Molecular basis for lethal cross-talk between two unrelated bacterial transcription factors - the regulatory protein of a restriction-modification system and the repressor of a defective prophage

Aleksandra Wisniewska¹, Ewa Wons¹, Katarzyna Potrykus², Rebecca Hinrichs³,⁴, Katarzyna Gucwa¹, Peter L. Graumann³,⁴ and Iwona Mruk¹,*

¹Department of Microbiology, Faculty of Biology, University of Gdansk, Wita Stwosza 59, Gdansk 80-308, Poland, ²Department of Bacterial Molecular Genetics, Faculty of Biology, University of Gdansk, Wita Stwosza 59, Gdansk 80-308, Poland, ³SYNMIKRO, LOEWE Center for Synthetic Microbiology, Philipps Universität Marburg, Germany and ⁴Department of Chemistry, Philipps Universität Marburg, Hans-Meerwein-Strasse 6, 35032 Marburg, Germany

Received July 25, 2022; Revised September 29, 2022; Editorial Decision September 30, 2022; Accepted October 06, 2022

ABSTRACT

Bacterial gene expression depends on the efficient functioning of global transcriptional networks, however their interconnectivity and orchestration rely mainly on the action of individual DNA binding proteins called transcription factors (TFs). TFs interact not only with their specific target sites, but also with secondary (off-target) sites, and vary in their promiscuity. It is not clear yet what mechanisms govern the interactions with secondary sites, and how such rewiring affects the overall regulatory network, but this could clearly constrain horizontal gene transfer. Here, we show the molecular mechanism of one such off-target interaction between two unrelated TFs in Escherichia coli: the C regulatory protein of a Type II restriction-modification system, and the RacR repressor of a defective prophage. We reveal that the C protein interferes with RacR repressor expression, resulting in derepression of the toxic YdaT protein. These results also provide novel insights into regulation of the racR-ydaST operon. We mapped the C regulator interaction to a specific off-target site, and also visualized C protein dynamics, revealing intriguing differences in single molecule dynamics in different genetic contexts. Our results demonstrate an apparent example of horizontal gene transfer leading to adventitious TF cross-talk with negative effects on the recipient’s viability. More broadly, this study represents an experimentally-accessible model of a regulatory constraint on horizontal gene transfer.

INTRODUCTION

Control of bacterial gene expression is one of the most intriguing examples of biological plasticity in response to changing environmental conditions (1,2). It is mediated by multiple combinatorial entities, including DNA sequence and topology, proteins, modifications of those proteins via co-regulatory molecules or post-translational modification, and non-coding RNAs. Complex regulatory networks coordinate communication between individual genes or their groups to exert certain cascades of signals (3). The interconnectivity of these networks relies mainly on small DNA binding proteins called transcription factors (TFs), which interact with their DNA target sites (TSs) to modulate gene expression over a wide range, from a complete shut-down to full gene induction (4). The action of most TFs is directed at the stage of transcription initiation, where repressors prevent binding and progression of RNA polymerase by steric hinderance, while activators stimulate gene expression, often by stabilizing association of RNA polymerase with promoter sites or facilitation promoter DNA opening (5,6). The characteristic mode of action of each TF is adjusted to operate either less specifically and affect large groups of genes (regulons) like global TFs, or to be highly specific and regulate a single gene or operon (dedicated gene target, local regulator) (7). TF action is determined in part by diffusion on DNA to reach their specific sequence recognition site (TS). However, only a small fraction of TFs have their TSs comprehensively characterized (8). In addition, most TFs also recognize secondary TSs (off-target sites), making the cellular regulatory landscape even more complex, and sometimes noisy (9). It is not clear yet to what extent regulatory interactions between the regulons of certain TFs overlap with one another within the cell’s regulatory network. For example, it is challenging to predict whether two TFs
MATERIALS AND METHODS

Strains, plasmids and oligonucleotides

E. coli K-12 strains used in this study are described in Supplementary Table S1 of Supplementary Materials. E. coli Δrac strain was used to carry out in vivo assays with plasmids harboring the ydaS-racR operon (37). The source of genes of Csp231I R-M system was Citrobacter sp. RFL231 (19,38). The plasmids used are listed in Supplementary Table S2, the oligonucleotides/DNA fragments used are shown in Supplementary Table S3, Supplementary Data.

Effect of TFs delivered in trans on racR expression

The genes for selected TFs were cloned downstream of the arabinose-inducible P araBAD promoter in vector pBAD33 (39) (Supplementary Table S2). A second compatible plasmid carried the racR gene with its natural promoter/operator sequence fused to a reporter gene (lacZ). Arabinose induction experiments were performed in M9-minimal medium with 0.2% glycerol as the carbon source (40). Briefly, single colonies were used to inoculate overnight cultures in M9 media supplemented with appropriate antibiotics. The cells were then resuspended and divided among flasks containing M9-minimal media with varied concentrations of L-arabinose. After about 5 h of subsequent cultivation, ONPG hydrolysis was measured and Miller units were calculated as previously described (23,41). The ONPG assay was done in the E. coli K-12 Δrac strain (37), so that genomic presence of the Rac repressor did not affect expression quantitation. Details on plasmid cloning and features are outlined in the Supplementary Data (Supplementary Table S2).

Overproduction and purification of the C protein and RacR repressor

For the C-terminally-His-tagged C protein (WT, or Cmut with A33G R34E Q37A, unable to bind its target site), as well as the RacR repressor, overproduction and purification were performed as before (19), except that the purified proteins were resuspended in Tris buffer: 10 mM Tris–HCl (pH 8.0) and 1 mM EDTA (Supplementary Table S2, Supplementary Figure S1).

Electrophoretic mobility shift assays (EMSA)

DNA substrates (271 bp) were PCR-amplified using fluorescently Cy5-labeled primer at one end (Supplementary Table S3 of Supplementary Data). They covered the entire racR promoter/operator/intergenic region (IGR) and part of its coding sequence (racR), as well as the ydaS gene; i.e. region −186 to +85 bp, where +1 refers to racR translational start site. Reactions containing 25 nM DNA and the indicated purified C protein/RacR repressor concentrations were prepared in a binding buffer [10 mM Tris–HCl (pH 8.0), 50 mM NaCl, 10 mM MgCl₂, 1 μg of poly(dIdC) as a non-specific competitor] in a final volume of 20 μl, and incubated for 20 min at 22°C. Reactions for competitive EMSA experiments were prepared in the same way, with
the first protein incubated for 20 min at identical concentrations, followed by addition on increased concentration of the second protein and 20 min incubation time. DNA fragments used as substrates run as two specific bands due to presence of many inverted repeats. Samples were electrophoresed on 5% native polyacrylamide gels in 0.5 × TBE buffer at 22°C or on 1% agarose gels. Detection of the Cy5-labeled DNA was performed using the Typhoon 9200 variable mode imager (Molecular Dynamics, USA).

**DNaseI footprinting assays**

Protein-DNA binding reactions were performed with 20 nM DNA in 20 μl of the following buffer: 10 mM Tris–HCl (pH 8.0), 10 mM MgCl₂, 50 mM NaCl. DNA substrates were PCR-amplified using fluorescently Cy5-labeled primer at one end (Supplementary Table S3 of Supplementary Data). For C protein footprints: two DNA fragments were used. One was 271 bp, covering region from −186 to +85 bp, and the second was 175 bp (−90 to +85 bp, where +1 refers to racR translational start site). For RacR footprints, the 271-bp DNA fragments were used (−186 to +85 bp) (Supplementary Table S3 in Supplementary Data). For reactions, the following protein concentrations were used: CWT or Cmut at 0, 264, 378, 422, 528 nM, and RacR at 0, 33, 231, 264 and 300 nM. The samples were incubated at 22°C for 20 min, followed by an addition of DNaseI (0.075 U; Eurx, Gdansk, Poland) and further incubation at 22°C for 4 min. Reactions were terminated by addition of EDTA to 25 mM, pH 8.0, and concentrated by vacuum evaporation. Next, the samples were resuspended in 40 μl of loading solution (30% formamide, 6 mM urea, 10 mM NaOH), denatured at 100°C for 2 min, and loaded (7 μl) on 6% urea, 8% acrylamide gels along with sequencing reactions (5′-Cy5-labeled primers) according to manufacturer’s recommendations (ddNTPs from Jena Bioscience, Jena, Germany and TERMIPol DNA Polymerase for efficient incorporation of ddNTPs from Solis BioDyne). K lanes indicate control reactions (DNA substrate with buffer and DNaseI), but with shorter incubation time than other samples, where time was dedicated to bind DNA. The Typhoon imaging system (GE Healthcare) was used for gel scanning and documentation.

**RNA isolation, determination of transcription start points and promoter activity**

*E. coli* MC1061 carrying plex3bracRlong (for racR) or plex3BrcRydaSinaktyracR (for ydaS) plasmids were cultivated to OD₆₀₀ = 0.4, and pellets from 15 ml of each bacterial culture were collected. These pellets were suspended in 100 μl of TE with lysozyme (0.1 mg/ml), subjected twice to immediate freezing at −80°C/thawing cycles, and 1 ml of TRI reagent (Merck) was added to ensure complete bacterial lysis. Next, 0.2 ml of chloroform was added, the tubes were vortexed, incubated for 5 min at RT and centrifuged (12 000 g, 10 min, 4°C). The top aqueous phase (0.6 ml) was
collected and total RNA was precipitated with 0.5 ml isopropanol (10 min RT), centrifuged (12 000 g, 10 min, 4°C), washed with 70% ethanol and dried. RNA was suspended in 30 μl of DEPC-water and its concentration was assessed spectrophotometrically.

Transcription start points of the racr and ydas genes were determined by primer extension (40). The mixture (12.5 μl) containing 5 pmol of appropriate Cy-5 labeled primer and 10 μg total RNA was incubated for 10 min at 75°C, and then 20 min at 58°C. In the next step, preheated buffer (final concentration: 50 mM Tris–HCl (pH 8.3), 50 mM KCl, 4 mM MgCl₂, 10 mM DDT), 1 mM (final concentration) of each dNTPs, 10 U of Ribolock RNase Inhibitor and 200 u of RevertAid H Minus Reverse Transcriptase (Fermentas) were added to a final volume of 20 μl, and the samples were incubated at 42°C for 1h, followed by RNA digestion with 0.25 U RNase H (EurX) (15 min, 37°C). Finally, the cDNA was precipitated overnight at −20°C with 300 μl of absolute ethanol and 40 μl of 3M sodium acetate (pH 5.5), centrifuged (12 000 g, 1h, 4°C), dried and suspended in 10 μl of water and 2 μl of loading buffer (80% formamide, 6 M urea, 10 mM NaOH). Sequencing reactions were performed in parallel on DNA templates using the same labeled primers. The samples were resolved on 8% acrylamide (acylamide:bis ratio 19:1) – 7 M urea 1×TBE gels (Sambrook1986) and visualized with Typhoon 9200 variable mode imager (Molecular Dynamics, USA).

Promoters activity or their inactivation by substitutions was confirmed by ONPG assay (40).

**In vitro transcription**

Single round *in vitro* transcription was performed on linear templates with the P_ydas_cPCR region spanning −184 to +113 bp, relative to the RacR translation start site. The expected transcript lengths were 81 nt, 111nt, 145 nt and 154 nt for the P_ydas_cPCRc, P_ydas_cPCRb and P_ydas_cPCRa promoters, respectively. The following conditions were used: 10 nM template, 30 nM RNAP holoenzyme (New England Biolabs), 75 mM potassium glutamate, 50 mM Tris-acetate (pH 8.0), 10 mM magnesium acetate, 10 mM β-ME, 0.01 mg/ml BSA, 100 μM ATP, CTP and GTP, 10 μM UTP (10 μCi/reaction [α-32P]UTP). The DNA template was pre-incubated with either RNAP or CWT/Cmut/RacR for 10 min at 37°C, which was followed by another 10 min incubation, but this time either CWT/Cmut/RacR or RNAP were added. Transcription was initiated by addition of NTPs and heparin (100 μg/ml, final). The reactions (final volume: 20 μl) were carried out for 10 min at 37°C, terminated by addition of an equal volume of stop solution (95% formamide, 20 mM EDTA, 0.05% bromphenol blue, and 0.05% xylene cyanol), run on 7 M urea 8% polyacrylamide sequencing gels in 1×TBE, and quantified by phosphorimaging with a Typhoon imaging system.

**Single-molecule tracking (SMT) and the C protein dynamics**

The C protein gene was fused to the mVenus fluorescent protein gene, and tested for use in SMT technology (Supplementary Figures S4 and S5 of Supplementary Materials).

The SMT setup we used is explained in (28,42). Briefly, the central part of a 514 nm laser beam was used for stream acquisition (20 ms integration time) of mVenus fusions, and movies were captured by a Hamamatsu ImageEM EM-CCD camera (128×128 pixel area of chip used). About 160 W cm⁻² were applied onto the image plane. Protein fusions were expressed at very low levels, such that the single molecule level was reached after few frames of bleaching of most molecules present in the cell, followed by single step bleaching of the fluorophores. Expressing very few molecules also avoids localization artefacts due to overproduction of the respective protein. The images were taken for an average of 8 time intervals, with about 10% of tracks being longer than 10 steps. Only tracks of 5 and more steps were included in the analyses. Data analyses were done using the SMTracker 2.0 program (43).

**RESULTS**

**Transcription factors and their effect on YdaST-RacR operon expression**

We showed previously that operon transfer of the Csp231I R-M system, carrying the C regulator as TF, heavily affected the expression of several genes within the rac cryptic prophage region. Dissection of detailed effects indicated that the C protein directly affected expression solely of the RacR repressor - the essential master regulator of the Rac prophage (18). To pursue the molecular basis of such cross-talk, here we first tested selected TFs related to the YdaST-RacR operon, as well as the C protein for comparison, to establish their potential activation/repression effects on this operon’s expression and their relationships in *in vivo* assays. Our experimental genetic system was based on two compatible plasmids, one of which produced a linear gradient of the selected TF in a controlled fashion, using the arabinose P_BAD promoter, while the second plasmid carried the lacZ reporter gene fused to the racr and ydas gene fragments preceded by the intergenic region (IGR) encompassing their promoters and putative TF target sites on both DNA strands (Figure 2). Based on genetic and structural analysis (our unpublished data and (36)), we predicted that YdAS is also a TF. Its autoregulatory properties and effect on racr expression could also be tested quantitatively in the above described assay. The results reveal, as expected, that WT RacR can inhibit its own transcription from the racr promoter, yielding a about a 3-fold decrease at higher TF concentration, as compared to the promoter activity in the absence of a functional RacR repressor (RacRmut) (Figure 2A). Interestingly, the profile of racr gene expression for WT RacR repressor and the unrelated C protein regulator are comparable, as the C protein is able to repress racr expression efficiently, to the same extent as the natural RacR repressor. Inactive C protein (Cmut – A33G R34E Q37A unable to bind its natural target site (19)) is also deficient in repressing the racr promoter. In addition, it seems that YdAS is not involved in RacR expression control, as a similar overall effect was observed on promoter activity over the gradient of TF concentrations for YdAS WT, as well as its inactive variant YdaSmut (K37A R40A) (Figure 2A).
The same set of TFs was used in an identical approach to test the ydaST expression from the IGR region oriented in reverse direction to the racR expression. Strikingly, the RacR TF was the most efficient repressor, able to almost completely inhibit ydaST expression at its highest concentration, whereas YdaS only reduced its own expression to 80% of its activity at the same concentration (0.001% arabinose) (Figure 2B). In contrast, increasing C protein concentrations had no effect on ydaST expression, indicating its indirect role in this operon’s regulation via the RacR repressor.

**C protein and RacR repressors bind to the racR region**

We next tested if the observed in vivo effects on transcription were due to direct C protein and RacR binding to the racR region, by conducting EMSA assays. As a substrate, we used DNA fragments corresponding to −186 to +85 bp with respect to the annotated racR translation start codon (A of ATG). The EMSA reactions were performed with the same amount of Cy5 end-labeled dsDNA (25 nM) and increasing concentrations of purified repressors. The natural target site for C protein is a C-box, which was established experimentally and by sequence comparison of C.Csp231I-like proteins (19). The C-box is formed by the two inverted repeats of CTAAG-n5-CTTAG sequence separated by an extended AT-rich spacer located in the C protein gene promoter (19,44,45). For the C protein assays with off-target sites, such as here in the rac region, we expected weak binding, and thus for control we also used its defective variant (Cmut-A33G, R34E, Q37A) with impaired helix 3 involved in specific DNA recognition, as tested previously (19). The shift in DNA-protein migration was observed only for the wild type C protein (CWT), but not Cmut. The retarded complexes for CWT appeared rather smeary, and at higher C protein concentration the DNA substrate band was invisible (Figure 3A). In contrast, the RacR repressor, which specifically interacted with its target site, showed distinct retardation complexes at low (≤50 nM) protein concentrations, indicating strong and specific binding (Figure 3B).

In addition, we also performed competitive EMSA reactions with one repressor pre-bound to DNA, in order to detect a possible interference in binding between these two repressors (Figure 3C). First, the C protein was added to DNA fragment to form a stable complex. Subsequent addition of increasing concentrations of the RacR repressor induced obvious higher-mobility complexes, indicating that C protein might not disturb RacR binding to its target site within the DNA substrate, nor does RacR repressor possibly displace the C protein (Figure 3C, left). In a similar way, RacR was first pre-bound to the same DNA substrate and later the C protein was added. Again, we observed that these two repressors formed visible, multiple low-mobility complexes, poorly resolved in the gel. In fact, the two lanes with 108 nM RacR and 1500 nM C-WT were indistinguishable, despite different binding orders. Thus, we can neither rule out that these two repressors compete for the same binding site, nor that their different stoichiometry impacts their DNA binding, at least under in vitro conditions tested here.

![Image](image_url)
Figure 3. The C protein and RacR bind to the rac region. The 271 bp substrate DNA covered the entire racR promoter/operator/ydaS intergenic region (IGR) and part of the racR coding sequence (positions −186 to +85 bp, in reference to the +1 translation start site of racR). DNA was amplified by PCR with one primer introducing the Cy5-label at the 5′ end. Each binding reaction was carried out with the same amount of DNA (25 nM) and increasing concentrations of the proteins: the C protein and its defective variant Cmut (0; 500; 650; 800; 1000; 1150; 1300; 1500; 1750 nM) (A) and RacR (0; 27; 54; 81; 108; 135; 162; 189; 243; 300 nM) (B). For competitive EMSA, the same substrate DNA was pre-bound with one protein, as indicated, followed by addition of the second protein under the same reaction conditions (C). The reactions were assembled and processed as outlined in Materials and Methods, resolved on 5% native polyacrylamide gels, and visualized by Typhoon scanner (GE Healthcare). Open and filled arrows denote positions of unbound DNA and shifted DNA–protein complexes, respectively.

Mapping the C protein off-target binding site by DNaseI footprinting

To reinforce the binding assay results, and to confirm that the reduction in RacR expression was due to direct C protein binding to the racR region, we performed binding site mapping assays using DNaseI footprinting. We used the same DNA fragment as a substrate that we used in EMSA studies. First, we searched the sequence to find a DNA site resembling the C protein consensus site CTAAG-n5-CTTAG with the first box being more conserved (shown as the LOGO in Figure 4A). Indeed, we found two exact CTAAG half-sites on the complementary DNA strand, which correspond to the coding sequence of racR gene, close to the start codon (named box1 and box2) (Figure 4A, top right). Analysis of the extended sequence of these boxes on the complementary strand indicated that box1 (5′-CTAAG-n5-CTTAA-3′; nearly inverted repeats) is much closer to the C protein consensus sequence than box2 (5′-CTAAG-n5-ACCAC-3′) (Figure 4A). Thus, we decided to focus on the two boxes 1 and 2 close to the racR gene start codon, due to our in silico search and some data obtained previously (18). We thus performed DNaseI footprinting assays using WT C protein and its Cmut variant. A distinct region of protection was observed from −10 to +15 bp in reference to the start codon of racR gene, which covered the inverted repeats of box1 (Figure 4BC, left, red bar). Such an area of protection was not detected when Cmut was used. To further verify that indeed the first CTTAG region within racR coding sequence (complementary 5′-CTAAG-n5-CTTAA-3′) is bound by the C protein, we prepared DNA substrates with altered box1 or box2 sequences (changing CTTAG into CGCAT in box1 or CTTAG in box2) and repeated DNaseI footprinting in the presence of the C protein (Figure 4C). We noticed sharp bands corresponding to DNaseI digestion products for the DNA fragment with CTTAG-box1 mutated, an area which had been protected in previous assays. However, this was not observed for the mutated CTTAG-box2; instead, a similar region seems to be protected as in the case of the wild type template, although the degree of protection is not as strong (Figure 4C).

In addition, we also performed DNaseI footprinting of C protein within the racR upstream region (IGR) to locate possible additional off-target sites there (Supplementary Figure S2). We did not see any clear area of protection comparable to the spot found for CTTAG-box1, however there was a characteristic pattern of hypersensitivity in the DNaseI footprinting indicating that C protein might bend the DNA strongly, unlike in the experiment with Cmut used. This region corresponds to an A-rich spot within the leader of the ydaS gene (5′-AAAAACACACA-3′) (Supplementary Figure S2).

The YdaS–RacR operon intergenic region contains multiple promoter sites with potential to be inhibited by TFs

To verify and further investigate the process of induction of cell toxicity via C protein interaction with the rac region, we decided to establish the molecular mechanism of gene
expression regulation at the intergenic region between ydaS and racR genes, and possible impact of the C protein as an external adventitious transcription factor. First, we located promoters within this region, identifying transcription start sites by primer extension on total RNA prepared from E. coli cells carrying the ydaS-racR operon. Putative positions of promoter sequences were identified based on comparison with E. coli consensus sequences: TATAAT (−10box) and TTGACA (−35box), with 17nt spacer (46). For the racR gene, we obtained several products, but we analyzed in detail only three of the most intense bands, in agreement with unpublished observations of others (34). We assigned probable −10 and −35 boxes for the racR promoters (PracR1; PracR2; PracR3), with the PracR3 promoter for the leaderless transcript being the strongest (as it seems from the primer extension experiment; Figure 5AC).

The positions of identified promoter sequences were confirmed by directed mutagenesis of −10 boxes followed by racR-lacZ gene expression assessment in the Δrac genetic context (Figure 5A). As a result, the leaderless PracR3 promoter seems to be essential for the racR promoter strength, but all promoters contribute to overall racR gene expression.

In contrast, the same approach identified only a single primer extension product for the ydaS transcript with exceptional nucleotide characteristics, and nearly perfect consensus sequences for −10 box (TAATAT) and a perfect one for −35box (TTGACA) (Figure 5BC). Substitution of the −10 box with the CCCGGG sequence completely abolished this promoter’s activity. We also determined that the ydaT gene does not have a separate promoter, as the cloned intergenic region between ydaS and ydaT does not yield any promoter activity in reporter assay (unpublished data). Thus, it seems that close-to-perfect promoter consensus sequence, P3yda, drives the bicistronic mRNA for the ydaS and ydaT genes.

In addition, we also asked how RacR repressor exerts its TF inhibiting property by binding to its promoters under in vitro conditions. Thus, we again conducted DNaseI footprinting using purified RacR repressor and the IGR between ydaS and racR as a substrate. Our data show large protection areas divided by short patches of DNaseI digested fragments, which may indicate several binding sites (Figure 5D, red bars on gel). Notably, gel analysis revealed that for the lowest RacR concentration, the distinct protection of sequence corresponded to the region of the ydaST

Figure 4. Mapping of the C protein off-target (Rac) site by DNaseI footprinting. (A) The genetic map of the ydaS-racR operon with two CTAAG/CTTAG boxes in the racR coding sequence (in blue font) is shown, as well as the LOGO of the C protein target site (C-box consists of two inverted repeats CTAAG-n5-CTTAG). 20 nM Cy5-labeled DNA fragments were incubated without or with increasing amounts of WT C protein or its inactive variant Cmut (0; 264; 378; 422; 528 nM) for 20 min at 22°C. The resulting complexes were briefly digested with DNaseI, and the products were run on 8% denaturing acrylamide gels along with sequencing reactions (GATC), as outlined in Materials and Methods. Panel (B) shows comparison of the protected area between the WT C protein and its defective Cmut variant for the same WT DNA fragments. Panel (C) focuses on the C protein interaction with WT and two mutated DNA substrates, where either CTTAG of box1 or of box2 is changed into CCGAT (box1mut or box2mut). Protected areas are indicated by red bars. K lanes represent control – DNA substrate subjected to DNaseI only.
promoter on the DNA complementary strand. At higher concentrations, a large region just upstream of the initiation codon of RacR was protected, where RacR promoters are located, mainly the PracR3 promoter. A summary of regulatory elements in the context of RacR repressor and C protein binding to the intergenic region is presented as Supplementary Figure S3.

We tried to determine a distinct sequence motif within the protected sequence (Figure 5D, red font in IGR sequence), which could be assigned as RacR binding site. We found several repeat sequences in this region (Figure 5D, nine blue boxes), which coincide with RacR protection areas. Five of them are located on the coding strand (in the 5′-TARGC orientation), whereas the other four are on the complementary strand. Interestingly, most of them are exactly covering the positions of all mapped promoters for the ydaS and racR genes, and the other two repeats are located within the racR coding sequence (Figure 5C, blue bars). We prepared positional alignment of these seven sequences to obtain a LOGO (47) and found a 5′-ATTAGGC-3′ motif, which is presumably the RacR monomer binding site (Figure 5D, LOGO). We also searched for the same motifs in extended sequences upstream of the racR gene present in several species of the Enterobacteriaceae family, and we found three highly conserved inverted repeats 5′-T/AGGCC-n3-GCCTA/3′ to be present in that region, which we assigned as putative RacR operators: O1, O2 and O3 (Figure 6A). The O1 position coincides with the ydaST promoter located on the complementary strand of DNA (Figure 5C), whereas the O3 position (its n3 spacer) overlaps with the initiating ATG and two more codons of the racR gene. In addition, there is another single 5′-TTAGGCC-3′ site further downstream in the racR coding sequence. Sequence comparison of these three inverted repeats by the LOGO software revealed a perfect core sequence in O1, but there is one nt difference in the other two operators, which may affect RacR binding affinity at these sites. The linker regions between these sites are A/T rich, which

Figure 5. Mapping of the transcription start sites within the racR and ydaS intergenic region, and mapping of the RacR binding sites. For each reaction, total RNA isolated from E. coli carrying racR-ydaS genes on a plasmid (Δrac background) was used as a template for the primer extension method using Cy5-labeled primers and was performed as indicated in Materials and Methods. The primer extension products (marked as +1) were resolved on a DNA/IGR track. mRNA start sites of different intensity are shown by arrows: PracR1, PracR2 and PracR3. (A) For the racR gene, three transcription start sites of different intensity are shown by arrows: PracR1, PracR2 and PracR3. (B) For the ydaS gene a single transcription start site was found, as indicated by an arrow. (C) The transcript start sites were analyzed in the context of upstream sequences and the appropriate O70 promoter sequences are indicated. The −10 and −35 promoter motifs are boxed (racR promoters) or in bold (ydaS promoter), respectively. The putative RacR repressor binding site motives are marked by blue bars on both DNA strands. (AB) The activity of identified promoters were tested in vivo in a Δrac genetic background. WT promoters and their mutants (−10 box substitutions) are as follows: PracR2 (WT AGTAAG→CCCCGGG), PracR3 (WT GAGAAT→CCCCGGG); PydaS (WT TAATAT→CCCCGGG). Vec abbreviates the promoterless vector. Transcription activity was measured as LacZ specific activity (Miller units) and error bars indicate the average (± SD) of at least three measurements. (D) Mapping of the RacR repressor binding sites using DNaseI footprinting with IGR DNA. Increasing amounts of RacR repressor (0, 33, 231, 264, 300 nM) were added to the DNA substrates and incubated for 20 min at 22°C, briefly digested by DNaseI, and the products were resolved on 8% denaturing acrylamide gel along with sequencing reactions (G,A,T,C), as outlined in Materials and Methods. Digestion protected areas are indicated by red bars. The IGR sequence (between the ydaS and racR genes, with start codons in bold) is shown with the protected three areas marked with red font. Putative nine RacR repressor binding sites were found and are boxed in blue; their sequence was used to prepare a LOGO as a possible RacR monomer binding site.
Figure 6. RacR repressor homologs and their putative operators within the racR and ydaS intergenic region. (A) Upstream regions of the racR gene, within several species of the Enterobacteriaceae family, revealed high identity sites, which coincide with inverted repeats, possibly RacR binding sites (operators: O1, O2 and O3), indicated by arrows. The positional nucleotide alignment is presented as a LOGO for three operators separately, showing slightly difference from established consensus sequences, as the inverted repeats separated by 3nt: 5′–TAGGC–3′ and 5′–TATAGG–3′. The translation initiation ATG codon for racR is present within O3, as marked. (B) Alignment of amino-acid sequences for RacR homologs using MultAlin software. Position of 8 helices is indicated, as well as the linker between two HTH domains. Species codes: E.coli (Escherichia coli), E.alb (Escherichia albertii), E.sp (Enterobacter sp.), E.mar (Escherichia marmotae), M.sp (Morganella sp.), C.am (Citrobacter amalonaticus), S.flex (Shigella flexneri), P.mir (Proteus mirabilis), K.qua (Klebsiella quasipneumonia), K.pne (Klebsiella pneumonia), S.mar (Serratia marcescens), S.ent (Salmonella enterica). (C) RacR repressor monomer structure from E.coli as predicted by AlphaFold software deposited at Uniprot P76062 database. The 8 helices marked with light and dark blue are predicted with highest model confidence. The 18-aa unstructured linker (low confidence) connects two protein domains: N-terminal α1–5 domain present in all RacR homologs and C-terminal α6–8 domain present only in some Enterobacteriaceae species, as indicated in protein alignment.
indicates their propensity towards DNA bending (Figure 6A, LOGO). In line with these observations, comparison of the RacR amino-acid sequences between the Enterobacteriaceae species revealed a high degree of identity. However, we noticed that the RacR orthologs found in the Escherichia, Morganella and Shigella genera consist of a much longer protein with 8 helices organized in a two domain architecture (Figure 6BC), where helix-turn-helix motives are connected by an 18-aa linker to the core protein fragment found among all species evaluated here, as predicted by AlphaFold at Uniprot P76062 database (Figure 6C). Notably, other RacR orthologs from Proteus, Salmonella, Serratia and Klebsiella are shorter, one-domain variants without an unstructured linker. In addition, we did not find any RacR ortholog in Citrobacter freundii/koseri, where the C regulatory gene was originally identified. This reinforces the notion that co-existence of the C protein and RacR repressor in one host may lead to possible viability defects (genetic conflict), which we proposed in our previous work (18).

In vitro transcription assays confirm direct RacR and C protein involvement in transcriptional regulation of the ydaST and racR region

Next, we performed in vitro transcription assays to corroborate our in vivo and in vitro findings obtained so far. Single round transcription assays were performed on linear templates with E. coli RNA polymerase in the presence and absence of RacR and the C protein. Two different conditions were tested – when RacR or CWT/Cmut were allowed to bind to DNA before or after RNA polymerase (RNAP).

We found, that P_{ydaS} is indeed a very strong promoter while, under in vitro conditions that were used, P_{racR1-3} promoters seem to be much weaker (Figure 7AB). Also, in accordance with our previous observations, RacR is a strong inhibitor of transcription initiating at the P_{ydaS}, as well as the P_{racR1} promoter (Figure 7BD). However, this effect is only observed when RacR is allowed to bind before (and not after) RNAP addition, indicating that this repressor probably acts by preventing RNAP binding to the P_{ydaS} and P_{racR1} promoter regions, thus inhibiting transcription initiation. Interestingly, RacR seems to activate transcription initiating from the P_{racR3} promoter when added after RNAP (with its maximum level reached at the lowest concentration tested, i.e. 50 nM), which may reflect a regulatory loop ensuring that appropriate RacR levels are always present in the cell.

In contrast, the C protein inhibits all P_{racR} promoters when added before RNAP (with full repression of P_{racR2} and P_{racR3}; P_{racR1} is repressed down to $\sim$10% of its full activity), and to a certain degree also the P_{ydaS} promoter (also down to $\sim$10% of its full activity). These effects are specific, as they were not observed for the Cmut variant (Figure 7AC). When RNAP was allowed to bind before the C protein, inhibition was not observed, which indicates that (similarly to RacR) the C protein-mediated repression is probably due to physical blocking of RNAP binding to DNA, possibly by inducing bending or distortions in the whole promoter region. Interestingly, under these conditions transcriptional activation is observed for the P_{racR1} and P_{racR2} promoters, which reflects complexity of this region and its regulation. We note that the effect of C protein requires higher protein concentrations than does the effect of RacR. This agrees with the EMSA experiments in Figure 3.

Single molecule tracking of C protein reveals considerable changes in mobility associated with the presence or absence of high affinity binding sites

It has recently been shown that mobility of DNA binding proteins is largely determined by non-specific DNA interactions, and that the size of the proteins has very little influence on their diffusion properties (48). The C protein binding investigated in this study provided a perfect system for determining how specific versus non-specific DNA binding affect a DNA-binding protein’s dynamics.

A typical TF’s action relies on a mixture of free diffusion and scanning of the cellular DNA via non-specific DNA binding; non-DNA-bound fractions of TFs can vary between about 1 and 40% in vivo (48,49). To reach a specific binding site, TFs diffuse in 3D and perform local motions such as 1D sliding, hopping or intersegmental transfer. We wished to gain further insights into the mode of action of C protein in vivo, in different genetic contexts, so we resolved protein dynamics at the single molecule level. We employed single molecule tracking (SMT), using YFP fusion proteins (42). Specifically, we expressed C protein-mVenus fusions (mVenus is a brighter variant of YFP; Supplementary Table S2 and Supplementary Figure S4 for corresponding western blot analysis) from a moderate-copy plasmid, using extremely low induction levels. Single fluorophores were obtained almost instantaneously, after few initial frames in which molecules bleached, and single molecule tracks could be captured with 20 ms stream acquisition. The C protein was fused to mVenus by a flexible linker – GSAGSAGSAGF – to allow for proper folding and achievement of the optimal biological activity of the fusion protein (50).

Different genetic contexts were designed to test the dynamics of C protein mobility: (a) the natural context of a single C-box target site in the $\sim$5 Mbp Citrobacter genome (which naturally lacks the Rac prophage, based on analysis of related sequenced Citrobacter genomes), (b) plasmid-borne C-box target sites in E. coli cells deleted for rac promoter binding sites (‘off-target site’), (c) the same as (b) but containing a genome-located off-target Rac site, (d) off-target sites within the genome of E. coli cells (but lacking C-box ‘target’ sites) and (e) E. coli cells devoid of both C-box and Rac sites, as presented in the schemes (Figure 8A). Whenever the target site is present on plasmids carrying the R-Ms system, the produced C regulatory protein is unable to bind a DNA (Cmut), so that only fluorescently labeled WT C protein is able to bind DNA (Supplementary Figure S5). Interestingly, one or two sites of strong confined motion (molecules not leaving a radius of 120 nm for at least 7 steps, red tracks, Figure 8B) were found in Citrobacter cells, while E. coli cells devoid of any binding sites (‘control’) largely contained freely diffusing molecules (blue tracks). Strong chromosomal binding can also be seen in the heat maps for Citrobacter cells, where the probability of finding a track is projected into an average sized-cell of $3 \times 1 \mu m$, while there
is almost completely homogeneous localization in the control strain.

These experiments suggest that C protein is strongly bound to the RM gene C-box locus, which is duplicated into two loci following DNA replication. Of note, any freely diffusing molecule can stay without motion for some time, so confinement will be somewhat convoluted with non-bound molecules, or molecules staying bound in a non-specific manner for some period of time. In Citrobacter cells, about half of molecules showed confined motion, and half free diffusion, while only a quarter to a third were confined in half of molecules showed confined motion, and half free diffusion, while only a quarter to a third were confined in

Figure 7. The effect of C protein and RacR on transcription initiating from the P<sub>PydaS</sub>-P<sub>racR</sub> promoter region depends on whether they are allowed to bind DNA before or after RNAP addition. (A, B) Sample gels; (C, D) quantified intensities of transcripts initiating from the corresponding promoters. Data points were normalized to no C/Cmut/RacR samples and reflect average values ± S.D. (n ≥ 3). Single round in vitro transcription was performed on linear templates, as described in the Materials and Methods section. (A, C) CWT or Cmut protein titration (0–4 μM) when RNAP was pre-bound to DNA before CWT/Cmut (RNAP→CWT/Cmut) or after CWT/Cmut addition (CWT/Cmut→RNAP). (B, D) The same as in A and C, but RacR was titrated (0–300 nM).

As a further means to characterize target DNA-binding binding versus off-target binding, we used squared displacement (SQD) analysis, in which the probability of a certain jump distance (JD) is plotted, and data are fit with Rayleigh distributions that explain a probability distribution for nonnegative-valued random variables. JDs observed for C protein could not be explained by a single Rayleigh fit, but quite well by two distributions (Supplementary Figure S7). Assuming three distributions (and thus three distinct populations) increased the quality of fitting only marginally: quantile/quantile plots in Supplementary Figure S7 show the difference between experimental data versus modelled data based on free Brownian diffusion, which shows very little deviation already for the two-population fit, but less deviation for three fits. In order to be able to distinguish between free protein diffusion, constrained motion due to non-specific DNA binding, and static motion based on specific DNA binding, we used three Rayleigh fits; results for two fits are shown in Supplementary Figure S8, which are qualitatively similar, but less discriminatory. SMTracker program searched for the best fit for average diffusion constants of populations for all 5 conditions, to better compare changes, because these are reflected in population size only, rather than in a combination of diffusions constants as well as population sizes.

SQD analyses revealed that in the Citrobacter native system, about 30% of C protein molecules showed a high, average diffusion constant of 1 μm<sup>2</sup> s<sup>−1</sup>, most compatible with free diffusion, and about 22% a > 50 fold lower value of 0.02 μm<sup>2</sup> s<sup>−1</sup> (Figure 8D and E). The low value can
Figure 8. Single molecule tracking of C protein::mVenus reveals considerable changes in mobility in the context of the target and off-target sites. (A) The fusion protein of C::mVenus (scheme) was introduced on a plasmid in the various cell genetic contexts, as presented on schemes: with target, natural sites upstream of C gene within Csp231I R-M system operon (green circles); with the off-target sites located within racR gene on genome (magenta circles). i) the natural context of a single target site in the Citrobacter genome coding for Csp231I R-M system; ii) plasmid-borne R-M systems carrying mutated C gene, but with WT target sites in E. coli cells, iii) genomic off-target sites as well as target plasmid-bornesites in E. coli cells; iv) only genomic off-target sites in E. coli cells, v) E. coli cells devoid of any binding sites. The red crossing on C gene indicates the presence of Cmut variant unable to bind its specific DNA site, so that only the C::mVenus protein can bind DNA. (B) Examples of trajectories in cells. Red tracks show confined motion, not leaving a radius of 120 nm (three times the localization error) for at least 6 steps, blue tracks are freely-diffusing molecules, green tracks molecules changing between confined and free diffusion. (C) Heat maps showing the probability of motion of the C protein. All tracks obtained were projected into a medium-size cell of 1×3 μm. To avoid asymmetry artefacts, heat maps were mirrored twice along the two central axes. Darker to yellow shading indicates higher to lower presence of molecules. (D) Bubbleplot showing size and diffusion coefficients of the three populations, determined from SQD analyses. (E) Summary of data obtained from SQD analyses, diffusion constants D1-3 correspond to populations: static, slow-mobile and fast-mobile.
only be explained by relatively tight binding to DNA target sites. 48% of molecules were found to diffuse with 0.13 \mu m^2 s^{-1}, which would fit to constrained motion through the nucleoids, based on fast on- and off binding to low-affinity sites. According to this analysis, more than 22% of C protein is bound to its genomic sites in *Citrobacter*, and close to half are moving through the genome in search of specific binding sites. The population of tightly DNA-bound molecules decreased to 20% in *E. coli* cells carrying target sites on plasmids, and similarly in cells with target and off-target sites (Figure 8D). Between these two systems, marginally more molecules showed medium-fast mobility, 43 versus 36%, in cells containing target sites but lacking off target sites, reflecting a higher degree of confined motion (Figure 8B). It was surprising to find that mobility of C protein was higher in cells with target and off-target sites than in cells with target sites only. Cells carrying off-target sites only had decreased slow and medium-fast mobility in favor of free diffusion, and still, 15% showed low mobility in the control cells, and 26% medium mobility. Possibly, there are additional, unknown high affinity binding sites for C protein other than target and off target sites in *E. coli*. Alternatively, even non-specific DNA binding may result in a significant number of TF molecules that remain static for an extended period of time.

In any event, the presence of a genomic off target site leads to a decrease in freely diffusing molecules, and an increase in medium – and slow-diffusing C protein. Of note, experiments were performed in rich medium, such that there are likely two to 4 off target sites on average in the growing culture (overlapping rounds of replication). Constrained motion (as derived from the medium mobile population) further increased by the addition of target sites on plasmids, and likewise slow/static motion. Thus, changes in binding-site availability lead to considerable, but not drastic changes in C protein mobility. Our data also suggest that specific binding sites on the chromosome (in this case the ‘off target’ sites) increase low protein mobility (specific DNA binding) to a similar extent as specific binding sites on several plasmids (C protein ‘target’ sites), or even more so in the *Citrobacter* cell. We propose that constrained motion through nucleoids based on non-specific DNA binding leads to efficient target search, while target sites on plasmid are less efficiently found, as most plasmids are excluded from the nucleoids (51).

**DISCUSSION**

**Transcriptional cross-talk**

Full knowledge of a complete bacterial transcriptional genetic network does not exist. While *E. coli* K-12 MG1655 has one of the most widely-studied genomes (52), even there about a third of its genes have not yet been characterized, including 50–80 TFs out of nearly 300 predicted TFs (17,53–55). *E. coli* K-12 TFs were classified into four groups based on the number of regulatory targets, which form a hierarchical structure resembling a ball of cobwebs with layers of different density. Usually the highest density nodes in the web are represented by global TFs with multiple target sites, in contrast to local and single-target TFs, which are a minority group (56,57). The other group consists of nucleoid-associated TFs engaged in organization and maintenance of nucleoid structure, which heavily affects gene expression, and some TFs play intermediate or dual roles in this respect (58). Altogether in concert, they control over 1000 genes, however it is hard to dissect a single TF’s in vivo effect due to the whole repertoire of co-existing regulators (14,59,60).

*In vitro* studies are also biased, as mapping the TF’s targets usually requires TF purification and tagging, and target determination is more likely to be established for strong binders, though new technologies have been developed to tackle this problem, as reported recently (61). Thus, it is still difficult to study the TF’s weak interactions and to discriminate between direct and indirect targets. Global systematic analyses are usually focused on direct targets and secondary interactions are ignored, and thus implication of the off-target actions seems to be underestimated. Moreover, TF/DNA interactions depend on multiple cellular parameters, such as the nucleoid environment, and in particular, the cooperativity with other DNA-binding proteins within the regulatory networks (62,63).

In the course of our previous work, we noticed that upon *Citrobacter*-derived TF introduction (as an R-M system operon) into a new genetic species (*E. coli*), the host genetic network was changed, which affected cellular fitness (18). Here, we aimed to characterize the molecular basis of this interesting phenomenon, as well as to map the cross-talk. We have established earlier via transcriptomics that the primary, direct causative effect is due to silencing of the RacR TF by the action of the C regulatory protein, which is an unrelated TF linked to the RM system. The C proteins are small, cis-acting local TFs working as autorepressors and/or autorepressors dedicated to control the level and timing of the restriction endonuclease action so that the DNA nucleolytic activity is not unleashed until genomic DNA is completely modified by a cognate DNA methyltransferase (28–30,64). The consensus binding site of the studied C regulator is relatively simple (CTAAG-n5-CCTAG) and reflects its dimeric nature (44). Naturally occurs within the C gene promoter (19). We have confirmed that this exact recognition site does not exist in the *E. coli* MG1655 genome. [This is not unexpected, as the 10-bp sequence should occur once in > 10^6 bp in a random sequence.]

It has been suggested that the transcriptional cross-talk is more likely to occur for TFs recognizing shorter DNA sequences or partially degenerate ones, where there is a high chance that many less specific sites are present at genome (65). In addition, target sites also face constant genetic pressure to adjust TF binding via point mutations, which drives regulatory network evolvability (66). Nevertheless, we expected to find C protein binding to a similar sequence, or to its half consensus sequence in the proper DNA context, as the C proteins also recognize phosphates of the non-contacted nucleotides in a process of indirect read-out, in a manner similar to some phage repressors (45,67,68).

Guided by the RNA-seq data we obtained earlier (18), as well as the sequence analysis, we determined that the C protein induces transcriptional cross-talk within the Rac region, where master regulator RacR was adventitiously prevented from binding to its target site, leading to perturbation in RacR regulation. Here, by *in vitro* and *in vivo* as-
saying we mapped the C regulator interaction region within the rac gene. We determined the C protein binding site as box1 (5′-CTAGAG)n-5′-CCTAA-3′; only 1 nt apart from consensus) located within discriminator of the rac promoter P_{Rac3} and extending downstream into the translation start codon of the rac gene (Figure 9, Supplementary Figure S3). The cross-talk spot located within rac transcription initiation site, and the tendency of C proteins to strongly bend DNA (69–71), may both affect rac repression, which we confirmed by an in vivo reporter assay (Figure 2). In addition, the C protein interaction did not directly affect ydaS expression in vivo, and no binding site was found near the ydaS promoter, within upper IGR sequence. In contrast, out in vitro transcription experiments indicate that although P_{ydas} is fully inhibited by RacR, it is also inhibited by the C protein, albeit not fully (transcript level is down to ~10%) (Figure 7). On the other hand, in vitro transcription experiments did confirm C protein inhibition of transcription initiating from the rac promoter. Perhaps discrepancies between in vivo and in vitro data result from involvement of an additional factor (missing in vitro) or presence of a DNA fragment (missing in the DNA template used in vitro) influencing interactions between the two factors and DNA.

We predict the C protein dimer rather than a monomer to interact with its off-target site, which is in accordance with the DNaseI protection region (Figure 4BC). Interestingly, it seems that the C protein and RacR binding site in operator 3 overlap, 5′-ACCTGGCTTAATTTGC-3′ (C protein consensus site underlined, RacR consensus site in bold) (Figure 9). Still, we do not know how other upstream RacR binding sites within the IGR region (operator 1 and 2, Figure 6) contribute to the autoinhibitory effect. This could be so because our in vitro binding tests show that pre-binding of one TF does not prevent binding of the other TF, at least using the WT DNA substrates where all recognition sites are present (Figure 3). Also, our multiround in vitro transcription data indicates that when both proteins (C and RacR) are present together, transcription initiation at P_{ydas} is no longer fully repressed by RacR, allowing for a low level (5–8%) of transcripts to initiate (Supplementary Figure S6). These results may reflect multi-layered interactions between the C and RacR proteins and the region surrounding their cross-talk spot, awaiting further verification.

Theoretical models have tried to outline possible scenarios for cross-talk interactions and to test whether gene misregulation can affect adaptive gene expression (72,73). In fact, the dynamics of TFs interaction creates a balance between maintaining binding specificity and allowing some relaxation in DNA interaction to enable adaptation of networks. If newly re-wired TF links are beneficial, they are likely to facilitate network evolution (13). Sometimes, as in our case, the cross-talk triggers toxic gene activity, which is silenced under normal conditions (18). In a similar example, another C regulatory protein linked to the Type I R-M system of Mycoplasma, showed cross-talk activity via binding to secondary sites within promoters for a protease gene and tRNA gene cluster leading to cell death (74). Overall, studies show that adding a new network connection into an existing network can result in a large gene network perturbation; the higher connections in the TF hierarchy that are re-wired, the larger the perturbation (12,75). Thus, as horizontal gene transfer events involving TF genes happen constantly, there is a genetic pressure on the host regulatory circuitry to keep it stable and to prevent any regulatory perturbation in accepting host cell (11,76–81).

**Single molecule dynamics of a TF in different genetic backgrounds**

The finding that the C protein binds with high affinity to both the promoter region of the R-M system as well as to the rac gene motivated us to study changes in single molecule dynamics in the presence or absence of these binding sites. In the absence of any known high affinity binding sites in E. coli, C protein shows predominantly free diffusion, however, we found 15% of molecules to be essentially static, likely corresponding to strong binding to DNA sequences. Whether there are additional, unknown high affinity sites is unknown. About a quarter of molecules showed dynamics that can be best described as constrained diffusion through the nucleoids, characteristic of any DNA-binding protein. The presence of the ‘off-target’ Rac site in the E. coli genome led to a modest but detectable increase in the static fraction, and interestingly, also an increase in slow motion, i.e. an increased number of molecules showing constrained motion. Thus, specific binding sites appear to not only increase static binding, but also target search via constrained diffusion. The presence of target sites on plasmids (between 15–20 based on plasmid copy number) lead to a considerable increase in static motion, as expected, as well as in medium-mobility, strongly decreasing the number of freely diffusing molecules. This is in line with our finding that specific binding sites on the genome increase general target search via constrained motion. Surprisingly, the presence of both, target and off target sites did not strongly change C protein dynamics compared with cells having target-sites only, but in fact somewhat increased protein mobility. We cannot rule out that this effect is secondary, caused by cell filamentation due to induction of the ydaST genes. The strongest effect on chromosome-based motion...
was seen in *Citrobacter* cells, where we found the highest number of static as well as medium mobile molecules. Compared with the *E. coli* control cells having 60% freely diffusing molecules, only 30% showed free diffusion (based on SQD analyses), and 70% low mobility in *Citrobacter* cells. Thus, genomic target sites have a stronger effect on DNA-based mobility of C protein when compared with plasmid-borne target sites in *E. coli* cells. These analyses show that few high affinity binding sites have a strong effect on TF mobility, and suggest that target search via constrained movement through the genome has a higher efficiency for finding target sites than for target sites on nucleoid-excluded plasmids (assuming that the shift of C protein abundance towards the cell periphery in the presence of plasmids observed in this study reflects the general rule of nucleoid occlusion for medium/high copy number plasmids).

**Novel insights into the ydaS-racR region with RacR as the master regulator**

The Rac prophage was the first defective lambdoid prophage discovered in *E. coli* K-12 genome, and has lost 60% of its original DNA content. Most *E. coli* sequenced genomes have at least part of the rac locus (34,82). The rac region comprises only 29 genes, although some of them have not yet been characterized and their genetic relationships have not been determined (83). It seems that the RacR repressor is a master regulator of the Rac expression network, which together with neighboring genes ydaS and ydaT form the operon in the *E. coli* genome, featured as essential (32–34). The Keio collection of *E. coli* single gene deletion mutants does not contain a ΔracR mutant (84), consistent with the fact that such mutation is possible only in ΔydaST background. In the course of our work, we found very close analogy of the IGR region located between the racR and ydaST genes to the lysis/lysogeny decision region between the CI and Cro genes of lambda bacteriophage (35), as it was also noted by others (34,85). We confirm here that RacR shares functional similarities to the CI repressor, as well as YdaS resembling λCro protein promoting the transcriptional switch into λ phage lytic cycle. In addition, RacR transcripts initiating from P$_{racR}$ are leaderless, similar to the λCI transcript (86). YdaS acts as an autorepressor, but it seems that RacR inhibits ydaS expression more strongly, even though racR promoters appear to be weaker than the single ydaS promoter (Figures 2, 5 and 7). Efficient ydaS inhibition was seen at very low RacR concentration (both, in *vitro* and in *vivo* – Figures 2 and 7), which may indicate that RacR’s affinity is greater for operator 1 within promoter sequence of ydaS (Figure 6A), similar to CI/Cro operator (O$_{R}$) binding (35). The strong ydaS promoter also controls expression of YdaT, as we did not detect any promoter activity upstream of ydaT gene (not shown). Interestingly, the ydaS promoter activity in *vivo* is barely detectable (even in the Δrac strain), in contrast to our in *vivo* transcription data, where we see a very strong ydaS promoter signal. Thus, more cellular factors could be involved in ydaS promoter activity regulation. We tested some global TFs, like HNS and FIS and did not find any effect on the ydaS promoter (not shown), but some other TFs or other regulatory elements might be engaged.

Based on the analogy to the intergenic region of λCI and λCro, as well as other observations indicating possible three binding sites for RacR repressor (34), we used footprinting methods to test their exact location (Figure 5). Indeed, we found multiple binding sites, such as three inverted repeats, which slightly differ among themselves (Figure 6A), just like the three λ operators O$_{R}$ for CI and Cro (35). It seems likely that RacR binding also relies on cooperative interactions with its operators, just like λCI, as well as on positive and negative feedback loops involving presence of the YdaS anti-repressor. However, unlike the CI/Cro system, YdaS does not seem to repress racR expression (while λCro inhibits CI) (Figure 2), which suggests that even though there are promoter and operator organization similarities with the λ system, the racR/ydaS region regulation is distinct from that of λ phage. In addition, the racR operator 3 site overlaps the coding sequence of the racR gene.

Recent studies on the *E. coli* CP-933P lambdoid cryptic prophage, which shares similarity with the Rac prophage (including racR-ydaST operon), also suggest an analogy to the λ genetic switch (36). PaaR2 seems to play a role of RacR TF, with an additional role of regulating a downstream toxin-antitoxin system, not present in the Rac prophage. There, PaaR2 binding sites are present in the intergenic region between paaR2 and ydaS genes, forming four inverted repeats that do not overlap the coding sequence. Interestingly, the YdaT-like protein of prophage CP-933P does not have any toxic effects (36). This is unlike the ydaT gene of Rac (18,33,34), which could be cloned solely under expression silencing conditions in certain *E. coli* genetic contexts (not shown). We have shown previously that Rac YdaT directly leads to cell filamentation, which does not result from induction of the SOS response (18). Thus again, Rac prophage regulation and the gene functions it encodes are distinct from other systems described so far, and require further studies to reveal its peculiarities.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We thank Drs Thomas K. Wood and Yunxue Guo for the gift of strains/plasmids. We also acknowledge the excellent help of Marian Sektas and Bob Blumenthal at the stage of critical reading our manuscript.

**FUNDING**

National Science Centre (Poland) [UMO-2019/35/B/NZ2/00701 to I.M.]; SMT studies were funded by a grant from the Deutsche Forschungsgemeinschaft and the state of Hessen (Research consortium MOSLA) (to P.L.G.). Funding for open access charge: National Science Centre (Poland) [UMO-2019/35/B/NZ2/00701].

Conflict of interest statement. None declared.

**REFERENCES**

1. Martínez-Antonio, A., Janga, S.C. and Thieffry, D. (2008) Functional organisation of *Escherichia coli* transcriptional regulatory network. *J. Mol. Biol.*, 381, 238–247.
van Hjum,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, 81–509.

Seshasayee,A.S., Sivaraman,K. and Luscombe,N.M. (2011) An overview of prokaryotic transcription factors: a summary of function and occurrence in bacterial genomes. Subcell. Biochem., 52, 7–23.

Todeschini,A.L., Georges,A. and Veitia,R.A. (2014) Transcription factors: specific DNA binding and specific gene regulation. Trends Genet., 30, 211–219.

Chen,J., Boyaci,H. and Campbell,E.A. (2021) Diverse and unified mechanisms of transcription initiation in bacteria. Nat. Rev. Microbiol., 19, 554–566.

Brown,D.F. and Busby,S.J. (2004) The propagation of perturbations in rewired bacterial gene networks. Biophys. J., 86, 4648–4657.

Ishihama,A. (2010) Prokaryotic genome regulation: multifactor mechanisms of transcription initiation. Curr. Opin. Microbiol., 13, 554–564.

Nug-Staron,A., Tomasek,K., Caruso Carter,C., Sonnleitner,E., Kavčič,B., Paixão,T. and Guet,C.C. (2021) Local genetic context shapes the function of a gene regulatory network. Elife, 10, e65993.

Rowland,M.A., Abdelzaher,A., Ghosh,P. and Mayo,M.L. (2017) The propagation of perturbations in rewired bacterial gene networks. Nat. Commun., 8, 10105.

Browning,D.F. and Busby,S.J. (2004) The regulation of bacterial transcription initiation. Nat. Rev. Microbiol., 2, 57–65.

Andersson,S.G., Zomorodipour,A., Andersson,J.O., Ives,C.L., Sohail,A. and Brooks,J.E. (1995) The regulatory C protein for production of EcoO109I restriction endonuclease, a protein for production of EcoO109I restriction endonuclease, associated with biological regulation of a restriction-modification gene expression after entry into a new host cell. Nucleic Acids Res., 33, 5622–5632.

Kavčiˇc,B., Paix˜ao,T. and Guet,C.C. (2021) Local genetic context shapes the function of a gene regulatory network. Elife, 10, e65993.

Jaeger,S.A., Chan,E.T., Metzler,G., Vedenko,A., Chen,X. and Kneale,G. and Severinov,K. (2009) Transcription regulation of the EcoRV restriction-modification system. Nucleic Acids Res., 37, 5622–5632.

Crosstalk and the dynamical modularity of feed-forward loops in transcriptional regulatory networks. Trends Genet., 25, 439–440.

Shen,T., Wang,J.J., Chen,K., Yurkovich,J.T., Cho,B.K. and Palsson,B.O. (2021) CRISPR-Cas-mediated gene silencing reveals RacR to be a negative regulator of YdaS and YdaT toxins in Escherichia coli. Nucleic Acids Research, 2022, Vol.50, No.19.
