Published online 14 December 2017

Nucleic Acids Research, 2018, Vol. 46, No. 2 ***873–885***
doi: 10.1093/nar/gkx1264

Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR–Cas effector complexes

Marnix Vlot[1], Joep Houkes[1], Silke J.A. Lochs[1], Daan C. Swarts[2], Peiyuan Zheng[3], Tim Kunne[1], Prarthana Mohanraju[1], Carolin Anders[2], Martin Jinek[2], John van der Oost[1], Mark J. Dickman[3] and Stan J.J. Brouns[1,4,*]

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands, [2]Department of Biochemistry, University of Zurich, Zurich, Switzerland, [3]ChELSI Institute Department of Chemical and Biological Engineering University of Sheffield, Sheffield, UK and [4]Department of Bionanoscience, Kavli Institute of Nanoscience, Van der Maasweg 9, Delft University of Technology, 2629 HZ Delft, The Netherlands

Received November 07, 2017; Revised December 04, 2017; Editorial Decision December 05, 2017; Accepted December 06, 2017

ABSTRACT

Prokaryotes encode various host defense systems that provide protection against mobile genetic elements. Restriction–modification (R–M) and CRISPR–Cas systems mediate host defense by sequence specific targeting of invasive DNA. T-even bacteriophages employ covalent modifications of nucleobases to avoid binding and therefore cleavage of their DNA by restriction endonucleases. Here, we describe that DNA glucosylation of bacteriophage genomes affects interference of some but not all CRISPR–Cas systems. We show that glucosyl modification of 5-hydroxymethylated cytosines in the DNA of bacteriophage T4 interferes with type I-E and type II-A CRISPR–Cas systems by lowering the affinity of the Cascade and Cas9–crRNA complexes for their target DNA. On the contrary, the type V-A nuclease Cas12a (also known as Cpf1) is not impaired in binding and cleavage of glucosylated target DNA, likely due to a more open structural architecture of the protein. Our results suggest that CRISPR–Cas systems have contributed to the selective pressure on phages to develop more generic solutions to escape sequence specific host defense systems.

INTRODUCTION

In many environments, bacteria are subject to strong selective pressure by bacteriophages (phages). The number of phages exceeds that of their hosts in most ecosystems, outnumbering them up to 150-fold (1). In response to this selective pressure, bacteria have developed a diverse palette of defense mechanisms including prevention of phage adsorption, blocking of DNA entry, restriction of phage DNA, and abortive infection mechanisms (2,3). Two of these defense mechanisms, restriction–modification (R–M) and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-CRISPR-associated (Cas) systems, act on the DNA level by selectively degrading invading DNA. Distinction between native and foreign DNA by R–M systems is often based on chemical modification (e.g. methylation) of adenines and cytosines in host genomic DNA, which protects host DNA from cleavage by specific restriction endonucleases. *Escherichia coli* K-12 harbors the type I restriction enzyme EcoK encoded by the *hsdR* gene that cleaves unmethylated DNA at 5′-AAC(N_6)GTGC and 5′-GCAC(N_6)GTT sequences and three type IV R–M systems encoded by *McrA*, *McrBC* and *mrr* that cleave methylated DNA (4,5). Together, these R–M systems limit horizontal gene transfer.

Recognition of invading DNA by type I, II and V CRISPR–Cas systems is based on base pairing of the invader DNA with CRISPR RNAs (crRNAs). These host encoded crRNAs guide effector complexes to complementary target DNA (also called protospacer) which is subsequently cleaved by the effector complex or by recruited nucleases (6,7). Even though all different types of CRISPR–Cas systems have a common role in immunity, they are structurally and mechanistically diverse. Based on this diversity, CRISPR systems are divided into two different classes and six different types (8,9). Type I CRISPR systems are the most abundant CRISPR type in nature (6), and comprise a multiprotein crRNA–effector complex (named Cascade) and a nuclease (Cas3) to bind target DNA. After target DNA recognition, Cascade recruits Cas3 that mediates tar-

*To whom correspondence should be addressed. Tel: +31 15 2783920; Fax: +31 15 2781202; Email: stanbrouns@gmail.com

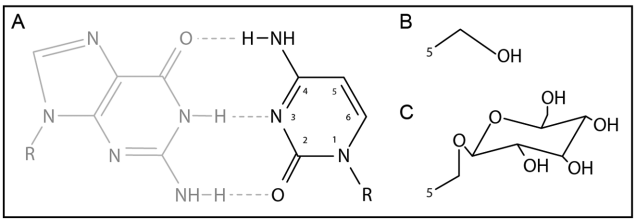

Figure 1. Modifications of nucleobases in phage T4 DNA. (**A**) Cytosine present in phage T4(C). (**B**) 5-Hydroxymethylation of cytosine present in phage T4(hmC). (**C**) Glucosyl-5-hydroxymethylation present in phage T4(ghmC).

get DNA degradation (10,11). The best characterized example of type II CRISPR–Cas systems is the Cas9 nuclease, a single effector protein which facilitates both crRNA-mediated DNA binding and DNA cleavage (8). Recently, type V-A CRISPR systems were discovered (12). Like type II systems, type V-A systems employ a single effector enzyme named Cas12a, which provides an interesting alternative to Cas9 in genome editing (12,13).

In response to the selective pressure posed by anti-viral defense systems in bacteria, phages have evolved several mechanisms to escape anti-viral defense systems. Phages can evade sequence-specific host defense systems by mutating target sequences (14–16). While this allows for efficient escape from restriction endonucleases, mutations in DNA sequences targeted by CRISPR–Cas systems can trigger a process called priming, which leads to an accelerated update of the CRISPR memory repertoire (17). Furthermore, phages may use recombination of their genomes to get rid of CRISPR target sites (18,19). In addition to mutation and recombination of DNA, phages can escape R–M systems by expressing inhibitory proteins. Such proteins can inhibit R–M immunity by degradation of R–M cofactors, masking of restriction sites, or modification of phage DNA (2). Additional strategies include DNA mimicry, as exemplified by the Ocr protein of phage T7 that mimics DNA to sequester EcoKI (20,21). Recently, phage-encoded proteins that inhibit CRISPR–Cas systems have been characterized (22–24). Inhibition of CRISPR–Cas systems by anti-CRISPR (ACR) proteins occurs via distinct mechanisms. Two Acr proteins (AcrF1 and AcrF2) were found to bind the type I-F Csy complex, inhibiting the binding of target DNA, while another Acr protein (AcrF3) binds to Cas3, blocking its interaction with Cascade to inhibit target degradation (25). Three recently identified families of ACRs are encoded by mobile elements in bacteria and are shown to inhibit *Neisseria meningitidis* Cas9 by direct binding (26). ACR proteins AcrIIA1-4 encoded by *Listeria monocytogenes* prophages prevent Cas9 binding and can be used to regulate the genome engineering activity of *Streptococcus pyogenes* Cas9 (27). ACR proteins that inhibit SpyCas9 have also been found in virulent bacteriophages (28).

Another way to escape host defense is found in T-even phages that infect *E. coli* (29). Phage T4 has evolved a pathway to bypass the R–M systems of *E. coli* by substitution of its genomic cytosines with 5-hydroxymethylcytosines (5-hmC) (Figure 1A and B). The dNTP synthesis complex (DSC) which comprises both host- and phage-encoded enzymes performs this substitution before DNA synthesis (30). Degradation of the host genome by T4-derived DenA and DenB nucleases provides an increased pool of dCTP, which is converted to dCMP by the phage-encoded dCTPase gp56. Subsequently, dCMP is converted into 5-hydroxymethyl-dCMP (5-hmdCMP) by phage-encoded Deoxycytidylate 5-hydroxymethyltransferase gp42 (31). Conversion of 5-hmdCMP to 5-hmdCDP is then catalyzed by the phage-encoded dNMP kinase gp1. Host-encoded nucleoside diphosphate kinase Ndk catalyzes the final conversion to 5-hmdCTP, which is incorporated by T4 DNA polymerase into the DNA as 5-hmC. To avoid immunity escape by 5-hmC modification, *E. coli* encodes the McrBC system, which specifically degrades 5-hmC modified DNA (32). However, T4 is impervious to this modification dependent system because 5-hmC residues are further modified with glucose moieties derived from uridine diphosphate glucose (UPDG). These are covalently attached to the 5-hydroxyl group of 5-hmC by phage-encoded alpha- and beta-glucosyl transferases, yielding glucosyl-5-hydroxymethylcytosine (5-ghmC) (33) (Figure 1C). As DNA glucosyl-5-hydroxymethylation provides an effective way to escape immunity by most R–M systems, the question arises whether it also protects against certain types of CRISPR–Cas immunity, and what the mechanistic basis for protection from CRISPR interference would be.

We have tested the effect of DNA glucosylation on the activity of type I-E, type II-A and type V-A CRISPR–Cas systems. We demonstrate that T4 with 5-ghmC DNA can escape type I-E and II-A interference by reducing target binding affinity of Cascade and Cas9, respectively. Interestingly, 5-ghmC modifications do not lower target binding affinity and cleavage efficiency of the type V-A effector nuclease Cas12a. The structural basis for the observed differences as well as potential applications for the described results are discussed.

MATERIALS AND METHODS

Bacterial strains and plasmid construction

Cells were made chemically competent using the RuCl method and transformed by applying a heat-shock as described in the QIA expressionist handbook (QIAGEN). For the experiments with Type I-E CRISPR–Cas, *E. coli* T7 Express (NEB) was transformed with pWUR400, pWUR797 and pWUR800, pWUR801 or pWUR802 (Supplementary Table S1) and used for plaque assays with phage T4. For the *in vivo* experiments with Type II-A CRISPR–Cas, *E. coli* T7 Express was transformed with pWUR805 and pWUR806, pWUR809 or pWUR810 (Supplementary Table S1).

Bacteriophage strains and propagation

Phage T4(ghmC) (CBS-KNAW Biodiversity Centre, Utrecht, Netherlands) was propagated in *E. coli* B834 (Su^-). Phage T4(C) (dCTPase$^-$, *denA*, *denB*, *alc*$^-$) was kindly provided by Prof. Elisabeth Kutter (Evergreen State College) and propagated in *E. coli* B834 or CR63 (Su^+). T4(hmC) (α-, β-glucosyltransferase) was kindly provided by Peter Weigele (NEB, Ipswich) and propagated in *E. coli* T7 Express (mcrC-mrr). The phage T4(C) mutant

we used in our analysis was a quintuple mutant of genes *alc*, *denB*, *denA* and *dCTPase*. Genes *denA* and *denB* encode endonuclease II and IV respectively and *denA* and *denB* mutations allow the phage to contain C DNA. The *alc* mutation allows for the production of late proteins, enabling bursts of phage.

Plaque assays and efficiency of plating calculation

Escherichia coli strains were grown at 37°C in Luria Broth (LB; 5 g/l yeast extract, 10 g/l tryptone, 5 g/l NaCl) at 180 rpm or on LB-agar plates containing 1.5% (w/vol) agar. When required, medium was supplemented with the following: ampicillin (Amp; 100 μg/l), chloramphenicol (Cam; 25 μg/ml), kanamycin (Kan; 50 μg/ml) or streptomycin (Str; 50 μg/ml). Bacterial growth was assessed spectrophotometrically at 600 nm (OD_{600}). To induce *cas* gene expression, IPTG (isopropyl β-D-1 thiogalactopyranoside) was added to the final concentration of 1 mM when the bacterial culture reached and OD_{600} of approximately 0.4. After 30 min of Cas protein and crRNA production, the cells were used in double agar layer plaque assays with phage T4(ghmC), T4(hmC), and T4(C). Plaque assays were performed in triplicate. The sensitivity of the host to phage infection was calculated as the efficiency of plaquing, which is the plaque count ratio of a strain containing an anti-T4 CRISPR, to that of a strain containing a CRISPR with non-targeting spacers. Anti-T4 spacers were designed by picking a PAM-flanked protospacer in two essential T4 genes (gene 19 and gene 22). CRISPR protospacer sequences are provided in Supplementary Table S2. A two tailed t-test was performed to calculate whether the differences were significant.

The presence or absence of modifications of T4 DNA was verified by analysis using restriction enzymes. Isolated DNA was incubated with enzymes that are either sensitive or not sensitive to 5-hmC or 5-ghmC. Our analysis confirmed that our T4 stock contains 5-ghmC DNA and our T4(C) stock contains C DNA. The presence of 5-hmC in phage T4(hmC) was verified by plating on *E. coli* expressing the McrBC restriction enzyme, which is active on 5-hmC DNA but not on 5-ghmC DNA. Inability of T4(hmC) to propagate reveals the presence of 5-hmC DNA.

Design and synthesis of modified dsDNA

All *in vitro* assays were done using a 98 bp dsDNA target containing the spacer 8 (sp8) sequence (Supplementary Table S2). The target was amplified using the primers BG8415 & BG8416 (Supplementary Table S2). PCR products were generated using both Q5 high-fidelity DNA polymerase and Taq DNA polymerase (NEB). hm^5-dCTP (Bioline) was used in the PCR reactions to obtain hydroxymethylated targets. PCR conditions were optimised for each of the modified dNTPs used. Typical PCR conditions were as follows: 50 μl PCR reactions with final concentrations dNTPs 200 μM, primers 0.5 μM, DNA template 100 ng, 1.25 U Taq polymerase. The following parameters were used: denaturation at 95°C for 30 s followed by 30 cycles of 95°C for 30 s, 46°C for 60 s, 68°C for 60 s with a final extension at 68°C for 3 min. For Q5 DNA polymerase (1 U) the following parameters were used; 98°C 30 s denaturation, followed by 30 cycles at 98°C for 30 s 61°C for 30 s, 72°C for 20 s with a final extension at 72°C for 2 min. The PCR products were subsequently purified by QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions. Purified samples were quantified using a NanoDrop 2000C (Thermo Scientific) and analyzed by native PAGE.

Three different spacers were designed for the target sequence for Type I-E, Type II-A and Type V CRISPR systems. Each of these spacers was designed to have a suitable PAM for the respective CRISPR Type (34). The designed spacers can be found in Supplementary Table S2.

Enzymatic glucosylation of the purified 5-hmC dsDNA was used to synthesize DNA with glucosylated 5-hmC. T4 Phage β-glucosyltransferase (NEB) 1 U was incubated with approximately 1 μg of 5-hmC dsDNA, 40 μM UDP-glucose at 37°C for 16 h. The DNA products were subsequently purified using QIAprep Spin Miniprep Kit (Qiagen) and analysed by native PAGE.

DNA targets containing 5-hmC modifications in the PAM region were produced by annealing oligonucleotides BG6508–6510 to BG6506. Oligonucleotides BG6508–6510, containing 5-hydroxymethyldeoxycytidine, were purchased from Integrated DNA Technologies (IDT).

Target DNA was subsequently glucosylated as described above.

Radiolabeling of target DNA

Target DNA containing unmodified cytosine, 5-hmC and 5-ghmC were 5′ radioactively labeled using T4 Polynucleotide Kinase (PNK) (NEB). Conditions are 1× PNK reaction buffer (NEB), 50 pmol [γ-32P] ATP (Perkin Elmer) and 25 pmol target DNA in a 50 μl reaction. The reaction was carried out at 37°C for 30 min and subsequently inactivated at 65°C for 20 min.

Preparation of crRNA and sgRNA

The Cas9 sgRNA was made by *in vitro* transcription (IVT) of partly overlapping primers. DNA oligonucleotides (oligos) used for IVT were PAGE purified (Supplementary Table S3). IVT was performed using the HiScribe™ T7 Quick High Yield RNA Synthesis Kit (NEB) in 30 μl containing 10 μl NTP buffer mix, 1 μg DNA and 2 μl T7 polymerase. The reaction was incubated for 4 h at 37°C, thereafter 20 μl Milli-Q and 2 units DNase I (NEB) were added to remove target DNA for 15 min at 37°C. 10 μl 6× loading dye (Thermo Scientific) was added to both samples for separation and isolation from a denaturing PAGE gel. The samples were heated for 5 min at 95°C and loaded on a 7 M urea 8% PAGE gel, together with a Low Range ssRNA Ladder (NEB). The gel was run for 3 h at 15 mA in 1× TBE. After 3 h, the gel was stained using 1× SYBR gold (Thermo Scientific) in 1× TBE for 10 min. For both RNA oligos, the band at 103 nt was cut from gel.

The gel fragments were ground using a pipet tip. 1 ml of RNA elution buffer (0.5 M NaAc, 10 mM $MgCl_2$, 1 mM EDTA, 0.1% SDS) was added and incubated for 2 h at 37°C. The sample was transferred to a gel extraction column (Zymoclean) and centrifuged at 13 200 × g for 1 min. The flow-through was purified on a Microcon 30 column (Millipore)

by centrifugation at 16 000 × g for 20 min at 4°C. RNA was washed twice using 500 μl Milli-Q and centrifuged at 16 000 × g for 20 min at 4°C. RNA was resuspended on a filter column in 50 μl Milli-Q, and retrieved by centrifuging the column upside down for 5 min at 6000 × g. RNA quantity and quality were analysed using NanoDrop and Denaturing PAGE. The 48 nt crRNA and non-complimentary crRNA for Cas12a assays were synthesized RNA Oligos (Sigma, Supplementary Table S3).

Cascade and Cas3 degradation reactions

Cascade purification was performed as described in (10). Cas3 was purified as described in (35). Target DNA constructs were incubated with 100 μM Cascade and 10 μM Cas3 in reaction buffer (10 mM HEPES pH 7.5, 60 mM KCl, 50 μM $CoCl_2$, 10 mM $MgCl_2$ and 2 mM ATP) at 37°C for 2 h. Reactions were stopped by transferring the tubes to ice and addition of 4 μl 6× loading dye (Thermo Scientific). Reaction products were run on a 6% acrylamide gel (with 7 M urea and 1× TBE). The gel was run in 1× TBE for ∼4 h at 15 mA and subsequently exposed for 48 hours in a phosphor imaging cassette (Molecular Dynamics) at –20°C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Bio-Rad).

Cas9 cleavage reactions

For Cas9 Cleavage reactions 33 nM Cas9 (NEB), 120 nM sgRNA and 1× Cas9 buffer (NEB) were pre-incubated for 20 min at 37°C. Subsequently 5′ ^{32}P-radiolabeled target DNA was added to a final concentration of 3 nM and the reaction was incubated for 2 h at 37°C. 30 μl 2× formamide loading dye (95% formamide, 0.125% bromophenol blue) was added and the complete samples were heated to 95°C for 5 min and loaded on an 8% acrylamide gel (with 7 M urea and 1× TBE). The gel was run in 1× TBE for ∼4 h at 15 mA and subsequently exposed for 16–48 h in a phosphor imaging cassette (Molecular Dynamics) at –20°C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Bio-Rad).

Cas12a protein purification

A codon optimized Cas12a gene was cloned into a bacterial expression vector [6-His-TEV-Cas12a, a pET-based vector that was a gift from Scott Gradia (Addgene plasmid # 29653)]. One liter of LB growth media with 100 μg/ml ampicillin was inoculated with 10 ml overnight culture Rosetta (DE3) (EMD Millipore) cells containing the Cas12a expression construct. Growth media plus inoculant was grown at 37°C until the cell density reached 0.5 OD_{600}, then the temperature was decreased to 20°C. Growth was continued until OD_{600} reached 0.6 when a final concentration of 500 μM IPTG was added to induce Cas12a expression. Expression took place for 14–18 h before harvesting cells and freezing them at –20°C until purification.

Cell paste was suspended in 20 ml of Lysis Buffer (50 mM NaH_2PO_4 pH 8, 500 mM NaCl, 1 mM 2-mercaptoethanol, 10 mM imidazole) supplemented with protease inhibitors (Roche complete, EDTA-free) and lysozyme. Once homogenized, cells were lysed by sonication (Bandelin Sonoplus) and then centrifuged at 16 000 × g for 1 h at 4°C to clear the lysate. The lysate was filtered through 0.22 μm filters (Mdi membrane technologies) and applied to a nickel column (HisTrap HP, GE lifesciences), washed and then eluted with 250 mM imidazole. Fractions containing protein of the expected size were pooled and dialyzed overnight into the dialysis buffer (250 mM KCl, 20 mM HEPES/KOH, 1 mM DTT). After dialysis, sample was diluted 1:1 in 10 mM HEPES/KOH pH 8, and loaded on a heparin FF column pre-equilibrated in IEX-A buffer (150 mM KCl, 20 mM HEPES/KOH pH 8). Column was washed with IEX-A and then eluted with a gradient of IEX-C (2 M KCl, 20 mM HEPES/KOH pH 8). The sample was concentrated to 700 μl prior to loading on a gel filtration column (HiLoad 16/600 Superdex 200) via FPLC (AKTA Pure). Fractions from gel filtration were analyzed by SDS-PAGE; fractions containing Cas12a were pooled and concentrated to 200 μl (50 mM Tris–HCl pH 8, 2 mM DTT, 5% glycerol, 500 mM NaCl) and either used directly for biochemical assays or frozen at –80°C for storage.

Cas12a cleavage reactions

Cleavage assays were done by pre-incubating 100 nM Cas12a with 400 nM crRNA in 1× Cas9 buffer (NEB) for 20 min at 37°C. Thereafter 8 nM 5′ ^{32}P radiolabeled target DNA was added in a 30 μl reaction and incubation was done at 37°C for 2 h. 30 μl 2× formamide loading dye was added and the complete samples were heated to 95°C for 5 min and loaded on an 8% acrylamide gel (with 7 M urea and 1× TBE). The gel was run in 1× TBE for ∼4 h at 15 mA and subsequently exposed for 16–48 h in a phosphor imaging cassette (Molecular Dynamics) at –20°C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Biorad).

Cascade electrophoretic mobility shift assays

Cascade electrophoretic mobility shift assays were performed based on (36). Incubation reactions were performed with Cascade–crRNA (sp8) complex concentrations of 0, 1, 3, 10, 30, 100, 300 and 1000 nM. Dilutions of the Cascade stock solution (2.78 mg/ml) were made in equilibration buffer (20 mM HEPES pH 7.5, 75 mM NaCl, 1 mM DTT). 3 nM 5′ ^{32}P radiolabeled target was incubated with Cascade complex in 1× Equilibration buffer in a 30 μl reaction. The reaction was incubated for 30 min at 37°C. 6 μl 6× loading dye (Thermo Scientific) was added and the complete samples were loaded on a native 8% acrylamide gel containing 1× TBE. The gel was run at 15 mA for 2–3 h, after which the gel was placed in a phosphor imaging cassette and stored for 16–48 h at –20°C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Bio-Rad).

dCas9 and dCas12a electrophoretic mobility shift assays

Electrophoretic mobility shift assays with Cas9 and Cas12a were done using catalytically dead versions of the proteins, dCas9 and dCas12a (12,37). 1 μM dCas9 and dCas12a were incubated with 4 μM sgRNA and crRNA respectively in

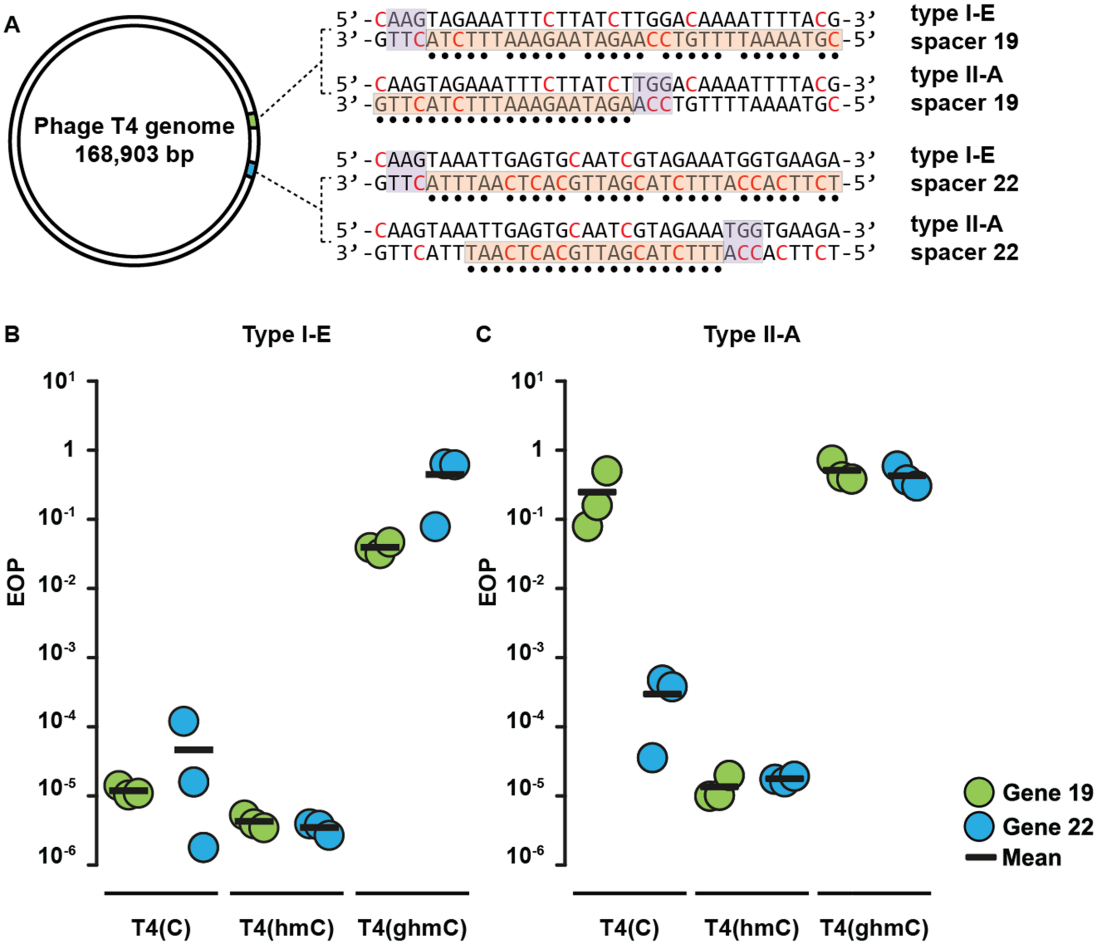

Figure 2. Efficiency of plaquing (EOP) assays of T4 phages on *E. coli* cells expressing CRISPR–Cas systems targeting T4 genes 19 or 22. (**A**) Representation of the circular genome of bacteriophage T4. Positions of protospacer sequences and PAMs are shown (indicated by orange bars and purple bars respectively). Glucosyl-5-hydroxymethylcytosines are indicated in red and nucleotides that base pair with crRNA are indicated by black dots. (**B**) EOP of T4 phages on *E. coli* cells expressing targeting Cascade complexes normalized to EOP on non-targeting strains. (**C**) EOP of T4 phages on *E. coli* cells expressing targeting Cas9 proteins normalized to EOP on non-targeting strains.

binding buffer (20 mM HEPES, pH 7.5, 250 mM KCl, 2 mM $MgCl_2$, 0.01% Triton X-100, 0.1 mg/ml bovine serum albumin, 10% glycerol) for 10 min at 25°C as described in (38). A dilution series was made with 0.1, 0.3, 1, 3, 10, 30, 100 and 300 nM final protein concentration using binding buffer. 3 nM 5′ ^{32}P-radiolabeled target DNA was added in a 30 μl reaction and incubated for 10 minutes at 37°C. 10 μl 6× loading dye (Thermo Scientific) was added and the complete samples were loaded on a 12% native acrylamide gel containing 1× TBE. The gel was run in 1× TBE at 15 mA for 2–3 h. Subsequently, the gel was exposed for 16–48 h in a phosphor imaging cassette (Molecular Dynamics) at –20°C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Biorad)

Structural Modeling of modified target DNA binding to Cascade, Cas9 and Cas12a

Structural coordinates of glucosyl-5-hydroxymethyl modified thymine was isolated from an existing model (PDB 308D; (39)). The modified nucleotide was statically modelled onto target strand (TS) and non-target strand (NTS) nucleotides in structures of *E. coli* Cascade (PDB: 5H9F; (40)), *S. pyogenes* Cas9 (PDB: 5F9R; (41)), and *F. novicida* Cas12a (PDB: 5NFV; (42)) in WinCoot 0.8.3 using Least Squares Fit Superpose. Clashes between the DNA modifications and the polypeptide chain were identified using the MolProbity (43) output of the Comprehensive Validation Tool within Phenix (44). Identified clashes between the glucosyl-5-hydroxymethyl modification and polypeptide residues are schematically indicated in Figure 7.

RESULTS

Phages escape CRISPR–Cas interference by DNA glucosylation

The majority of phages that contain 5-ghmC modifications in their DNA infect gammaproteobacteria (45). Hence, anti-viral defense systems that are effective against 5-ghmC DNA-containing phages may be expected among this class of bacteria. To this end we assessed the ability of the native type I-E CRISPR–Cas system of the gammaproteobacterium *E. coli* to provide resistance against phage T4. Two

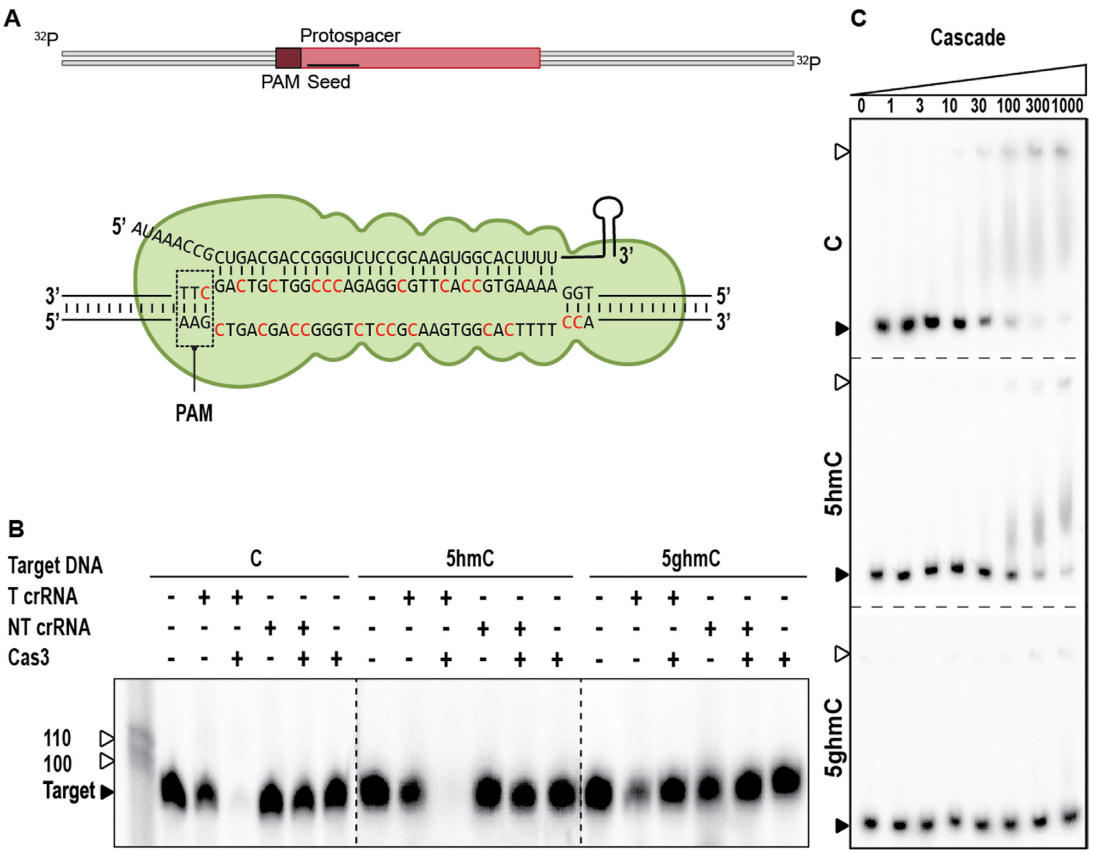

Figure 3. Effect of T4 DNA modifications on type I-E CRISPR–Cas sgRNA mediated DNA targeting. (**A**) Schematic of DNA targeting and R-loop formation by Cascade. Modified cytosine residues are indicated in red. (**B**) Cleavage assay of Cas3 in conjunction with Cascade on 98 bp modified targets, indicated by black arrow. The marker is indicated by white arrows. Cascade effector complexes are loaded with either targeting crRNA (T crRNA) or non-targeting crRNA (NT crRNA). Restriction products of Cas3 are of undefined length. (**C**). Electrophoretic Mobility Shift Assay (EMSA) of Cascade on target DNA containing C, 5-hmC or 5-ghmC (indicated by black arrow) at increasing protein concentrations [nM]. Fraction of bound target is indicated by white arrows, dotted lines represent separate gels.

plasmids expressing crRNAs against essential T4 genes were constructed (Supplementary Table S2). The crRNAs target T4 gene 19 (crRNA19), encoding a tail tube protein, and T4 gene 22 (crRNA22), encoding the prohead core protein. *Escherichia coli* cells expressing Cascade and Cas3 were transformed with the plasmid encoding the crRNAs and were subsequently tested for their sensitivity to infection by a panel of T4 phages. Phage T4(C) contains unmodified cytosine, T4(hmC) contains 5-hmC, and T4(ghmC) contains 5-ghmC. Infectivity of the phages was normalized against *E. coli* cells containing a CRISPR array harboring a non-targeting spacer (Figure 2). Efficiency of plaquing (EOP) by T4(C) was reduced 10^4-fold by cells expressing targeting Cascade as compared to cells expressing non-targeting Cascade. Similarly, EOP by T4(hmC) was reduced 10^5-fold by cells expressing targeting Cascade. In contrast, EOP by T4(ghmC) was lowered only 10-fold by cells expressing targeting Cascade. We therefore conclude that CRISPR–Cas type I-E resistance is severely inhibited by 5-ghmC modifications, but not by 5-hmC modifications of phage DNA.

Next, we tested whether the type II-A CRISPR–Cas system of *Streptococcus pyogenes* mediates immunity against T4 phages in *E. coli*. *E. coli* cells expressing Cas9 and tracrRNA were transformed with plasmids encoding crRNAs targeting T4 genes 19 and 22 and subsequently tested for sensitivity to infection by phages T4(C), T4(hmC), and T4(ghmC) (Figure 2B). Efficiency of plaquing (EOP) by T4(C) and T4(hmC) is decreased 10^4-fold in cells expressing targeting Cas9 compared to non-targeting Cas9. Although both chosen crRNAs mediated efficient targeting of T4(hmC), crRNA19 provided lower resistance than crRNA22, indicating spacer-specific interference efficiencies on normal DNA. In contrast to EOP of T4(C) and T4(hmC), EOP of T4(ghmC) is similar between cells expressing targeting or non-targeting Cas9. We therefore conclude that, like type I-E systems, type II-A CRISPR–Cas systems are severely impaired by phage DNA glucosylation.

Escaping phages carry seed mutations

We noticed that each of the three T4 phages generated plaques on CRISPR targeting strains, indicating that these phages escaped CRISPR–Cas immunity. To determine the nature of this escape, we sequenced the target site of the different plaque-forming T4 phages (Tables S4 and S5). The majority (23/34) of T4(C) and T4(hmC) phages that escaped from type I-E or type II-A immunity contained a

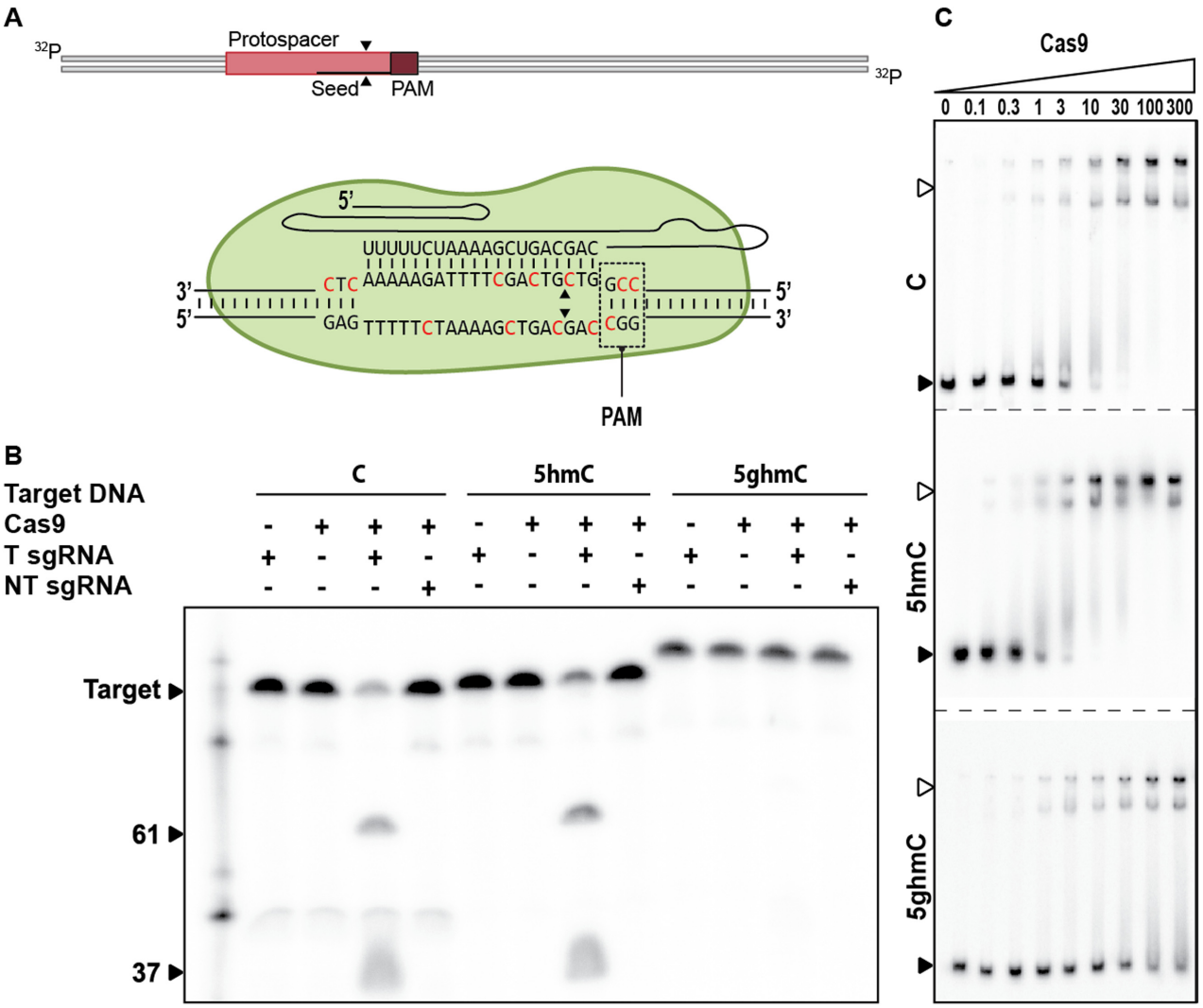

Figure 4. Effect of T4 DNA modifications on type II-A CRISPR–Cas sgRNA mediated DNA targeting. (**A**) Schematic of DNA targeting by Cas9. Modified cytosine residues are indicated in red. Cleavage sites are indicated by black arrows. (**B**) Cleavage assay of Cas9 on 98 bp modified targets (indicated by black arrow). Cas9 is loaded with either targeting sgRNA (T sgRNA) or non-targeting sgRNA (NT sgRNA). Restriction products of Cas9 are 61 and 37 bp. (**C**) EMSA of dCas9 on target DNA containing C, 5-hmC or 5-ghmC (indicated by black arrow) at increasing protein concentrations [nM]. Fraction of bound target is indicated by white arrows.

single mutation in the target sequences. Almost all mutations (20/23) were located in the seed region of the target sequences. This is a region in the target site that is highly intolerant for mismatches during CRISPR mediated immunity (46). Surprisingly, for both systems none of the mutations were located in the PAM (Protospacer adjacent motif). The PAM is a short motif adjacent to the DNA target sequence (PAM) important for efficient target recognition and avoidance of self-targeting, and is commonly prone to accumulating escape mutations.

Thus, several target DNA sequences in T4(C) and T4(hmC) phages escaping type I-E or II-A systems did not contain mutations. To ensure these phages were not revertants in which the T4(ghmC) genotype was restored, the dCTPase gene of the escape mutants was PCR amplified and sequenced. This confirmed that the T4(C) phages still had the dCTPase deficient genotype, which prevents incorporation of 5-ghmC, and suggests the mutants escaped CRISPR immunity in another way (47). In contrast to T4(C) and T4(hmC) phages that successfully infected *E. coli*, none of the T4(ghmC) phages infecting *E. coli* expressing type I-E and type II-A systems had mutations in the DNA targeted by the CRISPR systems. Sequencing of escape mutant phages shows that T4(C) and T4(hmC) phages mostly escape CRISPR immunity by mutating the target sequence. In contrast, T4(ghmC) infection does not require mutation of DNA for efficient infection. This indicates that the 5-ghmC modification itself allows efficient escape from type I-E and type II-A CRISPR–Cas systems

DNA glucosylation inhibits Cascade target binding

To understand the molecular basis for escape from CRISPR–Cas immunity, we reconstituted CRISPR inter-

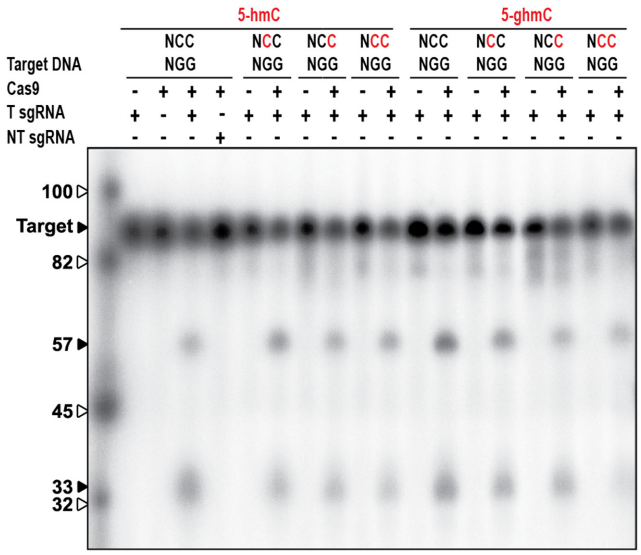

Figure 5. Effect of T4 DNA modifications of PAM-complementary cytosines on type II-A CRISPR–Cas sgRNA mediated DNA targeting. Cleavage assay of Cas9 on target DNA containing 5-hmC (indicated in red). Cas9 is loaded with either targeting sgRNA (T sgRNA) or non-targeting sgRNA (NT sgRNA). Restriction products of Cas9 are 57 and 33 bp. The marker is indicated by white arrows.

ference with type I-E components *in vitro* (Figure 3A). After pre-incubation of crRNA-bound Cascade with unmodified, 5-hmC, or 5-ghmC target DNA, the capacity of Cas3 to degrade these targets was assessed. Cas3, in conjunction with Cascade, was able to degrade 5-hmC DNA and the unmethylated control DNA, but no degradation of 5-ghmC target DNA was observed (Figure 3B). To determine whether this was due to inhibition of Cas3 or rather due to inhibition of target binding by Cascade, we performed electrophoretic mobility shift assays (EMSAs) with Cascade (Figure 3C). The EMSAs showed that sequence-specific binding of targets containing 5-hmC is consistently reduced compared to targets containing unmodified cytosines. In addition, we were unable to observe binding of 5-ghmC targets, indicating that phages with glucosylated DNA escape from type I-E CRISPR immunity by preventing binding of Cascade to the target DNA. Even though reduced binding affinity of Cascade to targets containing 5-ghmC is reduced, thereby allowing escape from CRISPR–Cas interference, inability of Cas3 to cleave glucosylated DNA cannot be excluded even when stable binding of target DNA by Cascade would occur.

Glucosylated DNA affects Cas9 binding and prevents target cleavage

Next, we investigated if DNA modifications can prevent DNA targeting by type II-A Cas9 *in vitro* (Figure 4A). The ability of Cas9 to degrade unmodified, 5-hmC and 5-ghmC targets was assessed by incubation of a reconstituted Cas9–sgRNA-loaded complex with target DNAs. As for Cascade/Cas3, Cas9–sgRNA complexes were able to cleave the 5-hmC DNA and the unmodified control DNA, but not the 5-ghmC DNA (Figure 4B). To determine the molecular basis for inhibition of 5-ghmC DNA cleavage, EMSAs were conducted using a catalytically inactive version of Cas9 (dCas9) (Figure 4C). Surprisingly, the assays showed that the sequence-specific binding of 5-hmC target DNA is consistently increased compared to binding of unmodified DNA targets. In contrast, the binding affinity of 5-ghmC target DNA is consistently decreased compared to targets with unmodified DNA. Combined, these results indicate that 5-ghmC partially affects type II-A interference by lowering the binding affinity of Cas9–sgRNA complexes to the target DNA. Additionally, glucosylation may affect the ability of Cas9 to restrict target DNA by protecting the scissile bonds from the catalytic residues, rendering Cas9–sgRNA complex-mediated interference even less effective.

Glucosyl-5-hydroxymethylation of the PAM does not abolish cleavage by Cas9

We next investigated whether glucosyl-5-hydroxymethylation of cytosines in the PAM region is sufficient to inhibit DNA cleavage by Cas9. Interactions between the target DNA and Cas9 are initiated at the PAM by two arginines probing the major groove for the two guanines in the 5′-NGG-3′ PAM (48). Upon PAM recognition, Cas9 sequentially unwinds the protospacer dsDNA and forms an R-loop structure between the target DNA strand and the guide RNA. Based on their location in the major groove of the DNA, we hypothesized that glucosyl-5-hydroxymethylation of the two cytosines (3′-NCC-5′) complementary to the guanines of the PAM could inhibit the PAM recognition process, and could abolish target cleavage. We utilized target DNA containing 5-hmC or 5-ghmC modifications either proximal to the protospacer (3′-N^{hm}CC-5′), distal to the protospacer (3′-NC^{hm}C) or at both cytosine residues in the PAM-complementary region (3′-N^{hm}C^{hm}C-5′). We found that Cas9-mediated cleavage of target DNA containing 5-hmC or 5-ghmC opposite the guanines in the 5′-NGG-3′ PAM was not inhibited by such modifications. Thus, cytosine glucosylating modifications of the PAM-complementary sequence are not sufficient and do not contribute to escape from Cas9-mediated interference (Figure 5).

DNA cleavage by Cas12a is not inhibited by glucosyl-5-hydroxymethylation

To investigate the activity of type V-A Cas12a systems to cleave glucosylated DNA, we performed DNA cleavage assays using Cas12a from *Francisella novicida* (FnCas12a; Figure 6A). Strikingly, DNA cleavage assays with Cas12a-crRNA complex and various DNA targets reveal Cas12a activity is not affected by 5-hmC or 5-ghmC modifications of the target (Figure 6B). Furthermore, EMSAs with catalytically inactive Cas12a containing mutations E1006A and R1218A (dCas12a) (42) shows similar binding affinities for unmodified, 5-hmC, and 5-ghmC target DNA (Figure 6C). These results indicate that DNA targeting by Cas12a–crRNA complexes, in contrast to Cascade and Cas9–crRNA effector complexes, is not inhibited by 5-hmC or 5-ghmC modifications in the DNA.

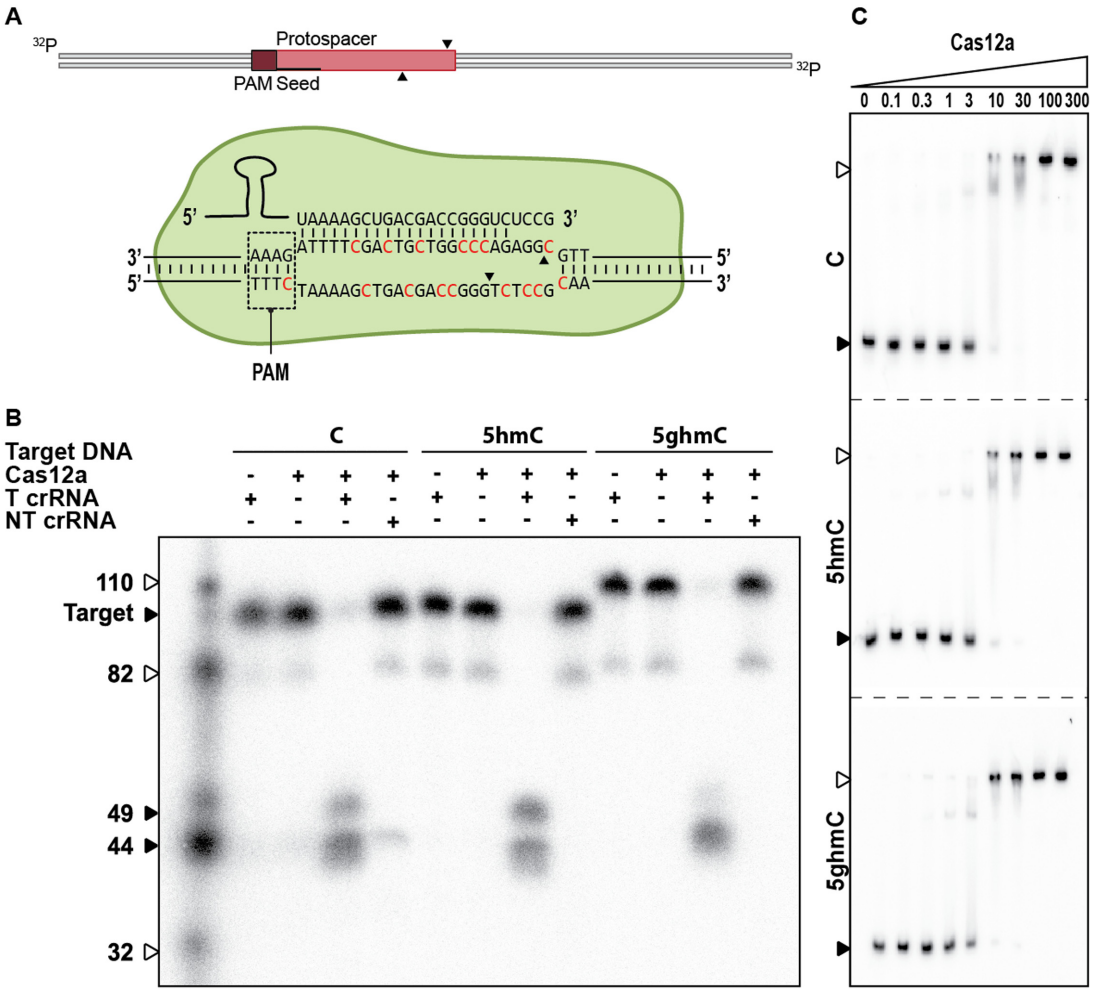

Figure 6. Effect of T4 DNA modifications on type V-A CRISPR–Cas sgRNA mediated DNA targeting. (**A**) Schematic of DNA targeting by Cas12a. Modified cytosine residues are indicated in red. Cleavage sites are indicated by black arrows. (**B**) Cleavage assay of Cas12a on 98 bp modified targets (indicated by black arrow). Cas12a is loaded with either targeting crRNA (C crRNA) or non-targeting crRNA (NC crRNA). Cleavage products of Cas12a are 49 and 44 bp. The marker is indicated by white arrows. (**C**) Electrophoretic Mobility Shift Assay (EMSA) of Cas12a on target DNA containing C, 5-hmC or 5-ghmC (indicated by black arrow) at increasing protein concentrations [nM]. Fraction of bound target is indicated by white arrows.

Structural modelling of modified target DNA into crRNA–effector complexes

In order to further understand the contrasting effects of DNA glucosylation on CRISPR–Cas effector complexes we analyzed potential steric clashes of DNA modifications in the target and non-target strand of the DNA at each nucleotide position by statically modelling glucosyl-5-hydroxymethyl modified DNA in existing structural models of ternary effector complexes (*E. coli* Cascade (PDB: 5H9F (40)), SpCas9 (PDB: 5F9R (41)) and FnCas12a (PDB: 5NFV (42))). As structures of 5-ghmC are not available, we used the structure of a 5-ghmT (PDB: 308D), in which the glucosyl-hydroxymethyl modification is also positioned on carbon 5 of the base. Static modelling does not take into account potential flexibility of polypeptide and nucleic acid chains, but allows for the identification of potential clashes and thereby could provide an idea about how effector complexes might accommodate modified DNA substrates.

In both the Cascade and Cas9 models, multiple clashes are observed between the polypeptide chain and 5-ghmC modifications of nucleotides in both target DNA strands (Figure 7A and B, Supplementary Table S6). Clashes with Cascade are observed for 17 out of 32 target strand nucleotides (TS, complementary to the crRNA) and for 3 out of 7 of the non-target strand (NTS) nucleotides of which the position is identified. Clashes with Cas9 are observed for 3 out of 20 TS nucleotides and for 5 out of 9 of the NTS nucleotides that are structurally defined in the crystal structure of Cas9. Several of the modifications in both Cascade and Cas9 that cause steric clashes are located in the segment of the DNA that base-pairs with the crRNA seed sequence. Because interactions with the seed sequence are essential for R-loop formation and efficient targeting by both Cascade and Cas9 (46), such clashes can explain why DNA targeting by Cascade and Cas9 is inhibited by 5-ghmC modifications of the target DNA. In contrast, such clashes are not observed in the Cas12a model, in which 5-ghmC modifica-

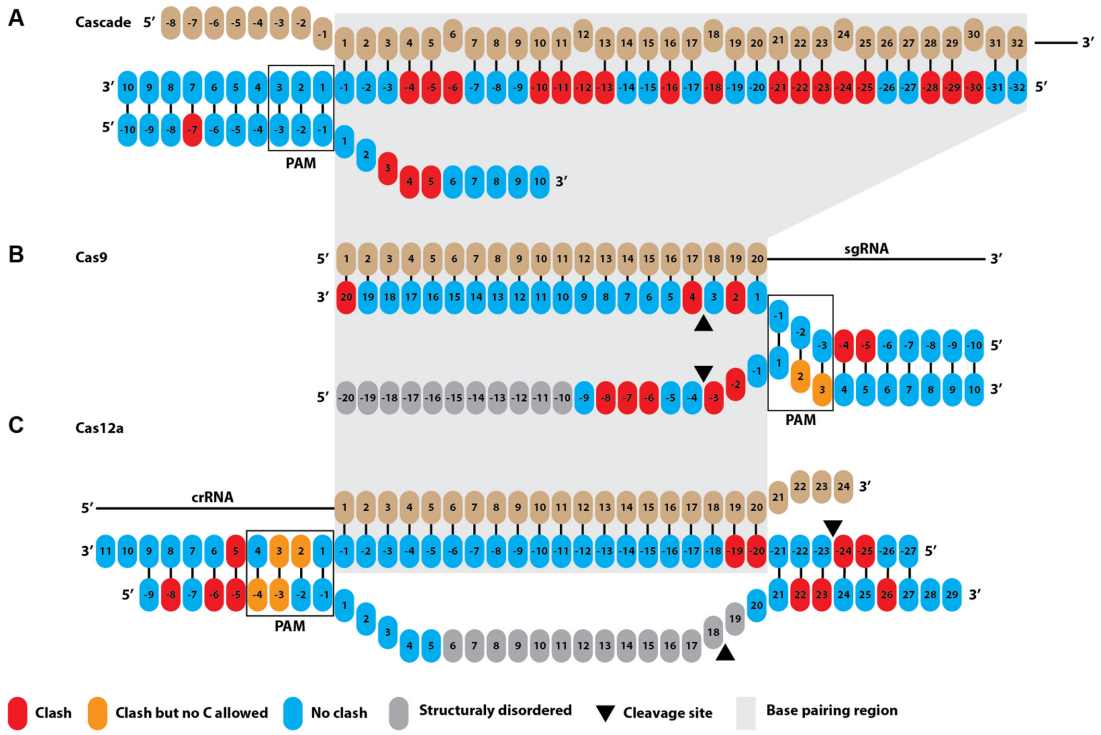

Figure 7. Potential steric clashes between target nucleotide 5-ghmC modifications and CRISPR effector proteins. (**A**) Multiple clashes (indicated in red) are observed between the polypeptide chains of Cascade and 5-ghmC modifications of nucleotides in the target strand (TS, complementary to the crRNA) and non-target strand (NTS). (**B**) Clashes are mostly observed between the polypeptide chains of Cas9 and 5-ghmC modifications of nucleotides in the seed region. (**C**) No clashes are observed in the seed region for Cas12a.

tions can be accommodated at almost all positions within the crRNA-target DNA heteroduplex (Figure 7C).

It should be noted that for both Cas9 and Cas12a, clashes between the 5-ghmC modifications and the polypeptide chain are observed when 5-ghmC modifications are modelled onto the PAM nucleobases (Figure 7B and C). However, specifically these PAMs cannot contain cytosines at these positions (Figure 7) since Cas9 and Cas12a recognize 5′-NGG-3′ and 5′-(T)TTN-3′ PAMs, respectively. In line with our biochemical experiments with Cas9 (Figure 5), modifications of the cytosine nucleotides opposite to the 5′-NGG-3′ PAM do not appear to introduce clashes with the polypeptide chain of Cas9 in our models. In both models, additional clashes can be found upstream the protospacer-PAM segment, and within the PAM-distal protospacer region. Because these segments of the DNA are generally not essential for efficient DNA targeting (Cas9: (49)/Cas12a: (42,50,51)), we hypothesize that the potential clashes between 5-ghmC modifications and the polypeptide chain in our models will have a limited effect on DNA targeting. In summary, we conclude that 5-ghmC modifications in the target DNA are likely to result in clashes with Cascade and Cas9 and impair their DNA binding activity. However, Cas12a appears to be able to accommodate modified nucleobases due to a much more open structure. These models corroborate our experiments that demonstrate Cascade and Cas9 activity, but not Cas12a activity, is inhibited by 5-ghmC modification of target DNA.

DISCUSSION

Phages have evolved diverse systems that allow them to counteract bacterial defense mechanisms. Evasion from CRISPR–Cas defense can occur by mutation of the protospacer or PAM sequence (14–16), target site deletion (52), genome recombination (19), or by production of anti-CRISPR proteins (23,25,26). Here, we show how chemical modifications of DNA nucleotides can render phages resistant to interference by specific CRISPR–Cas types. Our results demonstrate that CRISPR–Cas type I-E and type II-A interference is severely impaired by glucosyl modification of the cytosine bases (5-ghmC) in T4 phage DNA. However, the type V-A effector protein Cas12a remains able to bind and cleave T4 DNA even if it is glucosylated, suggesting the possibility that type V systems have evolved to combat covalently modified phages.

Phage encoded DNA modifications are well known antagonists of microbial defense systems, most notably restriction–modification systems (53). Apart from the known inhibitory effect of DNA modifications on R–M systems, the effect on CRISPR–Cas systems has been studied for some DNA modifications. Small modifications such as N6-methylation of adenine in 5′-GATC-3′ sequences do not prevent interference by the type II-A CRISPR–Cas system in *Streptococcus thermophilus* (54). Later, Yaung and colleagues showed that DNA modifications of T4 phages do not protect the phage from Cas9 interference (55). By contrast and in line with our results, Bryson and colleagues showed that Cas9 interference can be inhibited by glucosy-

lation of DNA when using the naturally occurring combination of crRNA and trans-activating crRNA (tracRNA), and to a lower extent using the engineered single guide RNA (sgRNA) (56). Also, the degree of inhibition is strongly dependent on the selected target DNA sequence, and therefore the positions of base modifications. While this manuscript was in preparation, Strotskaya and colleagues showed that no interference nor CRISPR array expansion was observed when T4 was targeted by the type I-E system in *E. coli* (57).The insensitivity of T4 to CRISPR interference was attributed to either DNA modifications or to yet unknown phage mechanisms. Here we demonstrate that the insensitivity is a result of DNA modification.

The inability of Cascade and Cas9 to bind and cleave 5-ghmC DNA is likely due to steric clashes between the glucose modifications of the protospacer segment of the target DNA and the polypeptide chain of the effector complexes. Our results show that Cas9 can still read the PAM nucleotides if only the PAM-complementary cytosines are glucosylated, which is in agreement with structural modeling. The inability of Cas9 to cleave fully glucosylated targets suggests that after successful PAM recognition, subsequent recognition of the protospacer is hampered.

The effect of DNA modifications on CRISPR–Cas interference is most likely dependent on the target DNA sequence and thereby the degree of DNA modifications. Although Cas9 and Cas12a target roughly the same region of our tested target DNA, the chosen target sites are different and result in different degrees of DNA modifications in the seed regions of the protospacer (12,58,59). However, our structural modeling suggests that Cas12a has an intrinsically more open architecture and we predict that it can accommodate and better deal with bulky DNA modifications regardless of their position in the protospacer.

In addition to steric hindrance effects, changes in chemical presentation of nucleobases can result in decreased binding affinities. Changes in the chemical signature of the major and minor groove by replacement of guanines in the PAM with purine analogs is shown to attenuate binding affinities of the type I-F Csy complex in *Pseudomonas aeruginosa* (60). Similarly, glucosylation of DNA is known to change the chemical signature and stability of the DNA duplex. Hydrogen bonds can be formed between the side groups of the glucosyl moieties and neighboring bases (61). These interactions alter the roll, slide, and twist angles of base pairs (62) which could impair R-loop formation and thereby prevent cleavage of the modified targets.

While phage T4 contains 70% α-glucosylated and 30% β-glucosylated DNA, this study only examined the effect of β-glucosyl linkages *in vitro* (33). The effect of α-glucosyl-5-hydroxymethylation on CRISPR–Cas interference remains to be determined. Nevertheless, based on a comparable occupation of the major groove by α-5-ghmC, we expect there to be little or no differences between the effects of α- or β-coupled glucosyls on CRISPR interference.

Our finding that DNA containing glucosyl-5-hydroxymethylated cytosine bases can be cleaved by a CRISPR–Cas type V-A system, but inhibits binding by crRNA–effector complexes from type I-E and II-A systems, has important implications that could be exploited for genome engineering applications (63). Firstly, Cas9 is not inhibited by 5-hmC modifications which allows editing of cells with 5-hmC modified regions such as some neuronal cells (64,65). Secondly, the inhibitory effect of 5-ghmC on Cas9 targeting could be used to introduce non-cleavable DNA templates for homology directed repair (HDR) applications. Lastly, the seemingly robust activity of Cas12a could be harnessed for genome engineering of organisms and bacteriophages that contain hypermodified bases such as the pathogen *Trypanosoma brucei* (66) and phage T4.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Netherlands Organization for Scientific Research VIDI Grant [864.11.005 to S.J.J.B.]; European Research Council Stg grant [638707 to S.J.J.B.]; M.J.D. acknowledges support from the Biotechnology and Biological Sciences Research Council UK [BB/M012166/1]. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. Funding for open access charge: Department of Bionanoscience, Kavli Institute of Nanoscience, Van der Maasweg 9, 2629 HZ, Delft University of Technology, Delft, The Netherlands.
Conflict of interest statement. None declared.

REFERENCES

1. Wigington,C.H., Sonderegger,D., Brussaard,C.P.D., Buchan,A., Finke,J.F., Fuhrman,J.A., Lennon,J.T., Middelboe,M., Suttle,C.A., Stock,C. *et al.* (2016) Re-examination of the relationship between marine virus and microbial cell abundances. *Nat. Microbiol.*, **1**, 15024.
2. Samson,J.E., Magadán,A.H., Sabri,M. and Moineau,S. (2013) Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.*, **11**, 675–687.
3. Stern,A. and Sorek,R. (2011) The phage-host arms race: shaping the evolution of microbes. *BioEssays*, **33**, 43–51.
4. Sain,B. and Murray,N.E. (1980) The hsd (host specificity) genes of E. coli K 12. *Mol. Gen. Genet.*, **180**, 35–46.
5. Loenen,W.A.M. and Raleigh,E.A. (2014) The other face of restriction: modification-dependent enzymes. *Nucleic Acids Res.*, **42**, 56–69.
6. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J.J., Charpentier,E., Haft,D.H. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
7. 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.
8. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J.J., Charpentier,E., Haft,D.H. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
9. Shmakov,S., Abudayyeh,O.O., Makarova,K.S., Wolf,Y.I., Gootenberg,J.S., Semenova,E., Minakhin,L., Joung,J., Konermann,S., Severinov,K. *et al.* (2015) Discovery and functional characterization of diverse class 2 CRISPR–Cas systems. *Mol. Cell*, **60**, 385–397.
10. 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.
11. Westra,E.R., van Erp,P.B.G., Künne,T., Wong,S.P., Staals,R.H.J., Seegers,C.L.C., Bollen,S., Jore,M.M., Semenova,E., Severinov,K.

et al. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by cascade and Cas3. *Mol. Cell*, **46**, 595–605.
12. 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.
13. Shmakov,S., Smargon,A., Scott,D., Cox,D., Pyzocha,N., Yan,W., Abudayyeh,O.O., Gootenberg,J.S., Makarova,K.S., Wolf,Y.I. *et al.* (2017) Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **15**, 169–182.
14. McGrath,S., Seegers,J.F.M.L., Fitzgerald,G.F. and Van Sinderen,D. (1999) Molecular characterization of a phage-encoded resistance system in Lactococcus lactis. *Appl. Environ. Microbiol.*, **65**, 1891–1899.
15. Semenova,E., Jore,M.M., Datsenko,K.A., Semenova,A., Westra,E.R., Wanner,B., van der Oost,J., Brouns,S.J.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.
16. Deveau,H., Barrangou,R., Garneau,J.E., Labonté,J., Fremaux,C., Boyaval,P., Romero,D.A., Horvath,P. and Moineau,S. (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.*, **190**, 1390–1400.
17. Jackson,S.A., McKenzie,R.E., Fagerlund,R.D., Kieper,S.N., Fineran,P.C. and Brouns,S.J.J. (2017) CRISPR–Cas: adapting to change. *Science*, **356**, eaal5056.
18. Andersson,A.F. and Banfield,J.F. (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*, **320**, 1047–1050.
19. Paez-Espino,D., Sharon,I., Morovic,W., Stahl,B., Thomas,B.C., Barrangou,R., Banfield,J.F. and Banfielda,J.F. (2015) CRISPR immunity drives rapid phage genome evolution in streptococcus thermophilus. *MBio*, **6**, e00262–15.
20. Krüger,D.H. and Bickle,T.A. (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.*, **47**, 345–360.
21. Walkinshaw,M.D., Taylor,P., Sturrock,S.S., Atanasiu,C., Berge,T., Henderson,R.M., Edwardson,J.M. and Dryden,D.T.F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics b-form DNA. *Mol. Cell*, **9**, 187–194.
22. Bondy-Denomy,J., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, **493**, 429–432.
23. Pawluk,A., Bondy-Denomy,J., Cheung,V.H.W., Maxwell,K.L. and Davidson,A.R. (2014) A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR–Cas system of pseudomonas aeruginosa. *MBio*, **5**, e00896.
24. Pawluk,A., Staals,R.H.J., Taylor,C., Watson,B.N.J., Saha,S., Fineran,P.C., Maxwell,K.L. and Davidson,A.R. (2016) Inactivation of CRISPR–Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat. Microbiol.*, **1**, 16085.
25. Bondy-Denomy,J., Garcia,B., Strum,S., Du,M., Rollins,M.F., Hidalgo-Reyes,Y., Wiedenheft,B., Maxwell,K.L. and Davidson,A.R. (2015) Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature*, **526**, 136–139.
26. Pawluk,A., Amrani,N., Zhang,Y., Garcia,B., Hidalgo-Reyes,Y., Lee,J., Edraki,A., Shah,M., Sontheimer,E.J., Maxwell,K.L. *et al.* (2016) Naturally occurring off-switches for CRISPR–Cas9. *Cell*, **167**, 1829–1838.
27. Rauch,B.J., Silvis,M.R., Hultquist,J.F., Waters,C.S., McGregor,M.J., Krogan,N.J. and Bondy-Denomy,J. (2017) Inhibition of CRISPR–Cas9 with bacteriophage proteins. *Cell*, **168**, 150–158.
28. Hynes,A.P., Rousseau,G.M., Lemay,M.-L., Horvath,P., Romero,D.A., Fremaux,C. and Moineau,S. (2017) An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. *Nat. Microbiol.*, **2**, 1374–1380.
29. Weigele,P. and Raleigh,E.A. (2016) Biosynthesis and function of modified bases in bacteria and their viruses. *Chem Rev.*, **116**, 12655–12687.
30. Kim,J. (2005) *Organization of the T4 dNTP Synthetase Complex at DNA Replication Sites*. Doctoral Thesis.
31. Mathews,C.K., North,T.W. and Prem veer Reddy,G. (1979) Multienzyme complexes in DNA precursor biosynthesis. *Adv. Enzyme Regul.*, **17**, 133–156.
32. Raleigh,E.A. and Wilson,G. (1986) Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 9070–9074.
33. Lehman,I.R. and Pratt,E.A. (1960) On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. *J. Biol. Chem.*, **235**, 3254–3259.
34. Leenay,R.T., Maksimchuk,K.R., Slotkowski,R.A., Agrawal,R.N., Gomaa,A.A., Briner,A.E., Barrangou,R. and Beisel,C.L. (2016) Identifying and visualizing functional PAM diversity across CRISPR–Cas systems. *Mol. Cell*, **62**, 137–147.
35. 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.
36. Künne,T., Westra,E.R. and Brouns,S.J.J. (2015) Electrophoretic mobility shift assay of DNA and CRISPR–Cas ribonucleoprotein complexes. In: Lundgren,M, Charpentier,E and Fineran,PC (eds). *CRISPR: Methods and Protocols*. Springer, NY, pp. 171–184.
37. 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.
38. Anders,C., Niewoehner,O. and Jinek,M. (2015) In vitro reconstitution and crystallization of Cas9 endonuclease bound to a guide RNA and a DNA target. In: *Methods in Enzymology*, Vol. **558**, Elsevier Inc., pp. 515–537.
39. Gao,Y.-G., Robinson,H., Wijsman,ER, Van Der Marel,G., Van Boom,J. and Wang,A.-J. (1997) Binding of daunomycin to beta-D-Glucosylated DNA found in protozoa trypanosoma brucei studied by X-ray crystallography. *J. Am. Chem. Soc.*, **119**, 1496–1497.
40. Hayes,R.P., Xiao,Y., Ding,F., van Erp,P.B.G., 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.
41. Jiang,F., Taylor,D.W., Chen,J.S., Kornfeld,J.E., Zhou,K., Thompson,A.J., Nogales,E. and Doudna,J.A. (2016) Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. *Science*, **351**, 867–871.
42. Swarts,D.C., van der Oost,J. and Jinek,M. (2017) Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR–Cas12a. *Mol. Cell*, **66**, 221–233.
43. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 12–21.
44. Adams,P.D., Afonine,P. V., Bunkóczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 213–221.
45. Warren,R.A. (1980) Modified bases in bacteriophage DNAs. *Annu. Rev. Microbiol.*, **34**, 137–158.
46. Künne,T., Swarts,D.C. and Brouns,S.J.J. (2014) Planting the seed: target recognition of short guide RNAs. *Trends Microbiol.*, **22**, 74–83.
47. Kutter,E.M. and Wiberg,J.S. (1969) Biological effects of substituting cytosine for 5-hydroxymethylcytosine in the deoxyribonucleic acid of bacteriophage T4. *J. Virol.*, **4**, 439–453.
48. 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.
49. Fu,Y., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822.
50. Zetsche,B., Heidenreich,M., Mohanraju,P., Fedorova,I., Kneppers,J., DeGennaro,E.M., Winblad,N., Choudhury,S.R., Abudayyeh,O.O., Gootenberg,J.S. *et al.* (2016) Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.*, **35**, 31–34.

51. Fonfara,I., Richter,H., Bratovič,M., Le Rhun,A. and Charpentier,E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, **532**, 517–521.
52. Pyenson,N.C., Gayvert,K., Varble,A., Elemento,O. and Marraffini,L.A. (2017) Broad targeting specificity during bacterial Type III CRISPR–Cas immunity constrains viral escape. *Cell Host Microbe*, **22**, 343–353.
53. Labrie,S.J., Samson,J.E. and Moineau,S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.*, **8**, 317–327.
54. Dupuis,M.-È., Villion,M., Magadán,A.H. and Moineau,S. (2013) CRISPR–Cas and restriction–modification systems are compatible and increase phage resistance. *Nat. Commun.*, **4**, 2087.
55. Yaung,S.J., Esvelt,K.M. and Church,G.M. (2014) CRISPR/Cas9-mediated phage resistance is not impeded by the DNA modifications of phage T4. *PLoS One*, **9**, 3–9.
56. Bryson,A.L., Hwang,Y., Sherrill-Mix,S., Wu,G.D., Lewis,J.D., Black,L., Tyson,A., Clark,F. and Bushman,F.D. (2015) Covalent modification of bacteriophage T4 DNA inhibits CRISPR-Cas9. *MBio*, **6**, 1–9.
57. Strotskaya,A., Savitskaya,E., Metlitskaya,A., Morozova,N., Datsenko,K.A., Semenova,E. and Severinov,K. (2017) The action of *Escherichia coli*CRISPR–Cas system on lytic bacteriophages with different lifestyles and development strategies. *Nucleic Acids Res.*, **45**, 1946–1957.
58. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
59. 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.
60. Rollins,M.F., Schuman,J.T., Paulus,K., Bukhari,H.S.T. and Wiedenheft,B. (2015) Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. *Nucleic Acids Res.*, **43**, 2216–2222.
61. Hunter,C.A. (1996) Sequence-dependent DNA structure. *Bioessays*, **18**, 157–162.
62. el Hassan,M.A. and Calladine,C.R. (1996) Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.*, **259**, 95–103.
63. Vlot,M. and Brouns,S.J.J. (2017) Dna modification. Patent WO2017055514 A1.
64. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
65. Münzel,M., Globisch,D. and Carell,T. (2011) 5-Hydroxymethylcytosine, the sixth base of the genome. *Angew. Chem. Int. Ed.*, **50**, 6460–6468.
66. Gommers-Ampt,J.H. and Borst,P. (1995) Hypermodified bases in DNA. *FASEB J.*, **9**, 1034–1042.