

**A new sugar for an old phage:
A c-di-GMP dependent polysaccharide pathway sensitizes *E. coli*
for bacteriophage infection**

Benjamin Sellner, Rūta Prakapaitė, Margo van Berkum, Matthias Heinemann¹, Alexander Harms,
and Urs Jenal

Biozentrum of the University of Basel
Spitalstrasse 41, 4056 Basel, Switzerland

¹ University of Groningen, Molecular Systems Biology
Nijenborgh 4, 9747 AG Groningen, Netherlands

For correspondence: urs.jenal@unibas.ch

Keywords: *Escherichia coli*, phage receptor, phage infection, N4 phage, NfrA, NfrB, c-di-GMP, allosteric activation, WecB, ECA, ManNAc, glycosyltransferase, exopolysaccharide, dCache domain, arginine, PdeL, DgcJ, DgcQ.

Abstract

Bacteriophages are ubiquitous parasites of bacteria and major drivers of bacterial ecology and evolution. Despite an ever-growing interest in their biotechnological and therapeutic applications, detailed knowledge of the molecular mechanisms underlying phage-host interactions remains scarce. Here, we show that bacteriophage N4 exploits a novel surface glycan, NGR, as a receptor to infect its host *Escherichia coli*. We demonstrate that this process is regulated by the second messenger c-di-GMP and that N4 infection is specifically stimulated by the diguanylate cyclase DgcJ while the phosphodiesterase PdeL effectively protects *E. coli* from N4-mediated killing. PdeL-mediated protection requires its catalytic activity to reduce c-di-GMP and includes a secondary role as a transcriptional repressor. We demonstrate that PdeL binds to and represses the promoter of the *wec* operon, which encodes components of the ECA exopolysaccharide pathway. However, only the acetylglucosamine epimerase WecB but none of the other ECA components is required for N4 infection. Based on this, we postulate that NGR is an N-acetylmannosamine-based carbohydrate polymer that is produced and exported to the cell surface of *E. coli* in a c-di-GMP dependent manner where it serves as a receptor for N4. This novel carbohydrate pathway is conserved in *E. coli* and other bacterial pathogens, serves as the primary receptor for a range of N4-like bacteriophages, and is induced at elevated temperature and by specific amino acid-based nutrients. These studies provide an entry point into understanding how bacteria use specific regulatory mechanisms to balance costs and benefits of highly conserved surface structures.

Importance

Because bacterial surface glycans are in direct contact with the environment they can provide essential protective functions during infections or against competing bacteria. But such structures are also ‘Achilles heels’ as they can serve as primary receptors for bacteriophages. Bacteria thus need to carefully control the exposure of conserved surface glycans to balance costs and benefits. Here, we identify a novel exopolysaccharide that is widely conserved in *E. coli* and is used by N4 and related bacteriophages as primary receptor. We demonstrate that the synthesis of NGR (N4 glycan receptor) is tightly controlled by the second messenger c-di-GMP in a highly specific manner and by a single diguanylate cyclase. These studies provide an example of how bacteria can alleviate the strong selective pressure imposed on them by bacteriophages entering through conserved surface structures by carefully regulating their synthesis and secretion.

Introduction

Bacteriophages are ubiquitous predators of their bacterial hosts and drive their ecology and evolution in a tight arms race (Dion et al., 2020). The host range of bacteriophages is pre-determined by the recognition of specific receptors on the bacterial cell surface using receptor-binding proteins that are displayed by tailed phages on their tail fibers, tailspikes, or similar structures (Nobrega et al., 2018). While exposed glycan structures are often used as a first “primary” receptor for host recognition, irreversible adsorption and injection of the phage genome are triggered by subsequent binding to a terminal or “secondary” receptor directly on the cell surface (Nobrega et al., 2018). For Gram-negative bacteria like the model organism *Escherichia coli*, all known types of glycans including capsules, the highly variable O-antigen chains of LPS, and the conserved yet enigmatic enterobacterial common antigen (ECA) have been described as primary receptors for phage docking (Maffei et al., 2021; Broeker and Barbirz, 2017; Kim et al., 2015; Silva et al., 2016; Washizaki et al., 2016). However, surface exposed polysaccharides also play major roles in bacterial defense against phages as they can shield terminal receptors on the cell surface (Broeker and Barbirz, 2017; Kim et al., 2015; Porter and Martens, 2015; Rousset et al., 2018). *E.g.*, in *E. coli* K12 O-antigen expression was shown to eliminate the adsorption of a wide range of bacteriophages that could bind diverse terminal receptors and infect productively in the absence of this barrier (Maffei et al., 2021). Likewise, overproduction of capsules can effectively protect bacteria from phage adsorption (Chaudhry et al., 2020; Mutalik et al., 2020). This dual role of surface glycans as barrier and receptor is mirrored on the phage side in form of tailspikes. These are tail fibers decorated with glycan-targeting enzymes that specifically recognize certain sugar motifs on host exopolysaccharides and then modify or degrade them unit by unit to drive translocation of the virion along the polysaccharide chain towards the cell surface (Knecht et al., 2020; Pires et al., 2016; Letarov and Kulikov, 2017).

Previous genetic studies indicated that the ECA glycan chains are exploited as host receptor by diverse and very common bacteriophages like the well-studied podovirus N4 (Kiino and Rothman-Denes, 1989; Maffei et al., 2021). This is remarkable since – unlike the almost 200 different types of O-antigens for *E. coli* alone (Liu et al., 2019) – the ECA glycan is invariable across enterobacteria, possibly due to functional constraints in their interaction with their animal hosts (Rai and Mitchell, 2020a). To ease the selective pressure imposed by phage predation via conserved surface structures like ECA, bacteria have evolved different strategies including tightly regulating such surface components (Batchelor et al., 1991; Morona et al., 1995; Murray et al.,

2003). Understanding how bacteria maintain the expression of highly conserved surface glycans, despite of phage predation, is not only relevant for phage ecology and evolution but could also have great value for the therapeutic use of bacteriophages. Not unexpectedly, the *bona fide* ECA-targeting phages studied in previous work systematically displayed the broadest host recognition of all phages tested (Maffei et al., 2021) which is, intuitively, a key property when selecting phages for therapeutic purposes (Hyman, 2019).

In this study we explored the molecular basis of host recognition by the podovirus N4 (**Fig. 1a**), a member of the *Schitoviridae* infecting *E. coli* (Wittmann et al., 2020). Selection for mutations conferring N4 resistance (*nfr*) had uncovered genes *nfrA*, *nfrB* and *nfrC* as candidates for phage entry (Kiino and Rothman-Denes, 1989; Kiino et al., 1993a). *nfrA* encodes an outer membrane protein that was described to interact with the tail sheath protein of N4 and might be the terminal receptor for N4 (McPartland and Rothman-Denes, 2008). The *nfrC* allele was mapped to *wecB*, a gene encoding a cytoplasmic UDP-N-acetylglucosamine 2-epimerase that is part of a large gene cluster involved in the synthesis of ECA (Meier-Dieter et al., 1992). Because ECA is the only known exopolysaccharide that depends on WecB, it was proposed that N4 uses ECA as its primary surface receptor to infect *E. coli* (Kiino et al., 1993a, 1993b). Apart from its requirement for N4 and related phages in the *Enquatrovirus* genus of *Schitoviridae*, WecB was recently also shown to be required for infectivity of myoviruses of the *Vequintavirinae* subfamily and their phi92-like relatives (Maffei et al., 2021). Intriguingly, all of these phages encode homologous glycan deacetylase tailspikes, indicating that they target surface glycans in a similar way.

Our results show that bacteriophage N4 and other phages previously linked to ECA do not target the ECA as their primary receptor, but instead use a novel surface glycan of *E. coli* that we call NGR (N4 glycan receptor). We present evidence that NGR is produced and exported by a conserved biosynthesis machinery including WecB, NfrA, and NfrB. Similar to ECA components, the genes encoding this machinery are widespread among enterobacteria and some related groups, providing an elegant explanation for the unusually broad host recognition of N4-like phages. Furthermore, we show that N4 infectivity critically depends on the second messenger c-di-GMP and that this requires the catalytic activity of a single diguanylate cyclase, DgcJ, possibly via a direct and local activation of the NfrB glycosyltransferase. An accompanying manuscript, strengthens this view by demonstrating that NfrB indeed binds c-di-GMP and that DgcJ directly interacts with the presumably glycosyltransferase (Junkermeier and Hengge, 2021). Thus, our study not only

sheds new light on the molecular mechanisms underlying bacteriophage host range but also provides an entry point into understanding how bacteria use local signaling via the second-messenger c-di-GMP to balance costs and benefits of surface glycan expression.

Results

Infection of *E. coli* by phage N4 depends on a putative exopolysaccharide pathway. To analyze the requirements for N4 infection, we first confirmed that chromosomal deletions of the known N4 resistance genes *nfrA*, *nfrB*, and *nfrC* (*wecB*) effectively protect *E. coli* from N4 infection (**Fig. 1b,c**) (Kiino and Rothman-Denes, 1989; Kiino et al., 1993a, 1993b). Since *nfr* mutants were shown to prevent phage adsorption (Kiino and Rothman-Denes, 1989), their products could either directly serve as receptors or could be involved in the production of surface exposed structures that are N4 receptors. Using structure-based protein comparison (Kelley et al., 2015; Zimmermann et al., 2018) and neural network-based structure prediction (Jumper et al., 2021) tools, we identified NfrA and NfrB as potential components of a novel exopolysaccharide secretion system. The N-terminal domain of NfrB shows strong homology to glycosyltransferases such as the cellulose synthase BcsA (Krasteva et al., 2017) (**Fig. 1d**), while the C-terminus contains a domain of unknown function and a small MshE-like domain, a c-di-GMP binding module involved in regulating diverse motor ATPases of type IV pili and type 2 secretion systems (Chou and Galperin, 2016; Wang et al., 2016b) (**Fig. 1d**). The modelled structure of NfrA shows strong homology to exopolysaccharide translocation pores located in the outer membrane of *E. coli* or *P. aeruginosa*, including PgaA (Wang et al., 2016a), BcsC (Acheson et al., 2019), or AlgE (Keiski et al., 2010). The NfrA N-terminus contains several tetratricopeptide repeat (TPR) units, which in other glycan translocation pores were hypothesized to interact with periplasmic polymer-modifying enzymes or with the synthase complex located in the inner membrane (Acheson et al., 2019). The C-terminus of NfrA is a 16-stranded β barrel pore with structural similarities to other glycan translocation pores including PgaA (Wang et al., 2016a) or BcsC (Acheson et al., 2019) (**Figs. 1e, S1**). Based on this, we postulate that NfrB and NfrA are part of a multi-component glycan synthase complex and that their strict requirement for N4 infection may indicate the existence of a novel *E. coli* exopolysaccharide that serves as the primary receptor for N4. Based on this assumption, we term this unknown exopolysaccharide N4 glycan receptor (NGR) and we use N4 infection assays from here on to probe the regulation of the Nfr-mediated NGR biogenesis/secretion.

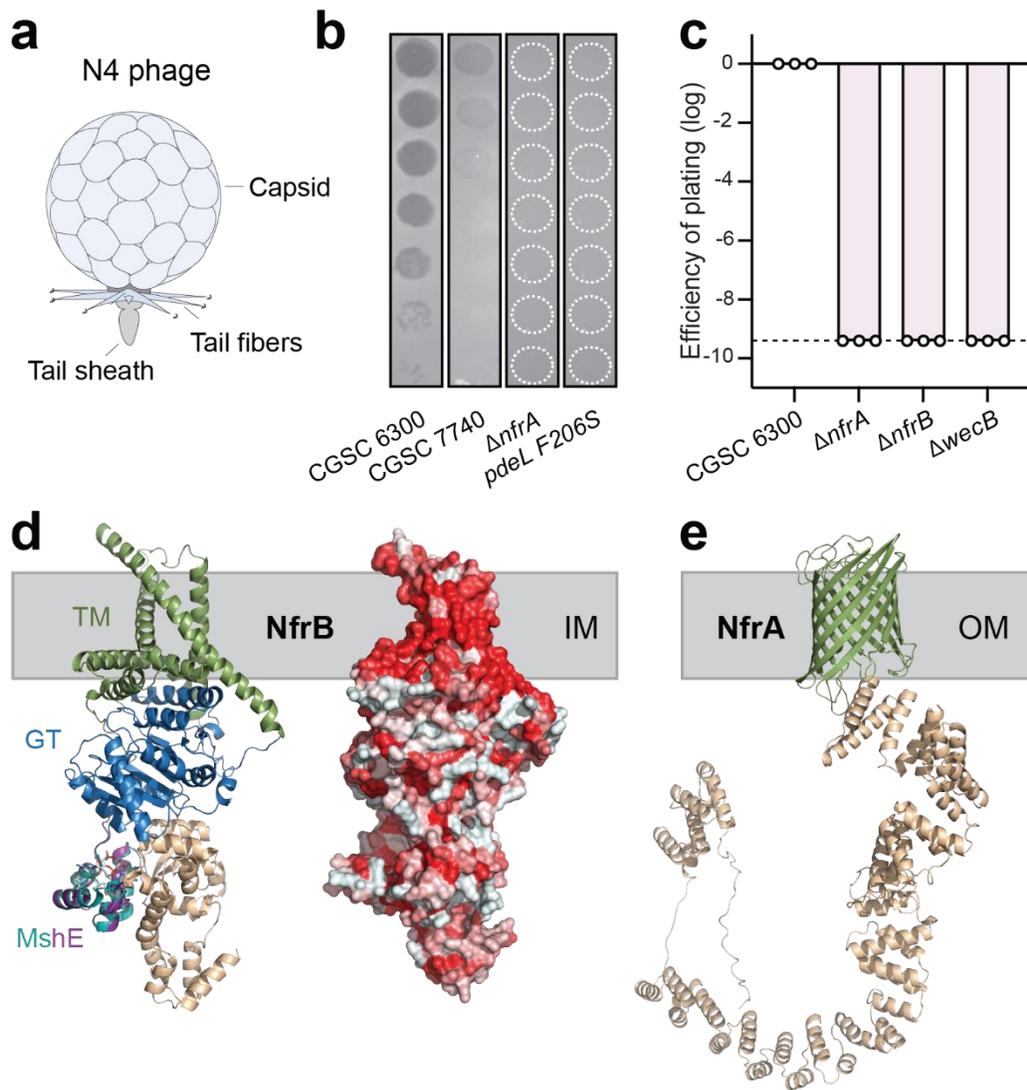


Figure 1: Infection of *E. coli* by bacteriophage N4 requires components of a putative surface glycan secretion system. (a) Schematic of bacteriophage N4. (b) Plaque assay with serial 10-fold dilutions of bacteriophage N4 spotted on lawns of different *E. coli* host strains as indicated. Stippled white circles indicate regions of phage application where no lysis was observed. (c) The efficiency of plating (EOP) is displayed for several *E. coli* host strains as the number of plaque-forming units (PFUs) relative to *E. coli* wild type strain CGSC6300. All mutants are in the CGSC 6300 background. Circles indicate the average of 2 technical replicates of one biological repeat and the bar indicates the mean of the log transformed EOP values. The stippled line marks the detection limit. (d) Model of the structure of NfrB as predicted by AlphaFold (Jumper et al., 2021). Left: colored domains of NfrB with homology to glycosyl transferases (GT, blue), the c-di-GMP binding domain (cyan, purple), and a domain with unknown function (sand). Putative transmembrane helices (TM) incorporated in the inner membrane (IM) are indicated in green. The putative c-di-GMP binding domain of NfrB (purple) is shown as overlap with the c-di-GMP binding domain of the MshE ATPase from *V. cholerae* (Wang et al., 2016b) with bound ligand (teal). Right: Depiction of the NfrB surface with hydrophobic amino acids indicated in red. (e) Structural model of NfrA as predicted by AlphaFold (Jumper et al., 2021). The outer membrane (OM) beta-barrel structure is indicated in green and TPR domains and unstructured regions are shown in sand.

N4 infection requires c-di-GMP. The presence of a MshE domain – typically mediating the allosteric regulation of proteins by c-di-GMP (Wang *et al.*, 2016; Floyd *et al.*, 2020) – in NfrB indicated that its activity may be controlled by c-di-GMP. To test this, we first mutagenized residues that are conserved between NfrB and MshE and that were shown to be involved in c-di-GMP binding (Wang *et al.*, 2016b) and analyzed their effect on N4 infection. This included Leu490, Gly491, Leu505, Leu509, Leu518 and Gly519 (**Fig. 2a**). While most substitutions showed no effect on N4 infection, G491L and G519S abolished N4-mediated killing completely or partially, respectively (**Fig. 2b**). This indicated that an intact MshE domain and c-di-GMP binding to NfrB are required for phage infection.

Next, we investigated N4 infection in several lab adapted strains of *E. coli*. We found that while the original *E. coli* K12 MG1655 strain (CGSC 6300) (Guyer *et al.*, 1981) was susceptible to N4, a closely related hyper-motile variant (CGSC 7740) (Blattner *et al.*, 1997) (**Fig. S2a**) showed strong resistance towards N4 infection (**Fig. 2c,d**). Strain CGSC 7740 carries an *IS1* insertion upstream of *flhDC*, which encodes the master regulator of the flagellar regulon (Barker *et al.*, 2004) (**Fig. S2b**). Because this insertion leads to the constitutive expression of flagellar genes and the phosphodiesterase gene *pdeH*, c-di-GMP levels are substantially reduced in strain CGSC 7740 as compared to strain CGSC 6300 (Reinders *et al.*, 2015). This strengthened the idea that c-di-GMP plays an important role in NGR biogenesis and argued that constitutive expression of *pdeH* may be responsible for N4 resistance of CGSC 7740. However, N4 sensitivity was not restored when deleting *pdeH* in the CGSC 7740 background (**Fig. 2c**) despite the fact that global c-di-GMP levels increased 10-fold and motility was strongly impaired (Reinders *et al.*, 2015). Likewise, restoring the original *flhDC* locus by removing of the *IS1* element, although effectively blocking motility (**Fig. S2b**), failed to restore N4 sensitivity (**Fig. 2d**). Finally, expression of *pdeH* from a plasmid in strain CGSC 6300 provided only limited protection against N4 (**Fig. 2c**). From these experiments, we concluded that although c-di-GMP is required for N4 infection, global changes of c-di-GMP levels do not strongly influence N4-mediated killing of *E. coli*.

The diguanylate cyclase DgcJ regulates Nfr-dependent N4 infection in a highly specific manner. To decipher the molecular determinants responsible for N4 resistance of strain CGSC 7740, we re-examined its chromosome sequences (U00096) and found two additional *IS1* insertions that are not present in strain CGSC 6300 (NC_000913.3).

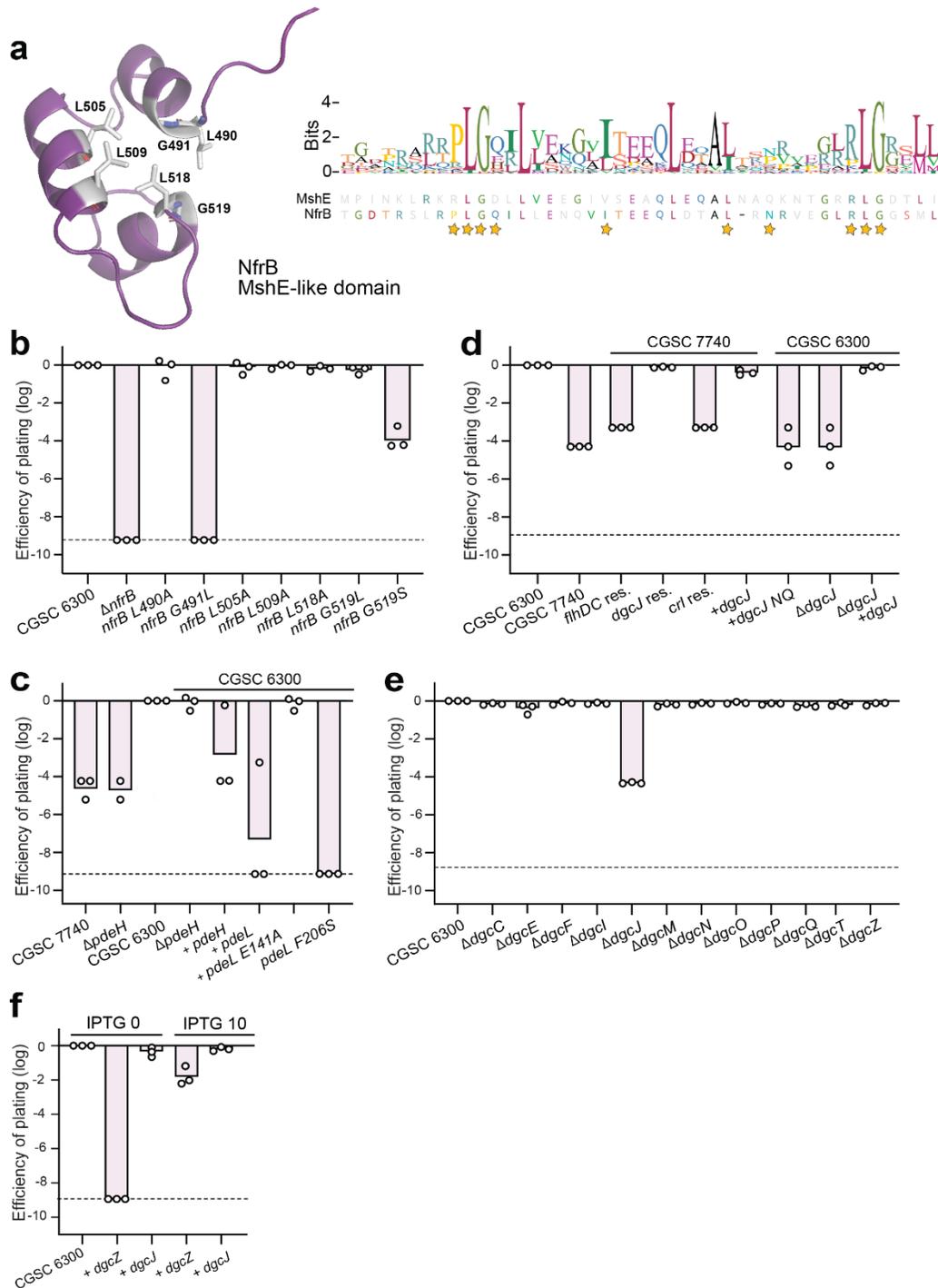


Figure 2: C-di-GMP is required for N4 infection. (a) NfrB harbors a domain with strong homology to a c-di-GMP binding domain of *V. cholerae* MshE. Left: Close-up of a structural model of the putative c-di-GMP binding domain of NfrB (purple). Conserved residues of the c-di-GMP binding pocket (Wang *et al.*, 2016b) that were used for the mutational analysis in (b) are indicated in sticks. Right: ClustalW sequence alignment of the 1000 closest homologues of *E. coli* NfrB using all *Enterobacteriales* but excluding the genus *E. coli*. The conservation of the putative c-di-GMP binding motif of NfrB homologs is shown by the sequence logo with residues of MshE involved in c-di-GMP binding marked by stars. The amino acid sequences of MshE and NfrB are shown below the logo. (b) Conserved residues of the MshE-like domain of NfrB are required for N4 infection. Circles indicate the efficiency of plating (EOP) with N4 phages; displayed for *E. coli* wild type and mutant strains as described in figure 1. The stippled line marks the detection limit. (c) The phosphodiesterase PdeL effectively protects *E. coli* against N4 infection.

EOP is displayed for *E. coli* wild type (strain CGSC 6300 or CGSC 7740) and mutant strains. Strains harboring a plasmid-born copy of *pdeL* wild type, *pdeL* mutant alleles or *pdeH* transcribed from an IPTG inducible promoter (+) are indicated. **(d)** An *IS1* insertion in *dgcJ* is responsible for N4 resistance of *E. coli* strain CGSC 7740. The EOP of phage N4 is displayed for strains CGSC 6300 and CGSC 7740 and for strains with restored wild-type sequences (res.) at the chromosomal loci *flhDC*, *crl* and *dgcJ* as indicated. Strains containing chromosomal deletions of *dgcJ* and strains carrying *dgcJ* alleles on a plasmid (+) are indicated, whereas *dgcJ* NQ indicates a catalytically inactive *dgcJ* allele. **(e)** Specific requirement of the DgcJ diguanylate cyclase for N4 phage infection. The EOP of phage N4 is displayed for *E. coli* wild type and deletion strains lacking individual diguanylate cyclases. **(f)** Ectopic expression of *dgcJ* sensitizes *E. coli* towards N4 infection in a *dgcJ/dgcQ* knockout strain. Expression of *dgcJ/dgcZ* was induced from the lactose promoter with 0 or 10 μ M IPTG.

Insertions mapped to *crl*, a gene encoding an activator of the stress sigma factor RpoS (Typas et al., 2007), and to *dgcJ*, which codes for one of several diguanylate cyclases of *E. coli* (Hengge et al., 2015) (**Fig. S2d**). Replacing *IS1* elements in strain CGSC 7740 individually with the corresponding chromosomal wild-type sequences from strain CGSC 6300 (**Fig. S2b**) revealed that N4 sensitivity was only re-established upon restoring *dgcJ* but not when the *flhDC* or *crl* loci were restored (**Fig. 2d**).

These findings indicated that DgcJ is a main driver of sensitivity to N4 infection. In line with this, deleting *dgcJ* in strain CGSC 6300 provided strong protection against N4, while deleting any other *dgc* gene in this background showed no effect (**Fig. 2e**). Ectopic expression of *dgcJ* restored N4 sensitivity of both the Δ *dgcJ* mutant in the CGSC 6300 background and of strain CGSC 7740 (*dgcJ::IS1*). In contrast, expression of *dgcJ* (*DE425NQ*) a mutant allele encoding a catalytically inactive variant of DgcJ, failed to restore phage sensitivity in strain CGSC 7740 (**Fig. 2d**). Ectopic expression of *dgcJ* also restored N4 susceptibility in a *nfrB G491L* mutant background, arguing that this mutation indeed compromised c-di-GMP binding to NfrB, a phenotype that is likely compensated by increasing the levels of this highly specific diguanylate cyclase (**Fig. S3**). Finally, basal level expression of *dgcJ* readily restored N4 sensitivity, while expression of *dgcZ*, a gene encoding a highly active diguanylate cyclase from *E. coli* (Zähringer et al., 2013), failed to restore N4 sensitivity in strain CGSC 7740 (**Fig. 2f**), despite of its potent inhibition of *E. coli* swimming motility under the same conditions (**Fig. S2c**) (Boehm et al., 2010). DgcZ could, however, partially restore N4 sensitivity when its transcription was increased by the addition of IPTG (**Fig. 2f**).

The above results indicated that DgcJ is a critical determinant for N4 infection of *E. coli* that activates NGR biogenesis in a highly specific manner. DgcJ is a membrane protein (Sueki et al., 2020) with a periplasmic dCache domain and a cytoplasmic catalytic GGDEF domain (**Figs. 3a, S2d**). The dCache domain of DgcJ is closely related to the periplasmic domain of the methyl-

accepting chemotaxis protein PctA from *Pseudomonas aeruginosa*, which was crystallized in complex with its amino acid ligands L-Met, L-Trp, and L-Ile (Gavira et al., 2020). Some residues involved in ligand binding (Y121, Y144, D146, D173) are conserved in DgcJ (Y168, Y210, D212, and D239) (**Fig. 3a**), indicating that DgcJ may bind similar ligand(s) via its periplasmic dCache domain. Consistent with this idea, isosteric substitutions of potential ligand-binding residues of DgcJ (Y210F, D212N, D239N) invoked strong protection against phage N4, similar to levels observed for the $\Delta dgcJ$ (**Fig. 3b**). Also, *E. coli* K12 MG1655 (CGSC 6300) was resistant to phage N4 when grown on defined media containing glycerol as sole carbon source, but was readily killed by N4 when grown in defined media supplemented with casamino acids (**Fig. 3c**). When testing amino acids individually, we found that the addition of arginine to minimal media restored phage infection in minimal medium (**Fig. 3c**). However, Arg-induced N4 killing under these conditions was not dependent on DgcJ, as supplementation of minimal glycerol media with arginine or with casamino acids also restored phage susceptibility in a $\Delta dgcJ$ mutant (**Fig. 3c**). From this, we concluded that arginine promotes N4 phage infection, possibly by activating a second diguanylate cyclase that promotes N4 infection specifically under these conditions.

To identify this second DGC, we generated all possible double mutant combinations lacking DgcJ and each of the other diguanylate cyclases of *E. coli*. This identified DgcQ as an additional diguanylate cyclase involved in N4 infection (**Fig. 3d**). While a *dgcJ* single mutant reduced N4 infection to intermediate levels in complex media, plaque formation was reduced below the detection limit in a $\Delta dgcJ \Delta dgcQ$ double mutant, similar to mutants lacking NfrA or NfrB. Surprisingly, a *dgcQ* single mutant was fully susceptible to N4 in complex media (**Fig. 3d**), arguing that it has an auxiliary role in activating the NGR pathway. DgcQ is a homolog of the diguanylate cyclase STM1987 from *Salmonella* Typhimurium, which was shown to sense arginine (Mills et al., 2015). Consistent with this, DgcQ was strictly required and sufficient for infection of *E. coli* by phage N4 in minimal media supplemented with arginine (**Fig. 3c**). These experiments demonstrated that DgcQ can compensate the lack of DgcJ activity in minimal media in response to extracellular arginine. Importantly, the addition of casamino acids to minimal media also re-potentiated DgcJ (**Fig. 3c**), suggesting that the dCache domain of DgcJ recognizes a nutritional signal that is contained in casamino acids.

Finally, we observed that *E. coli* was considerably less sensitive to N4 infections when grown at 30°C as compared to 37°C (**Fig. 3e**). Ectopic expression of *dgcJ* fully restored sensitivity to phage N4 at 30°C, arguing that DgcJ levels may be limiting at 30°C. In line with this, *dgcJ*

transcription was significantly reduced at 30°C as compared to 37°C (**Fig. 3e**). Together, these results demonstrate that DgcJ expression and activity are stimulated by elevated temperatures and sensing of yet unknown ligands.

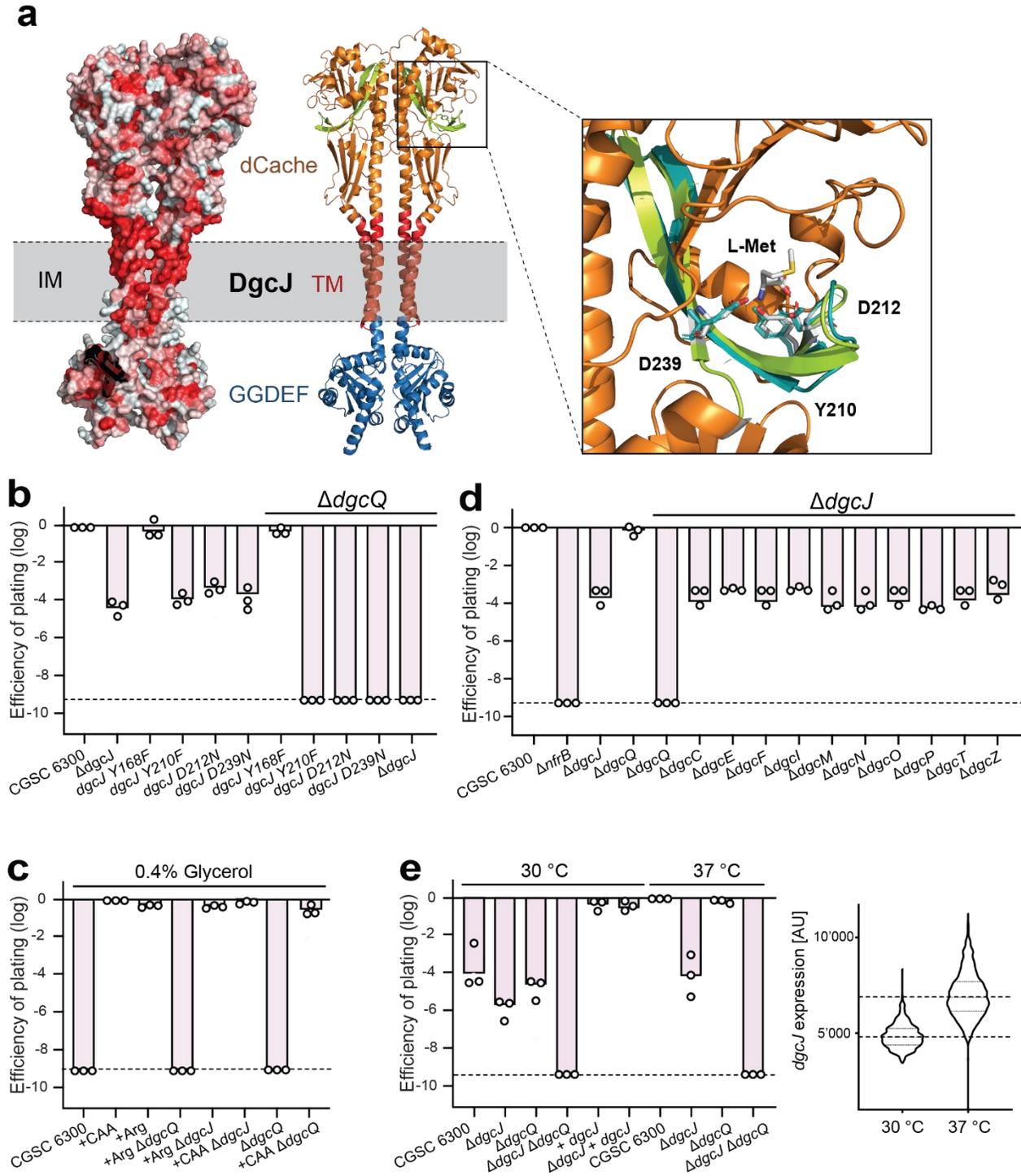


Figure 3: DgcJ and DgcQ, two homologous diguanylate cyclases that specifically sensitize *E. coli* for N4 infection. (a) Model of the dimer structure of DgcJ as predicted by AlphaFold (Jumper *et al.*, 2021). Left: hydrophobicity is displayed by red coloring, highlighting the TM domain. Right: Colored domains of DgcJ with homology to

dCache domain (orange) with the putative ligand-binding domain (LBD, green) and the GGDEF domain (blue). Putative TM helices incorporated in the inner membrane (IM) are indicated in red. The close-up view of the putative LBD reveals homology of the DgcJ residues Y210, D212 and D239 (displayed as sticks) to the ligand-interacting residues (teal) of PctA (*Gavira et al., 2020*) coordinating L-methionine. **(b)** Isosteric mutations of putative ligand-binding residues in DgcJ phenocopy a *dgcJ* knockout. Strains carrying an additional *dgcQ* mutation are indicated above. Circles indicate the efficiency of plating (EOP) with N4 phages is displayed for *E. coli* wild type and mutant strains as described in Fig. 1. **(c)** Phage N4 infection requires extracellular amino acids. Phage infection assays were performed as described before at 37°C, but in MOPS minimal medium supplemented with 0.4% glycerol and, if indicated, with 0.4% casamino acids or arginine. While no N4 infection could be observed in glycerol minimal medium, supplementation with casamino acids completely rescued phage susceptibility. Arginine restored N4 susceptibility in a *dgcQ* dependent manner **(d)** DgcJ specifically sensitizes *E. coli* for N4 infection in combination with DgcQ. Double-mutants of DgcJ with other cyclases do not change N4 infection except a combination with DgcQ. EOP is determined as in (b). **(e)** N4 infection and *dgcJ* expression is reduced at lower temperatures. Left: Phage infection assays were performed as described before but either at 30°C or 37°C as indicated. Right: A plasmid-borne fluorescent *dgcJ* promoter reporter was used to read out *dgcJ* expression with the microscope. Y-axis indicates the pixel intensity from the fluorescent cells. Violin plots contain >1000 individually quantified cells.

The PdeL phosphodiesterase efficiently protects *E. coli* against N4 phage infection. From the experiments above we concluded that c-di-GMP binds to NfrB to stimulate secretion of the NGR exopolysaccharide and that this process is regulated by the diguanylate cyclases DgcJ in a highly specific manner. The reduction of the global c-di-GMP pool through the constitutive expression of the phosphodiesterase PdeH did not effectively protect *E. coli* from N4 infections. Surprisingly, we found that expression of the phosphodiesterase gene *pdeL* from a plasmid phenocopied the N4 protection level observed for the $\Delta dgcJ$ mutant (**Fig. 2c**). In contrast, expression of *pdeL* *E141A* encoding a catalytic inactive variant had no protective effect (**Fig. 2c**). Moreover, replacing the chromosomal wild-type copy of *pdeL* with the *pdeL* allele F206S encoding a constitutively active PdeL variant (Reinders et al., 2015), provided complete protection against N4, similar to mutants lacking NfrA or NfrB (**Figs. 1b, 2c**).

The observation that PdeL, but not PdeH, is able to effectively protect *E. coli* against N4 led us to investigate the molecular details of PdeL specificity. We have shown previously that PdeL is both an active phosphodiesterase and a c-di-GMP dependent transcription factor that autoregulates its own expression (Reinders et al., 2015). We thus hypothesized that PdeL influences N4 infection through a combination of effectively lowering c-di-GMP levels and regulating the transcription of genes involved in N4 infection. To define additional promoters regulated by PdeL, we performed ChIP-Seq experiments using a strain expressing HA-tagged PdeL from the chromosome. These experiments not only confirmed that PdeL binds to the *pdeL* promoter region, but also identified eight additional binding sites that were mapped to the promoter regions of *cstA*, *fruB*, *xanP/gltS*, *sufA*, *wecA*, *yafC/yafD*, *sslE*, and *yqaB* (**Fig. S4**). While several of these genes encode components involved in nutrient scavenging and uptake (CstA, pyruvate uptake;

FruB, fructose uptake; XanP, xanthine uptake; GltS, glutamate uptake; SslE: mucin degradation), we focused our attention on the *wecA* promoter, which drives a large 12-gene operon involved in the synthesis of enterobacterial common antigen (ECA), a complex glycan polymer associated with the cell surface of *Enterobacteriales* (**Fig. 4a**) (Rai and Mitchell, 2020a).

The third gene of the *wecA* operon is *wecB* (*nfrC*), which was shown to be strictly required for N4 infection of *E. coli* (Kiino et al., 1993a). To test if the ECA glycan polymer serves as primary receptor for phage N4, we analyzed the contribution of other *wec* genes to N4-mediated killing. Strains containing defined chromosomal deletions of *wecB* or of the entire ECA operon (Δ *wecA*) were indeed resistant to N4 infection. However, deletion of the *wecA* gene alone showed no effect (**Fig. 4b**). This was surprising as *wecA* encodes the undecaprenyl-phosphate α -N-acetylglucosaminyl transferase, which catalyzes the initial step of O-antigen and ECA biogenesis (Meier-Dieter et al., 1992). Importantly, expression of *wecB* alone from a plasmid fully restored N4 susceptibility of the Δ *wecA* strain (**Fig. 4b**). This excludes the ECA as primary receptor for N4 and argues that UDP-ManNAc, the product of the WecB-mediated epimerase reaction, serves as precursor for the as yet uncharacterized NGR glycan polymer. In line with this, *wecB* K15A, encoding a catalytically inactive WecB variant (Samuel and Tanner, 2004), failed to restore N4 susceptibility (**Fig. 4b**).

Together, this implied that PdeL modulates *wecB* expression and that this may contribute to its strong protective effect against N4. To test this, we determined the exact binding site of PdeL in the *wecA* promoter region using surface plasmon resonance (SPR) and short overlapping DNA sequences covering the entire *wecA* promoter region (**Fig. 4a**). Strong binding of PdeL (K_D 83 nM) was observed for a region overlapping two of the four transcription start sites upstream of *wecA* (Salgado et al., 2013) (**Fig. 4c,d**). This region contains a short sequence with similarity to the PdeL binding site upstream of the *pdeL* promoter (Reinders et al., 2015) and is positioned upstream of the binding site for NsrR, the only known transcription factor of the *wec* gene cluster (Partridge et al., 2009) (**Fig. 4a**). To examine if PdeL influences *wecA* promoter activity, we engineered a strain carrying a reporter for *wecA* transcription on the chromosome. While *wecA* promoter activity was increased in a Δ *pdeL* mutant, it was reduced below wild-type levels in a strain expressing *pdeL* from a plasmid (**Fig. 4e**). Together, this suggested that PdeL is a repressor of the *wec* operon and argued that it contributes to N4 resistance by reducing c-di-GMP and as a transcriptional repressor of *wecB*, thereby limiting the availability of key components or precursors of the NGR glycan polymer.

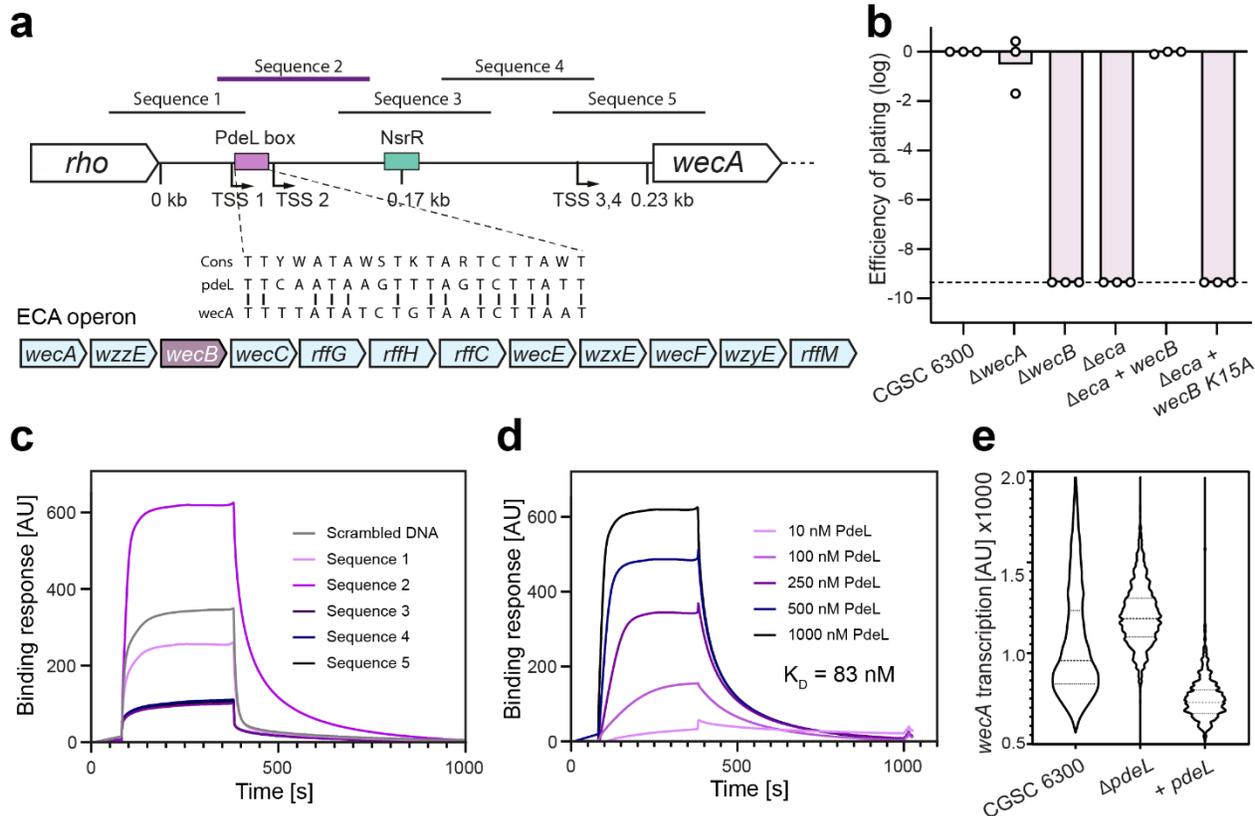


Figure 4: PdeL is a transcriptional repressor of the ECA operon. (a) Schematic of the *wecA* promoter region and the entire ECA operon. Transcriptional start sites (TSS) are based on (Salgado *et al.*, 2013). The binding site of NsrR repressor and of PdeL are indicated in green and purple, respectively. The sequence of the putative PdeL binding site and homology to the PdeL binding site in the *pdeL* promoter region (Reinders *et al.*, 2015) are indicated. Sequences 1-5 mark the five DNA fragments used for in-vitro binding studies using PdeL. The position of the *wecB* gene in the ECA cluster is highlighted in purple. (b) The *wecB* gene is the only gene of the ECA cluster required for N4 infection of *E. coli*. EOP is displayed for *E. coli* wild type (CGSC 6300) and mutants with deletions in *wecA*, *wecB* or the entire ECA gene cluster (*eca*) as described in figure 1. Plasmid-born copies of *wecB* alleles used for complementation of the Δ *eca* strain are indicated. Expression was induced with 1 mM IPTG. (c) PdeL binding to different fragments of the *wecA* promoter region as determined by surface plasmon resonance (SPR). DNA fragments (see: a) were individually immobilized on SPR chips and purified PdeL (1 μ M) flushed through the flow cell. The amount of bound PdeL is shown on the y-axis. PdeL was added 60 secs after the start of recording and after 420 secs the flow cell was flushed with buffer, resulting in PdeL dissociation from the DNA. Sequence 2 showed the strongest binding and the slowest dissociation, indicating the formation of a stable DNA-protein complex. (d) PdeL binds to the *wecA* promoter with high affinity. Experiments as outlined in (c) were carried out with immobilized DNA sequence 2 and with increasing concentrations of PdeL. The maximal binding response at around 300 secs was used to calculate the binding affinity. (e) PdeL represses the ECA operon. Activity of the *wecA* promoter was determined with a strain carrying a transcriptional *mCherry* reporter downstream of *wecA* on the chromosome. Violin plots show fluorescence distribution of at least 3000 individual cells of *E. coli* wild type and mutant strains as indicated.

The N4 infection mechanism is widely conserved in pathogenic and non-pathogenic *E. coli*.

N4-like phages of the family of *Schitoviridae* were shown to infect alpha-, beta-, and gammaproteobacteria (Wittmann *et al.*, 2020). To investigate the phylogenetic distribution of proteins facilitating N4 infection, we screened 1688 bacterial genomes of the OMA database (Altenhoff *et al.*,

2020) for co-occurrence of genes encoding NfrA, NfrB, WecB and DgcJ. This analysis revealed that these genes are strongly conserved in *E. coli*. However, some *E. coli* strains (e.g., all *E. coli* B strains) carry insertions elements, deletions or premature stop codons in *nfrB* or *dgcJ*, indicating that expressing these genes under laboratory conditions is associated with fitness costs.

Because most bacteriophages display a narrow host range (Nobrega *et al.*, 2018), the remarkable conservation of N4-associated proteins in *E. coli* prompted us to investigate if pathogenic *E. coli* strains are also susceptible to bacteriophage N4 and if so, if that process depends on c-di-GMP and on components of the Nfr pathway. We chose the uropathogenic *E. coli* K1 strain UTI89, which in contrast to the K12 strain MG1655 produces O-antigen and group 1 capsular polysaccharides (Goh *et al.*, 2017), surface structures that can provide effective phage protection. As shown in **Fig. 5a**, UTI89 was indeed infected by phage N4, a process that was dependent on *wecB*, but not on *wecA* (ECA, O-antigen) or *kpsT* (capsule) genes. Also, expression of *pdeL* from a plasmid resulted in complete phage protection (**Fig 5a**). Together, this indicated that the primary receptor for bacteriophage N4 is widely conserved in *E. coli* and that surface exposure of the N4 glycan receptor follows similar regulatory logic in the pathogenic UTI89 strain as in the lab adapted, non-pathogenic *E. coli* K12 MG1655.

In other gram-negative bacteria homologs of NfrA, NfrB, WecB and DgcJ are sporadically encoded making it difficult to assess their overall conservation and role. To identify phage representatives that exploit similar structures on the surface of their respective prey bacteria, we analyzed the available genome sequences of N4-like phages (Wittmann *et al.*, 2020). While most proteins of N4-like phages are strongly conserved, tail fibers (Gp64) and tail sheaths (Gp65) are highly variable, reflecting the diversity of surface receptors of their prey (McPartland and Rothman-Denes, 2008). BLAST analysis with *gp64* and *gp65* from N4 on the family of *Schitoviridae* (Wittmann *et al.*, 2020) identified a phage subgroup with conserved tail fibers and sheaths (**Table S1**). This included a group of N4-like phages belonging to the subfamily of *Rothmandenesvirinae* infecting *Achromobacter xylosoxidans*, an opportunistic human pathogen and member of the betaproteobacteria that causes a wide range of infections including bacteremia, meningitis, urinary tract infections, endocarditis, or pneumonia (Hansen *et al.*, 2006; Namnyak *et al.*, 1985; Rodrigues *et al.*, 2017; Tena *et al.*, 2009). Intriguingly, in *A. xylosoxidans* *nfrA*, *nfrB* and *wecB* are cluster together with additional genes encoding putative components of exopolysaccharide biogenesis, modification and secretion. This includes a homolog of BcsB, a component of the cellulose synthase complex (Acheson *et al.*, 2019) and homologs of periplasmic O-acetyltransferases

(AT) involved in glycan polymer modification (Chanasit et al., 2020; Sychantha et al., 2018) (Fig. S5). This raises the intriguing possibility that the products of these genes are functionally linked and that NGR-like glycan polymers may serve as surface receptor for N4-like phages in a diverse range of bacterial pathogens.

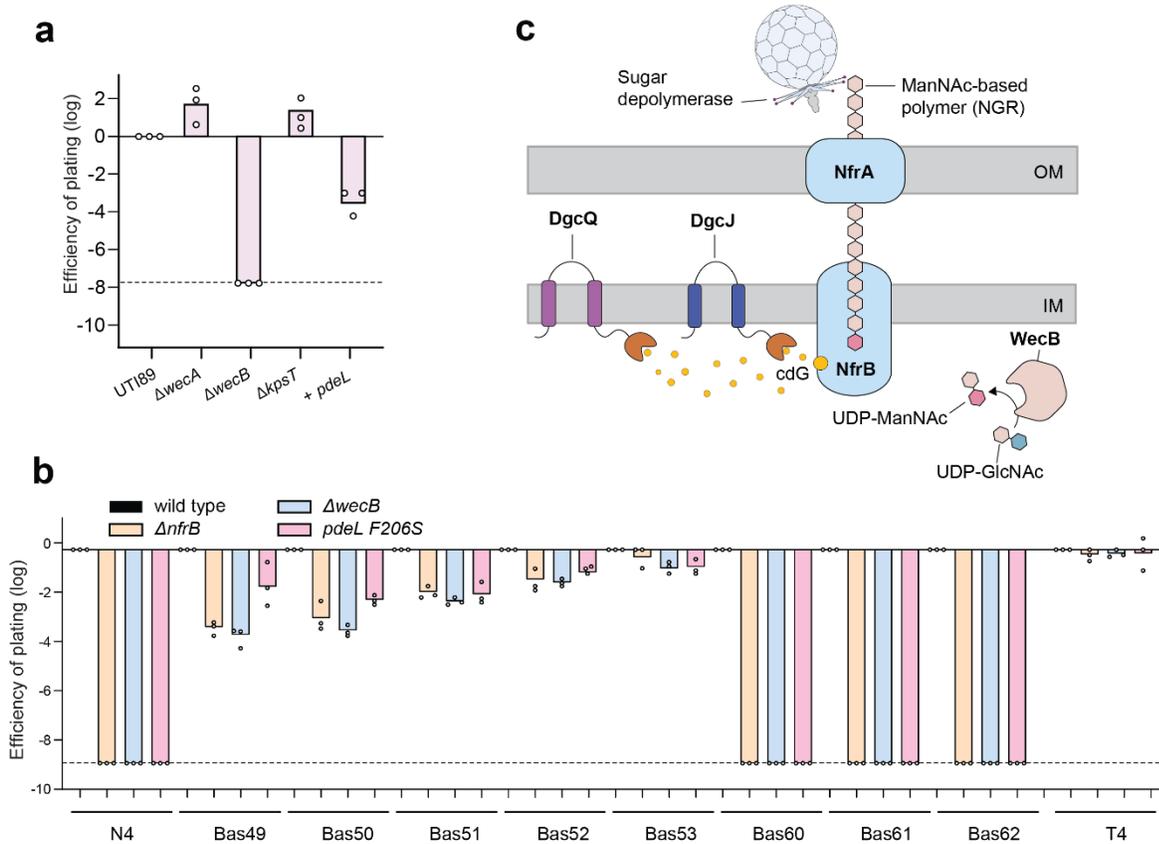


Figure 5: NGR is a conserved phage receptor in the pathogenic *E. coli* strain UTI89. (a) Phage N4 infections with the uropathogenic *E. coli* strain UTI89 were as described in Materials and Methods. (b) Phages from the BASEL phage collection (Maffei et al., 2021) that were shown to depend on *wecB* were used to infect *E. coli* CGSC 6300 wild type and selected mutant strains as indicated. (c) Model for NGR regulation and N4 adsorption (see text for details).

Recent work on the BASEL collection, a representative set of isolates from all major groups of *E. coli* phages, showed that phages of the *Vequentavirinae* subfamily of *Myoviridae*, their phi92-like relatives, and N4 as well as a close relative within *Enquatrovirus* depend partially or completely on *wecB* for infectivity on the *E. coli* K12 host (Maffei et al., 2021). Since WecB had previously been described as being specifically required for the ECA but for no other surface glycan of *E. coli* (Rai and Mitchell, 2020a), these results were seen as evidence for a role of ECA as primary receptor of these phages. However, our finding that at least phage N4 does not use

ECA but possibly a new surface glycan, NGR, prompted us to revisit this interpretation. As expected, susceptibility of *E. coli* K12 MG1655 CGSC 6300 to all tested isolates of *Vequintavirinae*, their phi92-like relatives, and *Enquatrovirus* was reduced or even abolished by a *wecB* knockout (**Fig. 5b**). However, none of these phages required WecA or any other part of the *wec* operon, except for *wecB* (**Fig. S6**). Also, a $\Delta nfrB$ mutation or the *pdeL* F206S allele showed full or partial resistance, exactly phenocopying the $\Delta wecB$ mutant host (**Fig. 5b**). These results strongly suggest that none of the WecB-dependent phages use ECA as their primary receptor but that all of them instead target the NGR.

Discussion

NGR, an enterobacterial surface glycan commonly exploited as a phage receptor. Our results show that N4 and other phages previously thought to target the ECA to infect *E. coli* use a novel surface-associated polysaccharide that we call NGR (N4 glycan receptor) as their primary receptor. We suggest that biosynthesis and export of NGR depend on a conserved machinery including WecB as well as NfrB and NfrA, which share strong homology to known polysaccharide export systems (**Fig. 5c**). The conservation of this *bona fide* NGR biosynthesis machinery may well explain the remarkably broad host recognition among phages targeting this surface structure among enterobacteria (Maffei et al., 2021). This raises the question why ECA does not seem to be used by phages to a similar extent because, to the best of our knowledge, there is no remaining bacteriophage thought to target this surface glycan. One possibility could be differences in expression or insufficient size of the ECA to be useful as a receptor and, therefore, availability of these glycans as a receptor in the true habitats of enterobacteria. We anticipate that future studies exploring the biology of these elusive polysaccharides might also help to understand their very different use as bacteriophage receptors.

Based on the observation that the sugar epimerase WecB is essential for phage infection, we propose that NGR is a polymer containing the monosaccharide N-acetylmannosamine (ManNac) that is produced by this enzyme. This is in line with *wecB* being the only gene of the ECA cluster required for N4 infection and the only EPS biosynthesis gene ever identified in genetic screens for N4 resistance (Kiino et al., 1993a; Mutalik et al., 2020). Uncovering the molecular identity of NGR, its overall conservation and the details of its biosynthesis and export will require further studies. Notably, ManNac-based EPS components are widely used by other bacteria. For instance, *Neisseria meningitidis* serogroup A produces a poly-ManNac capsule, which serves as a primary virulence factor to promote host colonization and serum resistance (Calloni et al., 2018; Liu et al., 1971). Similarly, neuro-invasive *E. coli* K1 (Tanner, 2005) requires ManNac as a building block for sialic acid, the repeat unit of their capsule (King et al., 2007). In *E. coli* K1 strains, UDP-GlcNAc is converted to ManNac by the epimerase NeuC, but this mechanism is catalytically distinct from WecB (Vann et al., 2004). The observation that deletion of the *wecB* gene in the K1 strain UTI89 leads to complete N4 resistance argues that the substrate for the polymerization of NGR is provided by WecB and not by NeuC.

The specific role of NGR remains unclear. Its widespread distribution and conserved surface exposure despite of potentially strong selection by phage predators argues that it has a vital

role in *E. coli* and *Enterobacteriales*. Similarly, the functional significance of ECA has remained enigmatic. Mutants lacking ECA are more sensitive to different forms of stress, have increased outer membrane permeability and show reduced virulence (Rai and Mitchell, 2020b). It is possible that NGR has similar protective roles in *E. coli* and its relatives or is involved in bacteria-host interaction. A specific role for NGR in the host environment is supported by the observation that *dgcJ* expression limits NGR biogenesis and N4 infection at 30°C but increases strongly when cells are shifted to 37°C. NGR expression in the host could protect from phagocytosis, similar to capsules, or through masking patterns on the bacterial surface to avoid recognition from the immune system. Alternatively, NGR may have evolved in non-pathogenic members of the microbiota to avoid encounters by the immune system. This would be in line with strong structural conservation of the NGR surface glycan because the structure would be defined through receptor binding to the eukaryotic cells in the gut.

NGR biogenesis is controlled by local c-di-GMP signaling. Based on its strong homology to membrane-embedded glycosyltransferases, we propose that NfrB is responsible for NGR polymerization (**Fig. 5c**). We propose that NfrB activity is regulated by c-di-GMP binding to the MshE-like domain, which is positioned adjacent to the glycosyltransferase domain. This is reminiscent of the *E. coli* cellulose synthase BcsA, in which the catalytically autoinhibited state is released by binding of c-di-GMP to an associated PilZ domain (Acheson et al., 2021; Morgan et al., 2014). Importantly, BcsA is specifically regulated by the diguanylate cyclase DgcC and the phosphodiesterase PdeK, which interact directly with the cellulose synthase complex. It was proposed that this arrangement provides a target-specific pool of c-di-GMP to locally boost the activity of BcsA, thereby sequestering the regulation of the cellulose synthase complex from global fluctuations of the second messenger (Richter et al., 2020). Our findings indicate that NfrB activity may be controlled in a similar manner. Not only did large changes of the global c-di-GMP concentration not affect N4 susceptibility (despite of their strong effect on other c-di-GMP-dependent cellular functions like motility), but also, N4 infection was critically dependent on *dgcJ* (but none of the other *dgc* genes) even in the hypermotile strain CGSC 7740, which harbors low levels of c-di-GMP. Thus, DgcJ likely acts as ‘local pacemaker’ (Ross et al., 1987) to specifically regulate NGR biogenesis in response to environmental cues (**Fig. 5c**). This idea is supported by a parallel study showing that c-di-GMP indeed binds to the MseH domain of NfrB and also demonstrating a direct interaction between NfrB and DgcJ (Junkermeier and Hengge, 2021).

It is possible that *E. coli* uncouples the regulation of NGR biogenesis from global cellular changes of c-di-GMP to avoid directly linking the exposure of this surface glycan to general life-style changes mediated by c-di-GMP (Hengge, 2021; Jenal et al., 2017). This would not only increase the precision of NfrB regulation, but would limit NGR surface exposure to appropriate and highly distinct environmental conditions and thereby strongly reduce the selective pressure exerted by a diverse range of bacteriophages that exploit strongly conserved surface structures as receptors. To better understand this ‘Achilles’ heel’ strategy it will be interesting to identify the environmental cues that activate DgcJ and stimulate NGR surface exposure. The periplasmic dCache domain of DgcJ shows homology to chemosensory domains of *P. aeruginosa*, which bind amino acids and autoinducer-2 (Gavira et al., 2020; Zhang et al., 2020). DgcJ may well respond to similar nutritional or cell density-related cues, which again raises the question of NGR functionality and why *E. coli* would need to expose the NGR glycan under such highly specific conditions. The role and specificity of DgcQ is less clear as it seems to have a auxiliary function in activating NGR biogenesis in response to the presence of arginine in the growth medium. Although we cannot exclude that DgcQ locally cooperates with DgcJ, we find it more plausible that it supports NGR biogenesis by increasing the global c-di-GMP pool of *E. coli*.

While the specific role of DgcJ can be explained by its spatial coupling to the NGR biosynthesis machinery (Junkermeier and Hengge, 2021), the highly effective role of PdeL in protecting *E. coli* from N4 lysis is likely due to its dual function as a phosphodiesterase and as a transcription factor (Reinders et al., 2015). Although the catalytic activity of PdeL is clearly important for N4 protection, PdeL is a transcription factor and thus unlikely to act in a spatially confined compartment. Rather, we speculate that PdeL interferes with NGR biogenesis by repressing *wecB* transcription and, consequently, by limiting WecB-dependent ManNac supply for NGR biosynthesis. The prominent role of PdeL in regulating NGR and N4 phage sensitivity may also relate to its bimodal expression that was shown to generate distinct *E. coli* subpopulations with high and low levels of c-di-GMP, respectively (Reinders et al., 2021). We therefore speculate that PdeL bimodality may be part of a bet-hedging mechanism that specifically protects a fraction of clonal bacterial populations from phage predation by preventing the production of NGR during conditions that would induce NGR biosynthesis.

Material & Methods

Bacterial strains and growth conditions.

The bacterial strains and plasmids used in this study are listed in table S1. *E. coli* K12 MG1655 *wildtype* strains were ordered from the Coli Genetic Stock Center and indicated with their accession number. Strains were grown in glass culture tubes with agitation at 170rpm on 37°C or 30°C. When needed antibiotics were present at following concentrations: 50 µg/mL kanamycin, 30 µg/mL for low or single-copy plasmids.

Culture media and solutions

Lysogeny Broth (LB) was prepared by dissolving 10 g/l tryptone, 5 g/l yeast extract and 10 g/l NaCl in Milli-Q H₂O. LB agar plates were prepared by supplementing LB medium with 1.5% (w/v) agar (AppliChem). Top agar was prepared by supplementing LB containing 0.4 % agarose with 20 mM MgSO₄ and 5 mM CaCl₂ and stored at 60°C for up to 4 weeks.

MOPS defined medium was prepared according to (Neidhardt, Bloch and Smith, 1974) in Milli-Q H₂O. Phosphate-buffered saline (PBS) was prepared as a solution containing 1.44 g/l Na₂HPO₄, 0.24 g/l KH₂PO₄, 0.2 g/l KCl and 8 g/l NaCl in Milli-Q H₂O adjusted to a pH of 7.4 using 10 M NaOH. SM buffer was prepared as a solution of 0.1 M NaCl, 10 mM MgSO₄ and 0.05 M Tris-HCl (pH 7.5). PdeL purification buffer contains 200 mM NaCl, 5 mM MgCl₂, 5 mM DTT and 20 mM Tris-HCl (pH 8.0). DNA hybridization buffer contains 150 mM NaCl and 15 mM trisodium citrate in Milli-Q H₂O adjusted to a pH of 7.

Chromosomal gene deletions and modifications

Gene deletions were carried out either by transduction from the Keio collection (Baba *et al.*, 2006) or using λ-red homologous recombination as described by (Datsenko and Wanner, 2000) using pKD46. Selection markers were removed by site-specific recombination using pCP20 (Datsenko and Wanner, 2000).

Plasmid construction

Plasmids were constructed either using classic restriction-based molecular cloning or Gibson Assembly (Gibson *et al.*, 2009). Plasmids were transformed into *E. coli* DH5α and purified using the GenElute Miniprep kit from Sigma-Aldrich.

Microscopy

Bacteria were grown to exponential phase in LB at 37°C and transferred on a 0.75 mm thick agarose pad containing PBS and 1% agarose. Images were acquired using an Eclipse Ti2 inverted microscope (Nikon) equipped with an ORCA-Flash4.0 CMOS camera C11440-22C (Hamamatsu), and an CFI PlanApo DM 100x Lambda Oil/1.45/0.13 objective (Nikon). Brightfield images were illuminated using the High-Power LED-100 Illumination system (Nikon) at 50 ms exposure time. Fluorescence of GFPmut2 was acquired at 470/24 nm with 100 ms exposure time. The open-source software Oufiti (Paintdakhi *et al.*, 2016) was used for automatic cell detection and WHISIT (Sprecher *et al.*, 2017) to quantify the fluorescence intensity.

Swarming assay

Swarming assays were performed as described in (Girgis *et al.*, 2007) with small modifications. In brief, 2.5 µl *E. coli* overnight culture was transferred on a plate containing 10 g/l tryptone, 5 g/l NaCl and 0.3 % agar (AppliChem) and incubated at 37°C. After 7 - 15 hours, swarm size was measured and relative swarming was calculated using the *wildtype* as reference.

Phage lysate preparation

P1 phage lysate preparation and transduction were performed as described in (Thomason, Costantino and Court, 2007). N4 and T5 phage lysates were prepared as described in (Maffei *et al.*, 2021) and stored in SM buffer.

Phage infection assay

Phage infections were adapted from (Maffei *et al.*, 2021). In brief, 100 µl *E. coli* overnight culture was mixed with 3ml top agar and poured on a LB agar plate pre-warmed to 60°C. The top agar solidifies after 15 min at room temperature, allowing to spot 2.5 µl of a 10-fold serial diluted phage solution on the double-agar overlay plate. After the spots were dried, the plate was incubated at 37°C. Plaque forming units (PFUs) were counted after 12 - 18 hours to calculate the efficiency of plating (EOP).

Protein purification

PdeL-6xHis was expressed from pET28a in BL21 cells grown at 22°C for 5h in 2l LB. Cells were harvested by centrifugation at 6000 g for 30 min at 4°C. The cell pellet was resuspended in 10 ml

PdeL purification buffer and one tablet of cOmplete mini EDTA-free protease inhibitor (Roche) and a spatula tip of DNase I (AppliChem) was added to the cell suspension. Cells were lysed by 3 passages of French press and the lysate was cleared at 100'000 g for 1h in a ultracentrifuge at 4°C. The supernatant was added to 2 ml Protino Ni-NTA agarose slurry in a 15 ml falcon tube and rotated slowly at 4°C for 30 min to allow for binding. The slurry was then filled in a gravity-flow column and washed with 10 ml PdeL purification buffer and subsequently 40 ml of the same buffer supplemented with 40 mM imidazole. Proteins were eluted with 10 ml PdeL purification buffer supplemented with 500 mM imidazole. The eluted protein was then loaded onto a HiLoad 16/600 Superdex 200 pg size exclusion column for fractionation using 140 ml PdeL purification buffer. Protein concentration of the appropriate fractions was determined by photo spectrometric absorption at 280 nm and used fresh or stored at -80°C.

SPR measurements

The affinity of PdeL to DNA sequences was determined by surface plasmon resonance (SPR) measurements using the ReDCaT method (Stevenson *et al.*, 2013). In brief, hybridized biotinylated DNA was immobilized on a streptavidin-coated SPR chip (Cytiva). Experiments were performed at 4°C in a GE Biacore T200 SPR instrument using a flow-rate of 10 µl/min. Washing and regeneration of the chip was performed using 1M NaCl and subsequently 50 mM NaOH at 10 µl/min flow-rate. PdeL buffer was supplemented with 0.1 mg/ml bovine serum albumin (BSA) and 20 ng/µl salmon sperm DNA to reduce unspecific interactions. K_D was determined using the Michaelis-Menten kinetic fitting model.

AUTHOR CONTRIBUTIONS

Conceptualization, B.S., R.P., A.H., M.H. and U.J.; Methodology, B.S., R.P., A.H., Formal Analysis, B.S., R.P., A.H., M.H., and U.J.; Investigation, B.S., R.P., A.H., and U.J.; Resources, M.H., A.H. and U.J.; Writing – Original Draft, B.S., A.H. and U.J. with contributions from all other authors; Funding Acquisition, A.H., M.H. and U.J.

CONFLICT OF INTEREST

The authors declare that they do not have a conflict of interest.

ACKNOWLEDGEMENTS

We thank Dr. Timothy Sharpe and Dr. Ludovit Zweifel of the Biophysics Core Facility of the Biozentrum (University of Basel) for technical assistance with SPR and Fabienne Hamburger for plasmid construction. We thank Prof. Regine Hengge for sharing unpublished data. This work was supported by the Swiss National Science Foundation grant 310030B_147090 to U.J. The authors declare no competing interests.

References

- Acheson, J.F., Derewenda, Z.S., and Zimmer, J. (2019). Architecture of the Cellulose Synthase Outer Membrane Channel and Its Association with the Periplasmic TPR Domain. *Structure* 27, 1855-1861.e3.
- Acheson, J.F., Ho, R., Goularte, N.F., Cegelski, L., and Zimmer, J. (2021). Molecular organization of the *E. coli* cellulose synthase macrocomplex. *Nat Struct Mol Biol* 28, 310–318.
- Altenhoff, A.M., Train, C.-M., Gilbert, K.J., Mediratta, I., Mendes de Farias, T., Moi, D., Nevers, Y., Radoykova, H.-S., Rossier, V., Warwick Vesztrocy, A., et al. (2020). OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res* Volume 49, D373–D379.
- Barker, C.S., Prüß, B.M., and Matsumura, P. (2004). Increased Motility of *Escherichia coli* by Insertion Sequence Element Integration into the Regulatory Region of the *flhD* Operon. *J Bacteriol* 186, 7529–7537.
- Batchelor, R.A., Haraguchi, G.E., Hull, R.A., and Hull, S.I. (1991). Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. *J Bacteriol* 173, 5699–5704.
- Blattner, F.R., Plunkett, G., Bloch, C.A., and Perna, N.T. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1462
- Boehm, A., Kaiser, M., Li, H., Spangler, C., Kasper, C.A., Ackermann, M., Kaefer, V., Sourjik, V., Roth, V., and Jenal, U. (2010). Second Messenger-Mediated Adjustment of Bacterial Swimming Velocity. *Cell* 141, 107–116.
- Broeker, N.K., and Barbirz, S. (2017). Not a barrier but a key: How bacteriophages exploit host's O-antigen as an essential receptor to initiate infection. *Mol Microbiol* 105, 353–357.
- Calloni, I., Unione, L., Jiménez-Osés, G., Corzana, F., Bino, L.D., Corrado, A., Pitirollo, O., Colombo, C., Lay, L., Adamo, R., et al. (2018). The Conformation of the Mannopyranosyl Phosphate Repeating Unit of the Capsular Polysaccharide of *Neisseria meningitidis* Serogroup A and Its Carba-Mimetic. *Eur J Org Chem* 2018, 4548–4555.
- Chanasit, W., Gonzaga, Z.J.C., and Rehm, B.H.A. (2020). Analysis of the alginate O-acetylation machinery in *Pseudomonas aeruginosa*. *Appl Microbiol Biot* 104, 2179–2191.
- Chaudhry, W., Lee, E., Worthy, A., Weiss, Z., Grabowicz, M., Vega, N., and Levin, B. (2020). Mucooidy, a general mechanism for maintaining lytic phage in populations of bacteria. *FEMS Microbiol Ecol* 96, fiae162-.
- Chou, S.-H., and Galperin, M.Y. (2016). Diversity of Cyclic Di-GMP-Binding Proteins and Mechanisms. *I98*, 32–46.

Dion, M.B., Oechslin, F., and Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nat Rev Microbiol* 18, 125–138.

Gavira, J.A., Gumerov, V.M., Rico-Jiménez, M., Petukh, M., Upadhyay, A.A., Ortega, A., Matilla, M.A., Zhulin, I.B., and Krell, T. (2020). How Bacterial Chemoreceptors Evolve Novel Ligand Specificities. *Mbio* 11.

Goh, K.G.K., Phan, M.-D., Forde, B.M., Chong, T.M., Yin, W.-F., Chan, K.-G., Ulett, G.C., Sweet, M.J., Beatson, S.A., and Schembri, M.A. (2017). Genome-Wide Discovery of Genes Required for Capsule Production by Uropathogenic *Escherichia coli*. *Mbio* 8, e01558-17.

Guyer, M.S., Reed, R.R., Steitz, J.A., and Low, K.B. (1981). Identification of a Sex-factor-affinity Site in *E. coli* as $\gamma\delta$. *Cold Spring Harb Symp* 45, 135–140.

Hansen, C.R., Pressler, T., Høiby, N., and Gormsen, M. (2006). Chronic infection with *Achromobacter xylosoxidans* in cystic fibrosis patients; a retrospective case control study. *J Cyst Fibros* 5, 245–251.

Hengge, R. (2021). High-Specificity Local and Global c-di-GMP Signaling. *Trends Microbiol.*

Hengge, R., Galperin, M.Y., Ghigo, J.-M., Gomelsky, M., Green, J., Hughes, K.T., Jenal, U., and Landini, P. (2015). Systematic Nomenclature for GGDEF and EAL Domain-Containing Cyclic Di-GMP Turnover Proteins of *Escherichia coli*. *198*, 7–11.

Hyman, P. (2019). Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. *Pharm* 12, 35.

Jenal, U., Reinders, A., and Lori, C. (2017). Cyclic di-GMP: second messenger extraordinaire. *Nature Reviews. Microbiology* 15, 271–284.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 1–11.

Junkermeier, E.H., and Hengge, R. (2021). A novel locally c-di-GMP-controlled exopolysaccharide synthase required for N4 phage infection of *E. coli*. Submitted.

Keiski, C.-L., Harwich, M., Jain, S., Neculai, A.M., Yip, P., Robinson, H., Whitney, J.C., Riley, L., Burrows, L.L., Ohman, D.E., et al. (2010). AlgK Is a TPR-Containing Protein and the Periplasmic Component of a Novel Exopolysaccharide Secretin. *Structure/Folding and Design* 18, 265–273.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10, 845–858.

Kiino, D.R., and Rothman-Denes, L.B. (1989). Genetic analysis of bacteriophage N4 adsorption. *J Bacteriol* 171, 4595–4602.

Kiino, D.R., Licudine, R., Wilt, K., Yang, D.H., and Rothman-Denes, L.B. (1993a). A cytoplasmic protein, NfrC, is required for bacteriophage N4 adsorption. *J Bacteriol* *175*, 7074–7080.

Kiino, D.R., Singer, M.S., and Rothman-Denes, L.B. (1993b). Two overlapping genes encoding membrane proteins required for bacteriophage N4 adsorption. *J Bacteriol* *175*, 7081–7085.

Kim, M.S., Kim, Y.D., Hong, S.S., Park, K., Ko, K.S., and Myung, H. (2015). Phage-Encoded Colanic Acid-Degrading Enzyme Permits Lytic Phage Infection of a Capsule-Forming Resistant Mutant *Escherichia coli* Strain. *Appl Environ Microb* *81*, 900–909.

King, M.R., Steenbergen, S.M., and Vimr, E.R. (2007). Going for baroque at the *Escherichia coli* K1 cell surface. *Trends Microbiol* *15*, 196–202.

Knecht, L.E., Veljkovic, M., and Fieseler, L. (2020). Diversity and Function of Phage Encoded Depolymerases. *Front Microbiol* *10*, 2949.

Krasteva, P.V., Bernal-Bayard, J., Travier, L., Martin, F.A., Kaminski, P.-A., Karimova, G., Fronzes, R., and Ghigo, J.-M. (2017). Insights into the structure and assembly of a bacterial cellulose secretion system. *Nature Communications* *8*, 2065.

Letarov, A.V., and Kulikov, E.E. (2017). Adsorption of bacteriophages on bacterial cells. *Biochem Mosc* *82*, 1632–1658.

Liu, B., Furevi, A., Perepelov, A.V., Guo, X., Cao, H., Wang, Q., Reeves, P.R., Knirel, Y.A., Wang, L., and Widmalm, G. (2019). Structure and genetics of *Escherichia coli* O antigens. *FEMS Microbiol Rev* *44*, 655–683.

Liu, T.Y., Gotschlich, E.C., Dunne, F.T., and Jonssen, E.K. (1971). Studies on the Meningococcal Polysaccharides II. COMPOSITION AND CHEMICAL PROPERTIES OF THE GROUP B AND GROUP C POLYSACCHARIDE. *J Biol Chem* *246*, 4703–4712.

Maffei, E., Shaidullina, A., Burkolter, M., Druelle, V., Willi, L., Estermann, F., Michaelis, S., Hilbi, H., Thaler, D.S., and Harms, A. (2021). Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. *bioRxiv* 2021.03.08.434280.

McPartland, J., and Rothman-Denes, L.B. (2008). The Tail Sheath of Bacteriophage N4 Interacts with the *Escherichia coli* Receptor NfrA. *J Bacteriol* *191*, 525–532.

Meier-Dieter, U., Barr, K., Starman, R., Hatch, L., and Rick, P.D. (1992). Nucleotide sequence of the *Escherichia coli* rfe gene involved in the synthesis of enterobacterial common antigen. Molecular cloning of the rfe-rff gene cluster. *J Biol Chem* *267*, 746–753.

Mills, E., Petersen, E., Kulasekara, B.R., and Miller, S.I. (2015). A direct screen for c-di-GMP modulators reveals a *Salmonella Typhimurium* periplasmic L-arginine-sensing pathway. *Science Signaling* *8*, ra57.

- Morgan, J.L.W., McNamara, J.T., and Zimmer, J. (2014). Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nature Structural & Molecular Biology* *21*, 489–496.
- Morona, R., Bosch, L. van den, and Manning, P.A. (1995). Molecular, genetic, and topological characterization of O-antigen chain length regulation in *Shigella flexneri*. *J Bacteriol* *177*, 1059–1068.
- Murray, G.L., Attridge, S.R., and Morona, R. (2003). Regulation of *Salmonella typhimurium* lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second Wzz. *Mol Microbiol* *47*, 1395–1406.
- Mutalik, V.K., Adler, B.A., Rishi, H.S., Piya, D., Zhong, C., Koskella, B., Kutter, E.M., Calendar, R., Novichkov, P.S., Price, M.N., et al. (2020). High-throughput mapping of the phage resistance landscape in *E. coli*. *PLoS Biol* *18*, e3000877.
- Namnyak, S.S., Holmes, B., and Fathalla, S.E. (1985). Neonatal meningitis caused by *Achromobacter xylosoxidans*. *J Clin Microbiol* *22*, 470–471.
- Nobrega, F.L., Vlot, M., Jonge, P.A. de, Dreesens, L.L., Beaumont, H.J.E., Lavigne, R., Dutilh, B.E., and Brouns, S.J.J. (2018). Targeting mechanisms of tailed bacteriophages. *Nat Rev Microbiol* *16*, 760–773.
- Omasits, U., Ahrens, C.H., Müller, S., and Wollscheid, B. (2014). Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* *30*, 884–886.
- Pires, D.P., Oliveira, H., Melo, L.D.R., Sillankorva, S., and Azeredo, J. (2016). Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. *Appl Microbiol Biot* *100*, 2141–2151.
- Porter, N.T., and Martens, E.C. (2015). The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology. *Annu Rev Microbiol* *71*, 1–21.
- Rai, A.K., and Mitchell, A.M. (2020). Enterobacterial Common Antigen: Synthesis and Function of an Enigmatic Molecule. *Mbio* *11*, e01914-20.
- Reinders, A., Hee, C.-S., Ozaki, S., Mazur, A., Boehm, A., Schirmer, T., and Jenal, U. (2015). Expression and Genetic Activation of Cyclic Di-GMP-Specific Phosphodiesterases in *Escherichia coli*. *Journal of Bacteriology* *198*, 448–462.
- Reinders, A., Sellner, B., Fadel, F., Berkum, M. van, Kaczmarczyk, A., Ozaki³, S., Rueher⁴, J., Manfredi, P., Sangermani⁵, M., Harms, A., et al. (2021). Digital control of c-di-GMP in *E. coli* balances population-wide developmental transitions and phage sensitivity. *BioRxiv*.
- Richter, A.M., Possling, A., Malysheva, N., Yousef, K.P., Herbst, S., Kleist, M. von, and Hengge, R. (2020). Local c-di-GMP Signaling in the Control of Synthesis of the *E. coli* Biofilm Exopolysaccharide pEtN-Cellulose. *J Mol Biol* *432*, 4576–4595.

Rodrigues, C.G., Rays, J., and Kanegae, M.Y. (2017). Native-valve endocarditis caused by *Achromobacter xylosoxidans*: a case report and review of literature. *Autopsy Case Reports* 7, 50–55.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., Vroom, E. de, Marel, G.A. van der, Boom, J.H. van, et al. (1987). Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* 325, 279–281.

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., and Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in *E. coli* identify essential genes and phage host factors. *PLoS Genet* 14, e1007749.

Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñoz-Rascado, L., García-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., et al. (2013). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res* 41, D203–D213.

Silva, J.B., Storms, Z., and Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiol Lett* 363, fnw002.

Sueki, A., Stein, F., Savitski, M.M., Selkrig, J., and Typas, A. (2020). Systematic Localization of *Escherichia coli* Membrane Proteins. *Msystems* 5, e00808-19.

Sychantha, D., Little, D.J., Chapman, R.N., Boons, G.-J., Robinson, H., Howell, P.L., and Clarke, A.J. (2018). PatB1 is an O-acetyltransferase that decorates secondary cell wall polysaccharides. *Nat Chem Biol* 14, 79–85.

Tanner, M.E. (2005). The enzymes of sialic acid biosynthesis. *Bioorg Chem* 33, 216–228.

Tena, D., González-Praetorius, A., Pérez-Balsalobre, M., Sancho, O., and Bisquert, J. (2009). Urinary tract infection due to *Achromobacter xylosoxidans*: Report of 9 cases. *Scand J Infect Dis* 40, 84–87.

Vann, W.F., Daines, D.A., Murkin, A.S., Tanner, M.E., Chaffin, D.O., Rubens, C.E., Vionnet, J., and Silver, R.P. (2004). The NeuC Protein of *Escherichia coli* K1 Is a UDP N-Acetylglucosamine 2-Epimerase. *J Bacteriol* 186, 706–712.

Wang, Y., Pannuri, A.A., Ni, D., Zhou, H., Cao, X., Lu, X., Romeo, T., and Huang, Y. (2016). Structural Basis for Translocation of a biofilm-supporting exopolysaccharide across the bacterial outer membrane*. *J Biol Chem* 291, 10046–10057.

Wang, Y.-C., Chin, K.-H., Tu, Z.-L., He, J., Jones, C.J., Sanchez, D.Z., Yildiz, F.H., Galperin, M.Y. and Chou, S.-H. (2016). Nucleotide binding by the widespread high-affinity cyclic di-GMP receptor MshEN domain. *Nature Communications* 7, 1–12.

Washizaki, A., Yonesaki, T., and Otsuka, Y. (2016). Characterization of the interactions between *Escherichia coli* receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. *Microbiology-open* 5, 1003–1015.

Wittmann, J., Turner, D., Millard, A.D., Mahadevan, P., Kropinski, A.M., and Adriaenssens, E.M. (2020). From Orphan Phage to a Proposed New Family—The Diversity of N4-Like Viruses. *Antibiotics* 9, 663.

Zähringer, F., Lacanna, E., Jenal, U., Schirmer, T., and Boehm, A. (2013). Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. *Structure* 21, 1149–1157.

Zhang, L., Li, S., Liu, X., Wang, Z., Jiang, M., Wang, R., Xie, L., Liu, Q., Xie, X., Shang, D., et al. (2020). Sensing of autoinducer-2 by functionally distinct receptors in prokaryotes. *Nat Commun* 11, 5371.

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* 430, 2237–2243.