assay, competent pre-induced cells were prepared and electroporated with 10 ng of pG8, pG8mut or control pT7blue plasmids. Transformation efficiency was determined as ampicillin-resistant colony numbers per microgram DNA. Mean values and standard deviations were obtained from three independent experiments.

To analyze plasmid loss, individual colonies of uninduced cells transformed with pG8 or pG8mut plasmids were inoculated in 3 ml LB broth supplemented with 100 μg/ml ampicillin for overnight growth at 37°C. Aliquots of overnight cultures were diluted 100-times into fresh LB without antibiotic and allowed to grow until the culture OD_{600} reached 0.5. At this point half of the culture was induced with 1 mM arabinose and 1 mM isopropyl-ß-D-thiogalactopyranoside (IPTG). Another half of the culture remained uninduced. Plasmid DNA was purified from 1.5-ml aliquots withdrawn at various times post-induction using Thermo Scientific GeneJET Plasmid Miniprep Kit. Purified plasmids were analyzed on 0.9% agarose gels.

To monitor CRISPR adaptation, KD263 or strains producing crRNA with altered spacer length transformed with the pG8, pG8mut or pT7blue plasmids were grown overnight at 37°C in LB broth supplemented with 100 μg/ml ampicillin. Aliquots of cultures were diluted 100-fold with LB broth without ampicillin, allowed to grow until OD_{600} 1.0 and then supplemented with 1 mM arabinose and 1 mM IPTG and grown overnight. To monitor spacer acquisition 1 μl of culture was added to 20 μl polymerase chain reaction (PCR) with primers Ec_LDR-F (5′-AAGGTTGGTGGGTTGTTTTTATGG-3′) and Ec_minR (5′- CGAAGGCGTCTTGATGGGTTTG-3′). PCR products corresponding to expanded CRISPR cassettes were gel purified using QIAquick Gel Extraction Kit (QIAGEN)

and sequenced with MiSeq Illumina System at Moscow State University Genomics facility.

High throughput sequence analysis

Raw sequencing data were analyzed using ShortRead and BioStrings packages (25,26). Illumina-sequencing reads were filtered for quality scores of ≥20 and reads containing two repeats (with up to two mismatches) were selected. Intervening sequences were considered as spacers. Spacers were next mapped on the pG8 or pG8mut plasmids with no mismatches allowed. R scripts and their package ggplot2 (27) were used for spacers statistics. Circle histograms of spacer mapping were obtained with EasyVisio tool created by E. Rubtsova.

Cascade expression and purification

Cascade subcomplexes lacking Cse1 were prepared from *E. coli* KD418 cells (28) co-expressing *cas* genes and appropriate CRISPR cassette from pCDF-based plasmids. Cascade subcomplexes containing N-terminal Strep-Tag II fused to Cse2 subunit were affinity-purified on *Strep*-Tactin® column (IBA) from cells grown at 37°C until OD_{550} reached 0.5 followed by 4-h induction with 1 mM IPTG. Purification buffers contained 100 mM Tris-HCl, pH 8, 150 mM NaCl, 5 mM β-mercaptoethanol and 1 mM ethylenediaminetetraacetic acid (EDTA). Binding buffer additionally contained 0.1 mM phenylmethanesulfonyl fluoride and Elution buffer contained 2.5 mM desthiobiotin and 1 mM (tris(2-carboxyethyl)phosphine) TCEP. Complexes were further separated using a Superdex 200 HiLoad 16/60 column (Amersham Biosciences) equilibrated by 50 mM Tris-HCl, pH 8, containing 150 mM NaCl, 1 mM EDTA and 1 mM TCEP.

N-terminally 6His-tagged Cse1 was purified by IMAC from *E. coli* KD418 strain transformed with an appropriate expression plasmid followed by gel-filtration on Superdex 200 HiLoad 16/60 column (Amersham Biosciences) equilibrated with 20 mM HEPES-K buffer (pH 7.5) containing 150 mM NaCl.

crRNA analysis

To verify molecular size of crRNA 1 μl of Cascade samples was added to 10 μl reaction mixture containing 2.5 μCi [γ-^{32}P]-ATP (6000 Ci/mmol), 5 units of T4 polynucleotide kinase (PNK, New England Biolabs) and 1x PNK reaction buffer (New England Biolabs). After 30 min incubation at 37°C, reactions were stopped with equal volume of formamide-containing loading buffer, boiled for 1 min and products resolved by denaturing 8 M urea, 20% polyacrylamide gel electrophoresis (PAGE) and revealed by autoradioagraphy.

Native gel analysis

Cascade samples were loaded on a native precast 4–12% gradient polyacrylamide gel (Novex™, Invitrogen) and run in Tris-Glycine buffer at 25 mA for 2 h. Protein bands were visualized by staining with SimplyBlue™ (Invitrogen) according to manufactorer's protocol. To identify protein bands carrying crRNA, Cascade samples were incubated with ^{32}P-end labeled oligonucleotide probe complementary to the 5′-part of crRNA (the same probe was used for Northern blotting) in the binding buffer (40 mM Tris-HCl, pH 8.0, 50 NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, 50 μg/ml bovine serum albumin (BSA)) at room temperature for 15 min prior to loading on native gel. Products were revealed by autoradiography.

Protein and native mass spectrometry

Cascade complex samples were sprayed from in-house prepared gold-coated borosilicate glass capillaries and analyzed on a SYNAPT G2-Si instrument (Waters) as reported before (18). Briefly, purified Cascade complexes were buffer exchanged to 100 mM ammonium acetate, pH 7 (Sigma) using 3 kDa molecular weight cutoff spin filters (Pall Corp.) and infused to electrospray source at protein concentration of 2–3 μM and the rate around 90 nl/min. To assure best instrument performance in high mass-to-charge range critical parameters were adjusted as follow: source temperature 30°C, capillary voltage 1.7 kV, trap bias voltage 16 V and argon pressure in collision cell (trap) 7 ml/min. Transfer collision energy was kept at constant level of 10 V while trap energy varied between 10–200 V. To obtain accurate mass measurement for individual protein subunits, Cascade samples were dissociated and completely unfolded using a 1% of formic acid (Sigma) in acetonitrile solution mixed with complex in 1:1 ratio. Data analysis was performed in MassLynx software version 4.1 (Waters).

Permanganate probing

Target dsDNA 209-bp fragment was PCR amplified from plasmid pG8 with g8-F 5′-CTTTAGTCCTCAAAGCCTCTG-3′ and g8-R 5′-GCTTGCTTTCGAGGTGAATTTC-3′ primers. The 5′ ends of the target DNA fragment were labeled with ^{32}P using T4 PNK for 30 min at 37°C (5–10 pmoles of 5′ termini in 30 μl reaction mixture containing 70 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM DTT, 10 pmoles [γ-^{32}P]-ATP (6000 Ci/mmol) and 10 units of T4 PNK (New England Biolabs). ^{32}P-labeled DNA fragments were purified on micro Bio-Spin™ chromatography columns packed with Bio-Gel P-6 (Bio-Rad). $KMnO_4$ probing was performed with 5 nM labeled DNA fragment, 0.2–0.4 mM of Cascade complex, in 10 μl binding buffer (40 mM Tris-HCl, pH 8.0, 50 NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, 50 μg/ml BSA). Before probing, the Cse2Cas7Cas5Cas6e-crRNA subassemblies were incubated with Cse1 (1:1.5 molar ratio) in binding buffer at 37°C for 15 min. After the addition of DNA, incubation at 37°C was continued for 30 min and probing reaction was initiated by adding $KMnO_4$ to a final concentration of 2.5 mM. Reactions were incubated for 15 s at 37°C, quenched by the addition of 10 μl of 1% β-mercaptoethanol, followed by 5 μg of calf thymus DNA in 50 μl of 10 mM Tris-HCl (pH 8.5). Reactions were extracted with phenol–chloroform mixture, followed by ethanol precipitation. DNA pellets were dissolved in 100 μl

of freshly prepared 1 M piperidine and heated in dry bath at 95°C for 20 min. Piperidine was removed by chloroform extraction and DNA was ethanol precipitated. Pellets were dissolved in 10–15 μl of formamide loading buffer and products were separated by denaturing 8% PAGE and revealed by autoradiography.

Cas3 cleavage

Cas3 digestion of DNA in preformed complexes with target DNA was performed in 10 mM HEPES pH 7.4, 60 mM KCl, 10 mM $MgCl_2$, 20 μM $CoCl_2$, 2 mM adenosine triphosphate (ATP) by 800 nM Cas3 for 50 min at 37°C. The reactions were stopped with 2 volumes of formamide-loading buffer, heated at 100°C for 1 min and products resolved by denaturing 8 M urea, 10% PAGE.

Fluorometric measurements

Cascade beacons were formed by mixing oligonucleotide 1 labeled with fluorescein at 5′ end, unmodified oligonucleotide 2 and oligonucleotide 3 labeled with Iowa Black® FQ at 3′ end (final oligonucleotide concentrations were within low μM range) in a buffer containing 40 mM Tris, pH 7.9, 100 mM NaCl by heating for 1 min at 90°C and slow cooling to 20°C. Oligonucleotides 2 and 3 were taken in 30% excess to oligonucleotide 1 to avoid the presence of free oligonucleotide 1 in samples. Control experiments verified that such excess of the oligonucleotides 2 and 3 had no effect on beacon binding.

Fluorescence measurements were performed using a QuantaMaster QM4 spectrofluorometer (PTI) in binding assay buffer (20 mM Tris HCl (pH 7.9), 50 mM NaCl, 5% glycerol, 0.1 mM DTT and 5 mM $MgCl_2$) containing 0.02% Tween 20 at 25°C. Final assay mixtures (800 μl) contained 40 or 80 nM of a Cascade-crRNA complex and 2 nM Cascade beacon construct. The fluorescein fluorescence intensities were recorded with an excitation wavelength of 498 nm and an emission wavelength of 520 nm. Time-dependent fluorescence changes were monitored after addition of negligible volume of Cascade beacon to a cuvette followed by manual mixing; the mixing dead-time was 15 s.

Analysis of Type I spacers from public database

CRISPRFinder (17) with default parameters were used to fetch all CRISPR arrays from archaeal and bacterial sequences downloaded from NCBI FTP site (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/). A total of 488 437 spacers were found. CRISPR-Cas system types were annotated using procedures described in (1,29). Only one CRISPR array per CRISPR-Cas subtype per species Tax ID were taken resulting in a set of 3214 spacers for Type I CRISPR-Cas systems. R library ggplot2 (27) was used to generate histograms for spacer length distribution.

RESULTS

E. coli crRNAs with altered spacer lengths are processed *in vivo* and cause CRISPR interference

E. coli KD263 contains inducible *cas* genes and an engineered CRISPR array with one 32-nt spacer (referred to as 'wt'), matching a protospacer in the M13 phage genome (Figure 1A). Derivative strains harboring CRISPR arrays containing a longer version of 32-nt wt spacer (+6, extended by 6 nucleotides at the leader-distal end), as well as shorter versions (−3, −6, −12 and −18, truncated, correspondingly, by 3, 6, 12 and 18 nucleotides at the leader-distal end) were constructed (Figure 1A). Northern blot analysis of total RNA prepared from induced cell cultures showed that each strain contained processed crRNA of expected length (Figure 1B). Thus, processing of pre-crRNA transcript by *E. coli* Cas6e occurs independently of spacer length.

A CRISPR interference test was performed with strain KD263 and its altered-length spacer derivatives using M13 phage infection. The wild-type phage did not form plaques on lawns of induced KD263, +6 and −3 cells. Other cells were permissive for infection (Figure 1C). As an alternative to phage infection assay, induced cells were transformed with plasmid pG8 containing a protospacer matching the +6 spacer and functional ATG PAM. The pG8 protospacer also matches all shorter spacers. Efficiency of transformation (EOT) of KD263 was decreased ∼100 times compared to the control cells harboring a single repeat in CRISPR array or EOT with vector plasmid, as expected (Figure 1C). A similar decrease was observed for cells producing +6 and −3 crRNA, indicating they form active effector complexes. The EOT of −6, −12 and −18 cells was not affected by *cas* genes induction (Figure 1C).

As a more sensitive interference test, we introduced the pG8 plasmid in uninduced cells and monitored plasmid loss 2 and 4 h after induction of *cas* gene expression (Figure 1D). The yield of plasmids in uninduced cells increased along with growth of each culture, as expected. In contrast the pG8 plasmid did not accumulate in induced cultures of KD263 and cells expressing +6, and −3 crRNA, in agreement with plasmid transformation and phage infection test results. Plasmid yield from −6 cultures was also strongly decreased compared to uninduced control. A much weaker, but highly reproducible decrease in plasmid yield was observed in cultures expressing −12 crRNA. The effect was specific since no such decrease was observed in cultures expressing −12 crRNA and carrying a pG8mut plasmid with a C1T escape mutation in the first position of the protospacer (Figure 1D, inset). The yield of plasmids from −18 induced cell culture was the same as that for control uninduced cells. Overall, we conclude that with the exception of −18 crRNA, altered-length spacer crRNAs tested here are capable of at least some interference with plasmid DNA targets *in vivo*.

crRNAs with altered spacer length induce primed adaptation *in vivo*

We next followed expansion of CRISPR arrays in induced cultures of cells transformed with the pT7blue vector (lanes labeled 1 in Figure 2A), the pG8 plasmid (lanes 2), or pG8mut (lanes 3). The KD263 control cultures behaved as expected: no expansion was detected in cells transformed with pT7blue; less than 10% of arrays were expanded in cultures transformed with pG8, but >80% of cells harboring pG8mut expanded their arrays. The +6 culture behaved as the KD263 control. In −3, −6 and −12 cultures compa-

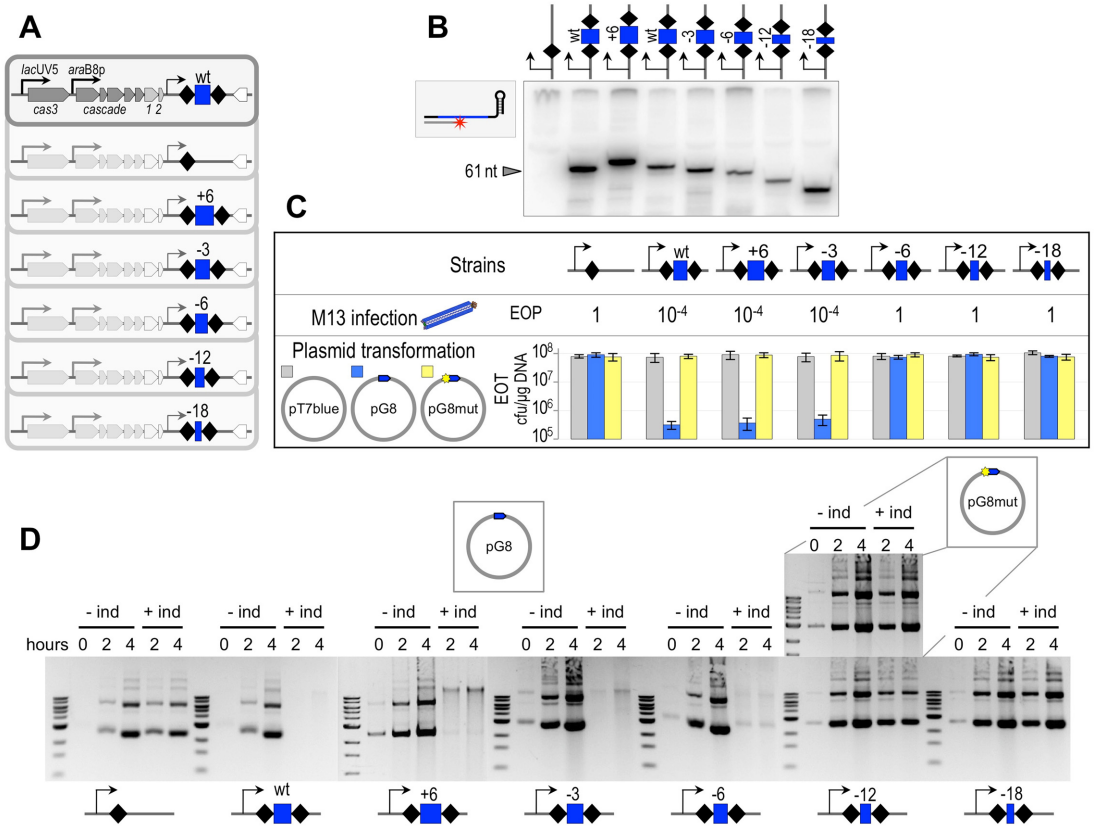

Figure 1. Cells expressing altered-length crRNAs are capable of CRISPR interference. (**A**) KD263 *E. coli* cells expressing wild-type, 32-nt spacer crRNA and derivatives with Clustered Regular Interspaced Short Palindromic Repeats (CRISPR) arrays containing no or altered-length spacers are schematically shown. (**B**) A Northern blot of total RNA prepared from induced cells shown in panel A showing material hybridizing to a radioactive probe complementary to 5′ part of crRNA. (**C**) Efficiency of plaque formation (EOP) by the M13 phage and plasmid transformation efficiency (EOT) of cells shown in A. In transformation experiments, each strain was transformed with blank vector pT7blue; the pG8 plasmid containing a protospacer matching +6 spacer and therefore all shorter versions of the spacers and a functional PAM or pG8mut, a pG8 derivative with a C1T substitution at the first position of the protospacer. (**D**) The pG8 plasmid was purified from equal volumes of cultures of indicated cells grown for 2 and 4 h with or without *cas* gene expression inducers and resolved by agarose gel-electrophoresis. For -12 cells, results of purification from cultures transformed with pG8mut are shown in the inset.

rable levels of adaptation were observed in cells carrying either pG8 or pG8mut. In −18 cultures, CRISPR arrays were expanded only in cultures harboring pG8.

Since no adaptation was detected in cultures transformed with control pT7blue vector, amplicons corresponding to expanded CRISPR arrays shown in Figure 2A must have arisen as a result of primed adaptation. To prove this, expanded CRISPR arrays were subjected to Illumina sequencing. After filtering, spacers (defined as sequences located between repeats) were extracted and mapped onto the pG8 plasmid. The results, presented in Figure 2B, show that the distribution and the efficiency of use of protospacers from which new spacers were acquired was very similar for cells expressing wild-type and altered-length crRNAs. The strong preference for AAG PAM in donor protospacers and a strand bias of spacer acquisition observed in each cell culture (Figure 2C) are indicative of primed adaptation. Analysis of Illumina reads indicated that in all cases the parental spacer was of expected (wild-type, longer or shorter) lengths. Yet, all acquired spacers were 32 ± 1 nucleotides long (Figure 2D). Overall, we conclude that altered-length crRNAs stimulate adaptation of standard-length spacers by priming at the target protospacer.

Purification of Cascades containing altered-length crRNAs

E. coli KD418 cells lacking their own *cas* gene operon were transformed with pCDF-based plasmid co-overexpressing *cse2*, *cas7*, *cas5*, *cas6e* genes and CRISPR arrays containing wt or altered-length spacers from inducible promoters. The plasmid-borne *cse2* gene encodes functional Cse2 N-terminally fused to StrepII-tag, which allows purification of Cascade $(Cse2)_2(Cas7)_6Cas5_1Cas6e_1$ subassembly with bound wild-type crRNA (Figure 3A, (23,28)). Purified Cascade subassemblies from cells expressing altered-length crRNAs contained all expected subunits as judged by visual inspection of SDS gels (Figure 3B). Nucleic acids in purified Cascades subassemblies were radioactively labeled using T4 nucleotide kinase and labeled products were revealed after denaturing urea-PAGE (Figure 3C). Each sample contained a major radioactive band of corresponding crRNA. Judging by labeling efficiency, crRNA content in −3, −6 and −12 Cascades was equal to that in the wild-type Cascade, while +6 and −18 Cascades contained less crRNA.

Cascades samples containing unlabeled crRNAs were combined with radioactive DNA oligonucleotide complementary to the 5′-part of crRNA and analyzed by native

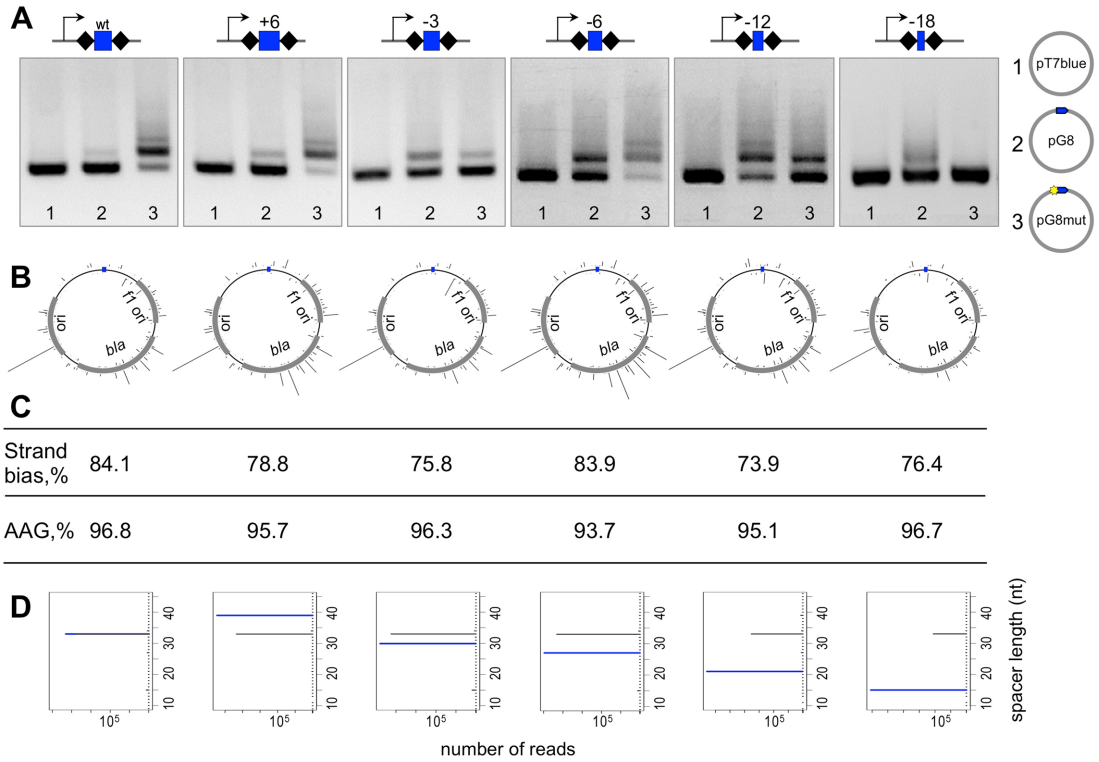

Figure 2. Primed adaptation in cells expressing altered-length crRNAs. (**A**) Cultures of KD263 or its derivatives were transformed with indicated plasmids, grown at conditions of *cas* gene induction and subjected to PCR analysis using a pair of primers annealing upstream and downstream of the genomic CRISPR array. (**B**) Mapping of newly acquired spacers on the pG8 plasmid. The height of bars emanating from plasmid circles correspond to the number of times a spacer corresponding to a protospacer in this location was revealed after Illumina sequencing. The priming protospacer (blue rectangle) is located at the top of each plasmid circle. Bars facing inside the circle correspond to spacers originating from the targeted strand of the priming protospacer. Bars facing outside correspond to spacers from non-targeted strand. (**C**) Statistics of newly acquired spacers. The overall percentage of AAG PAMs associated with newly acquired spacers and the strand bias are presented. (**D**) Acquired spacer lengths. The blue horizontal lines show the length of the spacer in priming crRNA. Black lines show the lengths of acquired spacers.

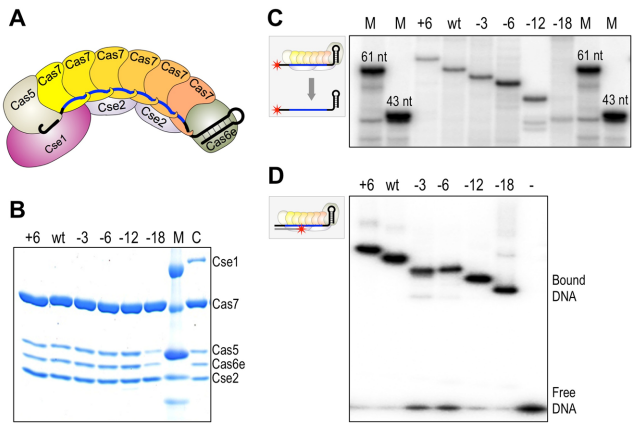

Figure 3. Altered-length crRNAs assemble into Cascade compelxes. (**A**) Schematic representation of complete *E. coli* Cascade complex containing wild-type, 32-nt spacer crRNA. (**B**) Protein content of Cascades purified from KD263 and derivatives with altered-length CRISPR spacers. A Coommasie-stained SDS gel is presented. (**C**) RNA content of purified Cascades. The 5′-ends of RNA in affinity purified Cascades were labeled with ^{32}P, resolved by denaturing gel and revealed by autoradiography. (**D**) Affinity purified Cascades were combined with radioactively labeled DNA oligonucleotide complementary to 5′ part of crRNA and reaction products were resolved by native PAGE and visualized by autoradiography.

gel electrophoresis. The wild-type crRNA Cascade bound the target oligonucleotide and migrated as a single sharp radioactive band (Figure 3D). In the preparation of +6 crRNA Cascade a similar band but with lower mobility was observed. The −3 and −6 crRNA Cascades had the same mobility and moved faster that the wild-type crRNA Cascade band. Separation of Cascades with −12 and 18 crRNAs revealed radioactive bands with progressively higher mobility on the gel.

The observed mobility shifts during native gel electrophoresis (Figure 3D and Supplementary Figure S1) are consistent with altered stoichiometry of resulting complexes, presumably caused by the addition of one (for +6 crRNA) or loss of one (for −3 and −6 crRNAs), two (for −12) crRNA or three (for −18 crRNA) Cas7 monomers. To directly determine the stoichiometry of altered crRNA Cascades native mass-spectrometry of each complex was performed. To simplify data interpretation of mass measurements carried out in native conditions that preserve complex integrity, accurate masses of all individual components were determined first. To enable that, Cascade complexes were fully denatured using combination of organic solvent and low pH before introduction to ionization source. Masses of individual subunits were consistent with predicted values calculated based on pri-

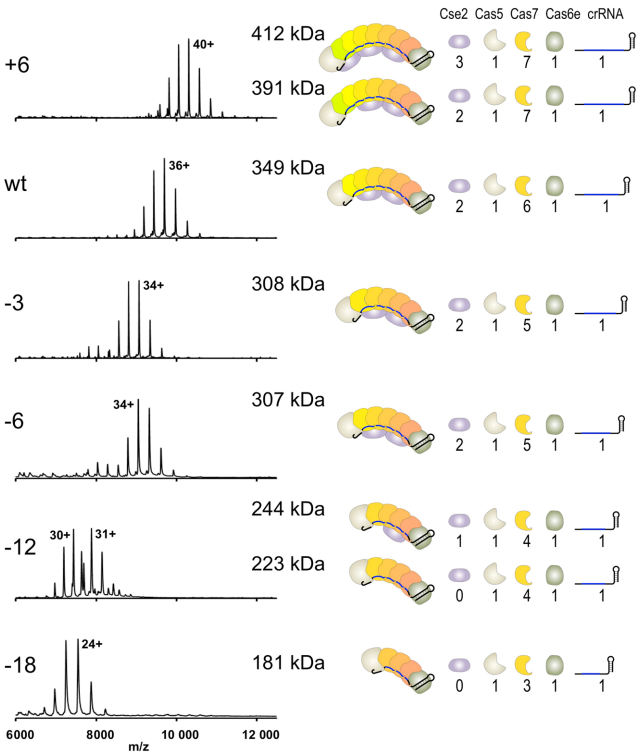

Figure 4. The stoichiometry of Cascade complexes assembled on variable spacer length crRNAs. On the left, native mass-spectra of indicated Cascade complexes are shown. Main charge state is denoted as a number by the most intense peak in charge state distribution. On the right, stoichiometries consistent with recorded complex masses are schematically shown.

mary protein sequence (Supplementary Table S3). In the native conditions Cascade complex containing standard length crRNA appeared primarily as a single species with a mass around 349 kDa, which is in agreement with predicted stoichiometry $(Cse2)_2(Cas7)_6Cas5_1Cas6e_1$-crRNA (Figure 4). Upon collision activation conditions, a 39.9 kDa monomer of Cas7 and 25.2 kDa Cas5 dissociated from the complex (Supplementary Figure S2), suggesting peripheral location or limited interaction of dissociated subunits with main complex. After dissociation of Cas7 and/or Cas5 (which appeared in the low mass region of the spectrum) corresponding residual Cascade forms—309 kDa ($(Cse2)_2(Cas7)_5Cas5_1Cas6e_1$-crRNA) and 324 kDa ($(Cse2)_2(Cas7)_6Cas6e_1$-crRNA)—were observed in high mass region.

Cascade containing the +6 crRNA was present in two forms: a less abundant 391 kDa complex (consistent with expected $(Cse2)_2(Cas7)_7Cas5_1Cas6e_1$-crRNA stoichiometry) and a heavier major 412 kDa complex (Figure 4). Based on calculated mass difference between expected and recorded mass, there appears to be an additional Cse2 subunit in this complex with resulting stoichiometry $(Cse2)_3(Cas7)_7Cas5_1Cas6e_1$-crRNA. Cascades containing the −3 and −6 crRNA appeared to be lacked on Cas7 monomer, consistent with a $(Cse2)_2(Cas7)_5Cas5_1Cas6e_1$-crRNA stoichiometry (Figure 4). Cascade containing the −12 crRNA existed in two main forms in gas phase: a 244 kDa complex matching expected $(Cse2)_1(Cas7)_4Cas5_1Cas6e_1$-crRNA stoichiometry with two Cas7 monomers missing and a 223 kDa complex (Figure 4). Based on the mass difference the latter complex appears to be missing the Cse2 subunit. Since our Cascade isolation relies on affinity tag located on Cse2, the presence of a complex without this subunit indicates lesser overall stability of purified effector complexes with shortened crRNA spacer. Cascade with −18 crRNA existed as one main form in gas phase (Figure 4). Based on mass difference with expected stoichiometry it also appears to be missing the Cse2 subunit and has a stoichiometry of $(Cas7)_3Cas5_1Cas6e_1$-crRNA, i.e. it lacks three Cas7 monomers. Comparing to spectra of other complexes at identical experimental conditions (low collision energy that facilitates ions transfer but preserves complex integrity) the −18 crRNA Cascade spectrum contained low charge state distributions of free Cse2 and Cas7 in the low mass region (Supplementary Figure S3). This may again suggest that Cse2 is rather loosely bound to Cascades with significantly shortened crRNAs. Overall, despite the complications caused by subunit dissociation, mass-spectrometric analysis confirmed that altering the crRNA spacer length leads to predicted changes in Cas7 stoichiometry of the Cascade complex, which are sometimes also accompanied by changes in the Cse2 copy number.

Functional analysis of Cascades containing altered-length crRNAs

After binding to their targets, Cascades with altered Cas7 stoichiometry are expected to form R-loop complexes that contain longer (or shorter) areas of localized DNA melting depending on the crRNA spacer length. We supplemented affinity purified Cascades with recombinant Cse1, incubated with radioactively labeled double-stranded DNA fragment with the target protospacer and then probed with $KMnO_4$, a chemical agent that modifies thymines present in single-stranded form. In complexes formed by the wild-type Cascade-crRNA, $KMnO_4$–sensitive thymines in positions 2, 4, 6, 7, 8, 12, 15, 21 and 27 in the non-targeted strand were observed (Figure 5A). Binding of the +6 Cascade led to $KMnO_4$ sensitivity of additional thymines at positions 34, 36 and 38, indicating that the R-loop has expanded in the downstream direction. The -3 Cascade had a footprint that was very similar or identical to wild-type. The sensitivity of thymine to modification 27 was decreased in the complex formed by the -6 Cascade, a result consistent with formation of an R-loop shortened due to loss of one Cas7 monomer. However, a new band further away (position 34) appeared and the intensity of modification of thymine at position 6 in the seed was increased, compared to the wild-type pattern. While the extra downstream band is likely caused by non-specific binding (see below), the change in the efficiency of T^6 modification suggests that the conformation of non-template strand in the -6 complex is altered. The -12 crRNA Cascade gave a weak footprint that did not extend past T^{15}. The absence of sensitivity of T^{21} is expected for an R-loop formed by an effector complex lacking two Cas7 monomers. A barely visible short footprint was observed with the −18 Cascade. Two additional bands present in −12 and −18 Cascades lanes outside the normal R-loop—T^{34}

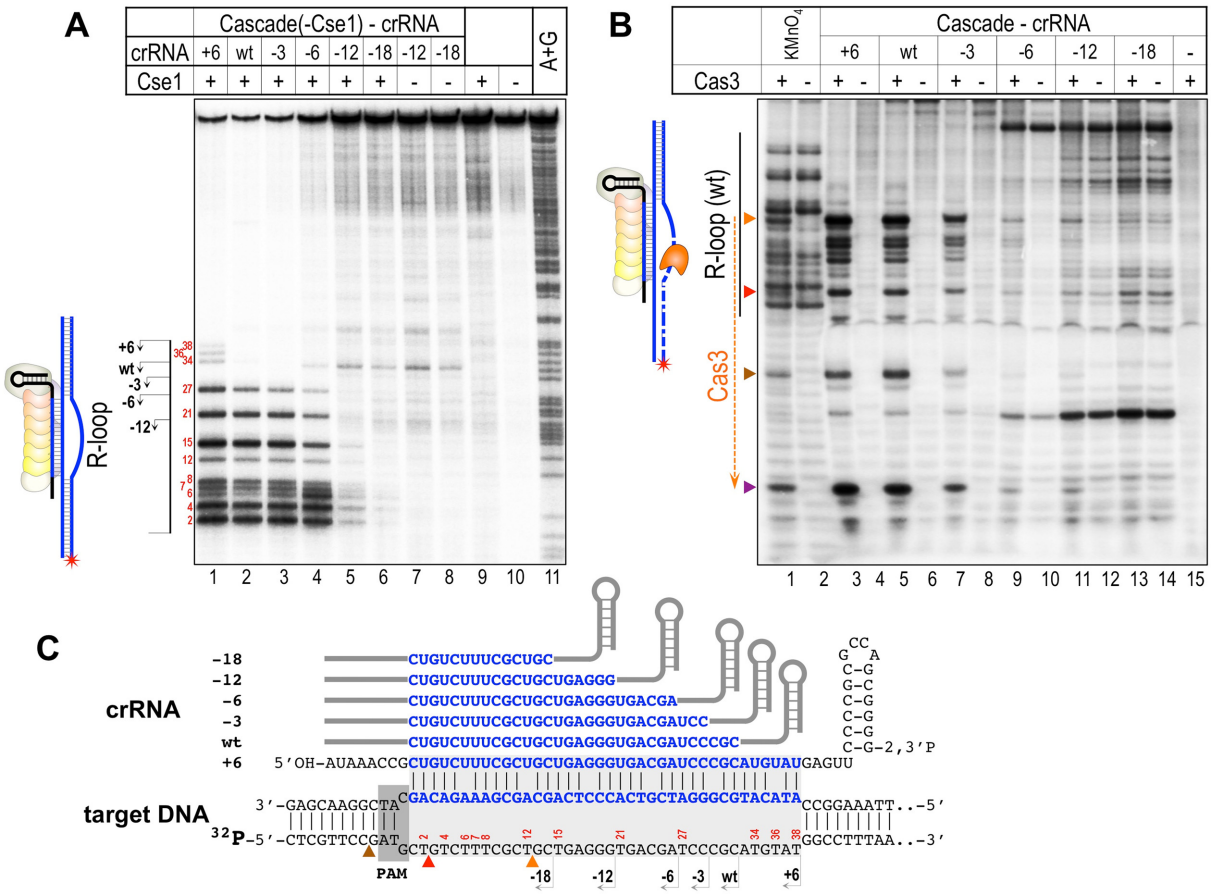

Figure 5. *In vitro* analyses of target binding by Cascades with altered-length crRNAs. (**A**) $KMnO_4$ probing of complexes formed by various Cascades. The varying boundaries of R-loop extension observed in various complexes (indicated in black color font) are shown at the left of the gel with black arrows and also summarized in panel C. Positions of $KMnO_4$-sensitive thymines discussed in the text are also indicated (red-colored numbers) at the left of the gel and in panel C scheme. The first (PAM proximal) residue of the protospacer has a number of 1. (**B**) Cas3 cleavage of complexes formed by various Cascades. The extent of R-loop extension in R-loop formed by wild-type crRNA Cascade is shown by a black vertical line on the left of the gel. The direction of Cas3 cleavage is indicated by an orange arrow. Positions of specific Cas3-induced cleavage sites are marked at the left of the gel with red-brown triangles and are also indicated in corresponding colors in the scheme in panel C. (**C**) The structure of an R-loop complex formed by the +6 crRNA Cascade is shown with changes in extent of the R-loop formed by Cascades containing other crRNAs indicated below. Positions of $KMnO_4$-sensitive thymines are shown by red-colored numbers. Sites of Cas3 cleavage are shown by red and brown triangles.

and T^{43}—were also observed when −12 and −18 Cascade preparations were combined with target DNA in the absence of Cse1. Since even highly active wt Cascade does not form R-loop in the absence of Cse1, we conclude that these bands are due to non-specific binding.

The R-loop complexes formed by various Cascades were also treated with Cas3 in the presence of ATP. As expected, prominent cleavages in the non-target strand of the R-loop as well as downstream (in the 3′-5′ direction) were observed in reactions containing wild-type Cascade (Figure 5B, (30)). Similar cleavages were observed in reactions containing the +6 and −3 crRNA Cascades. Much weaker Cas3-induced cleavages were also present in reactions containing the −6 and −12 Cascades. No Cas3 cleavages were detected with the −18 Cascade. It should be noted that in −6, −12 and −18 reactions there were multiple cleavages that did not depend on the addition of Cas3 and were practically absent in other reactions. The cause of this degradation was not investigated; it could have been caused by the presence of contaminating endonucleases in our Cascade preparations. Efficient R-loop complex formation by the wild-type and +6 and −3 Cascades could have protected bound DNA from the action of the endonuclease, allowing to clearly see Cas3 degradation products.

Quantitative analysis of altered-length crRNA Cascades binding to target DNA

To obtain quantitative information on target binding affinity of Cascades with different crRNAs, a fluorescent beacon method was implemented, following a strategy that was recently used to monitor the interaction of the Cas9 effector complex with target DNA (31). The assay relies on fluorescent signal from a model target DNA construct referred to as 'Cascade beacon' (Figure 6A), a ∼50 bp DNA fragment encompassing complete or partial protospacer matching crRNA spacer, a PAM and nucleotides upstream of PAM. The non-target strand of the beacon contains discontinuity between the PAM (oligo 1) and protospacer (oligo 2) segments. Oligos 2 and 3 are labeled with a fluorescent label and

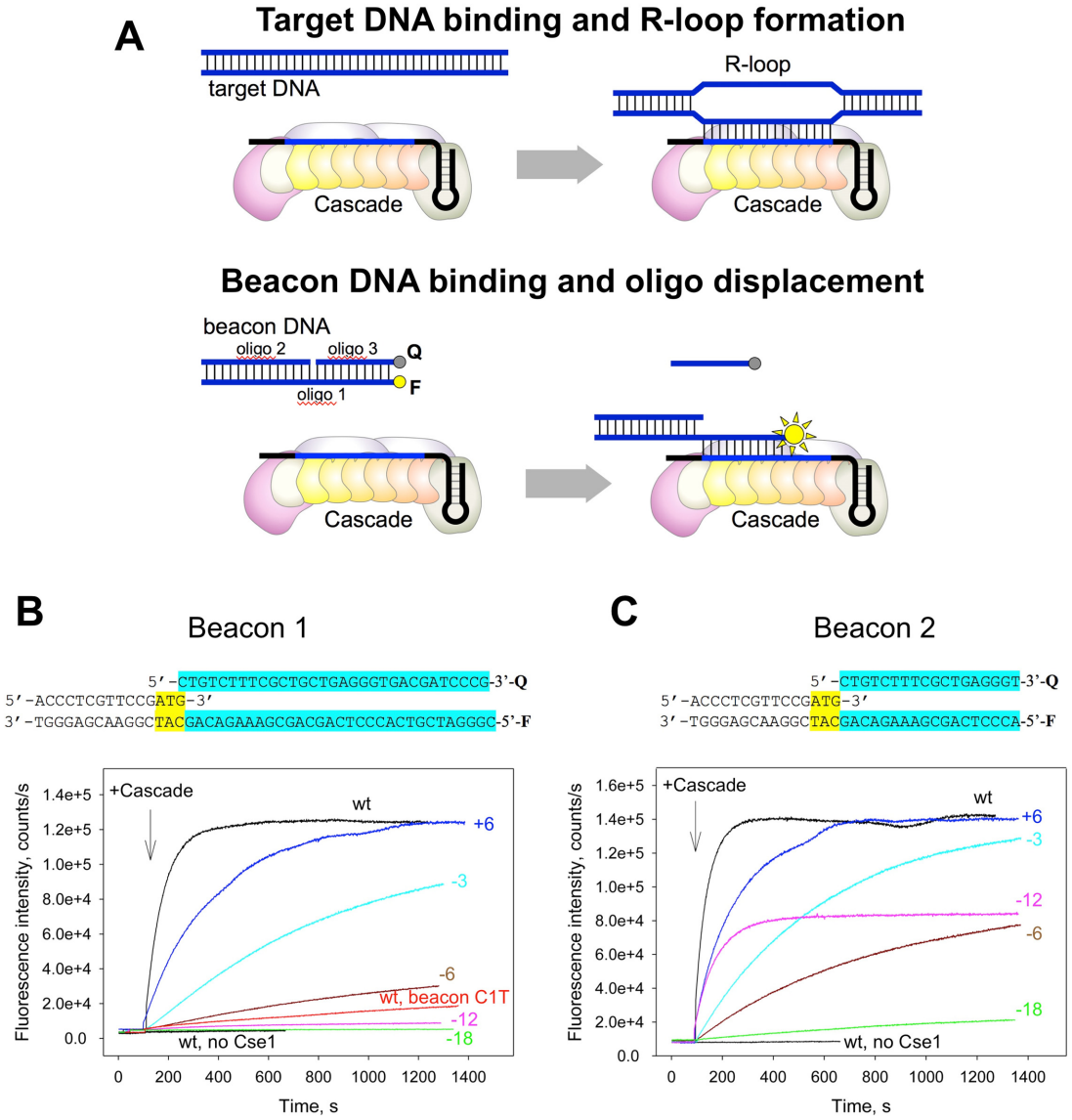

Figure 6. Quantitative measurement of target binding using Cascade beacons. (**A**) The principle of the Cascade beacon assay is schematically shown. (**B**) The binding kinetics of various Cascades to Cascade beacon 1 (shown at the top). In addition to binding of wild-type and altered-length crRNA Cascades to fully matching beacon 1, binding of wild-type crRNA Cascade without added Cse1 to beacon 1 and binding of wild-type crRNA Cascade supplemented to Cse1 to beacon 1 derivative with a C1T substitution is shown. (**C**) The binding kinetics of various Cascades to Cascade beacon 2 (shown at the top).

a fluorescence quencher, correspondingly (Figure 6A). The baseline beacon fluorescence intensity is low due to quenching of the fluorescence label by the nearby quencher. Formation of the R-loop promoted by Cascade-crRNA should lead to oligo 3 dissociation, increasing fluorescence intensity due to the disappearance of the quenching effect. We tested these expectations using Cascade beacon 1, which contains the ATG PAM and a full-sized protospacer segment. Upon the addition of 40 nM Cascade loaded with wild-type crRNA to 2 nM beacon 1, fluorescence intensity increased by ∼30-fold, reaching peak values in about 3 min (Figure 6B). The fluorescence signal increase of beacon 1 derivative harboring a C1T substitution in the first position of the protospacer seed was ∼400-fold slower than that observed with beacon 1 (Figure 6B). No change in fluorescence intensity was observed upon the addition to beacon 1 of Cascade-crRNA complex that lacked the Cse1 subunit (Figure 6B). Together, these experiments indicate that increases of beacon fluorescence intensity in the presence of complete Cascade with cognate wild-type crRNA are due to specific interactions of Cascade-cRNA with the beacon followed by beacon melting.

Next, we measured beacon 1 signal increases upon the addition of Cascade complexes containing crRNA of various lengths. As can be seen from Figure 6B, fluorescence intensity increases could be readily measured with +6, −3 and −6 Cascades. The kinetics of fluorescence increase indicated that the binding efficiency was decreased in the order wt >

+ 6 > −3 > −6. The −12 and −18 Cascades did not reveal appreciable binding with beacon 1.

We surmised that some of the observed differences with short crRNA Cascades may be a consequence of hindered dissociation of oligo 3. Therefore, we carried out similar measurements using beacon 2 that bears a shorter 18 bp protospacer segment (Figure 6C). Compared to the wild-type, the binding behavior of +6, −3 and −6 Cascades was similar to that observed with beacon 1. The -12 Cascade also rapidly bound beacon 2, though fluorescence intensity plateaued at a level that was less than that observed with wild-type Cascade. The −18 Cascade also bound beacon 2—slowly, but specifically—with a binding curve similar to that observed for wild-type Cascade binding to the C1T beacon.

The binding traces discussed above may be due not just to changes in binding specificities of Cascade variants but can also be caused by different specific activities of Cascade preparations. To characterize beacon complexes formed by active Cascade variants, we measured the kinetics of beacon 2 binding to Cascades preincubated with either a 74 bp dsDNA containing functional PAM and the target protospacer (g8cons) or a similar DNA fragment bearing the protospacer with mutated seed region (g8mut) (Supplementary Figure S4). The rate of beacon 2 binding to wild-type Cascade preincubated with 80 nM g8cons was 60-fold slower than that observed in the absence of competitor (the binding halftimes are 40 s and 34 min, respectively). In contrast, g8mut only slightly affected the binding kinetics, indicating that the competition effect caused by g8cons results from its specific binding to Cascade. Assuming that the beacon binding rate in the presence of g8cons is proportional to concentration of Cascade molecules unoccupied by g8cons, a Kd for g8cons binding to wt Cascade is 0.7 nM, consistent with earlier reported Kd values (18,32). Competition binding experiments with other Cascade-crRNA complexes (Supplementary Figure S4 and Table S4) showed that affinities of +6 Cascade and −3 Cascade to g8cons were similar to wild-type (Kd values ∼1 nM), whereas −6 Cascade and −12 Cascade formed less stable complexes (Kd values ∼15 nM). Only about 1.5-fold drop in the rate of beacon 2 binding to −18 Cascade was observed upon the addition of g8cons, suggesting that the −18 Cascade-g8cons complex is unstable (Kd > 120 nM) under conditions of the experiment.

DISCUSSION

Very recently, evidence has been presented that crRNAs from two different type I CRISPR-Cas systems with longer than normal spacers assemble into effector complexes containing larger than the normal complement of six Cas7 monomers and capable of CRISPR interference (18,19). Cascades with shortened crRNA spacers were reported to be inactive, however (19). Here, we complement these observations and show, using *E. coli* type I-E system, that decrease of crRNA spacer size leads to formation of Cascades complexes with altered (reduced) Cas7 stoichiometry that recognize target DNA *in vitro* and interfere, fully or partially, with foreign DNA. In addition, all Cascade complexes studied here promote efficient primed adaptation *in vivo*. Earlier *in vivo* analyses showed that primed adaptation depends on weak but specific interactions of Cascade-crRNA with target protospacer (22,33). Recent *in vitro* work shows that Cas3 nuclease degradation initiated after target recognition by Cascade generates substrates for primed adaptation (34). Thus, each complex analyzed, including the one missing three Cas7 subunits, must be able to interfere with the target at least to some extent.

Existing *E. coli* Cascade structures show that one Cas7 monomer binds 6 nt of spacer (14–16,35). Indeed, removal of six crRNA spacer nucleotides, or the addition of six extra nucleotides, cause, respectively, a loss or recruitment of extra Cas7 subunit to crRNA-bound Cascade complex. Further Cas7 monomer losses are seen when crRNA spacer is shortened by 12 and 18 nucleotides. The order of assembly of crRNA repeat binding Cascade components is unknown. It is also not known whether assembly happens before or after processing of the primary pre-crRNA CRISPR array transcript. Both ends of mature crRNA are generated by Cas6e, which alone can bind to and process pre-crRNA hairpins formed by CRISPR repeats and remains bound at the 3′ terminal segment of crRNA repeat in the context of effector complex (3,36). So, in principle Cas6e could initiate Cascade assembly. However, Cas6e is dispensable for Cascade assembly and function if pre-processed crRNA is provided (23). The fact that the amount of Cas7 monomers in the complex follows that of crRNA spacer length suggests that the Cas7 oligomer must be assembling on the crRNA spacer, somehow 'sensing' its length. The deposition of Cas7 may be initiated by Cas5, which binds at the 5′ proximal repeat fragment of crRNA, and proceed until the entire length of the spacer is covered. The finding that crRNA lacking three spacer nucleotides binds to a Cascade containing 5 Cas7 monomers suggests that there is a significant flexibility in the length of the crRNA spacer that can be accommodated by the Cascade. The fact that the amount of Cse2 bound to different Cascades can vary also points out to flexibility of the complex.

Mutational analysis of various spacer–protospacer pairs shows that introduction of mismatches can have complex effects on the interference and primed adaptation efficiencies (37). The strength of target interaction should clearly affect these outcomes, however, heretofore, there was no convenient method to quantify Cascade-crRNA binding to targets and determine the binding rate constants. The Cascade beacon method, developed along the lines of the beacon described earlier (31) offers a wide dynamic range of measurements and, when conducted in the completion mode, allows to measure target complex dissociation rates. The beacon assay thus open way for systematic biophysical analysis of wild-type and mutant Cascade-crRNA complexes interactions with matching and mismatched targets.

Primed spacer adaptation requires the recognition of priming protospacer in target DNA by the effector complex (22,33) and Cas3-induced target degradation (34). The adaptation happens even when wild-type effector interaction is weakened by escape mutations in the protospacer seed or PAM that decrease the binding by 100-times or more (22,24). Decreased interaction with the target slows down the rate of interference/target destruction by Cas3, allowing plasmid copy number maintenance mechanisms to com-

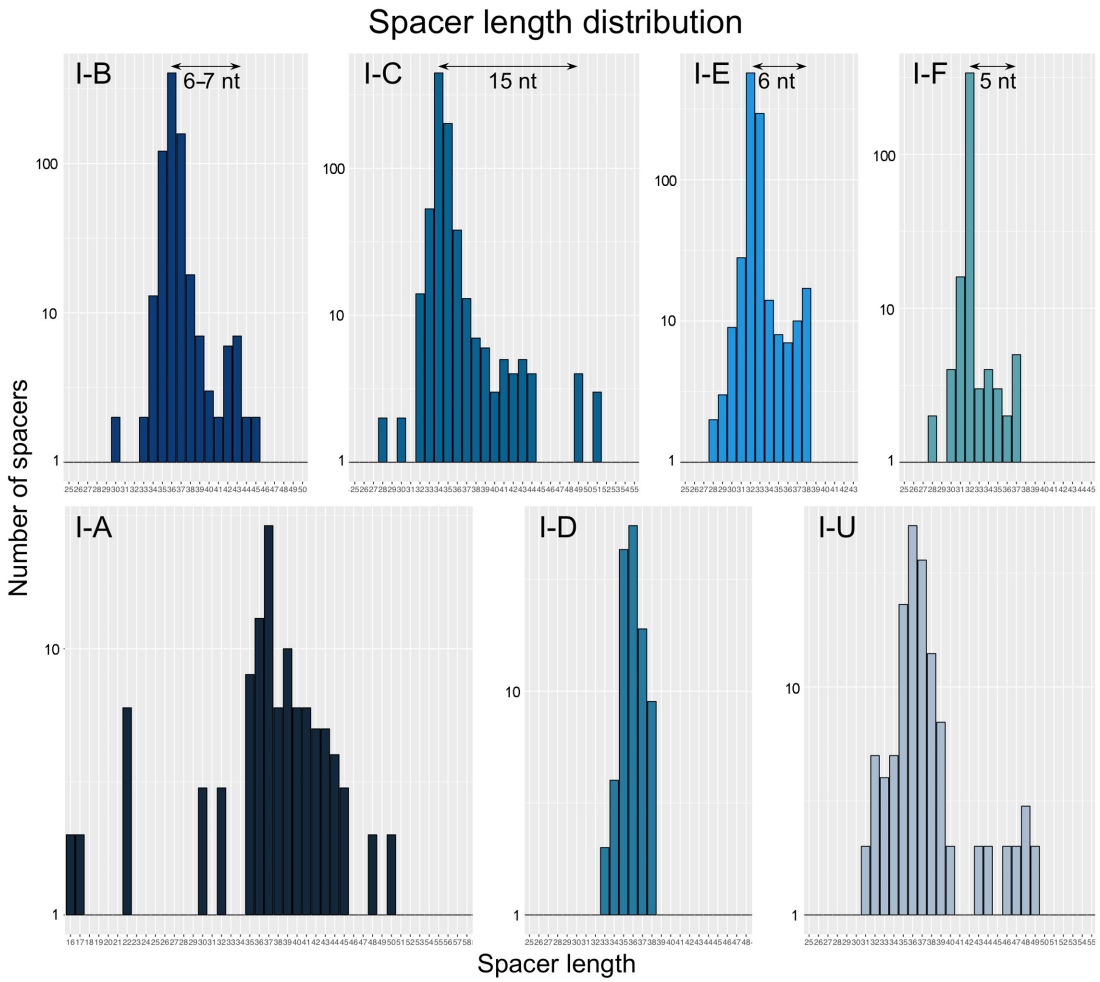

Figure 7. Size heterogeneity of spacers from type I CRISPR arrays. Type I CRISPR-Cas system subtypes were annotated using procedures described in (1,29). One CRISPR array per CRISPR-Cas subtype per species Tax ID was analyzed (a total of 3214 arrays). Median spacer length for each array was determined and used to plot histograms shown.

pensate, establishing a new equilibrium state where products of target DNA degradation are maintained in the cell over extended periods of time and enabling gradual acquisition of spacers (38). As such, primed adaptation appears to be a much more sensitive assay to monitor effector complex interaction with its targets. The results with the −18 crRNA Cascades, lacking three Cas7 monomers, illustrate this point well. While no interference in −18 cells was observed, primed adaptation was robust. However, while in wild-type crRNA Cascade adaptation was most effective in the presence of an escape mutation introducing a mismatch between the spacer and protospacer seed, for −18 crRNA Cascade adaptation was only observed when there was a full match between crRNA spacer and the target. Presumably, the presence of a mismatch weakened the interaction even further, preventing target recognition and thus abolishing both the interference and adaptation.

Analysis of newly acquired spacers shows that different effector complexes that prime adaptation have no discernable effect on spacer choice or length of acquired spacers. The result is consistent with a view that spacer acquisition is carried out solely by Cas1 and Cas2, while the substrates for acquisition are produced by Cas3, which is able to recognize R-loop complexes formed by various Cascades tested in this work (38–40).

Removal or addition of one Cas7 monomer caused by, respectively, a 6-nt increase and 3-nt decrease of crRNA spacer do not have a profound effect on CRISPR interference against viral or plasmid protospacer targets at our conditions *in vivo*. Quantitative measurements using the highly sensitive beacon assay show, however, that either extension or reduction of crRNA spacer length lead to decreased binding affinity and facilitate dissociation from the target. Thus, the hexamer of Cas7 bound to a 32-bp spacer may in fact represent the optimal arrangement for target recognition by and/or stability of *E. coli* Cascade complex. The optimal linear hexamer arrangement of Cas7 Type I-E Cascade must have evolved from an earlier less complex state. Our analysis suggests that simpler functional complexes, perhaps having as little as one Cas7 subunit might have existed. Such complexes could have also contained Cas5 and could bind and protect short RNAs generated by Cas6 or by other means. These complexes could have then evolved by recruiting additional Cas7 monomers to recruit RNAs with

longer spacers to allow tighter interaction with target DNA and, therefore, better protection from genetic parasites. The length of spacer RNA (and the stoichiometry of Cascade) must have also been influenced by the preferred size of protospacers recognized by the Cas1-Cas2 acquisition complex and inserted in to CRISPR array as spacers. Systematic analysis of type I spacers from CRISPR arrays present in public databases indicates that their length varies (Figure 7). With the exception of I-A subtype, spacers in the remaining 6 recognized type I subtypes have a strong preference for spacers of specific lengths: 34–37 nucleotide long spacers (I-B, I-C, I-D and I-U) and 31–33 nt long spacers for I-E and I-F. Yet, there appears to be a larger than expected number of organisms with type I-B, I-E and I-F whose CRISPR arrays harbor spacers which are 5–6 nucleotides longer than the average. Some I-C spacers are 15 nucleotides longer. Our results and the data of others would suggest that Cascades from organisms with such arrays should naturally contain Cas7 heptamers or octamers rather than hexamers present in effectors studied to date. Shorter than average spacers are generally more rare with the notable exception of type I-A systems, where the distribution of preferred spacer lengths is unusually broad with spacers ranging in size from 35 to 45 nucleotides being equally common, but multiple arrays organisms with 16/17, 22 and 30/32 nucleotides arrays also exist. It should be interesting to determine the subunit composition of Cascades in these systems as well as to establish how the Cas1-Cas2 proteins from these systems differ from the *E. coli* homologs that seem to be structurally constrained for selection of 33-bp DNA fragments for integration into CRISPR array (41,42).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Institutes of Health (NIH) [GM10407 to K.S; 5P20RR02437 to B.B.; GM097330 to S.B.]; Russian Science Foundation [14-14-00988 to K.S.] and Ministry of Education and Science of the Russian Federation [14.B25.31.0004 to K.S.]; Russian Foundation for Basic Research [16-04-00767 to E. Savitskaya]; Department of Energy [DE-SC0012518 to B.B.]; M.J. Murdock Charitable Trust [to B.B.]; The Charles and Johanna Busch Memorial Fund [predoctoral fellowship to I.J.]; Natural Sciences and Engineering research Council of Canada (NSERC) [Strategic Network Grant IBN and Discovery grant RGPIN-2015-04745 to S.L. and A.F.Y.]. Funding for open access charge: Skolkovo Institute of Science and Technology; The Russan Science Foundation [14-14-00988], I.J. was supported by Charles and Johanna Busch predoctoral fellowship.

Conflict of interest statement. None declared.

REFERENCES

1. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J., Charpentier,E., Haft,D.H. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
2. Mohanraju,P., Makarova,K.S., Zetsche,B., Zhang,F., Koonin,E.V. and van der Oost,J. (2016) Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*, **353**, aad5147.
3. Brouns,S.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J., Snijders,A.P., Dickman,M.J., Makarova,K.S., Koonin,E.V. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
4. Charpentier,E., Richter,H., van der Oost,J. and White,M.F. (2015) Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.*, **39**, 428–441.
5. Carte,J., Wang,R., Li,H., Terns,R.M. and Terns,M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.*, **22**, 3489–3496.
6. Haurwitz,R.E., Jinek,M., Wiedenheft,B., Zhou,K. and Doudna,J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355–1358.
7. Deltcheva,E., Chylinski,K., Sharma,C.M., Gonzales,K., Chao,Y., Pirzada,Z.A., Eckert,M.R., Vogel,J. and Charpentier,E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, **471**, 602–607.
8. Zetsche,B., Gootenberg,J.S., Abudayyeh,O.O., Slaymaker,I.M., Makarova,K.S., Essletzbichler,P., Volz,S.E., Joung,J., van der Oost,J., Regev,A. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759–771.
9. Hatoum-Aslan,A., Samai,P., Maniv,I., Jiang,W. and Marraffini,L.A. (2013) A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J. Biol. Chem.*, **288**, 27888–27897.
10. Rouillon,C., Zhou,M., Zhang,J., Politis,A., Beilsten-Edmands,V., Cannone,G., Graham,S., Robinson,C.V., Spagnolo,L. and White,M.F. (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol. Cell*, **52**, 124–134.
11. Zhang,J., Rouillon,C., Kerou,M., Reeks,J., Brugger,K., Graham,S., Reimann,J., Cannone,G., Liu,H., Albers,S.V. *et al.* (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell*, **45**, 303–313.
12. Carte,J., Pfister,N.T., Compton,M.M., Terns,R.M. and Terns,M.P. (2010) Binding and cleavage of CRISPR RNA by Cas6. *RNA*, **16**, 2181–2188.
13. Wiedenheft,B., Lander,G.C., Zhou,K., Jore,M.M., Brouns,S.J., van der Oost,J., Doudna,J.A. and Nogales,E. (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, **477**, 486–489.
14. Jackson,R.N., Golden,S.M., van Erp,P.B., Carter,J., Westra,E.R., Brouns,S.J., van der Oost,J., Terwilliger,T.C., Read,R.J. and Wiedenheft,B. (2014) Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science*, **345**, 1473–1479.
15. Mulepati,S., Heroux,A. and Bailey,S. (2014) Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*, **345**, 1479–1484.
16. Zhao,H., Sheng,G., Wang,J., Wang,M., Bunkoczi,G., Gong,W., Wei,Z. and Wang,Y. (2014) Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. *Nature*, **515**, 147–150.
17. Grissa,I., Vergnaud,G. and Pourcel,C. (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.*, **35**, W52–W57.
18. Luo,M.L., Jackson,R.N., Denny,S.R., Tokmina-Lukaszewska,M., Maksimchuk,K.R., Lin,W., Bothner,B., Wiedenheft,B. and Beisel,C.L. (2016) The CRISPR RNA-guided surveillance complex in Escherichia coli accommodates extended RNA spacers. *Nucleic Acids Res.*, **44**, 7385–7394.
19. Gleditzsch,D., Muller-Esparza,H., Pausch,P., Sharma,K., Dwarakanath,S., Urlaub,H., Bange,G. and Randau,L. (2016) Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res.*, **44**, 5872–5882.
20. Shmakov,S., Savitskaya,E., Semenova,E., Logacheva,M.D., Datsenko,K.A. and Severinov,K. (2014) Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.*, **42**, 5907–5916.

21. Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 6640–6645.
22. Datsenko,K.A., Pougach,K., Tikhonov,A., Wanner,B.L., Severinov,K. and Semenova,E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.*, **3**, 945.
23. Semenova,E., Kuznedelov,K., Datsenko,K.A., Boudry,P.M., Savitskaya,E.E., Medvedeva,S., Beloglazova,N., Logacheva,M., Yakunin,A.F. and Severinov,K. (2015) The Cas6e ribonuclease is not required for interference and adaptation by the E. coli type I-E CRISPR-Cas system. *Nucleic Acids Res.*, **43**, 6049–6061.
24. Semenova,E., Jore,M.M., Datsenko,K.A., Semenova,A., Westra,E.R., Wanner,B., van der Oost,J., Brouns,S.J. and Severinov,K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10098–10103.
25. Morgan,M., Anders,S., Lawrence,M., Aboyoun,P., Pages,H. and Gentleman,R. (2009) ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, **25**, 2607–2608.
26. Pages,H., Aboyoun,P., Gentleman,R. and DebRoy,S. (2012) *R package version 2.24.1*.
27. Wickham,H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer, NY.
28. Beloglazova,N., Kuznedelov,K., Flick,R., Datsenko,K.A., Brown,G., Popovic,A., Lemak,S., Semenova,E., Severinov,K. and Yakunin,A.F. (2015) CRISPR RNA binding and DNA target recognition by purified Cascade complexes from Escherichia coli. *Nucleic Acids Res.*, **43**, 530–543.
29. Makarova,K.S. and Koonin,E.V. (2015) Annotation and Classification of CRISPR-Cas Systems. *Methods Mol. Biol.*, **1311**, 47–75.
30. Mulepati,S. and Bailey,S. (2013) In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.*, **288**, 22184–22192.
31. Mekler,V., Minakhin,L., Semenova,E., Kuznedelov,K. and Severinov,K. (2016) Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3′-terminal segment of guide RNA. *Nucleic Acids Res.*, **44**, 2837–2845.
32. Westra,E.R., Semenova,E., Datsenko,K.A., Jackson,R.N., Wiedenheft,B., Severinov,K. and Brouns,S.J. (2013) Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.*, **9**, e1003742.
33. Swarts,D.C., Mosterd,C., van Passel,M.W. and Brouns,S.J. (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One*, **7**, e35888.
34. Kunne,T., Kieper,S.N., Bannenberg,J.W., Vogel,A.I., Miellet,W.R., Klein,M., Depken,M., Suarez-Diez,M. and Brouns,S.J. (2016) Cas3-Derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell*, **63**, 852–864.
35. Hayes,R.P., Xiao,Y., Ding,F., van Erp,P.B., Rajashankar,K., Bailey,S., Wiedenheft,B. and Ke,A. (2016) Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature*, **530**, 499–503.
36. Jore,M.M., Lundgren,M., van Duijn,E., Bultema,J.B., Westra,E.R., Waghmare,S.P., Wiedenheft,B., Pul,U., Wurm,R., Wagner,R. *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, **18**, 529–536.
37. Xue,C., Seetharam,A.S., Musharova,O., Severinov,K., SJ,J.B., Severin,A.J. and Sashital,D.G. (2015) CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.*, **43**, 10831–10847.
38. Severinov,K., Ispolatov,I. and Semenova,E. (2016) The influence of copy-number of targeted extrachromosomal genetic elements on the outcome of CRISPR-Cas defense. *Front. Mol. Biosci.*, **3**, 45.
39. Kunne,T., Kieper,S.N., Bannenberg,J.W., Vogel,A.I., Miellet,W.R., Klein,M., Depken,M., Suarez-Diez,M. and Brouns,S.J. (2016) Cas3-Derived target DNA degradation fragments fuel primed CRISPR aAdaptation. *Mol. Cell*, **63**, 852–864.
40. Semenova,E., Savitskaya,E., Musharova,O., Strotskaya,A., Vorontsova,D., Datsenko,K.A., Logacheva,M.D. and Severinov,K. (2016) Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 7626–7631.
41. Nunez,J.K., Harrington,L.B., Kranzusch,P.J., Engelman,A.N. and Doudna,J.A. (2015) Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature*, **527**, 535–538.
42. Wang,J., Li,J., Zhao,H., Sheng,G., Wang,M., Yin,M. and Wang,Y. (2015) Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*, **163**, 840–853.