9542–9556 *Nucleic Acids Research, 2019, Vol. 47, No. 18* *Published online 2 August 2019*
doi: 10.1093/nar/gkz665

Transcriptome analyses of cells carrying the Type II Csp231I restriction–modification system reveal cross-talk between two unrelated transcription factors: C protein and the Rac prophage repressor

Alessandro Negri[1], Marcin Jąkalski[2], Aleksandra Szczuka[1], Leszek P. Pryszcz[3] and Iwona Mruk[1,*]

[1]Department of Microbiology, Faculty of Biology, University of Gdansk, Wita Stwosza 59, Gdansk 80-308, Poland, [2]Department of Plant Taxonomy and Nature Conservation, University of Gdansk, Wita Stwosza 59, 80-308 Gdansk, Poland and [3]Laboratory of Zebrafish Developmental Genomics, International Institute of Molecular and Cell Biology, Warsaw, ul. Trojdena 4, 02-109 Warsaw, Poland

Received February 25, 2019; Revised July 16, 2019; Editorial Decision July 17, 2019; Accepted July 23, 2019

ABSTRACT

Restriction-modification (R–M) systems represent an effective mechanism of defence against invading bacteriophages, and are widely spread among bacteria and archaea. In acquiring a Type II R–M system via horizontal gene transfer, the new hosts become more resistant to phage infection, through the action of a restriction endonuclease (REase), which recognizes and cleaves specific target DNAs. To protect the host cell's DNA, there is also a methyltransferase (MTase), which prevents DNA cleavage by the cognate REase. In some R–M systems, the host also accepts a *cis*-acting transcription factor (C protein), which regulates the counteracting activities of REase and MTase to avoid host self-restriction. Our study characterized the unexpected phenotype of *Escherichia coli* cells, which manifested as extensive cell filamentation triggered by acquiring the Csp231I R–M system from *Citrobacter* sp. Surprisingly, we found that the cell morphology defect was solely dependent on the C regulator. Our transcriptome analysis supported by *in vivo* and *in vitro* assays showed that C protein directly silenced the expression of the RacR repressor to affect the Rac prophage-related genes. The *rac* locus *ydaST* genes, when derepressed, exerted a toxicity indicated by cell filamentation through an unknown mechanism. These results provide an apparent example of transcription factor cross-talk, which can have significant consequences for the host, and may represent a constraint on lateral gene transfer.

INTRODUCTION

Bacteria are highly diverse organisms, that can adapt to a wide range of habitats mostly due to the plasticity of their genomes, which is driven mainly by horizontal gene transfer (HGT), as well as by other mechanisms, such as point mutations, and DNA rearrangements. However, HGT is the most important mechanism, which strongly affects the evolution and speciation of prokaryotes (1,2). Among many factors that modulate this process, restriction–modification (R–M) systems play a crucial role. R–M systems limit the flow of genetic material into the host cell (3–5), and produce recombinogenic ends on the acquired DNA, to facilitate their integration into the genome (6,7). However, the most prominent role of R–M systems involves cellular defence against invasive DNAs, such as bacteriophages (8). It is possible that this beneficiary feature for hosts resulted in the R–M systems being prevalent and diverse in bacteria and archaea. R–M systems are found in nearly all bacterial genomes, and are especially numerous in naturally competent cells, which suggests that R–M systems not only control, but also circulate using HGT routes (6,9–12). Among the four types, the Type II is the most frequent and also the simplest in structure. It is composed of two independent enzymes, which involve a restriction endonuclease (REase) and a DNA methyltransferase (MTase). Both enzymes recognize the same short specific DNA sequences, where MTase adds a methyl group to modify such sites, to protect them from further cleavage by the cognate REase (13). Such counteracting activities often are compared to the action of toxin–antitoxin systems (14).

Mobile Type II R–M systems, when successfully introduced into new hosts, lead to global changes in the host cell physiology associated with the actions of their two enzymatic entities: MTase and REase. First, the cell genome

*To whom correspondence should be addressed. Email: shamrock127@hotmail.com

acquires the new epigenetic status related to the specificity of the introduced MTase. As a result, all genomic target sites are methylated, forming a new, unique set of epigenetic markers, which generates a cell-specific methylome dependent on the repertoire of active MTases (15–17). The methyl group may change expression of a single gene if it is located within the promoter/operator region, by blocking either RNA polymerase recruitment or binding by transcription factors. An increasing number of studies have reported that methylation may cause global transcriptome changes, yielding distinct cell phenotypes related to stress response, fitness, motility, or production of virulence factors (18–25).

Second, the new REase might serve as an efficient antiphage defence as long as its activity is precisely controlled to minimize genome damage (14). Nevertheless, global response to DNA damage (SOS response) is often triggered when the R–M system is not balanced (26) or not transmitted properly to progeny cells, resulting in post-segregational cell killing (27). In the latter case, the remaining REase may cleave the genome no longer fully protected by MTase, and the cell may die unless DNA repair occurs (28,29). In this context, the bacterial hosts remain in an intimate and dependent relationship with their acquired R–M systems.

A large number of Type II R–M systems also possess a specific transcription factor, C protein, dedicated to the control of their own gene expression (30). C proteins are relatively small proteins (8–11 kDa), which bind to a specific DNA operator sequence called the C-box (31,32). Their helical structure, comprising helix-turn-helix (HTH) DNA-binding motifs, resembles that of the Xre family of transcription regulators, including the λ and 434 phage repressors. This suggests a common mechanism of DNA recognition and their effect on transcription by direct contact with σ^{70} RNA polymerase (33,34). The controlling effect of a C protein on R–M system expression was first found in the PvuII system and then in several others (31,35–41). C protein action is dependent on location and structure of its C-box-DNA recognition site. It is usually located within the promoter of its own gene and of controlled genes (REase and/or MTase) (36,42). The C-box comprises two palindromic binding sites (proximal and distal) for two C protein homodimers. The C protein cooperatively binds the C-box sequence in a highly concentration-dependent manner (43,44), and has greater affinity for the proximal C-box sequence, which facilitates its interaction with RNA polymerase and promotes the activation step (ON) during transcription. Subsequent C protein accumulation leads to binding both proximal and distal sequences, resulting in transcriptional repression representing the mode of action of a typical genetic switch (35,40,44). The C-box DNA sequence varies for different families of C proteins; some consensus sequences have been determined (32), and C protein:C-box co-crystal structures have been determined (45–47). The box structures of C proteins are specific and complex (due to nearly palindromic repeated sequences), enough to be unique in the genome. Although cross-complementation can occur (41), if two R–M operons with the C proteins of the same family exist in the same cell, exclusion of one R–M system can occur (48).

The main function of C proteins is providing temporal regulation, which is essential during R–M system transfer to the new host. Specifically, C protein mediates a delay in REase expression to enable the MTase to complete the modification process of the host genome (48–51).

Although previous studies have shown the effect of MTase or REase on global gene expression of the bacterial hosts, there has been no report characterizing the effect of the introduced transcription factor linked to the R–M system. In the present study, using a combination of genetics and transcriptomics, we characterized the unusual phenomenon of *Escherichia coli* cell filamentation that is manifested when acquiring genes of the Csp231I R–M system, and in particular its C protein. We show that this transfer phenomenon results from the adventitious cross-talk between two transcription factors. One of these factors, the C protein of the incoming R–M system reduced expression of RacR, the essential regulator of the cryptic Rac prophage (52–54). As a result, two Rac genes, *ydaS* and *ydaT* were derepressed and had a toxic effect manifested by cell filamentation, as well as loss of cell viability and fitness. In general, such transcription factor interconnectivity may have fatal consequences for the host and may become a serious constraint in lateral gene transfer. This process may also exemplify genetic transfer as a possible deleterious event, which is subsequently lost from the cell population.

MATERIALS AND METHODS

Bacterial strains and plasmids

The source of Csp231I and EcoO109I R–M systems were *Citrobacter* sp. RFL231 (kindly supplied by MBI Fermantas, Lithuania) and *E. coli* H709c (55,56). Despite *E. coli* and *Citrobacter* both being members of the *Enterobacteriaceae*, expression of the Csp231I R–M system in *E. coli* needs additional protection by a second MTase of the same specificity. Accordingly, any manipulation with the wild-type (WT) Csp231I R–M system requires competent cells prepared from MM294 strain of *E. coli*, which expresses the MTase gene from pEcoVIIIM to ensure the protection of the host genome. M.Csp231I and M.EcoVIII both recognize the same specific nucleotide sequence (56). The MG1655 strain was used for chromosomal gene knockouts. The other strains and plasmids used are listed in Supplementary Table S1 (Supplementary File 1).

High-throughput sequencing

Cultures of *E. coli* MG1655 carrying pMCsp231 with an additional protective MTase gene on the pACYC177 derivative, and with a plasmid with RMs of interest, WT (p18) and two variants of C-R+ (p30) and C+R– (p24), were grown in LB medium. At mid log phase, the cells were harvested and resuspended in RNALater solution (Sigma-Aldrich, USA). Total RNA was extracted with RiboPure (Ambion, USA) in two biological replicates. The amount and quality of the resulting RNA extracts were determined using a BioAnalyzer instrument (Agilent, USA) with an RIN >8.4. The RNA was then treated with DNase I (Ambion) to remove residual DNA, inactivated with a chemical reagent from the Ambion kit, centrifuged, and precipitated according to recommendations from Macrogen (Seoul, Republic of Korea). RNA samples in two biological replicates per vari-

ant were sent in ethanol to Macrogen. Prior to sequencing, RNA was rRNA-depleted using a Ribo-Zero rRNA Removal Kit for bacteria (Thermo Fisher, USA). A RNA-seq library was prepared according to the TruSeq RNA Sample Preparation, version 2 Guide (Illumina, San Diego, CA, USA) and sequencing of the libraries was performed using an Illumina HiSeq2500 platform at Macrogen (paired-end and non-strand specific 101 bp long reads with an average median insert size across all samples of 155 bp). The total read bases averaged 10.6 Gbp per sample. The results have been deposited in the NCBI GEO (accession number GSE126248). Additional details are included in the Supplementary File 1.

Search for homologs of Rac prophage region and C protein

Escherichia coli genomes available at GenBank (495 complete genomes with protein annotations; access date 20 July 2018) were obtained from the NCBI ftp site. Amino acid sequences of 21 Rac region genes and the C protein (C.Csp231I and C.EcoO109I) were used as queries for searches against the *E. coli* proteomes using the phmmer search tool (version 3.2.1) from the HMMER package (57) with the e-value parameter (–E) set to $1e^{-20}$, analogous to the analyses by Krishnamurthi *et al.* (54).

Chromosomal gene knock-outs

The knockout strains were constructed using the lambda-red recombination method with a pSIM5 plasmid carrying the recombineering proteins, Gam, Exo, and Beta (58,59), and using pKD46 as a template plasmid for ampicillin resistance cassette amplification. The constructed strains and primers used are listed in Tables S1 and S2 (Supplementary File 1).

Fluorescence assay and microscopy

To measure the SOS response, the MP060 and MP064 cells were grown with shaking to the exponential phase in LB or M9-glucose medium, then gently pelleted, washed once with phosphate-buffered saline (PBS) and resuspended in 500 μl of PBS. Half of the sample was read to monitor the optical density (600 nm) of bacteria and the other half used to read the yellow fluorescence (YFP) intensity (excitation at 515 nm with emission at 545 nm) in a 96-well plate reader (EnSpire Multimode; Perkin Elmer, USA). Relative fluorescence was corrected by subtracting the level of fluorescence of non-YFP bacterial cells and dividing by the optical density. For microscopy analyses, the cells were stained with SYTO9 dye (Invitrogen, USA) alone or with a combination of SynaptoRed (Sigma-Aldrich) and 4′,6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich), then visualized using a fluorescence microscope (DMB400B; Leica, Germany).

Relative restriction activity assay

The restriction activity of *E. coli* cells carrying the Csp231I R–M system and its variants was measured using the efficiency of plaque formation (EOP) of phage λ*vir*. There are six recognition sites in the λ*vir* genome. The EOP of λ*vir* was calculated as the ratio of plaques formed on *E. coli* MG1655 containing plasmids with no R–M system to those formed on the same strain containing a plasmid with the Csp231I R–M system or their variants.

Cell viability and LacZ reporter assay

Cell survival was measured using the spotting assay. The overnight cultures were subcultured in LB /M9 medium, and when they reached an OD of 0.3, they were split into two cultures, with and without 0.1% L-arabinose; both cultures continued to grow. They were then serially diluted and spotted on LB agar plates with and without arabinose. CFU values were calculated and cell survival was determined as the ratio of induced cells divided by uninduced cells. For the LacZ assay, MG1655 Δrac cells were grown in M9 medium with glycerol as the carbon source. At the induction time, glucose or arabinose were added at the indicated concentrations. The o-nitrophenyl-β-D-galactopyranoside (ONPG) test was performed as previously described (60).

Electrophoretic mobility shift assays (EMSA)

DNA substrates were double-stranded PCR-amplified fragments that were fluorescently labelled from one end by the Cy5 tag introduced by the primer (Table S2 of Supplementary File 1). WT substrate (176 bp) covered the entire *racR* promoter/operator region and part of its coding sequence (*racR*), but its mutated variant (*racRmut*) had CTTAG sequence within coding sequence replaced by CGCAT to eliminate possible binding. Reactions containing 25 nM of DNA and the indicated purified C protein concentrations (60) were prepared in binding buffer [10 mM Tris–HCl (pH 8.0), 50 mM NaCl, 10 mM $MgCl_2$, 1 μg of poly(dIdC)] to a final volume of 20 μl, and incubated for 20 min at 22°C. Samples were electrophoresed on 5% native polyacrylamide gels in 0.5× TBE buffer at 22°C. Detection of the Cy5-labeled DNA was performed using the Typhoon 9200 variable mode imager (Molecular Dynamics, USA).

Testing the *rac*+/C+ conflict by biological assay

A single colony of *E. coli* MG1655 (*rac*+) carrying the p24 plasmid, with the active *csp231IC* gene, was picked from a fresh transformation on LB-agar (with appropriate selective antibiotic) and inoculated into 5 ml of LB medium in triplicate. The colony was confirmed to present the expected extensive filamentation. Every 15–18 h of incubation at 37°C with shaking, the culture was diluted 10^2 into fresh LB medium with antibiotic. Samples of culture were spread quantitatively to count the CFUs based on their size (small versus large) to calculate their ratio, as well as the generation number. At each time, we inspected cells by microscopy to assess cell filamentation.

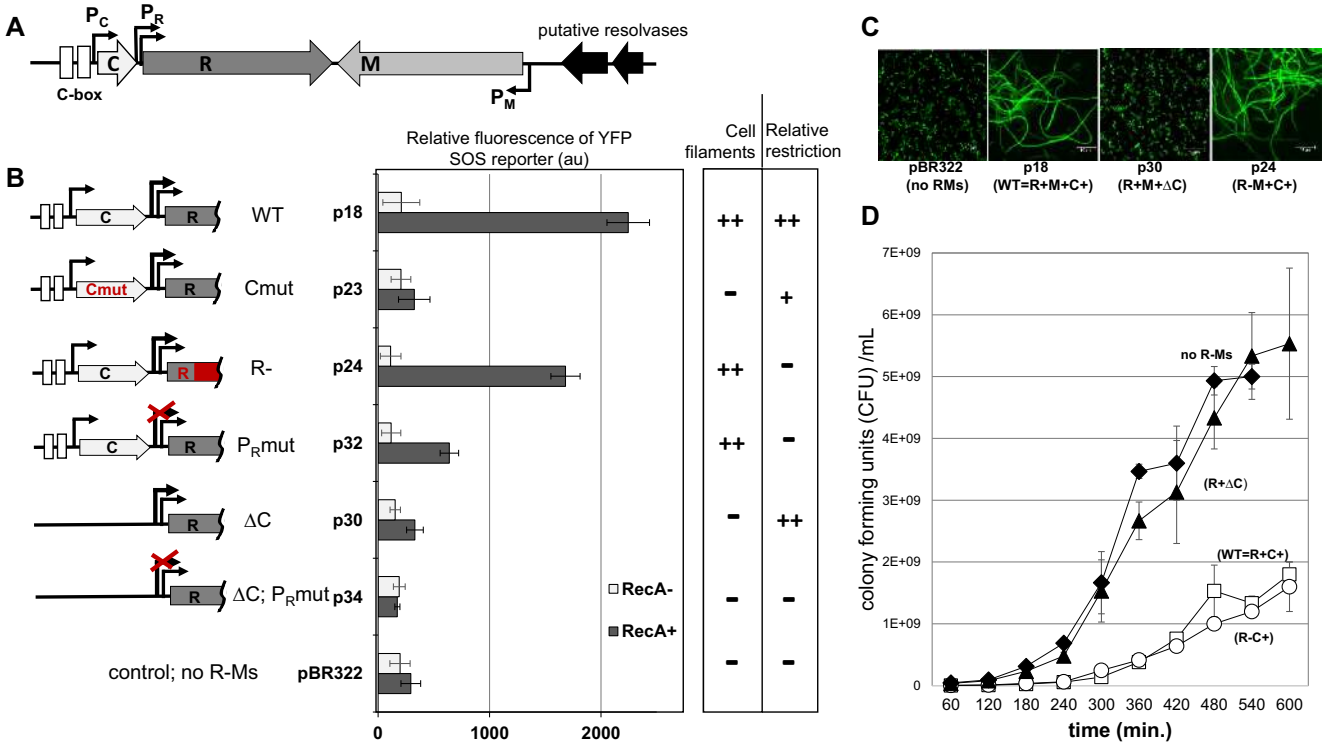


Figure 1. Presence of the active C protein within the Csp231I R–M system changes the host cell morphology from rod-shaped into extensively elongated regardless of restriction endonuclease activity. (**A**) Schematic diagram of the wild-type (WT) Csp231I R–M system (not to scale). (**B**) Induction of the SOS response for cells carrying the indicated plasmids were measured by relative fluorescence in context of: *recA* positive (dark bars) and *recA* negative (white bars), where the yellow fluorescent protein was fused to the *sulA* promoter (P_{sulA}-yfp). Tested R–M system variants are presented on the left and data in panel B are adjusted to be read horizontally. The unchanged MTase gene is not shown, while varying elements are depicted in red. Briefly: p18 (WT); p23 (Cmut; substitutions: A33G; R34E; Q37A in C protein); p24 (knock-out of REase gene, XhoI cut and Klenow filling); p32 (REase main promoter P_R mutated; -10 hexamer TTAAAT→CCCGGG); p30 (deletion of C gene and its upstream region including C-box and P_C); p32 (mutation of -10 box of P_{R1}, TTAAAT→CCCGGG); p34 (C-deleted variant of p30 mutated as in p32). The standard deviation from four experiments is shown. Cell morphology for tested strains were determined by microscopy: normal rod-shaped cells (–) or filamentous (++). Relative restriction was measured using λ*vir* phage, and determined from highly restrictive (++), low restrictive (+), to restriction-negative (–); as also previously shown by EOP values (60). (**C**) Cell morphology typical for filamentous and non-filamentous phenotypes. (**D**) Growth curves for cells cultured in LB media carrying plasmids presented in panel C; pBR322 (black diamonds); p30 (black triangles); p18 (white squares); p24 (white circles). Error bars represent the standard deviation from three replicates of each culture.

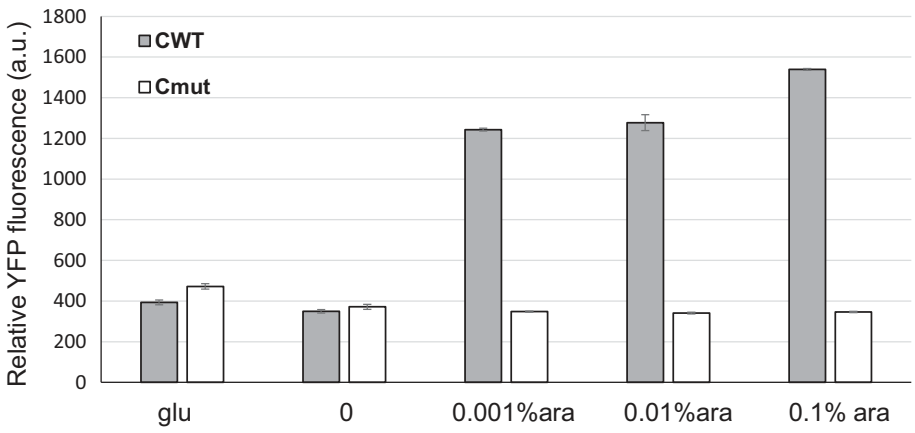
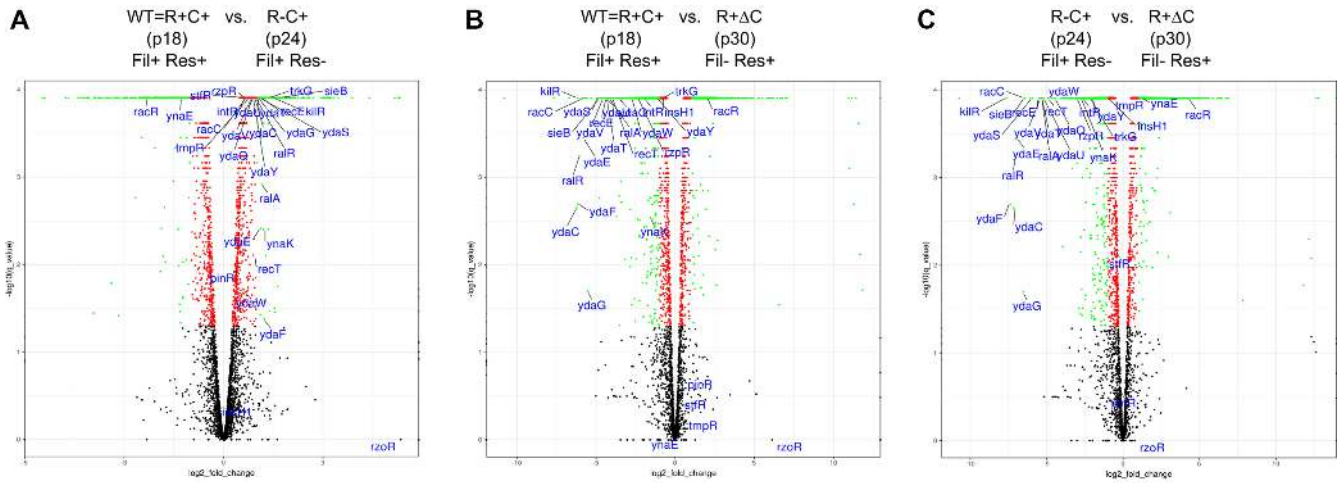


Figure 2. C protein alone triggers cell filamentation. Cells carrying the yellow fluorescent protein fused to the promoter of *sulA* (P_{sulA}-*yfp*) (*E. coli* MP064) was used to measure the global response to DNA damage. Wild-type C gene under inducible P_{araBAD} promoter or its C gene mutant (Cmut) unable to bind DNA was delivered by plasmids. The cells were grown in minimal media with glycerol as the carbon source and with glucose (glu) or arabinose (ara) at gradually increasing concentrations (from 0 to 0.1%). The error bars indicate the averages (±SD) of four independent experiments.



D

	Number of detected significant DEGs	% up-regulated	% down-regulated
p18 vs. p24	577	38.65	61.35
p18 vs. p30	1591	50.97	49.03
p24 vs. p30	1847	50.84	49.16

Figure 3. Transcription-wide analyses. Global overview of differential gene expression between the wild-type R–M system (p18, R+M+C+) and its variants lacking restriction endonuclease (REase) activity (p24, R–M+C+) and C gene activity (p30, R+M+ΔC). Fil+ indicates the filamentation present in cells; Res+, the presence of restriction activity (active REase). (**A–C**) Volcano plots represent scatter plots of significance versus fold-change of expression of individual transcripts (in two biological replicates). Points representing transcripts with significant (a value of $P < 0.05$) expression change are in red or green (if expression fold change was ≥2). Finally, the points representing transcripts of the Rac prophage region are in blue. (**D**) Summary statistics for detected significant differentially expressed genes, with percentages of up- and downregulated transcripts for tested pairs of strains.

RESULTS

Peculiarity of the Csp231I R–M system: *E. coli* host cells filament after accepting the R–M system regardless of REase activity

The Csp231I R–M system was cloned from chromosomal DNA of *Citrobacter* sp. RFL231 and has the same specificity as its prototype HindIII R–M system (56,60). The REase and MTase genes are convergently oriented and the REase is preceded by a regulatory C gene (Figure 1A). Two putative resolvase genes were located downstream of the MTase gene (GenBank AY787793). It is likely this genetic unit was acquired by HGT, as the genes have approximately 34% GC content compared to ∼50% of its host *Citrobacter* sp. In addition, codon usage is biased for rare codons in *Citrobacter*, such as arginine (AGG and AGA) and isoleucine (AUA), which also suggests that the R–M genes might be of foreign origin. The AGG/A arginine codon usage in the R–M system unit was 61%, while it was only 15% in *Citrobacter* overall. In a similar manner, the AUA of isoleucine codon fraction was 34% in the Csp231I R–M system versus 14% for its host.

We observed that the WT Csp231I R–M system, when transferred to *E. coli* cells, triggered extensive cell filamentation (Figure 1B and C). Initially, we linked this phenomenon to the global response to DNA damage (SOS response) due to the REase activity. Such an effect is typical for some R–M systems, that cause a low level of self-restriction, which leads to cell division arrest and manifestation of cell elongation (26,28,29,61,62). We thus determined if eliminating REase activity caused corresponding changes in the level of filamentation. For quantitative analysis, we used MP60/MP64 *E. coli* host cells (28), where the promoter of the SOS responsive *sulA* gene was fused to the YFP reporter (P_{sulA}-*yfp*), allowing fluorescence measurement of the SOS response. Plasmids with the WT R–M system (p18) or its variants were introduced into these cells, and the relative SOS response and cell morphology were determined (Figure 1B and C). First, absence of the C gene (p30), unlike WT, resulted in loss of both cell filamentation and SOS induction, regardless of having the highest relative restriction, as previously measured (60) (Figure 1C). Among other R–M system variants, only a REase-negative mutant (p24) showed an SOS response comparable to that of the WT, as well as exhibiting similar filamentation (Figure 1B and C). A slightly smaller effect was shown by the p32 variant (filaments-present), which had an inactive promoter for REase, but an intact C gene. Most notably, the R–M system variant that displayed reduced restriction due to mutation of the C gene (p23), which disabled C protein binding to its C-box, gave a cell phenotype identical to cells with no R–M system, and also showed a loss of SOS response when compared with the WT. Overall, these data indicated that cell filamentation correlated with the function of the active C gene, and not with REase activity. We compared the *E. coli* MG1655 growth curve for cells carrying the WT R–M system (p18) and its two variants: filamentous, but restriction-negative (p24), and nonfilamentous, C-absent, but with a highly restrictive phenotype (p30). There was a slower growth for C-present cells and a delay in reach-

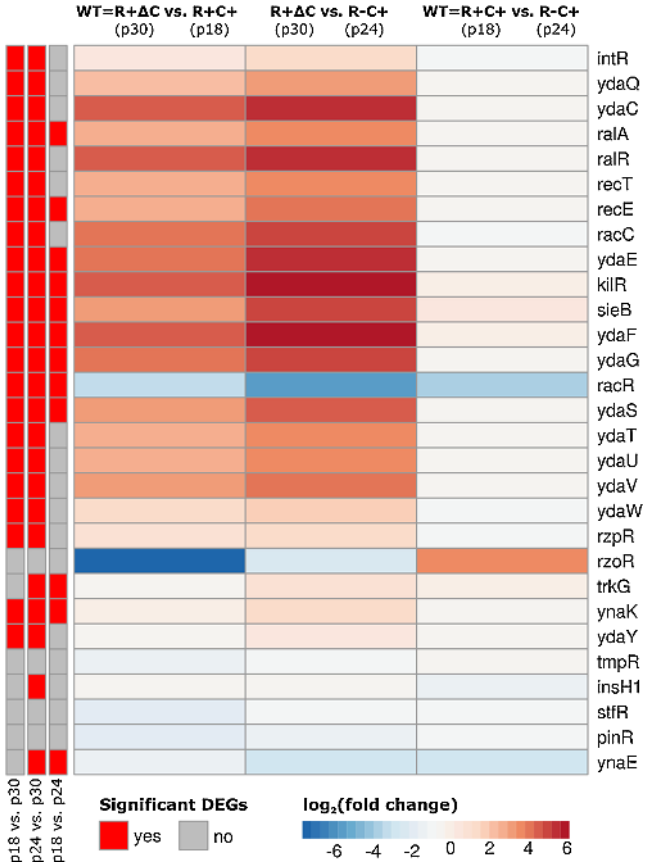


Figure 4. Heat map representation of differentially regulated genes of Rac prophage in the K12 strain of *E. coli*. The $\log_2$-transformed fold change of FPKM values (fragments per kilobase per million mapped reads) are shown. Upregulated genes are shown in red, downregulated genes are shown in blue. Differentially expressed genes (FC $\geq$ 2, q-value $<$ 0.05) in each comparison are shown in red. The native order of Rac prophage genes is presented (no clustering has been applied).

ing the early exponential phase (Figure 1D). In addition, we observed C-protein associate cell filamentation in RecA-negative host cells (Figure S1 of Supplementary File 1), indicating that the C protein effect is not RecA dependent.

C protein alone triggers cell filamentation

We sought to test whether the C gene alone resulted in a specific cell phenotype. We used a previously tested plasmid (pBAD-CWT), where C expression is controlled by the P_{BAD} arabinose-inducible promoter, which is repressed by glucose and induced over a wide range of arabinose concentration (60). As a control, we also used a C protein variant (C-ARQ = Cmut), which is unable to bind its target site, to serve as an inactive regulator (pBAD-Cmut). Plasmids were introduced into cells bearing the YFP reporter of SOS response. The C protein effect was observed only with the WT, but not the mutated C gene, which remained at the same level regardless of amount of the arabinose inducer (Figure 2). Taken together, these results suggest that C protein acts by binding to an unknown DNA target in a concentration-dependent manner. Cell filamentation was also detected only for the WT C gene, which correlated with the measured SOS response (not shown). However, we were puzzled by the observation that filamentation did not occur during overexpression of the C gene using the T7 promoter system, which we used for C protein purification in the BL21(DE3) derivative host in our other report (not shown; and considered more below (60)).

Differential expression analysis of transcriptomes with the C gene reveal strong upregulation of Rac prophage-related genes

To help establish the molecular basis for C-dependent filamentation of host cells, we performed total RNA sequencing (RNA-seq) for cells carrying plasmids with the WT R–M system (p18, R+M+C+), or one of two other variants, namely p30 (R+M+ΔC) and p24 (R–M+C+), each growing in the exponential phase in a rich medium. The Illumina HiSeq reads produced for the three samples (each with two biological replicates) ranged from 7.8 million to 11.6 million, with over 98% of reads properly mapping to the reference genome in each sample (Tables S3 and S4, Figure S2 of Supplementary File 1).

In total, 2238 differentially expressed genes (DEGs) were identified among the three conditions (Supplementary File 2, Table S5 and Figure S3 of Supplementary File 1). In general, the volcano plots (Figure 3) representing the comparison of the filamentous phenotype with the non-filamentous phenotype showed similar dispersion of expression values for WT (p18) versus R+ΔC (p30), and R-C+ (p24) versus R+ΔC (p30); (Figure 3B versus C). Accordingly, a similar number of DEGs were detected (1580 versus 1826), both having similar numbers of down- and upregulated genes (Figure 3D; Table S5 of Supplementary File 1). Thus, in this fraction of genes, we could expect to find gene candidates determining the cell filamentation phenotype. In contrast, the pair that differed in restriction activity, but had the same filamentous phenotype, WT R+C+ (p18) versus R-C+ (p24), showed different patterns of changes in expression (Figure 3A). Fewer DEGs were detected (568), with 39% being upregulated and 61% downregulated. RNA-seq results were validated by qPCR for selected transcripts (Supplementary File 1, Table S6). More detailed transcriptome-wide analyses are described in the Supplementary Materials including gene ontology analysis (Supplementary File 4 and 5, Figures S4–S6 of Supplementary File 1).

Importantly, 15 transcripts of the 40 most highly upregulated in the two C+/ΔC comparisons (p30 versus p18 and p30 versus p24) (Figure 3B and C) were related to the defective Rac bacteriophage locus of *E. coli.* The entire Rac region contains 20 genes and nine pseudogenes encompassing approximately 23 kb of DNA (63). We next examined the expressions of all 29 Rac genes. The majority of these genes showed elevated expression when filamentous versus non-filamentous samples were compared (Figure 4). For example, *ydaC*, *kilR* and *ralA* increased their expressions by more than 70-fold in p30 versus p18 or p30 versus p24, while *ydaF, ydaG, ydaE* and *racC* increased their expressions approximately 40-fold. Rac genes with unchanged expressions included Rac structural genes, located at the 3' region of the prophage. Notably, only one gene, *racR*, showed significantly decreased expression in the presence of the C protein.

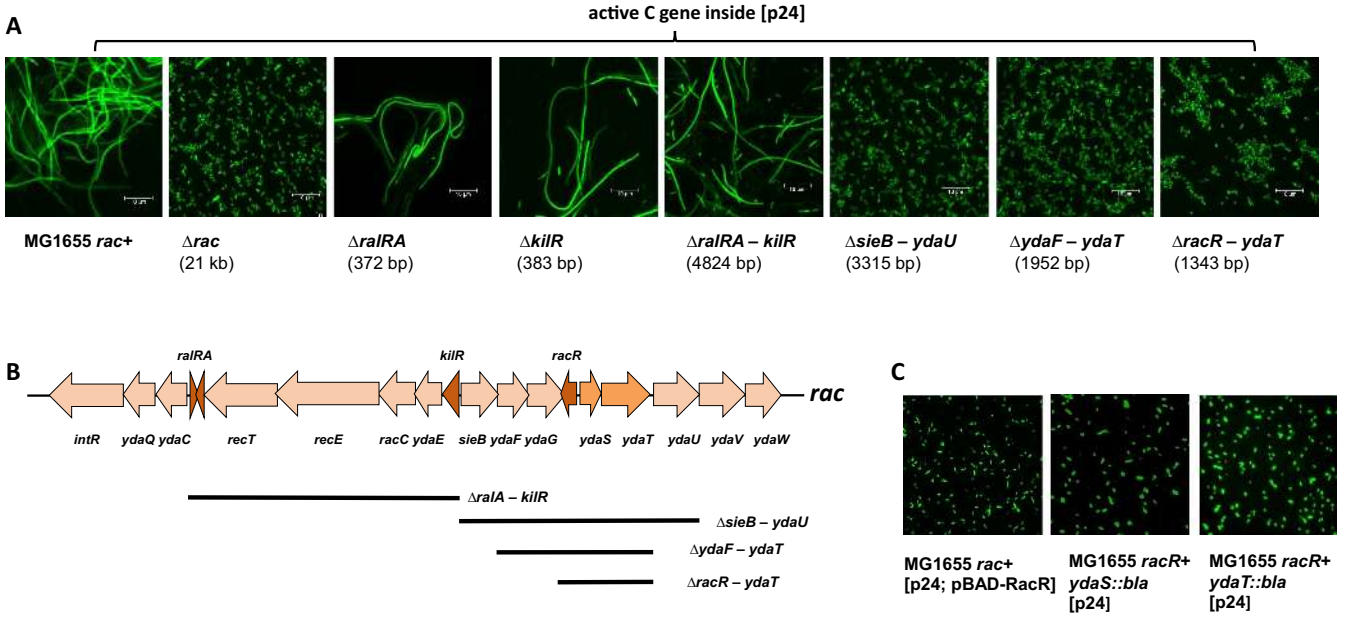


Figure 5. Deletion screen of Rac prophage genes reveals that the presence of cell filamentation is associated with RacR and YdaST genes. (**A**) The morphology of cells carrying the plasmid with the wild-type C gene (p24) was tested in the genetic background of different deletion mutants of *E. coli* MG1655 as indicated under the pictures. The deletion length is noted in brackets. (**B**) The distribution of genes of the *rac* locus around the *racR* gene is presented (not in scale) with their deletions indicated below by black bars. (**C**) When the MG1655 strain (rac+) with p24 plasmid was supplemented with a second plasmid overexpressing the *racR* gene, the cell filamentous phenotype was suppressed. In addition, inactivating *ydaS* or *ydaT* by *bla* gene insertion (with the *racR* gene maintained intact) resulted in loss of the filamentation phenotype when the WT C gene was introduced.

The *racR* gene is towards the center of the Rac locus (Figures 4 and 5B). It is predicted to function as a prophage transcription factor, because it contains a DNA binding motif (54). Overall, these results suggest that RacR functions as the master regulator of the Rac prophage locus, and that C protein indirectly induces Rac transcription by somehow reducing RacR levels.

Deletion scanning of Rac prophage genes shows that C protein targets the *racR-ydaST* region to trigger cell filamentation

To test whether the Rac region contains a C protein binding site, and whether the resultant interaction induces cell filamentation, we used an *E. coli* MG1655 mutant with a deletion of the entire *rac* locus (approximately 23 kb; Δrac (64)), and introduced the C gene on a plasmid (p24). The results showed no cell elongation (Figure 5A). To localize a target site for C protein binding within the Rac region, we first deleted DNA segments carrying genes related to cell division and viability, which might be upregulated in the C+ context. We deleted a *kilR*, which inhibits the major cell division protein, FtsZ (63,65), as well as genes coding for the toxin–antitoxin system, RalR/RalA, producing nonspecific endonuclease activity (64). None of these deletions changed the cell morphology, and the cells remained elongated in the presence of C protein (Figure 5). Next, we used three strains with deletions encompassing the *racR* gene. The *racR* gene alone could not be deleted, and others have observed (52–54). The minimal deletion region, that showed loss of cell filamentation for cells carrying the C gene, removed the *racR* and *ydaST* genes along with the intergenic region of the putative operator for the RacR repressor (Figure 5A and B). In addition, we constructed plasmids overexpressing the *racR* repressor (pBAD-RacR) to test its effect *in trans*. When Rac+ cells carrying the C gene on the p24 plasmid were additionally supplemented with the *racR* repressor, their cell filamentation was suppressed and they presented the same cell phenotype as Δrac cells (Figure 5C). These results indicated that C-dependent cell filamentation was linked to low expression of the *racR* repressor gene. We also separately inactivated *ydaS* and *ydaT* by *bla* gene insertion, keeping the *racR* gene intact. In both insertion mutants, introducing the active C gene no longer triggered filamentation, making it very likely that the toxic effect comes from YdaST activity (Figure 5C).

C protein affects cell viability only in the presence of Rac prophage genes

To further understand the role of C protein in *E. coli rac*+ cells, we determined whether the C protein affected cell viability as well as cell elongation. Previous studies have reported that the RacR repressor controls the expression of adjacent *ydaST* genes, which probably cause cell toxicity by an unknown mechanism (54). We induced C protein in WT *E. coli* cells (*rac*+) and in the deletion variant (*rac*–) using plasmids with the C gene (pBAD-WTC or pBAD-Cmut) with 0.01% arabinose. The same strains without arabinose induction served as negative controls. Cell viability was estimated as the ratio of colony forming units (CFUs) at induction with the CFUs under uninduced conditions (Figure 6A and B). The cells expressing WT C protein in the *rac*+ context showed significant reduction of viability, of about

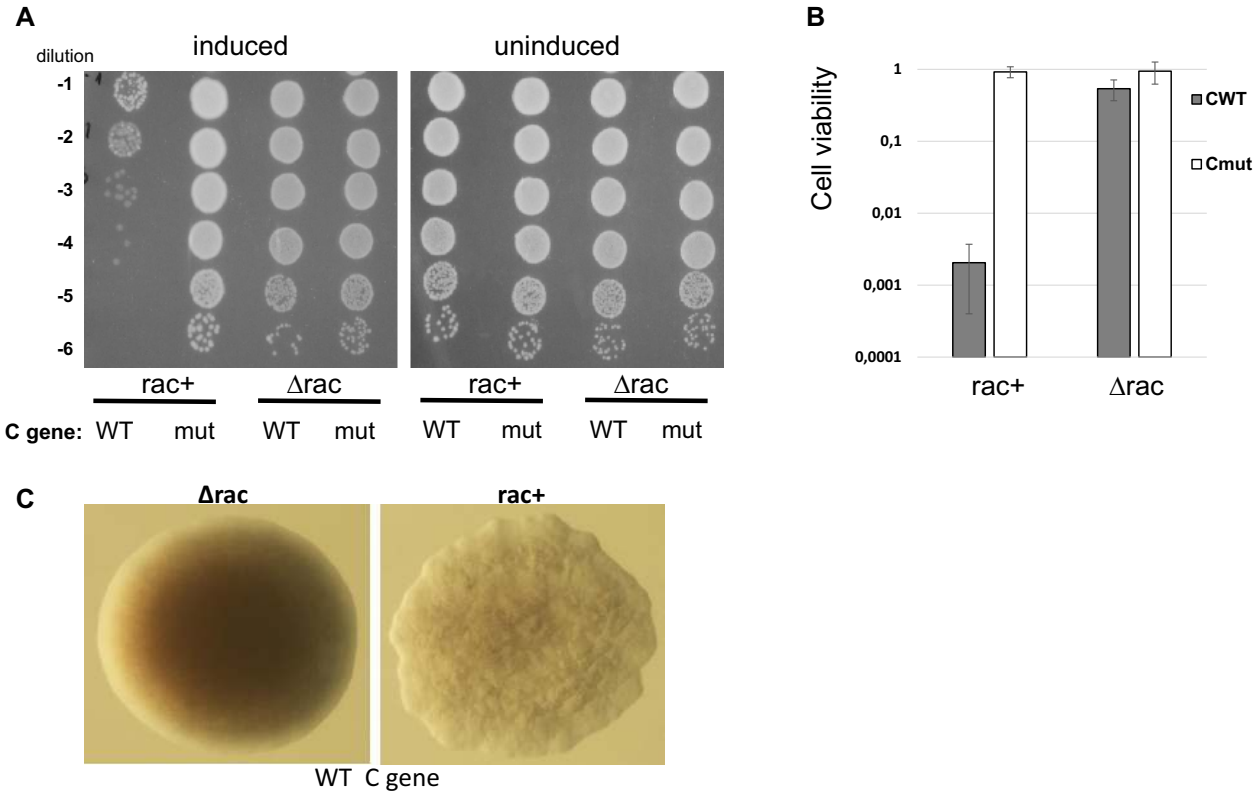


Figure 6. The C protein effect on *E. coli rac*+ cell viability. (**A**) The *rac*+ and *Δrac* cells carrying the wild-type (WT) C gene with the inducible P_{araBAD} promoter or its C gene mutant (Cmut) unable to bind DNA, were grown until they reached log phase. Then, they were split and one half was induced in 0.01% of arabinose, and the second half remained uninduced. After 5 h, they were serially diluted and spotted on LB agar plates with and without arabinose. CFU values were calculated and cell survival was determined for the *rac*+ and *Δrac* cells as the ratio of induced divided by uninduced cells, which are presented as panel (**B**). Grey bars show the effect of WT C protein, whereas the white bars show the Cmut variant. The results are the averages (±SD) of three independent experiments. (**C**) The representative single colony of *rac*+ cells with active C gene was magnified and the photograph revealed its distinct morphology as compared to the *Δrac* background. Such morphology with a flat center and irregular margins was typical for highly filamentous cells.

three orders of magnitude in comparison to cells producing the inactive C protein variant. However, surprisingly we noticed altered morphology of such *E. coli* colonies. Closer inspection revealed translucent, flat colonies with irregular surfaces, which contrasted with the opaque, regular, cone-shaped colonies formed by the Rac-absent strain (Figure 6C).

Microscopy of single cell morphology from such colonies again confirmed extensive cell filamentation. In addition, we used DAPI staining to visualize the localization of nucleic acids, and found that it was nearly equally distributed at each non-separated segment of elongated cells. The segments were similar in length with visible septa. Some longer segments had DNA still being replicated, but overall, the cell division defect seemed to be related to the cell separation stage (Figure 7C). Failure to separate, as opposed to altered viability, might explain at least part of the reduction in CFUs seen in Figures 6B, 7A and B.

We also tested the induction of C protein in the *rac*+ context with time. The CFU values decreased for ∼2 h after arabinose induction, but prolonged induction led to the recovery of cell viability and even slight growth (Figure 7A). We also found that this recovery was halted when cells were deficient in RecA function (Figure 7B). Although we determined the cell elongation is RecA-independent, the SOS response could also be induced as shown by the transcriptomic data with genes of SOS regulon upregulated in context of cells with C protein (Table S7 of Supplementary File 1).

C protein reduces expression of RacR repressor by binding its gene sequence

Next, we searched for a potential C protein binding sequence within the DNA adjacent to *racR*. A C.Csp231I binding consensus sequence was determined by extensive DNA analysis and classified as a distinct motif known as motif 8: 5′-ACTAAGGA-T-TnCTTAGT-3′ (32). Simpler inverted repeats are bound by C.Csp231I and its ortholog C.EcoO109I (consensus: CTAAG–N_5–CTTAG) (39,66). There were no exact sequences matches near *racR*, but some sites resembled the consensus with just one or two changes. They were located in two spots: one close to the *racR* operator and other just after the initiation codon of the *racR* coding sequence (Figure 8A). To further test whether C protein binding occurred and affected the expression of the *racR* gene, we fused its natural promoter/operator sequence to a reporter gene (*lacZ*), to create the pLex-racRL plasmid. In this genetic system, the second compatible plas-

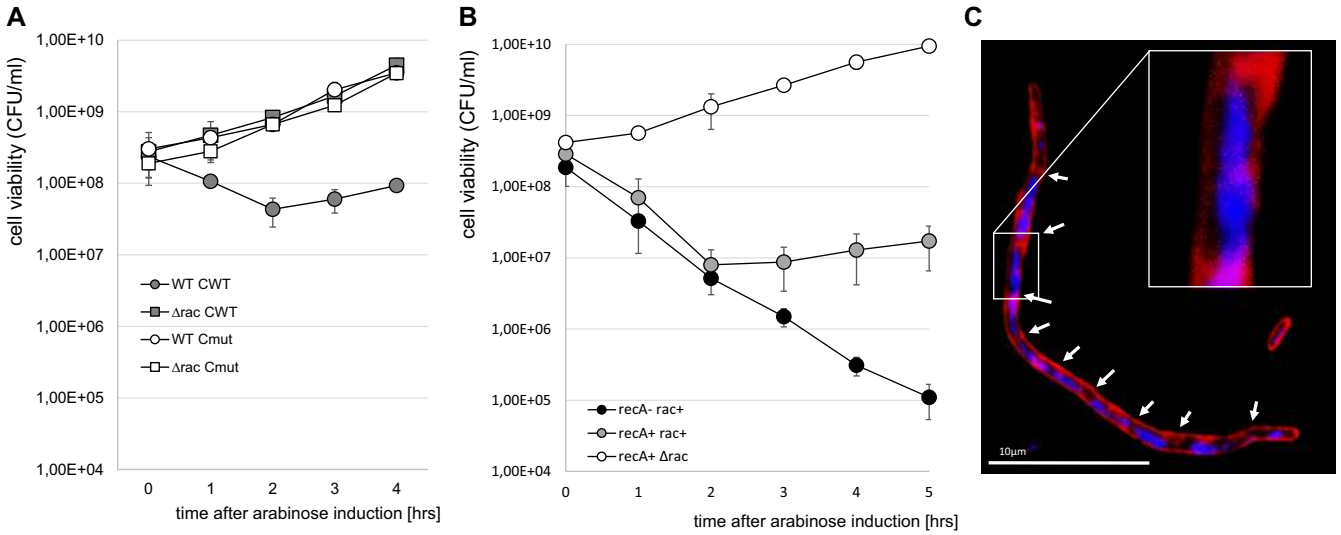


Figure 7. Cell viability kinetics of C protein induction in the genetic context of *rac*+ versus *Δrac*. (**A**) The *rac*+/*Δrac* cells carrying plasmids with the C wild-type (WT) gene or its mutant (Cmut) were induced with 0.01% of arabinose. The CFUs for four combinations of cultures (as in the legend) were measured just before induction and every hour after induction. (**B**) The same experiment was performed only for C WT in a different genetic context with the presence or absence of a combination of rac prophage and *recA* gene active or inactive, as indicated in the figure. The additional windows indicate representative single cell morphologies observed for tested cultures. The results are the averages (±SD) of three independent experiments. (**C**) The filamentous cells were stained with 4′,6-diamidino-2-phenylindole to visualize the nucleus (DAPI, blue) and SynaptoRed to visualize bacterial membranes (red). The arrows indicate the non-separated cell segments within the elongated single cell. The internal window shows the magnified single segment with still dividing DNA prior septum formation. A non-filamentous cell is also shown for reference.

mid produced the WT C protein (or its mutant unable to bind DNA; Cmut) under an inducible arabinose promoter, as shown in Figure 2. The MG1655 Δrac cells carrying the two plasmids were gradually induced with or without arabinose, and LacZ activity was assessed to reveal the level of *racR* expression (Figure 8B). The results showed that the RacR level was reduced ∼4-fold by only the WT C protein, but not by its mutant. The effect was dependent on the C protein concentration and was reached with induction by 0.01% of arabinose, which was consistent with experimental data from Figure 2.

Next, we questioned, which of these two putative spots of C protein binding were affecting RacR expression. We could not make any changes within the *racR* gene promoter/operator sequence without affecting *racR* expression, so we made substitutions within the coding sequence without disrupting the open reading frame (CTTAG→CGCAT; *racR*mut). We again performed the same experiments to measure the level of RacR expression (Figure 8C), which showed that although the level of expression was lower than WT, but stayed similar regardless of C protein activity, and showed no arabinose concentration-dependent reduction as was found for the WT *racR::lacZ* variant.

To reinforce the results and confirm that the effect was due to direct C protein binding to the *racR* region, we performed the EMSA. We used the same DNA fragments as a substrate, that had been cloned in front of the reporter gene, and had been used in the *in vivo* assay: *racR and racR*mut fragments. The EMSA reactions were performed with the same amount of Cy5-labeled ds DNA (25 nM) and increasing concentrations of purified C protein involving WT or its defective binding mutant Cmut (0–1750 nM) (Figure 9 and Figure S9 and S10 of Supplementary File 1). The shift in DNA−protein migration was observed only for a combination of *racR* and CWT, but not other combinations, such as *racR* and Cmut or *racRmut* and CWT/Cmut. The retarded DNA appeared using a fairly high C protein concentration (≥ 1.3 μM) indicating weak binding, which was not surprising because the *racR* gene sequence was not a natural target for C protein.

Possible disturbance of two cross-talking regulators

We screened the available *E. coli* genome sequences (Supplementary File 3) to find the co-occurrence of a gene with high amino acid sequence identity to a C protein, either C.Csp231I or C.EcoO109I, with the constituents of the Rac prophage region. Overall, we identified homologs of C protein in 23 genomes out of 495 tested (∼5%). These homologs were usually annotated as putative transcription factors (Supplementary File 3). Among the genomes with C.Csp231I homologs, 16 contained no *racR* or *ydaS* genes. In the case of genomes with C.EcoO109I homologs, seven genomes had *racR* and *ydaS* genes present, but at the same time *ydaT* was absent in six of them. Overall, we could detect only one genome out of the 23, in which the C homolog co-existed with *racR*/*ydaS*/*ydaT* genes. We could not confirm if all identified genes were active, but the co-occurrence appears to be rare. We also tested the original strain *E. coli* H709c, from which the EcoO109I R–M system was isolated, and found that the *racR* gene could not be amplified by PCR, suggesting its possible absence (Figure S7 of Supplementary File 1). In addition, we wondered why we never found any cell viability problem while overproducing C protein for its purification (60). The clear result showed no am-

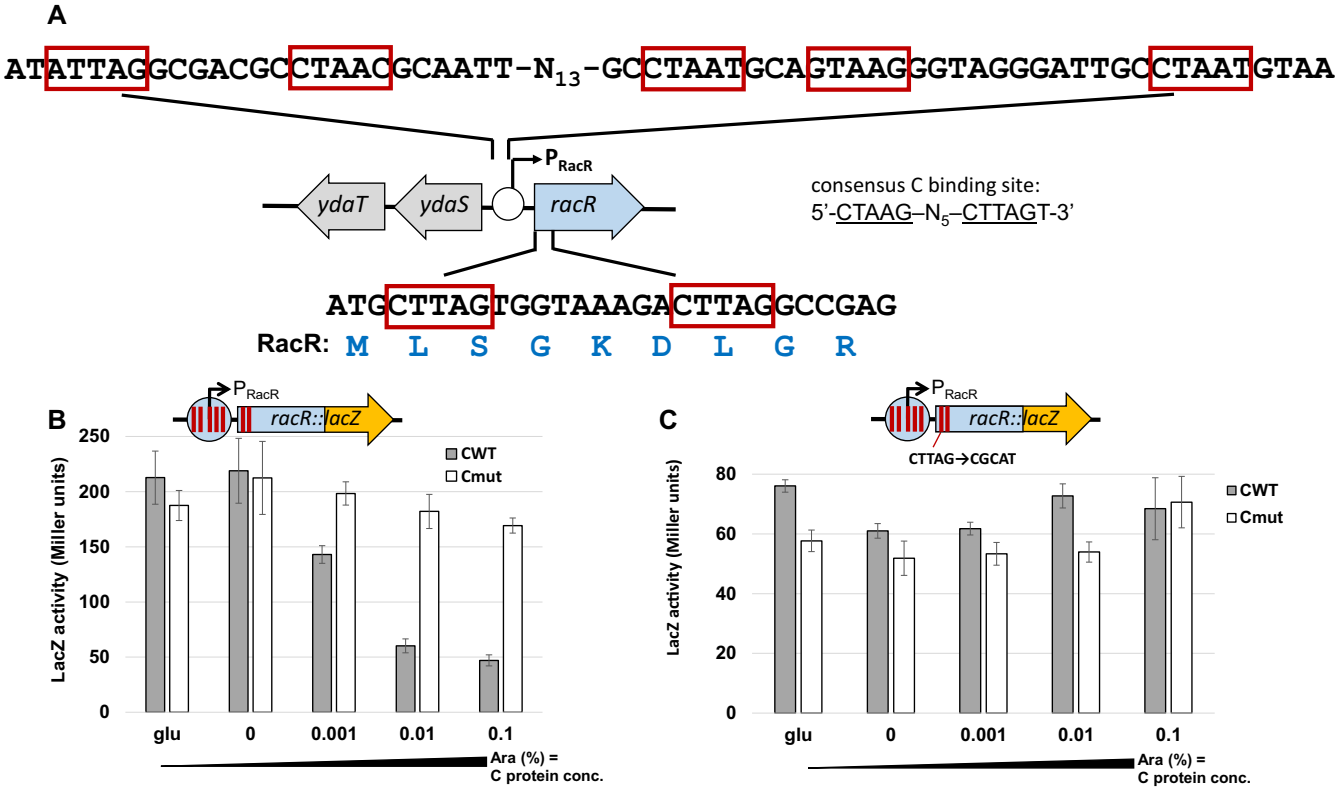


Figure 8. C protein is responsible for silencing the repressor of Rac prophage. (**A**) The *racR* and *ydaS*/*ydaT* genes are separated by the 123 bp intergenic region containing the promoter and binding site for RacR repressor. There are putative binding sites for C protein resembling its consensus sequence scattered within this intergenic region, as well as in the RacR coding sequence, both indicated by red frames. (**B**) The *in vivo* effect of C protein on reduction of RacR expression. In the experimental system, the cells contained two compatible plasmids: one with *racR* promoter fused in the same open reading frame to reporter gene *lacZ*; and a second plasmid delivered wild-type (WT) C gene under the inducible P_{araBAD} promoter or its C gene mutant (Cmut) unable to bind DNA. The cells were grown in minimal media with glycerol as the carbon source and with glucose (glu) or arabinose (ara) at gradually increasing concentrations (from 0 to 0.1%). The level of transcription activity was measured by LacZ activity in Miller units. (**C**) The same experiment was performed for the *racR::lacZ* variant, where the sequence of CTTAG within the *racR* coding sequence was replaced by CGCAT to break the putative C protein binding. For clarity, the fusion constructs are presented as schemes indicating the presence of the possible C binding sites marked by red bars in panel A and B. Grey bars show the effect of WT C protein, whereas the white bars show the Cmut variant. The results are the averages (± SD) of three independent experiments.

plification of *racR* gene in the *E. coli* BL21(DE3) genome or its derivative (Figure S7 of Supplementary File 1).

We also addressed the possible conflict of the two transcription factors, using a biological assay. We started a culture of *rac*+ cells with the p24 plasmid (carrying the active C gene) from a single colony. Each day, the cells were diluted into medium containing the appropriate antibiotic to maintain the plasmid, and screened for the filamentation phenotype. We noticed that the diluted, passaged cells formed two distinct types of colonies. Small colonies presented the initially-observed cell filamentation, while the large colonies lost that feature as revealed by microscopy (Figure 10A). There were no colonies with intermediate size. We measured the loss of filamentous phenotype by counting the proportion of large colonies and found that, after ∼75 generations, nearly 90% of cells had lost their filamentation (Figure 10B). We isolated the p24 (C+) plasmid DNA from large colonies, and showed that the C protein is still active, as these plasmids still induced filamentation in fresh *rac*+ cells. We concluded that some suppressor mutations appeared, possibly at the *rac* locus. However, the sequenced *ydaT-ydaS-racR* region isolated from large, non-filamentous colonies did not reveal mutation hot spots. It seems the mutations might suppress YdaST toxicity, though their location is hard to predict as we know neither YdaST function nor its target.

DISCUSSION

Adventitious cross-talk between two unrelated transcription factors

Fundamental processes in all living systems, such as the regulation of gene expression and coordination of genetic networks, rely on transcription factors. Those transcription factors are key elements involved in host adaption to a broad range of environments, including response to stresses and the presence of dynamically changing conditions (67). Transcription factors are also responsible for the flexibility of the genetic systems, which allows the gene expression circuits to evolve much faster than the cell's genetic content (68). The action of each transcription factor is tuned either to operate less specifically to affect large groups of genes (regulons), or to be highly specific and dedicated to regulating a single gene or operon (gene target). In addition, most

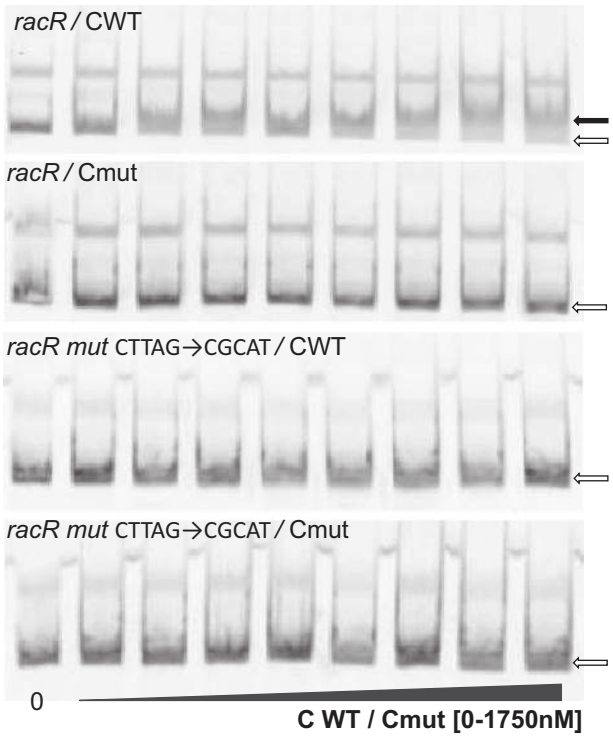


Figure 9. C protein and RacR cross-talk is mediated by C protein binding to the *racR* coding sequence. The 176 bp DNA substrates covering the *racR* gene operator and coding sequence (*racR* WT and its variant with CTTAG→CGCAT, *racR*mut) were prepared by PCR amplification, with one primer introducing the Cy5-label. Each binding reaction was carried out with the same amount of DNA (25 nM) and increasing concentrations of proteins (0; 500; 650; 800; 1000; 1150; 1300; 1500; 1750 nM). The reactions were assembled and processed as outlined in Materials and Methods, resolved on 5% native polyacrylamide gels, and visualized by a fluorescence scanner (Typhoon, GE Healthcare). Open and filled arrows denote positions of unbound DNA and shifted DNA–protein complex, respectively.

transcription factors are also able to recognize secondary target sites, making the regulatory systems even more complex (69). Such a multilayered regulatory network is prone to regulatory cross-talk, a situation where the transcription factor could have off-target effects (positive or negative), with potentially serious consequences for the cell (70–74). This phenomenon is likely to occur, especially for transcription factors having short recognition sites that include a degenerated sequence component. This process is still not well understood and many theoretical models have been built to describe such cross-talk (75). If we consider the entry of the genetic unit coding for transcription factor into the new host by means of HGT, it is even more complicated. In order to be integrated, the genetic module needs to exert its function, but, overall, it needs to fit into and sustain the host genetic network interconnectivity (76–78).

In this report, we present an example of regulatory cross-talk between a transcription factor (C protein) linked to the R–M system, and RacR repressor controlling the genes of cryptic *rac* prophage. As a result the connectivity of RacR regulon is disrupted. This in turn led to decreased cell viability associated with induction of the YdaST products. As a possible result, HGT of C protein-linked operons may kill new hosts carrying the *rac* locus. Some studies have shown that the gene incompatibility or their interfered network connectivity may form a significant functional barrier for HGT (4,52). Similar transcriptional incompatibility likely caused by transcription factor cross-talk has been proposed to limit interspecies hybridisation between various groups of organisms (79) and can—on a smaller scale—mediate competition between selfish genetic elements (80). Consistent with this possibility, we detected only one case out of 23 analyzed *E. coli* genomes, where genes for C.Csp231I/EcoO109I-like proteins co-occurred with *racR*/*ydaS*/*ydaT* genes. This also might support the possibility that transcription factors in active cross-talk may result in incompatibility leading to death of the host cell.

In general, the DNA target sites (C-boxes) of C proteins are formed from four inverted repeats to be occupied as a tetramer. The known consensus sequences for the C.Csp231I dimer and its homolog C.EcoO109I, (5'-CTAAG-N_5-CTTAG-3'; the most conserved bases are underlined) indicate significant potential of these C proteins interacting with DNA at secondary sites (32). We searched for such sites in the *E. coli* genome and found six sites with the spacer *N*, where *N* was $3 < N < 11$ (but $N \neq 5$ as in primary sites; data not shown). None of them were located at the *rac* locus. This suggests that the C protein may bind to even less-similar sequences possibly as a monomer. The *in vivo* tests showed the effect of C protein on *racR* expression (Figure 9), which prompted us to screen DNA sequences in the *racR* operator/promoter region. We found five motifs with single differences from the consensus sequence (5′-CTAAG-3′), in addition to two such sites close to the translation site of the *racR* gene. We confirmed the weak C protein binding to RacR gene region by an *in vitro* assay, and found that the disruption of the latter site affected the C-dependent silencing of the *racR* expression in the *in vivo* test. All studied C proteins exerted their regulatory function by binding to the promoter region and interfering with transcription initiation, but recent studies of C.Kpn2I have reported that the same function could be achieved by blocking transcription at the elongation stage (81). C.Kpn2I has two binding sites within its coding sequence (unlike other known C proteins) very close to its ATG codon, which probably creates a strong roadblock for the elongating RNA polymerase complex. It is possible that the Csp231I C protein affects RacR expression *via* the same mechanism. In addition, there is another documented example of C regulatory proteins linked to the Type I R–M system of *Mycoplasma*, which showed its ability to interact with nonspecific targets, leading to deleterious events. These mycoplasmal C proteins efficiently bind to additional occurrences of their consensus sequence $(GTGTTAN_5)_2$, which are located in the *Mycoplasma* genome within promoters for the protease gene, *clpB,* and the tRNA gene cluster. Overexpression of this C gene is lethal demonstrating the serious consequences such gene expression cross-talk can have (82). Nonspecific interactions are not limited to transcription factors.

Implication for Rac prophage gene expression regulation

Rac prophage (lambda-like) is highly conserved in the *E. coli* genome (83). Although the phage is cryptic, it affects host cell physiology by helping to withstand osmotic, oxida-

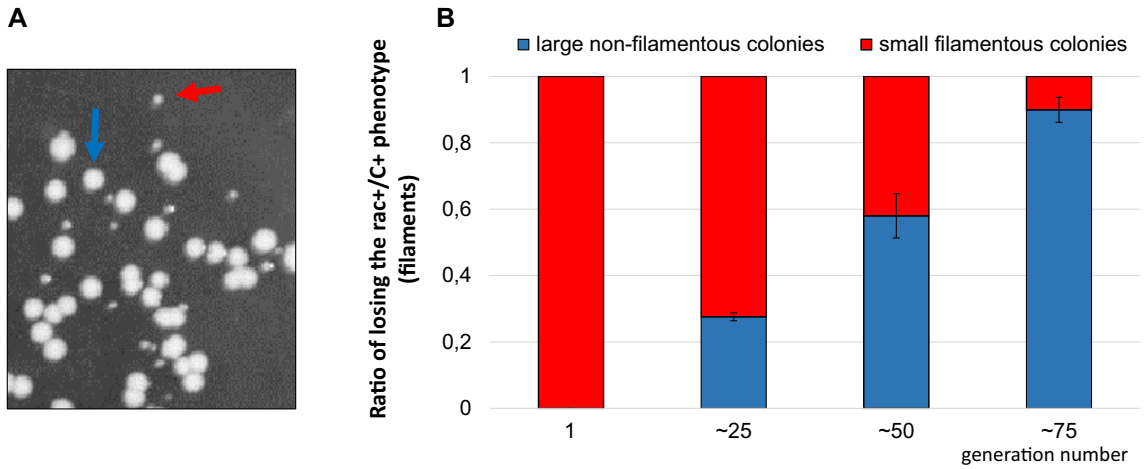


Figure 10. Possible conflict of the two transcription factors revealed as C gene can induce a selective pressure for *rac*+ cells to lose the filamentation phenotype. The *rac*+ cells carrying the active C gene passaged into fresh medium every ~25 generations. At the start of experiments all cells were filamentous and formed uniformly small colonies. With time, two distinct types of colony morphology were observed: small colonies with extensive filamentation of constituent cells, and large colonies containing normal, rod-shaped cells (**A**). Filamentation phenotype loss was measured as fraction of CFU_{small} to CFU_{large} (**B**).

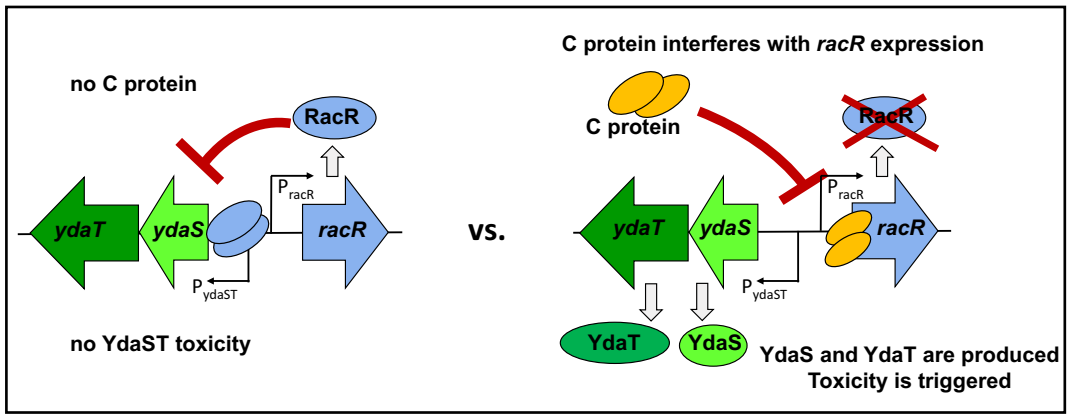


Figure 11. Possible scenario of transcriptional cross-talk between the C protein and RacR repressor. At normal conditions, when C protein is not present, the RacR repressor binds within the intergenic region, which blocks the possible common promoter/operator region for *ydaS* and *ydaT* genes on the bottom DNA strand. As a result, the YdaS and YdaT are completely silenced. When DNA fragment carrying the gene of C protein enters the cell, the excess of C protein, which cannot occupy its natural target site, starts to bind within the *racR* gene sequence. This can lead to the inhibition of gene expression. Insufficient RacR repressor is not capable of inhibiting *ydaST* expression. When YdaST is produced, the likely toxicity of YdaT triggers the cell division defect (or other) and cell filamentation occurs.

tive, and acid stresses, to forming a biofilm and to increasing fitness (63). Most *E. coli* sequenced genomes have at least part of the *rac* locus (54). The *rac* region comprises only 29 genes, although some of them have not yet been characterized, and their genetic relationships have not been determined. Our transcriptome analyses indicated that the *racR* gene may play a central role in Rac prophage gene expression. We found that the *racR* gene was the only gene downregulated when C protein was expressed; in contrast several nearby genes were highly upregulated. This may suggest a repressor function for RacR, not only for adjacent and neighboring genes, but also for distantly located genes, like for the RalRA toxin–antitoxin system (6 kb apart). Usually phage genomes present highly compact gene expression networks, mediated by numerous polycistronic transcription units and regulators responsible for driving the phage life cycles (84). RacR (158 aa) is similar in function to C protein (98 aa), though with no clear sequence similarity (60,66) (Figure S8 of Supplementary File 1). It is also a transcription factor of lambdoid phage origin, with an HTH motif to interact with DNA targets of its gene upstream region located amid promoter boxes with a feedback loop potential. This 123 bp region is shared with the *ydaS* gene divergently coded on the opposite DNA strand (Figure 8A), with a strong negative effect of RacR on *ydaST* expression (53,54). YdaS and the adjacent YdaT are annotated with only putative functions (85). Both have been suggested to act in a similar manner to Cro and CII repressors, which coordinate lambda bacteriophage gene expression related to lysis *versus* lysogeny decision during its life cycle (83,86). However, only the YdaS protein, but not the YdaT protein, contains an HTH motif. It is possible that the inseparable yet toxic activity of the *ydaST* operon might be due to the vitally important controlling switch of YdaS (regulator) to activate

the *ydaT* gene via a common promoter and possible polycistronic transcript. The exertion of toxicity manifested by extensive cell filamentation might therefore originate solely from the YdaT protein, although the overexpression of transcription factors, like lambda CII can also lead to cell elongation (87,88). The present study showed the indispensable role of the RacR repressor in the Rac prophage gene network and showed how an external transcription factor is able to interfere with this network via transcriptional crosstalk (Figure 11). The mechanism of YdaT toxicity alone is not understood, as well as its putative function in the Rac prophage context. Previously, it was linked indirectly to cell division inhibition by acting on DNA replication or chromosome segregation due to the absence of well-segregated nucleoids in filamentous cells (89). However, in contrast, our results showed a strong DAPI signal distributed at each non-separated segment of elongated cells, which was consistent with other studies (53). It is probable that the cell division defect was related to the cell separation stage (Figure 7C).

DATA AVAILABILITY

The results have been deposited in the NCBI GEO (accession number GSE126248).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Drs Katarzyna Potrykus, Calin Guet, Yunxue Guo, Revathy Krishnamurthi and Aswin Seshasayee for the gift of strains/plasmids. We also acknowledge the technical help of Ania Maliszewska, and Dr Malgorzata Kapusta.

FUNDING

National Science Center (Poland) [2015/19/B/NZ2/01835 to I.M.]; European Union's Horizon 2020 Research and Innovation Program under the Marie Sklodowska-Curie [665778 to L.P.P.]. Funding for open access charge: National Science Center (Poland) [2015/19/B/NZ2/01835].
Conflict of interest statement. None declared.

REFERENCES

1. Vos,M., Hesselman,M.C., Te Beek,T.A., van Passel,M.W.J. and Eyre-Walker,A. (2015) Rates of lateral gene transfer in prokaryotes: high but why? *Trends Microbiol.*, **23**, 598–605.
2. Treangen,T.J. and Rocha,E.P. (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLos Genet.*, **7**, e1001284.
3. Jeltsch,A. (2003) Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene*, **317**, 13–16.
4. Thomas,C.M. and Nielsen,K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.*, **3**, 711–721.
5. Johnston,C.D., Skeete,C.A., Fomenkov,A., Roberts,R.J. and Rittling,S.R. (2017) Restriction-modification mediated barriers to exogenous DNA uptake and incorporation employed by *Prevotella intermedia*. *PLoS One*, **12**, e0185234.
6. Vasu,K. and Nagaraja,V. (2013) Diverse functions of restriction–modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.*, **77**, 53–72.
7. Oliveira,P.H., Touchon,M. and Rocha,E.P. (2016) Regulation of genetic flux between bacteria by restriction–modification systems. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 5658–5663.
8. Labrie,S.J., Samson,J.E. and Moineau,S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.*, **8**, 317–327.
9. Oliveira,P.H., Touchon,M. and Rocha,E.P. (2014) The interplay of restriction–modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.*, **42**, 10618–10631.
10. Makarova,K.S., Wolf,Y.I., Snir,S. and Koonin,E.V. (2011) Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.*, **193**, 6039–6056.
11. Furuta,Y., Abe,K. and Kobayashi,I. (2010) Genome comparison and context analysis reveals putative mobile forms of restriction–modification systems and related rearrangements. *Nucleic Acids Res.*, **38**, 2428–2443.
12. Werbowy,O. and Kaczorowski,T. (2016) Plasmid pEC156, a naturally occurring *Escherichia coli* genetic element that carries genes of the EcoVIII restriction-modification system, is mobilizable among *Enterobacteria*. *PLoS ONE*, **11**, e0148355.
13. Loenen,W.A., Dryden,D.T., Raleigh,E.A., Wilson,G.G. and Murray,N.E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.*, **42**, 3–19.
14. Mruk,I. and Kobayashi,I. (2014) To be or not to be: regulation of restriction–modification systems and other toxin-antitoxin systems. *Nucleic Acids Res.*, **42**, 70–86.
15. Mruk,I. and Kobayashi,I. (2016) In: De,B and Frans,J (ed). *Epigenetics Mediated by Restriction Modification Systems. Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria*. Wiley Blackwell, cop, Hoboken, Vol. **1**, pp. 425–436.
16. Furuta,Y., Namba-Fukuyo,H., Shibata,T.F., Nishiyama,T., Shigenobu,S., Suzuki,Y., Sugano,S., Hasebe,M. and Kobayashi,I. (2014) Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLos Genet.*, **10**, e1004272.
17. Krebes,J., Morgan,R.D., Bunk,B., Spröer,C., Luong,K., Parusel,R., Anton,B.P., König,C., Josenhans,C., Overmann,J. *et al.* (2014) The complex methylome of the human gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.*, **42**, 2415–2432.
18. Shell,S.S., Prestwich,E.G., Baek,S.H., Shah,R.R., Sassetti,C.M., Dedon,P.C. and Fortune,S.M. (2013) DNA methylation impacts gene expression and ensures hypoxic survival of *Mycobacterium tuberculosis*. *PLoS Pathog.*, **9**, e1003419.
19. Gonzalez,D. and Collier,J. (2013) DNA methylation by CcrM activates the transcription of two genes required for the division of *Caulobacter crescentus*. *Mol. Microbiol.*, **88**, 203–218.
20. Mouammine,A. and Collier,J. (2018) The impact of DNA methylation in Alphaproteobacteria. *Mol. Microbiol.*, **110**, doi:10.1111/mmi.14079.
21. Atack,J.M., Tan,A., Bakaletz,L.O., Jennings,M.P. and Seib,K.L. (2018) Phasevarions of bacterial pathogens: methylomics sheds new light on old enemies. *Trends Microbiol.*, **26**, 715–726.
22. Gauntlett,J.C., Nilsson,H.O., Fulurija,A., Marshall,B.J. and Benghezal,M. (2014) Phase-variable restriction/modification systems are required for *Helicobacter pylori* colonization. *Gut Pathog*, **6**, 35.
23. Casselli,T., Tourand,Y., Scheidegger,A., Arnold,W.K., Proulx,A., Stevenson,B. and Brissette,C.A. (2018) DNA methylation by restriction modification systems affects the global transcriptome profile. *J. Bacteriol.*, **200**, e00395-18.
24. Nell,S., Estibariz,I., Krebes,J., Bunk,B., Graham,D.Y., Overmann,J., Song,Y., Spröer,C., Yang,I., Wex,T. *et al.* (2018) Genome and methylome variation in *Helicobacter pylori* with a cag pathogenicity island during early stages of human infection. *Gastroenterology*, **154**, 612–623.
25. Doberenz,S., Eckweiler,D., Reichert,O., Jensen,V., Bunk,B., Spröer,C., Kordes,A., Frangipani,E., Luong,K., Korlach,J. *et al.* (2017) Identification of a *Pseudomonas aeruginosa* PAO1 DNA methyltransferase, Its targets, and physiological roles. *MBio*, **8**, e02312-16.
26. Katna,A., Boratynski,R., Furmanek-Blaszk,B., Zolcinska,N. and Sektas,M. (2010) Unbalanced restriction impairs SOS-induced DNA repair effects. *J. Microbiol. Biotechnol.*, **20**, 30–38.

27. Naito,T., Kusono,K. and Kobayashi,I. (1995) Selfish behavior of restriction–modification systems. *Science*, **267**, 897–899.
28. Pleška,M., Qian,L., Okura,R., Bergmiller,T., Wakamoto,Y., Kussell,E. and Guet,C.C. (2016) Bacterial autoimmunity due to a restriction-modification system. *Curr. Biol.*, **26**, 404–409.
29. Handa,N. and Kobayashi,I. (1999) Post-segregational killing by restriction modification gene complexes: observations of individual cell deaths. *Biochimie*, **81**, 931–938.
30. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **43**, D298–D299.
31. Tao,T., Bourne,J.C. and Blumenthal,R.M. (1991) A family of regulatory genes associated with type II restriction–modification systems. *J. Bacteriol.*, **173**, 1367–1375.
32. Sorokin,V., Severinov,K. and Gelfand,M.S. (2009) Systematic prediction of control proteins and their DNA binding sites. *Nucleic Acids Res.*, **37**, 441–451.
33. McGeehan,J.E., Streeter,S.D., Papapanagiotou,I., Fox,G.C. and Kneale,G.G. (2005) High-resolution crystal structure of the restriction–modification controller protein C.AhdI from *Aeromonas hydrophila*. *J. Mol. Biol.*, **346**, 689–701.
34. Aggarwal,A.K., Rodgers,D.W., Drottar,M., Ptashne,M. and Harrison,S.C. (1988) Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science*, **242**, 899–907.
35. Bogdanova,E., Djordjevic,M., Papapanagiotou,I., Heyduk,T., Kneale,G. and Severinov,K. (2008) Transcription regulation of the type II restriction–modification system AhdI. *Nucleic Acids Res.*, **36**, 1429–1442.
36. Bogdanova,E., Zakharova,M., Streeter,S., Taylor,J., Heyduk,T., Kneale,G. and Severinov,K. (2009) Transcription regulation of restriction–modification system Esp1396I. *Nucleic Acids Res.*, **37**, 3354–3366.
37. Cesnaviciene,E., Mitkaite,G., Stankevicius,K., Janulaitis,A. and Lubys,A. (2003) Esp1396I restriction–modification system: structural organization and mode of regulation. *Nucleic Acids Res.*, **31**, 743–749.
38. Mruk,I., Rajesh,P. and Blumenthal,R.M. (2007) Regulatory circuit based on autogenous activation-repression: roles of C-boxes and spacer sequences in control of the PvuII restriction–modification system. *Nucleic Acids Res.*, **35**, 6935–6952.
39. Kita,K., Tsuda,J. and Nakai,S.Y. (2002) C.EcoO109I, a regulatory protein for production of EcoO109I restriction endonuclease, specifically binds to and bends DNA upstream of its translational start site. *Nucleic Acids Res.*, **30**, 3558–3565.
40. Semenova,E., Minakhin,L., Bogdanova,E., Nagornykh,M., Vasilov,A., Heyduk,T., Solonin,A., Zakharova,M. and Severinov,K. (2005) Transcription regulation of the EcoRV restriction–modification system. *Nucleic Acids Res.*, **33**, 6942–6951.
41. Ives,C.L., Sohail,A. and Brooks,J.E. (1995) The regulatory C proteins from different restriction–modification systems can cross-complement. *J. Bacteriol.*, **177**, 6313–6315.
42. Vijesurier,R.M., Carlock,L., Blumenthal,R.M. and Dunbar,J.C. (2000) Role and mechanism of action of C.PvuII, a regulatory protein conserved among restriction–modification systems. *J. Bacteriol.*, **182**, 477–487.
43. McGeehan,J.E., Papapanagiotou,I., Streeter,S.D. and Kneale,G.G. (2006) Cooperative binding of the C.AhdI controller protein to the C/R promoter and its role in endonuclease gene expression. *J. Mol. Biol.*, **358**, 523–531.
44. Mruk,I. and Blumenthal,R.M. (2009) Tuning the relative affinities for activating and repressing operators of a temporally regulated restriction–modification system. *Nucleic Acids Res.*, **37**, 983–998.
45. Ball,N.J., McGeehan,J.E., Streeter,S.D., Thresh,S.J. and Kneale,G.G. (2012) The structural basis of differential DNA sequence recognition by restriction–modification controller proteins. *Nucleic Acids Res.*, **40**, 10532–10542.
46. Ball,N., Streeter,S.D., Kneale,G.G. and McGeehan,J.E. (2009) Structure of the restriction–modification controller protein C.Esp1396I. *Acta Crystallogr. D. Biol. Crystallogr.*, **65**, 900–905.
47. Martin,R.N., McGeehan,J.E., Ball,N.J., Streeter,S.D., Thresh,S.J. and Kneale,G.G. (2013) Structural analysis of DNA-protein complexes regulating the restriction–modification system Esp1396I. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **69**, 962–966.
48. Nakayama,Y. and Kobayashi,I. (1998) Restriction-modification gene complexes as selfish gene entities: roles of a regulatory system in their establishment, maintenance, and apoptotic mutual exclusion. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 6442–6447.
49. Mruk,I. and Blumenthal,R.M. (2008) Real-time kinetics of restriction–modification gene expression after entry into a new host cell. *Nucleic Acids Res.*, **36**, 2581–2593.
50. Williams,K., Savageau,M.A. and Blumenthal,R.M. (2013) A bistable hysteretic switch in an activator-repressor regulated restriction–modification system. *Nucleic Acids Res.*, **41**, 6045–6057.
51. Morozova,N., Sabantsev,A., Bogdanova,E., Fedorova,Y., Maikova,A., Vedyaykin,A., Rodic,A., Djordjevic,M., Khodorkovskii,M. and Severinov,K. (2016) Temporal dynamics of methyltransferase and restriction endonuclease accumulation in individual cells after introducing a restriction–modification system. *Nucleic Acids Res.*, **44**, 790–800.
52. Sorek,R., Zhu,Y., Creevey,C.J., Francino,M.P., Bork,P. and Rubin,E.M. (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, **318**, 1449–1452.
53. Bindal,G., Krishnamurthi,R., Seshasayee,A.S.N. and Rath,D. (2017) CRISPR-Cas-mediated gene silencing reveals RacR to be a negative regulator of YdaS and YdaT toxins in. *mSphere*, **2**, e00483-17.
54. Krishnamurthi,R., Ghosh,S., Khedkar,S. and Seshasayee,A.S.N. (2017) Repression of YdaS toxin is mediated by transcriptional repressor RacR in the cryptic. *mSphere*, **2**, e00392-17.
55. Kita,K., Tsuda,J., Kato,T., Okamoto,K., Yanase,H. and Tanaka,M. (1999) Evidence of horizontal transfer of the EcoO109I restriction–modification gene to *Escherichia coli* chromosomal DNA. *J. Bacteriol.*, **181**, 6822–6827.
56. Mruk,I. and Kaczorowski,T. (2007) A rapid and efficient method for cloning genes of type II restriction–modification systems by use of a killer plasmid. *Appl. Environ. Microbiol.*, **73**, 4286–4293.
57. Eddy,S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.*, **7**, e1002195.
58. Sharan,S.K., Thomason,L.C., Kuznetsov,S.G. and Court,D.L. (2009) Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.*, **4**, 206–223.
59. Datta,S., Costantino,N. and Court,D.L. (2006) A set of recombineering plasmids for gram-negative bacteria. *Gene*, **379**, 109–115.
60. Rezulak,M., Borsuk,I. and Mruk,I. (2016) Natural C-independent expression of restriction endonuclease in a C protein-associated restriction–modification system. *Nucleic Acids Res.*, **44**, 2646–2660.
61. Nagamalleswari,E., Rao,S., Vasu,K. and Nagaraja,V. (2017) Restriction endonuclease triggered bacterial apoptosis as a mechanism for long time survival. *Nucleic Acids Res.*, **45**, 8423–8434.
62. Handa,N., Ichige,A., Kusano,K. and Kobayashi,I. (2000) Cellular responses to postsegregational killing by restriction–modification genes. *J. Bacteriol.*, **182**, 2218–2229.
63. Wang,X., Kim,Y., Ma,Q., Hong,S.H., Pokusaeva,K., Sturino,J.M. and Wood,T.K. (2010) Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.*, **1**, 147.
64. Guo,Y., Quiroga,C., Chen,Q., McAnulty,M.J., Benedik,M.J., Wood,T.K. and Wang,X. (2014) RalR (a DNase) and RalA (a small RNA) form a type I toxin-antitoxin system in *Escherichia coli*. *Nucleic Acids Res.*, **42**, 6448–6462.
65. Conter,A., Bouché,J.P. and Dassain,M. (1996) Identification of a new inhibitor of essential division gene *ftsZ* as the *kil* gene of defective prophage Rac. *J. Bacteriol.*, **178**, 5100–5104.
66. Shevtsov,M.B., Streeter,S.D., Thresh,S.J., Swiderska,A., McGeehan,J.E. and Kneale,G.G. (2015) Structural analysis of DNA binding by C.Csp231I, a member of a novel class of R–M controller proteins regulating gene expression. *Acta Crystallogr. D. Biol. Crystallogr.*, **71**, 398–407.
67. van Hijum,S.A., Medema,M.H. and Kuipers,O.P. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol. Mol. Biol. Rev.*, **73**, 481–509.
68. Lozada-Chávez,I., Janga,S.C. and Collado-Vides,J. (2006) Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res.*, **34**, 3434–3445.
69. Wunderlich,Z. and Mirny,L.A. (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.*, **25**, 434–440.
70. Rowland,M.A., Abdelzaher,A., Ghosh,P. and Mayo,M.L. (2017) Crosstalk and the dynamical modularity of feed-forward loops in transcriptional regulatory networks. *Biophys. J.*, **112**, 1539–1550.

71. Todeschini,A.L., Georges,A. and Veitia,R.A. (2014) Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet.*, **30**, 211–219.
72. San Millan,A., Toll-Riera,M., Qi,Q. and MacLean,R.C. (2015) Interactions between horizontally acquired genes create a fitness cost in *Pseudomonas aeruginosa*. *Nat. Commun.*, **6**, 6845.
73. Miyazaki,R., Yano,H., Sentchilo,V. and van der Meer,J.R. (2018) Physiological and transcriptome changes induced by *Pseudomonas putida* acquisition of an integrative and conjugative element. *Sci. Rep.*, **8**, 5550.
74. Harrison,E., Guymer,D., Spiers,A.J., Paterson,S. and Brockhurst,M.A. (2015) Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Curr. Biol.*, **25**, 2034–2039.
75. Friedlander,T., Prizak,R., Guet,C.C., Barton,N.H. and Tkačik,G. (2016) Intrinsic limits to gene regulation by global crosstalk. *Nat. Commun.*, **7**, 12307.
76. Davids,W. and Zhang,Z. (2008) The impact of horizontal gene transfer in shaping operons and protein interaction networks–direct evidence of preferential attachment. *BMC Evol. Biol.*, **8**, 23.
77. Dorman,C.J. (2009) Regulatory integration of horizontally-transferred genes in bacteria. *Front. Biosci. (Landmark Ed)*, **14**, 4103–4112.
78. Lercher,M.J. and Pál,C. (2008) Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.*, **25**, 559–567.
79. Tulchinsky,A.Y., Johnson,N.A., Watt,W.B. and Porter,A.H. (2014) Hybrid incompatibility arises in a sequence-based bioenergetic model of transcription factor binding. *Genetics*, **198**, 1155–1166.
80. Kobayashi,I. (2001) Behavior of restriction–modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742–3756.
81. Klimuk,E., Bogdanova,E., Nagornykh,M., Rodic,A., Djordjevic,M., Medvedeva,S., Pavlova,O. and Severinov,K. (2018) Controller protein of restriction–modification system Kpn2I affects transcription of its gene by acting as a transcription elongation roadblock. *Nucleic Acids Res.*, **46**, 10810–10826.
82. Fisunov,G.Y., Evsyutina,D.V., Manuvera,V.A. and Govorun,V.M. (2017) Binding site of restriction–modification system controller protein in *Mollicutes*. *BMC Microbiol.*, **17**, 26.
83. Casjens,S. (2003) Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.*, **49**, 277–300.
84. Atsumi,S. and Little,J.W. (2004) Regulatory circuit design and evolution using phage lambda. *Genes Dev.*, **18**, 2086–2094.
85. Jobling,M.G. (2018) Ectopic expression of the *ydaS* and *ydaT*genes of the cryptic prophage Rac of *Escherichia coli* K-12 may be toxic but do they really encode toxins?: a case for using genetic context to understand function. *mSphere*, **3**, e00163-18.
86. Ptashne,M., Johnson,A.D. and Pabo,C.O. (1982) A genetic switch in a bacterial virus. *Sci. Am.*, **247**, 128–130.
87. Kedzierska,B., Glinkowska,M., Iwanicki,A., Obuchowski,M., Sojka,P., Thomas,M.S. and Wegrzyn,G. (2003) Toxicity of the bacteriophage lambda cII gene product to *Escherichia coli* arises from inhibition of host cell DNA replication. *Virology*, **313**, 622–628.
88. Rajamanickam,K. and Hayes,S. (2018) The bacteriophage lambda CII phenotypes for complementation, cellular toxicity and replication inhibition are suppressed in CII-OOP constructs expressing the small RNA OOP. *Viruses*, **10**, E115.
89. Campos,M., Govers,S.K., Irnov,I., Dobihal,G.S., Cornet,F. and Jacobs-Wagner,C. (2018) Genomewide phenotypic analysis of growth, cell morphogenesis, and cell cycle events in. *Mol. Syst. Biol.*, **14**, e7573.