Nucleic Acids Research, 2025, **53**, gkaf113
https://doi.org/10.1093/nar/gkaf113
Gene regulation, Chromatin and Epigenetics

A transcription factor from the cryptic *Escherichia coli* Rac prophage controls both phage and host operons

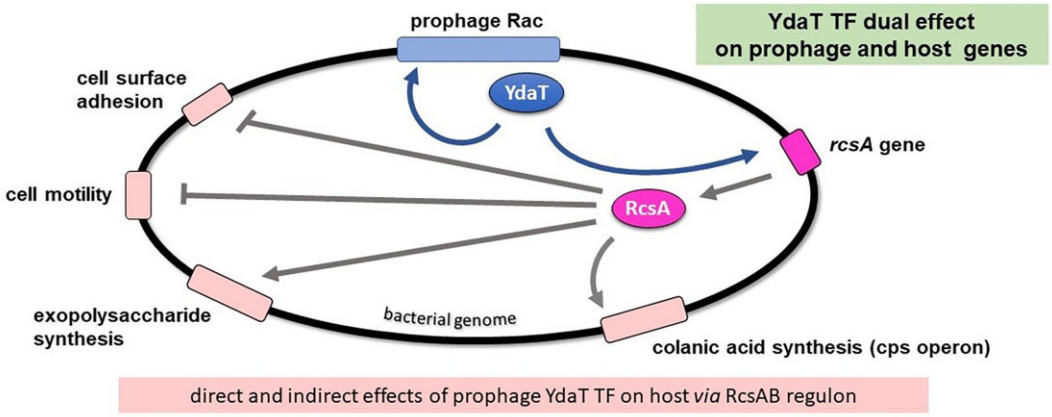
Ewa Wons[1], Katarzyna Gucwa[1], Natalia Lewandowska[1], Aleksandra Wisniewska[1], Lukasz Pawel Kozlowski[2], Iwona Mruk[1,*]

[1]Department of Microbiology, Faculty of Biology, University of Gdansk, Wita Stwosza 59, Gdansk 80-308, Poland
[2]Institute of Informatics, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, 02-097 Warsaw, Poland
[*]To whom correspondence should be addressed. Email: shamrock127@hotmail.com

Abstract

Bacterial genomes are shaped by cryptic prophages, which are viral genomes integrated into the bacterial chromosome. *Escherichia coli* genomes have 10 prophages on average. Though usually inactive, prophage genes can profoundly impact host cell physiology. Among the phage genes in the *E. coli* chromosome, there are several putative transcription factors (TFs). These prophage TFs are predicted to control only phage promoters; however, their regulatory functions are not well characterized. The cohabitation of prophages and bacteria has led to conditions under which the majority of prophage genes are unexpressed, at least under normal growth conditions. We characterized a Rac prophage TF, YdaT, expression of which is normally inhibited by Rac TFs and, surprisingly, by the host global regulator OxyR. YdaT, when expressed, leads to a toxic phenotype manifested by drastic cell filamentation and cell death. We determined the binding sites and regulatory action for YdaT, finding two sites within the Rac locus, and one upstream of the host *rcsA* gene, which codes for the global regulator RcsA. The resulting increase in RcsA strongly impacts the bacterial RcsA/B regulon, which includes operons related to motility, capsule biosynthesis, colanic acid production, biofilm formation, and cell division. Our results provide novel insights into the host's genetic network, which appears to integrate YdaT in a complex manner, favoring its maintenance in the silenced state. The fact that the potentially toxic YdaT locus remains unmutated suggests its importance and potential benefits for the host, which may appear under stress conditions that are not yet known.

Graphical abstract



Introduction

Bacterial genetic networks include regulatory scaffolding, fundamental for coordination and maintenance of bacterial physiology and communication with the dynamic environment. The main components of the genetic networks are DNA-binding transcription factors (TFs), which modulate gene expression within operons and regulons. They can quickly adjust gene expression levels as needed. Approximately 300 TFs were detected or predicted in *Escherichia coli* K-12 substr. MG1655 genome. However, still some fraction of TFs are unknown and not all TFs' functions are comprehensively characterized, including determination of their DNA interaction sites [1–4]. Some of the predicted TFs belong to a set encoded by cryptic prophages and are not well characterized [5]. Cryptic or defective prophages are segments of DNA, which have remained in a host's genome after domestication of active lysogenic

Received: November 13, 2024. **Revised:** January 17, 2025. **Editorial Decision:** January 31, 2025. **Accepted:** February 5, 2025

phages, which has usually resulted from their loss of genes necessary to excise from the bacterial chromosome and to produce phage progeny [6, 7]. The well-studied bacterium *E. coli* K-12 has ∼10 cryptic prophages, representing almost 4% of the host genome [8, 9]. Some of them, like DLP12, e14, Rac, CPZ-55, and Qin, are fairly well characterized [6, 10]. Some other *E. coli* strains, like *E. coli* O157:H7 str. Sakai, has even bigger number of prophages (up to 18) that can constitute ∼16% of the total genome [11]. Many such cryptic prophages contain TFs of unknown function, many of which are not even predicted or properly annotated. However, the collective data obtained so far on known prophage TFs suggest that a great majority of them regulates the expression of their own prophage's genes [12, 13]. Surprisingly, only in rare cases have impacts of prophage TFs on host global gene expression been characterized, like an AppY from prophage DLP12, which affects >200 genes of *E. coli* K-12. In DLP12, some genes were directly regulated by AppY, including these involved in the acid stress response and biofilm formation. Some other genes were indirectly affected resulting in strong motility defect [14]. It seems clear that the number of prophage TFs having strong effects on the host may be underestimated, as most prophage genes are unexpressed under normal conditions [15, 16], as exemplified by DLP12 prophage [17]. There is growing body of evidence that cryptic phages modulate cell growth, resistance to antibiotics, stress responses, and biofilm formation [9, 18]. Prophage genes allow the host to survive harsh conditions and to adapt to the new ecological niches, which is crucial for microbial evolution [19]. It seems that understanding the action of single cryptic prophage operons is not enough, without the broader context of the genetic relationships between phages and their hosts. It is very likely that long-term persistence of phage TFs inside of their host genomes might have selected for binding site adaptation so as to establish new regulatory links.

One of the cryptic prophage unknown TFs is YdaT, whose gene is located in the immunity region of Rac in *E. coli* K-12 [10]. Despite Rac prophage being highly conserved in the *E. coli* genomes [10], most of its 29 genes have not yet been characterized, with some exceptions like: active type I toxin–antitoxin system, RalRA [20] or KilR, a cell division inhibitor [9]. Rac prophage genetic relationships have not been determined [21]. YdaT gene was computationally predicted, and initially annotated as a putative bacterial toxin that is neutralized by the putative antitoxin YdaS [22]. In contrast, we have shown that both YdaT and YdaS are Rac prophage TFs, involved in the Rac lysis/lysogeny decision [23–25]. In similar cryptic CP-933P phage, in *E. coli* O157:H7 strain EDL933, regulatory genes analogous to *racR*, *ydaS*, and *ydaT* were found and compared to the λ phage repressors: CI, Cro, and CII, respectively [26] (Fig. 1A).

Of note, all three Rac genes (*racR*, *ydaS*, and *ydaT*) are considered to be essential in *E. coli* [21, 27–29]. The Keio collection of *E. coli* single gene deletions mutants does not contain *ΔracR*, and that deletion could only be obtained in *ΔydaST* background. YdaT gene expression is undetectable under normal growth conditions. When YdaT is produced, there is strong cellular filamentation leading to cell death, if prolonged (Fig. 1A) [23, 24]. Moreover, *ydaT* expression due to insufficient RacR repressor level also results in Rac DNA excision, and the excised Rac DNA is rapidly lost from cells due to its defect in replication. Thus, the lethal YdaT effect is counteracted by Rac prophage induction. We have shown how *E. coli* rewires its regulatory network, so as to minimize the adverse regulatory effects of adventitiously unleashed YdaT [25]. This suggest possible YdaT engagement in some fundamental host processes, though this remains to be tested experimentally.

So far, the exact role of YdaT in *E. coli* cell physiology remains elusive, thus our objective was to characterize YdaT regulatory interactions with its host's genetic networks and biochemical pathways. Our results indicate that YdaT has expanded beyond its local phage regulatory function, impacting the host RcsAB regulon. The fact that *E. coli* has maintained the active *ydaT* gene, while losing some other Rac genes, might indicate an important YdaT function within a specialized, but yet undetermined pathway.

Materials and methods

Bacterial strains, plasmids, and oligonucleotides

The bacterial strains and plasmids used in this study are listed in Supplementary Table S1 of Supplementary Data. Oligonucleotides used are listed in Supplementary Table S2.

Reporter assays

The tested promoters or genes with their native promoter/operator sequences were cloned as fusions to a *lacZ* or red fluorescence (mKate) reporter gene (Supplementary Table S1). When indicated, a second compatible plasmid was introduced - a pBAD33 derivate carrying the *ydaT* gene or other tested TF gene, cloned after an arabinose-inducible P_{araBAD} promoter. All experiments were performed in a Δrac background, to eliminate expression of the Rac repressors. Briefly, fresh transformants were inoculated into LB supplemented with appropriate antibiotics, and when the cultures reached mid-log phase (OD_{600} = 0.5–0.6), ONPG (ortho-nitrophenyl-β-galactoside) hydrolysis was measured and Miller units were calculated. When YdaT variant or OxyR effect was tested, the cells were incubated to early log phase (OD_{600} = 0.1–0.2), then the expression of TF was induced with varied concentration of arabinose, and cultures were further incubated for 60 min, followed by ONPG test. To measure fluorescence of *racR* fused to *mKate* the bacterial pellet was collected and resuspended in physiological saline (0.85%) to (OD_{600} = 0.3–0.6), and the relative red fluorescence (mKate2 gene) was measured as the red intensity (emission at 633 nm with an excitation at 588 nm) using a 96-well plate reader (EnSpire Multimode; Perkin Elmer) divided by the bacteria cells optical density. Statistical tests were conducted, as described in figure legends. Details on plasmid cloning and features are outlined in the Supplementary Data (Supplementary Table S1).

Cell viability analysis

Overnight cultures of *E. coli* MG1655 *Δrac* carrying pBAD33 or pBAD33-ydaT were subcultured in LB supplemented with chloramphenicol (34 μg/ml) until early log phase (OD_{600} 0.1–0.2), followed by induction of gene expression with 0.1% L-arabinose. Control cultures remained uninduced (actually repressed, treated with glucose). Bacteria were incubated with shaking for an additional 3 h, samples were taken at the indicated times and, after serial dilution, spotted onto LB-agar

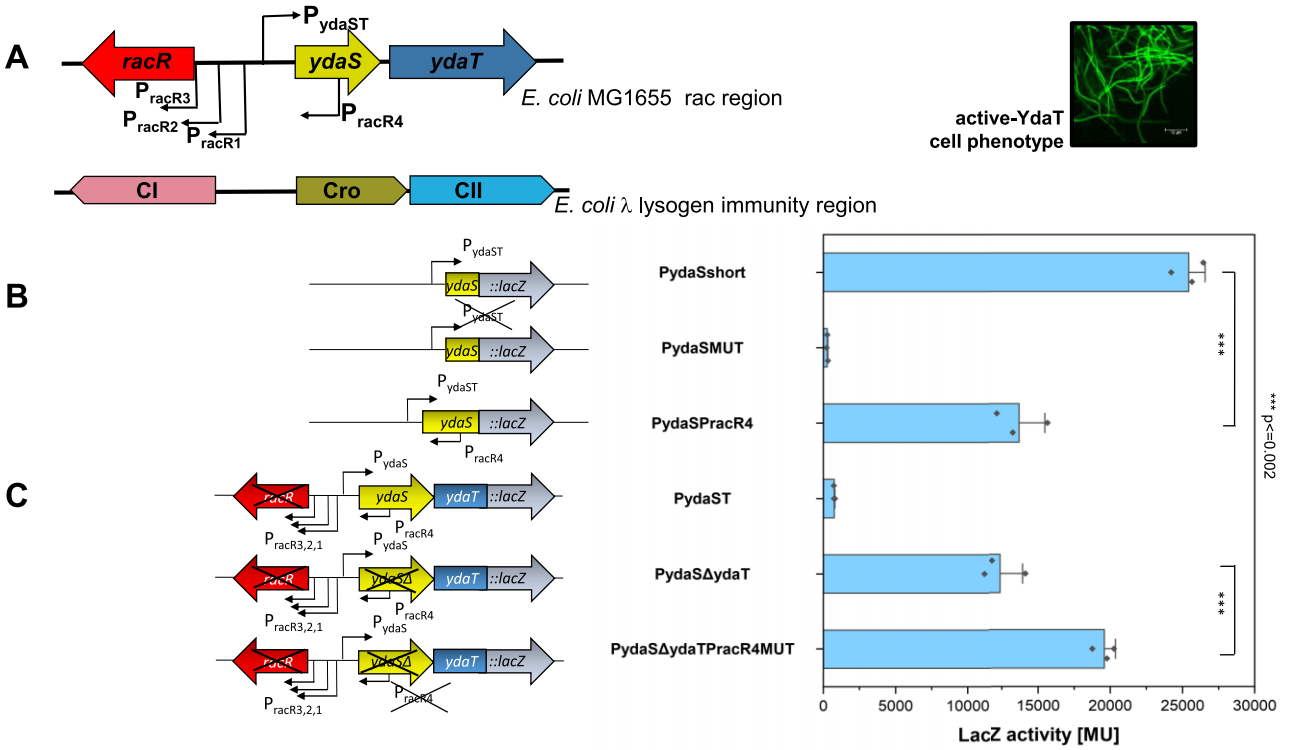


Figure 1. Gene organization of three Rac prophage TFs: RacR, YdaS, and YdaT and its analogy to λ bacteriophage with CI, Cro, and CII proteins (not in scale). Upper right: representative *E. coli* MG1655 single cells phenotype under induction of *ydaT* expression, stained with SYTO 9 (green fluorescence). (**A**) The experimentally verified promoters are indicated. (**B**) Quantitative LacZ assay shows the activity of P_{ydaST}, a shared promoter (−35box TTGACA and −10box TAATAT) driving expression of *ydaS* and *ydaT* genes (pLexydaSshort), its mutated variant (−35 box TTGACA into CCCGGG; pLexydaSshortMUT), and the negative effect of reverse promoter P_{rac4} activity located within *ydaS* coding sequence on the complementary DNA strand (pLexydaSPracR4). (**C**) The expression of *ydaT* gene depends on its operon background under presence of inactive RacR and inactive YdaS (pLexydaST versus pLexydaSΔydaT), as well as knocked-down reverse promoter (P_{racR4} promoter −35 box TTGACG into CCCGGG; pLexydaSΔPracR4MUT). The error bars indicate the averages (±SD) of three independent measurements. Statistical significance calculations were performed using a two-tailed unpaired Student's *t*-test (*** *P*-value ≤ 0.002).

(LA) containing antibiotics and arabinose inducer. The same samples were used for microscopic analysis and for the flow cytometry to visualize live and dead cells using LIVE/DEAD BacLight™ Bacterial Viability Kit (Thermo Fisher, Invitrogen) that includes a mixture of SYTO 9 and propidium iodide (PI) fluorescent dyes. The green fluorescent dye SYTO 9 permeates both intact and damaged membranes of the cells, binding to nucleic acids and fluorescing green when excited by a 485 nm wavelength laser. Conversely, the red fluorescent dye PI alone enters only cells with significant membrane damage, which are considered to be nonviable, and binds with higher affinity to nucleic acids than SYTO 9. DAPI (4′,6-diamidino-2-phenylindole) was also used to stain DNA in viable cells. Bacterial suspensions were stained according to the manufacturer's protocol and imaged using a Leica DMI4000B microscope equipped with a DFC365FX camera (Leica).

For the flow cytometry, a dead cell control was prepared by treating *E. coli* MG1655 Δrac pBAD33 with 70% ethanol, followed by centrifugation and resuspension in 0.85% NaCl to match the other samples and minimize background fluorescence. Data were acquired using an Amnis FlowSight Imaging Flow Cytometer (Luminex) with a 488 nm laser. Laser power was adjusted to avoid signal saturation in the sample with the highest fluorescence intensity. The IDEAS 6.2 software was used to analyze the data, allowing for the gating of single cells to prevent the counting of fluorescence from aggregates. This was particularly important due to the increased slime production and aggregation propensity of *ydaT*-induced cells. IDEAS 6.2 also enabled the creation of a dividing line based on a dead-only control (ethanol-killed cells), differentiating between dead cells (PI fluorescence > 1000) and live cells (PI fluorescence < 1000). Percentage evaluations were also performed using IDEAS 6.2. Raw data are accessible at RePOD repository: https://doi.org/10.18150/1BZXHU.

TEM (transmission electron microscopy) and SEM (scanning electron microscopy) were employed to investigate structural changes in cell morphology and ultrastructure subsequent the *ydaT* induction. For ultrastructure (TEM), the collected bacteria samples were fixed overnight in 2.5% glutaraldehyde (Agar) and postfixed in 1% osmium tetroxide (Agar), then gradually dehydrated in ethanol and embedded in EPON (Agar). Samples were cut (65 nm) on ultramicrotome (Leica UC7), stained with Uranyless and Reynold's lead citrate (Delta Microscopies). Analyses were performed using Tecnai Spirit BioTWIN TEM (FEI Company, Eindhoven, Netherlands) at 120 kV. For SEM, samples were fixed overnight in 2.5% glutaraldehyde (Agar) and postfixed in 1% osmium tetroxide (Agar), then gradually dehydrated in ethanol. Bacteria were mounted onto SEM stubs, air dried, and coated with gold using a spi-module sputter coater prior SEM observation. The samples were examined and photographed using a Prisma E Thermo SEM (Fisher Scientific).

YdaT overexpression, purification, and size-exclusion chromatography

The N-terminally 6× His-tagged YdaT variants were purified as follows. Plasmids carrying *ydaT* gene variants were pLATE51 derivatives (Thermo Scientific), where the *ydaT* gene is cloned under control of the P_{T7} promoter, with T7 RNA polymerase induced by IPTG (isopropyl β-D-1-thiogalactopyranoside). Plasmid p51ydaT carried wild-type YdaT, while YdaTLA and YdaTmut variants carried mutations in *ydaT* gene affecting (respectively) two oligomerization residues (L107P and A110P) or two resides responsible for DNA binding (H56A R57A). Freshly transformed *E. coli* ER2566 carrying p51ydaT, p51ydaTLA, or p51ydaTHR were used to inoculate 10 ml of LB supplemented with ampicillin (100 μg/ml) and incubated overnight following dilution (1:50) in 200 ml LB supplemented with ampicillin. These were incubated until they reached early log phase. Then *ydaT* expression was induced with 1 mM IPTG, cultures were incubated for an additional 3 h, and the cell pellets were collected and stored at −80°C.

Frozen cells were thawed in NPi10 buffer (50 mM K/PO_4, pH 8.0; 0.5 M NaCl; 10 mM imidazole; 10% glycerol) supplemented with 0.25 mM phenylmethylsulfonyl fluoride and sonicated (40 × 10 s). The lysates were cleared by centrifugation and applied to a column packed with TALON cobalt resin (Clontech), and washed with NPi10 buffer and NPi20 (as NPi10, but with 20 mM imidazole). Proteins were eluted with 200 mM imidazole in NPi10 buffer; protein-containing fractions were pooled and dialyzed overnight either against DE buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA), for electrophoretic mobility shift assay (EMSA) and footprint analysis, or against SEC (size-exclusion chromatography) buffer (50 mM K/PO_4, pH 7.0; 150 mM NaCl; 5% glycerol) for determination of oligomerization status. For that purpose, SEC on a Superdex 75 10/300 GL column using the ÄKTA Pure 25 system (GE Healthcare) were performed. The predicted size of 6His-tagged YdaT monomer is 18.81 kDa. The 1 mg proteins sample was loaded onto the column equilibrated with 50 mM potassium phosphate buffer (pH 7.0), 150 mM NaCl, and eluted at a flow rate of 0.5 ml/min in the same buffer. The column was calibrated with proteins of known molecular masses: alcohol dehydrogenase, 146.8 kDa; bovine serum albumin, 66 kDa; ovalbumin, 43 kDa; trypsin inhibitor, 22 kDa; and cytochrome C, 12.4 kDa (Sigma–Aldrich, St. Louis, MO, USA).

In vitro binding assays: electrophoretic mobility shift assay and spectral shifts

For both assays, DNA substrates were polymerase chain reaction (PCR)-amplified using fluorescently Cy5-labeled primer at one end (Supplementary Table S2).

For EMSA, reactions containing 20 nM DNA and the indicated purified protein concentrations were prepared in a binding buffer [10 mM Tris–HCl (pH 8.0), 50 mM NaCl, 10 mM $MgCl_2$, and 1 μg of poly(dI-dC) as a nonspecific competitor] in a final volume of 20 μl, and incubated for 20 min at 22°C. Rac DNA substrate within intergenic region (IGR, 271 bp) covered region −186 to +85 bp, where +1 refers to *racR* translational start site. Its negative control was an irrelevant sequence, a 262 bp long from *bla* gene. DNA substrates to test $YdaT_{Rac}$ and $YdaT_{CP\text{-}933P}$ cross-reactivity covered the equivalent region of Rac and CP-933P prophages, spanning from the *ydaS* gene end to the start of the *ydaT* coding sequence (206 bp). The *rcsA* substrate covered the *rcsA* gene region from −390 to + 51 (441 bp) in relation to the *rcsA* translation start. 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).

The spectral shifts (SpS) procedure was performed using Monolith X (NanoTemper Technologies, Germany) [30]. SpS quantifies the strength of the interaction between fluorescently labeled DNA and protein in a rapid and sensitive manner. The detection of SpS is achieved through dual-wave length detector at 650 nm and 670 nm in isothermal environment. The reactions contained 5 nM Cy5-labeled DNA (Supplementary Table S2) and increasing concentration of protein over a dilution series (for YdaTWT in a range from 1.1×10^{-10} to 3.59×10^{-6} M; for nonbinding YdaTmut in a range from 5.95×10^{-10} to 1.95×10^{-5} M) in binding buffer [10 mM Tris–HCl (pH 8.0), 50 mM NaCl, 10 mM $MgCl_2$, and 1 μg of poly(dI-dC) as a nonspecific competitor] in final volume of 15 μl. The standard MST (microscale thermophoresis) glass capillaries were used in the auto-detected channel with automatic power (mostly 100%) at 25°C temperature. The dissociation constant (K_D), representing binding affinity, was determined by plotting a fluorescence intensity ratio (670 nm/650 nm) against the increasing concentration of proteins. The data were processed with MO.Control 2 software v.2.5.4 (NanoTemper Technologies, Germany). All measurements were performed in triplicates.

DNaseI footprinting assay

Protein–DNA binding reactions were performed with 40 nM DNA in 20 μl of the following buffer: 10 mM Tris–HCl (pH 8.0), 10 mM $MgCl_2$, and 50 mM NaCl. DNA substrates were PCR-amplified using fluorescently Cy5-labeled primer at one end (Supplementary Table S2). For footprints three DNA fragments were used. First one was to determine the YdaT-binding site within the *racR–ydaS* IGR, from −105 to +60 bp, relative to translation start for *ydaS* (165 bp). The second one was to map a binding site upstream of *ydaT* coding sequence, region from −136 to +70 bp, relative to translation start for *ydaT* (206 bp). Third, DNA substrate covered sequence upstream of *rcsA* gene from −390 to +51 bp, relative to translation start for *rcsA* (441 bp). For EMSA reactions, the following protein concentrations were used: YdaTWT and YdaTmut at 0, 360, 720, 1440, and 2880 nM. The samples were incubated at 22°C for 20 min, followed by an addition of DNaseI (0.075U; Eurx, Gdansk, Poland) and further incubation at 22°C for 4 min. Reactions were terminated by adding EDTA to 25 mM, pH 8.0, and concentration by vacuum evaporation. Next, the samples were resuspended in 20 μl of loading solution (30% formamide, 6M urea, and 10 mM NaOH), denatured at 100°C for 2 min, and loaded (7 μl) on 6M 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). The Typhoon 9200 variable mode imager (Molecular Dynamics, USA) was used for gel scanning.

Motility, biofilm, and membrane sensitivity assays

E. coli MG1655 Δrac strains, transformed either with the pBAD33-ydaT or pBAD33 vector as a control, were cultured in LB medium supplemented with chloramphenicol until they reached mid-log phase (ca. OD_{600} 0.2), when expression of genes under P_{BAD} promoter was induced with 0.1% arabinose. Two hours after induction 1 μl of cell culture was injected in the center of thick LB soft agar plates (0.3% agar) enriched with chloramphenicol and 0.1% arabinose to ensure that the agar is punctured halfway to prevent the bacterial suspension from escaping to the surface or coming into contact with the bottom of the plate. Additionally, the cells were applied at a precise right angle to ensure uniform cell diffusion. Plates were incubated 20 h at 32°C. After that time, motility was determined as the diameter of the zone migrated by bacteria inoculated into plate center. Each growth spot was measured twice by diameter (mm) and the average was taken.

For biofilm assay crystal violet staining was applied. Freshly transformed *E. coli* MG1655 Δrac strains harboring pBAD33 or pBAD33-ydaT were grown on LB agar plates supplemented with chloramphenicol and 0.1% arabinose for 48 h at 37°C. Bacterial suspensions were prepared from both transformants in LB broth containing 0.1% arabinose to the OD_{600} of 0.08. Two hundred microliters of each bacterial suspension and a blank control (LB broth only) were transferred to individual wells of a 96-well polystyrene microtiter plates (Sarstedt, Germany) and were cultured for 48 h at 37°C. The plates were then washed twice with distilled water to remove planktonic cells. The attached biofilm was stained using 125 μl of 0.1% crystal violet solution for 30 min and washed three times with distilled water to remove the excess stain. The biofilm-bound dye was resolubilized by adding 200 μl of 33% acetic acid to each well. Absorbance at 595 nm was measured after 20 min incubation at room temperature using the Tecan plate reader.

For membranes sensitivity assay, the same strain was induced with 0.1% arabinose, serially diluted and spotted on LB agar supplemented with 0.1% SDS (sodium dodecyl sulfate) and 0.8 mM EDTA or on MacConkey medium supplemented with chloramphenicol and glucose/arabinose, as indicated. Plates were incubated 20 h at 37°C.

DNA affinity purification assay for the identification of DNA-binding proteins

The assay utilized the natural affinity of streptavidin to biotin and was performed to analyze the interaction between tested DNA and cellular proteins [31]. Two DNA regions, one encompassing the locus of interest (165 bp) and another serving as a negative control (172 bp, promoter region of Csp231I R-M system [32]) were generated by PCR using primer pairs BIOydaSpdfor and ydaSpdrev; and Fbiot1 and pdnrev, respectively (Supplementary Table S2). One primer of each pair was biotinylated. The reaction products were visualized by agarose gel electrophoresis and purified with Clean up kit (A&A Biotechnology). Approximately 5 μg of DNA was immobilized to 1 mg of Dynabeads M280 streptavidin (Invitrogen) in buffer A (DNA binding and washing buffer: 50 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 1 M NaCl) according to the manufacturer's instructions. The cell extract was prepared from the pellet of 25 ml log phase culture of MG1655 by sonication in buffer (50 mM Tris–HCl, pH 8.0, 100 mM NaCl, and 0.5 mM EDTA). Dynabeads were prewashed with buffer B (protein binding buffer: 20 mM Tris, pH 8, 1 mM EDTA 10% glycerol, 1 mM DTT (dithiothreitol), 100 mM NaCl, and 0.05% Triton X100). The cell extract was also enriched with poly(dI-dC) competitor (Thermo Scientific), which binds unspecific proteins. The cell extract was loaded four times in 300 μl aliquots (20 min shaking with 500 RPM followed by magnet separation). The magnetic particles were washed four times with buffer B (20 min, 500 RPM). The final fraction was collected to assess unbound proteins. Bound proteins were eluted with 60 μl of buffer C (elution buffer, 20 mM Tris pH 8, 1 mM EDTA, 10% glycerol, 1 mM DTT, 1 M NaCl, and 0.05% Triton X100).

LC-MS/MS analysis

Samples for LC-MS/MS (liquid chromatography-mass spectrometry) analysis were prepared according to a standard protocol FASP [33] on a 10 kDa membrane. Obtained peptide fractions were subjected for a final clean-up on C18 (exchange disks 3M EmporeTM) StageTips following the described protocol [34]. LC-MS/MS analysis was performed on a hybrid Triple-TOF 5600+ mass spectrometer with DuoSpray Ion Source (AB Sciex LLC, Framingham, MA, USA) coupled with Ekspert MicroLC 200 Plus System (Eksigent, Dublin, CA, USA). The system was controlled by SCIEX Analyst TF 1.7.1 software. The chromatographic gradient for the analysis was 11–42.5% B (solvent A: 0% aqueous solution 0.1% formic acid; solvent B: 100% acetonitrile, 0.1% formic acid) with a flow rate of 10 μl/min for 60 min. Chromatographic separation was carried out on a ChromXP C18CL column (3 μm, 120 Å, 150 × 0.3 mm; Eksigent). The data-dependent acquisition (DDA) analyses comprise precursor spectra accumulation in 100 ms in the range of 400–1200 m/z, followed by top 20 candidate ions per scan in 50 ms in the range of 100–1800 m/z resulting in a total cycle time of 1.15 s. Experiment was performed in a looped product ion and high sensitivity modes.

LC-MS/MS spectra were processed using PEAKS Studio version 11 (Bioinformatics Solutions Inc., Waterloo, ON, Canada) with the settings previously described [35]. The data were analyzed against the *E. coli* K-12 database (Uniprot, January 14, 2024). The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [36] partner repository with the dataset identifier PXD059580.

Bioinformatics analysis and molecular modeling

The 3D structures of YdaT and its homologs were predicted utilizing a range of tools: AlphaFold3 [37], Chai-1 [38], AlphaFold2 [39], ESM [40], OmegaFold [41], RGN2 [42], Phyre2 [43], HHsearch [44], and Modeller [45]. Comparative analysis, structural alignment, and modeling part were conducted using UCSF Chimera [46]. Putative oligomers and DNA–protein interactions were modeled using AlphaFold3 [37] and Chai-1 [38] comparison with DNA-containing templates from homologous structures (e.g. λCII repressor) in the PDB, identified by FoldSeek [47] and DALI web server [48]. The files related to the bioinformatics protein modeling are accessible at RePOD repository: https://doi.org/10.18150/1BZXHU.

Results

The *racR–ydaS–ydaT* operon regulatory insights

Organization of the *racR–ydaST* operon, along with its promoters is shown in Fig. 1A. We previously demonstrated positions of the experimentally identified promoters for *racR* gene ($P_{racR1-3}$) and the RacR repressor multiple binding sites (four operators) within the IGR, as well as a single promoter for *ydaS* gene [24]. In the present work, we aimed to dissect the activity of the P_{ydaST} promoter in its natural context, which is silent under normal conditions, in spite of the exceptional, nearly perfect consensus sequence for −10 box (TAATAT) and a perfect one for −35box (TTGACA), and perfect 18 bp spacing between the two [49]. Such sequence characteristics suggests strong expression from P_{ydaST} promoter, which was confirmed previously by *in vitro* transcription [24]. We wondered whether *ydaT* expression *in vivo* also depends on the apparently strong P_{ydaST} promoter. We first tested whether P_{ydaST} drives production of the bicistronic mRNA for *ydaS–ydaT* genes. We used reverse-transcription PCR (RT-PCR) with primers encompassing the *ydaS–ydaT* region. PCR products were generated for each primer set located within the *ydaS* and *ydaT* genes, clearly indicating that a bicistronic transcript is made (Supplementary Fig. S1). We also determined that *ydaT* does not have a separate promoter, as the cloned IGR between *ydaS* and *ydaT* does not yield any promoter activity in reporter assay.

Next, we also noticed that there is a reverse-oriented promoter within the *ydaS* coding sequence, as predicted by promoter finder (BProm; www.softberry.com), and also indicated in EcoCyc (www.ecocyc.org). Based on the consensus sequence for *E. coli* σ^{70} RNA polymerase promoters, the deduced −10 and −35 boxes of the reverse P_{racR4} are TACGCT (consensus TATAAT) and TTGACG (consensus TTGACA), respectively. Two convergent promoters being 107 bp apart (from the respective transcription starts) usually results in negative effect on transcription [50]. We examined this by construction of *ydaS*'::*lacZ* fusions with the separate P_{ydaST} promoter extracted out of its operon, its knocked-down variant (−35 box TTGACA into CCCGGG) and *ydaS* fusion variant with extended coding sequence containing the reverse promoter (schemes at Fig. 1B). The *in vivo* data show that indeed P_{ydaST} WT drives high gene expression (up to 25 000 Miller units), typical for exceptionally strong promoters, as predicted by the *in silico* analysis. The reverse P_{racR4} promoter action apparently inhibits transcription from the P_{ydaST} by 2-fold (Fig. 1B). We also wondered if the reverse P_{racR4} promoter affects *racR* expression, being far from but in the same orientations as promoters $P_{racR1-3}$. However it seems, at least under our tested conditions, that there were effects on gene expression, so the P_{racR4} promoter activity is linked to *ydaS/ydaT* transcription, but not to that of the *racR* gene (Supplementary Fig. S2).

Next, we used the same approach to test gene expression, but in the native genetic context of the *racR–ydaS–ydaT* operon. All plasmid fusion constructs had *racR* gene inactivated (restriction site Klenow filled), so as not to interfere with P_{ydaST}-dependent transcription, as we showed previously that RacR binds upstream of *ydaS*, within IGR [24] (Fig. 1B). In the *racR*-negative context, the P_{ydaST} promoter is quite weak, however inactivating the *ydaS* (deletion of 10 codons downstream of P_{racR4} reverse promoter), increases transcription from P_{ydaST} by 16-fold (Fig. 1C). Additional knock-down of the reverse promoter (−35 box TTGACG into CCCGGG) resulted in a further increase of gene expression up to 20 000 Miller units. Altogether, we demonstrated that at least the TFs RacR and YdaS together with the reverse promoter within *ydaS* collectively play a negative role in transcription initiation from the P_{ydaST} promoter, as measured by *ydaT* gene expression.

Two separate YdaT-binding sites within the Rac locus

In order to demonstrate specific binding of YdaT, and to identify its DNA targets, we performed *in vitro* gel shift assay (EMSA) to detect complex formation between purified YdaT and selected DNA fragments. For this purpose, a set of Cy5-labeled targets were selected, where finally two separate DNA fragments showed the specific binding.

First, we used PCR-amplified DNA substrate containing the IGR between *racR* and *ydaS* genes and another, irrelevant, similar-size DNA fragment as a negative control (Fig. 2A). The assay clearly exhibited the YdaT dose-dependent formation of single shifted complexes. Binding was not observed with a non-specific probe targeting the DNA fragment that codes for the *bla* gene (Fig. 2A).

To identify YdaT-binding sites within its bound DNA fragment, we next performed DNaseI footprinting analysis, with YdaT WT and its nonbinding variant YdaTmut (H56A R57A). We found a 100 bp long DNA segment, which was protected from DNaseI treatment, unlike the same substrate with YdaTmut protein. The protected DNA was located in the range −60 bp to +40 bp where +1 refers to A of the start codon of the *ydaS* coding sequence (Fig. 2B). We observed two regions: one 20 bp long, more distinct region (red bar, Fig. 2B) and the other much wider region, more downstream of *ydaS* coding sequence, where YdaT binding may be weaker (blue bar, Fig. 2B). The protected areas correspond to the location of promoter for *ydaS* (completely overlapping the −35 and −10 box of P_{ydaST}), as well as ribosome binding sites and several codons. We could not distinguish any palindrome or evident sequence motif within protected region. However, we attempted to assign the binding sites as three operators based on data obtained in following text sections below (Fig. 2C).

For the second DNA substrate that had yielded a positive gel shift assay, we also performed DNaseI footprints analysis, in this case close to the start of *ydaT* coding sequence. A distinct region of DNA protection was observed, overlapping the translation initiation codon for *ydaT* (Fig. 3A). Within the protected region, this time we identified the inverted repeats 5′-ATGATTCATGAAAATCAA-3′, which are nearly perfect and spaced by 6 nt. The ATG in the spacer is the ATG initiation codon of *ydaT*. Another report, on YdaT homolog from prophage CP-933P in the genome of *E. coli* O157:H7, showed the binding site as being 5′-TTGATT-N_6-AATCAA-3′, which is only 1 nt different at the first position of the first repeat [51]. We named this binding site as "YdaT-box," unlike YdaT operators within IGR.

Given the similarity of the apparent binding sites, we sought to test whether YdaT of Rac binds the CP-933P prophage substrate and vice versa, and we were not surprised to see that the cross-binding did occur (Fig. 3B and C). We used an *in vitro* approach to test cross-reactivity of purified $YdaT_{Rac}$ with different prophage DNA substrates (Fig. 3B) and *in vivo* assays to confirm that both YdaT homologs can activate the

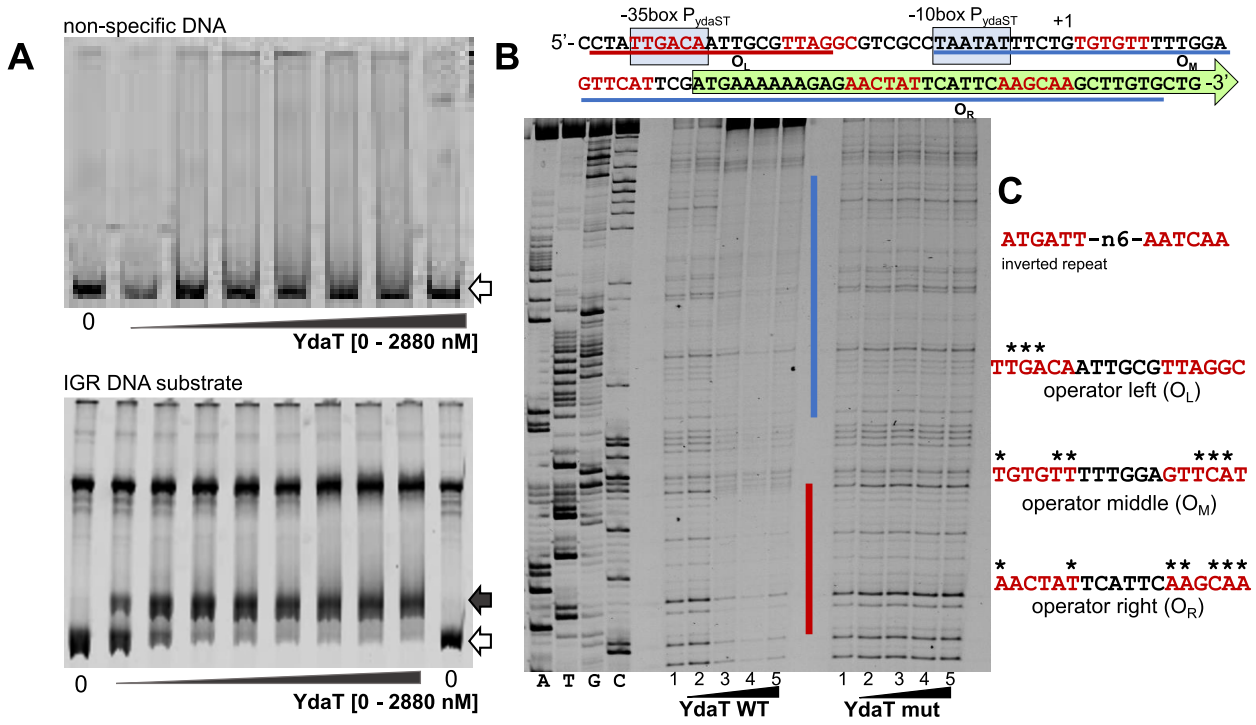


Figure 2. Mapping the YdaT binding sites within Rac prophage *racR–ydaS* IGR. (**A**) YdaT binds specifically to the 271 bp *rac* region spanning IGR and part of coding sequence of *ydaS* and *racR*. As a negative control, DNA substrate of a similar size was used (DNA coding *bla* gene). DNA substrates were 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 (20 nM) and increasing concentrations of YdaT, as indicated. "0" refers to lane control with no protein. Open and filled arrows denote positions of unbound DNA and shifted DNA–protein complexes, respectively. (**B**) Determination of the YdaT-binding sites by DNaseI footprinting. The reactions contained 40 nM Cy5-labeled DNA fragment (IGR substrate) and increasing amounts of purified YdaT WT protein or its inactive variant YdaTmut [H56A R57A] (lanes 1–5 correspond to protein concentration: 0, 360, 720, 1440, and 2880 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 (ATGC), as outlined in "Materials and methods" section. The protected areas, resistant to DNaseI digestion, are indicated by red and blue bars, and the corresponding DNA sequence is presented above the gel picture, including part of the *ydaS* coding sequence (in green) and elements of P_{ydaST} promoter (blue boxes). (**C**) Protected sequence was analyzed and three 18 bp operators (O_L; O_M; O_R) were distinguished (binding sites in red font spaced by 6 bp in black) based on data presented in Fig. 3A. The inverted repeat 5′-ATGATT-n6-AATCAA-3′ served as a reference sequence (Fig. 3A) to assign the similar sequences as operators within indicated protected regions (identities marked by asterisks).

YdaT-dependent promoter to the same extent (Fig. 3C). Compared prophage-derivative YdaT homologs share only 30% amino acid identity, though with some highly conserved regions (Fig. 3D).

YdaT binding within the intergenic region and its effect on *ydaS* and *racR* expression

To test the effect of YdaT binding to its three operators, located between *racR* and *ydaS*, but in the absence of the YdaT-box within *ydaT*, we used plasmids where both genes are fused to a reporter *lacZ* gene. In the second plasmid, the *ydaT* gene was cloned under the control of inducible P_{BAD} promoter, where YdaT gradient is correlated with arabinose concentration added to the culture medium.

First, we tested the level of *racR* expression, when the YdaT binding occurred in the IGR when YdaS and RacR are absent. Clearly, even a very low level of YdaT decreases *racR* expression (Fig. 4A). However, when the *lacZ* reporter reflects expression of the other side of the IGR (*ydaS*), we observe no effect of YdaT binding. The shorter or longer genetic context did not change the results (Fig. 4B and C). Overall, it seems the role of the YdaT binding within the IGR is rather to control the level of RacR, and not of YdaS.

YdaT binding to the YdaT-box activates *racR* expression from a distance, via a reverse promoter present within *ydaS*

To explore the potential regulatory role of the YdaT binding site (YdaT-box) within *ydaT* (as opposed to the one in the IGR), we prepared plasmids to test the activity of the reverse P_{racR4} promoter by fusing it to a *lacZ* gene with an inserted SD (Shine-Dalgarno) sequence. In this plasmid, the YdaT-box is localized at its natural distance, ~190 bp upstream from the tested reverse P_{racR4} promoter (Fig. 5A). In addition, neither *ydaS* nor *ydaT* is active, and their shared promoter is absent. We also prepared a variant of this plasmid as control, where YdaT-box is mutated (YdaT-boxWT = ATGATTCATGAAAATCAA changed to YdaT-boxmut = CCCGGGTTCATGGCGCGC) *via* site-directed mutagenesis. Neither of these constructs showed any change in promoter activity (Fig. 5B). However, when YdaT expression was induced in *trans* by arabinose, only the reverse promoter with WT YdaT binding sites drove transcription strongly (up to a 29-fold increase) (Fig. 5B). This result clearly indicates a dependence of the high reverse promoter activity on the YdaT interaction with its YdaT-box.

We wondered if such great stimulation of the reverse promoter could affect nearby convergent promoter driving the

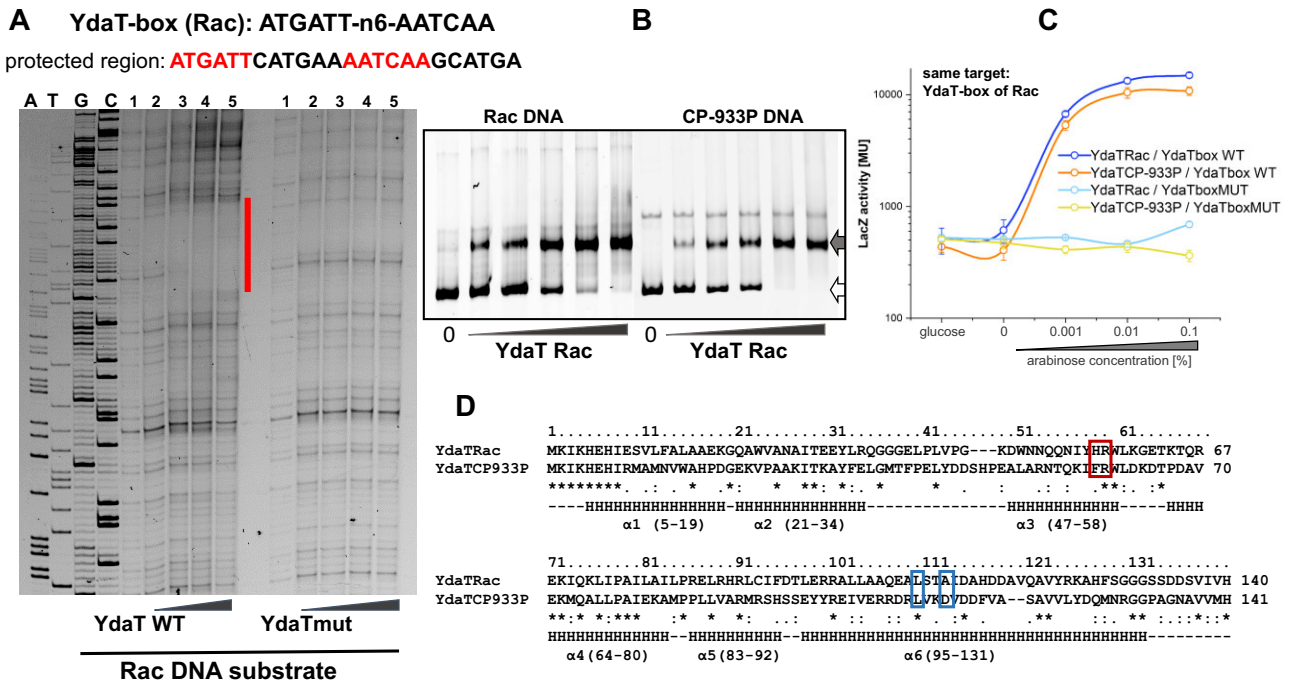


Figure 3. Identification of the YdaT binding site and YdaT homologs cross-reactivity. (**A**) Mapping YdaT binding sites within *ydaT* upstream region of Rac locus (assigned here as YdaT-box) by DNaseI footprinting. The reactions with purified YdaTWT and YdaTmut were composed and treated with DNaseI as described at Fig. 2. Protected area is indicated by a bar and corresponding DNA sequence is presented above the gel picture. The recognition sequence, an inverted repeat, assigned as YdaT-box is shown. (**B**) YdaT of Rac can cross-interact with two DNA substrates, natural and of CP-933P prophage, using *in vitro* assay. The Cy5-labeled DNA substrate contained the recognition site, YdaT-box of Rac and of CP-933P prophage. Each binding reaction was carried out with the same amount of DNA (20 nM) and increasing concentrations of $YdaT_{Rac}$ (0, 90, 180, 360, 720, and 1440 nM). "0" refers to lane control with no protein. Open and filled arrows denote positions of unbound DNA and shifted DNA–protein complexes, respectively. (**C**) The YdaT homologs from two cryptic prophages of *E. coli* (Rac and CP-933P) were tested *in vivo* for ability to cross-activate YdaT-box-dependent promoter. The same plasmid with reporter LacZ gene carrying YdaT-box of Rac (WT or mutated) (pLexPracR4YdaTbox versus pLexPracR4YdaTboxMUT) was used to measure the promoter activity following induced expression of YdaT from either Rac or CP-933P (pBAD33-ydaT or pBAD-ydaTCP, respectively). (**D**) Alignment of amino acid residues of YdaT homologs from *E. coli* using CLUSTAL Omega (1.2.4): $YdaT_{Rac}$ (Uniprot P76064) and $YdaT_{CP933P}$ (Uniprot Q8 × 4Y2). The conserved residues H56 and R57, located within the α-helix (α3) and crucial for DNA binding are boxed. In addition, residues L107 and A110, marked in boxes, were mutated to prolines to disrupt the long α-helix (α6) that serves as an oligomerization scaffold. The secondary structure of YdaT is also shown.

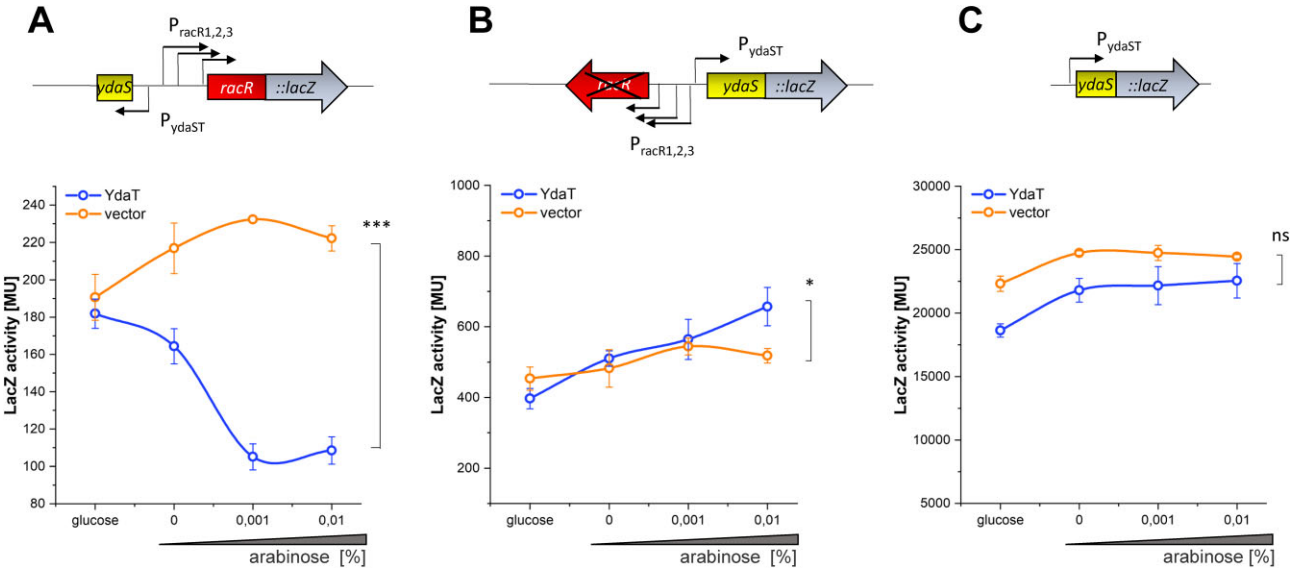


Figure 4. YdaT binding within its operators in the IGR affects *racR* expression rather than *ydaS* expression. A two-plasmid genetic system was used to estimate the gene expression by quantitation of LacZ. The *racR* (pLexPracR1-3) (**A**) or *ydaS* [wider context, pLexydaST (**B**) and shorter context (**C**), pLexydaSshort] genes were fused to *lacZ* reporter on one plasmid, whereas the second plasmid delivered empty vector (orange) or a wild-type YdaT gene (blue) under the inducible P_{araBAD} promoter. Gene expression was measured as LacZ activity (Miller units) at increasing YdaT concentrations (under glucose, or from 0 to 0.01% arabinose). The error bars indicate the averages (±SD) of at least three independent measurements. Statistical significance calculations were performed using a two-tailed unpaired Student's *t*-test (ns, not significant; **P*-value ≤ 0.05; ****P*-value ≤ 0.001).

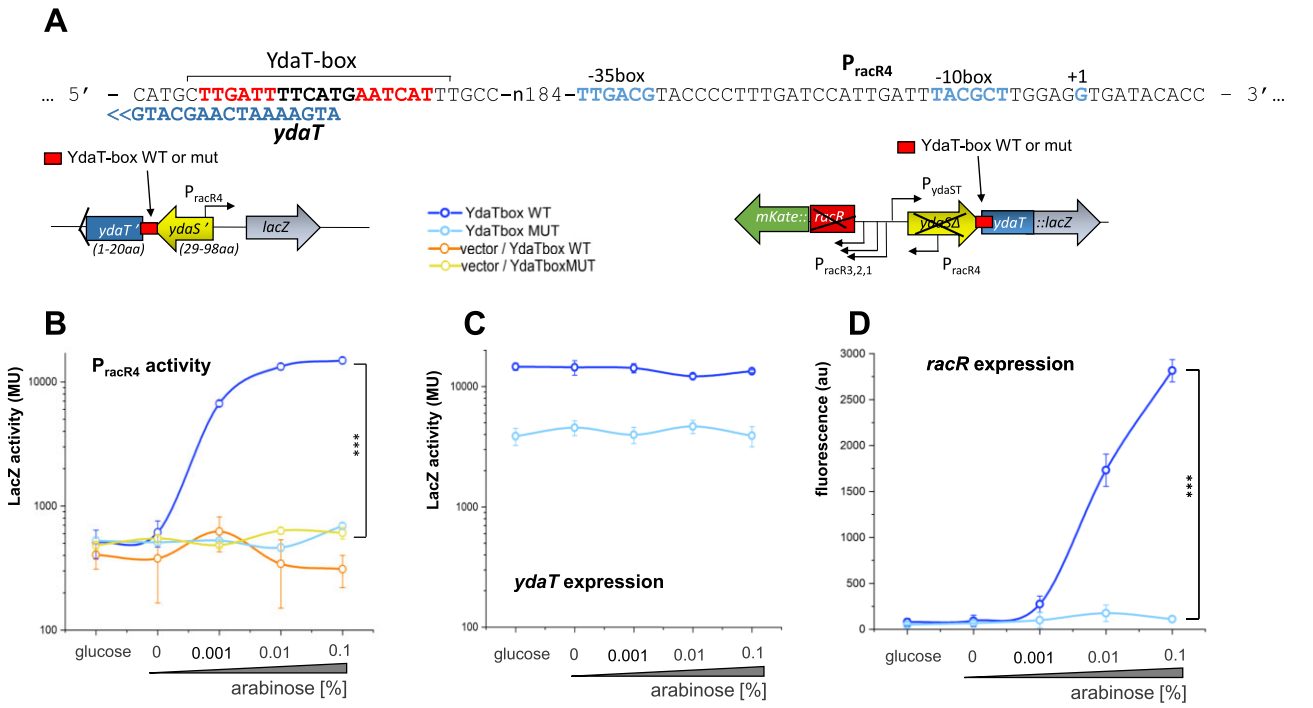


Figure 5. YdaT increases *racR* expression by binding to its YdaT-box and boosting the reverse promoter. (**A**) The YdaT-box, and elements of the reverse P_{racR4} elements, are shown at the top, along with schematics of the plasmids used. In all experiments, this two-plasmid system was used to test the effects of YdaT delivered *in trans*, using the same procedures as described in Fig. 4. (**B**) P_{racR4} promoter activity was measured in the absence of active *ydaS* and *ydaT* genes, but preserving the natural distance between promoter elements and YdaT-box, as indicated in the plasmid scheme. Two constructs were used with either the WT YdaT-box or its mutated variant (pLexPracR4 or pLexPracR4MUT, respectively). The second plasmid delivered either inducible *ydaT* over a gradient of inducer (arabinose) or empty vector. (**C** and **D**) A two-reporter construct (pLexydaSΔracR::mKate) was prepared to measure *ydaT* expression via LacZ activity (**C**) and *racR* expression was assessed from a red fluorescent reporter (mKate) (**D**). Again, the same plasmid delivering inducible *ydaT* to test whether binding to YdaT-box (or its mutated variant) affected the divergent P_{racR4} or P_{ydaST} promoters, located between the two reporters. Colors indicate which plasmid pair was used. The error bars indicate the averages (±SD) of at least three independent measurements. Statistical significance calculations were performed using a two-tailed unpaired Student's t-test (*** (*P*-value ≤ 0.001).

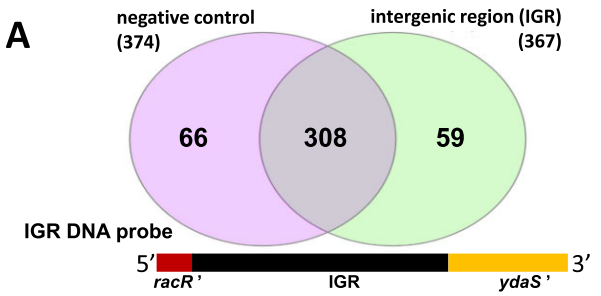
ydaST transcript and/or same direction promoters for *racR* gene. Accordingly, we constructed a two-reporter plasmid. Though none of TF is active, all of the binding sites for YdaT are present. One reporter measures *ydaT* expression (LacZ units), while the second shows *racR* expression (red fluorescence from mKate) (Fig. 5A). In addition, the same plasmid but with a mutated YdaT-box served as control. We varied YdaT expression in *trans* again, and measured both reporters with WT and mutated YdaT-box in parallel. The level of *ydaT* expression did not change much over the entire arabinose gradient, so it seems, the P_{ydaST} promoter activity did not change much regardless of YdaT binding to its YdaT-box (Fig. 5C). However, when we measured expression of *racR*, we observed a steep increase only in case of WT YdaT-box, and not within the mutated variant (Fig. 5D).

Altogether, we clearly demonstrated the importance of YdaT binding to the YdaT-box for activation of *racR* expression, which is mediated by the P_{racR4} promoter located at *ydaS* coding sequence. It seems this may explain the cooperation of RacR and YdaT to achieve a common goal, which is a stable maintenance of prophage within a genome.

YdaT expression is blocked by the host oxidative stress TF–OxyR

Expression of *ydaS*/*ydaT* is undetectable under physiological growth conditions [21]. We previously determined the start of a single bicistronic transcript for *ydaS* and *ydaT*, and mapped their common P_{ydaST} promoter [24]. This promoter has nearly perfect sequences as compared to the *E. coli* σ^{70} consensus promoter [49], including the −10box (TAATAT), −35box (TTGACA), and 18 bp spacing between them (Fig. 2B), indicating strong promoter potential, as we have actually seen *in vitro* [24]. This raises the question of why P_{ydaST} is normally silent *in vivo*. Thus, we sought to find some host cellular TFs, which might be involved in suppressing toxicity of *ydaT*. We already found the abundant presence of RacR repressor in the IGR between *ydaS* and *racR* genes, but P_{ydaST} promoter had weak activity even in *ΔracR* background [21, 24]. To identify proteins that might bind to the promoter region upstream of *ydaS*, we employed a DNA pull down assay. We prepared two biotinylated DNA probes, one spanning the entire IGR region (between the *racR* and *ydaS* genes, Fig. 6A) and the second serving as a negative control, encompassing an irrelevant promoter region of similar length (from the Csp231I restriction-modification system [32]). Three biological replicates of *E. coli* cell extracts were applied and the pulled-down proteins were analyzed by mass spectrometry.

The LC-MS/MS analysis showed >300 proteins bound to both DNA probes. For further analysis, we considered only peptides identified solely in the IGR region (absent in negative control) and present in all three pull-down replicates (Fig. 6A and B; Supplementary Table S1). We also limited our analysis to proteins with regulatory potential, such as TFs. Only two proteins met all these criteria: RacR and OxyR. The RacR



B

UniProt acc. number		IGR DNA probe replicates			Negative control
P0ACQ4	**OxyR**	√	√	√	—
P76062	**RacR**	√	√	√	—
P37767	YfhH	√	—	—	—
P0A780	NusB	√	—	—	—
P67662	AaeR	√	—	—	—
P22186	MraZ	√	—	—	—
P0ACU5	FabR	√	—	—	—
P0A9G2	NhaR	√	—	—	—
P46118	HexR	√	—	—	—
P77171	YdcI	√	—	—	—
P07604	TyrR	√	—	—	—
P39376	YjiE	—	√	√	—
P0ACF8	H-NS	√	√	—	√

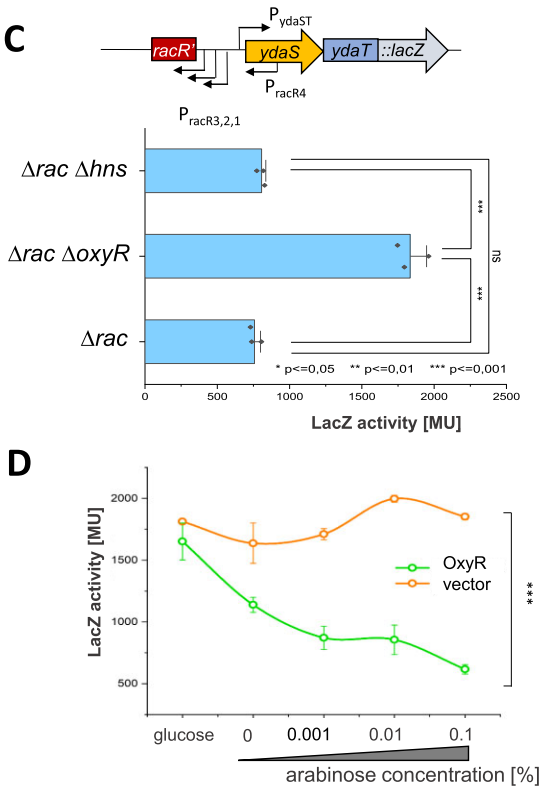


Figure 6. Negative impact of global transcriptional regulator OxyR on *ydaST* expression. (**A**) A Venn diagram (not to scale) illustrates the number of protein hits obtained from a DNA pull-down assay using two DNA substrates (probes): one with the IGR between *racR* and *ydaS* genes (a scheme below Venn diagram), and a negative probe with DNA similar length. (**B**) Identified proteins which are TFs are listed in table form, in columns representing the three tested probe replicates and one of the negative probe. Only two TFs fulfilled the criteria of (i) being present in all three replicates, (ii) absent from the negative control: RacR and OxyR. (**C**) The *ydaT* gene expression was measured in a *Δrac* background with either *ΔoxyR* or *Δhns Δhha* (as a control). Expression was measured from a plasmid construct, in which *ydaT* was transcriptionally fused to the *lacZ* reporter gene (pLexydaST on the scheme). (**D**) Effect of inducible *oxyR* expression, under control of the arabinose-inducible promoter, showing its negative effect on *ydaT* expression. The *ydaT* expression level was measured in *E. coli Δrac ΔoxyR* strain with two plasmids: one with *ydaT* fused to *lacZ* (as in schematic), and the other delivering OxyR from arabinose-inducible P_{BAD} promoter (pBAD-oxyR plasmid). Empty vector served as a negative control. LacZ assay was performed in triplicate, over a gradient of arabinose, and the averages and standard deviations (± SD) are shown. Statistical significance calculations were performed using a two-tailed unpaired Student's *t*-test (ns, not significant; **P*-value ≤ 0.05; ***P*-value ≤ 0.01; ****P*-value ≤ 0.001).

repressor appearance was expected and served as an internal positive control. The OxyR regulator is associated with global oxidative stress [52], though the experiment was performed under physiological conditions in LB medium.

To determine whether OxyR indeed is involved in negative regulation of *ydaS* and *ydaT* operon, we constructed a double mutant strain of *E. coli* (*Δrac ΔoxyR*) and quantitated P_{ydaST} promoter activity. We used also *Δhns Δhha* mutant strain as a control, which eliminated a global transcription regulator H-NS and Hha (hemolysin expression modulating protein), from among protein candidates to bind IGR region (H-NS bound to both DNA segments, Fig. 6B). Hha acts as a helper to H-NS, enhancing its global gene silencing efficiency [53, 54]. To measure P_{ydaST} activity, the reporter *lacZ* gene was fused to *ydaT* gene in presence of active *ydaS* and in the absence of *racR* gene (Fig. 6C). Our results confirmed that in the *Δrac* background only OxyR, but not H-NS, affects expression from P_{ydaST}, as the expression level increases 2.5-fold in the absence of OxyR, while it is not changed in *Δhns* strain. In addition, we also tested the effects of overexpressing *oxyR*, from an arabinose-inducible promoter, on P_{ydaST} in the same *Δrac ΔoxyR* strain. As expected, increasing OxyR levels reduced P_{ydaST} promoter activity (Fig. 6D). In fact, we tried to find a sequence typical for OxyR binding sites, but such sequence is poorly conserved [55].

Evidence for direct YdaT binding to the host *rcsA* promoter region and its regulatory impact

In the previous section, we found that a host TF (OxyR) controls a Rac prophage promoter (P_{ydaST}). Here we did the reciprocal experiment, looking for host genes controlled by the Rac prophage TF YdaT. Previously, we found an additional effector gene for YdaT outside the *rac* locus by creating a Tn*5* genome library without Rac genes and selecting for growth under YdaT overexpression [25]. The selected Tn*5*-disrupted gene was *rcsA*, coding a multifunctional transcriptional activator, which controls a phosphorelay system and is involved in biofilm production, cell division, motility, swarming, and other functions [56, 57]. We used an essential complementation assays and *de novo rcsA* mutagenesis, to confirm the *ydaT–rcsA* genetic associations responsible for inducting cell filamentation and death under prolonged *ydaT* expression [25]. To determine whether YdaT affects *rcsA* expression directly, we used gel shift assay to test whether YdaT protein binds a DNA fragment containing the upstream region of *rcsA*

gene and part of its coding sequence (Fig. 7). YdaT binding was evident, in contrast to its nonbinding variant YdaTmut (H56A R57A) (Fig. 7A). We fused the same DNA fragment with the native *rcsA* promoter, to the *lacZ* reporter gene in a *Δrac* background, and induced *ydaT* expression in *trans* (Fig. 7B). The results show that higher YdaT levels led to increased *rcsA* expression, by almost 5-fold. EMSA and LacZ fusion experiments, together, indicate that YdaT activates *rcsA* directly. However, the same test when conducted in *Δrac Δhns Δhha* cells resulted in high *rcsA* gene expression with no response of *rcsA* promoter for YdaT action. This confirms that *rcsA* expression is under negative regulation by H-NS (Fig. 7B), as reported before [58].

Next, to reinforce the binding assay results, and to confirm that the change in RcsA expression was due to direct YdaT binding to the *rscA* region, we performed binding site mapping assays using again DNaseI footprinting (Fig. 8A). We used the same DNA fragment as a substrate that we used in EMSA studies. We also checked that the *rcsA* promoter region did not reveal any DNA inverted repeats like the YdaT-box site determined within *ydaT* sequence 5′-ATGATT-n_6-AATCAA-3′, so we expected some deviation from this recognition sequence. Indeed, a distinct region of DNA protection was observed far upstream of *rcsA* promoter (∼170 bp upstream from *rcsA* transcription start site). Of note, the protection site overlapped the recognition sites for the TF heterodimer RcsA–RcsB [59], as an important element of *rcsA* gene autoregulation. In fact, their binding sites are formed by two boxes, where box1, but not box2, is present within the protected area determined by footprint assay analysis (red bar at Fig. 8A). Such an area of protection was not detected when YdaTmut was used. We identified the sequence resembling YdaT-box within the protection sequence, as 5′-TGAATT-n_6-AATATA-3′, where only 7 nt out of 12 nt is identical with YdaT-box.

Of note, we determined that YdaT homolog from CP-933P prophage is unable to activate the *rcsA* promoter, indicating that some protein flexibility is achieved toward targets, like conserved YdaT-box sequences, but not others (Fig. 7C). This property may illustrate some limitations in TF specialization and effect on host.

Next, we also asked whether YdaT has any significant preferences among its molecular targets. We used the sensitive biophysical method of SpS analysis, for rapid and easy determination of affinity constants (K_D) for the three determined DNA substrates having YdaT-binding sites. Plotting the ratio of the fluorescence signals (670 nm/650 nm) against the concentration of YdaT protein provided binding curves that could be used to generate K_D values (Fig. 8B). The obtained dose-response curves were not observed using YdaTmut protein, which confirmed its inability to bind DNA. Overall, we did not see any great differences in binding, as all tested dissociation constants (K_D) were in the same order of magnitude, being 1.23 μM for *rcsA* promoter; 0.66 μM for *racR–ydaS* IGR; 1.1 μM for the upstream *ydaT* coding sequence (Fig. 8B). This suggests that, at any given cellular concentration of YdaT and under our growth conditions, all molecular targets are bound without any significant preference.

YdaT homologs, its predicted structure, and analogy to λCII repressor

YdaT is usually annotated in gene and protein databases as a putative bacterial toxin that is neutralized by the putative antitoxin YdaS. We and others have shown previously that both YdaT and YdaS are rather Rac prophage TFs [21, 24–26].

A BLAST database search revealed that YdaT is easily identifiable in most *E. coli* strains, particularly within prophage regions, and more broadly within the *Enterobacteriaceae* family. The predicted structure of YdaT and its homologs indicates that the 3D structure is highly conserved (Fig. 9 and Supplementary Fig. S3).The protein is composed of an N-terminal DNA-binding domain (DBD), which contains a classical helix-turn-helix (HTH) motif that binds to the major groove of DNA (Fig. 9A–C). This HTH motif is formed by five α-helices packed into a compact, globular domain. Specifically, the α3 helix interacts with the DNA's major groove, while helices α1–α2 and α3–α4 stabilize the interaction with the DNA backbone (Fig. 9C). To confirm this interaction, we mutated residues H56 and R57 to alanine, which resulted in the loss of DNA binding. Using size-exclusion chromatography, combined with co-chromatographed molecular size markers, we confirmed that YdaT forms tetramers (Fig. 9B). More precisely, the individual DBDs pair into two sets of dimers, with monomers A and B forming one dimer, and C and D forming another. At the C-terminus, the protein features a long α-helix (α6, residues 95–128), which serves as a scaffold for tetramerization (Supplementary Fig. S4A and B). The significance of this α6 helix was demonstrated by mutating residues L107 and A110 to prolines, which disrupted oligomerization (Supplementary Fig. S4C–E). These mutations, known for their helix-breaking properties, likely caused a disruption in secondary structure, as suggested by molecular modeling. The YdaT tetramer behaves similarly to its homolog CP-933P, forming a tetrameric structure. Each YdaT unit contains two independently binding sites within two inverted repeats, such as those found in YdaT operator recognition sites (Supplementary Fig. S5) [51]. YdaT also appears capable of recognizing single inverted repeats as a dimer, such as those in the *rcsA* promoter. Alternatively, it may require additional binding sites at a suitable distance to function as a tetramer, if possible. It can recognize two double DNA motifs located, for instance, at operator left (O_L) and operator right (O_R) from the IGR, or a YdaT-box at one site and the operator middle (O_M) from the IGR at another site (Supplementary Fig. S6).

Interestingly, the mode of action of YdaT resembles that of CII from Lambda phage, another well-studied TF. Structurally, these two proteins also share a similar architecture, with a crucial α-helix (α3 in both YdaT and CII) showing high sequence similarity (Fig. 9D). In both proteins, arginine residue stabilizes DNA binding (YdaT R57; CII R43), while next tryptophan residue maintains the overall DBD fold, with both residues conserved across the two proteins. This α3 helix from the HTH motif binds the major groove of DNA, while the rest of the DBD stabilizes the interaction by binding to the DNA backbone from above (Fig. 9E).

YdaT-mediated toxicity in subpopulation of *E. coli* cells

Our previous work demonstrated first that YdaT protein expression is lethal to *E. coli* MG1655 cells [23, 24] and second that to counteract the toxicity, the entire Rac prophage DNA is excised and eliminated from cells [25]. The cells cured of Rac completely recovered, provided that they could resist

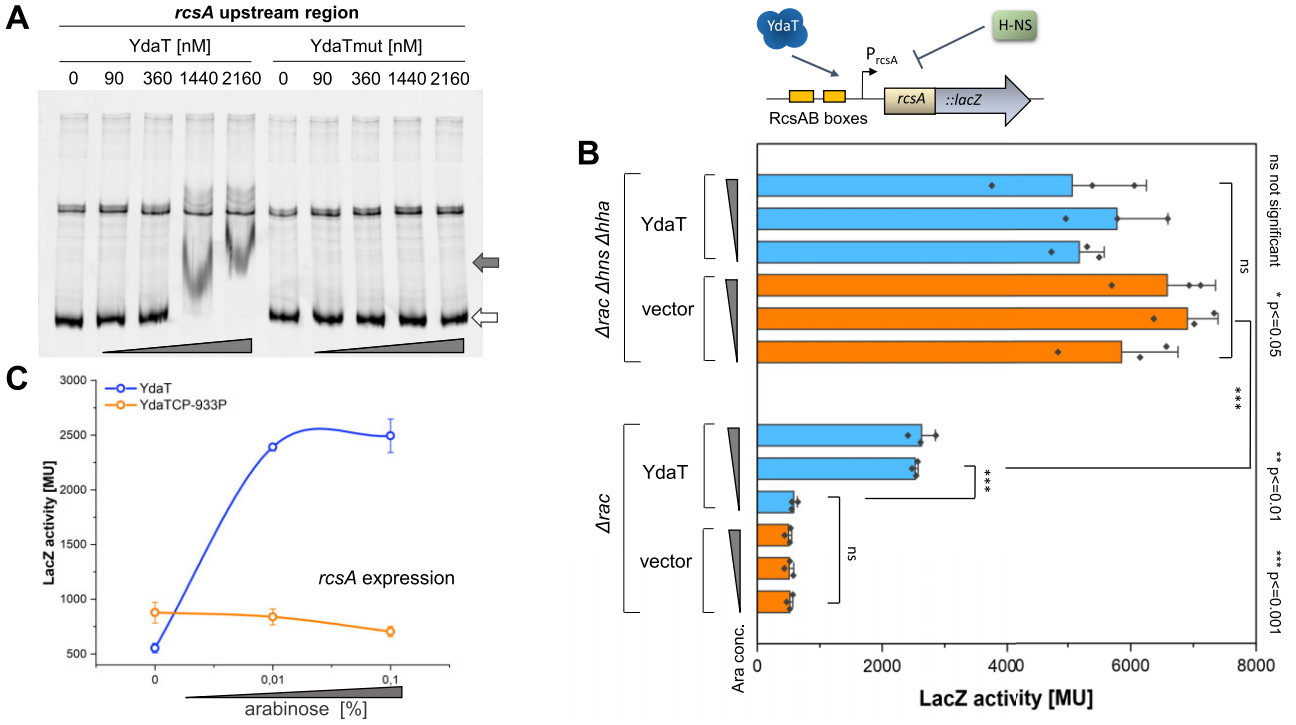


Figure 7. Rac prophage TF YdaT activates expression of *rcsA* gene coding for *E. coli* host RcsA TF. (**A**) YdaT binds the promoter region of *rcsA* (*in vitro* assay). The Cy5-labeled DNA substrate (20 nM), carrying the extended upstream region of *rcsA* gene and part of *rcsA* coding sequence, was mixed with increasing amounts of purified His-tagged YdaT or its nonbinding variant YdaTmut (H56A R57A) at the indicated concentrations. The reactions were processed as outlined in "Materials and methods" section, 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. (**B**) *In vivo* effect of YdaT binding on *rcsA* gene expression. In this experimental system, the *E. coli* K-12 *Δrac* or *Δrac Δhns Δhha* strain cells contained two compatible plasmids: one with *rcsA* promoter fused in the same open reading frame with the *lacZ* reporter gene (pLexUTRrcsA); the second plasmid delivered empty vector (pBAD33) or a wild-type YdaT gene (pBAD33-ydaT) under the inducible P_{araBAD} promoter. The *rcsA* gene expression was measured as LacZ activity (Miller units), at increasing YdaT or YdaTmut concentrations (0, 0.01, and 0.1% arabinose). Scheme of regulatory factors affecting *rcsA* expression: negative role of H-NS and positive role of YdaT. (**C**) Only YdaT of Rac can affect the *rcsA* promoter, but not a homolog of CP-933P. Two plasmid system was used as in panel (B), but $YdaT_{Rac}$ (pBAD33-ydaT) or $YdaT_{CP933P}$ (pBAD-ydaTCP) was delivered in *trans*. The *rcsA* gene expression was measured as LacZ activity. The error bars indicate the averages (± SD) of at least three independent measurements. Statistical significance calculations were performed using a two-tailed unpaired Student's *t*-test (ns, not significant; **P*-value ≤ 0.05; ***P*-value ≤ 0.01; ****P*-value ≤ 0.001).

the transient *ydaT* expression for several generations. We observed that the lethal effect was preceded by cell elongation, but was not observed at every cell. Because a subset of cells were resistant, eventually the entire population was restored.

In addition, previous transcriptomic analysis [23], revealed that high *ydaT* expression was also associated with repression of an entire group of genes responsible for lipopolysaccharide (LPS) synthesis, such as operon *waaBGLOPRSUYZ*, *wbbIJK*, *gmhB*, *glf*, *lpxP*, and *nlpA*.

Thus, we aimed to explore the toxicity more closely, but in Δrac cells to avoid the effect of other Rac regulators and genes. We employed the LIVE/DEAD™ BacLight™ Bacterial Viability Kit (Thermo Fisher, Invitrogen) to differentiate the viable (green, SYTO 9 staining) and dead cells (red, PI staining) using fluorescence techniques at both population and single cell levels. In this assay, the dead cell ratio is defined as a number of red cells (that have lost the membrane integrity) to the total number of green cells stained with SYTO 9. The *E. coli Δrac* cells carrying the pBAD-ydaT plasmid were exposed either to glucose (*ydaT* OFF) or to arabinose (*ydaT* ON), and samples were collected over 180 min (Fig. 10). We observed that cell viability (CFUs) started to decrease about 50 min after *ydaT* expression. By 90 min the viable cell number dropped by four orders of magnitude, but remained at that level over the next 3 h (Fig. 10A). This is in agreement with a parallel test measuring dead cells via PI fluorescence, where the highest level of dead cells was observed ~90 min after *ydaT* induction, again followed by a leveling off (Fig. 10B). The microscopic and flow cytometry analysis revealed that within 3 h approximately 20 ± 5% of cells are already dead, as PI-staining (red) is only seen in cells that lost membrane integrity (Fig. 10C and D). However, with prolonged observation it is hard to calculate the exact number of dead cells by this method, as the dead cells undergo the lysis. In addition, at that point the cell morphology undergoes a transition from rod-shaped to slightly elongated, spindle-shaped cells with bright center, where DNA has condensed (Fig. 10D, internal window). This stage precedes the full filamentation observed several hours after *ydaT* induction (Fig. 1A).

We also carried out more single-cell studies on YdaT-induced cells using TEM and SEM, respectively. The TEM images indicate chromatin condensation, revealed as white blobs within cells (Fig. 10E indicated by arrows). In accord, the SEM images capture the membrane collapse stages, displaying the profound membrane damage and cellular content leakage (Fig. 10F indicated by arrows). The fluorescent microscopy also shows the cytoplasm efflux, as a haze of red PI dye leaking from dead cells (Fig. 10G marked by arrows).

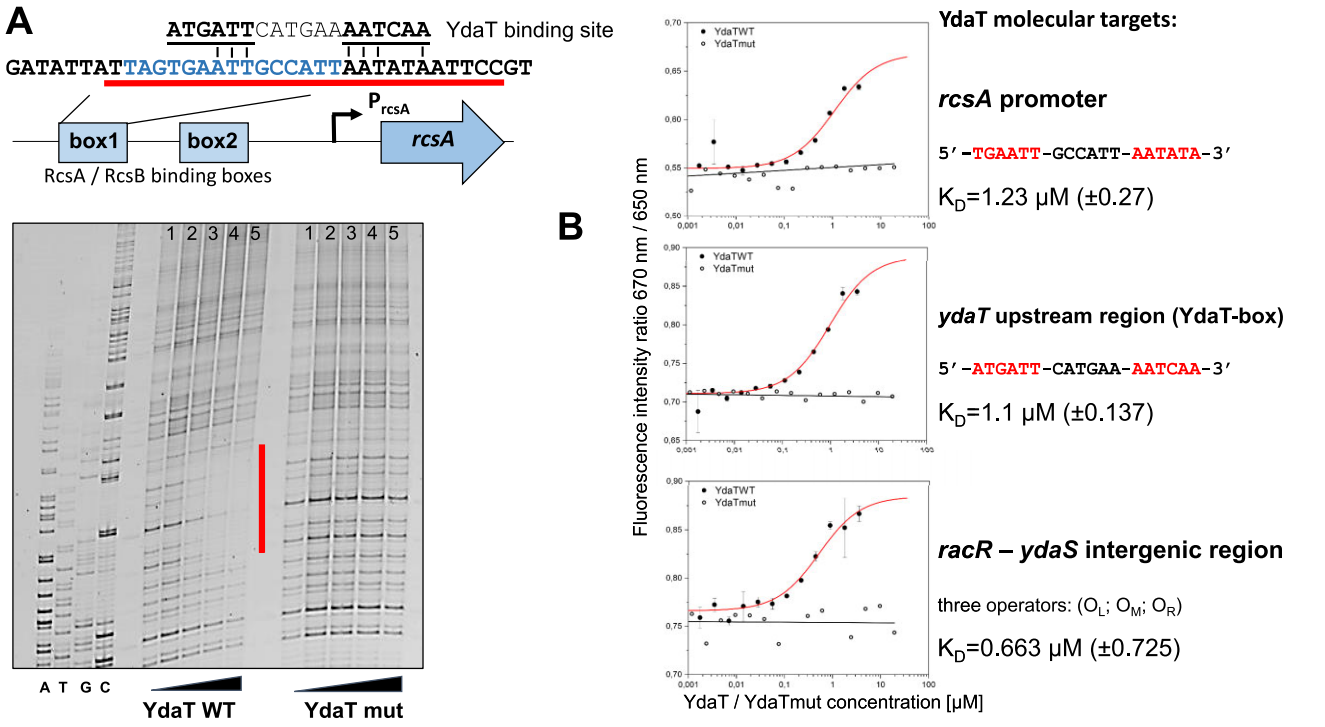


Figure 8. Identification of the YdaT-binding site within *rcsA* upstream region and comparison of YdaT binding to its three molecular targets. (**A**) Mapping the YdaT binding by DNaseI footprinting. The reactions containing 40 nM Cy5-labeled DNA fragments carrying the upstream region of *rcsA* gene were incubated with increasing amounts of purified YdaT WT protein or its inactive variant YdaTmut [H56A R57A] (lanes 1–5 correspond to protein concentration: 0, 360, 720, 1440, and 2880 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" section. The protected areas (resistant to DNaseI digestion) are indicated by a red bar, and the corresponding DNA sequence is presented above the gel picture. Blue font in the protected sequence indicates box1, where RcsA/B binds. Box1 overlaps the inverted repeats determined as YdaT-binding sites: 5′-TGAATTGCCATTAATATA-3′. (**B**) Quantitative comparison of YdaT binding at its three natural substrates using the SpS technology. The same Cy5-labeled substrates were used as in EMSA assay. The binding affinity was determined by intensity of fluorescence as 670/650 nm ratio in gradient of increasing amounts of YdaTWT and its variant YdaTmut (negative control). The dissociation constants (K_D) were calculated for each substrate. The error bars indicate the averages (±SD) of at least three independent measurements. To ease visualization, least-squares linear fits are shown for the control DNAs.

In addition, we tested also whether the decreased LPS synthesis associated with *ydaT* expression affects membrane permeability. Usually such cells may be characterized by hypersensitivity to a low level of SDS (0.1%) and bile salts. Though the YdaT-expressing cells are less viable than empty vector controls in the absence of SDS, we attempted to test the cell growth (as CFUs), and confirmed that *ydaT* expression makes cells more sensitive to SDS (Fig. 10H). Similar results were obtained when the growth was tested on McConkey agar containing bile salts (Fig. 10I). Taken together, the viability test, including SEM, TEM and membrane sensitivity test are clear evidence of the toxic high-level YdaT effect for the majority of cells, but independent of genes within the Rac prophage. However it seems some cells are able to escape the genetic pressure of YdaT, suggesting the strong host response to survive.

Impact of YdaT expression on *E. coli* cell motility and biofilm formation

It was reported before that RcsA/RcsB heterodimer can efficiently repress expression of the *flhDC* operon, which encodes the master regulator of flagella production, by binding to Rcs boxes located just downstream of *flhDC* transcription start [60, 61]. In addition, RcsA alone can repress the *flh* operon, and reduce motility [62]. Thus we wondered if, under conditions of high RcsA expression following induction of *ydaT*, the cell motility is inhibited enough to show a difference in a standard motility assay. We prepared soft agar plates with arabinose, and spotted the cells after *ydaT* induction to evaluate the motility-associated phenotypes. The results clearly showed that motility was significantly reduced (about 5-fold) with *ydaT* expression, as measured by migration diameter (Fig. 11A and B). In order to confirm that weaker motility is not only the effect of lower cell viability under induction of *ydaT* expression, we took the bacteria sample from the spot center on soft-agar and stained with PI (red; dead cells) and DAPI (blue; all cells). We observed a strong majority of live cells, with some isolated elongated dead cells (Fig. 11C).

The Rcs phosphorelay is known to negatively impact cell attachment to surfaces, an early stage of biofilm formation, by repressing genes responsible for surface adhesins and inhibiting the expression of the *flhDC* operon, which encodes the master regulators of flagella biosynthesis—both essential for the initiation of biofilms [60, 63–65]. On the other hand, the RcsA TF stimulates expression of the *cps* operon, leading to the production of extracellular colanic acid, which is essential for the maturation of biofilms in *E. coli* during the late stages of biofilm formation [56, 66]. Thus, we also predicted that YdaT production may indirectly affect biofilm formation, and we addressed this issue experimentally. Again, we grew the

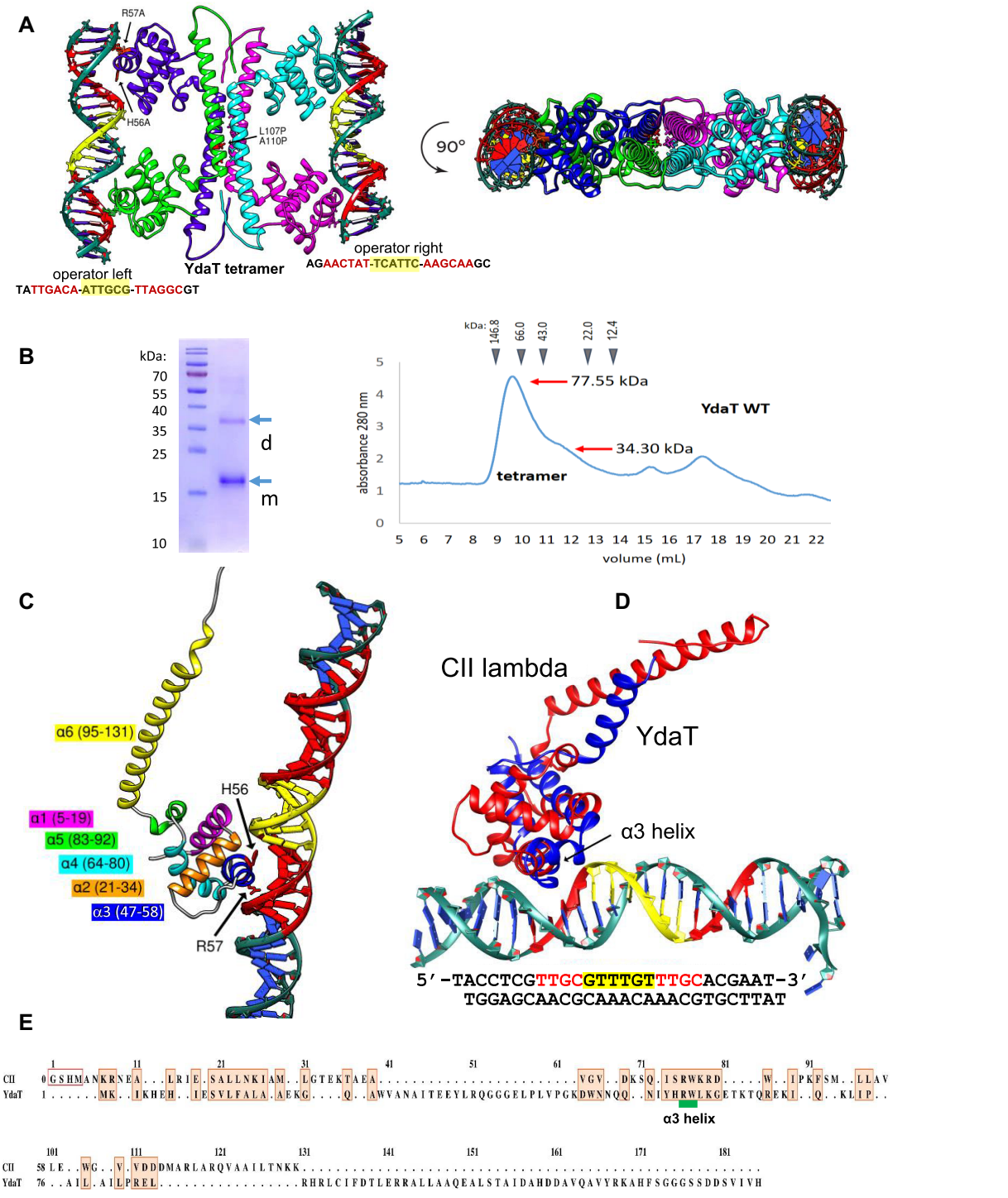


Figure 9. YdaT and CII protein from bacteriophage λ are analogs. (**A**) YdaT 3D structure as predicted by AlphaFold3. Experimental data from size-exclusion chromatography indicate that YdaT forms a tetramer capable of binding two DNA strands, each containing two recognized sites (the symmetrical inverted repeats 5′-TTGATT-n6-AATCAA-3′). The oligomerization scaffold is formed by a long α-helix (α6). Substitutions of residues L107 and A110 to prolines, marked with arrows, reduce oligomerization. (**B**) The purified 6His-tagged YdaT reveals monomer (arrow "m": 18.8 kDa) and dimer (arrow "d": 37.6 kDa) showing high propensity to oligomerization even under denaturing conditions (SDS–PAGE gel) in comparison to protein markers. Size-exclusion chromatography elution profile of 6His-tagged YdaT (monomer: 18.8 kDa) on a Superdex 75 10/300 GL column (GE Healthcare) supports the tetrameric structure. The YdaT peak at 77.55 kDa aligns with the predicted mass of the tetramer (75.24 kDa), indicating a close fit to the curve. Molecular mass markers are indicated by arrows. (**C**) The DBD features a HTH motif, with α3 playing a crucial role in DNA binding, as confirmed by mutations of key residues (H56A and R57A). (**D**) Comparison of YdaT to the CII protein from bacteriophage λ. In both proteins, the DNA interaction occurs via a HTH motif. Specifically, the α3 helix in both YdaT and CII plays a central role in binding the major groove of DNA. This interaction is conserved at both the sequence and structural levels, with key residues such as YdaT R57 (CII R43) stabilizing DNA binding, and YdaT W58 (CII W44) maintaining the protein fold. The rest of the DBD in each protein further stabilizes the interaction by anchoring to the DNA backbone from above. (**E**) The indicated DNA sequence refers to the recognition site for λCII. Alignment of amino acids sequences for λCII and YdaT is also shown with secondary structure elements marked in boxes.

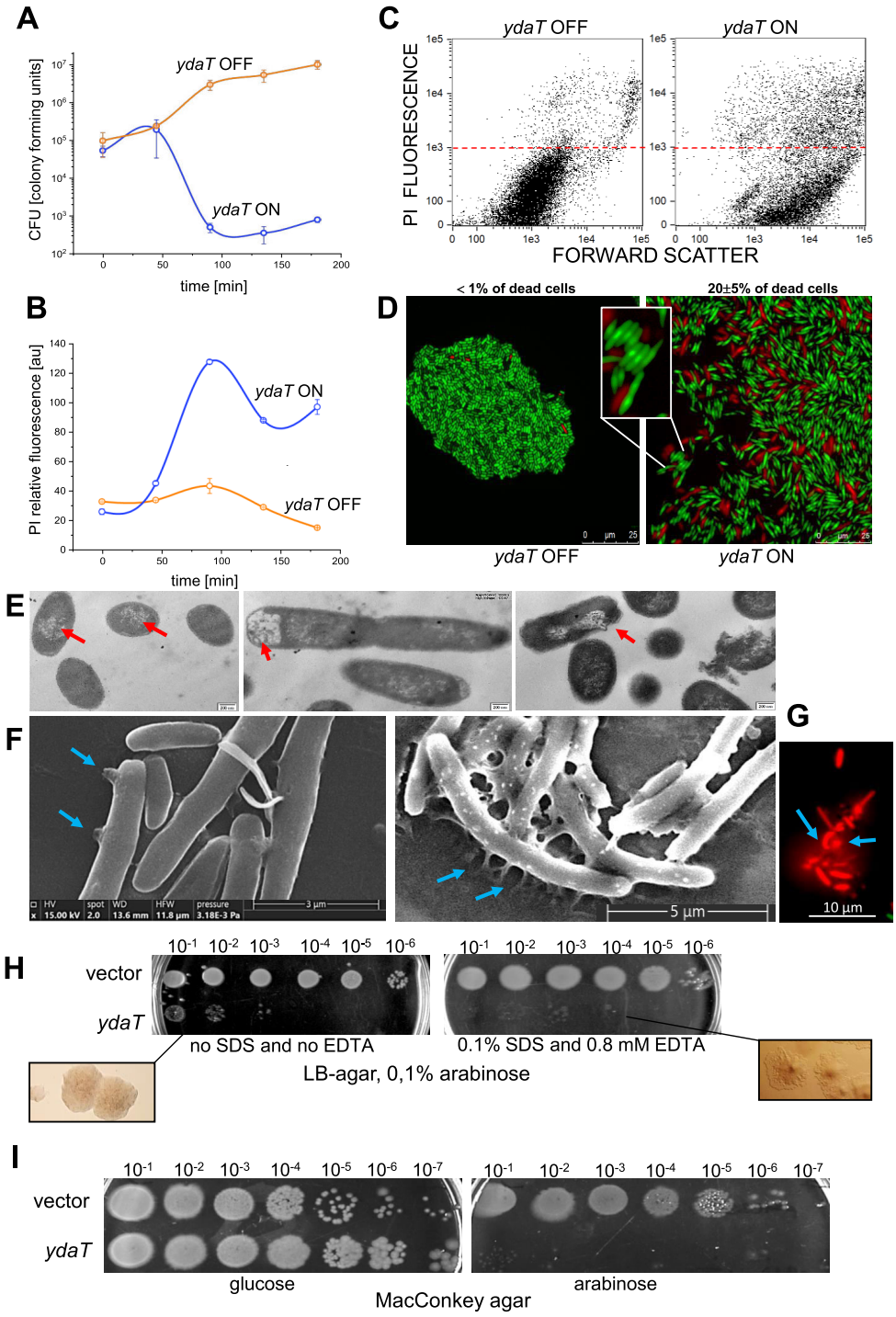


Figure 10. YdaT-mediated toxicity on *E. coli* cells. (**A**) Kinetics of YdaT toxicity on *E. coli Δrac* cells monitored within 3 h of *ydaT* induction. Cell survival evaluated by colony-forming units (CFU), with *ydaT* expression ON (+arabinose) or OFF (+glucose). (**B**) Measuring dead cells numbers by PI fluorescence staining. Data shown in panels (A) and (B) are the mean ± standard deviation (SD), $n = 3$. (**C**) Flow cytometry scatter plot displaying the dead cells stained with PI. The arbitrary dashed line refers to a control of dead cells only (ethanol-treated) and separates the dead cells (PI fluorescence > 1000) and live cells (PI < 1000). Forward scatter discriminates the cells by size and visualizes the entire cell population. The increased forward scatter presumably reflects cell filamentation. (**D**) Representative sample of fluorescence microscopy images of the LIVE/DEAD BacLight™ assay for cells with *ydaT* expression (ydaT ON) and without (ydaT OFF). Green color indicates live cells, red, dead cells (scale bar: 25 µm). Internal window shows the change of still live single cell from rod-shaped into spindle-shape cell with bright center, where the chromatin starts to condense. This stage may precede cell death. (**E**) TEM images show enlargement of the periplasmic space and nucleoid condensation (arrows), which precedes the eventual disruption of cell membranes and cell lysis. (**F**) SEM images showing leakage of cytoplasmic material as indicated by arrows. (**G**) Collapse of membrane integrity and release of fluorescence red dye (PI) with the cell contents. (**H**) Hypersensitivity of cell membranes to 0.1% SDS. Cells were induced with arabinose and spotted at indicated dilutions onto agar plates supplemented with 0.1% SDS and 0.8 mM EDTA, or just plain agar. Underneath: microcolonies of *ydaT*-expressing cells that grew on SDS agar revealed shadows (right), in contrast to cells from plain agar (left). (**I**) Sensitivity of *ydaT*-expressing cells on bile salts tested on MacConkey agar. Cells carrying the plasmid with *ydaT* gene were spotted at indicated dilutions on agar plates supplemented with glucose (*ydaT* repression) or arabinose (*ydaT* induction).

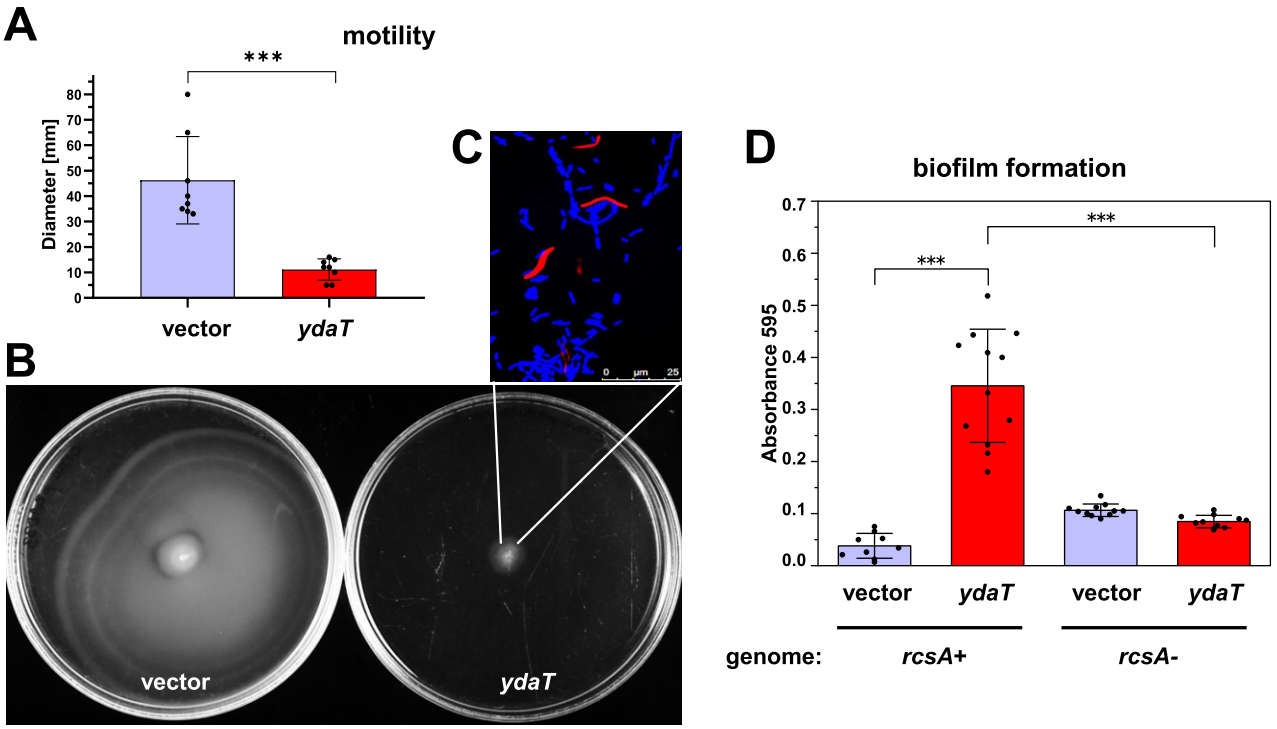


Figure 11. Motility and biofilm phenotypes of cells expressing *ydaT* gene. The Rac YdaT reduces a cell motility. (**A**) *E. coli Δrac* cells, carrying empty vector (pBAD33) or plasmid with inducible *ydaT* (pBAD33-ydaT), were exposed to arabinose and inoculated onto the surface of motility (soft-agar) plates, then incubated at 32°C for 20 h. The growth spots were measured as migration diameters. (**B**) Representative plates are shown. (**C**) A fluorescence microscopic image is shown, of a cell sample with *ydaT* overexpression, derived from the spot center after incubation on soft-agar. Cells were stained with DAPI (blue) to show all cells and PI (red) to show only dead cells. (**D**) The Rac YdaT favors biofilm formation in WT cells. The same cells expressing *ydaT* were tested for biofilm formation in the background of *ΔrcsA* genome or in WT. Empty vector served as a control. Cells were grown in LB with 0.1% arabinose at 37°C without shaking for 48 h. Biofilm formation was visualized using crystal violet staining and quantified by measuring the optical density at 595 nm. The mean from several replicates is presented; the error bars represent the standard deviation (SD). Statistical significance calculations were performed using a two-tailed unpaired Student's *t*-test (*** *P*-value ≤ 0.001).

two types of cells: WT and *ΔrcsA* cells with empty vector, or plasmid carrying arabinose-inducible *ydaT* at 37°C without shaking for 48 h, then measured biofilm formation *via* crystal violet staining (Fig. 11D). The quantification clearly showed that YdaT induction led to a 7-fold increase in biofilm formation in the presence of genomic *rcsA*, while failed to enhance biofilm formation in the absence of *rcsA* (Fig. 11D). Hence, we claim that YdaT stimulates biofilm production *via* RcsA activation.

Discussion

To date, the exact role of the Rac prophage TF YdaT in *E. coli* cell physiology remains elusive, due to the fact that this TF is not expressed under normal growth conditions. However, we created a genetic context in which YdaT expression was induced *via* adventitious transcriptional cross-talk between a C protein (regulator of a restriction-modification system) and the Rac master regulator - RacR [24, 67]. We previously described this phenomenon and demonstrated that the toxic effect on *E. coli* cells is associated with YdaT induction, and that this toxicity may be obviated by Rac prophage excision and elimination of prophage genes [25].

Our major objective in this report was to determine YdaT action as a TF, and to characterize YdaT regulatory properties, both within its prophage *racR–ydaST* operon, as well as outside of its native locus.

YdaT expression and its prophage locus regulatory landscape

Although Rac prophage is cryptic, and many of its structural genes are disabled, the *racR–ydaST* regulatory operon is still active. However expression of this operon is blocked by multilayered molecular mechanisms, which we attempted to characterize. This Rac region was previously investigated and noted to have an analogy to the immunity regions of other lambdoid prophages in terms of gene organization, operator locations, and gene regulation [21, 24, 26]. Specifically, this Rac region resembles the lysis/lysogeny decision region between the CI and Cro genes in the case of λ bacteriophage [68, 69], the *dicA* and *dicC* genes in case of *E. coli* Qin prophage [16], and many other prophages. We previously identified several promoters between *racR* and *ydaS* genes (IGR), as well as multiple RacR (CI homolog) binding sites [24] (Fig. 1). The stable production of RacR assures the prophage maintenance, and its loss drives the Rac excision and subsequent elimination. The very strong P_{ydaST} promoter (equivalent of P_R of λ phage) drives expression of both TFs genes: *ydaS* and *ydaT*, and yet under normal growth conditions its transcript is barely detectable. Previously we showed that YdaT expression does not induce Rac excision, unlike YdaS, supporting the notion of YdaT and RacR cooperation in maintaining the lysogen status of Rac [24]. We determined two areas of YdaT binding, one within the *ydaT* coding sequence (YdaT-box), and the other within the *racR–ydaS* IGR. The YdaT-box has in-

verted repeats that are nearly perfect in symmetry, but is quite variable in other regions, like the intergenic Rac locus (three operators) and *rcsA* promoter at genomic site. Of note, YdaT-box is almost the same as that for the YdaT homolog of CP-933P prophage [51]. YdaT forms a tetramer, and similarly to λCII, activates transcription from the reverse P_{racR4} (equivalent of P_{RE} of λ phage). This stimulates RacR repressor expression via binding to the YdaT-box, and possibly by RNAP recruitment, to maintain the lysogenic state [70, 71]. We tested the role of YdaT-box in Rac immunity region and found out that YdaT binds it to achieve stimulation of the reverse P_{racR4} promoter, just like in lambda and CP-933P phage. However YdaT in CP-933P prophage does not seem to affect the analog of λCI (PaaR2) [26]. In addition, YdaT of CP-933P does not bind within the *paaR2–ydaS* region (CI–Cro analog in lambda phage), as we observe in Rac. The λCII binding to P_{RE} results in the elevated levels of CI repressor [72], what we also observe.

Moreover, we showed that RacR and YdaS strongly depress *ydaT* expression, as does the global host regulator OxyR. These three TFs have binding sites in the IGR [24]. OxyR is a major sensor of oxidative and nitrosative stress, and controls over hundred *E. coli* genes [52, 73], but also affects prophage genes and resultant lysis/lysogeny decisions in *E. coli* (and in other hosts having OxyR homologs) [74–76]. So far, there is no evidence, of any lambdoid phage repressor having expanded its role beyond the phage locus, aside from the observation of poor *E. coli* growth and reduced fitness upon overexpression of λCII [77, 78], or of λCI in *E. coli* growing on succinate as a carbon source (due to direct inhibition of host *pckA* gene by CI [79]. Overall, we see much greater similarity in Rac immunity region regulation with lambda phage than with CP-933P prophage (which apparently lacks YdaT binding sites between *ydaS* and *paaR2*) [26].

Transcriptional prophage–host relationships

Many studies have focused on prophage genes that can directly contribute to the repertoire of host genes that affect host physiology or metabolism [6, 7, 9, 10, 18, 80, 81]. However, much less attention has been paid to prophage regulatory elements, like TFs or sRNAs, that may modulate expression of host genes. The co-existence of a prophage and its host's genes would seem to offer opportunities for positive selection of prophage TFs' gaining the ability to regulate some host operons. Likewise the host would seem to be under selective pressure to use its regulatory mechanisms to domesticate its prophages for its own benefit [6, 7]. (Taking this a step further, some prophage regulatory genes also control other co-resident prophages [82], and even communicate within a bacterial population in a phage specific code to make a coordinated lysogeny decisions [83]).

This report is a continuation of our previous studies, showing that a single prophage TF is able to induce a phenotypic changes through regulation of host-encoded genes [23, 25]. We mapped the crucial molecular interaction between prophage YdaT and a DNA sequence within so called RcsAB boxes, affecting expression of the host TF RcsA. RcsA is an auxiliary protein that helps to form more stable RcsA/RcsB heterodimer, with greater affinity for RcsAB boxes, which are present in many bacterial promoters [59, 84]. However, not much is known about function of box1 and box2 in the RcsAB binding regions, except that binding to both boxes activates target gene expression. YdaT binding to box1 enhances the activation mode via box2 only, to an even higher extent. Interestingly, the same DNA region is also bound by the bacterial histone-like protein H-NS [85], and yet YdaT can interfere with *rcsA* expression. However, in a *Δhns* context, full *rcsA* expression occurs regardless of YdaT action. This distant upstream *rcsA* region contain long A/T-rich tracts, inducing a DNA curvature that affects some proteins' preferential binding, resulting in gene expression changes. These effects may include binding of regulatory proteins like H-NS, RcsA/RcsB, and YdaT, or the small regulatory RNA DsrA [58], which together modulate the level of *rcsA* transcript. In addition, RcsA level is achieved and controlled through direct processing by the Lon protease [57].

The RcsA, together with RcsB, affects hundreds of genes within its regulon that are related to the exopolysaccharide and colanic acid biosynthesis (cps), cell motility, lipoprotein synthesis, and other gene regulators [56, 57, 86]. In addition, Rcs system can sense cell envelope damage, and regulate gene expression to counteract this stress. In this sense, YdaT might stand as an interesting prophage-encoded example of a TF with pleiotropic effects on *E. coli* physiology. Furthermore, since YdaT is notably able to interact with rather poorly conserved DNA sites, it is highly possible that interaction within the genome is not limited solely to RcsAB box, but it may appear also at other genomic sites affecting other genes expression.

Another TF, the DLP12 prophage protein AppY, was found to interact directly with 11 molecular targets, which collectively increase bacterial survival under low pH conditions, while also boosting biofilm formation and decreasing motility [14]. Other *E. coli* O157:H7, CP-933H prophage-encoded regulatory protein, PatE, increases the expression of genes associated with acid resistance and to inhibit the ones belonging to the heat shock family and type III secretion pathways [87]. In similar fashion, the *E. coli* Qin prophage TF - DicC (an analog of Cro repressor of lambda immunity region) binds hundreds of genomic loci related to acid and oxidative stress resistances, as well as host cell growth defects [88]. Though Qin-related TFs affect closely located genes controlling cell division, they seem not be linked to YdfX (Qin prophage YdaT homolog), which is coded for in the same operon as DicC (Rac prophage YdaS homolog). However some reports, related to the Qin prophage, describe growth defect mechanisms connected to the action of *dicF* (which specifies a small RNA that inhibits cell division by preventing *ftsZ* translation [89]; *dicB* encoding a protein that inhibits cell division by enhancing MinC's inhibitory activity on FtsZ [90, 91]; and *ydfD* encoding a protein capable of lysing cells [92].

Our results are also in accord with other reports showing that some prophage TFs' actions promote biofilm formation. Biofilms result in direct "cell-to-cell" and "cell-to-surface" interactions, *via* a matrix environment produced by the cells, including extracellular polymeric substances, certain proteins, curli pili, flagella, and extracellular DNA (eDNA) [93]. Here, the *rcsA* upregulation by YdaT of Rac leads to inhibition of the cell motility, as the Rcs phosphorelay components inhibit the expression of the *flhDC* operon, encoding the master regulators of flagella biosynthesis [60, 61]. In addition, RcsA also play a negative role in curli synthesis [63]. On the other hand, production of the colanic acids *via* activation of *rcsA* also may contribute to growth of the exopolysaccharides facilitating the attachment to the surface, that is crit-

ical for maturation of the biofilm [66]. YdaT is toxic and seems to be lethal for a high proportion of cells, and may thus deliver a crucial biofilm component - the extracellular DNA from lysed cells. The function of eDNA in biofilm formation and accumulation [94] was highlighted in number of reports [95], also linking the induction of lytic cycles with biofilm formation [75, 76, 96]. Although we don't understand the detailed YdaT role in pathways leading to biofilm promotion, we predict the involvement of the mentioned factors, which could collectively contribute to this modulation in host physiology.

Taken together, our studies provide molecular insights into a prophage-encoded TF that has been integrated into the *E. coli* regulatory network. Specifically, YdaT activity has expanded beyond its local phage regulatory function, impacting the global host regulatory network controlled by RcsAB. We highlighted the YdaT negative effects on host cells, which might indicate why *ydaT* expression is silenced. However, the fact that host cells maintained the active *ydaT* gene, while letting some other Rac genes decay mutationally, may reflect an important YdaT function within a yet undetermined pathway. This phenomenon could indicate that YdaT provides significant benefits to the host, particularly under non-optimal conditions, such as stress. Further testing with conditions of mimicking the environmental stress, such as oxidative and osmotic stress, heat shocks, exposure to low pH, could help elucidate the specific roles of YdaT in cellular survival mechanisms and its potential contributions to host resilience in challenging scenarios.

Acknowledgements

We thank Dr. Remy Loris, Dr. Thomas Wood, Dr. Monika Glinkowska, and Dr. Katarzyna Potrykus for the gift of strains/plasmids; Dr. Sebastian Dorawa for technical help in protein chromatography; Dr. Magdalena Narajczyk for TEM/SEM images, and Julia Waszak for genomic search of putative YdaT-binding sites. We appreciate also excellent help of Prof. Robert M. Blumenthal at the stage of critically reading our manuscript.

Author contributions: Formal analysis (E.W., K.G., N.L. A.W., L.P.K., and I.M.); Investigation (E.W., K.G., N.L., A.W., and L.P.K.); Methodology (E.W., K.G., N.L., A.W., and L.P.K.); Validation (E.W., K.G., N.L., A.W., and L.P.K.); Visualization (E.W., K.G., N.L., L.P.K., and I.M.); Resources (I.M.); Writing - original draft preparation (E.W., K.G., L.P.K., and I.M.); Writing - review and editing (E.W., K.G., L.P.K., and I.M.); Conceptualization (I.M.); Supervision (E.W. and I.M.); Funding acquisition (I.M.); Project administration (I.M.)

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

National Science Centre (Poland) [UMO-2019/35/B/NZ2/00701] to I.M. Funding to pay the Open Access publication charges for this article was provided by the University of Gdansk (Poland) and National Science Centre (Poland), grant number (UMO-2019/35/B/NZ2/00701).

Data availability

The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium *via* the PRIDE (36) partner repository with the dataset identifier PXD059580. The complete data from the bioinformatics analyses are available through the Repository for Open Data (RePOD) at https://doi.org/10.18150/1BZXHU.

References

1. Seshasayee AS, Sivaraman K, Luscombe NM. An overview of prokaryotic transcription factors: a summary of function and occurrence in bacterial genomes. *Subcell Biochem* 2011;**52**:7–23. https://doi.org/10.1007/978-90-481-9069-0_2
2. Ishihama A, Shimada T, Yamazaki Y. Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Res* 2016;**44**:2058–74. https://doi.org/10.1093/nar/gkw051
3. Gao Y, Lim HG, Verkler H *et al.* Unraveling the functions of uncharacterized transcription factors in *Escherichia coli* using ChIP-exo. *Nucleic Acids Res* 2021;**49**:9696–710. https://doi.org/10.1093/nar/gkab735
4. Lally P, Gómez-Romero L, Tierrafría VH *et al.* Predictive biophysical neural network Modeling of a compendium of *in vivo* transcription factor DNA binding profiles for *Escherichia coli*. bioRxiv, https://doi.org/10.1101/2024.05.23.594371, 24 May 2024, preprint: not peer reviewed.
5. Ghatak S, King ZA, Sastry A *et al.* The y-ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. *Nucleic Acids Res* 2019;**47**:2446–54. https://doi.org/10.1093/nar/gkz030
6. Ramisetty BCM, Sudhakari PA. Bacterial 'grounded' prophages: hotspots for genetic renovation and innovation. *Front Genet* 2019;**10**:65. https://doi.org/10.3389/fgene.2019.00065
7. Bobay LM, Touchon M, Rocha EP. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci USA* 2014;**111**:12127–32. https://doi.org/10.1073/pnas.1405336111
8. Lawrence JG, Ochman H. Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* 1998;**95**:9413–17. https://doi.org/10.1073/pnas.95.16.9413
9. Wang X, Kim Y, Ma Q *et al.* Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* 2010;**1**:147. https://doi.org/10.1038/ncomms1146
10. Casjens S. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 2003;**49**:277–300. https://doi.org/10.1046/j.1365-2958.2003.03580.x
11. Asadulghani M, Ogura Y, Ooka T *et al.* The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog* 2009;**5**:e1000408. https://doi.org/10.1371/journal.ppat.1000408
12. Shimada T, Ogasawara H, Kobayashi I *et al.* Single-target regulators constitute the minority group of transcription factors in *Escherichia coli* K-12. *Front Microbiol* 2021;**12**:697803. https://doi.org/10.3389/fmicb.2021.697803
13. Shimada T, Momiyama E, Yamanaka Y *et al.* Regulatory role of XynR (YagI) in catabolism of xylonate in *Escherichia coli* K-12. *FEMS Microbiol Lett* 2017;**364**:fnx220. https://doi.org/10.1093/femsle/fnx220
14. Derdouri N, Ginet N, Denis Y *et al.* The prophage-encoded transcriptional regulator AppY has pleiotropic effects on *Escherichia coli* physiology. *PLoS Genet* 2023;**19**:e1010672. https://doi.org/10.1371/journal.pgen.1010672

15. Pfeifer E, Hünnefeld M, Popa O *et al.* Impact of xenogeneic silencing on phage-host interactions. *J Mol Biol* 2019;**431**:4670–83. https://doi.org/10.1016/j.jmb.2019.02.011
16. Ragunathan PT, Ng Kwan Lim E, Ma X *et al.* Mechanisms of regulation of cryptic prophage-encoded gene products in *Escherichia coli*. *J Bacteriol* 2023;**205**:e0012923. https://doi.org/10.1128/jb.00129-23
17. Rueggeberg KG, Toba FA, Bird JG *et al.* The lysis cassette of DLP12 defective prophage is regulated by RpoE. *Microbiology* 2015;**161**:1683–93. https://doi.org/10.1099/mic.0.000115
18. Wang X, Kim Y, Wood TK. Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *ISME J* 2009;**3**:1164–79. https://doi.org/10.1038/ismej.2009.59
19. Stern A, Sorek R. The phage-host arms race: shaping the evolution of microbes. *Bioessays* 2011;**33**:43–51. https://doi.org/10.1002/bies.201000071
20. Guo Y, Quiroga C, Chen Q *et al.* RalR (a DNase) and RalA (a small RNA) form a type I toxin–antitoxin system in *Escherichia coli*. *Nucleic Acids Res* 2014;**42**:6448–62. https://doi.org/10.1093/nar/gku279
21. Krishnamurthi R, Ghosh S, Khedkar S *et al.* Repression of YdaS toxin is mediated by transcriptional repressor RacR in the cryptic Rac prophage of *Escherichia coli*. *mSphere* 2017;**2**:e00392-17. https://doi.org/10.1128/mSphere.00392-17
22. Sevin EW, Barloy-Hubler F. RASTA-Bacteria: a web-based tool for identifying toxin–antitoxin loci in prokaryotes. *Genome Biol* 2007;**8**:R155. https://doi.org/10.1186/gb-2007-8-8-r155
23. Negri A, Jąkalski M, Szczuka A *et al.* 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. *Nucleic Acids Res* 2019;**47**:9542–56. https://doi.org/10.1093/nar/gkz665
24. Wisniewska A, Wons E, Potrykus K *et al.* 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. *Nucleic Acids Res* 2022;**50**:10964–80. https://doi.org/10.1093/nar/gkac914
25. Gucwa K, Wons E, Wisniewska A *et al.* Lethal perturbation of an *Escherichia coli* regulatory network is triggered by a restriction-modification system's regulator and can be mitigated by excision of the cryptic prophage Rac. *Nucleic Acids Res* 2024;**52**:2942–60. https://doi.org/10.1093/nar/gkad1234
26. Jurėnas D, Fraikin N, Goormaghtigh F *et al.* Bistable expression of a toxin-antitoxin system located in a cryptic prophage of *Escherichia coli* O157:H7. *mBio* 2021;**12**:e0294721. https://doi.org/10.1128/mBio.02947-21
27. Bindal G, Krishnamurthi R, Seshasayee ASN *et al.* CRISPR-Cas-mediated gene silencing reveals RacR to be a negative regulator of YdaS and YdaT toxins in *Escherichia coli* K-12. *mSphere* 2017;**2**:e00483-17. https://doi.org/10.1128/mSphere.00483-17
28. Sorek R, Zhu Y, Creevey CJ *et al.* Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 2007;**318**:1449–52. https://doi.org/10.1126/science.1147112
29. Gerdes SY, Scholle MD, Campbell JW *et al.* Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* 2003;**185**:5673–84. https://doi.org/10.1128/JB.185.19.5673-5684.2003
30. Langer A, Bartoschik T, Cehlar O *et al.* A new spectral shift-based method to characterize molecular interactions. *Assay Drug Dev Technol* 2022;**20**:83–94. https://doi.org/10.1089/adt.2021.133
31. Gabrielsen OS, Hornes E, Korsnes L *et al.* Magnetic DNA affinity purification of yeast transcription factor tau - a new purification principle for the ultrarapid isolation of near homogeneous factor. *Nucl Acids Res* 1989;**17**:6253–67. https://doi.org/10.1093/nar/17.15.6253
32. Rezulak M, Borsuk I, Mruk I. Natural C-independent expression of restriction endonuclease in a C protein-associated restriction-modification system. *Nucleic Acids Res* 2016;**44**:2646–60. https://doi.org/10.1093/nar/gkv1331
33. Wiśniewski JR, Zougman A, Nagaraj N *et al.* Universal sample preparation method for proteome analysis. *Nat Methods* 2009;**6**:359–62. https://doi.org/10.1038/nmeth.1322
34. Mendes ML, Fischer L, Chen ZA *et al.* An integrated workflow for crosslinking mass spectrometry. *Mol Syst Biol* 2019;**15**:e8994. https://doi.org/10.15252/msb.20198994
35. Ma B, Zhang K, Hendrie C *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Comm Mass Spectrometry* 2003;**17**:2337–42. https://doi.org/10.1002/rcm.1196
36. Perez-Riverol Y, Bai J, Bandla C *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 2022;**50**:D543–52. https://doi.org/10.1093/nar/gkab1038
37. Abramson J, Adler J, Dunger J *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;**630**:493–500. https://doi.org/10.1038/s41586-024-07487-w
38. Discover C, Boitreaud J, Dent J *et al.* . Chia I: decoding the molecular interaction of life. bioRxiv, https://doi.org/10.1101/2024.10.10.615955, 15 October 2024, preprint: not peer reviewed.
39. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–89. https://doi.org/10.1038/s41586-021-03819-2
40. Lin Z, Akin H, Rao R *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;**379**:1123–30. https://doi.org/10.1126/science.ade2574
41. Wu R, Ding F, Wang R *et al.* High-resolution *de novo* structure prediction from primary sequence. bioRxiv, https://doi.org/10.1101/2022.07.21.500999, 22 Jult 2022, preprint: not peer reviewed.
42. Chowdhury R, Bouatta N, Biswas S *et al.* Single-sequence protein structure prediction using a language model and deep learning. *Nat Biotechnol* 2022;**40**:1617–23. https://doi.org/10.1038/s41587-022-01432-w
43. Kelley LA, Mezulis S, Yates CM *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 2015;**10**:845–58. https://doi.org/10.1038/nprot.2015.053
44. Steinegger M, Meier M, Mirdita M *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinf* 2019;**20**:473. https://doi.org/10.1186/s12859-019-3019-7
45. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Protein Science* 2016;**86**, 2.9.1–37. https://doi.org/10.1002/cpps.20
46. Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605–12. https://doi.org/10.1002/jcc.20084
47. van Kempen M, Kim SS, Tumescheit C *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 2024;**42**:243–6. https://doi.org/10.1038/s41587-023-01773-0
48. Holm L, Laiho A, Törönen P *et al.* DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci* 2023;**32**:e4519. https://doi.org/10.1002/pro.4519
49. Shultzaberger RK, Chen Z, Lewis KA *et al.* Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res* 2007;**35**:771–88. https://doi.org/10.1093/nar/gkl956
50. Shearwin KE, Callen BP, Egan JB. Transcriptional interference—a crash course. *Trends Genet* 2005;**21**:339–45. https://doi.org/10.1016/j.tig.2005.04.009
51. Prolič-Kalinšek M, Volkov AN, Hadži S *et al.* Structural basis of DNA binding by YdaT, a functional equivalent of the CII repressor in the cryptic prophage CP-933P from *Escherichia coli* O157:H7. *Acta Crystallogr D Struct Biol* 2023;**79**:245–58. https://doi.org/10.1107/S2059798323001249
52. Chiang SM, Schellhorn HE. Regulators of oxidative stress response genes in *Escherichia coli* and their functional

conservation in bacteria. *Arch Biochem Biophys* 2012;**525**:161–69. https://doi.org/10.1016/j.abb.2012.02.007
53. Baños RC, Vivero A, Aznar S *et al.* Differential regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS. *PLoS Genet* 2009;**5**:e1000513. https://doi.org/10.1371/journal.pgen.1000513
54. Hong SH, Wang X, Wood TK. Controlling biofilm formation, prophage excision and cell death by rewiring global regulator H-NS of *Escherichia coli*. *Microb Biotechnol* 2010;**3**:344–56. https://doi.org/10.1111/j.1751-7915.2010.00164.x
55. Zheng M, Wang X, Doan B *et al.* Computation-directed identification of OxyR DNA binding sites in *Escherichia coli*. *J Bacteriol* 2001;**183**:4571–79. https://doi.org/10.1128/JB.183.15.4571-4579.2001
56. Meng J, Young G, Chen J. The Rcs system in *Enterobacteriaceae*: envelope stress responses and virulence regulation. *Front Microbiol* 2021;**12**:627104. https://doi.org/10.3389/fmicb.2021.627104
57. Majdalani N, Gottesman S. The Rcs phosphorelay: a complex signal transduction system. *Annu Rev Microbiol* 2005;**59**:379–405. https://doi.org/10.1146/annurev.micro.59.050405.101230
58. Sledjeski D, Gottesman S. A small RNA acts as an antisilencer of the H-NS-silenced *rcsA* gene of *Escherichia coli*. *Proc Natl Acad Sci USA* 1995;**92**:2003–7. https://doi.org/10.1073/pnas.92.6.2003
59. Wehland M, Bernhard F. The RcsAB box. Characterization of a new operator essential for the regulation of exopolysaccharide biosynthesis in enteric bacteria. *J Biol Chem* 2000;**275**:7013–20. https://doi.org/10.1074/jbc.275.10.7013
60. Francez-Charlot A, Laugel B, Van Gemert A *et al.* RcsCDB His-Asp phosphorelay system negatively regulates the *flhDC* operon in *Escherichia coli*. *Mol Microbiol* 2003;**49**:823–32. https://doi.org/10.1046/j.1365-2958.2003.03601.x
61. Krin E, Danchin A, Soutourina O. RcsB plays a central role in H-NS-dependent regulation of motility and acid stress resistance in *Escherichia coli*. *Res Microbiol* 2010;**161**:363–71. https://doi.org/10.1016/j.resmic.2010.04.002
62. Fredericks CE, Shibata S, Aizawa S *et al.* Acetyl phosphate-sensitive regulation of flagellar biogenesis and capsular biosynthesis depends on the Rcs phosphorelay. *Mol Microbiol* 2006;**61**:734–47. https://doi.org/10.1111/j.1365-2958.2006.05260.x
63. Vianney A, Jubelin G, Renault S *et al.* *Escherichia coli tol* and *rcs* genes participate in the complex network affecting curli synthesis. *Microbiology* 2005;**151**:2487–97. https://doi.org/10.1099/mic.0.27913-0
64. Ferrières L, Clarke DJ. The RcsC sensor kinase is required for normal biofilm formation in *Escherichia coli* K-12 and controls the expression of a regulon in response to growth on a solid surface. *Mol Microbiol* 2003;**50**:1665–82. https://doi.org/10.1046/j.1365-2958.2003.03815.x
65. Huang YH, Ferrières L, Clarke DJ. The role of the Rcs phosphorelay in *Enterobacteriaceae*. *Res Microbiol* 2006;**157**:206–12. https://doi.org/10.1016/j.resmic.2005.11.005
66. Danese PN, Pratt LA, Kolter R. Exopolysaccharide production is required for development of *Escherichia coli* K-12 biofilm architecture. *J Bacteriol* 2000;**182**:3593–96. https://doi.org/10.1128/JB.182.12.3593-3596.2000
67. Negri A, Werbowy O, Wons E *et al.* Regulator-dependent temporal dynamics of a restriction-modification system's gene expression upon entering new host cells: single-cell and population studies. *Nucleic Acids Res* 2021;**49**:3826–40. https://doi.org/10.1093/nar/gkab183
68. Ptashne M. Principles of a switch. *Nat Chem Biol* 2011;**7**:484–87. https://doi.org/10.1038/nchembio.611
69. Court DL, Oppenheim AB, Adhya SL. A new look at bacteriophage lambda genetic networks. *J Bacteriol* 2007;**189**:298–304. https://doi.org/10.1128/JB.01215-06
70. Ho YS, Mahoney ME, Wulff DL *et al.* Identification of the DNA binding domain of the phage lambda cII transcriptional activator and the direct correlation of cII protein stability with its oligomeric forms. *Genes Dev* 1988;**2**:184–95. https://doi.org/10.1101/gad.2.2.184
71. Ho YS, Wulff DL, Rosenberg M. Bacteriophage lambda protein cII binds promoters on the opposite face of the DNA helix from RNA polymerase. *Nature* 1983;**304**:703–8. https://doi.org/10.1038/304703a0
72. Schmeissner U, Court D, Shimatake H *et al.* Promoter for the establishment of repressor synthesis in bacteriophage lambda. *Proc Natl Acad Sci USA* 1980;**77**:3191–95. https://doi.org/10.1073/pnas.77.6.3191
73. Seth D, Hausladen A, Stamler JS. Anaerobic transcription by OxyR: a novel paradigm for nitrosative stress. *Antioxid Redox Signal* 2020;**32**:803–16. https://doi.org/10.1089/ars.2019.7921
74. Glinkowska M, Loś JM, Szambowska A *et al.* Influence of the *Escherichia coli oxyR* gene function on lambda prophage maintenance. *Arch Microbiol* 2010;**192**:673–83. https://doi.org/10.1007/s00203-010-0596-2
75. Binnenkade L, Teichmann L, Thormann KM. Iron triggers λSo prophage induction and release of extracellular DNA in *Shewanella oneidensis* MR-1 biofilms. *Appl Environ Microb* 2014;**80**:5304–16. https://doi.org/10.1128/AEM.01480-14
76. Li D, Liang W, Huang Z *et al.* The spontaneously produced lysogenic prophage phi456 promotes bacterial resistance to adverse environments and enhances the colonization ability of avian pathogenic *Escherichia coli* strain DE456. *Vet Res* 2024;**55**:37. https://doi.org/10.1186/s13567-024-01292-z
77. Kedzierska B, Glinkowska M, Iwanicki A *et al.* Toxicity of the bacteriophage lambda cII gene product to *Escherichia coli* arises from inhibition of host cell DNA replication. *Virology* 2003;**313**:622–28. https://doi.org/10.1016/S0042-6822(03)00376-3
78. Rajamanickam K, Hayes S. The bacteriophage lambda CII phenotypes for complementation, cellular toxicity and replication inhibition are suppressed in cII-oop constructs expressing the small RNA OOP. *Viruses* 2018;**10**:115. https://doi.org/10.3390/v10030115
79. Chen Y, Golding I, Sawai S *et al.* Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *PLoS Biol* 2005;**3**:e229. https://doi.org/10.1371/journal.pbio.0030229
80. Liu X, Li Y, Guo Y *et al.* Physiological function of Rac prophage during biofilm formation and regulation of Rac excision in *Escherichia coli* K-12. *Sci Rep* 2015;**5**:16074. https://doi.org/10.1038/srep16074
81. Beggs GA, Bassler BL. Phage small proteins play large roles in phage–bacterial interactions. *Curr Opin Microbiol* 2024;**80**:102519. https://doi.org/10.1016/j.mib.2024.102519
82. Pavlin A, Lovše A, Bajc G *et al.* A small bacteriophage protein determines the hierarchy over co-residential jumbo phage in *Bacillus thuringiensis* serovar israelensis. *Commun Biol* 2022;**5**:1286. https://doi.org/10.1038/s42003-022-04238-3
83. Erez Z, Steinberger-Levy I, Shamir M *et al.* Communication between viruses guides lysis-lysogeny decisions. *Nature* 2017;**541**:488–93. https://doi.org/10.1038/nature21049
84. Pristovsek P, Sengupta K, Löhr F *et al.* Structural analysis of the DNA-binding domain of the *Erwinia amylovora* RcsB protein and its interaction with the RcsAB box. *J Biol Chem* 2003;**278**:17752–59. https://doi.org/10.1074/jbc.M301328200
85. Meenakshi S, Karthik M, Munavar MH. A putative curved DNA region upstream of *rcsA* in *Escherichia coli* plays a key role in transcriptional regulation by H-NS. *FEBS Open Bio* 2018;**8**:1209–18. https://doi.org/10.1002/2211-5463.12348
86. Guo XP, Sun YC. New insights into the non-orthodox two component Rcs phosphorelay system. *Front Microbiol* 2017;**8**:2014. https://doi.org/10.3389/fmicb.2017.02014
87. Bender JK, Praszkier J, Wakefield MJ *et al.* Involvement of PatE, a prophage-encoded AraC-like regulator, in the transcriptional

activation of acid resistance pathways of enterohemorrhagic *Escherichia coli* strain EDL933. *Appl Environ Microb* 2012;**78**:5083–92. https://doi.org/10.1128/AEM.00617-12
88. Lally P, Tierrafría VH, Gómez-Romero L *et al.* A cryptic prophage transcription factor drives phenotypic changes *via* host gene regulation. bioRxiv, https://doi.org/10.1101/2024.09.21.614188, 21 September 2024, preprint: not peer reviewed.
89. Balasubramanian D, Ragunathan PT, Fei J *et al.* A prophage-encoded small RNA controls metabolism and cell division in *Escherichia coli*. *Msystems* 2016;**1**:e00021-15. https://doi.org/10.1128/mSystems.00021-15
90. Yang S, Pei H, Zhang X *et al.* Characterization of DicB by partially masking its potent inhibitory activity of cell division. *Open Biol* 2016;**6**:160082. https://doi.org/10.1098/rsob.160082
91. Pan H, Dong K, Rao L *et al.* The association of cell division regulated by DicC with the formation of viable but non-culturable. *Front Microbiol* 2019;**10**:2850. https://doi.org/10.3389/fmicb.2019.02850
92. Masuda H, Awano N, Inouye M. *ydfD* encodes a novel lytic protein in *Escherichia coli*. *FEMS Microbiol Lett* 2016;**363**:fnw039. https://doi.org/10.1093/femsle/fnw039
93. Beloin C, Roux A, Ghigo JM. *Escherichia coli* biofilms. *Curr Top Microbiol Immunol* 2008;**322**:249–89.
94. Panlilio H, Rice CV. The role of extracellular DNA in the formation, architecture, stability, and treatment of bacterial biofilms. *Biotech Bioengin* 2021;**118**:2129–41. https://doi.org/10.1002/bit.27760
95. Zhao J, Wang Q, Li M *et al.* *Escherichia coli* toxin gene *hipA* affects biofilm formation and DNA release. *Microbiology* 2013;**159**:633–40. https://doi.org/10.1099/mic.0.063784-0
96. Carrolo M, Frias MJ, Pinto FR *et al.* Prophage spontaneous activation promotes DNA release enhancing biofilm formation in *Streptococcus pneumoniae*. *PLoS One* 2010;**5**:e15678. https://doi.org/10.1371/journal.pone.0015678

Received: November 13, 2024. **Revised:** January 17, 2025. **Editorial Decision:** January 31, 2025. **Accepted:** February 5, 2025