Regulatory interactions between *Corynebacterium glutamicum* and its prophages

Inaugural Dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Max Hünnefeld

from Georgsmarienhütte

Jülich, July 2019

The thesis has been conducted at the Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, from December 2015 until July 2019 under the supervision of Prof. Dr. Julia Frunzke.

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf.

Supervisor: Prof. Dr. Julia Frunzke
Institute of Bio- and Geosciences, IBG-1: Biotechnology,
Forschungszentrum Jülich, Jülich

Co-supervisor: Prof. Dr. Michael Feldbrügge
Institute of Microbiology,
Heinrich-Heine-University Düsseldorf, Düsseldorf

Date of oral examination: 02.10.2019

"The world is a thing of utter inordinate complexity and richness and strangeness that is absolutely awesome. I mean the idea that such complexity can arise not only out of such simplicity, but probably absolutely out of nothing, is the most fabulous extraordinary idea. And once you get some kind of inkling of how that might have happened, it's just wonderful."

Douglas N. Adams (1952 – 2001)

Response to the question "What is it about science that really gets your blood running?"

The studies presented in this dissertation have been published in the following articles and manuscripts:

Pfeifer, E., **Hünnefeld, M.**, Popa, O., Polen, T., Kohlheyer, D., Baumgart, M., et al. (2016). Silencing of cryptic prophages in *Corynebacterium glutamicum*. Nucleic Acids Res. 44, 10117–10131. doi:10.1093/nar/gkw692.

Pfeifer, E., **Hünnefeld, M.**, Popa, O., and Frunzke, J. (2019). Impact of Xenogeneic Silencing on Phage–Host Interactions. J. Mol. Biol. doi:10.1016/J.JMB.2019.02.011.

Hünnefeld, M., Persicke, M., Kalinowski, J., and Frunzke, J. (2019). The MarR-Type Regulator MalR Is Involved in Stress-Responsive Cell Envelope Remodeling in *Corynebacterium glutamicum*. Front. Microbiol. 10, 1039. doi:10.3389/fmicb.2019.01039.

Abbreviations

ATCC	American Type Culture Collection
BHI(S)	Brain heart infusion (+ sorbitol)
CRISPR	Clustered regularly interspaced short palindromic repeats
ChAP-seq	Chromatin affinity purification and sequencing
ChIP-seq	Chromatin immunoprecipitation and sequencing
CGP1-4	*Corynebacterium glutamicum* prophage 1-4
CgpS	*C. glutamicum* prophage silencer
DNA	Deoxyribonucleic acid
DNAP	DNA polymerase
e.g.	*exempli gratia*
et al.	*et alii*
etc.	*et cetera*
eYFP	Enhances yellow fluorescent protein
FACS	Fluorescence-activated cell sorting
HGT	Horizontal gene transfer
IPTG	Isopropyl-thio-β-D-galactopyranosid
LC-MS	Liquid chromatography – mass spectrometry
MALDI-TOF	Matrix-assisted laser desorption / ionization – time of flight
MMC	Mitomycin C
NAP	Nucleoid-associated protein
NGS	Next generation sequencing
OD_x	Optical density at x nm
ORF	Open reading frame
RM	Restriction modification
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNA-seq	RNA sequencing
SPI	Spontaneous prophage induction
TCA	Tricarboxylic acid
TLS	Translation start site
TSS	Transcription start site
UV	Ultraviolet
v/v	Volume per volume
wt	Wild type
w/v	Weight per volume
XS	Xenogeneic silencing

Further abbreviations not included in this section are according to international standards, as, for example, listed in the author guidelines of the Journal of Cell Biology (http://jcb.rupress.org/content/standard-abbreviations).

1 Abstracts

1.1 Summary

Viruses preying on bacteria, so-called bacteriophages, are significantly involved in shaping the evolution of microorganisms. More than 10^{31} phage virions are involved in the transduction of an estimated number of 10^{25} - 10^{28} base pairs of DNA, already in the marine environment. Consequently, it is not surprising that viral DNA represents a major cause for strain-specific differences within bacterial species.

In around 46 % of genomes of all completely sequenced bacteria, at least one active prophage element could be identified, whereas the number of cryptic prophage elements is expected to be even higher. The integration of virus-derived DNA material (e.g. in the form of prophages) into the host genome involves the risk of toxic gene products and thus requires stringent regulation. However, the presence of these prophages not only bears risks for the host cell but also may encode potential beneficial traits for the recipient cell. To gain an adaptive advantage from newly acquired DNA, successful integration into host regulatory circuits is mandatory.

The actinobacterial strain *C. glutamicum* ATCC 13032 contains a total of four prophage elements (CGP1-4). The inducible CGP3 prophage covers nearly 7 % (~219 kbp) of the entire *C. glutamicum* genome and contains the prophage element CGP4. Previous studies revealed that CGP3 is spontaneously induced in a small fraction of cells and can be triggered in an SOS-dependent manner. Recently, we identified the Lsr2-like protein CgpS as an essential xenogeneic silencer of the cryptic prophages in *C. glutamicum*. This thesis aimed at shedding light on how CgpS is governing the lysogenic state of CGP3 and studying the regulatory interaction between the host and the prophage.

Chromatin affinity purification and sequencing (ChAP-Seq) revealed a redistribution of CgpS under prophage-inducing conditions towards different targets in the host genome coinciding with a lower coverage at the prophage region. Under prophage-inducing conditions, CgpS binds to multiple host targets, comprising genes with important functions in DNA replication and repair mechanisms, cell envelope biosynthesis and global transcriptional regulators. While previous approaches relied on a snapshot view of XS binding, these data present the first time-resolved analysis of XS protein binding behavior under prophage-inducing conditions.

Furthermore, the regulatory interactions between host-encoded regulators and the CGP3 prophage were studied. DNA affinity chromatographies with different CGP3 promoters as well as global binding profile analyses of different *C. glutamicum* regulators revealed that several host regulators bind inside of the CGP3 region. One of these regulators, the MarR-type regulator MalR, was a subject of further studies. Regulators of the MarR-type family have been shown to be involved in environmental stress responses, regulation of virulence genes, and degradation of aromatic compounds. Furthermore, different studies showed that MarR-type regulators are involved in the counter-silencing of H-NS silenced horizontally acquired genes in *Escherichia coli* and *Salmonella enterica*. A combination of ChAP-seq based binding profiling and transcriptome analysis using DNA microarrays revealed the function of MalR as a regulator involved in the stress-responsive remodeling of the cell envelope of *C. glutamicum* with several binding sites inside of CGP3. A *malR* deletion strain showed higher sensitivity towards different β-lactam antibiotics, and overexpression of *malR* led to a significantly altered cell envelope. Increased levels of MalR impaired inducibility of the CGP3 prophage, indicating a link between the regulation of cell envelope composition and prophage induction.

Overall, this thesis provides valuable insights into the dynamic binding behavior of the XS protein CgpS under prophage-inducing conditions and reveals a high degree of regulatory interaction between host regulators and the CGP3 prophage in *C. glutamicum.*

1.2 Zusammenfassung

Viren, die Bakterien infizieren, sogenannte Bakteriophagen, nehmen maßgeblich Einfluss auf die Evolution von Mikroorganismen. Alleine in marinen Umgebungen sind mehr als 10^{31} Phagenvirionen an der Transduktion von ungefähr 10^{25} - 10^{28} Basenpaaren DNA beteiligt. Angesichts dessen erscheint es nicht verwunderlich, dass virale DNA eine der Hauptursachen stammspezifischer Unterschiede innerhalb von Bakterienarten darstellt. Obwohl die Integration von viralem DNA-Material in das Wirtsgenom (beispielweise in Form von Prophagen) ein hohes Risiko in sich birgt, enthalten rund 46 % aller vollständig sequenzierten bakteriellen Genome mindestens ein aktives Prophagenelement. Hierbei nicht beachtet ist die Zahl von kryptischen Prophagen, welche weitaus höher geschätzt wird. Interessanterweise stellen integrierte Prophagen nicht nur ein hohes Risiko für die Bakterienzellen dar, da sie toxische Genprodukte kodieren oder eine Wirtszelllyse auslösen könnten, sondern können der Wirtszelle auch einige vorteilhafte Eigenschaften gewähren. Um einen Nutzen aus diesen vorteilhaften Eigenschaften zu ziehen und einen evolutiven Vorteil zu gewinnen, müssen die neuerworbenen Gene in die regulatorischen Kreisläufe der Wirtszelle integriert werden.

Das Actinobakterium *Corynebacterium glutamicum* ATCC 13032 enthält insgesamt vier integrierte Prophagenelemente (CGP1-4), wobei der induzierbare CGP3 Prophage mit circa 219 kb fast 7 % des gesamten Genoms ausmacht und zusätzlich den Prophagen CGP4 enthält. Frühere Untersuchungen konnten zeigen, dass CGP3, welcher über eine bakterielle SOS-Antwort induziert werden kann, in einem kleinen Teil der Zellen spontan induziert. Weiterhin konnte kürzlich das Lsr2-artige Protein CgpS als ein essentieller xenogener *silencer* (XS) des kryptischen Prophagen CGP3 identifiziert werden. Die vorliegende Dissertation zielte darauf ab, aufzuklären, wie CgpS den lysogenen Zustand von CGP3 steuert, und regulatorische Interaktionen zwischen dem Wirt und dem Prophagen zu untersuchen.

Eine Chromatin-Affinitätsreinigung und Sequenzierung (ChAP-Seq) zeigte eine Umverteilung von CgpS unter Prophagen-induzierenden Bedingungen vom CGP3 Prophagen hin zu verschiedenen Zielen im Wirtsgenom, was eine geringere Abdeckung der Prophagenregion zur Folge hatte. Die durch die Umverteilung gebundenen Wirtsgene codieren unter anderem Transkriptionsregulatoren und Proteine mit wichtigen Funktionen in DNA-Replikations- und Reparaturmechanismen und in der Zellhüllenbiosynthese. Während frühere Experimente lediglich eine Momentaufnahme der XS-Bindung darstellten, präsentiert diese Arbeit Daten einer ersten zeitaufgelösten Analyse des Bindungsverhaltens von XS-Proteinen unter Prophagen-induzierenden Bedingungen.

Zusätzlich wurden im Zuge dieser Arbeit die regulatorischen Interaktionen zwischen Wirtsregulatoren und dem CGP3 Prophagen untersucht. DNA-Affinitätschromatographien mit unterschiedlichen CGP3-Promotoren und globale Bindeprofilanalysen zeigten, dass mehrere Wirtsregulatoren Bindestellen innerhalb der CGP3-Region aufwiesen. Einer dieser Regulatoren war der MarR-artige Transkriptionsregulator MalR, welcher anschließend weiter untersucht wurde.

Regulatoren aus der MarR-Familie sind nachweislich an Umweltstressreaktionen, der Regulation von Virulenzgenen und dem Abbau von Aromaten beteiligt. Zusätzlich zeigten verschiedene Studien, dass Regulatoren aus der MarR-Familie in *Escherichia coli* und *Salmonella enterica* an einem *counter-silencing* von H-NS-reprimierten horizontal-erworbenen Genen beteiligt sind.

Im Zuge dieser Arbeit konnte mit Hilfe von Bindeprofilanalysen und Transkriptionsanalysen gezeigt werden, dass es sich bei MalR um einen Regulator handelt, der an einer Stress-bedingten Änderung der Zellhülle von *C. glutamicum* beteiligt ist und darüber hinaus einige Bindestellen im Prophagen CGP3 aufweist. Ein *malR*-Deletionsstamm zeigte eine erhöhte Empfindlichkeit gegenüber verschiedenen β-Lactam-Antibiotika, während die Überexpression von *malR* zu einer signifikant veränderten Zellhülle und einer Beeinträchtigung der CGP3-Induktion führte. Insgesamt deutet dies auf einen Zusammenhang zwischen der Stress-abhängigen Zellwandänderung und der Prophagen-Induktion hin.

Diese Arbeit liefert wertvolle Erkenntnisse über das dynamische Bindungsverhalten des XS-Proteins CgpS unter Prophagen-induzierenden Bedingungen und zeigt ein hohes Maß an regulatorischer Interaktion zwischen Wirtsregulatoren und dem CGP3-Prophagen in *C. glutamicum*.

2 Scientific context and key results of this thesis

2.1 Bacteriophages

2.1.1 Bacteriophages and phage life styles

Bacteriophages (phages), viruses that infect bacteria, are the most abundant biological entity on this planet. With an estimated number of more than 10^{31} phage particles on earth, they are found in every environment in which microbes occur (Hendrix et al., 1999; Weinbauer, 2004). Their ability to transduce DNA horizontally to their host makes them important drivers of microbial evolution. Based on their lifestyle, phages are classically categorized in virulent and temperate phages. Virulent phages infect microbes and immediately hijack the microbial replication machinery in order to replicate themselves. After successful reproduction, phage components self-assemble within the host cytoplasm. New virions are subsequently released via cell lysis to infect further bacteria. Temperate phages, on the other hand, have two different options. They are similarly able to enter the lytic cycle, but they can also take another path: the lysogenic cycle (Figure 1). One very recently discovered example influencing this lytic-lysogenic decision is the so-called 'Arbitrium-system' from *B. subtilis* (Erez et al., 2017; Gallego Del Sol et al., 2019; Wang et al., 2018). This system is the first described phage-communication system, where different levels of the AimR peptide lead to either entering the lysogenic cycle or starting the lytic cycle after infection with the temperate phage SPβ.

The lysogenic cycle is already known since 1953 and describes the integration of the phage DNA into the host cells chromosome (Lwoff, 1953). Here, the phage DNA remains as a so-called prophage in a dormant state and is replicated along with the chromosome. Thus, it maintains a long-term association with its host. In addition to integration into the host genome, the episomal state represents a special form within the group of lysogenic phages. Here, the prophage DNA remains extra-chromosomally in a stable circular state and is transferred to daughter cells in a plasmid-like manner, using a plasmid-like segregation system (Sengupta et al., 2010). Examples for phages showing episomal states are the coliphage P1 and the *Staphylococcus aureus* phage ΦBU01 (Ikeda and Tomizawa, 1968; Utter et al., 2014). However, the stable state of the episomal prophages must be distinguished from the state of pseudolysogeny. Pseudolysogeny can occur when a phage infects the host cells during unfavorable conditions (e.g. under starvation). The phage exhibits stalled development and phage DNA remains in a circular state in the cell until more favorable conditions for the host

cell arise, so that lysogenization or the lytic cycle can begin (Łoś and Wegrzyn, 2012). Interestingly, with so-called chronic infection or chronic cycle, another special case for the life cycle of temperate phages exists (Figure 1). The filamentous coliphage M13 is a model for this chronic cycle (de Paepe et al., 2010). During the chronic cycle new phage particles are assembled inside the cytoplasm and continuously released into the environment without lysis of the host cell. Furthermore, the intracellular form of the M13 phage is replicated together with the host genome and thus can be vertically transferred to daughter cells during cell division.

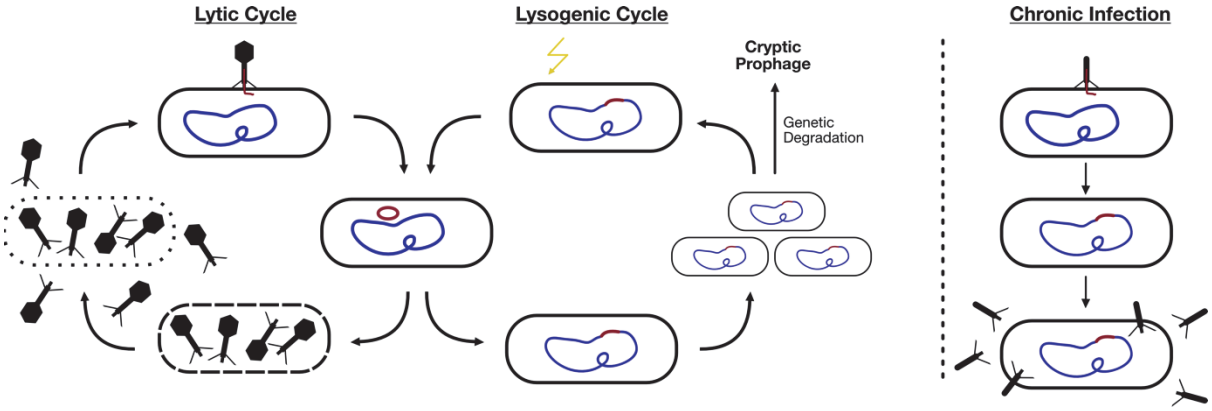


Figure 1: Phage life cycles and lytic-lysogenic switch. After a temperate phage injects the phage DNA into the host cell, different molecular factors influence the decision if the phage will follow the lytic or the lysogenic path of infection (Casjens and Hendrix, 2015). During the lytic cycle the circular phage DNA remains freely inside the host, phage genes are expressed, and new phage particles are assembled. Subsequent to this assembly, phage encoded lysins and holins are expressed in order to lyse the host cell. Through this lysis, new phage progeny is released into the environment and can attack new host cells. If the decision is in favor of lysogeny, the circular phage DNA is integrated into the host cell genome via site-specific recombination. The integrated phage is now present as a dormant prophage in the genome. This lysogenic state is stabilized with maintenance mechanisms and the prophage is replicated together with the host chromosome and thus passed vertically to daughter cells. A long-term integrated prophage can – due to selection pressure – be cryptified (described in more detail in Chapter 2.1.5). External triggers (e.g. DNA damage caused by UV radiation, marked with a lightning) can trigger an SOS response in the host cell, which can lead to induction of the prophage. This induction describes the switch from lysogenic to lytic state. The prophage 'leaves the sinking ship'. Another life cycle that some phages (e.g. coliphage M13) exhibit is the chronic cycle (also called 'chronic infection'). Here, the phage DNA is integrated into the genome of the host cell and - as in the lytic cycle - new phage particles are assembled. However, these are released into the environment without cell lysis. An example of this life cycle is known in coliphage M13 (de Paepe et al., 2010).

2.1.2 Lysogeny: integrating horizontally acquired DNA into the host chromosome

According to Howard-Varona and colleagues the proceeding of lysogeny can be separated into three successive steps, further explained on the coliphage λ (Howard-Varona et al., 2017). The first step is the establishment of the lysogeny by a very complex regulatory interaction (Casjens and Hendrix, 2015). This interaction is mainly based on the presence of the repressor CI and the activity and stability of the CI regulator CII. These are modified in response to the nutritional- and growth state of the host cell as well as to the number of phages infecting one cell, and several other factors (reviewed by Casjens and Hendrix, 2015). Further important for the establishment of the lysogenic state is the presence of a bacterial attachment site (*attB*) that matches to the attachment site of the temperate phage (*attP*). This allows a successful site-specific recombination to integrate the phage DNA into the host chromosome using the λ integrase (Int) together with the host-encoded integration host factor (IHF) (Landy, 2015; Landy and Ross, 1977). However, these attachment sites are not required for episomal prophages, which do not integrate into the host genome. The second step to proceed lysogeny is the maintenance step. This step describes the stable maintenance of the lysogenic state over a longer period of time. Again, the lambda repressor CI plays a decisive role here, because it maintains the lysogenic state (Court et al., 2007). Under stressful conditions, cells can exit the lysogenic cycle. This third step is the so-called induction and will be in the detailed spotlight in the following chapter.

2.1.3 SOS-dependent induction of prophages: the lytic-lysogenic switch

The lytic-lysogenic switch (based on the example of the λ-phage, Figure 1) can be triggered by different mechanisms. The most prominent and best-known mechanism is the bacterial SOS-response (Michel, 2005; Oppenheim et al., 2005). The key components of this SOS-response are the proteins RecA and LexA. LexA is – under normal conditions – bound to its target promoter regions where it represses the expression of several SOS genes. These genes are involved in DNA repair mechanisms, like nucleotide excision repair (e.g. *uvrA*, *uvrB* and *uvrD*), homologous recombination (e.g. *recA*) as well as the repair by polymerase PolV (*umuC*, *umuD*), and genes involved in growth arrest (*sfiA*) (Janion, 2008; Michel, 2005). As a result of spontaneous DNA damage or stalled replication forks, the occurrence of ssDNA triggers the activation of RecA (Figure 2). This activated RecA forms a nucleoprotein filament at the ssDNA regions. The nucleoprotein filament, in turn, activates the co-protease function of RecA,

enabling the autoproteolytic cleavage of the LexA repressor (Little, 1984). Simultaneously, as illustrated in Figure 2, RecA triggers the autoproteolytic cleavage of the CI-λ-repressor, which subsequently leads to prophage induction (Figure 1 and Mustard and Little, 2000). In addition to this model, there are also examples known, where not the cleavage of a prophage repressor is triggering the SOS-dependent prophage induction, but an anti-repressor system is interposed (Kim and Ryu, 2013; Mardanov and Ravin, 2007; Shearwin et al., 1998). In these examples, the anti-repressor proteins are under direct repression of LexA and thus formed in consequence of an SOS-response. Anti-repressors may function via a direct interaction with the corresponding prophage repressor thereby interfering with its function.

Prophage induction can be triggered in the course of a cellular SOS response. This SOS response is caused e.g. by DNA damage, which can be provoked by the addition of certain agents (e.g. MMC) (Otsuji et al., 1959), but also occurs spontaneously in a small fraction of cells (Alexeeva et al., 2018; Helfrich et al., 2015; Nanda et al., 2014). Nevertheless, the negatively afflicted (spontaneous) prophage induction causing cell lysis or cell death can also have beneficial effects on population level that can be of importance in biofilm formation, for virulence or for horizontal gene transfer (Nanda et al., 2015).

In addition to the SOS-dependent prophage induction, also examples of SOS-independent prophage induction have been described. One example of this SOS-independent prophage induction was shown in *E. coli* (Rozanov et al., 1998). Rozanov and colleagues demonstrated that the overexpression of the transcriptional regulator RcsA or the overabundance of the RNA DsrA could lead to a prophage induction independent of RecA. Two other examples connect quorum sensing systems with a prophage induction: (i) Ghosh and colleagues demonstrated that acyl-homoserine, an essential quorum sensing molecule produced in *P. aerigunosa* PAO1, is able to trigger λ induction in *E. coli* (Ghosh et al., 2009). (ii) A very recent publication described an inactivation of the CI prophage repressor in *Vibrio cholorae* in an SOS-independent manner that caused induction of the VP882 prophage (Silpe and Bassler, 2019). Interestingly, this CI inactivation is caused by an anti-repressor protein (Qtip), that is controlled by a phage-encoded homolog of the quorum sensing receptor VqmA.

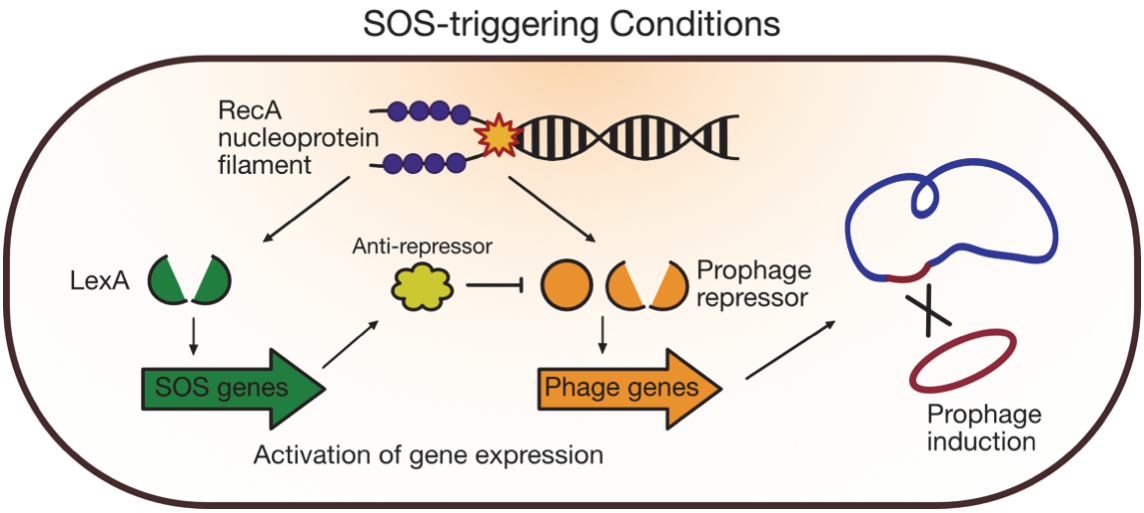


Figure 2: SOS-dependent prophage induction (on the example of *E. coli* phage λ). Under physiological conditions in a lysogenic cell, prophage genes are repressed by a phage repressor protein, SOS genes are repressed by the SOS-repressor LexA, and RecA is available freely in the cytosol. Environmental conditions (e.g. exposure to UV radiation) or the addition of DNA damaging compounds (e.g. MMC) lead to DNA damage as well as a stalling of the replication fork, which result in the occurrence of ssDNA in the cell. This ssDNA is bound by the SOS-protein RecA and consequently a nucleoprotein filament is formed. This nucleoprotein formation activates the co-protease functionality of RecA, which subsequently triggers the autocatalytic cleavage of LexA. Due to LexA cleavage, the former repressed SOS-genes are de-repressed (shown in green). Furthermore, the expression of a LexA-repressed anti-repressor is activated. This anti-repressor antagonizes the function of the phage repressor, which subsequently leads to an activation of the phage gene expression and the prophage induction (Kim and Ryu, 2013; Mardanov and Ravin, 2007; Shearwin et al., 1998). Another mode of action activating the prophage induction due to an SOS-response is the RecA-triggered autocatalytic cleavage of the prophage repressor itself (shown in orange), similar as described for LexA. (Figure adapted from Nanda et al., 2015).

2.1.4 Lysogenic conversion

Horizontal gene transfer caused by transduction of phage DNA and the subsequent integration into the host genome is a very common event in nature. In a screening approach for prophage elements among all sequenced bacterial genomes it was found that 46 % of these genomes contained at least one active prophage element (Touchon et al., 2016). In locations with a very high occurrence of virus like particles this frequency was even higher. A study from 2011 showed that the human fecal metagenome consists of 4-17 % DNA of virus-like particles or phages (Minot et al., 2011). Another study analyzed the abundancy of lysogeny in the gut microbiome of mice. A metagenome analysis conducted in the course of this study, revealed that in 119 of totally 181 bacterial bins at least one putative prophage genome fragment could be detected, which constitutes a relative number of 65.8 % (Kim and Bae, 2018). These results illustrate that there is a frequent occurrence of prophages in commensal bacteria populations.

Many prophage elements in bacterial chromosomes do not only encode genes important for the prophage life cycle or the prophage itself, but also contain accessory genes that can be beneficial for the host cell survival. These genes are also termed 'morons' (Brüssow et al., 2004; Juhala et al., 2000) and they are able to cause a lysogenic conversion of the host cells. Some of these morons help host cells to cope with diverse changing environments by enhancing e.g. resistance to oxidative stress or acid (Wang et al., 2010). Other morons are involved in the protection of the host cell against infections by further phages, both homotypic (closely related or same phages) and heterotypic (unrelated phages). In an analysis of ten temperate mycobacteriophages, Dendrick and colleagues discovered different prophage encoded defense mechanisms. These defense mechanisms can cause i.a. a modification of the cell envelope in order to destroy the recognition by further phages (e.g. Charlie gp32 defence against Che9c), they can cause a repressor-mediated immunity (homologs found in all ten mycobacteriophages) or function via specific restriction modification systems (e.g. Panchio gp28) (Dedrick et al., 2017). Further, prophage morons are also included in pathogenicity-relevant host modifications like enhancing antibiotic resistance, encoding exotoxins, improving bacterial adhesion and invasion, modulating the immune system of infected organisms and enhancing survival inside infected organisms (Balcazar, 2014; Davies et al., 2016). A recent interesting example of lysogenic conversion was presented by Gerlach and colleagues (Gerlach et al., 2018). Here, they could show that one crucial strategy for healthcare-associated methicillin-resistant *Staphylococcus aureus* (HA-MRSA) strains to evade the immune system of patients is an alteration of their wall teichoic acids (WTA). This alteration is caused by a prophage-encoded alternative WTA glycosyltransferase, TarP, that changes the surface structure of the prophage-carrying MRSA strains and could lead to up to 40-fold decreased immune responses (Gerlach et al., 2018).

In addition to these advantages given by lysogenic conversion, the integration of prophages can also have other effects for the host cells. The integration of prophages into specific genomic regions as well as the subsequent excision can e.g. act as a regulatory switch (Feiner et al., 2015). Furthermore, prophage encoded recombination systems can be used for remodeling the chromosome of the host cells and structural rearrangements, which can enhance the evolutionary adaptivity (Menouni et al., 2015). Finally, prophages can also be very important on the population level. Phage-mediated transduction processes like specialized transduction (areas enclosing the prophage area are also excised out of the genome) or generalized

transduction (random pieces of the host genome are co-transduced) increase the genetic variability and are also useful for the transduction of genes to other cells within the community (Harrison and Brockhurst, 2017).

2.1.5 Domestication of prophage elements: minimizing risks, keeping advantages

The long-term relationship between host cells and integrated prophages with the ubiquitous risk of an induction leading to cell death or cell lysis – like molecular time bombs – puts an enormous selection pressure on the cells. An important target to disarm these molecular time bombs are factors required for an induction of prophages. Because prophages require intact attachment sites as well as an integrase, excisase and accessory host factors (IHF, Fis) for excision, these represent interesting targets for a first inactivation (Ramisetty and Sudhakari, 2019). Furthermore, Bobay et al. could show in bioinformatic analyses regarding the domestication of prophages, that there is a strong bimodal distribution of prophages. This is suggested to result from a rapid gene loss that leads to a prophage inactivation followed by a slower genetic degradation (Bobay et al., 2014). Ramisetty and Sudhakari also describe the domestication process as a two-step process with a first 'grounding' event, that inactivates the inducibility (e.g. by disruption of attachment sites or integrases) and a subsequent slower degradation process (Figure 3, Ramisetty and Sudhakari, 2019). The second step in both described processes leads to cryptic prophages inside bacterial genomes, that harbor beneficial functions for the host cells but are not able to cause harsh damage like cell lysis or cell death any more.

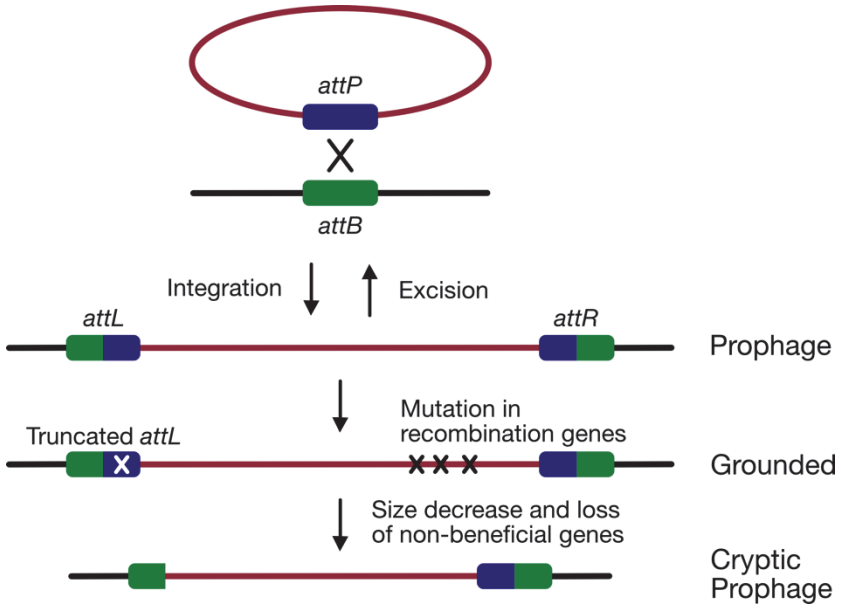


Figure 3: Development of cryptic prophages. The genomic integration of phage DNA during the lysogenic cycle occurs via a site-specific recombination between the phage attachment site (*attP*) and the corresponding attachment site in the bacterial genome (*attB*). The recombined attachment sites flanking the integrated prophage are subsequently designated as *attL* and *attR*. As the inducible prophage element exerts a high selection pressure on the cells (an induction would lead to cell lysis or cell death), mutations affecting the induction-required elements (like attachment sites, recombinases, etc.) are very likely. These mutations lead in a first step to the so-called 'grounding' of the prophage. After a grounding process, further evolutionary processes lead to mutations and recombination inside the grounded prophage in order to reduce the size of this element and to dispose non-beneficial or even detrimental properties encoded in this area. This further reduction leads to the generation of a cryptic prophage (information and figure adapted from Ramisetty and Sudhakari, 2019).

Lysogenic conversion can provide many advantages to the host cell (as described in 2.1.4). However, if a bacterial host cell domesticates prophages, it can gain several further beneficial traits. Functional prophages inside the host genome can be used as weapons of immune lysogens to kill non-lysogenic competitors (Brown et al., 2006; Gama et al., 2013). However, harboring these functional prophages additionally bears a high risk and it is very likely, that due to infections the competitor also becomes a lysogen at a certain point. Therefore, it is more efficient to cope with competitors by using only specific components of prophages as biological weapons. One way is to use phage-derived bacteriocins that can kill other bacterial cells that do not encode cognate immune features (Hardy, 1975; Michel-Briand and Baysse, 2002). Phage-derived bacteriocins are e.g. R-type or F-type pyocins from *P. aeruginosa*, which both resemble parts of bacteriophage tails (Michel-Briand and Baysse, 2002). In addition, there are also phage-derived particles known that act similar as bacteriocins but were named 'phage killer particles' or 'protophages' in former publications (Bobay et al., 2014).

Another interesting function that can be gained from domesticated prophages is the ability to transduce genetic material using phage-derived transduction mechanisms (Lang and Beatty, 2007). Lang and Beatty describe the elements used for this purpose as so-called 'gene transfer agents' (GTAs), that resemble virus-like particles lacking integrases. These are able to transduce random pieces of the genome without any detrimental properties. Further, it has been shown that these GTAs are derived from a domestication process of phages which show the ability of generalized transduction, like e.g. P22 (Casjens et al., 1992; Lang and Beatty, 2007).

Finally, a benefit of harboring domesticated prophages is the protection against superinfection. This can be achieved by an interference of the gene products from domesticated prophages with the viable progeny of infecting phages. A prior study could demonstrate, that a double-infection of an already lysogenic *E. coli* strain results in less produced virions (Refardt, 2011).

2.1.6 The cryptic prophages of *Corynebacterium glutamicum*

This doctoral thesis focusses on investigations of prophage-host interactions using exemplarily the model organism *Corynebacterium glutamicum*. This Gram-positive soil bacterium is a non-pathogenic bacterium with high relevance in biotechnological applications and is used as a model for different human-pathogenic *Corynebacterineae* like *Mycobacterium tuberculosis* or *Corynebacterium diphtheriae*. *C. glutamicum* was firstly isolated in 1957 as a strain naturally producing high amounts of glutamate (Kinoshita et al., 2004).

Decades ago, in different *C. glutamicum* strains, various inducible prophages could be detected. Their induction was successfully triggered using UV light or MMC. Some of those (like Φ15, Φ16 and ΦGA1) were characterized regarding their host range and life cycle (Moreau et al., 1995; Sonnen et al., 1990b, 1990a). Genome sequencing of the *C. glutamicum* strain ATCC 13032, the model strain of this thesis, also revealed the presence of prophages. This *C. glutamicum* strain ATCC 13032 exists in two different sub-strains. One of those (NCBI reference: NC_003450.3) was sequenced by Ikeda and Nakagawa (Ikeda and Nakagawa, 2003) and the other one (NCBI reference: BX927147) was sequenced by Kalinowski et al. (Kalinowski et al., 2003). Both strains contain three cryptic prophages in their genome: CGP1 – 3. Nevertheless, the main difference between those strains is an integration of a fourth prophage (CGP4) inside the CGP3 prophage region of the NC_003450.3 strain. While CGP1 and CGP2 are small (13.5 kb and 3.9 kb) and highly degenerated cryptic prophages that do not show any activity, the biggest prophage CGP3

(including CGP4, ~ 219 kb) is inducible. It can consequently excise of the genome, circularize and replicate autonomously inside the host cell (Donovan et al., 2015; Frunzke et al., 2008). However, CGP3-triggered cell lysis or active phage particles have never been observed. Besides many hypothetical proteins, some of the CGP3 genes encode different putative prophage elements: a primase, a resolvase, a lysin, and an integrase (Annotation, SI in Baumgart et al., 2018). Another interesting CGP3-encoded element is a restriction modification system, spanning the genes *cglIM*, *cglIR*, *cglIIR*. It is suggested to be i.a. important for the maintenance of the integrated CGP3, because a deletion of this prophage area was only possible after an inactivation of this system (Baumgart et al., 2013; Schäfer et al., 1997). The CGP3 prophage, furthermore, encodes the *alpAC* operon, which codes for an actin-like protein (AlpC) that is in combination with a prophage DNA binding protein (AlpA) important for efficient viral replication (Donovan et al., 2015). Donovan and colleagues demonstrated that AlpC forms filamentous structures upon prophage induction, which interact with AlpA-bound CGP3 DNA and might be involved in the spatiotemporal organization of the viral replication.

Studies regarding the inducibility of CGP3 revealed that spontaneous prophage induction occurs in up to 1 % of a *C. glutamicum* culture (Nanda et al., 2014). Another study focusing on this spontaneous induction of CGP3 revealed that around 30 % of the spontaneous induced cells are induced independent of RecA (SOS-independently), whereas the other 70 % of the spontaneous induced cell show a preceding SOS-response followed by the prophage induction (Helfrich et al., 2015). This study could further show, that the SOS-dependent CGP3 induction can be triggered by using MMC as previously described for other prophages.

In order to decipher the molecular mechanism behind the CGP3 induction, previous experiments aimed at finding regulators involved in this process (Pfeifer, 2013). A DNA affinity chromatography with the promoter region of the *alpAC* operon, which belongs to the group of early phage genes (Donovan et al., 2015), resulted in the identification of a 13.4 kDa protein. This protein could be identified by MALDI-TOF analysis as the CGP3 encoded protein Cg1966, later renamed to CgpS (Pfeifer et al., 2016). CgpS is a Lsr2-like nucleoid-associated protein (NAP) belonging to the group of xenogeneic silencers (XS). It was a focus of this thesis to characterize the functionality of CgpS and to disentangle its regulatory role in the life cycle of CGP3. Furthermore, this thesis aimed at detecting CGP3-binding host regulators in order to characterize prophage-host interactions. These host regulators were further characterized

regarding their binding behavior, their physiological function and their impact on CGP3 induction.

2.2 Xenogeneic Silencing

2.2.1 Phage defense systems and the importance of xenogeneic silencing

Every second, around 10^{24} productive phage infections occur on Earth. This is an astonishing number, but considering the approximated number of 10^{31} individual phages on Earth, this frequent occurrence of infections is not surprising (Hendrix, 2003). In order to survive and to limit the number of productive infections, bacteria have evolved different defense mechanisms during the ongoing molecular arms race between the prey and the predator. Each step of the phage infection and replication cycle can be targeted by the host cell to disrupt productive infections (Labrie et al., 2010). The first barrier against phage infections is represented by the cell surface. Before phages can inject their DNA into the host cell, they have to recognize the host cell and adsorb to the cell surface. Labrie et al. collected several examples of how the cell surface recognition and adsorption can be modified via blocking of phage receptors, production of extracellular matrix or the production of competitive inhibitors that outcompete phage binding (Labrie et al., 2010). However, if the phage could successfully adsorb to the cell surface and injected its DNA, two other well-known defense systems can protect the host cells of further damage: restriction-modification systems (RM systems) and CRISPR-Cas systems (Figure 4, Stern and Sorek, 2011). The common effect of both systems is a degradation of the phage DNA, either using restriction endonucleases that differentiate between host and foreign DNA via the methylation pattern (Bickle and Krüger, 1993), or using the adaptive bacterial anti-virus system CRISPR-Cas (Sorek et al., 2008). Interestingly, also many phage defense mechanisms are encoded on prophages in the host genome and contribute to the benefits of carrying prophages (Chapter 2.1.4 and Dedrick et al., 2017). Although these defense systems are essential for the survival of bacteria, they have an important disadvantage: While the cells are protected from infections and are not exposed to the danger of lysis, they cannot benefit from potentially valuable genes encoded on the horizontally acquired DNA genes at the same time. At this point XS becomes important: XS represents a system that allows specific silencing of foreign DNA inside bacterial genomes to protect the host (Navarre, 2016). However, a direct involvement of XS in the defense against phage infections has not yet been proven. XS allows additionally to

the protection the incorporation of parts of the silenced regions into host regulatory circuits in order to use beneficial genes (Navarre, 2016).

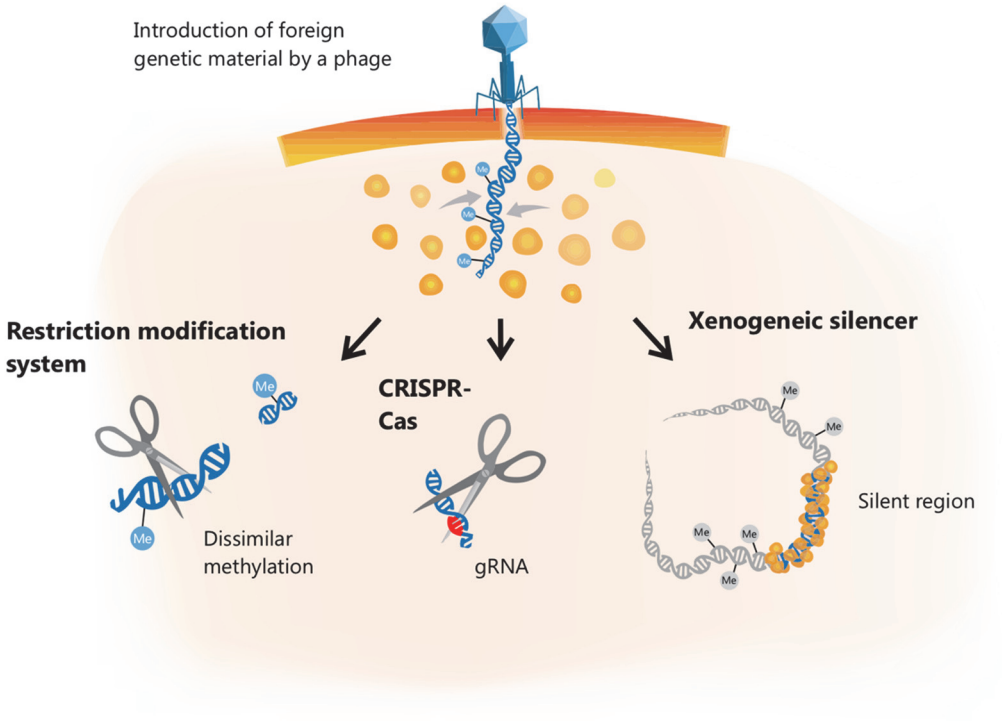


Figure 4: Bacterial mechanisms coping with foreign DNA elements. The infection of bacterial cells with phage DNA can be very detrimental for the bacteria, because it can lead to phage replication followed by cell lysis or other detrimental effects of the encoded genes (Chapter 2.1.1). Thus, defense mechanisms are required that are able to defuse the high risks. Two of those defense mechanisms aim at a destruction of the phage DNA: the restriction modification system that senses dissimilar methylation patterns of the foreign DNA and the bacterial adaptive 'immune system' CRISPR-Cas (Stern and Sorek, 2011). Another mechanism by which bacteria can deal with horizontal-acquired DNA is constituted by XSs. These repress the expression of horizontally acquired DNA but do not disrupt this DNA. Thus, putative beneficial traits of these acquired elements stay usable for the host cell (Navarre, 2016). (Figure taken from Pfeifer et al., 2019).

2.2.2 How does xenogeneic silencing work?

2.2.2.1 The four known XS classes feature a similar mode of action

The key players of the xenogeneic silencing are small (< 15 kDa) nucleoid-associated proteins, so-called xenogeneic silencers. These silencers are known to bind to adenine- and thymine-(AT)-rich DNA stretches, that mostly derive from foreign DNA elements, which usually feature a higher content of AT in comparison to the host genome (Moran, 2004; Navarre, 2016). The binding of the XSs leads to the formation of a tight nucleoprotein complex which in turn leads to repression of gene expression (Figure 5A, B). All currently known XSs can be categorized into four different groups: H-NS and H-NS-like proteins from different proteobacteria like

Escherichia and *Salmonella* (Navarre et al., 2006; Oshima et al., 2006), MvaT/U from different *Pseudomonas* strains (Tendeng et al., 2003), Lsr2 and Lsr2-like proteins from actinobacteria (Ali et al., 2012; Pfeifer et al., 2016) and Rok from *Bacillus subtilis* (Smits and Grossman, 2010). Interestingly, only MvaT and H-NS are homolog protein families; the other classes do not share any sequence similarities (Singh et al., 2016). However, structural as well as functional analysis revealed a similar mode of action for these XS proteins that suggest a convergent evolution of these silencers. The silencer CgpS from *C. glutamicum*, which is the main focus of this work, belongs to the group of Lsr2-like proteins. Interestingly, this silencer is the first published XS encoded on the genome of a cryptic prophage (Pfeifer et al., 2016).

2.2.2.2 Recognition of foreign DNA and formation of nucleoproteins complexes

As mentioned before, XSs of all four classes bind specifically to AT-rich DNA sequences. In the case of Lsr2 and H-NS, high resolution structural analyses of the C-termini could show that this binding is mediated by a prokaryotic AT-hook motif (Gordon et al., 2011). The core of this AT-hook consists of a specific "XGR" (Lsr2: X = R, H-NS: X = R/Q), that can bind into the narrow minor groove of AT-rich DNA (Figure 5C, Ali et al., 2012). This specific sequence can also be found in eukaryotic AT-hooks inside of HMG-I(Y) eukaryotic nuclear proteins as well as high-mobility-group A proteins (Aravind and Landsman, 1998; Fonfría-Subirós et al., 2012). Interestingly, the AT-hook motif plays also an important role in the maintenance of viral DNA in human cells, infected latently with the Epstein-Barr Virus (Chakravorty and Sugden, 2015). Here, the protein EBNA1 contains these motifs and mediates the binding of the viral plasmid towards the host genome, so the viral DNA is co-replicated and can be spread to daughter cells.

We performed an amino acid sequence analysis with CgpS from *C. glutamicum* and could show, that this XS also contains an AT-hook-like motif in its C-terminus, flanked by two predicted α-helices (Pfeifer et al., 2016). However, the motif was slightly different compared to the motifs of Lsr2 and H-NS, because the third arginine of CgpS is substituted with an isoleucine. Nevertheless, it was previously shown, that the AT-hook motif can vary without losing the binding ability of the silencer and that a naturally occurring variability between different XSs of the same group exist in their AT-hook (Cordeiro et al., 2011; Gordon et al., 2011). One possible reason for these variations in the AT-hook motif could be an adaptation towards different GC-contents of distinct host genomes. As XS proteins bind specifically towards high AT areas,

different host GC-profiles could require different AT-binding specificities. The remaining two groups of XSs show different AT-binding mechanisms. For MvaT it was shown that instead of an AT-hook, this protein contains a so-called 'AT-pincer' motif (Ding et al., 2015). Ding and colleagues could show that this AT-pincer consists of the motif KGGNH, that can be inserted into the minor groove of AT-rich DNA, and several lysine residues, that can interact with the sugar-phosphate backbone of the DNA and alter the conformation of the DNA molecule. It is suggested that this specific constitution of MvaT leads to a higher tolerance towards GC-interruptions inside the recognized AT-stretches of the recognized sequences (Ding et al., 2015). For the *B. subtilis* XS Rok no AT-hook or AT-pincer like motif could be discovered (Duan et al., 2018). However, Duan et al. showed that Rok contains three non-consecutive residues (N-T-R). These residues recognize the DNA minor groove. Additionally, Rok contains four lysine residues, that interact with the phosphate groups of the DNA, which – similarly to MvaT – leads to conformational changes of the DNA molecule. Furthermore, the recognition mechanics lead to a preference of Rok towards AT-rich DNA sequences with G or C insertions and also to some sequence specific motifs which are apparently underrepresented in the respective host genome (Duan et al., 2018).

In a very recent study, we analyzed a ChAP-seq experiment of the Lsr2-like silencer CgpS with regard to the binding behavior at DNA regions containing different AT-contents as well as GC-interruptions (Wiechert et al., 2019, *submitted*). With this analysis we could show that CgpS exhibits a high affinity towards AT-rich stretches with multiple A-T-steps: the more AT-steps per fragment, the higher the fraction of CgpS-bound regions. Furthermore, our data revealed that an increasing amount of GC-interruption inside the AT-stretches causes less frequent CgpS binding. Investigation of CgpS bound promoter regions indicated the presence of a potential motif. AT-rich DNA regions covering this binding motif displayed an increased CgpS-bound fraction (Wiechert et al., 2019, *submitted*). A combination of the described findings could lead to the assumption that CgpS probably starts DNA-binding from a specific nucleation site, similar to H-NS (described in more detail in the following chapter).

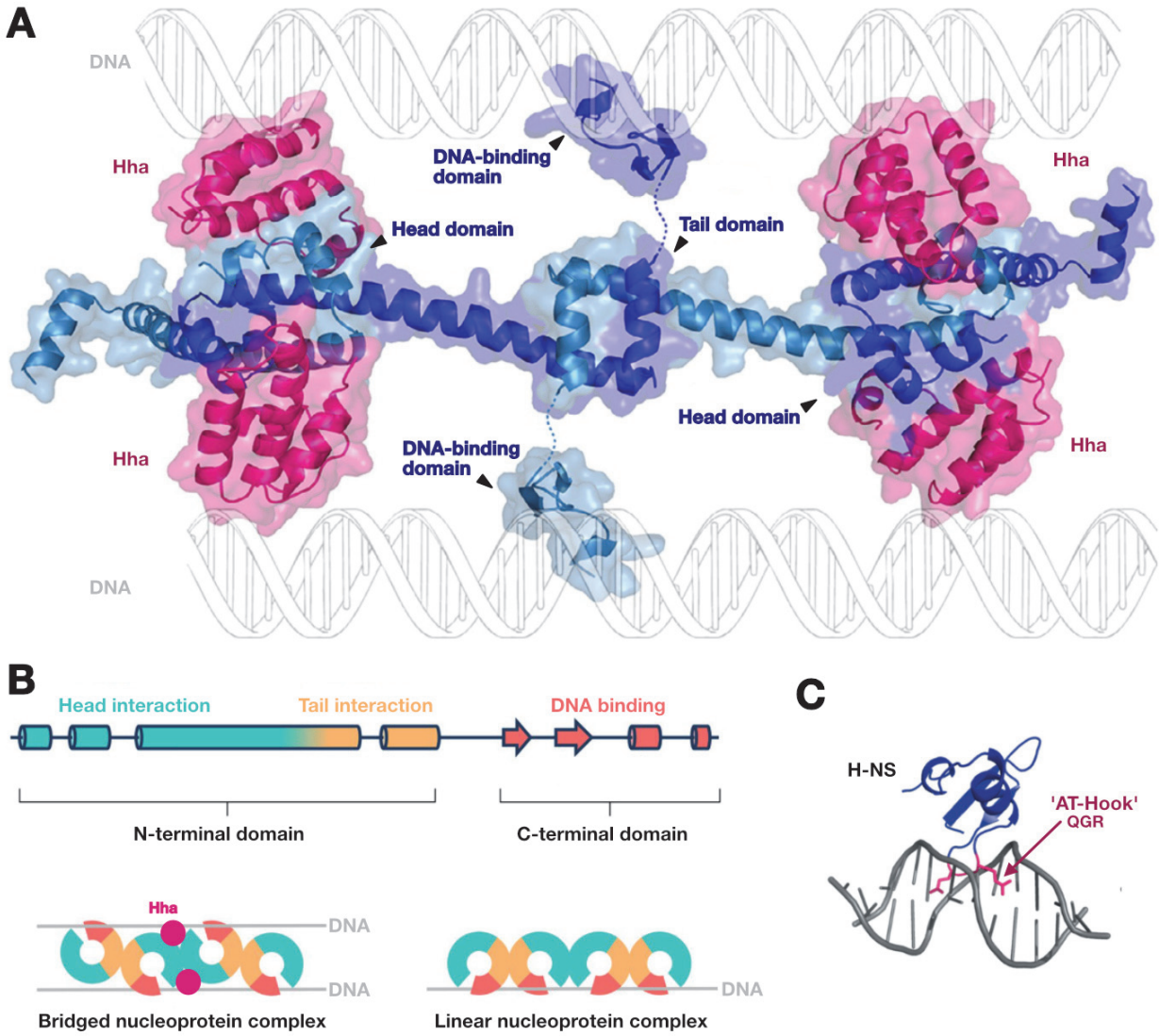


Figure 5: Nucleoprotein complex formation and DNA binding properties of the XS protein H-NS. (A) This figure presents a model of the H-NS/Hha nucleoprotein complex spanning four H-NS protomers (blue) and four Hha molecules (pink). Except for the linker region (dashed lines), the complete model is built based on multiple structural analyses (as described in Navarre, 2016). The interaction with the DNA helices (not in scale) shows how DNA bridging can be established. H-NS forms oligomers with both N-terminal domains: the head and the tail. The oligomer is composed of head-to-head and tail-to-tail interactions. Hha associates to the H-NS head domain and contains specific positively charged residues facing outwards the complex. This positive charge is required for interaction with negatively charged DNA. (Information and figure adapted from Navarre, 2016). (B) Schematic overview of the composition of H-NS and the interactions of H-NS protomers with each other and with DNA. This figure represents an illustration of the model presented in (A). Turquoise represents the N-terminal head domain, yellow shows the N-terminal tail domain and the C-terminal DNA-binding domain is illustrated in red. The accessory protein Hha is presented in pink. (Information and figure adapted from Grainger, 2016). (C) H-NS binds to the minor groove of AT-rich DNA using a conserved prokaryotic AT-hook (QGR). This figure shows the interaction of the AT-hook site chains with the DNA. For the high-resolution models, Ali and colleagues used heteronuclear single-quantum correlation (HSQC) experiments (Information and figure adapted from Ali et al., 2012).

2.2.2.3 Nucleoprotein complex formation leads to DNA structuring and silencing

In addition to the DNA-binding mechanism described above, XSs require a multimerization via their N-terminal domain, to enable successfully nucleoprotein complex formation and thus silencing. A general picture that emerged is that after an initial binding to a 'nucleation-site', more silencer proteins are recruited and oligomerize cooperatively along the AT-stretches in the DNA, to form a nucleoprotein filament (Gordon et al., 2011).

Possible nucleation sites could be identified for H-NS in former studies and described to require as a critical feature a TpA-step (Bouffartigues et al., 2007; Lang et al., 2007). The high affinity of H-NS towards certain T-A-steps could also be modelled mathematically in a very recent approach (Riccardi et al., 2019). Riccardi et al. presented the probabilities of finding H-NS at different positions in an AT-rich DNA fragment. These probabilities were highest at certain A-T-steps.

Following the binding at the nucleation site, additional XS proteins bind cooperatively and oligomerize to form the nucleoprotein complex. This oligomerization could be shown to be essential for the function of H-NS, Lsr2, Rok, MvaT/U and CgpS (Arold et al., 2010; Duan et al., 2018; Gordon et al., 2010; Pfeifer et al., 2016; Winardhi et al., 2012, 2014a). We could demonstrate that the CgpS protein can be purified using only the N-terminal oligomerization domain in pulldown assays. Furthermore, an overproduction of the N-terminal oligomerization domain lacking the DNA-binding domain lead to a disruption of the CgpS-mediated silencing (Pfeifer et al., 2016). This results fit to the observation of Williamson et al., that in some pathogenic *E. coli* strains a truncated version of H-NS (lacking the DNA-binding domain) was found to disrupt the correct silencing function of H-NS (Williamson and Free, 2005). Additionally, very recently published results from Gehrke et al. elucidated that an Lsr2 variant, that is defective in DNA binding, could also be used to interfere with the native silencing function of Lsr2 in *Streptomyces coelicolor* and *Streptomyces venezuelae* (Gehrke et al., 2019). The idea behind this different silencing disruption methods is that heteromeric complexes are created with the native XS protein and the DNA-binding-defective XS protein, which as a result are not able to form a correct nucleoprotein complex.

The establishment of a silencing nucleoprotein complex is described in the following using the to date best analyzed XS protein H-NS. Binding of H-NS to a high-affinity site initiates the protein multimerization and thus the nucleoprotein complex formation. The cooperative binding of

additional H-NS molecules, also to proximate low-affinity sites, leads to a nucleoprotein structure that enables silencing of the bound region (Lang et al., 2007). The N-terminal domain of H-NS contains two homodimerization domains: a head domain and a tail domain (Figure 5 AB, Ueguchi et al., 1996). While H-NS exists as a head-to-head dimer in solution, very high protein concentrations or binding to DNA leads to the formation of longer head-to-head / tail-to-tail filaments (Figure 5 AB, Arold et al., 2010; Badaut et al., 2002). Figure 5A shows a DNA region bridged by an H-NS-Hha nucleoprotein complex. Additionally, it was found that this nucleoprotein complex can be present in a linear filamentous form (Figure 5B, Grainger, 2016). Hha and its paralog YdgT (35 % identical) were shown to associate with the N-terminal domain of H-NS and contribute to the silencing of large AT-rich DNA region via an DNA-association using positively charged site chains (More details are given in the following chapter). However, in absence of H-NS, Hha does not show any DNA-binding abilities (Ueda et al., 2013).

The way how oligomerization of the XS proteins along the AT-rich DNA regions leads to silencing of the target genes can be explained in three different modes of action (Lim et al., 2012): The nucleoprotein complex can occupy the promoter region, so that the RNA-polymerase (RNAP) is not able to bind; the nucleoprotein complex can obstruct the elongation process (ρ-dependent termination); the bridging of more distant nucleoprotein complexes can trap the RNAP (Figure 6). For MvaT, Lsr2 and H-NS both bridging of DNA and filamentous DNA binding could be demonstrated as mode of action in different studies (Dame et al., 2005; Qu et al., 2013; Winardhi et al., 2012). Furthermore, detailed studies of H-NS revealed a connection between the availability of divalent cations and the mode of DNA binding *in vitro* (Liu et al., 2010). A bridging of more distant nucleoprotein complexes with each other is also important for a condensation of DNA and thus chromosome compaction, which leads to an effect reminiscent of the role fulfilled by histones in eukaryotic cells (Navarre, 2016). How the XS protein CgpS influences the DNA condensation, if it is involved in chromosome compaction and if it is able to form bridges and stiff filaments has to be analyzed in further experiments. An interesting method for this purpose would be the use of single-molecule magnetic tweezers, as described by Gulvady et al. (Gulvady et al., 2018). Here, the distance changes triggered by XS-mediated bridging of specific DNA fragments can be measured. The most common method for analyzing changes in DNA structure triggered by XS proteins is atomic force microscopy (AFM), which was used already to visualize DNA binding properties of H-NS (and H-NS-like proteins), MvaT and Lsr2 (Chen et al., 2008; Dame et al., 2000, 2005). However, first

investigations of the CgpS DNA binding mode using atomic force microscopy (Pfeifer and Frunzke, data not published) did not lead to clear results. Another interesting alternative to those two *in vitro* methods would be a live cell imaging using super resolution microscopy (SRM), which would allow to study the mode of CgpS binding *in vivo*. The visualization of H-NS using stochastic optical reconstruction microscopy (STORM) and photoactivated localization microscopy (PALM) was already used for the analyses of different NAPs in *E. coli* in 2011 by Wang et al. (Wang et al., 2011). Wang and colleagues could demonstrate, that the four NAPs IHF, HU, StpA and HU exhibit a scattered binding throughout the complete nucleoid. In contrast, H-NS formed compact clusters in the cells caused by its oligomerization ability, which underlines the involvement of H-NS in chromosome organization (Wang et al., 2011). However, the used microscopical methods only show the NAPs, but do not visualize the DNA itself, which would be a useful addition in order to analyze the structural changes caused from XSs. Some former experiments were conducted to visualize DNA with super-resolution microscopy using either DNA-intercalating (e.g. YOYO-1) or minor groove-binding dyes (e.g. SYTO-13), but all approaches were done *in vitro* (Miller et al., 2015; Persson et al., 2011). As reviewed by Kozma and Kele, life cell imaging with *in vivo* stained DNA is a really challenging approach, because DNA binding dyes can destroy the DNA topology and thus the binding behavior of different DNA-binding proteins as well as the cell physiology. Nevertheless, live cell images of *in vivo* stained DNA together with fluorescently labeled XS proteins could shed light into the whole nucleoprotein complex formation.

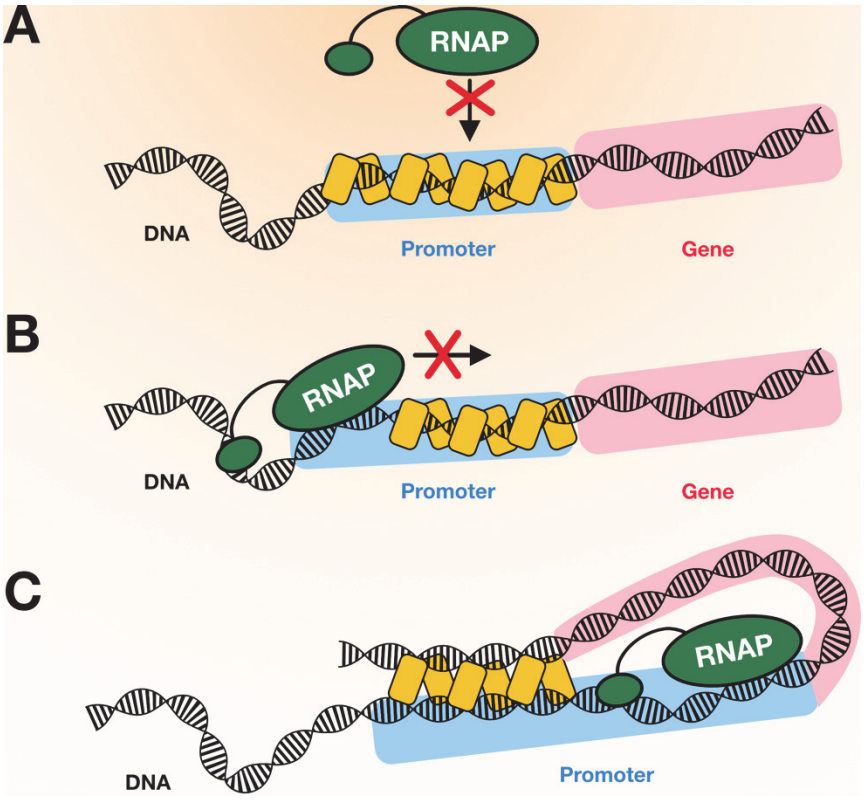


Figure 6: Possible promoter silencing mechanisms of XS proteins based on nucleoprotein complex formation. Three different modes of action have been proposed for XS proteins in order to inhibit transcription of target genes: (A) The XS protein occupies the promoter region of the target gene and thus inhibits the binding of the RNA polymerase (RNAP). (B) The promoter region of the target gene is only partially occupied by the XS protein, so the RNAP can bind and transcription starts, but the transcription elongation is blocked and cannot be successfully completed. (C) The binding of the XS protein leads to a bridging of the DNA close to the promoter region of a target gene and a proximal area. This bridging may result in a trapping of the RNAP. (Adapted from Lim et al., 2012).

2.2.2.4 Accessory proteins involved in xenogeneic silencing

In the enteric bacteria *Salmonella*, *Yersinia* and *E. coli*, H-NS-mediated silencing often requires additional accessory factors, like the protein Hha or other Hha-like proteins (Madrid et al., 2007). These proteins are able to interact with the N-terminal oligomerization domain and thus lead to a stabilization of the H-NS oligomer – especially when covering large AT-rich DNA regions (Figure 5A and 5B, Navarre, 2016). It could be shown by transcriptomic analyses that binding to low affinity regions and the establishment of large nucleoprotein complexes is not possible without Hha (Baños et al., 2009). However, short AT-rich DNA regions with high affinity binding of H-NS (e.g. inside the core genome) are bound Hha-independently and thus are not influenced in *hha* mutant strains. Furthermore, H-NS and its paralog StpA (58% identity) from *E. coli* are able to form heteromeric oligomers (Johansson et al., 2001). Boudreau et al. could

show recently, that Hha and StpA can stimulate the bridging of nucleoprotein filaments created by H-NS and thus affect the gene regulation of the target regions (Boudreau et al., 2018).

Regarding the binding mode of CgpS in *C. glutamicum*, we could not yet identify any other NAPs or accessory proteins associated to the silencer using a protein pull-down approach (Pfeifer et al., 2016). Furthermore, we could show *in vitro* using electrophoretic mobility shift assays (EMSAs) and surface plasmon resonance, that the purified CgpS protein is sufficient for DNA binding (Pfeifer et al., 2016; Wiechert et al., 2019, *submitted*). Additional studies demonstrated that CgpS is able to silence target promoters in a strain lacking the CGP3 prophage, when the *cgpS* gene was reintegrated. Thus, no other prophage factors appear to be necessary for silencing. Taking together these findings with the results of the pull-down assay and the in vitro binding analysis, it can be hypothesized that for a correct functioning silencing under our tested laboratory conditions, the XS protein CgpS could be sufficient. However, we analyzed the silencing ability of CgpS using only one single promoter in the prophage-lacking strain. Maybe more proteins are involved in the silencing of larger areas or low-affinity areas, as described for H-NS and Hha, or the silencing requires host-genes (Baños et al., 2009).

2.2.3 XS proteins display auto-regulatory properties

Negative autoregulation is very common in transcriptional regulation. This autoregulation may help speeding up the reactivity of regulatory circuits towards different inputs. It may further be useful to increase the gene expression stability against small perturbations and to set the expression rate of the specific transcription factor in a controlled equilibrium (Becskei and Serrano, 2000; Rosenfeld et al., 2002). In *E. coli*, 39 out of 96 tested transcriptional factors (~40 %) have negative autoregulatory properties (Thieffry et al., 1998). For the XS proteins H-NS and Rok a negative autoregulation could also be shown: different *in vivo* and *in vitro* studies elucidated H-NS as a protein that is involved in the regulation of its own gene (Dersch et al., 1993; Ueguchi et al., 1993). Here, H-NS acts as a repressor of its own expression and it was shown that the autoregulation is highly growth-phase dependent. Dersch et al. demonstrated that, during log-phase, the expression of *hns* is four-fold decreased. However, in the stationary phase, the H-NS amount was ten-fold increased compared to the start of the cultivation (Dersch et al., 1993). For Rok from *B. subtilis* it was published that this XS protein also shows negative auto-regulatory properties (Smits and Grossman, 2010). Smits and Grossman suggested that

this autoregulation could help adjusting the levels of Rok in case new AT-rich DNA is acquired horizontally, because higher amounts of foreign DNA would titrate Rok from its own promoter, so expression is increased until it reaches an equilibrium. The two paralogous XS proteins MvaT and MvaU of *P. aeruginosa* also show auto-regulatory properties (Castang et al., 2008). The authors could show that between both genes a reciprocal regulatory mechanism exists, because they also cross-bind the promoter region of each other. Furthermore, Vallet-Gely and colleagues could show in another study that a deletion of either *mvaT* or *mvaU* let to an increased expression of the other genes, suggesting a negative auto-regulation of each other and hints on a compensation mechanism during loss of one XS (Vallet-Gely et al., 2005).

Our data also suggested a negative auto-regulation for the XS CgpS, since ChAP-seq analysis revealed that CgpS completely covers its own promoter region (Pfeifer et al., 2016). Interestingly, for this promoter region - the intergenic area between *cgpS* and its neighboring gene (cg1967) - three transcriptional start sites (TSS) were published (Pfeifer-Sancar et al., 2013). However, in a recent study, new TSS were analyzed under prophage-inducing conditions, namely in samples treated with MMC (Wiechert et al., 2019, *submitted*). Here, in total 8 TSS were found inside the intergenic area between *cgpS* and cg1967 on the same strand as *cgpS*. A previous study from Mentz et al. predicted two small RNAs encoded in this intergenic area (Mentz et al., 2013). However, the TSS (both, new and old data analyses) do not fit exactly to the predicted small RNAs. We performed time-resolved chromatin-affinity precipitation and sequencing (ChAP-seq) analyses with the prophage encoded XS CgpS, under SOS-dependent prophage inducing conditions, to analyze the changes inside the bound regions (Hünnefeld et al., 2019a). A deeper focus on the intergenic region between *cgpS* and cg1967 revealed an interesting peak shape. The single large peak appears to be composed of several smaller peaks (Figure 7). These smaller peaks display different peak maxima in all different analyzed time points. Thus, also the shape of the summed up big peak changes through the time series. This finding could hint on either further small regulatory RNAs, encoded in between these two genes, or also on different possible transcript sizes for the mRNA of CgpS. Both of these options imply interesting further layers of regulation of the *cgpS* expression. One method to further analyze the presence of different *cgpS* mRNAs or other small regulatory RNAs would be to conduct transcriptome analysis (e.g. with mRNA-sequencing) under the same conditions and a comparison with the binding data.

In order to investigate the very detailed changes of subpeaks inside of a major peak, the normalization parameters for this analysis had to be adapted. Normally, to get manageable datasets for visualization and to get broad insights into the ChAP-seq data sets, we used a 'moving average' approach with a step-size of 50 bp and a window-size of 500 bp (Hünnefeld et al., 2019a; Hünnefeld et al., 2019b; Pfeifer et al., 2016). For the detailed analysis of the *cgpS* promoter region, however, a step-size of 1 was used together with a window-size of 10. This increases the resolution of the presented area highly and allows conclusions also on a single peak level that is nearly completely obscured using the larger window- and step-sizes. For a detailed analysis of large peaks, which probably span multiple smaller peaks, it should always be considered to decrease the moving average parameters in order to get a clearer insight.

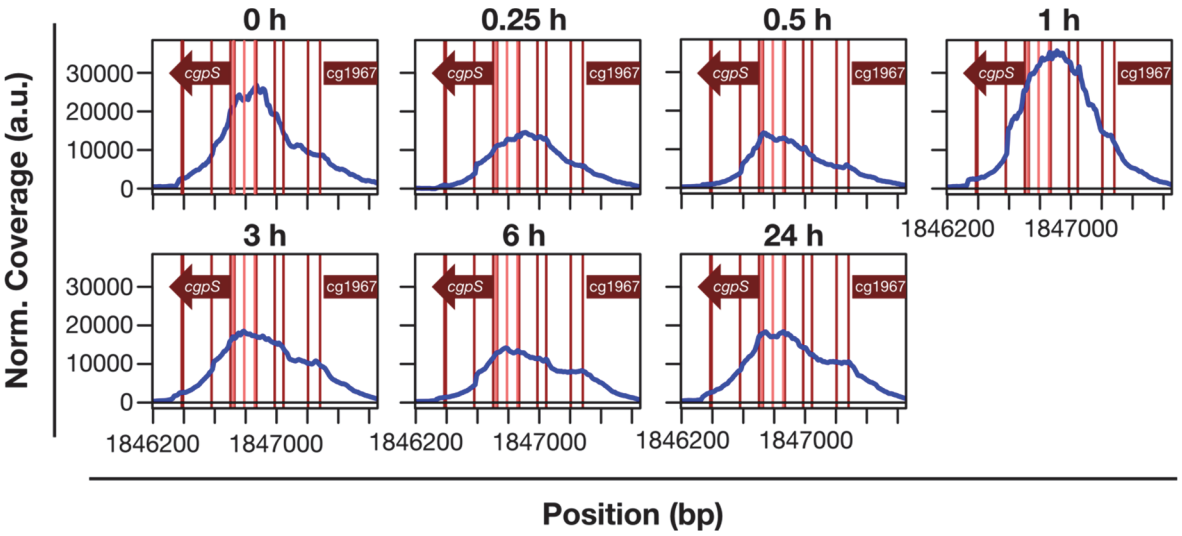


Figure 7: CgpS binding to its own promoter region. The binding of CgpS towards the intergenic region between itself and the neighbor gene cg1967 is displayed with the different time points tested in the ChAP-seq time series. The graphs represent zoomed-in versions of the ChAP-seq experiments conducted in Hünnefeld et al., 2019a. Further details regarding the growth conditions and experimental procedures can be found in this manuscript. The blue lines represent the binding coverage of CgpS, the light-red lines represent the TSS published by Pfeifer-Sancar and colleagues (Pfeifer-Sancar et al., 2013) and the dark-red lines are newly analyzed TSS under MMC addition extracted from Wiechert et al., 2019, *submitted*.

2.2.4 Physiological impact of different binding regions of XSs

The main focus of previous research on the function of XS proteins was - besides the structure-biological analysis of the binding mechanisms – to identify the targets regulated by XSs. The most obvious approach of this investigation is to identify binding sites in promoter regions, as these may have a direct influence on gene expression. However, in addition to the binding of

XS proteins to promoter regions of horizontally acquired or core genes, global binding pattern analyses showed that additional intergenic binding is also detectable. Depending on the silencing mechanism these bindings could have different functions. As described previously (Chapter 2.2.2.3), there are different ways how XS proteins can repress certain genes: (i) Blocking of the promoter region, so that RNAPs are not able to bind, (ii) bridging of the DNA, (iii) intragenic binding to disrupt correct elongation of transcription. The two latter options could involve intra- and intergenic binding. Nevertheless, intra- and intergenic binding positions could also hint on further functions of XSs. During our first characterization of the XS CgpS and its binding mechanism in *C. glutamicum*, the distribution of the peak maxima was: 60 % in promoter regions, 31 % intragenic and 9 % intergenic (Pfeifer et al., 2016). For H-NS in *E. coli* it could be shown that 46 % of all suppressed transcripts are intragenic transcripts. These intragenic transcripts can, on the one hand, be small regulatory RNAs or regulatory anti-sense transcripts, but on the other hand it can also be silencing of spurious transcription of non-coding RNAs (ncRNAs) or non-sense RNAs, which can be very detrimental for the bacterial cell (Singh et al., 2014). Besides the risks associated with horizontally acquired detrimental genes (as described previously, Chapter 2.2.1), AT-rich DNA poses another threat that can lead to a reduction in the fitness of bacterial cells. Because of the high AT-content, these DNA regions often contain intragenic promoter-like structures that sequester RNA polymerases (Lamberte et al., 2017). Lamberte and colleagues could demonstrate that this leads to a titration of the RNA polymerases and thus to a decreased expression of host genes. Furthermore, they could show that these intragenic promoter structures are special targets of H-NS, which in turn can prevent this expression and contributes to the fitness of the host cell. Another study from Singh et al. could demonstrate that this targeting of intragenic promoter-like structures also occurs in case of Lsr2 and MvaT (Singh et al., 2014). Here, they stated that 58 % of the MvaT bound putative promoter regions and 78 % of the Lsr2 bound putative promoter regions where located inside of genes. From this it can be deduced that the 31 % intragenic binding of CgpS in *C. glutamicum* might have a similar function in suppression of intragenic transcription.

Another hypothesis regarding the intra- and intergenic binding of XS proteins is that these bindings contribute to the DNA structuring behavior of XSs. For the CgpS global binding pattern analysis under uninduced conditions (Pfeifer et al., 2016) as well as under inducing conditions during the time series (Hünnefeld et al., 2019a), we could find binding sites mainly close to the terminus (where also the CGP3 prophage is located) as well as closer to the ori region.

Furthermore, in the areas flanking the CGP3 area no binding could be detected. This is more clearly visible in the prophage induced state, six hours after MMC induction, when the binding is shifted towards host targets. By comparing these results with global binding pattern analyses of other XS proteins, a clear trend towards this behavior is detectable. For Lsr2 in *S. venezuelae*, ChIP-seq analysis revealed an over-abundancy of binding sites inside the core region (close to the origin of replication) as well as some binding inside the right and left arm (Gehrke et al., 2019). However, in between those 'binding-clusters' there are also unbound areas detectable, similar to the CgpS binding pattern. For the XS protein Rok from *B. subtilis* a pattern is also detectable, which shows high binding inside the prophage areas, close to the terminus as well as binding close to the ori region (Smits and Grossman, 2010). Interestingly, for Rok an interaction with the ori-binding replication initiation protein DnaA could be demonstrated, which is required for the successful binding of DnaA at four different genomic regions (Smith and Grossman, 2015). For the binding of Rok, furthermore, areas flanking the prophages are detectable, where nearly no binding occurs. The described specialized binding patterns of these three XS proteins hint on an involvement in the DNA structuring and condensation process, in addition to the known silencing effect on foreign DNA. The special binding patterns might lead to dense packed regions alternating with less packed regions, leading to a specific pattern. For H-NS and MvaT/U, however, no such patterns could be detected. The binding peaks of these XS proteins are very evenly distributed along the whole genomes (Castang et al., 2008; Fukui et al., 2016).

2.2.5 Role of XS in the regulation of prophage life cycles

XS proteins preferentially bind AT-rich horizontally acquired DNA regions in the bacterial genome and are necessary for a successful integration of viral DNA into host regulatory circuits (as described previously). This properties hint on an involvement of XS proteins in the regulation of the phage life cycle. In a recent review, we summarized what is already known regarding the effects of XS of all four categories on prophages and partially on their induction (Table 1). Surprisingly, the impact of XS proteins on the phage life cycle has not been explicitly studied. For the *M. tuberculosis* XS Lsr2, for example, only binding sites were reported but no further effects were characterized (Gordon et al., 2010). Furthermore, it was shown that Rok from *B. subtilis* has an influence on gene expression of the SPβ-phage and is able to repress the excision of the mobile element ICEBs1 via excisionase-repression (Albano et al., 2005; Smits

and Grossman, 2010). A direct connection between prophage induction and the XS protein H-NS from *E. coli* and *Shewanella oneidensis* as well as the XS proteins MvaT/U from *P. aeruginosa* could also be demonstrated in previous studies (Castang and Dove, 2012; Hong et al., 2010; Li et al., 2009; Zeng et al., 2016).

Table 1: Silencing of phage elements in bacterial genomes. This table was extracted from (Pfeifer et al., 2019).

Type of silencer	Host strain	GC of host (%)	Prophage-like element	Length (kb) of phage	GC of phage (%)	Reference
H-NS	*E. coli* K-12 BW25113	50.8	Rac (cryptic)	23.1	47.1	(Hong et al., 2010)
	S. onediensis MR-1	45.9	CP4So (cryptic)	36	43	(Zeng et al., 2016)
MvaT	*P. aeruginosa* PAO1	66.6	Filamentous phage Pf4	15.7	58.7	(Castang and Dove, 2012; Li et al., 2009)
Rok	*B. subtilis* 168	43.5	Prophage region 4	8	35.8	(Smits and Grossman, 2010)
			Prophage region 5	20.7	37.5	
			Prophage region 6	34.8	36.1	
			SPβ	134.4	34.6	
Lsr2	*C. glutamicum* ATCC 13032	53.8	Cryptic prophage CGP1	13.5	47.1	(Pfeifer et al., 2016)
			Cryptic prophage CGP3	186.0	48.4	
	M. tuberculosis H37Rv	65.6	Prophage region 1, Rv1573-1588c, (Rv1582c)	10.5 (1.4*)	66.2 (62.5*)	(Gordon et al., 2010)
			Prophage region 2, Rv2645-2664, (Rv2658-2659c)	12.3 (1.5*)	66.2 (63.5*)	

*In case of *M. tuberculosis* also the bound genes were considered and are indicated in brackets.

Interestingly, depending on the genetic composition of the bacterial strain, some XS proteins are essential for the viability of the bacterial host. In *E. coli* K12, H-NS silences the prophage Rac (Table 1). It was shown that a deletion of *hns* leads to a de-repression and thus to an induction of Rac, which as a consequence leads to cell lysis (Hong et al., 2010). Furthermore, in *P. aeruginosa*, a deletion of both XS protein encoding genes, *mvaT* and *mvaU*, is only possible

in mutants with an impaired Pf4 prophage. A double deletion in a strain with an intact prophage Pf4 led to increased expression of prophage genes, to the production of phage particles and to cell lysis (Castang and Dove, 2012). Here, the viability of the strains is dependent on the silencing ability of each XS protein. The XS protein CgpS is only essential for the survival of *C. glutamicum* in the presence of the cryptic prophage CGP3, because CgpS silences this prophage element and an induction would lead to cell death (Pfeifer et al., 2016). In a strain lacking CGP3 (e.g. MB001), CgpS is not required (Baumgart et al., 2013).

In summary, it can be deduced that the essentiality of XS proteins is tightly coupled to the presence and the type of prophages of the particular host strain. If a de-repression of this prophage element would lead to and induction causing cell death or cell lysis, the respective XS protein appears to be essential.

In our studies, we could characterize the prophage-encoded Lsr2-like XS protein CgpS from *C. glutamicum* ATCC 13032 as a prophage silencer, especially targeting the cryptic prophage CGP3 (Pfeifer et al., 2016). We could show that a disturbance of the silencing ability (Chapter 2.2.6) leads to prophage induction and thus to cell death, which makes the XS essential to maintain the lysogenic state of CGP3. However, a genome reduction project with *C. glutamicum* ATCC 13032 could show that a complete removal of CGP3 (including *cgpS*) is also possible without any significant impact on cellular fitness (Baumgart et al., 2013). Genome-wide binding profiling of CgpS showed that besides the main target, CGP3, also some other (putatively) horizontally acquired elements are bound by CgpS (Pfeifer et al., 2016). But this insight was very static, because the binding status was only observed at one time point and experimental condition. Hence, no conclusions regarding the dynamic processes behind prophage induction could be made. In order to get a comprehensive insight into the binding dynamics of CgpS during the process of CGP3 induction, we performed a time series ChAP-Seq experiment. For the same samples, we also performed whole genome sequencing to have a suitable input control for the ChAP-seq experiments and to determine the prophage copy number, and a proteome analysis (using LC-MS) to determine amount of CgpS at the different tested time points (Hünnefeld et al., 2019a).

Our results indicated a dynamic binding behavior of CgpS in response to the SOS-dependent prophage induction using MMC. During the first 30 minutes after the MMC-addition, we could observe a direct correlation between the CgpS protein level and the binding of CgpS. Both

simultaneously decreased. Interestingly, the protein level dropped fast (already after 15 minutes), which would mean that CgpS is quickly degraded by a so far unknown mechanism. However, to draw conclusions in this direction, it would be necessary to perform protein synthesis and degradation assays.

Trötschel and colleagues performed a protein turnover quantification experiment with *C. glutamicum* ATCC 13032, cultivated at 30°C and 40°C, using a pulse-chase experiment with proteins labeled with both ^{15}N and ^{13}C (Trötschel et al., 2012). The XS protein CgpS could only be identified, with their methods, in one sample at one specific time point, so nothing can be concluded regarding the turnover of this protein. Furthermore, it would be necessary to perform these experiments under CGP3-inducing conditions to compare the turnover rate with our experimental data. Analyses of the Clp-proteases in *C. glutamicum* ATCC 13032 revealed the presence of two putative Clp sub-units encoded on the CGP3 prophage: NCgl1689 (Cg1982) and NCgl1716 (Cg2009)(Engels, 2004; Lüdke, 2007). However, these two proteins were not yet characterized in detail. The host-encoded Clp-protease proteolytic subunits ClpP1 and ClpP2, as well as the regulatory subunit ClpC were further analyzed (Engels et al., 2004). Engels and colleagues could show that the transcriptional regulator ClpR is an activator for those three Clp-protease subunits. In her doctoral thesis, Sabine Engels tested the effect of an MMC addition to wild type cells and compared these with a Δ*clpR* strain. Here, differences in growth as well as in the cell morphology could be detected (Engels, 2004). Taken together, these results could hint indirectly on a connection between SOS-dependent prophage induction and Clp-protease activity. However, the effect on CgpS and a direct connection have to be addressed in further experiments.

Following the course of our ChAP-seq time series further, a high increase in CgpS binding towards the CGP3 area, could be detected 1 h after MMC addition (Hünnefeld et al., 2019a). Astonishingly, this increase happened without a clear increase of CgpS protein amounts. This increased binding without increased protein levels could hint on the presence of higher CgpS amounts inside the bacterial cell that are not completely bound to DNA. Thus, the present CgpS can – triggered by a certain stimulus – associate to DNA without the need of previous production of high protein amounts. A certain stimulus could be, for example, the MMC-induced SOS-response that subsequently activates a yet undiscovered co-regulatory protein involved in CgpS binding or triggers a post-translational modification (PTM) of CgpS, altering the affinity towards DNA. This change in affinity could alternatively be responsible for higher

amounts of purified DNA, one hour after induction, and would explain the elevated total peaks size without requiring higher amounts of CgpS. An affinity change could further explain, the observation we made in the following time points of our ChAP-seq analyses: Six hours after MMC addition, the CgpS-level as well as the prophage copy number increased. Additionally, we could detect a redistribution of CgpS from mostly horizontally acquired DNA regions, towards binding inside the host genome (Hünnefeld et al., 2019a).

For H-NS in *Salmonella* it was shown that depending on the affinity towards different binding sites, this XS protein requires accessory proteins (like e.g. Hha) or it forms hetero-complexes (e.g. with StpA) to enable silencing of different targets (Chapter 2.2.2.4 and Baños et al., 2009). Up to now, we performed protein pull-down assays with CgpS under standard cultivation conditions only (independent of prophage induction), which did not reveal any clear evidence for CgpS-associated proteins (Pfeifer et al., 2016). Nevertheless, a further LC-MS analysis of CgpS pull-down samples hinted slightly on a possible association of CgpS and the replication initiation protein DnaA (Hünnefeld and Frunzke, data not published). An interaction of the XS protein Rok and DnaA in *B. subtilis* was already reported in literature and a similar principle could be present in *C. glutamicum* (Seid et al., 2017). However, a future pull-down time series with CgpS under SOS-dependent prophage-inducing conditions using MMC is planned, similarly to the time-resolved experiments already performed (Hünnefeld et al., 2019b).

Another hypothesis why we could detect an increase in CgpS binding sites towards host genome regions is that due to the higher amounts of CgpS inside the cell, the complete CGP3 region and the native binding sites of CgpS are completely saturated. Thus, CgpS not associated to CGP3 binds unspecifically towards different DNA regions, because of a general affinity towards DNA. A general affinity towards DNA was also shown for H-NS (Gulvady et al., 2018). What speaks against this hypothesis, however, is that although the CgpS level increases, the number of prophage copies increases further and therefore, CGP3 is always present in excess. Furthermore, the distribution of the CgpS binding sites does not appear arbitrarily, but specific targets are bound (described in more detail in Chapter 2.3.1).

In summary, with regard to the CgpS binding dynamics it can be said that although a CGP3 induction takes place, CgpS was always bound over all investigated points in time. This is consistent with various observations on the XS protein H-NS. Although there are many different events that lead to a loosening of the nucleoprotein complex or destroy a bridging by H-NS, it

always remains in a DNA-bound state (reviewed by Stoebel et al., 2008). However, the changes within the binding of CgpS may explain how the induction occurs. An initial loosening of the tight nucleoprotein complex and a later redistribution of the xenogeneic silencer seem to be sufficient for a successful induction. The fact that partial loosening of the nucleoprotein complex can be sufficient to enable further induction can be illustrated with the following recent example: Rangarajan and Schnetz demonstrated 2018 that an elongating RNAP is able to disrupt the binding of H-NS towards a repressed promoter region and thus lead to a de-repression of this promoter (Figure 8, Rangarajan and Schnetz, 2018). Subsequent to this H-NS-DNA complex disruption, the de-repressed promoter can be bound by further RNAPs. However, this mechanism is suggested to require high levels of RNAP directed towards the repressed DNA region, maybe given by a very active transcription of a neighboring gene.

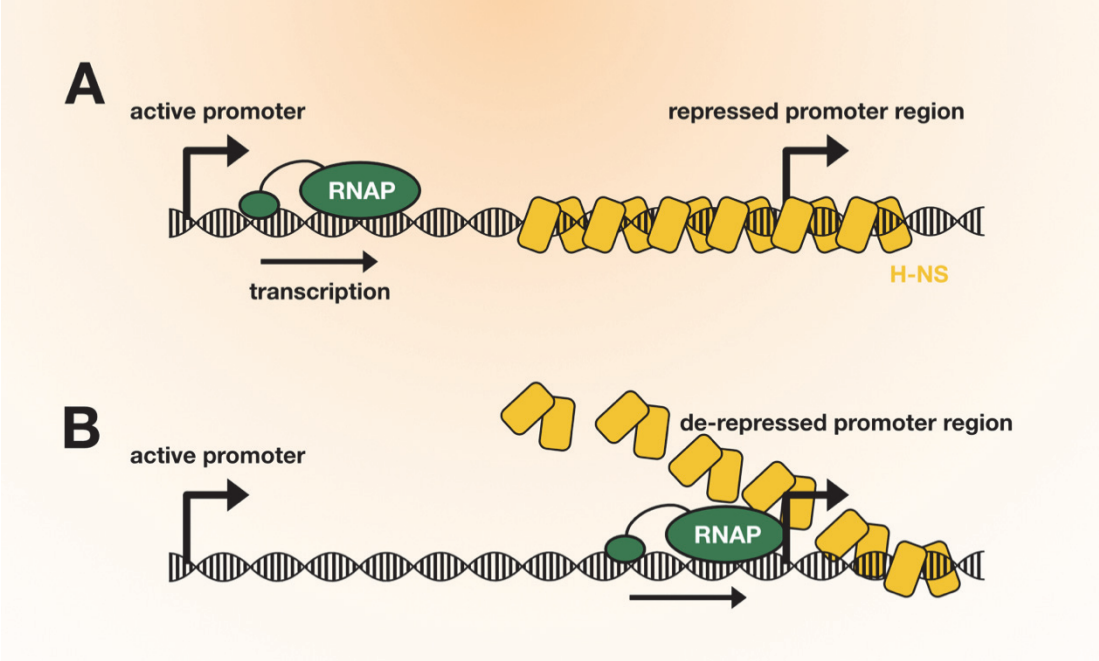


Figure 8: Transcription directed towards a repressed promoter leads to a de-repression. (A) An RNAP (green) started transcription at an active promoter and elongates towards a H-NS (yellow) repressed promoter region. (B) The transcription continues and leads to an elongation across the H-NS-DNA-complex, which leads to a disassembly that further on allows additional polymerases to bind this promoter. Thus, the promoter region is de-repressed. (Information and figure adapted from Rangarajan and Schnetz, 2018).

In order to have a suitable comparison and to get deeper insights into the dynamics of CgpS, it would be an interesting approach to conduct the time-resolved ChAP-Seq, genome sequencing and LC-MS measurements with further strains: the SOS-independent counter-silencing

approach using the truncated version of CgpS (Pfeifer et al., 2016), both under induced and uninduced conditions, as well as the strain with a strep-tagged CgpS under non-inducing conditions. This would help to disentangle effects resulting from CgpS nucleoprotein complex formation, which is impaired upon overproduction of CgpS-N, effects which derive from an SOS-response and effects directly correlated with prophage induction.

2.2.6 Counter-Silencing

The following chapter was written by me as a contribution to the review article "Impact of Xenogeneic Silencing on Phage-Host Interactions" from E. Pfeifer, M. Hünnefeld, O. Popa and J. Frunzke (2019), which is also part of this thesis (see Chapter 2.2). Here, an adapted extract is presented.

The formation of a nucleoprotein complex nucleating at AT-rich regions is a prerequisite for XS. Different studies focusing on counter-silencing mechanisms in various species have revealed that, upon activation of gene expression, XS proteins are not released from their target DNA; instead, remodeling of the XS-DNA complex enables RNA polymerase to bind and activate transcription. Different counter-silencing mechanisms are mainly based on other proteins binding in the upstream promoter region, thereby counteracting XS silencing. For instance, this mechanism has been nicely demonstrated by synthetic counter-silencing approaches, where operator sequences of specific transcription factors (TFs) were inserted into the upstream promoter region to counter-silence gene expression upon binding of the particular TF (Caramel and Schnetz, 1998). Several further studies demonstrated that different host-encoded TFs have been coopted – in the course of evolution – to act as counter-silencers. Examples include the response regulator PhoP, an essential activator of *Salmonella* virulence (Will et al., 2014), the AraC-family TF ToxT of *Vibrio cholerae* (Yu and DiRita, 2002), LeuO from *S. enterica* (De La Cruz et al., 2007), and the two MarR-type regulators RovA and SlyA of *S. enterica* and *Yersinia pseudotuberculosis*, which were shown to antagonize H-NS-dependent silencing of horizontally acquired genes (Heroven et al., 2004; Navarre et al., 2005). In a recent study, we also could show that the MarR-type regulator MalR of *C. glutamicum*, which controls genes involved in stress-responsive cell envelope remodeling, binds to several regions within the CGP3 prophage and is able to counteract SOS-dependent prophage induction (Hünnefeld et al., 2019b).

An alternative route for counter-silencing lies in the interference between XS proteins belonging to the same protein family. An interesting example has been provided for the unusual H-NS paralog Ler, which functions as a regulator of pathogenicity islands (locus of enterocyte effacement, LEE) in enteropathogenic (EPEC) and enterohemorrhagic (EHEC) *E. coli* strains (Bustamante et al., 2001; García et al., 2012; McDaniel et al., 1995; Mellies et al., 2011). Structural analysis emphasized that its function as a counter-silencer lies in differences in protein oligomerization as both H-NS and Ler bind to AT-rich regions (Winardhi et al., 2014b). Ler shows two different modes of DNA interaction: At low concentrations, Ler is able to increase DNA folding and wraps DNA; otherwise, with increasing concentration, Ler binds DNA in an unwrapped mode where Ler increases the rigidity of DNA similarly to the nucleoprotein filament formed by H-NS (Winardhi et al., 2014b). At these high concentrations, Ler displaces H-NS from the bound DNA and therefore overcomes the silencing of target regions. A further interesting example is provided with H-NST, a truncated derivative of H-NS lacking the DNA-binding domain. This XS protein was found to antagonize H-NS in enteropathogenic (EPEC) and uropathogenic *E. coli* by interfering with its oligomerization domain (Levine et al., 2014; Williamson and Free, 2005). Remarkably, this principle of silencer interference can be harnessed to study the function of essential XS proteins. Overproduction of the N-terminal oligomerization domain CgpS could be used – as described previously - to counteract CgpS silencing in vivo (Pfeifer et al., 2016). Our recent study, further demonstrated interference between different XS proteins. Here, the expression of other mycobacterial Lsr2 genes as well as introduction of *E. coli* H-NS led to XS interference at AT-rich regions, resulting in prophage induction. These findings are also supported by a bioinformatics analysis showing that different classes of silencers do not occur in the same species (Perez-Rueda and Ibarra, 2015). Altogether, these examples provide important insights how silencing and counter-silencing facilitate the expansion of regulatory networks in bacteria (Will et al., 2014).

A few studies suggest a variety of different mechanisms used by phages to counteract XS proteins. One example is provided by the 5.5 protein of the *E. coli* phage T7, which is able to antagonize H-NS function upon phage infection by interfering with the central oligomerization domain of H-NS (Ali et al., 2011; Arold et al., 2010; Liu and Richardson, 1993). Another example has been reported with the Mip protein (MvaT inhibiting protein), encoded by the LUZ24 phage of *P. aeruginosa*, which is able to inhibit the binding of the XS protein MvaT to DNA (Wagemans et al., 2015). In the case of *E. coli* T4 phage, interestingly, two different proteins were reported

to interfere with H-NS silencing. Deletion of the *motB* gene led to a decreased burst size (Patterson-West et al., 2018). The T4 protein Arn represents an interesting example of a phage-encoded DNA mimic protein and was shown to directly interact with *E. coli* H-NS (Ho et al., 2014). Here, Ho and colleagues showed that Arn mimics the shape and charge of double-stranded DNA, and thus interferes with the function of DNA-binding proteins, like H-NS. Interestingly, the DNA mimic proteins Ocr of the phage T7 and ArdA of the plasmid Collb-P9 were also reported to antagonize H-NS in a similar way (Melkina et al., 2016).

Finally, the direct interference with TFs or other proteins likely does not represent the only way to fight off XS. In a very recent study, Kronheim et al. highlighted the important role of small molecules secreted by bacterial hosts as weapons against phage infection (Kronheim et al., 2018). A link between these compounds and XS proteins does not necessarily exist, but a few examples suggest that small natural compounds – other than proteins – may also counteract XS. One class of compounds is represented by polyamides containing a biaryl motif. These polyamides especially target the minor groove of AT-rich DNA sequences and manipulate their topology (Brucoli et al., 2015). In their study, Brucoli et al. therefore suggest an effect of these chemical compounds on XS. A further example is the antiasthma medical zafirlukast, which was shown to inhibit the DNA-binding ability of Lsr2 in *M. tuberculosis* and *M. smegmatis* (Pinault et al., 2013). This compound was found to inhibit the growth of both mycobacterial strains and led to clarification of the bacterial cultures after three days. A direct interaction between zafirlukast and Lsr2 was revealed by MALDI-TOF analysis. However, we suggest that such interactions are specific for the particular protein since, in our hands, zafirlukast does not counteract the silencing mediated by the Lsr2-like protein CgpS (Hünnefeld and Frunzke, unpublished data).

2.3 Regulatory interactions between prophages and their hosts

2.3.1 The prophage encoded XS CgpS binds multiple targets in the host genome

The main targets of XS proteins are horizontally acquired elements inside the bacterial chromosome. These elements range from (pro-)phages or pathogenicity islands down to singles genes or small gene clusters that are completely integrated into host physiology and regulatory circuits. Compared with the other previously discussed XSs (Lsr2, H-NS, MvaT/U and Rok), one peculiarity of CgpS from *C. glutamicum* is, that it is encoded on the prophage CGP3

and not inside the host genome. Therefore, an interaction between CgpS and host-encoded genes is - strictly speaking - a phage-host interaction. For CgpS we found approximately one-third of all binding positions outside of the CGP3 area under uninduced conditions (Pfeifer et al., 2016). However, their majority was associated with horizontally acquired genes. The ChAP-seq studies for the characterization of CgpS were conducted during the mid-exponential growth phase. In order to get a clearer insight into binding of the XS under prophage-inducing conditions and into its dynamic binding patterns, we performed further ChAP-seq analyses at different cultivation time points (Hünnefeld et al., 2019a). Here, we observed a redistribution of CgpS from the CGP3 region towards multiple targets inside the core genome (as described previously, 2.2.5). At the maximal point of prophage activity (six hours after induction with MMC), the proportion of binding peaks inside and outside of CGP3 was nearly 1:1. To get an insight into their physiological function, we categorized the host-encoded CgpS-targets (Hünnefeld et al., 2019a). Here, nearly half of all the binding peaks were associated to genes involved in the following four functions: (i) signal transduction mechanisms (e.g. *divS*, *dtxR*, *glxR* and further transcriptional regulators), (ii) DNA replication/repair (e.g. *dnaN*, *dnaB*, *recF*), (iii) cell envelope biogenesis (e.g. *murA/B*, *wzx*), and (iv) amino acid transport and metabolism (e.g. *ilvN*, *gapA*, *dapA*). Taken together, our findings suggest that CgpS has a strong influence on DNA maintenance, signal transduction, metabolism, transport and the cell envelope. Interestingly, Gordon and colleagues could show 2010, that the XS Lsr2 from *M. tuberculosis* binds to genes associated with functions of similar categories (Gordon et al., 2010). They grouped these categories together as ‘intermediary metabolism/respiration’ and ‘cell-wall/cell processes’.

Furthermore, the XS protein Rok from *B. subtilis* regulates genes involved in cell surface and extracellular functions (Albano et al., 2005). MvaT from *P. aerigunosa* also showed binding sites inside genes that are involved in DNA replication, transport mechanisms, metabolism and the cell exterior (e.g. *wzz*, *wzy* and *wzx*, which are involved in O-antigen establishment) (Castang et al., 2008). For H-NS from *E.* coli, the categories of bound genes at mid-exponential growth are with 20 different ones very diverse. However, binding sites in promoter regions of genes coding for proteins involved in signal transduction (e.g. transcriptional regulators: *rcsB*, *evgA*, *bglJ*), transport mechanisms (e.g. arabinose-H^+-symporter, *araE*; multidrug efflux pump, *mdfA;* or shikimate-H^+-symporter, *shiA*), metabolism (e.g. *ilvH*, *gadA* and *gadB*) and cell envelope biogenesis (e.g. *wza*, *ompC* and *ftsQ*) were also represented (data extracted and analyzed from Kahramanoglou et al., 2011). Nevertheless, binding inside genes of these categories was only

detectable additionally to the typical XS binding behavior to AT-rich and horizontally acquired genes. Furthermore, it is an interesting fact that, in contrast to all host-encoded silencers, which bind to host genes under normal growth conditions, the prophage encoded XS protein CgpS binds host genes under prophage inducing conditions only. Maybe this binding behavior allows the CGP3 prophage to manipulate its host for its own purposes. In a recent study we could show that Lsr2s and Lsr2-like proteins exist in host-encoded as well as phage-encoded variants (Pfeifer et al., 2019). While the bacterial versions of Lsr2 are highly conserved among different actinomycetes, there is a high variability in the predicted secondary structure of the phage-encoded versions. This might indicate that phage-encoded XS proteins were adapted to the needs of the phage, or adapted to fit to the genomic composition of the particular host, during evolution and gained new functions. CgpS, however, belongs to the 'bacterial-cluster' of Lsr2 proteins. This could indicate a recombination event in the course of which *cgpS* was integrated into the CGP3 prophage region. Thus, this XS protein might represent an example for a hijacked version of an Lsr2-like protein.

2.3.2 Interactions of XS proteins and the bacterial origin of replication

One identified CgpS target of special interest is the origin of replication (ori) from *C. glutamicum* (Hünnefeld et al., 2019a). The ori region is characterized by a particularly high AT-content and contains multiple so-called DnaA boxes - areas that are recognized by the replication initiation protein DnaA (consensus for *C. glutamicum*: 5'-TTATCCACA-3') (Luo and Gao, 2019; Mott and Berger, 2007). Already in the beginning of our time series ChAP-seq experiment (t = 0 h) minimal binding of CgpS towards the ori was detectable. However, for the following time points, this binding increased and reached its maximum six hours after the addition of MMC. Interestingly, binding towards the particular ori region could also be detected for other XS proteins and NAPs (reviewed by Wolanski et al., 2015). Recent ChIP-seq analysis of Seid et al. demonstrated overlapping binding patterns for the XS Rok and DnaA in *B. subtilis* (Seid et al., 2017). One of these overlapping binding sites was the ori of this bacterium. Since no effect on replication was detected under physiological conditions for a tested *rok* deletion mutant, the physiological function of XS binding to the ori region remains unclear. Interestingly, Atlung and Hansen found that the disruption of *hns* leads to a decreased ori-per-cell ratio and an increase in *dnaA* expression (Atlung and Hansen, 2002). The authors suggested an importance of H-NS for the binding of DnaA to low-affinity sites but were not able to show high-affinity binding of

H-NS towards the ori. Thus, they postulated an indirect context between H-NS and the detected replication differences. Further studies were able to demonstrate that binding of H-NS to the ori region of *E. coli* requires the accessory protein Hha as well as the Hha-paralog Cnu (previously known as YdgT) (Kim et al., 2005; Paytubi et al., 2004; Yun et al., 2012b, 2012a). Kim and colleagues further deduced out of their finding that a mutant lacking *hha* and *cnu* also showed a slight decrease in ori content. This hints towards a possible contribution of the Cnu/Hha/H-NS complex at *oriC* to generate optimal activity of replication. Two further examples of NAPs binding to *oriC* are the histone-like protein HU and the integration host factor IHF from *E. coli* (Bonnefoy and Rouvière-Yaniv, 1992; Filutowicz and Roll, 1990; Polaczek, 1990). IHF binds to a single binding site inside of *oriC* leading to stimulation of DnaA binding at low affinity sites and bending of the DNA (Bonnefoy and Rouvière-Yaniv, 1992; Grimwade et al., 2000). In contrast, binding of HU shows a more unspecific nature. However, Polaczek and colleagues demonstrated that a combination of HU and IHF binding leads to stabilization of the strand-opening at *oriC* prior to replication initiation (Polaczek et al., 1998).

The impact of CgpS binding to the ori in the process of prophage induction still remains elusive and represents an interesting target for future studies. One possible explanation would be that CgpS is involved in a protection mechanism. The addition of MMC leads to a detrimental DNA damage that can cause severe problems for the bacterium, especially at its replication initiation region. However, if CgpS would have a DNA protecting property, cells containing CgpS would show an increased fitness in the presence of MMC compared to cells lacking CgpS. Investigations from Baumgart and colleagues revealed that a prophage-free strain (MB001) lacking CgpS did not have any detriments compared to the *C. glutamicum* wild type strain (Baumgart et al., 2013). In contrast, the MB001 strain even featured a fitness advantage under MMC stress when compared to the wild type, because all prophage elements (including the inducible CGP3) were removed in this strain. Another hypothesis is that CgpS blocks the host *oriC* in the course of prophage induction in order to hijack the cellular machinery for its own replication. To investigate this, SRM with a fluorescently tagged DNA polymerase (DNAP), CGP3 area and ori-region could provide further evidence for this mechanism. However, as described previously, in vivo tagging of DNA for SRM was not successful. Thus, tagging of the ori-region could be conducted using a tagged ParB protein, that was shown in a recent study to bind an *oriC*-proximal *parS* cluster region (Böhm et al., 2019, not peer-reviewed preprint). For the detection of the CGP3 area it was shown to be possible to use a tagged variant of the adaptor

proteins AlpA, which binds specifically in the promoter region of the CGP3-encoded operon *alpAC* (Donovan et al., 2015). These tagged proteins in combination with a tagged variant of DNAP might give an insight into a possible re-localization of the DNAP upon prophage-inducing conditions. Another possibility, inspired by studies of Davis and colleagues, would be to profile the binding of the DNAP using ChIP-seq analysis with DNAP-specific antibodies (Davis et al., 2011).

Furthermore, CgpS may lead to structural change of the DNA that brings the ori region in close proximity to the CGP3 prophage element, which is in good agreement with the before mentioned binding of XS proteins close to ori and terminus. This close proximity would increase the probability that both, the replication initiator DnaA as well as the DNAP, bind to the CGP3 region to improve the replication of the prophage. A suitable method to analyze changes in genome architecture as well as to identify genome regions which are in close proximity, is represented by Hi-C (Lieberman-Aiden et al., 2009; van Berkum et al., 2010). The Hi-C method uses the ability of formaldehyde to covalently link adjacent DNA regions leading to internal cross-links in the chromosome. Afterwards the chromosome is digested with a restriction enzyme, the resulting sticky ends are filled with biotinylated nucleotides and the fragments are ligated. The earned products consist of fragments that were originally in close spatial proximity and can subsequently be sheared, purified with streptavidin and sequenced to earn a data set of adjacent DNA regions (Lieberman-Aiden et al., 2009)

Finally, binding of the XS protein CgpS towards the *C. glutamicum* ori could also be an indirect effect, because of the high AT-content of the ori region. Overproduction of CgpS and analysis of the effects on the chromosome replication, e.g. by whole genome sequencing and ori-to-terminus ratio calculation, can exclude this. For comparison, the CGP3 and *cgpS* lacking strain MB001 as well as MB001 harboring a plasmid-based *cgpS* can be used.

2.3.3 Identification of host regulators involved in prophage regulation

After characterizing the Lsr2-like protein CgpS as a XS involved in silencing of horizontally acquired DNA in *C. glutamicum* and analyzing its binding behavior, the next step was the search for further factors involved in the inducibility of the CGP3 prophage. For this purpose, we performed DNA affinity chromatographies with the promoter regions of five different CGP3 encoded genes and subsequently analyzed the obtained bound proteins with MALDI-TOF and

LC-MS (Chapter 4.5). In total, we were able to determine seven regulatory proteins binding towards CGP3. Their binding patterns and their effects on prophage induction were tested for three of them. However, for one regulator, we could find hints that it represents a putative connection between stress-responsive cell envelope remodeling and regulation of CGP3. Therefore, further studies focused on this regulator. This protein is the MarR-type transcriptional regulator MalR (Cg3315). MalR from *C. glutamicum* was already published in 2012 as a repressor involved in the regulation of the expression of the *malE* gene, which encodes the malic enzyme (Krause et al., 2012).

2.3.4 The MarR-type regulator MalR binds to the CGP3 prophage in *C. glutamicum*

Due to observed MalR binding to the CGP3 prophage, we suggested that it not only regulates the expression of the malic enzyme but also is involved in controlling further cellular processes. Thus, we conducted global binding profiling with ChAP-seq analysis and transcriptome profiling to gain genome-wide insights into the regulon of MalR (Hünnefeld et al., 2019b). In summary, our results emphasized MalR as a regulator involved in the stress-responsive cell envelope remodeling. Many direct targets of MalR that showed differentially regulation upon overexpression of *malR* are involved in mechanisms responsible for the composition of the corynebacterial cell envelope. Additionally, we demonstrated that a *malR* defective strain showed increased sensitivity towards different β-lactam antibiotics and that MalR overproduction led to alterations of the cell surface compared to the wild type. The most important outcome with regard to this thesis, however, was the verification of binding as well as a regulation of thirteen genes inside the CGP3 prophage and the proof that overexpression of *malR* can impede the SOS-dependent CGP3-inducibility (Hünnefeld et al., 2019b).

2.3.5 Regulation of horizontally acquired genes by MarR-type regulators

The presence of binding sites of the MarR-type regulator MalR inside the CGP3 prophage in *C. glutamicum* led to the assumption, that this might have a regulatory effect on CgpS silenced prophage genes. So far, different combinations of XS proteins and other proteins binding in similar DNA regions are known. This co-binding could either have cooperative effects improving silencing, maintain the established silencing (like e.g. the H-NS-StpA heterocomplex; Müller et al., 2006) or interfere with the XS nucleoprotein complex leading to counter-silencing (like e.g.

PhoP from *Salmonella enterica* serovar Typhimurium; Will et al., 2014) (illustrated in Figure 9). The latter mechanism is very important for microbial strain diversification because counter-silencing enables the host cell to expand its genetic capacity by activation of XS-silenced horizontally acquired DNA regions. Counter-silencing, either provoked by the host cell or by the foreign element itself, can happen in different manners (as discussed previously chapter 2.2.6). This chapter focusses on interaction of MarR-type transcriptional regulators with silenced DNA regions. A direct influence of a MarR-type regulator on the expression of horizontally acquired elements was so far only reported in one recent publication: The MarR-type regulator PamR from *B. subtilis* was shown to be involved in the regulation of prophage-encoded genes (De San Eustaquio-Campillo et al., 2017). However, more publications dealt with the analysis of the influence of gene expression via counter-silencing. The MarR-type regulator SlyA from *S.* Typhimurium (Buchmeier et al., 1997) and its homolog RovA from *Yersinia pseudotuberculosis* (Nagel et al., 2001) constitute the best analyzed examples of MarR-type regulators involved in counter-silencing of H-NS. Interestingly, it was shown that these regulators can function in two ways: on the one hand they act as classical transcription factors (activators/repressors), but on the other hand they can act via counter-silencing as activators of horizontally acquired DNA including genes important for virulence (Ellison and Miller, 2006; Heroven et al., 2004; Navarre et al., 2005, 2006). H-NS counter-silencing by SlyA requires PhoP, the response regulator of the PhoPQ two component system. Both proteins cooperatively remodel the H-NS-DNA nucleoprotein complex (Will et al., 2014).

In a very recent study, Will and colleagues conducted analyses regarding the evolution process of SlyA and RovA from transcription factors to counter-silencers (Will et al., 2019). The authors suggest that SlyA from *S.* Typhimurium and the protein RovA from *Y. pseudotuberculosis* are an example of parallel evolution from a former ancestor MarR-type regulator. Will and colleagues analyzed multiple homologs of the highly abundant (orthologs in 55 organisms) regulators SlyA/RovA regarding their structure and functionality. They found regulators showing functions fitting well to those of the ancestral MarR protein, like auto-regulation, regulating efflux pumps and inhibition by aromatic compounds. However, also different homologs that act as counter-silencers were found. The counter-silencing function of SlyA/RovA proteins requires high levels of bound protein because otherwise it is not possible to overcome silencing (Will et al., 2019). Thus, the authors suggested that the high expression levels of repressors would lead in some organisms to a 'hyperrepression' of existing genetic functions which are under control of those

proteins. Subsequently, these genes would get lost during evolution. In total, it can be concluded from the study of Will et al. that the regulatory networks of bacterial cells are highly adaptive. Evolution shapes the function of different regulatory units, so they fit perfectly to the current requirements of the cells. Furthermore, they suggested that in case of SlyA/RovA, these proteins inherited specific features which made them the most favorable candidates for the evolution towards counter-silencers. One of these specific features it their ability to react towards external stimuli and to recognize AT-rich DNA sequences. This reactivity towards external stimuli is a common feature of MarR-type regulators, to which also MalR belongs (Wilkinson and Grove, 2006). Furthermore, MalR shows binding towards AT-rich sequences, which would make it a suitable candidate for the evolution towards a counter-silencer (Hünnefeld et al., 2019b).

In order to investigate the evolution of counter-silencing, a recent study of our group using a transcription factor (TF)-based synthetic approach demonstrated that TF binding inside a CgpS-silenced promoter disturbed the silencing mechanism (Wiechert et al., 2019, *submitted*). In this study, we could demonstrate that slight changes in silenced promoter regions could lead to the presence of binding motifs recognized by transcription factors. Furthermore, we elucidated that the relative position of the TF binding (compared to the TSS and the nucleation site of the XS protein) showed strong influence on the inducibility. Thus, the counter-silencing is a good possibility for the bacterium to make use of beneficial traits encoded on horizontally acquired DNA using its own TFs in a very finely adjustable manner.

However, although MalR from *C. glutamicum* meets important criteria with its affiliation to the MarR-type regulators and binding sites inside the CGP3 prophage, the protein did not show any counter-silencing activities under our tested conditions (Hünnefeld et al., 2019b). On the contrary, the inducibility of CGP3 was reduced during overproduction of MalR under SOS-dependent prophage-inducing conditions, and, on transcriptome level, the majority of MalR-bound prophage genes was downregulated. There are four hypotheses explaining the behavior of MalR regarding the non-detectable counter-silencing of CgpS-silenced CGP3 genes: (i) MalR represents a co-silencing protein that is important under specific cellular or environmental conditions to keep CGP3 in a silent state. (ii) MalR is a transcriptional regulator recognizing very AT-rich palindromic motifs inside the genome, which leads to an accidentally binding inside of the prophage CGP3. Nevertheless, this could be a first step towards the evolution of as new counter-silencer. (iii) MalR is a counter-silencing protein that requires a cooperation partner to

successfully overcome silencing, like it was demonstrated for SlyA and PhoP (Will et al., 2014). In order to find such a co-counter-silencer, protein pull-down assays could be conducted with MalR. (iv) Another possibility would be that MalR – as typical for MarR-type regulators – can bind an effector molecule which leads to a conformational change. Typically this effector binding would lead to a dissociation of the regulator and thus to a de-repression of its targets genes (Deochand and Grove, 2017). However, for the MarR-type regulator CbaR from *Comamonas testosteroni* it was shown, that one specific ligand (modified benzoate) improves its DNA-binding (Providenti and Wyndham, 2001). It is possible that MalR exhibits a similar mechanism and that ligand binding could alter DNA affinity, which in turn might favor counter silencing. To prove this, an affinity change as reaction toward different putative ligands could be tested using e.g. surface plasmon resonance with a selected prophage DNA fragment.

MalR does not represent a homologue of SlyA but a BLASTp search (Altschul et al., 1990) could identify a regulator in *C. glutamicum* that shows 30 % identity with the intermediate part of SlyA from *S.* Typhimurium: RosR (Cg1324). RosR is a MarR-type regulator already published as an important regulator for the resistance of *C. glutamicum* against H_2O_2, because a strain lacking *rosR* showed a higher sensitivity towards this compound (Bussmann et al., 2010). Bussmann et al. analyzed the RosR regulon on transcriptome level using DNA microarrays of the wild type and the *rosR* deletion strain, however, nothing was stated regarding the CGP3 prophage. Because of the small sequence identity compared to SlyA, the reactivity towards environmental triggers and the presence in the class of MarR-type regulators, RosR could be an interesting candidate for further analysis of putative counter-silencer candidates in *C. glutamicum*. Nevertheless, *C. glutamicum* encodes in total nine MarR-type regulators, which could all be interesting targets for further studies in order to identify further potential counter-silencing or cooperatively prophage-binding proteins (Brinkrolf et al., 2007).

For future approaches, the significance of MalR binding towards the CGP3 region would be another interesting aim to investigate. To clarify if there is a direct connection between the stress-dependent cell-envelope remodeling and the prophage CGP3, different stresses could be tested with regard to the prophage inducibility, in cells lacking and overexpressing *malR*. Furthermore, investigations regarding the physiological role of MalR should be conducted with focus on which conditions trigger *malR* expression and under which growth conditions (except of the beforementioned β-lactam antibiotics) this MarR-type regulator is essential for cell viability. These experiments would on the one hand shed light into the physiological function

of MalR, but on the other hand it could also hint on the relevance of MalR-binding inside of CGP3 and if there are cellular conditions that lead i.a. to an active counter-silencing. One further approach to analyze the interplay of CgpS and MalR at specific promoters would be, to construct a reporter, which consists of a promoter that is known to be bound by CgpS and MalR and a fluorescent protein under the control of this promoter. This construct can then be tested under different physiological conditions for a functional silencing or counter-silencing or co-silencing.

In conclusion, we demonstrated that MalR binds in the cryptic prophage CGP3 and represses single prophage genes. At elevated protein levels, it is furthermore able to reduce SOS-dependent CGP3 induction. Nevertheless, the exact interaction between MalR, CGP3 and CgpS and the question, if MalR is inside an intermediate evolutionary step towards becoming a counter-silencer, still remains unclear and should be addressed in future investigations.

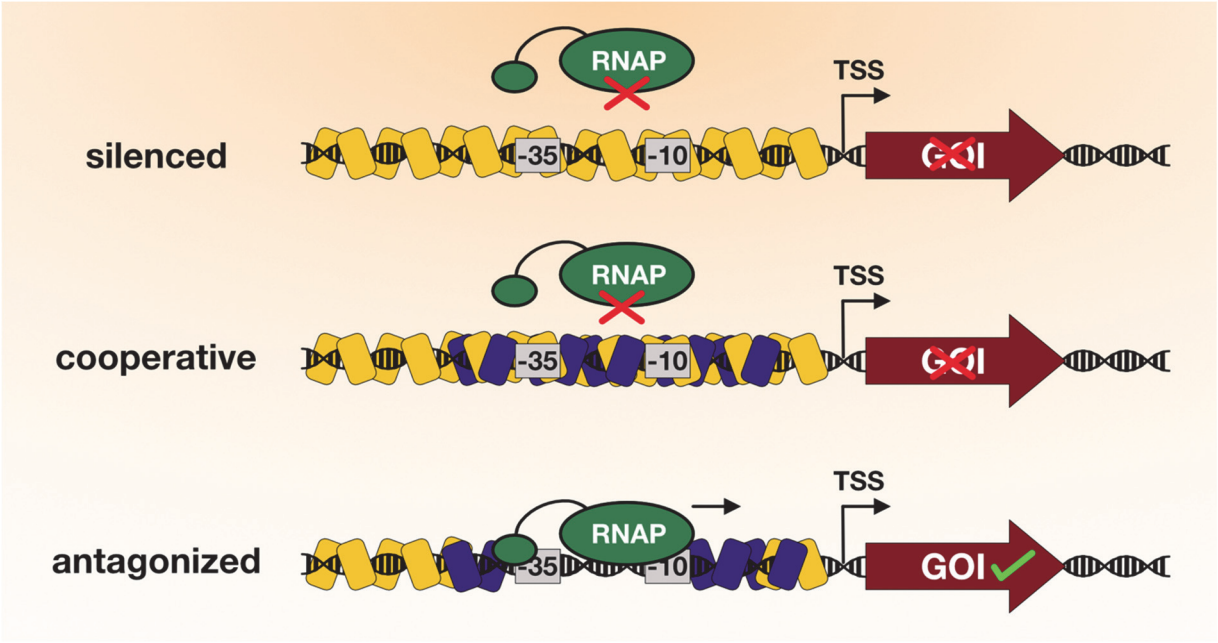


Figure 9: Different modes of interaction between a XS protein and another DNA-binding protein. The three examples represent different modes of interaction. The first model shows exemplarily the silenced state, where a XS protein binds to a promoter region and inhibits the binding of the RNA polymerase (RNAP) and thus leads to transcription inhibition of the following gene of interest (GOI). In the second row a possible co-silencing protein is displayed. This could be e.g. an accessory protein or a XS homolog that cooperatively bind with the XS protein and stabilizes the silencing of the GOI. In the third row, a possible counter-silencer is represented. Here, the binding of the additional proteins leads to a loosening or disruption of the silencing complex and hence enables the correct binding of the RNAP and an activation of transcription of the GOI.

2.4 Conclusion and future perspectives

The present thesis provides insights into the regulatory mechanism underlying prophage silencing and prophage induction in *C. glutamicum*. Special attention is directed to the small nucleoid-associated protein CgpS, which was characterized as an Lsr2-like protein responsible for the xenogeneic silencing of the prophage CGP3 (Pfeifer et al., 2016). While previous approaches only provided a snap-shot view on the binding of XS proteins, this thesis presents for the first time a time-resolved analysis of the global binding profile of an XS protein under prophage-inducing conditions to decipher the dynamic role of XS proteins in the process of prophage induction (Hünnefeld et al., 2019a). Additionally, this work presents a comprehensive overview of the state-of-art regarding xenogeneic silencing in bacteria, counter-silencing, and the distribution of Lsr2-like xenogeneic silencers on phage genomes (Pfeifer et al., 2019). In this review, the currently known mechanisms regarding XS binding abilities and their importance for (pro-)phage-host interactions are highlighted. In addition, different ways of how silencing can be overcome, and a bioinformatic analysis regarding the distribution of genes encoding XS proteins inside bacterial and phage genomes are presented.

Another section of this thesis, an analysis of host-encoded regulators with binding sites and thus regulatory influence on the CGP3 prophage, elaborates the deep integration of the prophage element CGP3 into the host regulatory network. Here, the MarR-type regulator MalR was described as a host-encoded regulator linking stress-responsive cell envelope remodeling to the regulation of prophage elements (Hünnefeld et al., 2019b). Generally, the study of genome-wide binding patterns of cellular transcription factors provides unprecedented insights into complex regulatory networks and network interference. In this context, the analysis of (pro-)phage-host regulatory interaction, sheds light on the evolutionary expansion of regulatory circuits and on the integration of horizontally acquired DNA into host regulatory networks.

The high abundance and broad distribution of XS proteins of all known classes in different bacteria and on MGEs make these proteins interesting targets for further studies (Pfeifer et al., 2019). To date, many things are known about the structure of XS proteins, their importance in horizontal genes transfer and their influence on target DNA. One of their most intensely analyzed functions is the influence on the regulation of pathogenicity genes because this is a crucial function involved in different medical and health problems (Wahl et al., 2018). What

remains touched only incidentally, however, is a complex network of interactions between XS proteins, host-encoded regulators and horizontally acquired DNA. The use of beneficial traits encoded on foreign DNA regions requires well organized regulatory hierarchies that enable a correctly coordinated gene expression. Multiple modern analysis methods (e.g. all -omics techniques) combined with a time-resolved experimental setup, as introduced in this thesis, enable the (re-)investigation of already described regulators and their modes of action, which leads to a more in-depth insight into these regulatory hierarchies. Thus, the method presented in this thesis can facilitate the gain of knowledge regarding the binding and regulatory activities of every XS protein and other involved regulators to study regulatory interference under different environmental conditions.

The function of XS proteins to silence horizontally acquired DNA regions, including prophages, makes them further an interesting target for antibacterial drug development. Counter-acting the silencing of a prophage e.g. by disrupting to function of the silencer, would lead to an induction of the prophage and thus potentially to lysis of the bacterial cell. For *M. tuberculosis*, it was published, that a particular anti-asthma drug (Zafirlukast) is also able to disrupt the Lsr2-DNA complexation and thus led to defective growth of the bacterium (Pinault et al., 2013). Future approaches could also aim at finding more XS disrupting chemicals that could function as potential antibacterial compounds.

XS proteins are involved in many different cellular processes. Gehrke et al. demonstrated, e.g. in a very recent publication, that Lsr2 of *S. venezuelae* is involved in the regulation of specialized metabolite production (Gehrke et al., 2019). Furthermore, in the course of this thesis, we could find that under specific conditions the Lsr2-like protein CgpS in *C. glutamicum* is redistributed from mainly horizontally acquired DNA areas towards different host targets (Hünnefeld et al., 2019a). Comprehensively, these examples show that XS proteins inherit a high relevance for the regulation of multiple cellular processes and that there is still a need for further characterizations of different XS protein or – more general – different NAPs, to deeper investigate this relevance.

A further important question regarding the role of XS proteins tackles the involvement of these proteins in phage defense. As described previously, it is a well-analyzed feature of all XS proteins to silence AT-rich foreign DNA, which protects the cells from detrimental genes encoded in these AT-rich areas. Nevertheless, no examples in literature can be found describing

a direct protective function of XS proteins against phage infection. Additionally, no studies investigating whether and how XS proteins bind to incoming phage DNA are currently available. Tackling this issue, spatiotemporal analyses of XS proteins during phage infection would help to elucidate a possible protective function and to discriminate whether XS proteins only show importance in maintaining lysogenic states and in silencing integrated foreign elements. Because XS proteins preferentially associate with AT-rich DNA regions, it would be further interesting to analyze the potential defense against phages with various GC-contents. Additionally, an interesting aim for future studies would be to elucidate the occurrence of interactions between XS and other phage defense systems. To date, many phage defense systems have been studied individually in different labs. A more integrative study would be required to analyze how different systems might potentially interact and complement each other. For example, the activity of XS proteins may potentially 'buy time' for the cell that RM or CRISPR systems could use to destroy the phage.

2.5 References

Albano, M., Smits, W. K., Ho, L. T. Y., Kraigher, B., Mandic-Mulec, I., Kuipers, O. P., et al. (2005). The rok protein of *Bacillus subtilis* represses genes for cell surface and extracellular functions. *J. Bacteriol.* 187, 2010–2019. doi:10.1128/JB.187.6.2010-2019.2005.

Alexeeva, S., Guerra Martínez, J. A., Spus, M., and Smid, E. J. (2018). Spontaneously induced prophages are abundant in a naturally evolved bacterial starter culture and deliver competitive advantage to the host. *BMC Microbiol.* 18, 120. doi:10.1186/s12866-018-1229-1.

Ali, S. S., Beckett, E., Bae, S. J., and Navarre, W. W. (2011). The 5.5 protein of phage T7 inhibits H-NS through interactions with the central oligomerization domain. *J. Bacteriol.* 193, 4881–4892. doi:10.1128/JB.05198-11.

Ali, S. S., Xia, B., Liu, J., and Navarre, W. W. (2012). Silencing of foreign DNA in bacteria. *Curr Opin Microbiol* 15, 175–181. doi:10.1016/j.mib.2011.12.014.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410. doi:10.1016/s0022-2836(05)80360-2.

Aravind, L., and Landsman, D. (1998). AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.* 26, 4413. doi:10.1093/NAR/26.19.4413.

Arold, S. T., Leonard, P. G., Parkinson, G. N., and Ladbury, J. E. (2010). H-NS forms a superhelical protein scaffold for DNA condensation. *Proc. Natl. Acad. Sci.* 107, 15728–15732. doi:10.1073/pnas.1006966107.

Atlung, T., and Hansen, F. G. (2002). Effect of different concentrations of H-NS protein on chromosome replication and the cell cycle in *Escherichia coli*. *J. Bacteriol.* 184, 1843–1850. doi:10.1128/JB.184.7.1843-1850.2002.

Badaut, C., Williams, R., Arluison, V., Bouffartigues, E., Robert, B., Buc, H., et al. (2002). The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA. *J. Biol. Chem.* 277, 41657–66. doi:10.1074/jbc.M206037200.

Balcazar, J. L. (2014). Bacteriophages as vehicles for antibiotic resistance genes in the environment. *PLoS Pathog.* 10, e1004219. doi:10.1371/journal.ppat.1004219.

Baños, R. C., Vivero, A., Aznar, S., García, J., Pons, M., Madrid, C., et al. (2009). Differential

regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS. *PLoS Genet.* 5, e1000513. doi:10.1371/journal.pgen.1000513.

Baumgart, M., Unthan, S., Kloß, R., Radek, A., Polen, T., Tenhaef, N., et al. (2018). *Corynebacterium glutamicum* Chassis C1∗: Building and Testing a Novel Platform Host for Synthetic Biology and Industrial Biotechnology. *ACS Synth. Biol.* 7, 132–144. doi:10.1021/acssynbio.7b00261.

Baumgart, M., Unthan, S., Rückert, C., Sivalingam, J., Grünberger, A., Kalinowski, J., et al. (2013). Construction of a prophage-free variant of *Corynebacterium glutamicum* ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl Env. Microbiol* 79, 6006–6015. doi:10.1128/aem.01634-13.

Becskei, A., and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature* 405, 590–593. doi:10.1038/35014651.

Bickle, T. A., and Krüger, D. H. (1993). Biology of DNA restriction. *Microbiol. Rev.* 57, 434–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8336674 [Accessed June 19, 2019].

Bobay, L.-M., Touchon, M., and Rocha, E. P. C. (2014). Pervasive domestication of defective prophages by bacteria. *Proc. Natl. Acad. Sci.* 111, 12127–12132. doi:10.1073/pnas.1405336111.

Böhm, K., Giacomelli, G., Schmidt, A., Imhof, A., Koszul, R., Marbouty, M., et al. (2019). Chromosome organization by a conserved condensin-ParB system in the actinobacterium *Corynebacterium glutamicum*. *bioRxiv*, 649749. doi:10.1101/649749.

Bonnefoy, E., and Rouvière-Yaniv, J. (1992). HU, the major histone-like protein of *E. coli*, modulates the binding of IHF to oriC. *EMBO J.* 11, 4489–4496. doi:10.1002/j.1460-2075.1992.tb05550.x.

Boudreau, B. A., Hron, D. R., Qin, L., van der Valk, R. A., Kotlajich, M. V, Dame, R. T., et al. (2018). StpA and Hha stimulate pausing by RNA polymerase by promoting DNA-DNA bridging of H-NS filaments. *Nucleic Acids Res.* 46, 5525–5546. doi:10.1093/nar/gky265.

Bouffartigues, E., Buckle, M., Badaut, C., Travers, A., and Rimsky, S. (2007). H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat. Struct. Mol. Biol.* 14, 441–448. doi:10.1038/nsmb1233.

Brinkrolf, K., Brune, I., and Tauch, A. (2007). The transcriptional regulatory network of the

amino acid producer *Corynebacterium glutamicum*. *J. Biotechnol.* 129, 191–211. doi:10.1016/j.jbiotec.2006.12.013.

Brown, S. P., Le Chat, L., De Paepe, M., and Taddei, F. (2006). Ecology of microbial invasions: amplification allows virus carriers to invade more rapidly when rare. *Curr. Biol.* 16, 2048–52. doi:10.1016/j.cub.2006.08.089.

Brucoli, F., Guzman, J. D., Maitra, A., James, C. H., Fox, K. R., and Bhakta, S. (2015). Synthesis, anti-mycobacterial activity and DNA sequence-selectivity of a library of biaryl-motifs containing polyamides. *Bioorganic Med. Chem.* 23, 3705–3711. doi:10.1016/j.bmc.2015.04.001.

Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602, table of contents. doi:10.1128/MMBR.68.3.560-602.2004.

Buchmeier, N., Bossie, S., Chen, C. Y., Fang, F. C., Guiney, D. G., and Libby, S. J. (1997). SlyA, a transcriptional regulator of *Salmonella typhimurium*, is required for resistance to oxidative stress and is expressed in the intracellular environment of macrophages. *Infect. Immun.* 65, 3725–30.

Bussmann, M., Baumgart, M., and Bott, M. (2010). RosR (Cg1324), a hydrogen peroxide-sensitive MarR-type transcriptional regulator of *Corynebacterium glutamicum*. *J. Biol. Chem.* 285, 29305–29318. doi:10.1074/jbc.M110.156372.

Bustamante, V. H., Santana, F. J., Calva, E., and Puente, J. L. (2001). Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Mol. Microbiol.* 39, 664–678. doi:10.1046/j.1365-2958.2001.02209.x.

Caramel, A., and Schnetz, K. (1998). Lac and λ repressors relieve silencing of the *Escherichia coli* bgl promoter. activation by alteration of a repressing nucleoprotein complex. 284. Available at: https://www.sciencedirect.com/science/article/pii/S002228369892191X?via%3Dihub [Accessed June 26, 2019].

Casjens, S. R., and Hendrix, R. W. (2015). Bacteriophage lambda: Early pioneer and still relevant. *Virology* 479–480, 310–330. doi:10.1016/j.virol.2015.02.010.

Casjens, S., Sampson, L., Randall, S., Eppler, K., Wu, H., Petri, J. B., et al. (1992). Molecular genetic analysis of bacteriophage P22 gene 3 product, a protein involved in the initiation of headful DNA packaging. *J. Mol. Biol.* 227, 1086–1099. doi:10.1016/0022-2836(92)90523-M.

Castang, S., and Dove, S. L. (2012). Basis for the essentiality of H-NS family members in *Pseudomonas aeruginosa*. *J. Bacteriol.* 194, 5101–9. doi:10.1128/JB.00932-12.

Castang, S., McManus, H. R., Turner, K. H., and Dove, S. L. (2008). H-NS family members function coordinately in an opportunistic pathogen. *Proc. Natl. Acad. Sci.* 105, 18947–18952. doi:10.1073/pnas.0808215105.

Chakravorty, A., and Sugden, B. (2015). The AT-hook DNA binding ability of the Epstein Barr virus EBNA1 protein is necessary for the maintenance of viral genomes in latently infected cells. *Virology* 484, 251–8. doi:10.1016/j.virol.2015.05.018.

Chen, J. M., Ren, H., Shaw, J. E., Wang, Y. J., Li, M., Leung, A. S., et al. (2008). Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. *Nucleic Acids Res* 36, 2123–2135. doi:10.1093/nar/gkm1162.

Cordeiro, T. N., Schmidt, H., Madrid, C., Juárez, A., Bernadó, P., Griesinger, C., et al. (2011). Indirect DNA readout by an H-NS related protein: Structure of the DNA complex of the C-terminal domain of Ler. *PLoS Pathog.* 7, e1002380. doi:10.1371/journal.ppat.1002380.

Court, D. L., Oppenheim, A. B., and Adhya, S. L. (2007). A new look at bacteriophage λ genetic networks. *J. Bacteriol.* 189, 298–304. doi:10.1128/JB.01215-06.

Dame, R. T., Luijsterburg, M. S., Krin, E., Bertin, P. N., Wagner, R., and Wuite, G. J. L. (2005). DNA bridging: a property shared among H-NS-like proteins. *J Bacteriol* 187, 1845–1848. doi:10.1128/jb.187.5.1845-1848.2005.

Dame, R. T., Wyman, C., and Goosen, N. (2000). H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.* 28, 3504–10. doi:10.1093/nar/28.18.3504.

Davies, E. V., Winstanley, C., Fothergill, J. L., and James, C. E. (2016). The role of temperate bacteriophages in bacterial infection. *FEMS Microbiol. Lett.* 363, fnw015. doi:10.1093/femsle/fnw015.

Davis, S. E., Mooney, R. A., Kanin, E. I., Grass, J., Landick, R., and Ansari, A. Z. (2011). Mapping *E. coli* RNA Polymerase and Associated Transcription Factors and Identifying Promoters

Genome-Wide. *Methods Enzymol.* 498, 449–471. doi:10.1016/B978-0-12-385120-8.00020-6.

De La Cruz, M. Á., Fernández-Mora, M., Guadarrama, C., Flores-Valdez, M. A., Bustamante, V. H., Vázquez, A., et al. (2007). LeuO antagonizes H-NS and StpA-dependent repression in *Salmonella enterica* ompS1. *Mol. Microbiol.* 66, 727–743. doi:10.1111/j.1365-2958.2007.05958.x.

de Paepe, M., de Monte, S., Robert, L., Lindner, A. B., and Taddei, F. (2010). Emergence of variability in isogenic *Escherichia coli* populations infected by a filamentous virus. *PLoS One* 5, e11823. doi:10.1371/journal.pone.0011823.

De San Eustaquio-Campillo, A., Cornilleau, C., Guerin, C., Carballido-Lopez, R., and Chastanet, A. (2017). PamR, a new MarR-like regulator affecting prophages and metabolic genes expression in *Bacillus subtilis*. *PLoS One* 12, e0189694. doi:10.1371/journal.pone.0189694.

Dedrick, R. M., Jacobs-Sera, D., Guerrero Bustamante, C. A., Garlena, R. A., Mavrich, T. N., Pope, W. H., et al. (2017). Prophage-mediated defence against viral attack and viral counter-defence. *Nat. Microbiol.* 2, 16251. doi:10.1038/nmicrobiol.2016.251.

Deochand, D. K., and Grove, A. (2017). MarR family transcription factors: dynamic variations on a common scaffold. *Crit. Rev. Biochem. Mol. Biol.* 52, 595–613. doi:10.1080/10409238.2017.1344612.

Dersch, P., Schmidt, K., and Bremer, E. (1993). Synthesis of the *Escherichia coli* K-12 nucleoid-associated DNA-binding protein H-NS is subjected to growth-phase control and autoregulation. *Mol. Microbiol.* 8, 875–889. doi:10.1111/j.1365-2958.1993.tb01634.x.

Ding, P., McFarland, K. A., Jin, S., Tong, G., Duan, B., Yang, A., et al. (2015). A Novel AT-Rich DNA Recognition Mechanism for Bacterial Xenogeneic Silencer MvaT. *PLoS Pathog* 11, e1004967. doi:10.1371/journal.ppat.1004967.

Donovan, C., Heyer, A., Pfeifer, E., Polen, T., Wittmann, A., Krämer, R., et al. (2015). A prophage-encoded actin-like protein required for efficient viral DNA replication in bacteria. *Nucleic Acids Res* 43, 5002–5016. doi:10.1093/nar/gkv374.

Duan, B., Ding, P., Hughes, T. R., Navarre, W. W., Liu, J., and Xia, B. (2018). How bacterial xenogeneic silencer rok distinguishes foreign from self DNA in its resident genome. *Nucleic*

Acids Res. 46, 10514–10529. doi:10.1093/nar/gky836.

Ellison, D. W., and Miller, V. L. (2006). H-NS represses inv transcription in *Yersinia enterocolitica* through competition with RovA and interaction with YmoA. *J. Bacteriol.* 188, 5101–5112. doi:10.1128/JB.00862-05.

Engels, S. (2004). Regulation der clp -Genexpression durch ClgR und Definition des ClgR-Regulons aus *Corynebacterium glutamicum*. Available at: https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=2976 [Accessed June 30, 2019].

Engels, S., Schweitzer, J.-E., Ludwig, C., Bott, M., and Schaffer, S. (2004). *clpC* and *clpP1P2* gene expression in *Corynebacterium glutamicum* is controlled by a regulatory network involving the transcriptional regulators ClgR and HspR as well as the ECF sigma factor σH. *Mol. Microbiol.* 52, 285–302. doi:10.1111/j.1365-2958.2003.03979.x.

Erez, Z., Steinberger-levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., et al. (2017). Communication between viruses guides lysis-lysogeny decisions. *Nature* 541, 488–493. doi:10.1038/nature21049.

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A. A. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol* 13, 641–650. doi:10.1038/nrmicro3527.

Fernández-de-Alba, C., Berrow, N. S., Garcia-Castellanos, R., García, J., and Pons, M. (2013). On the Origin of the Selectivity of Plasmidic H-NS towards Horizontally Acquired DNA: Linking H-NS Oligomerization and Cooperative DNA Binding. *J. Mol. Biol.* 425, 2347–2358. doi:10.1016/J.JMB.2013.03.006.

Filutowicz, M., and Roll, J. (1990). The requirement of IHF protein for extrachromosomal replication of the *Escherichia coli oriC* in a mutant deficient in DNA polymerase I activity. *New Biol.* 2, 818–827.

Fonfría-Subirós, E., Acosta-Reyes, F., Saperas, N., Pous, J., Subirana, J. A., and Campos, J. L. (2012). Crystal structure of a complex of DNA with one AT-hook of HMGA1. *PLoS One* 7, e37120. doi:10.1371/journal.pone.0037120.

Frunzke, J., Bramkamp, M., Schweitzer, J. E., and Bott, M. (2008). Population Heterogeneity in *Corynebacterium glutamicum* ATCC 13032 caused by prophage CGP3. *J Bacteriol* 190, 5111–5119. doi:10.1128/jb.00310-08.

Fukui, N., Oshima, T., Ueda, T., Ogasawara, N., and Tobe, T. (2016). Gene activation through the modulation of nucleoid structures by a horizontally transferred regulator, Pch, in enterohemorrhagic *Escherichia coli*. *PLoS One* 11, e0149718. doi:10.1371/journal.pone.0149718.

Gallego Del Sol, F., Penadés, J. R., and Marina, A. (2019). Deciphering the Molecular Mechanism Underpinning Phage Arbitrium Communication Systems. *Mol. Cell* 74, 59–72.e3. doi:10.1016/j.molcel.2019.01.025.

Gama, J. A., Reis, A. M., Domingues, I., Mendes-Soares, H., Matos, A. M., and Dionisio, F. (2013). Temperate bacterial viruses as double-edged swords in bacterial warfare. *PLoS One* 8, e59043. doi:10.1371/journal.pone.0059043.

García, J., Cordeiro, T. N., Prieto, M. J., and Pons, M. (2012). Oligomerization and DNA binding of Ler, a master regulator of pathogenicity of enterohemorrhagic and enteropathogenic *Escherichia coli*. *Nucleic Acids Res.* 40, 10254–10262. doi:10.1093/nar/gks846.

Gehrke, E. J., Zhang, X., Pimentel-Elardo, S. M., Johnson, A. R., Rees, C. A., Jones, S. E., et al. (2019). Silencing cryptic specialized metabolism in *Streptomyces* by the nucleoid-associated protein Lsr2. *Elife* 8. doi:10.1101/614727.

Gerlach, D., Guo, Y., De Castro, C., Kim, S. H., Schlatterer, K., Xu, F. F., et al. (2018). Methicillin-resistant *Staphylococcus aureus* alters cell wall glycosylation to evade immunity. *Nature* 563, 705–709. doi:10.1038/s41586-018-0730-x.

Ghosh, D., Roy, K., Williamson, K. E., Srinivasiah, S., Wommack, K. E., and Radosevich, M. (2009). Acyl-homoserine lactones can induce virus production in lysogenic bacteria: An alternative paradigm for prophage induction. *Appl. Environ. Microbiol.* 75, 7142–7152. doi:10.1128/AEM.00950-09.

Gordon, B. R. G., Li, Y., Cote, A., Weirauch, M. T., Ding, P., Hughes, T. R., et al. (2011). Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc Natl Acad Sci U S A* 108, 10690–10695. doi:10.1073/pnas.1102544108.

Gordon, B. R. G., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., et al. (2010). Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* 107, 5154–5159. doi:10.1073/pnas.0913551107.

Grainger, D. C. (2016). Structure and function of bacterial H-NS protein. *Biochem. Soc. Trans.* 44, 1561–1569. doi:10.1042/BST20160190.

Grimwade, J. E., Ryan, V. T., and Leonard, A. C. (2000). IHF redistributes bound initiator protein, DnaA, on supercoiled *oriC* of *Escherichia coli*. *Mol. Microbiol.* 35, 835–844. doi:10.1046/j.1365-2958.2000.01755.x.

Gulvady, R., Gao, Y., Kenney, L. J., and Yan, J. (2018). A single molecule analysis of H-NS uncouples DNA binding affinity from DNA specificity. *Nucleic Acids Res.* 46, 10216–10224. doi:10.1093/nar/gky826.

Hardy, K. G. (1975). Colicinogeny and related phenomena. *Bacteriol. Rev.* 39, 464–515. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1108869 [Accessed July 24, 2019].

Harrison, E., and Brockhurst, M. A. (2017). Ecological and Evolutionary Benefits of Temperate Phage: What Does or Doesn't Kill You Makes You Stronger. *BioEssays* 39, 1700112. doi:10.1002/bies.201700112.

Helfrich, S., Pfeifer, E., Krämer, C., Sachs, C. C., Wiechert, W., Kohlheyer, D., et al. (2015). Live cell imaging of SOS and prophage dynamics in isogenic bacterial populations. *Mol Microbiol* 98, 636–650. doi:10.1111/mmi.13147.

Hendrix, R. W. (2003). Bacteriophage genomics. *Curr. Opin. Microbiol.* 6, 506–511. doi:10.1016/J.MIB.2003.09.004.

Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E., and Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2192–7. doi:10.1073/pnas.96.5.2192.

Heroven, A. K., Nagel, G., Tran, H. J., Parr, S., and Dersch, P. (2004). RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and *rovA* expression in *Yersinia pseudotuberculosis*. *Mol. Microbiol.* 53, 871–888. doi:10.1111/j.1365-2958.2004.04162.x.

Ho, C. H., Wang, H. C., Ko, T. P., Chang, Y. C., and Wang, A. H.-J. (2014). The T4 phage DNA mimic protein arn inhibits the DNA binding activity of the bacterial histone-like protein H-NS. *J. Biol. Chem.* 289, 27046–27054. doi:10.1074/jbc.M114.590851.

Hong, S. H., Wang, X., and Wood, T. K. (2010). Controlling biofilm formation, prophage excision and cell death by rewiring global regulator H-NS of *Escherichia coli*. *Microb. Biotechnol.* 3, 344–356. doi:10.1111/j.1751-7915.2010.00164.x.

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., and Sullivan, M. B. (2017). Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. Nature Publishing Group doi:10.1038/ismej.2017.16.

Hünnefeld, M., Filipchyk, A., Polen, T., and Frunzke, J. (2019a). Spatiotemporal binding dynamics of the xenogeneic silencer CgpS during prophage induction in *Corynebacterium glutamicum*. *Inside this Thesis*. To be submitted.

Hünnefeld, M., Persicke, M., Kalinowski, J., and Frunzke, J. (2019b). The MarR-Type Regulator MalR Is Involved in Stress-Responsive Cell Envelope Remodeling in *Corynebacterium glutamicum*. *Front. Microbiol.* 10, 1039. doi:10.3389/fmicb.2019.01039.

Ikeda, H., and Tomizawa, J. -i. (1968). Prophage P1, and extrachromosomal replication unit. *Cold Spring Harb. Symp. Quant. Biol.* 33, 791–798. doi:10.1101/SQB.1968.033.01.091.

Ikeda, M., and Nakagawa, S. (2003). The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl. Microbiol. Biotechnol.* 62, 99–109. doi:10.1007/s00253-003-1328-1.

Janion, C. (2008). Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. 4. doi:10.7150/ijbs.4.338.

Johansson, J., Eriksson, S., Sondén, B., Wai, S. N., and Uhlin, B. E. (2001). Heteromeric interactions among nucleoid-associated bacterial proteins: localization of StpA-stabilizing regions in H-NS of *Escherichia coli*. *J. Bacteriol.* 183, 2343–7. doi:10.1128/JB.183.7.2343-2347.2001.

Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F., and Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* 299, 27–51. doi:10.1006/JMBI.2000.3729.

Kahramanoglou, C., Seshasayee, A. S. N., Prieto, A. I., Ibberson, D., Schmidt, S., Zimmermann, J., et al. (2011). Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.* 39, 2073–2091. doi:10.1093/nar/gkq934.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., et al. (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol* 104, 5–25.

Kim, M.-S., and Bae, J.-W. (2018). Lysogeny is prevalent and widely distributed in the murine

gut microbiota. *ISME J.* 12, 1127–1141. doi:10.1038/s41396-018-0061-9.

Kim, M., and Ryu, S. (2013). Antirepression system associated with the life cycle switch in the temperate podoviridae phage SPC32H. 87. doi:10.1128/JVI.02173-13.

Kim, M. S., Bae, S.-H., Yun, S. H., Lee, H. J., Ji, S. C., Lee, J. H., et al. (2005). Cnu, a novel *oriC*-binding protein of *Escherichia coli*. *J. Bacteriol.* 187, 6998–7008. doi:10.1128/JB.187.20.6998-7008.2005.

Kinoshita, S., Udaka, S., and Shimono, M. (2004). Studies on the amino acid fermentation. Part 1. Production of L-glutamic acid by various microorganisms. *J Gen Appl Microbiol* 50, 331–343.

Krause, J. P., Polen, T., Youn, J.-W., Emer, D., Eikmanns, B. J., and Wendisch, V. F. (2012). Regulation of the malic enzyme gene *malE* by the transcriptional regulator MalR in *Corynebacterium glutamicum*. *J Biotechnol* 159, 204–215. doi:10.1016/j.jbiotec.2012.01.003.

Kronheim, S., Daniel-Ivad, M., Duan, Z., Hwang, S., Wong, A. I., Mantel, I., et al. (2018). A chemical defence against phage infection. *Nature* 564, 283–286. doi:10.1038/s41586-018-0767-x.

Labrie, S. J., Samson, J. E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327. doi:10.1038/nrmicro2315.

Lamberte, L. E., Baniulyte, G., Singh, S. S., Stringer, A. M., Bonocora, R. P., Stracy, M., et al. (2017). Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat. Microbiol.* 2, 16249. doi:10.1038/nmicrobiol.2016.249.

Landy, A. (2015). "The λ Integrase Site-specific Recombination Pathway," in *Mobile DNA III* (American Society of Microbiology), 91–118. doi:10.1128/microbiolspec.MDNA3-0051-2014.

Landy, A., and Ross, W. (1977). Viral integration and excision: structure of the lambda *att* sites. *Science (80-.).* 197, 1147–1160. doi:10.1126/science.1094538.

Lang, A. S., and Beatty, J. T. (2007). Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends Microbiol.* 15, 54–62. doi:10.1016/J.TIM.2006.12.001.

Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, C. O., et al. (2007). High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Res.* 35, 6330–7. doi:10.1093/nar/gkm712.

Levine, J. A., Hansen, A. M., Michalski, J. M., Hazen, T. H., Rasko, D. A., and Kaper, J. B. (2014). H-NST induces LEE expression and the formation of attaching and effacing lesions in enterohemorrhagic *Escherichia coli*. *PLoS One* 9, e86618. doi:10.1371/journal.pone.0086618.

Li, C., Wally, H., Miller, S. J., and Lu, C.-D. (2009). The multifaceted proteins MvaT and MvaU, members of the H-NS family, control arginine metabolism, pyocyanin synthesis, and prophage activation in *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* 191, 6211–8. doi:10.1128/JB.00888-09.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (80-.).* 326, 289–293. doi:10.1126/science.1181369.

Lim, C. J., Lee, S. Y., Kenney, L. J., and Yan, J. (2012). Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. *Sci Rep* 2, 509. doi:10.1038/srep00509.

Little, J. W. (1984). Autodigestion of *lexA* and phage lambda repressors. *Proc Natl Acad Sci U S A* 81, 1375–1379.

Liu, Q., and Richardson, C. C. (1993). Gene 5.5 protein of bacteriophage T7 inhibits the nucleoid protein H-NS of *Escherichia coli*. *Proc Natl Acad Sci U S A* 90, 1761–1765. doi:10.1073/pnas.90.5.1761.

Liu, Y., Chen, H., Kenney, L. J., and Yan, J. (2010). A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev.* 24, 339–344. doi:10.1101/gad.1883510.

Łoś, M., and Wegrzyn, G. (2012). "Pseudolysogeny," in *Advances in Virus Research* (Academic Press), 339–349. doi:10.1016/B978-0-12-394621-8.00019-4.

Lüdke, A. (2007). AAA+ Proteasen in *Corynebacterium glutamicum* und ihr Einfluss auf die Stickstoffregulation.

Luo, H., and Gao, F. (2019). DoriC 10.0: an updated database of replication origins in prokaryotic

genomes including chromosomes and plasmids. *Nucleic Acids Res.* 47, D74–D77. doi:10.1093/nar/gky1014.

Lwoff, A. (1953). "Lysogeny.," in *Bacteriological reviews* (American Society for Microbiology (ASM)), 269–337. Available at: http://www.ncbi.nlm.nih.gov/pubmed/13105613 [Accessed June 14, 2019].

Madrid, C., García, J., Pons, M., and Juárez, A. (2007). Molecular Evolution of the H-NS Protein: Interaction with Hha-Like Proteins Is Restricted to *Enterobacteriaceae*. *J. Bacteriol.* 189, 265–268. doi:10.1128/JB.01124-06.

Mardanov, A. V., and Ravin, N. V. (2007). The Antirepressor Needed for Induction of Linear Plasmid-Prophage N15 Belongs to the SOS Regulon. *J. Bacteriol.* 189, 6333. doi:10.1128/JB.00599-07.

McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S., and Kaper, J. B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. U. S. A.* 92, 1664–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7878036 [Accessed October 29, 2018].

McNutt, M. K., Bradford, M., Drazen, J. M., Hanson, B., Howard, B., Jamieson, K. H., et al. (2018). Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proc. Natl. Acad. Sci.* 115, 2557–2560. doi:10.1073/pnas.1715374115.

Melkina, O. E., Goryanin, I. I., and Zavilgelsky, G. B. (2016). The DNA–mimic antirestriction proteins ArdA ColIB-P9, Arn T4, and Ocr T7 as activators of H-NS-dependent gene transcription. *Microbiol. Res.* 192, 283–291. doi:10.1016/j.micres.2016.07.008.

Mellies, J. L., Benison, G., McNitt, W., Mavor, D., Boniface, C., and Larabee, F. J. (2011). Ler of pathogenic *Escherichia coli* forms toroidal protein-DNA complexes. *Microbiology* 157, 1123–1133. doi:10.1099/mic.0.046094-0.

Menouni, R., Hutinet, G., Petit, M.-A., and Ansaldi, M. (2015). Bacterial genome remodeling through bacteriophage recombination. *FEMS Microbiol. Lett.* 362, 1–10. doi:10.1093/femsle/fnu022.

Mentz, A., Neshat, A., Pfeifer-Sancar, K., Pühler, A., Rückert, C., and Kalinowski, J. (2013). Comprehensive discovery and characterization of small RNAs in *Corynebacterium*

glutamicum ATCC 13032. *BMC Genomics* 14, 714. doi:10.1186/1471-2164-14-714.

Michel-Briand, Y., and Baysse, C. (2002). The pyocins of *Pseudomonas aeruginosa*. *Biochimie* 84, 499–510. doi:10.1016/S0300-9084(02)01422-0.

Michel, B. (2005). After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biol* 3, e255. doi:10.1371/journal.pbio.0030255.

Miller, H., Zhou, Z., Wollman, A. J. M., and Leake, M. C. (2015). Superresolution imaging of single DNA molecules using stochastic photoblinking of minor groove and intercalating dyes. *Methods* 88, 81–88. doi:10.1016/J.YMETH.2015.01.010.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., et al. (2011). The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616. doi:10.1101/GR.122705.111.

Moran, N. A. (2004). Microbial Minimalism. *Cell* 108, 583–586. doi:10.1016/s0092-8674(02)00665-7.

Moreau, S., Leret, V., Le Marrec, C., Varangot, H., Ayache, m., Bonnassie, S., et al. (1995). Prophage distribution in coryneform bacteria. *Res. Microbiol.* 146, 493–505. doi:10.1016/0923-2508(96)80295-6.

Mott, M. L., and Berger, J. M. (2007). DNA replication initiation: Mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* 5, 343–354. doi:10.1038/nrmicro1640.

Müller, C. M., Dobrindt, U., Nagy, G., Emödy, L., Uhlin, B. E., and Hacker, J. (2006). Role of histone-like proteins H-NS and StpA in expression of virulence determinants of uropathogenic *Escherichia coli*. *J. Bacteriol.* 188, 5428–38. doi:10.1128/JB.01956-05.

Mustard, J. A., and Little, J. W. (2000). Analysis of *Escherichia coli* RecA interactions with LexA, lambda CI, and UmuD by site-directed mutagenesis of *recA*. *J Bacteriol* 182, 1659–1670.

Nagel, G., Lahrz, A., and Dersch, P. (2001). Environmental control of invasin expression in *Yersinia pseudotuberculosis* is mediated by regulation of RovA, a transcriptional activator of the SlyA/Hor family. *Mol. Microbiol.* 41, 1249–69.

Nanda, A. M., Heyer, A., Krämer, C., Grünberger, A., Kohlheyer, D., and Frunzke, J. (2014). Analysis of SOS-Induced Spontaneous Prophage Induction in *Corynebacterium glutamicum* at the Single-Cell Level. *J. Bacteriol.* 196, 180–188. doi:10.1128/JB.01018-13.

Nanda, A. M., Thormann, K., and Frunzke, J. (2015). Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J Bacteriol* 197, 410–419. doi:10.1128/jb.02230-14.

Navarre, W. W. (2016). The Impact of Gene Silencing on Horizontal Gene Transfer and Bacterial Evolution. *Adv. Microb. Physiol.* 69, 157–186. doi:10.1016/BS.AMPBS.2016.07.004.

Navarre, W. W., Halsey, T. A., Walthers, D., Frye, J., McClelland, M., Potter, J. L., et al. (2005). Co-regulation of *Salmonella enterica* genes required for virulence and resistance to antimicrobial peptides by SlyA and PhoP/PhoQ. *Mol. Microbiol.* 56, 492–508. doi:10.1111/j.1365-2958.2005.04553.x.

Navarre, W. W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S. J., et al. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science (80-.).* 313, 236–238. doi:10.1126/science.1128794.

Niu, C., Wang, D., Liu, X., Liu, H., Liu, X., Feng, E., et al. (2017). An H-NS Family Protein, Sfh, Regulates Acid Resistance by Inhibition of Glutamate Decarboxylase Expression in *Shigella flexneri* 2457T. *Front. Microbiol.* 8, 1923. doi:10.3389/fmicb.2017.01923.

Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., and Adhya, S. (2005). Switches in Bacteriophage Lambda Development. *Annu. Rev. Genet.* 39, 409–429. doi:10.1146/annurev.genet.39.073003.113656.

Oshima, T., Ishikawa, S., Kurokawa, K., Aiba, H., and Ogasawara, N. (2006). *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res* 13, 141–153. doi:10.1093/dnares/dsl009.

Otsuji, N., Sekiguchi, M., Iijima, T., and Takagi, Y. (1959). Induction of phage formation in the lysogenic *Escherichia coli* K-12 by mitomycin C. *Nature* 184, 1079–1080. doi:10.1038/1841079b0.

Patterson-West, J., Arroyo-Mendoza, M., Hsieh, M.-L. L., Harrison, D., Walker, M. M., Knipling, L., et al. (2018). The bacteriophage T4 MotB protein, a DNA-binding protein, improves phage fitness. *Viruses* 10, 343. doi:10.3390/v10070343.

Paytubi, S., Madrid, C., Forns, N., Nieto, J. M., Balsalobre, C., Uhlin, B. E., et al. (2004). YdgT, the Hha paralogue in *Escherichia coli*, forms heteromeric complexes with H-NS and StpA. *Mol. Microbiol.* 54, 251–263. doi:10.1111/j.1365-2958.2004.04268.x.

Perez-Rueda, E., and Ibarra, J. A. (2015). Distribution of putative xenogeneic silencers in prokaryote genomes. *Comput. Biol. Chem.* 58, 167–172. doi:10.1016/j.compbiolchem.2015.06.007.

Persson, F., Bingen, P., Staudt, T., Engelhardt, J., Tegenfeldt, J. O., and Hell, S. W. (2011). Fluorescence nanoscopy of single DNA molecules by using stimulated emission depletion (STED). *Angew. Chem. Int. Ed. Engl.* 50, 5581–3. doi:10.1002/anie.201100371.

Pfeifer-Sancar, K., Mentz, A., Rückert, C., and Kalinowski, J. (2013). Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. *BMC Genomics* 14, 888. doi:10.1186/1471-2164-14-888.

Pfeifer, E. (2013). Untersuchungen zur Induktion und Infektiosität des CGP3 Prophagen in *Corynebacterium glutamicum*. Available at: http://juser.fz-juelich.de/record/141200 [Accessed June 17, 2019].

Pfeifer, E., Hünnefeld, M., Popa, O., and Frunzke, J. (2019). Impact of Xenogeneic Silencing on Phage–Host Interactions. Academic Press doi:10.1016/j.jmb.2019.02.011.

Pfeifer, E., Hünnefeld, M., Popa, O., Polen, T., Kohlheyer, D., Baumgart, M., and Frunzke, J. (2016). Silencing of cryptic prophages in *Corynebacterium glutamicum*. *Nucleic Acids Res.* 44, 10117–10131. doi:10.1093/nar/gkw692.

Pinault, L., Han, J. S., Kang, C. M., Franco, J., and Ronning, D. R. (2013). Zafirlukast inhibits complexation of lsr2 with DNA and growth of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* 57, 2134–2140. doi:10.1128/AAC.02407-12.

Polaczek, P. (1990). Bending of the origin of replication of *E. coli* by binding of IHF at a specific site. *New Biol.* 2, 265–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2279030 [Accessed July 1, 2019].

Polaczek, P., Kwan, K., and Campbell, J. L. (1998). Unwinding of the *Escherichia coli* Origin of Replication (*oriC*) Can Occur in the Absence of Initiation Proteins but Is Stabilized by DnaA and Histone-like Proteins IHF or HU. *Plasmid* 39, 77–83. doi:10.1006/PLAS.1997.1328.

Providenti, M. A., and Wyndham, R. C. (2001). Identification and functional characterization of CbaR, a MarR-like modulator of the *cbaABC*-encoded chlorobenzoate catabolism pathway. *Appl. Environ. Microbiol.* 67, 3530–41. doi:10.1128/AEM.67.8.3530-3541.2001.

Qu, Y., Lim, C. J., Whang, Y. R., Liu, J., and Yan, J. (2013). Mechanism of DNA organization by

Mycobacterium tuberculosis protein Lsr2. *Nucleic Acids Res.* 41, 5263–72. doi:10.1093/nar/gkt249.

Ramisetty, B. C. M., and Sudhakari, P. A. (2019). Bacterial 'Grounded' Prophages: Hotspots for Genetic Renovation and Innovation. *Front. Genet.* 10, 65. doi:10.3389/fgene.2019.00065.

Rangarajan, A. A., and Schnetz, K. (2018). Interference of transcription across H-NS binding sites and repression by H-NS. *Mol. Microbiol.* 108, 226–239. doi:10.1111/mmi.13926.

Refardt, D. (2011). Within-host competition determines reproductive success of temperate bacteriophages. *ISME J.* 5, 1451–1460. doi:10.1038/ismej.2011.30.

Riccardi, E., van Mastbergen, E. C., Navarre, W. W., and Vreede, J. (2019). Predicting the mechanism and rate of H-NS binding to AT-rich DNA. *PLoS Comput. Biol.* 15, e1006845. doi:10.1371/journal.pcbi.1006845.

Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002). Negative Autoregulation Speeds the Response Times of Transcription Networks. *J. Mol. Biol.* 323, 785–793. doi:10.1016/S0022-2836(02)00994-4.

Rozanov, D. V., D'Ari, R., and Sineoky, S. P. (1998). RecA-independent pathways of lambdoid prophage induction in *Escherichia coli*. *J. Bacteriol.* 180, 6306–6315.

Schäfer, A., Tauch, A., Droste, N., Pühler, A., and Kalinowski, J. (1997). The *Corynebacterium glutamicum cglIM* gene encoding a 5-cytosine methyltransferase enzyme confers a specific DNA methylation pattern in an McrBC-deficient *Escherichia coli* strain. *Gene* 203, 95–101. doi:http://dx.doi.org/10.1016/S0378-1119(97)00519-2.

Seid, C. A., Smith, J. L., and Grossman, A. D. (2017). Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in *Bacillus subtilis*. *Mol Microbiol* 103, 798–817. doi:10.1111/mmi.13590.

Sengupta, M., Nielsen, H. J., Youngren, B., and Austin, S. (2010). P1 plasmid segregation: accurate redistribution by dynamic plasmid pairing and separation. *J. Bacteriol.* 192, 1175–83. doi:10.1128/JB.01245-09.

Shearwin, K. E., Brumby, A. M., and Egan, J. B. (1998). The Tum protein of coliphage 186 is an antirepressor. *J. Biol. Chem.* 273, 5708–15. doi:10.1074/jbc.273.10.5708.

Silpe, J. E., and Bassler, B. L. (2019). A Host-Produced Quorum-Sensing Autoinducer Controls a

Phage Lysis-Lysogeny Decision. *Cell* 176, 268–280.e13. doi:10.1016/j.cell.2018.10.059.

Singh, K., Milstein, J. N., and Navarre, W. W. (2016). Xenogeneic Silencing and Its Impact on Bacterial Genomes. *Annu. Rev. Microbiol.* 70, 199–213. doi:10.1146/annurev-micro-102215-095301.

Singh, S. S., Singh, N., Bonocora, R. P., Fitzgerald, D. M., Wade, J. T., and Grainger, D. C. (2014). Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev* 28, 214–219. doi:10.1101/gad.234336.113.

Smith, J. L., and Grossman, A. D. (2015). In Vitro Whole Genome DNA Binding Analysis of the Bacterial Replication Initiator and Transcription Factor DnaA. *PLoS Genet.* 11, e1005258. doi:10.1371/journal.pgen.1005258.

Smits, W. K., and Grossman, A. D. (2010). The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in *Bacillus subtilis*. *PLoS Genet.* 6, e1001207. doi:10.1371/journal.pgen.1001207.

Sonnen, H., Schneider, J., and Kutzner, H. J. (1990a). Characterization of ϕGA1, An Inducible Phage Particle from *Brevibacterium flavum*. *J. Gen. Microbiol.* 136, 567–571. doi:10.1099/00221287-136-3-567.

Sonnen, H., Schneider, J., and Kutzner, H. J. (1990b). Corynephage Cog, a virulent bacteriophage of *Corynebacterium glutamicum*, and its relationship to ΦGA1, an inducible phage particle from *Brevibacterium flavum*. *J. Gen. Virol.* 71, 1629–1633. doi:10.1099/0022-1317-71-8-1629.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186. doi:10.1038/nrmicro1793.

Stern, A., and Sorek, R. (2011). The phage-host arms race: Shaping the evolution of microbes. *BioEssays* 33, 43–51. doi:10.1002/bies.201000071.

Stoebel, D. M., Free, A., and Dorman, C. J. (2008). Anti-silencing: Overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology* 154, 2533–2545. doi:10.1099/mic.0.2008/020693-0.

Tendeng, C., Soutourina, O. A., Danchin, A., and Bertin, P. N. (2003). MvaT proteins in Pseudomonas spp.: a novel class of H-NS-like proteins. *Microbiology* 149, 3047–3050.

doi:10.1099/mic.0.C0125-0.

Thieffry, D., Huerta, A. M., Pérez-Rueda, E., and Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *BioEssays* 20, 433–440. doi:10.1002/(SICI)1521-1878(199805)20:5<433::AID-BIES10>3.0.CO;2-2.

Touchon, M., Bernheim, A., and Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.* 10, 2744–2754. doi:10.1038/ismej.2016.47.

Trötschel, C., Albaum, S. P., Wolff, D., Schröder, S., Goesmann, A., Nattkemper, T. W., et al. (2012). Protein turnover quantification in a multilabeling approach: from data calculation to evaluation. *Mol. Cell. Proteomics* 11, 512–26. doi:10.1074/mcp.M111.014134.

Ueda, T., Takahashi, H., Uyar, E., Ishikawa, S., Ogasawara, N., and Oshima, T. (2013). Functions of the Hha and YdgT proteins in transcriptional silencing by the nucleoid proteins, H-NS and StpA, in *Escherichia coli*. *DNA Res.* 20, 263–71. doi:10.1093/dnares/dst008.

Ueguchi, C., Kakeda, M., and Mizuno, T. (1993). Autoregulatory expression of the *Escherichia coli hns* gene encoding a nucleoid protein: H-NS functions as a repressor of its own transcription. *MGG Mol. Gen. Genet.* 236, 171–178. doi:10.1007/BF00277109.

Ueguchi, C., Suzuki, T., Yoshida, T., Tanaka, K., and Mizuno, T. (1996). Systematic mutational analysis revealing the functional domain organization of *Escherichia coli* nucleoid protein H-NS. *J Mol Biol* 263, 149–162. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8913298.

Utter, B., Deutsch, D. R., Schuch, R., Winer, B. Y., Verratti, K., Bishop-Lilly, K., et al. (2014). Beyond the chromosome: The prevalence of unique extra-chromosomal bacteriophages with integrated virulence genes in pathogenic *Staphylococcus aureus*. *PLoS One* 9, e100502. doi:10.1371/journal.pone.0100502.

Vallet-Gely, I., Donovan, K. E., Fang, R., Joung, J. K., and Dove, S. L. (2005). Repression of phase-variable *cup* gene expression by H-NS-like proteins in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11082–7. doi:10.1073/pnas.0502663102.

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., et al. (2010). Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. *J. Vis.*

Exp. doi:10.3791/1869.

Wagemans, J., Delattre, A. S., Uytterhoeven, B., Smet, J. De, Cenens, W., Aertsen, A., et al. (2015). Antibacterial phage ORFans of *Pseudomonas aeruginosa* phage LUZ24 reveal a novel MvaT inhibiting protein. *Front. Microbiol.* 6, 1242. doi:10.3389/fmicb.2015.01242.

Wahl, A., Battesti, A., and Ansaldi, M. (2018). Prophages in *Salmonella enterica* : a driving force in reshaping the genome and physiology of their bacterial host? *Mol. Microbiol.* 111, mmi.14167. doi:10.1111/mmi.14167.

Wang, Q., Guan, Z., Pei, K., Wang, J., Liu, Z., Yin, P., et al. (2018). Structural basis of the arbitrium peptide–AimR communication system in the phage lysis–lysogeny decision. *Nat. Microbiol.* 3, 1266–1273. doi:10.1038/s41564-018-0239-y.

Wang, W., Li, G.-W., Chen, C., Xie, X. S., and Zhuang, X. (2011). Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* 333, 1445–9. doi:10.1126/science.1204697.

Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., et al. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.* 1, 147. doi:10.1038/ncomms1146.

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28, 127–181. doi:10.1016/j.femsre.2003.08.001.

Wiechert, J., Filipchyk, A., Hünnefeld, M., Gätgens, C., Heermann, R., and Frunzke, J. (2019). Deciphering the rules underlying xenogeneic silencing and counter-silencing of Lsr2-like proteins. *Nucleic Acids Res.* ***submitted***.

Wilkinson, S. P., and Grove, A. (2006). Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr. Issues Mol. Biol.* 8, 51–62.

Will, W. R., Bale, D. H., Reid, P. J., Libby, S. J., and Fang, F. C. (2014). Evolutionary expansion of a regulatory network by counter-silencing. *Nat Commun* 5, 5270. doi:10.1038/ncomms6270.

Will, W. R., Brzovic, P., Le Trong, I., Stenkamp, R. E., Lawrenz, M. B., Karlinsey, J. E., et al. (2019). The Evolution of SlyA/RovA Transcription Factors from Repressors to Countersilencers in *Enterobacteriaceae*. *MBio* 10, e00009-19. doi:10.1128/mbio.00009-19.

Williamson, H. S., and Free, A. (2005). A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist. *Mol Microbiol* 55, 808–827. doi:10.1111/j.1365-2958.2004.04421.x.

Winardhi, R. S., Castang, S., Dove, S. L., and Yan, J. (2014a). Single-molecule study on histone-like nucleoid-structuring protein (H-NS) paralogue in *Pseudomonas aeruginosa*: MvaU Bears DNA organization mode similarities to MvaT. *PLoS One* 9, e112246. doi:10.1371/journal.pone.0112246.

Winardhi, R. S., Fu, W., Castang, S., Li, Y., Dove, S. L., and Yan, J. (2012). Higher order oligomerization is required for H-NS family member MvaT to form gene-silencing nucleoprotein filament. *Nucleic Acids Res.* 40, 8942–8952. doi:10.1093/nar/gks669.

Winardhi, R. S., Gulvady, R., Mellies, J. L., and Yan, J. (2014b). Locus of enterocyte effacement-encoded regulator (Ler) of pathogenic *Escherichia coli* competes off histone-like nucleoid-structuring protein (H-NS) through noncooperative DNA binding. *J. Biol. Chem.* 289, 13739–13750. doi:10.1074/jbc.M113.545954.

Wolanski, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwinska, J. (2015). *oriC*-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.* 6, 735. doi:10.3389/fmicb.2014.00735.

Yu, R. R., and DiRita, V. J. (2002). Regulation of gene expression in *Vibrio cholerae* by ToxT involves both antirepression and RNA polymerase stimulation. *Mol. Microbiol.* 43, 119–134. doi:10.1046/j.1365-2958.2002.02721.x.

Yun, S. H., Ji, S. C., Jeon, H. J., Wang, X., Kim, S. W., Bak, G., et al. (2012a). The CnuK9E H-NS Complex Antagonizes DNA Binding of DicA and Leads to Temperature-Dependent Filamentous Growth in *E. coli*. *PLoS One* 7, e45236. doi:10.1371/journal.pone.0045236.

Yun, S. H., Ji, S. C., Jeon, H. J., Wang, X., Lee, Y., Choi, B.-S., et al. (2012b). A mutational study of Cnu reveals attractive forces between Cnu and H-NS. *Mol. Cells* 33, 211–216. doi:10.1007/s10059-012-0006-5.

Zeng, Z., Liu, X., Yao, J., Guo, Y., Li, B., Li, Y., et al. (2016). Cold adaptation regulated by cryptic prophage excision in *Shewanella oneidensis*. *Isme j* 10, 2787. doi:10.1038/ismej.2016.85.

3 Publications and Manuscripts

The “Contributor Roles Taxonomy (CRediT)”, published 2018 by McNutt et al., represents a new standard for declaring authorships of scientific publications and to clarify contributions (McNutt et al., 2018). The following extracted table from McNutt’s publication was used to describe the roles of the authors of the manuscripts in this chapter:

#	Role	Definition
1	Conceptualization	Ideas; formulation or evolution of overarching research goals and aims.
2	Data curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.
3	Formal analysis	Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data.
4	Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.
5	Methodology	Development or design of methodology; creation of models.
6	Project administration	Management and coordination responsibility for the research activity planning and execution.
7	Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.
8	Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.
9	Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.
10	Writing – original draft	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).
11	Writing – review & editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.

3.1 Silencing of cryptic prophages in *Corynebacterium glutamicum*

Pfeifer E., **Hünnefeld M.**, Popa O., Polen T., Kohlheyer D., Baumgart M., and Frunzke J.

Published in *Nucleic Acids Research,* 2016

Contributor Role	Contributor
Conceptualization	EP (70 %), JF (20%), TP (5%), DK (5%)
Formal Analysis	EP (65 %), MH (15 %), OP (10 %), JF (5 %), TP (2.5 %), MB (2.5 %)
Investigation/Experiments	EP (70 %), MH (30 %)
Methodology	EP (70 %), JF (25 %), MB (5 %)
Project Administration	EP (50 %), JF (50 %)
Software	OP (60 %), EP (40 %)
Supervision	EP (50 %), JF (50 %)
Visualization	EP (70 %), OP (20 %), MH (10 %)
Writing – Original Draft Preparation	EP (65%), JF (25%), MB (2.5%), TP (2.5%), DK (2.5%), OP (2.5%)
Writing – Review & Editing	JF (50 %), EP (20 %), MB (10 %), OP (7.5 %), MH (7.5 %), TP (2.5 %), DK (2.5 %)

Overall contribution MH: 10 %

The main contribution of MH to this manuscript was on the experimental and editing site. Many of the presented experiments were conducted by MH under the supervision of EP during the Master thesis time of MH. To be more precise, MH conducted the experiments on which the following figures are based; the work of MH was supervised by EP: Figure 2 B and C, Figure 3 A, Figure 4, Figure 5 A, B, D, and Figure 6 A and B. Additionally, MH's experimental work contributed to the Supplementary Tables S1, S2 and S3. Figure S7 and S8 were completely created by MH. Figure S4 and S5 as well as Table S4 are, furthermore, based on Figure 3. Figure 5, moreover, was the basis for Table S5. For the described figures, MH was also involved in the visualization.

Nucleic Acids Research Advance Access published August 4, 2016

Nucleic Acids Research, 2016 **1**
doi: 10.1093/nar/gkw692

Silencing of cryptic prophages in *Corynebacterium glutamicum*

Eugen Pfeifer[1], Max Hünnefeld[1], Ovidiu Popa[2], Tino Polen[1], Dietrich Kohlheyer[1], Meike Baumgart[1] and Julia Frunzke[1,*]

[1]Institute of Bio- und Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, 52425 Jülich, Germany and [2]Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, 40225, Düsseldorf, Germany

Received May 11, 2016; Revised July 25, 2016; Accepted July 26, 2016

ABSTRACT

DNA of viral origin represents a ubiquitous element of bacterial genomes. Its integration into host regulatory circuits is a pivotal driver of microbial evolution but requires the stringent regulation of phage gene activity. In this study, we describe the nucleoid-associated protein CgpS, which represents an essential protein functioning as a xenogeneic silencer in the Gram-positive *Corynebacterium glutamicum*. CgpS is encoded by the cryptic prophage CGP3 of the *C. glutamicum* strain ATCC 13032 and was first identified by DNA affinity chromatography using an early phage promoter of CGP3. Genome-wide profiling of CgpS binding using chromatin affinity purification and sequencing (ChAP-Seq) revealed its association with AT-rich DNA elements, including the entire CGP3 prophage region (187 kbp), as well as several other elements acquired by horizontal gene transfer. Countersilencing of CgpS resulted in a significantly increased induction frequency of the CGP3 prophage. In contrast, a strain lacking the CGP3 prophage was not affected and displayed stable growth. In a bioinformatics approach, *cgpS* orthologs were identified primarily in actinobacterial genomes as well as several phage and prophage genomes. Sequence analysis of 618 orthologous proteins revealed a strong conservation of the secondary structure, supporting an ancient function of these xenogeneic silencers in phage-host interaction.

INTRODUCTION

Viral DNA, in the form of functional prophages or degenerated (cryptic) phage elements, is ubiquitously found in bacterial genomes and may constitute up to 20% of the host genome (1–3). The mosaic-like structure of bacterial genomes indicates that phage-mediated horizontal gene transfer is a pivotal driver of bacterial evolution (4). Recent studies demonstrated that these elements might contribute significantly to the fitness of their respective host by improving stress tolerance, antibiotic resistance, biofilm formation or virulence (5,6). Phage-mediated gene transfer may provide the cell with novel adaptive traits, improving the fitness of the receptor cell, but this does not occur without risks. The integration of selfish replicators, including transposable elements, integrative/conjugative elements (ICE) or phages, can lead to high transcriptional and translational costs or even cell death (7,8). Hence, bacteria possess a number of different systems that confer resistance to foreign genetic elements, e.g. CRISPR/Cas and restriction modification (R-M) systems (9,10).

However, to harness the adaptive potential of foreign DNA and enable its integration into the host regulatory circuitry, bacteria have evolved a rather mediative mechanism called xenogeneic silencing (XS) (11–13). This mechanism relies on the function of small nucleoid-associated proteins (NAPs) to target and inhibit the expression of foreign DNA, which is recognizable by its typically higher AT content in comparison to the host genome (1,14). The major role of XS proteins is the binding of foreign DNA elements and the inhibition of transcription by a complex formation of AT-rich DNA stretches causing either the occlusion or trapping of the RNA polymerase (15,16). Currently known XS proteins belong to one of four classes, consisting of H-NS-type proteins found in several proteobacteria (12,17), Lsr2-like proteins of the actinomycetes (18), MvaT of *Pseudomonas* species (16) and Rok of *Bacillus subtilis* (19).

To date, most studies have focused on host-encoded XS proteins acting as silencers of foreign DNA. However, it may also be of benefit for the foreign element to bring its own silencer protein to improve tolerance within the host cell. Here, we describe a novel prophage-encoded XS protein of the Lsr2-type in *Corynebacterium glutamicum* ATCC 13032. The genome of this important industrial amino acid producer contains three cryptic prophages (20,21). Whereas CGP1 and CGP2 are highly degenerated, CGP3 comprises almost 6% of the entire genome (187 kb) and is inducible in an SOS-dependent manner (22,23). Even under

*To whom correspondence should be addressed. Tel: +49 2461 615430; Email: j.frunzke@fz-juelich.de

non-inducing conditions, spontaneous prophage induction (SPI) was observed, preceded by a spontaneous activation of the SOS response in >60% of cases (20,22,23). However, the precise regulatory control of CGP3 induction has not been studied thus far.

In this study, we demonstrate the essential role of a prophage-encoded NAP, which is a homolog to the mycobacterial Lsr2 protein and functions as a silencer of cryptic phage elements in *C. glutamicum* (CgpS, *C. glutamicum* prophage silencer). Genome-wide profiling of the CgpS–DNA interaction revealed its association with AT-rich DNA regions located primarily within prophage regions. Countersilencing of CgpS activity *via* the expression of its truncated oligomerization domain resulted in the induction of CGP3, causing cell death. A bioinformatics analysis revealed homologous proteins mainly in actinomycetes, but, interestingly, also in several phage and prophage genomes. These data demonstrate the importance of XS proteins for the tolerance of viral DNA and indicate that this mechanism is exploited by both the host and the virus.

MATERIALS AND METHODS

Bacterial strains and growth conditions

The bacterial strains and plasmids used in this study are listed in Supplementary Table S1. *Corynebacterium glutamicum* ATCC 13032 was used as wild-type strain (24). *E. coli* DH5α was used as host for cloning procedures and cultivated in Lysogeny Broth (LB) medium or on agar plates at 37°C (25). For growth studies and fluorescence assays (e.g. preparation of cells for fluorescence microscopy), *C.glutamicum* cells were pre-cultivated in BHI (brain heart infusion, DifcoTM BHI, BD, Heidelberg, Germany) medium at 30°C for 6 h. This first preculture was used to inoculate an overnight culture in CGXII minimal medium (26) containing 2% (w/v) glucose and 30 $mg \cdot l^{-1}$ protocatechuat acid. The CGXII culture was finally used to inoculate the main culture in the same medium (CGXII with 2% (w/v) glucose) to a start OD_{600} of 1, unless specified otherwise. If necessary, 50 $\mu g \cdot ml^{-1}$ (*E. coli*) or 25 $\mu g \cdot ml^{-1}$ (*C. glutamicum*) kanamycin and/or 34 $\mu g \cdot ml^{-1}$ (*E. coli*) or 10 $\mu g \cdot ml^{-1}$ (*C. glutamicum*) chloramphenicol were added.

Recombinant DNA work

Plasmids and oligonucleotides used in this study are listed in Supplementary Table S2, respectively. Standard methods including PCR, DNA restriction and ligation, were performed according to established protocols (25). In some cases, Gibson assembly (27) was used for the constructions of plasmids. DNA sequencing and oligonucleotides synthesis were conducted by Eurofins MWG Operon (Ebersberg, Germany). The chromosomal integration of the Strep tagged *cgpS* gene variant was performed using the two-step homologous recombination method (28). The 500 bp up and downstream regions of *cgpS* were amplified using the oligonucleotides LF_cgpS_pK19_fw and LF_cgpS_rv and, accordingly, RF_cgpS_fw and RF_cgpS_pK19_rv. Amplification of the Strep-tagged *cgpS* gene was done by using the plasmid pAN6-*cgpS-strep* as template for the oligonucleotide pair cgpS_strep_fw and cgpS_strep_rv. The three resulting PCR products and the digested pK19*mobsacB* plasmid (with *Bam*HI, *Eco*RI) were assembled using Gibson assembly (27). Correct integration into the *cgpS* locus was confirmed by sequencing of the colony PCR product with the oligonucleotides Cgps_indel-fw and CgpS_indel_rv.

Cultivation in the BioLector System

Growth experiments were performed predominantly in the BioLector® microcultivation system of m2p-labs (Aachen, Germany) as described by (29). Cultivation was performed in 48-well FlowerPlates (m2p labs, Germany) at 30°C and a shaking frequency of 1200 rpm. The cells were cultivated in 750 µl of CGXII minimal media with 2% (w/v) glucose containing different additives (e.g. Isopropyl β-D-1-thiogalactopyranoside (IPTG), MMC, kanamycin), as indicated. Measurements were taken at 15-min intervals.

DNA affinity chromatography with the promoter region of *alpAC*

The promoter region of *alpAC* was amplified by PCR with the oligonucleotides PalpAC-Biotin-Tag-fw and PalpAC rv (product size 516 bp). To flag the amplified product further PCRs were performed but with the Biotin-Primer (MWG Eurofins, Ebersberg, Germany) and the PalpAC rv. At least 220 pmol of the biotinylated products were purified by size exclusion chromatography with the usage of an 8 ml sepharose s400-HR column from GE Healthcare (Freiburg, Germany). A total of 5 mg of the M-280 Streptavidin Dynabeads® (Invitrogen, Carlsbad, CA, USA) were washed twice with the binding and wash (BW) buffer (10 mM Tris-HCl pH 7.5, 2 M NaCl), subsequently suspended in BW buffer containing biotinylated products and incubated for 1 h at room temperature. To eliminate unbound DNA fragments the beads were washed three times with the BW buffer and finally suspended in the binding and storage (BS) buffer (20 mM Tris-HCl pH 7.5, 1 mM EDTA, 10% (v/v) glycerin, 0.01% (v/v) Triton-X-100, 100 mM NaCl, 1 mM DTT). A total fo 500 ml of cells were grown in CGXII minimal media with glucose as carbon source (as described in bacterial strains and growth conditions) to an OD_{600} of ∼5. After the cells were harvested by centrifugation (20 min, 5300*g* and washed once with phosphate buffered saline (PBS) buffer (137 mM NaCl, 2.7 mM KCl, 20 mM Na_2HPO_4, 1.8 mM KH_2PO_4), cell pellets were suspended in BS buffer supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF). Cell disruption was performed by five passages at 172 MPa through a French pressure cell (Heinemann, Schwaebisch Gmuend, Germany). The DNA binding reactions were set up with complete prepared crude extracts, the DNA-coupled beads and 500 µg of chromosomal DNA for 45 min at room temperature. After the binding reaction, beads were washed once with BS buffer, twice with BS buffer and 400 µg chromosomal DNA and, as a final washing step, again with BS buffer. The elution was fulfilled in two subsequent steps with BS buffer containing 2 M sodium chloride. After TCA precipitations (30) of the pooled elution fractions the samples were analyzed via

sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (31). Identification of proteins was conducted by MALDI-ToF analysis as described in the section below.

Preparation of ChAP-Seq samples

Cells of the wild-type strain ATCC 13032 and the variant containing the Strep-tagged CgpS protein (WT::*cgpS-strep*) were first grown in BHI for 6 h and then 1 ml was used to inoculate minimal media cultures (CGXII with 2% (w/v) glucose). After cultivation overnight, these precultures were used to inoculate 500 ml of the same minimal medium, were grown to an OD_{600} 5 to 6, and finally harvested by centrifugation (10 min, 11 325*g* at 4°C). After washing the cells with CGXII medium without (w/o) MOPS, the cells were resuspended in 10 ml MOPS-free CGXII containing 1% (v/v) formaldehyde. The fixation was conducted by incubation at room temperature for 20 min. Subsequently, glycine was added to a final concentration of 125 mM and the cells were incubated for further 5 min at room temperature. Then, the cells were washed twice with buffer A (100 mM Tris-HCl, pH 8.0, 1 mM EDTA) and resuspended in 10 ml buffer A supplemented with *cOmplete* Protease Inhibitor (Roche, Basel, Switzerland) and 5 mg RNase A. Cell disruption was performed as described in the DNA affinity chromatography section (five passages through a French Press cell). The chromosomal DNA of the lysates were sheared by sonication 3 × 30 s with a Branson sonifier 250 (Heinemann, Schwaebisch Gmuend, Germany) using a pulse length of 40% and an intensity of one to give an average fragment size of 200–1500 bp as confirmed by agarose gel electrophoresis. Cell debris was first removed by centrifugation at 5300*g* for 20 min and then centrifuged for 1 h at 150 000*g* both steps at 4°C. The supernatant was used for protein–DNA purification according to the standard Strep-tag® purification protocol (see below, protein purification). The pooled elution fractions were incubated overnight at 65°C, followed by a treatment with proteinase K (final concentration 400 $mg \cdot ml^{-1}$) for 3 h at 55°C. Finally, the DNA of the samples was purified by phenol–chloroform extraction (32), precipitated with ethanol, washed with 70% (v/v) ethanol, dried and resuspended in 50–100 µl ddH_2O.

ChAP-Seq

The obtained DNA fragments of each sample (2 µg) were used for library preparation and indexing using the TruSeq DNA PCR-free sample preparation kit according to the manufacturer's instruction, yet omitting the DNA size selection steps (Illumina, Chesterford, UK). The resulting libraries were quantified using the KAPA library quant kit (Peqlab, Bonn, Germany) and normalized for pooling. Sequencing of pooled libraries was performed on a MiSeq (Illumina, San Diego, US) using paired-end sequencing with a read-length of 2 × 150 bases. Data analysis and base calling were accomplished with the Illumina instrument software and stored as fastq output files. The obtained sequencing data of each sample were imported into CLC Genomics Workbench (Version 7.5.1, Qiagen Aarhus A/S) for trimming and base quality filtering. The output was mapped to accession BX927147 as *C. glutamicum* reference genome (21). For peak detection the resulting mapping coverage of each sample was exported and imported into the in-house software Genome Data Viewer (unpublished). A peak was automatically annotated if the coverage of a region is above the 3-fold average of the averaged genome coverage. All peaks were inspected and confirmed manually.

qPCR

The relative amount of circular phage DNA was determined *via* quantitative PCR (qPCR). Therefore, *C. glutamicum* wild type cells containing empty pAN6 plasmid (control), pAN6-*cgpS* gene or pAN6-*N-cgpS* were grown in 48-well FlowerPlates containing CGXII minimal medium at 30°C and 900 rpm in a microtron (Infors-HT, Bottmingen, Switzerland). The overexpression of *cgpS* and the N-terminal part were induced with 150 µM IPTG (for control samples no IPTG was added). After 24 h, 750 µl of the cells were harvested and the DNA was extracted using the NucleoSpin microbial DNA Kit (Macherey Nagel, Dueren, Germany) and DNA concentration was quantified using a nanophotometer (Implen, München, Germany). Each sample contained 1 µg total DNA as a template. For the reaction an innuMIX qPCR MasterMix SyGreen (Analytic Jena, Jena, Germany) and a qTOWER 2.2 (Analytic Jena) was used. The reaction protocol was divided into two parts (i) polymerase chain reaction (PCR) ((a) 3 min preincubation at 95°C, (b) 5 s denaturation at 95°C, (c) 25 s elongation at 62°C, 40x repetition of step (b) to (c)) and a (ii) melting curve analysis ($\Delta T = 1°C/6$ s). The PCR product size using oligonucleotides belonging to the circular phage product is 150 bp (listed in Supplementary Table S2). As reference gene *ddh* was used with the oligonucleotides listed in Supplementary Table S2 resulting in a 150 bp product. For data analysis the qPCR software qPCR 3.1 (Analytik Jena) and the Livak method were used (33) to determine the $2^{-\Delta\Delta Ct}$ based on the measured C_T-values.

DNA microarrays

For a comparative transcriptome analysis of *C. glutamicum* ATCC 13032/pAN6 with cells carrying the pAN6-*N-cgpS*- (used for countersilencing) were cultivated in CGXII with 2% (w/v) glucose and 100 µM IPTG as described in bacterial strains and growth conditions. The preparation of labeled cDNA and DNA microarray analysis was performed as described previously (34). Array data were deposited in the GEO database (ncbi.nlm.nih.gov/geo) under accession number GSE80674.

Cultivation and perfusion in microfluidic device

For single-cell analysis an in-house developed microfluidic platform was used (22,35–37). Phase-contrast and fluorescence time-lapse imaging was performed at 6 min intervals. Medium was supplied continuously to ensure stable and constant environmental conditions. CGXII minimal medium with 2% (w/v) glucose and 25 $\mu g \cdot ml^{-1}$ kanamycin was infused at a rate of 300 $nl \cdot min^{-1}$ using a high-precision syringe pump (neMESYS, Cetoni GmbH, Korbussen, Germany). For the expression of the N-terminal part of CgpS

150 μM IPTG were added to the medium. A constant cultivation temperature of 30°C was ensured (PeCon GmbH, Erbach, Germany). The cells were cultivated for 16 h.

Fluorescence microscopy

The cultivations were done as described in bacterial strains and growth conditions. After 6 h of cultivation, 1–3 μl were pipetted on a microscope slide coated with a thin 1% (w/v) agarose layer that was based on tris-acetate buffer. To stain the DNA with the Hoechst Dye, 33 342 1 ml cells were harvested (5300*g*, 5 min), subsequently resuspended in PBS buffer containing 100 ng·ml^{-1} Hoechst 33342 and incubated at room temperature for 20 min. Images were taken on an AxioImager M2 (Zeiss, Oberkochen, Germany) equipped with a Zeiss AxioCam MRm camera. Fluorescence was monitored with the filter set 46 HE YFP for eYFP, 63 HE filter was used for mCherry fluorescence and Hoechst fluorescence was examined with the filter set 49. An EC Plan-Neofluar 100x/1.3 Oil Ph3 objective was used. Images were acquired and analyzed with the AxioVision 4.8 software (Carl Zeiss).

Protein purification

CgpS tagged C-terminal with a Strep-tag® was heterologously produced in *E. coli* BL21 (DE3). Cells were grown to an OD_{600} of 0.4 at 37°C. Upon induction with 50 μM IPTG the cultivation was continued at 16°C overnight. Cells were harvested by centrifugation at 5300*g* and 4°C for 10 min and resuspended in buffer B (250 mM NaCl, 50 mM Tris-HCl, pH 7.5). Cell disruption was performed by two passages through a French pressure cell at 172 MPa. Cell debris was removed by centrifugation at 20 min, 5300*g* and 4°C, followed by an ultracentrifugation (60 min, 229 000*g*, 4°C). The supernatant was applied to an equilibrated 1 ml Strep-Tactin®-Sepharose® (IBA, Göttingen, Germany) column. It was subsequently washed with 10 ml buffer B and the protein was eluted with 10 ml buffer B containing 1 mM d-desthiobiotin (Sigma Aldrich).

Electrophoretic mobility shift assays (EMSA)

EMSA studies of CgpS and selected DNA regions identified by ChAP-Seq were performed with selected regions (500 bp fragments, for oligo sequences see Supplementary Table S3). The corresponding regions were amplified by PCR and purified by using the PCR clean-up Kit of Macherey Nagel (Dueren, Germany). The promoter region of *gntK* was used as control fragment (560 bp). A total of 90 ng DNA per lane were incubated with different concentrations (1 μM and 2 μM) of purified CgpS protein for 20 min in EMSA buffer (250 mM Tris-HCl pH 7.5, 25 mM $MgCl_2$, 200 mM KCl, 25% (v/v) glycerol). Subsequently, samples were loaded onto a native 10% polyacrylamide gel (TBE-based, TBE (89 mM Tris base, 89 mM boric acid, 2 mM Na_2EDTA, loading dye: 0.01% (w/v) xylene cyanol dye, 0.01% (w/v) bromophenol blue dye, 20% (v/v) glycerol, 1x TBE). The DNA was stained with SYBR Green I (Sigma Aldrich, St. Louis, MO, USA).

Protein pull down and MALDI-TOF analysis

C. glutamicum cells containing the plasmids pAN6, pAN6-*cgpS*-strep or pAN6-N-*cgpS*-strep were cultivated as described in bacterial strains and growth conditions. The cultures were grown in 500 ml CGXII with 2% (w/v) glucose to an OD_{600} of 5 and subsequently induced with 150 μM IPTG for further 4 h. The cells were harvested (5300*g*, 20 min, 4°C), washed in buffer B (see protein purification) and disrupted as descripted in the DNA affinity chromatography section. Purification was performed as described in the section above. The eluted fractions were analyzed by SDS-PAGE (31) using a 4–20% Mini-PROTEAN® gradient gel (Bio Rad, Munich, Germany). The gels were stained with a Coomassie dye based RAPIDstain solution (G-Biosciences, St. Louis, MO, USA). MALDI-TOF-MS measurements were performed with an Ultraflex III TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany) for the identification of the proteins as described (38).

Homology search

BLAST 'nr' database (ver. February 2015) was downloaded from NCBI (http://www.ncbi.nlm.nih.gov/). CgpS amino acid sequence was extracted from the GenBank file *Corynebacterium glutamicum* ATCC 13032, accession: NC_006958.1 and locus_tag: cg1966. A PSI-BLAST ((39)) search with CgpS sequence as the query was executed against the ncbi nr database. The e-value threshold was set to 0.005, the number of iteration was not limited and the search iteration was performed until it converged. A total of 5230 (1920 unique) homologous hits were achieved from which 618 could be allocated to a particular bacterial species or a phage. Sequence global identity was calculated by pairwise comparison between the CgpS sequence with all 618 PSI-BLAST hits using the Needleman–Wunsch algorithm (40) implemented in the EMBOSS package (41) needle.

Secondary structure prediction

The amino acid sequence of the CgpS protein and the sequences of the 618 homologous hits were used to predict the secondary structure by psipred (42). The visualization of the psipred output was done in R (43).

Statistics and visualization

All statistical analysis and data visualization from the bioinformatic section was performed in R (43).

RESULTS

A small nucleoid-associated protein encoded by a cryptic prophage element

To decipher the control of prophage induction and activation of cryptic elements in *C. glutamicum* ATCC 13032, we performed DNA affinity chromatography with the promoter of the early phage operon *alpAC* using the crude extract of log-phase cells grown in glucose minimal medium ((34), Figure 1A). SDS-Page analysis of the proteins bound

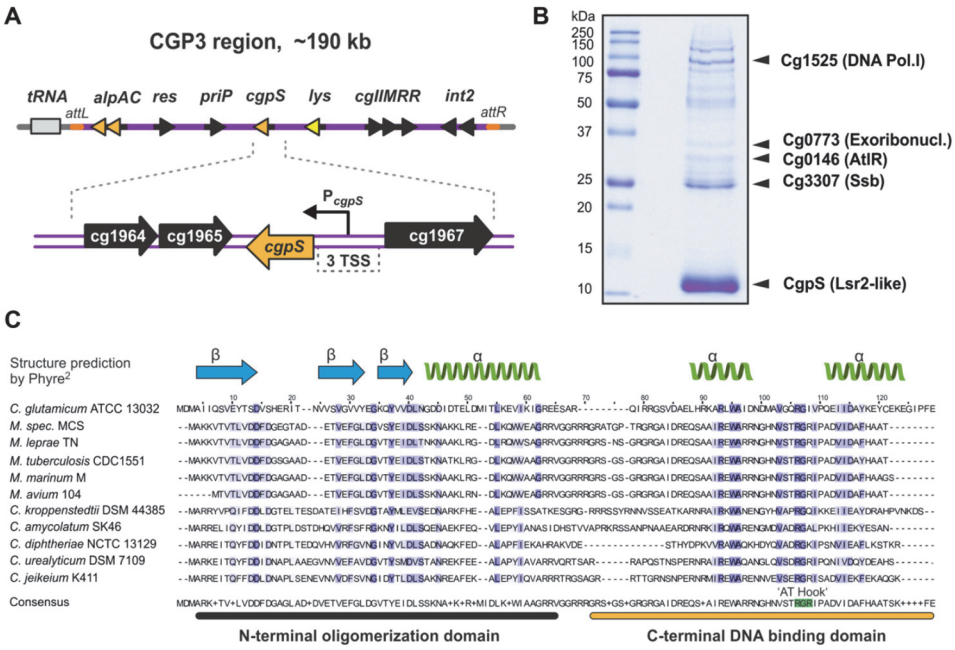


Figure 1. CgpS is a prophage-encoded nucleoid-associated protein in *C. glutamicum* ATCC 13032. (**A**) Genomic organization of the CGP3 prophage region containing the *cgpS* gene in *C. glutamicum* strain ATCC 13032. (**B**) DNA affinity chromatography was used to identify putative regulatory proteins bound to the early phage promoter of *alpAC* (34). Protein samples were separated by SDS-PAGE, and enriched proteins were identified using MALD-ToF analysis. The nucleoid-associated protein CgpS (Cg1966) was detected as a prominent band with a molecular mass of 13.4 kDa. (**C**) Amino acid sequences of CgpS, five related proteins of *Corynebacteria* and five Lsr2 homologs of *Mycobacteria* were used for a multiple sequence alignment conducted using the Clustal Omega platform (73). The predicted domain organization is depicted above the sequence alignment. The Blosum62 identity score is highlighted in blue and matches between 21–26% across the species. The secondary structure of CgpS was predicted with 99.7% confidence of 54 residues (46% of CgpS sequence) by Phyre^2 (74). The motif 'RGI,' which is similar to the AT hook motif 'RGR' of Lsr2 and H-NS (44,59), was identified between two predicted alpha helices.

to the *alpAC* promoter revealed a prominent band corresponding to the 13.4 kDa protein Cg1966 encoded within the CGP3 prophage region (Figure 1B). In particular, the C-terminal domain of Cg1966 shares significant sequence similarity with the nucleoid-associated protein Lsr2 of *Mycobacterium tuberculosis* (Supplementary Figure S1). This domain corresponds to the DNA binding domain of Lsr2 (IPR024412), which was previously found to bind AT-rich DNA *via* an AT-hook motif and functions as a silencer of xenogeneic DNA (44,45). Based on the data described in the following sections, we renamed Cg1966 as CgpS (*Corynebacterium glutamicum* prophage silencer). Secondary structure predictions of CgpS as well as of CgpS homologs suggest a significant structural similarity with Lsr2 and reveal the presence of an AT-hook-like motif 'RGI' between the two predicted C-terminal alpha helices (Figure 1C) (18,45).

CgpS functions as a silencer of CGP3 activity

To study the impact of *cgpS* expression on the activity of the CGP3 prophage, we overexpressed *cgpS* in a strain carrying a reporter construct (WT-P$_{lys}$-*eyfp*) indicative for the activation of CGP3 by the production of the yellow fluorescent protein eYFP under the control of a phage promoter (22). Upon induction with mitomycin C, the control strain carrying the empty plasmid displayed increased reporter activity. Consistent with our assumption, overexpression of *cgpS* reduced the reporter output to nearly the background level (Figure 2A).

To study the intracellular localization of CgpS in *C. glutamicum* cells, we C-terminally fused this protein to mCherry and analyzed its distribution *via* fluorescence microscopy. As shown by Hoechst staining, this NAP appeared associated with the nucleoid but formed distinct foci in the cell (Figure 2B and C). Remarkably, CgpS-mCherry foci co-localized with foci of an AlpA-eYFP fusion that was previously described as a CGP3 DNA adaptor protein binding to the *alpAC* promoter region (34) (Figure 2C). The

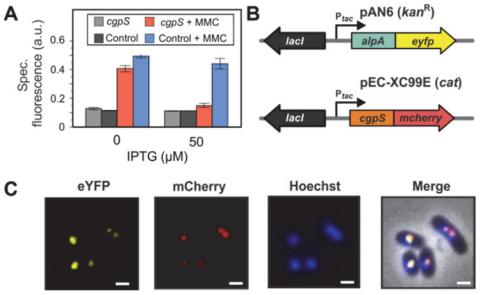


Figure 2. CgpS functions as a silencer of CGP3 prophage activity. (**A**) Silencing of CGP3 induction. The prophage reporter strain *C. glutamicum* ATCC 13032::P_{lys}-*eyfp* containing the *cgpS* overexpression plasmid pAN6-*cgpS*-*Strep* was cultivated in CGXII minimal medium in the presence or absence of IPTG (50 μM) and MMC (600 μM). The prophage reporter strain carrying the empty plasmid pAN6 served as a control. EYFP fluorescence was measured after 20 h of microplate cultivation. (**B** and **C**) CgpS is located in the nucleoid and displays colocalization with the phage adaptor protein AlpA (34). For co-localization studies, a C-terminal CgpS-mCherry fusion and a C-terminal fusion of the prophage adaptor protein AlpA to eYFP (pEC-XC99E) were analyzed. Both gene fusions were under control of the inducible *tac* promoter. DNA was stained with Hoechst Dye 33342. Fluorescence images were taken after 2.5 h of cultivation in CGXII with 150 μM IPTG. Scale bar, 2 μm.

functionality of this CgpS-mCherry fusion was confirmed by the counteraction of CGP3 activation upon addition of MMC (Supplementary Figure S2).

Genome-wide binding profile of CgpS

The data of the co-localization experiments suggest binding of CgpS to the CGP3 prophage region. In the following, the genome-wide binding profile was analyzed by combining affinity chromatography purification of crosslinked CgpS–DNA complexes followed by sequencing of associated DNA (ChAP-Seq). For this purpose, we replaced the native *cgpS* gene in the genome of ATCC 13032 with *cgpS-Strep* encoding a C-terminal Strep-tagged CgpS variant. This analysis revealed that CgpS associates with 1.5% of the ATCC 13032 genome and with ∼20.5% of the cryptic CGP3 prophage region (Supplementary Figure S3). In total, 90 peaks were detected, 58 of which were within and 32 were located outside the CGP3 prophage (Figure 3A, Supplementary Table S4). The majority of the peak maxima were located within promoter regions (60%), but CgpS binding was also observed within genes (31%) or intergenic regions (9%) (Supplementary Figure S4B and C). To deduce a binding motif of CgpS, sequences of the 90 peaks (Supplementary Table S4) were extracted and analyzed using the MEME-ChIP software platform (46). A 21-bp long AT-rich motif was predicted, which was present in 87 of 90 sequences (Figure 3B). The occurrences of the found DNA binding sites were validated using a FIMO search (Find Individual Motif Occurrences, (47)) in the ATCC 13032 genome, which revealed significant matches (>75%) of the predicted and experimentally identified CgpS binding sites (Supplementary Figure S5). Remarkably, the %GC content of the 90 peak sequences is considerably lower than the average GC content of the ATCC 13032 strain, indicating the preferred binding of CgpS to AT-rich DNA (Figure 3C). Moreover, the GC contents of the CgpS bound regions within the prophage revealed no significant differences from that of the regions bound outside the prophage (Figure 3C).

Most of the identified CgpS targets were located within the CGP3 prophage and code for hypothetical proteins. The two strongest signals were found within transposase-encoding genes (cg1950-cg1951) and in the promoter region of *cgpS* itself, indicating a negative autoregulation similar to that of H-NS (48). Other potential target genes encode the actin-like protein and the corresponding adaptor protein (*alpAC*, cg1890 and cg1891 (34)), a resolvase (cg1929), a prophage primase (cg1959), a putative phage lysin (cg1974) and a phage integrase (cg2071), which are spread across the cryptic prophage element. In addition to regions within CGP3, CgpS target sites are located in the low GC island 1 (LCG1), in the cryptic phage element CGP1, or proximal to transposases encoding genes. Furthermore, promoter regions of genes coding for R-M systems (Pcg1028 and P*cglIM*, (Pcg1996)) are also bound by CgpS, which in several studies were shown to be transferred horizontally (49–52). A considerably high peak was observed for the promoter region of cg0150 that encodes a putative regulatory protein or toxin possessing a predicted fido domain (IPR003812).

The binding profile obtained by the ChAP-Seq analysis was validated by EMSAs (Supplementary Figure S6). For this purpose, CgpS was purified as a C-terminal Strep-tag fusion and incubated with DNA fragments covering selected putative CgpS binding sites as identified by ChAP-Seq (Figure 3D and Supplementary Figure S6). This *in vitro* approach confirmed the binding of CgpS for all selected target regions (including the promoters of cg0150, *alpAC* and *cgpS* itself) in comparison to the control fragment (*gntK* promoter) (Figure 3D). Overall, these data are consistent with CgpS acting as a xenogeneic silencer by targeting AT-rich DNA regions, several of which have likely been acquired by HGT.

Countersilencing of CgpS activity

Several independent efforts to inactivate the *cgpS* gene failed (data not shown), suggesting that *cgpS* represents an essential gene for *C. glutamicum* ATCC 13032. However, previous studies revealed that deletions of all three cryptic phage elements, including the *cgpS* gene, are possible and do not lead to a significant growth defect of the particular strain (53). In fact, trials to construct an in-frame deletion of *cgpS* resulted in the isolation of strains lacking large parts of the CGP3 prophage, indicating that the essentiality of *cgpS* is a consequence of the de-repression of toxic phage genes in the absence of CgpS.

For the conditional inactivation of CgpS, we adapted a countersilencing approach similar to the H-NST system described by Williamson and Free (54). This protein was reported as a truncated H-NS derivative that antagonizes H-NS function by interfering with the multimerization of H-NS. Co-purification assays with the N-terminal domain of CgpS confirmed the interaction of this truncated variant

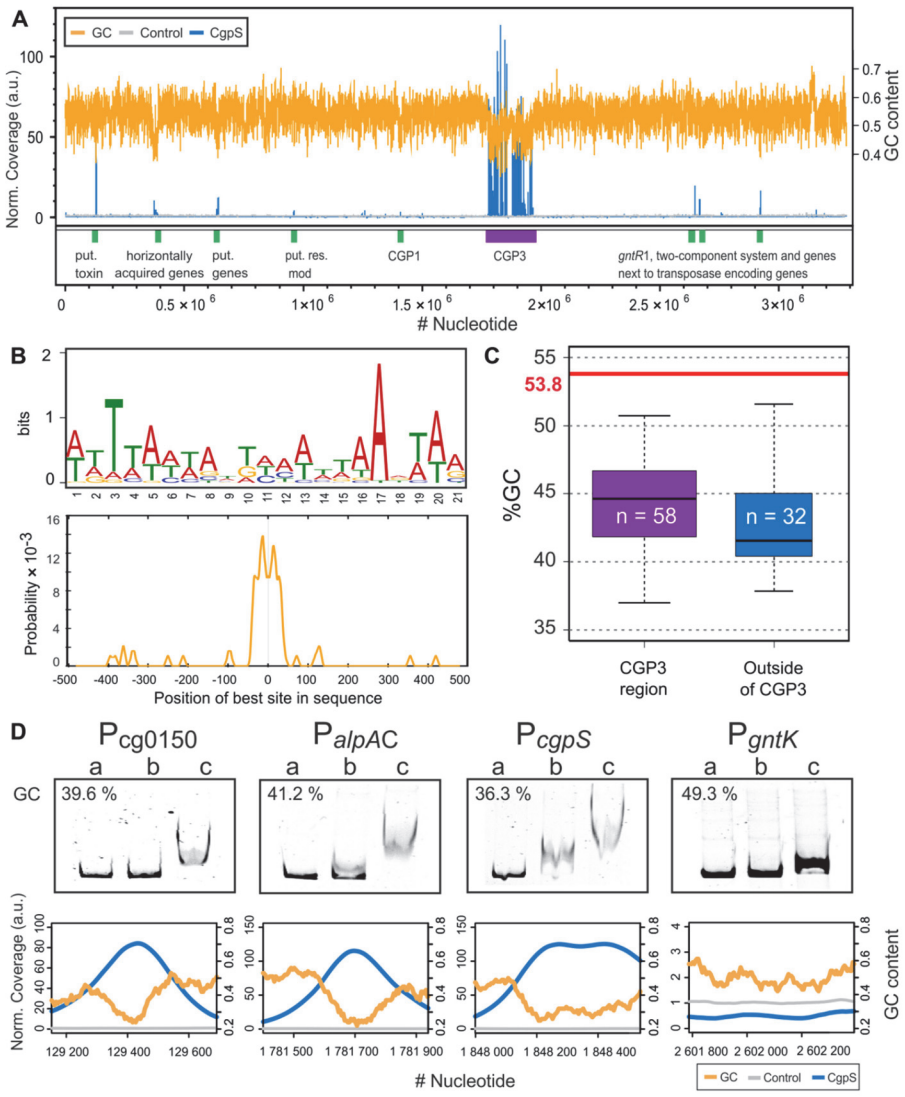


Figure 3. Genome-wide profiling of CgpS-binding using ChAP-Seq. (**A**) Genome wide binding profile of CgpS obtained by ChAP-Seq experiments. Enriched DNA regions purified in complex with CgpS (blue) or with the control sample (gray, empty vector control) were normalized to their mean and plotted against the ATCC 13032 genome. For calculation of the GC content (orange) and the coverages, a step size of 50 bp and a window size of 500 bp were used. Regions with high coverage are indicated by green (outside of CGP3 region) and purple boxes (CGP3 region). A total of 90 peaks were detected by applying a threshold of 3-fold of the mean coverage (SD of the control sample = ± 0.28 (a.u.)). (**B**) Sequences of the 90 peaks were used to derive a DNA binding motif using MEME-ChIP (75). A 21-bp long AT-rich motif was identified (E-value = 1.1×10^{-58}), and the highest probability was centered approximately ±38 bp around the maximum peak position. (**C**) A total of 58 detected signals were within and 32 were outside of the CGP3 region. GC contents of the peak sequences were calculated and compared to the average GC content of ATCC 13032 (21). (**D**) Electrophoretic mobility shift assays (EMSAs) were performed with promoter regions of the putative target genes. DNA fragments (around 500 bp) covering the promoter regions of cg0150, *alpAC*, *cgpS* and *gntK* (negative control) were incubated without (lane a) or with (lane b: 1 μM; c: 2 μM) purified CgpS protein. The corresponding ChAP-seq results of the particular regions are shown below the EMSA pictures (step size: 5 and windows size: 50).

with the full-length protein (Figure 4A). Based on previous data and the H-NST mechanism (Figure 4B), we constructed the pAN6-*N-cgpS* plasmid to overproduce a truncated variant of CgpS (amino acids 1–65) under the control of P_{tac}. Homology studies indicated that amino acids 1–65 cover the domain of CgpS required for the oligomerization of this NAP. Remarkably, production of the truncated CgpS-N domain in the wild-type strain resulted in a significant growth defect, whereas no impact on growth was observed in a strain lacking the CGP3 prophage (Figure 5A). This finding was supported by single-cell analysis of a strain containing a prophage reporter construct (P_{lys}-*eyfp*) (22) and the countersilencing construct pAN6-*N-cgpS*. Production of the N-terminal domain of CgpS led to a strong increase in fluorescence accompanied by growth arrest and a branched cell morphology (Figure 5C, Video S1 and S2). Quantitative real-time PCR revealed a 3-fold increase in the level of circular CGP3 DNA in comparison to uninduced cells, which is consistent with the induction of this cryptic prophage (Figure 5B) (20).

To monitor the impact of countersilencing CgpS activity on gene expression, we performed a comparative transcriptome analysis (Figure 5D, Supplementary Table S5). More than 194 genes were affected, 12 of which exhibited a reduced mRNA level (mRNA ratio ≤ 0.5, P-value < 0.05), and 182 genes were upregulated (mRNA ratio ≥ 2, P-value < 0.05). The majority of upregulated genes (148) were genes of the prophage CGP3. Additional genes that displayed an increased mRNA level were the ferritin gene (*ftn,* cg2782) and cg1517 of the CGP1 prophage (Supplementary Table S5), both of which were also identified as putative CgpS targets by ChAP-Seq. Together, these data demonstrate that CgpS is an essential NAP due to its function as a silencer of cryptic phage elements in *C. glutamicum*.

CgpS homologs are found in actinomycetes and their phages

Our data support a function for CgpS as a xenogeneic silencer that binds to AT-rich DNA similar to the Lsr2 of *M. tuberculosis* as well as the H-NS of *E. coli.* This is underlined by the fact that both proteins, Lsr2 and CgpS, are able to complement the phenotype of an *hns* mutant strain ((18), Supplementary Figure S7). These findings highlight the conserved mechanism of a highly diverse set of proteins.

In the following, we overexpressed the N-terminal oligomerization domains of CgpS orthologs from *Corynebacterium amycolatum* DSM 44737 (CORAM0001_2081) and *Corynebacterium diphtheria* DSM 44123 (CDC7B_2240) and the Lsr2 from *M. tuberculosis* H37R (Rv3597c; Lsr2) (Figure 6A and B). Whereas the production of the oligomerization domain strongly affected cellular growth in all cases (Figure 6A), only the N-terminal domain of the ortholog of *C. amycolatum* (DSM 44737) led to a significant induction of CGP3 (Figure 6B). No significant reporter output was observed with production of the truncated orthologs of *C. diphtheria* or *M. tuberculosis*, suggesting a high level of plasticity within this family of xenogeneic silencers (Figure 6B).

Furthermore, we used a bioinformatics approach to obtain a more general overview of the distribution of CgpS orthologous proteins. For this purpose, a PSI-BLAST (Position-Specific Iterated BLAST) search was performed on CgpS and resulted in 5230 hits, of which 1920 protein sequences were unique (threshold e-value ≤ 0.005). Of these, 98.3% were found in the domain of bacteria and 1.7% in phages, mostly belonging to the *Siphoviridae* (Figure 7A, Supplementary Table S6). Of 302 bacterial genomes containing prophage regions predicted by PhiSpy (55), 22 contain *cgpS* orthologs (Supplementary Table S6). The remaining 280 hits were found outside of any predicted prophage region. Moreover, secondary structure predictions were performed for 618 unique sequences, which were clearly assigned to bacterial or phage species, exhibiting high resemblances. The structural similarity suggests a common function, although the identity of the amino acid sequences is low (~23%) (Figure 7B, C and Supplementary Figure S10).

XS exclusion hypothesis

A recent bioinformatics study on the distribution of XS genes revealed that members of the same family can appear within a particular species but that members of different families are never found together (56). To test the proposed exclusion mechanism, we expressed the *hns* gene from *E. coli* MG1655 in a *C. glutamicum* ATCC 13032 strain containing the prophage reporter (::P_{lys}-*eyfp*). As expected, the overexpression of *hns* caused a severe growth defect, coinciding with a highly increased output of the prophage reporter (Figure 6C and D). The effect of *hns* overexpression was comparable to the countersilencing of CgpS activity with the production of a truncated CgpS variant (Figure 6E). When *hns* was expressed in a ΔCGP3 background the effect on growth was only moderate (Supplementary Figure S8). However, *hns* expression still negatively affected the growth of the CGP3 mutant strain which can likely be explained by unspecific binding and interference of H-NS at other genomic regions. These findings are in agreement with the hypothesis that different XS proteins interfere at AT-rich DNA regions, leading to a disruption of silencing complexes and thereby to an activation of foreign DNA elements. Nevertheless, in some cases the scenario is clearly more complex, as illustrated by the finding that the expression of *cgpS* in the *E. coli* wild-type strain was not able to counteract H-NS expression at the *bgl* operon (Supplementary Figure S9).

DISCUSSION

CgpS functions as a silencer of cryptic phage elements

In this study, we identified the prophage-encoded XS protein CgpS that inherits an essential role as a silencer of cryptic prophages in *C. glutamicum*. Genome-wide profiling of CgpS binding sites reveals an association of this protein to AT-rich DNA stretches primarily located within horizontally acquired genomic islands and shows a remarkable accumulation of binding sites within the large and cryptic CGP3 prophage. Countersilencing of CgpS activity by overproduction of its N-terminal oligomerization domain resulted in a strong increase in CGP3 activity leading to cell death. Furthermore, several CgpS binding sites were identified outside the CGP3 region, and the essentiality of the *cgpS* gene was attributed to the presence of the CGP3

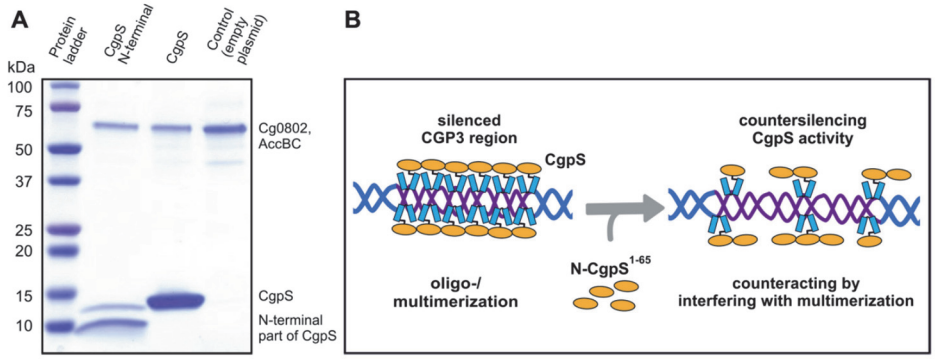


Figure 4. Principle of CgpS countersilencing. (**A**) Protein pulldown assays were conducted with *C. glutamicum* cells containing the plasmids pAN6-*cgpS-Strep* and pAN6-*N-cgpS-Strep*, which were used for the overexpression of Strep-tagged *cgpS* and its truncated variant. The pulldown of full-length CgpS by the truncated variant (aa 1–65) confirmed the N-terminal part of CgpS as oligo-/multimerization domain of CgpS proteins. (**B**) Model of CgpS silencing and countersilencing. The expression of genes depends on the accessibility of their particular promoter regions to the RNA polymerase. AT-rich regions such as CGP3 are bound by CgpS, likely resulting in an oligomerization of the CgpS protein (18,76,77), thereby interfering with the binding of RNA polymerase. The production of an N-terminal truncated CgpS variant interferes with the silencing ability of the native protein by binding to its N-terminal oligomerization domain while lacking the site for DNA binding.

prophage. This is consistent with the finding that the *cgpS* gene is located on the CGP3 island, suggesting that evolution favored a physical association between this XS and its main target.

Sequence analysis of CgpS revealed a low sequence identity (27%, Supplementary Figure S1) with the mycobacterial Lsr2 protein that was described in previous studies as an H-NS-like protein targeting AT-rich sequences in *M. tuberculosis* (18). Both XS proteins, Lsr2 and CgpS, complemented the *bgl*-based phenotype (57) of an *Escherichia coli* Δ*hns* strain, supporting the overall analogous functions of these XS proteins (Supplementary Figure S7) (18). Whereas both *lsr2* and *cgpS* are essential for viability in their native hosts, *E. coli hns* mutant strains are viable although exhibiting severe growth defects (58). *Salmonella Typhimurium* null mutants of *hns* are not viable unless mutations in *rpoS* (general stress response) or *phoP* (virulence gene regulator) counteract this deletion (12). Because the presence and diversity of phage elements contributes to major strain-specific differences within a bacterial species, our study illustrates that the essentiality of XS genes is highly dependent on the particular strain background. The *C. glutamicum* strain MB001, cured of all prophage regions as well as the *cgpS* gene located on prophage CGP3, displays wild-type-like growth behavior (53).

CgpS binds AT-rich xenogeneic DNA regions

Secondary structure predictions of CgpS-related proteins evince two α-helices flanking an 'RGI' motif (Figures 1C and 7C). This motif resembles the prokaryotic AT-hook motif 'Q/RGR' found in H-NS and Lsr2 and may also be responsible for the binding of AT-rich DNA as a general rule for XS functioning (44,59). A certain plasticity of the AT-hook motif is supported by experiments with AT-hook muteins of H-NS and Lsr2, showing that the exchange of a single arginine residue to an alanine reduces DNA binding but does not completely abolish it (59). Moreover, another member of the H-NS family, the Ler protein, has a hydrophobic amino acid ('VGR' motif) instead of an arginine at this position (60).

However, significant differences were observed for the number of target genes affected by the binding of the particular XS proteins. ChIP-on-Chip analysis revealed a direct influence of *S. Typhimurium* H-NS on the expression of more than 740 ORFs (12,61), and the binding of Lsr2 affected more than 800 regions within the *M. tuberculosis* genome and >900 in *Mycobacterium smegmatis* (45). ChAP-Seq profiling of CgpS binding, however, yielded only 90 potential target regions. Typical for XS function, an AT-rich DNA motif was derived from the ChAP-Seq results, which clusters at a high density within the CGP3 prophage region (Supplementary Figure S5). In general, promoter regions are more often bound by CgpS than genes or intergenic regions (Supplementary Figure S4), which is not surprising because promoter regions usually possess a higher AT content (62,63). CgpS targets outside the CGP3 region show a similar or lower GC content (Figure 3C) but less altered expression levels, and this may suggest the importance of motif density for XS function. Here, a variation of the AT-hook motif likely represents a mechanism to adjust the binding behavior of the XS protein to meet the needs of a particular host species.

In addition to CGP3 as a main CgpS target, further targets were identified which were also likely acquired by horizontal gene transfer, such as the LCG1 island, the cryptic prophage CGP1 (21), R-M systems, transposases and also regulatory proteins such as putative transcriptional regulators (Cg0725, Cg1340, Cg2426), the gluconate-responsive repressor GntR1 (Cg2783) (64) and an operon encoding the two-component system CgtSR6 (Cg3060) (Supplementary Table S4). Several previous studies reported similar tar-

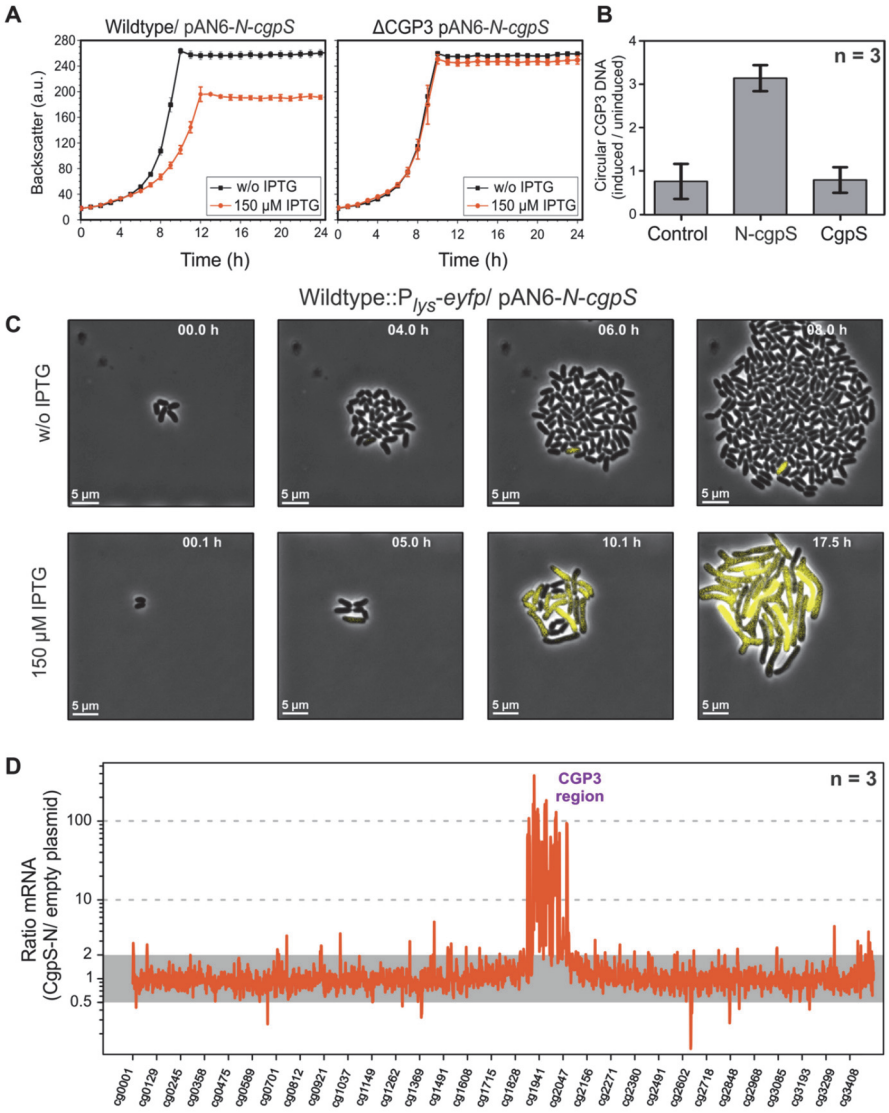


Figure 5. Countersilencing by overexpression of the N-terminal oligomerization domain of CgpS. (**A**) Growth studies of the wild-type ATCC 13032 and a strain lacking the CGP3 prophage (ΔCGP3), both carrying the pAN6-N-*cgpS* overexpression plasmid. Strains were grown in CGXII minimal medium with and without IPTG (150 µM). Data represent average values and standard deviations of three biological replicates. (**B**) Relative quantification of CGP3 excision using qPCR (20). The N-terminal domain of CgpS and the full-length protein were overproduced as described in (A). Samples for qPCR analysis were taken after 24 h. The relative amounts of circular phage DNA of induced and uninduced samples were compared. As a control wild-type cells with the empty plasmid were used. Data represent average values and standard deviations of three biological replicates. (**C**) Time-lapse fluorescence microscopy of the *C. glutamicum* prophage reporter strain (ATCC 13032::P*lys-eyfp*) carrying the pAN6-*N-cgpS*. Cells were grown in PDMS-based microfluidic chip devices under continuous supply of CGXII with 25 µg·ml^{-1} kanamycin and with or without 150 µM IPTG to induce the expression of the truncated CgpS variant (36) (300 nl·min^{-1}) (Video S1 and S2). (**D**) Comparative transcriptome analysis of *C. glutamicum* ATCC 13032 containing the overexpression plasmid pAN6-*N-cgpS* and a strain containing the empty vector control was performed as described in the Materials and Methods section.

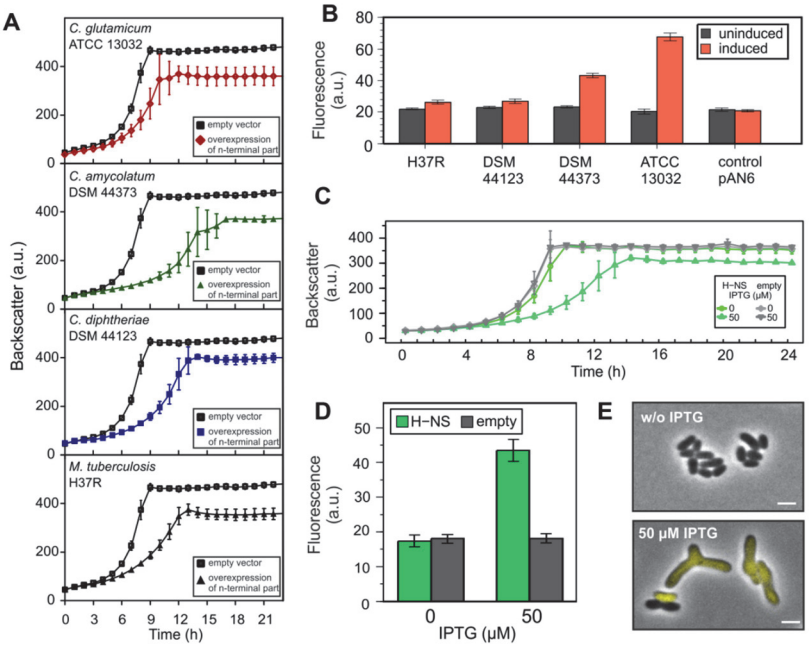


Figure 6. Impact of H-NS and Lsr2/Lsr2-like truncated variants on CGP3 prophage induction. (**A** and **B**) Impact of truncated CgpS orthologs on the growth and CGP3 prophage activity using the reporter strain ATCC 13032::P$_{lys}$-*eyfp*. Shown is the cultivation in microtiter plates in CGXII minimal media with 25 µg·ml^{-1} kanamycin and 150 µM IPTG. As a control the reporter strain containing the empty plasmid was used. In (**B**), the fluorescence output after 20 h is shown in comparison to the uninduced samples. (**C** and **D**) Growth experiments were performed with WT::P*lys-eyfp* cells carrying *hns* on the overexpression plasmid pAN6 under the conditions described in (**A**). Expression of *hns* was induced with 50 µM IPTG. Fluorescence output of the prophage reporter after 20 h is shown in (**D**). (**E**) After 24 h, fluorescence images were taken of cells (of C and D) placed on agar pads. Scale bar is 2 µm. Data represent average values and standard deviations of three biological replicates.

get genes or regions for H-NS, Lsr2 and MvaT, demonstrating the convergent evolution of XS in bacterial species (12,47,61,65).

Overall, more that 80% of CgpS-bound regions also exhibited a more than 2-fold altered expression level under countersilencing conditions (Figure 5D) confirming the postulated silencing effect of CgpS. Several potential targets outside of the CGP3 region, however, showed only a moderate impact on the expression level suggesting a more complex regulatory scheme at the corresponding promoter regions. Therefore, the role of CgpS for the control of these potential targets, including, e.g. the *gntR1* gene or the *cgtSR6* operon, remains to be elucidated in further studies.

How to overcome CgpS silencing?

Several different mechanisms were described to counteract H-NS-mediated silencing, including structural interference with H-NS-bound nucleoids by transcription factors, temperature or osmolarity effects, and the binding of alternative sigma factors or other NAPs preventing multimerization of the XS protein (11,66,67). To interfere with CgpS XS activity, we produced a truncated part of the native protein covering the N-terminal oligomerization domain of CgpS (Figure 5). This overcomes the problem of *cgpS* being essential in the presence of CGP3 and was inspired by the study of Williamson and Free, who described the antagonistic function of a truncated H-NS variant found in an enteropathogenic *E. coli* strain (54). As expected, production of the N-terminal CgpS domain resulted in strong activation of CGP3, leading to cell death.

In recent studies we described the spontaneous induction of the CGP3 prophage occuring in the absence of an external trigger (20,22,23). Single-cell analysis demonstrated that a considerable fraction of this SPI is preceded by an activation of the SOS response, which is likely the result of spontaneous DNA damage during replication (68,69). However, these studies also highlighted a certain (>30%) fraction of SOS-independent SPI, suggesting that other factors influence this common phenomenon of bacterial populations (5). The present study shows the sensitive reaction of *C. glutamicum* cells to the downregulation of CgpS activity (Video S2). It is therefore interesting to determine whether cells can adjust the level of XS proteins to manipulate the frequency of SPI according to their particular requirements.

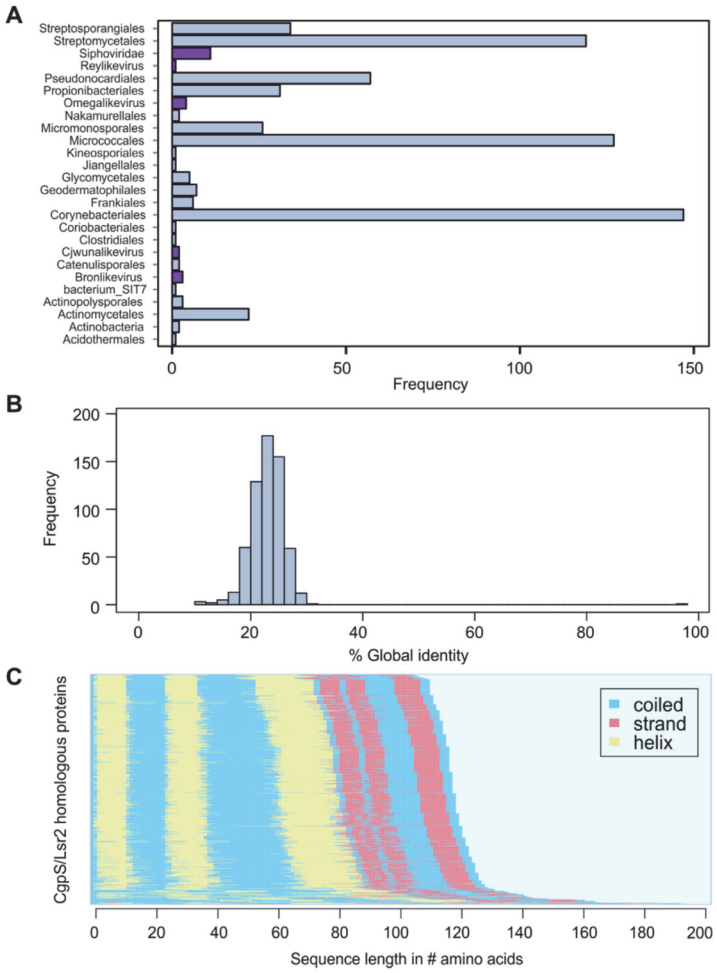


Figure 7. Orthologous sequences of the CgpS protein in actinomycetes and their phages. (**A**) Bar chart depicting the frequency of orthologous hits (n = 618) as predicted by PSI-BLAST (39) (e-value ≤ 0.005) across several orders of the phylum Actinobacteria and phages as annotated in the NCBI database (http://www.ncbi.nlm.nih.gov/). Orthologous found in phages are highlighted in purple. (**B**) Histogram of pairwise global identities between the CgpS amino acid sequence and its orthologous counterparts. The distribution reveals an overall low similarity (mean x = 23.07 and standard deviation σ = 4.05) to the orthologous sequences. Global identity was calculated using the Needleman–Wunsch (40) algorithm from the EMBOSS package (41). (**C**) Secondary structure prediction calculated by psipred (42) shows conserved protein structure for CgpS and the orthologous amino acid sequences. The corresponding secondary structure of each sequence was ordered in the direction from C-Terminus to N-terminus. Predicted coiled structures are shown in blue, strand regions in red and helices are colored in yellow.

CgpS homologs in phage genomes

Sequence analysis revealed the presence of CgpS/Lsr2 homologs in phage and prophage genomes displaying a low sequence identity but highly conserved secondary structure prediction (Figure 7). This finding is not surprising because bacterial evolution has been shaped by a tight interaction with bacteriophages. For the integration of viral DNA into the host genome, both the bacterium and phage benefit from tolerance and a smooth integration into the host genetic circuitry. Because the activation of silent prophages or mobile elements often causes serious detrimental effects to host cells (11,70,71), the stringent control of xenogeneic elements is required.

Several examples of XS proteins involved in the control of mobile elements or phages have been described in the recent literature, including H-NS of *S. Typhimurium* (12), Rok from *B. subtilis* (19) and MvaT from *P. aeruginosa* (72). Their corresponding genes, however, are all located on the host chromosome and are characterized as a type of immunity system protecting hosts against foreign DNA (11,66). A PSI-BLAST search of CgpS-related proteins revealed that the majority (>98% of all hits, >92% of prophage containing strains) are found in bacterial genomes (Supplementary Table S6). However, several examples located in phages or prophage regions were identified. The functions of these phage-encoded XS-like proteins remain to be studied, but their presence suggests the following: (i) like CgpS, they may be required to secure tolerance of their carrier DNA within the respective host; (ii) they may, however, also function as antagonistic proteins, interfering with the host XS protein similar to the situation described for H-NST (54); or (iii) they may interfere with the function of another class of XS proteins. This hypothesis is based on the exclusion theory suggested by Perez-Rueda and Ibarra, who postulated that XS from different families do not appear in the same bacterial organism (56). Consistent with this bioinformatics study, our data show that the expression of *E. coli hns* results in strong activation of the cryptic prophage CGP3 and consequently cell death. The finding that expression of the *C. glutamicum cgpS* gene in *E. coli* MG1655 does not counteract H-NS-mediated silencing at the *bgl* operon shows, however, that the scenario is more complex and strongly depends on the particular strain and its regulatory equipment. However, our data on prophage activation in *C. glutamicum* provide evidence for an interference of analogous XS proteins at AT-rich DNA regions. Here, likely the incompatibility of the oligomerization domains inhibits the formation of XS multimeric structures required for silencing. Considering the presence of XS encoding genes in phage and prophage genomes, this principle is likely to be harnessed by any phage predator by encoding an interfering XS.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors thank Karin Schnetz (University of Cologne) for helpful advice and for providing us with the *E. coli* Δ*hns* mutant strain.

FUNDING

Deutsche Forschungsgemeinschaft priority program SPP1617 [FR 2759/2-2 and KO 4537/1-2]; Helmholtz Association [VH-NG-716]. Funding for open access charge: Helmholtz Association [VH-NG-716].
Conflict of interest statement. None declared.

REFERENCES

1. Juhas,M., van der Meer,J.R., Gaillard,M., Harding,R.M., Hood,D.W. and Crook,D.W. (2009) Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.*, **33**, 376–393.
2. Canchaya,C., Proux,C., Fournous,G., Bruttin,A. and Brussow,H. (2003) Prophage genomics. *Microbiol. Mol. Biol. Rev.*, **67**, 238–276.
3. Casjens,S. (2003) Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.*, **49**, 277–300.
4. Soucy,S.M., Huang,J. and Gogarten,J.P. (2015) Horizontal gene transfer: Building the web of life. *Nat. Rev. Genet.*, **16**, 472–482.
5. Nanda,A.M., Thormann,K. and Frunzke,J. (2015) Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J. Bacteriol.*, **197**, 410–419.
6. Wang,X., Kim,Y., Ma,Q., Hong,S.H., Pokusaeva,K., Sturino,J.M. and Wood,T.K. (2010) Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.*, **1**, 147.
7. d'Hérelle,F. (1917) Sur un microbe invisible antagoniste des bacilles dysentérique. *Acad. Sci. Paris*, **165**, 373–375.
8. Diaz Ricci,J.C. and Hernandez,M.E. (2000) Plasmid effects on *Escherichia coli* metabolism. *Crit. Rev. Biotechnol.*, **20**, 79–108.
9. Labrie,S.J., Samson,J.E. and Moineau,S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.*, **8**, 317–327.
10. Samson,J.E., Magadan,A.H., Sabri,M. and Moineau,S. (2013) Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.*, **11**, 675–687.
11. Navarre,W.W., McClelland,M., Libby,S.J. and Fang,F.C. (2007) Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. *Genes Dev.*, **21**, 1456–1471.
12. Navarre,W.W., Porwollik,S., Wang,Y., McClelland,M., Rosen,H., Libby,S.J. and Fang,F.C. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science*, **313**, 236–238.
13. Will,W.R., Navarre,W.W. and Fang,F.C. (2015) Integrated circuits: how transcriptional silencing and counter-silencing facilitate bacterial evolution. *Curr. Opin. Microbiol.*, **23**, 8–13.
14. Langille,M.G., Hsiao,W.W. and Brinkman,F.S. (2010) Detecting genomic islands using bioinformatics approaches. *Nat. Rev. Microbiol.*, **8**, 373–382.
15. Dame,R.T., Luijsterburg,M.S., Krin,E., Bertin,P.N., Wagner,R. and Wuite,G.J. (2005) DNA bridging: A property shared among H-NS-like proteins. *J. Bacteriol.*, **187**, 1845–1848.
16. Ding,P., McFarland,K.A., Jin,S., Tong,G., Duan,B., Yang,A., Hughes,T.R., Liu,J., Dove,S.L., Navarre,W.W. *et al.* (2015) A novel AT-rich DNA recognition mechanism for bacterial xenogeneic silencer MvaT. *PLoS Pathog.*, **11**, e1004967.
17. Oshima,T., Ishikawa,S., Kurokawa,K., Aiba,H. and Ogasawara,N. (2006) *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res.*, **13**, 141–153.
18. Gordon,B.R., Imperial,R., Wang,L., Navarre,W.W. and Liu,J. (2008) Lsr2 of *Mycobacterium* represents a novel class of H-NS-like proteins. *J. Bacteriol.*, **190**, 7052–7059.
19. Smits,W.K. and Grossman,A.D. (2010) The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in *Bacillus subtilis*. *PLoS Genet.*, **6**, e1001207.
20. Frunzke,J., Bramkamp,M., Schweitzer,J.E. and Bott,M. (2008) Population Heterogeneity in *Corynebacterium glutamicum* ATCC 13032 caused by prophage CGP3. *J. Bacteriol.*, **190**, 5111–5119.
21. Kalinowski,J., Bathe,B., Bartels,D., Bischoff,N., Bott,M., Burkovski,A., Dusch,N., Eggeling,L., Eikmanns,B.J., Gaigalat,L. *et al.* (2003) The complete *Corynebacterium glutamicum* ATCC 13032

genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J. Biotechnol.*, **104**, 5–25.
22. Helfrich,S., Pfeifer,E., Krämer,C., Sachs,C.C., Wiechert,W., Kohlheyer,D., Nöh,K. and Frunzke,J. (2015) Live cell imaging of SOS and prophage dynamics in isogenic bacterial populations. *Mol. Microbiol.*, **98**, 636–650.
23. Nanda,A.M., Heyer,A., Krämer,C., Grünberger,A., Kohlheyer,D. and Frunzke,J. (2014) Analysis of SOS-induced spontaneous prophage induction in *Corynebacterium glutamicum* at the single-cell level. *J. Bacteriol.*, **196**, 180–188.
24. Kinoshita,S., Udaka,S. and Shimono,M. (1957) Studies on the amino acid fermentation - Part I. Production of L-glutamic acid by various microorganisms. *J. Gen. Appl. Microbiol.*, **50**, 331–343.
25. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
26. Keilhauer,C., Eggeling,L. and Sahm,H. (1993) Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon. *J. Bacteriol.*, **175**, 5595–5603.
27. Gibson,D.G., Young,L., Chuang,R.Y., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 343–345.
28. Niebisch,A. and Bott,M. (2001) Molecular analysis of the cytochrome bc1-aa3 branch of the *Corynebacterium glutamicum* respiratory chain containing an unusual diheme cytochrome c1. *Arch. Microbiol.*, **175**, 282–294.
29. Heyer,A., Gätgens,C., Hentschel,E., Kalinowski,J., Bott,M. and Frunzke,J. (2012) The two-component system ChrSA is crucial for haem tolerance and interferes with HrrSA in haem-dependent gene regulation in *Corynebacterium glutamicum*. *Microbiology*, **158**, 3020–3031.
30. Kim,S.C., Chen,Y., Mirza,S., Xu,Y., Lee,J., Liu,P. and Zhao,Y. (2006) A clean, more efficient method for in-solution digestion of protein mixtures without detergent or urea. *J. Proteome Res.*, **5**, 3446–3452.
31. Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
32. Evans,G.A. Molecular cloning: A laboratory manual. Second edition. Volumes 1, 2, and 3. Current protocols in molecular biology. Volumes 1 and 2. *Cell*, **61**, 17–18.
33. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
34. Donovan,C., Heyer,A., Pfeifer,E., Polen,T., Wittmann,A., Krämer,R., Frunzke,J. and Bramkamp,M. (2015) A prophage-encoded actin-like protein required for efficient viral DNA replication in bacteria. *Nucleic Acids Res.*, **43**, 5002–5016.
35. Grünberger,A., Paczia,N., Probst,C., Schendzielorz,G., Eggeling,L., Noack,S., Wiechert,W. and Kohlheyer,D. (2012) A disposable picolitre bioreactor for cultivation and investigation of industrially relevant bacteria on the single cell level. *Lab Chip*, **12**, 2060–2068.
36. Grünberger,A., Probst,C., Helfrich,S., Nanda,A., Stute,B., Wiechert,W., von Lieres,E., Nöh,K., Frunzke,J. and Kohlheyer,D. (2015) Spatiotemporal microbial single-cell analysis using a high-throughput microfluidics cultivation platform. *Cytometry*, **87A**, 1101–1115.
37. Grünberger,A., van Ooyen,J., Paczia,N., Rohe,P., Schiendzielorz,G., Eggeling,L., Wiechert,W., Kohlheyer,D. and Noack,S. (2013) Beyond growth rate 0.6: *Corynebacterium glutamicum* cultivated in highly diluted environments. *Biotechnol. Bioeng.*, **110**, 220–228.
38. Bussmann,M., Baumgart,M. and Bott,M. (2010) RosR (Cg1324), a hydrogen peroxide-sensitive MarR-type transcriptional regulator of *Corynebacterium glutamicum*. *J. Biol. Chem.*, **285**, 29305–29318.
39. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J.H., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
40. Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.
41. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
42. Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
43. R Development Core Team (2016) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, http://www.R-project.org.
44. Ali,S.S., Xia,B., Liu,J. and Navarre,W.W. (2012) Silencing of foreign DNA in bacteria. *Curr. Opin. Microbiol.*, **15**, 175–181.
45. Gordon,B.R., Li,Y., Wang,L., Sintsova,A., van Bakel,H., Tian,S., Navarre,W.W., Xia,B. and Liu,J. (2010) Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 5154–5159.
46. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
47. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
48. Dersch,P., Schmidt,K. and Bremer,E. (1993) Synthesis of the *Escherichia coli* K-12 nucleoid-associated DNA-binding protein H-NS is subjected to growth-phase control and autoregulation. *Mol. Microbiol.*, **8**, 875–889.
49. Khan,F., Furuta,Y., Kawai,M., Kaminska,K.H., Ishikawa,K., Bujnicki,J.M. and Kobayashi,I. (2010) A putative mobile genetic element carrying a novel type IIF restriction-modification system (PluTI). *Nucleic Acids Res.*, **38**, 3019–3030.
50. Nobusato,A., Uchiyama,I. and Kobayashi,I. (2000) Diversity of restriction-modification gene homologues in *Helicobacter pylori*. *Gene*, **259**, 89–98.
51. Jeltsch,A. and Pingoud,A. (1996) Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *J. Mol. Evol.*, **42**, 91–96.
52. Furuta,Y. and Kobayashi,I. (2014) *Bacterial Integrative Mobile Genetic Elements*. Austin.
53. Baumgart,M., Unthan,S., Rückert,C., Sivalingam,J., Grünberger,A., Kalinowski,J., Bott,M., Noack,S. and Frunzke,J. (2013) Construction of a prophage-free variant of *Corynebacterium glutamicum* ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl. Environ. Microbiol.*, **79**, 6006–6015.
54. Williamson,H.S. and Free,A. (2005) A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist. *Mol. Microbiol.*, **55**, 808–827.
55. Akhter,S., Aziz,R.K. and Edwards,R.A. (2012) PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.*, **40**, e126.
56. Perez-Rueda,E. and Ibarra,J.A. (2015) Distribution of putative xenogeneic silencers in prokaryote genomes. *Comput. Biol. Chem.*, **58**, 167–172.
57. Dole,S., Kühn,S. and Schnetz,K. (2002) Post-transcriptional enhancement of *Escherichia coli bgl* operon silencing by limitation of BglG-mediated antitermination at low transcription rates. *Mol. Microbiol.*, **43**, 217–226.
58. Yamada,H., Yoshida,T., Tanaka,K., Sasakawa,C. and Mizuno,T. (1991) Molecular analysis of the *Escherichia coli hns* gene encoding a DNA-binding protein, which preferentially recognizes curved DNA sequences. *Mol. Gen. Genet.*, **230**, 332–336.
59. Gordon,B.R., Li,Y., Cote,A., Weirauch,M.T., Ding,P., Hughes,T.R., Navarre,W.W., Xia,B. and Liu,J. (2011) Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10690–10695.
60. Cordeiro,T.N., Schmidt,H., Madrid,C., Juarez,A., Bernado,P., Griesinger,C., Garcia,J. and Pons,M. (2011) Indirect DNA readout by an H-NS related protein: structure of the DNA complex of the C-terminal domain of Ler. *PLoS Pathog.*, **7**, e1002380.
61. Lucchini,S., Rowley,G., Goldberg,M.D., Hurd,D., Harrison,M. and Hinton,J.C. (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.*, **2**, e81.
62. Pedersen,A.G., Jensen,L.J., Brunak,S., Staerfeldt,H.H. and Ussery,D.W. (2000) A DNA structural atlas for *Escherichia coli*. *J. Mol. Biol.*, **299**, 907–930.
63. Ussery,D.W., Tindbaek,N. and Hallin,P.F. (2004) Genome update: promoter profiles. *Microbiology*, **150**, 2791–2793.
64. Frunzke,J., Engels,V., Hasenbein,S., Gätgens,C. and Bott,M. (2008) Co-ordinated regulation of gluconate catabolism and glucose uptake in *Corynebacterium glutamicum* by two functionally equivalent

transcriptional regulators, GntR1 and GntR2. *Mol. Microbiol.*, **67**, 305–322.
65. Castang,S., McManus,H.R., Turner,K.H. and Dove,S.L. (2008) H-NS family members function coordinately in an opportunistic pathogen. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 18947–18952.
66. Navarre,W.W. (2009) In: Dame,RT and Dorman,CJ (eds). *Bacterial Chromatin*. Springer, Netherlands.
67. Stoebel,D.M., Free,A. and Dorman,C.J. (2008) Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology*, **154**, 2533–2545.
68. Cox,M.M., Goodman,M.F., Kreuzer,K.N., Sherratt,D.J., Sandler,S.J. and Marians,K.J. (2000) The importance of repairing stalled replication forks. *Nature*, **404**, 37–41.
69. Pennington,J.M. and Rosenberg,S.M. (2007) Spontaneous DNA breakage in single living *Escherichia coli* cells. *Nat. Genet.*, **39**, 797–802.
70. Buckling,A. and Rainey,P.B. (2002) Antagonistic coevolution between a bacterium and a bacteriophage. *Proc. Biol. Sci.*, **269**, 931–936.
71. Lee,S.W. and Edlin,G. (1985) Expression of tetracycline resistance in pBR322 derivatives reduces the reproductive fitness of plasmid-containing *Escherichia coli*. *Gene*, **39**, 173–180.
72. Li,C., Wally,H., Miller,S.J. and Lu,C.D. (2009) The multifaceted proteins MvaT and MvaU, members of the H-NS family, control arginine metabolism, pyocyanin synthesis, and prophage activation in *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.*, **191**, 6211–6218.
73. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Soding,J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
74. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
75. Ma,W., Noble,W.S. and Bailey,T.L. (2014) Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat. Protoc.*, **9**, 1428–1450.
76. Cam,E.L., Culard,F., Larquet,E., Delain,E. and Cognet,J.A. (1999) DNA bending induced by the archaebacterial histone-like protein MC1. *J. Mol. Biol.*, **285**, 1011–1021.
77. Chen,J.M., Ren,H., Shaw,J.E., Wang,Y.J., Li,M., Leung,A.S., Tran,V., Berbenetz,N.M., Kocincova,D., Yip,C.M. *et al.* (2008) Lsr2 of *Mycobacterium tuberculosis* is a DNA-bridging protein. *Nucleic Acids Res.*, **36**, 2123–2135.

3.2 Impact of Xenogeneic Silencing on Phage-Host Interactions

Pfeifer E., **Hünnefeld M.**, Popa O., and Frunzke J.

Published in *Journal of Molecular Biology,* 2019

Contributor Role	Contributor
Conceptualization	EP (40 %), JF (30 %), MH (25 %), OP (5 %)
Formal Analysis	EP (55 %), OP (40 %), MH (5 %)
Investigation/Experiments	-
Methodology	EP (50 %), OP (50 %)
Project Administration	EP (50 %), JF (40 %), MH (10 %)
Software	OP (50 %), EP (50 %)
Supervision	EP (50 %), JF (40 %), MH (10 %)
Visualization	EP (60 %), OP (25 %), MH (15 %)
Writing – Original Draft Preparation	EP (50 %), MH (35 %), JF (10%), OP (5%)
Writing – Review & Editing	JF (40 %), EP (30 %), MH (25 %), OP (5 %)

Overall contribution MH: 30 %

MH was involved in the writing process and contributed here with writing the chapter ‘How to overcome silencing?’. Furthermore, MH took part in the conceptualization of the review article, and participated in the complete editing process. Table 2 was mainly prepared by MH with some additions from EP.

ARTICLE IN PRESS

Review

KDC YJMBI-66015; No. of pages: 14; 4C:

Impact of Xenogeneic Silencing on Phage–Host Interactions

Eugen Pfeifer[1], Max Hünnefeld[1], Ovidiu Popa[2] and Julia Frunzke[1]

1 - Forschungszentrum Jülich GmbH, Institute for Bio- and Geosciences 1, IBG1, 52425 Jülich, Germany
2 - Heinrich Heine Universität Düsseldorf, Institute for Quantitative and Theoretical Biology, 40223 Düsseldorf, Germany

Correspondence to Eugen Pfeifer and Julia Frunzke: *e.pfeifer@fz-juelich.de, j.frunzke@fz-juelich.de.*
https://doi.org/10.1016/j.jmb.2019.02.011
Edited by Kirsten Jung

Abstract

Phages, viruses that prey on bacteria, are the most abundant and diverse inhabitants of the Earth. Temperate bacteriophages can integrate into the host genome and, as so-called prophages, maintain a long-term association with their host. The close relationship between host and virus has significantly shaped microbial evolution and phage elements may benefit their host by providing new functions. Nevertheless, the strong activity of phage promoters and potentially toxic gene products may impose a severe fitness burden and must be tightly controlled. In this context, xenogeneic silencing (XS) proteins, which can recognize foreign DNA elements, play an important role in the acquisition of novel genetic information and facilitate the evolution of regulatory networks. Currently known XS proteins fall into four classes (H-NS, MvaT, Rok and Lsr2) and have been shown to follow a similar mode of action by binding to AT-rich DNA and forming an oligomeric nucleoprotein complex that silences gene expression. In this review, we focus on the role of XS proteins in phage–host interactions by highlighting the important function of XS proteins in maintaining the lysogenic state and by providing examples of how phages fight back by encoding inhibitory proteins that disrupt XS functions in the host. Sequence analysis of available phage genomes revealed the presence of genes encoding Lsr2-type proteins in the genomes of phages infecting Actinobacteria. These data provide an interesting perspective for future studies to elucidate the impact of phage-encoded XS homologs on the phage life cycle and phage–host interactions.

> "*You have a grand gift for silence, Watson. It makes you quite invaluable as a companion.*"
> [(Sherlock Holmes)]

Introduction

Phages, viruses that prey on bacteria, represent the most abundant biological entities on this planet and are a major driver of horizontal gene transfer (HGT). Phages are not only present as infectious particles in the environment but are also found as integrated elements (prophages) within the genomes of their bacterial hosts. In some cases, DNA of viral origin accounts for up to 20% of an organism's entire genome [1–3]. Some of this DNA originates from fully functional prophages, which are capable of undergoing a lytic life cycle. However, a considerable part is made up of prophage-like elements, including phage remnants left after incomplete excision events, cryptic (degenerated) prophages or other genetic material acquired by HGT. In fact, this genetic material has significantly shaped microbial evolution due to the development of mutually beneficial interactions between prophage and host [4]. Nevertheless, the safe integration of viral elements into bacterial genomes demands stringent regulation of phage gene expression.

Upon integration into the host genome, a functional phage can exit the prophage state and enter the lytic cycle, which is typically triggered by severe DNA damage that activates the cellular SOS response. Even under non-inducing conditions, cells may encounter spontaneous DNA damage [5], leading to the SOS-dependent induction of prophages in a small fraction of the lysogenic population [6,7].

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

Originally, this spontaneous prophage induction (SPI) was considered a potentially detrimental process, but recent research in the fields of microbial biofilms, host–pathogen interaction and population dynamics emphasizes that SPI is an important contributor to the social behavior of microbes [6]. In a recent study, we quantified SPI in populations of *Corynebacterium glutamicum* using reporter promoter fusions. While we observed a positive correlation between SPI and the DNA damage (SOS) response, a significant fraction of the cases also occurred in an SOS-independent manner [8,9]. Thus, the molecular factors influencing SPI in single individuals and how the host modulates its frequency remain largely unknown.

Compared to point mutations or genomic rearrangements, HGT allows bacteria to acquire new traits much more rapidly, but the downside of this medal is that the new information is encoded on foreign genetic material [10,11]. The encounter with xenogeneic (foreign) DNA is a "high risk–high gain" situation: While the acquisition provides the potential for the fast gain of new beneficial traits, the activity of selfish genetic elements or bacterial viruses (bacteriophages) represents a perpetual threat to bacterial cells. Gene expression from xenogeneic material can strongly impair cellular fitness by sequestering RNA polymerase [12,13], by producing toxic proteins and, in the case of phages, by causing cell lysis. With more than 10^{24} productive viral infections on earth [14], the activity of bacteriophages plays a vital role in HGT, which is also reflected by the variety of phage defense mechanisms encoded in bacterial genomes, with restriction modification (RM) systems and CRISPR–Cas being among the most prominent mechanisms [15]. With the expansion of the phage genomic space, many more examples of phage defense systems have been described and have been covered in a number of recent reviews [16,17]. In contrast to the destructive mode of action of RM and CRISPR–Cas systems, where nucleases are employed to wipe out incoming foreign DNA [18], xenogeneic silencing (XS) represents a mechanism promoting tolerance of foreign genetic material [19,20]. The mechanism of XS is based on the activity of small, nucleoid-associated proteins (NAPs) that recognize and bind foreign, AT-rich DNA stretches and silence gene expression due to the formation of a tight nucleoprotein complex [19] (Fig. 1). By this means, XS proteins provide an important basis for the safe acquisition of new genetic material and foster evolutionary network expansion [21]. Hitherto, all known XS proteins fall into one of four different classes: H-NS in Proteobacteria [19,22], MvaT/U in *Pseudomonas* species [23], Rok in *Bacillus subtilis* [24] and Lsr2 in Actinobacteria [25]. Despite the low sequence similarity between different silencers, these proteins appear to fulfill very similar functions in their respective host, and examples of cross-complementation have been found for Lsr2, MvaT and H-NS [9,23,25].

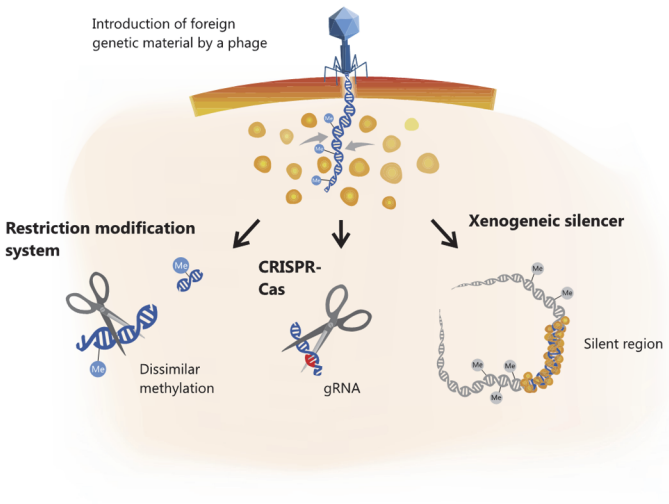


Fig. 1. Xenogeneic silencing of foreign DNA. Microbial cells have evolved a variety of different defense mechanisms to deal with viral DNA and to counteract potential detrimental effects. Schematically included examples are the RM systems and CRISPR–Cas. In contrast, XS proteins are able to recognize foreign, AT-rich DNA and form an oligomeric nucleoprotein complex that silences gene expression at the particular target regions [19].

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

Several recent studies focused, in particular, on the impact of H-NS on bacterial genome evolution and network expansion (for recent reviews, see Refs. [19,20,26]). Given the mosaic-like structure of bacterial genomes, (pro-)phages apparently account for a significant fraction of bacterial strain diversification [1]. In a recent study, we provided the first example of a prophage-encoded Lsr2-like protein functioning as an essential silencer of a cryptic prophage in the actinobacterium *Corynebacterium glutamicum*. Considering the generic role of XS proteins in the silencing of foreign DNA, it is reasonable to assume that these proteins play an important role and may adopt different functions in phage–host regulatory interaction. In this review, we summarize the literature focusing on the role of XS proteins on phage–host interactions, including examples of phage-mediated counter-silencing as a defense strategy during infection. Furthermore, we provide a comprehensive bioinformatics analysis of bacterial and phage genomes revealing that genes encoding Lsr2-like proteins are ubiquitously found in the genomes of actinobacteriophages, suggesting an adoption of XS function in the lifestyles of virulent and temperate phages.

XS: Recognition and Binding to Foreign, AT-rich DNA

The basis for XS is provided by the domain organization of the XS proteins, which is remarkably similar among the different classes. Typically, these proteins exhibit a small size of <15 kDa and consist of an N-terminal oligomerization domain and a C-terminal DNA-binding domain [20,27].

All silencers have the common feature of preferentially binding DNA regions that are more AT rich than the host genome. This feature appears to be the basis for the targeting of foreign elements since the vast majority of exogenous DNA has been found to be more AT-rich than the host core genome [19,28,29].

High-resolution structural analysis of the C-termini of H-NS and Lsr2 resulted in the identification of the "prokaryotic AT-hook" with the "Q/RGR" motif, which is reminiscent of the "AT-hook" motif found in eukaryotic HMG-I(Y) proteins [30]. AT sequences lack the exocyclic 6-amino group, which consequently allows for a deeper interaction with protein side-chain residues. In addition, the narrower minor groove of AT-rich sequences harbors a surface with a higher electronegative potential (than mixed or GC-rich sequences), enabling a stronger interaction with positively charged residues. From the evolutionary point of view, it is astonishing that although H-NS and Lsr2 do not share any structural similarity, these proteins show the same binding mechanism. In contrast, MvaT/U proteins, found in *Pseudomonas* species, lack the AT-hook motif but instead recognize target DNA via a so-called "AT-pincer" motif consisting of a conserved lysine residue and a downstream KGGN motif interacting with the minor groove [31]. Recently, Duan *et al.* [27] elucidated how the *Bacillus* silencer Rok distinguishes between host and foreign DNA. The authors could show that Rok directly binds to the minor groove of AT-rich sequences in a novel mode that, so far, has not been described for any other winged helix protein. Using *in vitro* protein binding microarrays and comparative genome analysis, the authors concluded that Rok preferentially recognizes a few distinct AT-rich DNA motifs present in horizontally acquired regions, which are significantly underrepresented in *Bacillus* host genomes [27]. However, for the other XS proteins, no sequence-specific recognition has been observed. This characteristic has been, most intensively, studied for H-NS, where AT-rich regions were suggested to function as initial nucleation regions [32]. In a more recent study, a single-molecule counting approach revealed that nucleation sites are crucial for recruiting H-NS molecules [33].

XS: Formation of the Nucleoprotein Complexes

However, binding to DNA is itself not sufficient for XS proteins to fulfill their function. Protein multimerization is an essential step that is required to enable silencing of the target regions. Generally, it is assumed that after the initial nucleation, additional silencing molecules are recruited and concurrently spread along adjacent AT-rich regions (filament formation) [32]. Finally, the high-order oligomerization between distal silencer–DNA complexes leads to strong condensation and DNA compaction that enables silencing of target gene expression [19]. In the current literature, three main mechanisms for silencing are suggested: (i) promoter occlusion, where the filamentous nucleoprotein complex prevents the binding of the RNA polymerase [19,34], (ii) trapping of the RNA polymerase, blocking promoter escape upon binding [34] and (iii) ρ-dependent transcriptional termination that occurs during the pausing of RNA polymerase [35].

Multimerization of XS proteins is facilitated by the N-terminal domains. This effect has been shown, experimentally, for H-NS-, MvaT- and Lsr2-like proteins [9,36–38] and was suggested for Rok [24]. Primarily, the mechanism behind this protein oligomerization was widely studied using H-NS and high-resolution approaches, like atomic force microscopy, electron microscopy and single-molecule magnetic tweezers experiments [37,39,40]. Depending on the concentration of the divalent cations (Mg^{2+} or Ca^{2+}), atomic force microscopy studies reported either a bridging or stiffened mode of the bound DNA

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

stretches [40,41]. These two different modes were also found for MvaT and Lsr2 [36,38,42,43], underlining a common silencing mechanism among the silencers and explain why these XS proteins are partly able to complement each other in the above-mentioned cross-complementation experiments [9,25]. Nevertheless, the dissimilar AT compositions of host genomes and the differences in DNA binding underline the specialization of the particular XS protein to the requirements in the particular host background.

For H-NS- and MvaT-like silencers, it was shown that XS proteins may form silencing nucleoprotein complexes by interacting with other NAPs. Efficient protein oligomerization is, however, still a prerequisite for the establishment of a silencing nucleoprotein complex. In particular, the small proteins Hha and its paralog YdgT (Cnu) were shown to structurally resemble the N-terminal domain of H-NS and are therefore capable of forming heteromeric complexes [44]. Further studies based on transcriptomics revealed that the binding of Hha to H-NS polymers is required for the efficient repression of a subset of the H-NS regulon [45]. At this point, it is worthwhile to mention that Cnu represents an *oriC* binding protein that binds within a DnaA binding box and has been suggested to contribute to optimal *oriC* activity [46]. In *B. subtilis*, a regulative dependency between the replication initiation protein DnaA and the silencer Rok was recently reported [47], evincing an interesting link between bacterial replication and this XS protein.

In the genomes of *Escherichia coli* and *Salmonella* Typhimurium, paralogs of H-NS, such as StpA, are encoded and are capable of forming heterodimers that in turn can interact again with the aforementioned small proteins Hha/YdgT [44]. Recently, the impact of Hha and StpA on the binding mode of the H-NS protein was investigated by *in vitro* biochemical and biophysical experiments. With their high-resolution approach, Boudreau *et al.* [48], nicely demonstrated how StpA and Hha modify H-NS-polymeric filaments to increase transcriptional pausing and provided evidence showing that the mixed nucleoprotein complexes (consisting of Hha/H-NS or StpA/H-NS, etc.) differentially affect gene regulation. Thus, these experiments further suggest that interactions with paralogs or accessory proteins likely specify the silencing characteristics also depending on the local concentrations of the respective proteins. The formation of heteromeric silencer complexes was also reported for MvaT and its paralog MvaU in *Pseudomonas aeruginosa* [49]. In *Pseudomonas putida*, MvaT-like proteins were shown to control distinct regulons revealing a functional specialization of these proteins [50–52]. Similar scenarios are also conceivable for Lsr2-like proteins, since the genomes of several Actinobacteria, like *Streptomyces coelicolor* A(3), *Streptomyces venezuelae* ATCC 10712 and *Mycobacterium smegmatis* MC^2 155, encode more than one copy of Lsr2-like proteins (based on a BLAST search with the sequence of Lsr2 (Rv3597c) and default parameters, *e*-value $< 1 \times 1010^{-5}$).

The genome compaction of XS proteins resembles the compaction by heteromeric nucleoprotein complexes formed by histones and DNA in eukaryotic cells. Hence, it is reasonable to assume that XS proteins, or in general NAPs, are targets of posttranslational modification (PTM) enzymes. In a very recent review, this topic was elucidated by Dilweg and Dame [53]. Interestingly, the authors reported approximately 29 PTMs for H-NS from *E. coli*; these PTMs were not experimentally investigated, but the physiological implication for DNA condensation and/or silencing was discussed [53].

Silencing of Prophages in Bacteria

The safe integration of viral elements into bacterial genomes demands stringent regulation of expression from highly active phage promoters to avoid the production of potentially toxic proteins. The function of XS proteins thereby provides a basis for integrating foreign genes into host regulatory networks by XS and counter-silencing (the latter is discussed in the next section).

For silencers of all four groups, an influence on the regulation of phage genes has been observed (Table 1). However, this effect is often noted only incidentally in these studies [9,54,56,57]. Interestingly, the essentiality of a particular XS-encoding gene strongly depends on the genetic setup of the particular strain as the induction of mobile genetic elements (MGEs), such as prophages, may lead to cell death. In the case of the Gram-positive actinomycete *C. glutamicum*, the Lsr2-like protein CgpS (*C. glutamicum* prophage silencer) was shown to be essential due to its function as a silencer of the cryptic but still inducible prophage CGP3 [9]. Counteracting of CgpS activity led to prophage induction and regional (in situ) replication at the CGP3 locus [58]. In contrast to other XS proteins described so far, CgpS is encoded on the CGP3 prophage itself and appears to act mainly as a silencer of CGP3 gene expression. Genome-wide profiling of CgpS confirmed the binding of this XS protein to AT-rich sequences in the CGP3 element. As a matter of fact, the essentiality of the *cgpS* gene is linked to the presence of CGP3, and *C. glutamicum* strains lacking the prophage do not require CgpS. Interestingly, integration of a second genomic copy of *cgpS* significantly reduced the fraction of spontaneously induced cells, emphasizing the role of XS proteins in the modulation of SPI in bacterial populations (Frunzke and Pfeifer, unpublished). In addition, for *Mycobacterium tuberculosis*,

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

Table 1. Silencing of phage elements in bacterial genomes

Type of silencer	Host strain	GC of host (%)	Prophage-like element	Length (kb) of phage	GC (%) of phage	Reference
H-NS	*E. coli* K-12 BW25113	50.8	Rac (cryptic)	23.1	47.1	[54]
	S. onediensis MR-1	45.9	CP4So (cryptic)	36	43	[55]
MvaT	*P. aeruginosa* PAO1	66.6	Filamentous phage Pf4	15.7	58.7	[56,57]
Rok	*B. subtilis* 168	43.5	Prophage region 4	8	35.8	[24]
			Prophage region 5	20.7	37.5	
			Prophage region 6	34.8	36.1	
			SPβ	134.4	34.6	
Lsr2	*C. glutamicum* ATCC 13032	53.8	Cryptic prophage CGP1	13.5	47.1	[9]
			Cryptic prophage CGP3	186.0	48.4	
	M. tuberculosis H37Rv	65.6	Prophage region 1, Rv1573-1588c, (Rv1582c)	10.5 (1.4[a])	66.2 (62.5[a])	[38]
			Prophage region 2, Rv2645-2664, (Rv2658-2659c)	12.3 (1.5[a])	66.2 (63.5[a])	

[a] In case of *M. tuberculosis*, also the bound genes were considered and are indicated in brackets.

which is also a member of the Actinobacteria, genome-wide binding studies conducted with a Lsr2-like protein not only confirmed the binding to prophage regions (and other MGEs) [9,38] but also revealed several additional targets in the host genome involved in virulence and immunogenicity [38]. Interestingly, in *B. subtilis*, the Rok protein was also shown to bind prophage genes and to be involved in the control of phage gene expression, as is the case for genes of the prophage SPβ [24,59]. A further interesting example involves *E. coli* H-NS, where a link between H-NS activity and enhanced biofilm formation was recently demonstrated. Here, H-NS was shown to repress the cryptic prophage Rac. Derepression of Rac resulting from *hns* deletion led to prophage induction and cell lysis in a toxin-dependent manner [54]. Moreover, in *Shewanella oneidensis*, H-NS was also reported to be involved in prophage induction during cold adaptation [55]. In addition, the H-NS orthologs MvaT and MvaU were shown to be essential due to the silencing of prophage elements in *P. aeruginosa* strains [56,57]. The depletion of silencers caused increased phage gene expression, the formation of infectious phage particles and cell lysis. Remarkably, only mutants impaired in phage production were capable of compensating the double deletion of *mvaT* and *mvaU* [56,57]. An overview of the studies showing the influence of XS proteins on the control of phage gene expression is provided in Table 1.

How to Overcome Silencing?

The formation of a nucleoprotein complex nucleating at AT-rich regions is a prerequisite for XS. Different studies focusing on counter-silencing mechanisms in various species have revealed that upon activation of gene expression, XS proteins are not released from their target DNA; instead, remodeling of the XS–DNA complex enables RNA polymerase to bind and activate transcription. Different counter-silencing mechanisms are mainly based on other proteins binding in the upstream promoter region, thereby counteracting XS silencing. For instance, this mechanism has been nicely demonstrated by synthetic counter-silencing approaches, where operator sequences of specific transcription factors (TFs) were inserted in the upstream promoter region to counter-silence gene expression upon binding of the particular TF [60]. Several further studies demonstrated that different host-encoded TFs have been coopted—in the course of evolution—to act as counter-silencers. Examples include the response regulator PhoP, an essential activator of *Salmonella* virulence [21], the AraC-family TF ToxT of *Vibrio cholerae* [61], LeuO from *Salmonella enterica* [62], and the two MarR-type regulators RovA and SlyA of *S. enterica* and *Yersinia pseudotuberculosis*, which were shown to antagonize H-NS-dependent silencing of horizontally acquired genes [63,64]. In a recent study, we also could show that the MarR-type regulator MalR of *C. glutamicum*, which controls genes involved in stress-responsive cell envelope remodeling, binds to several regions within the CGP3 prophage and is able to counteract SOS-dependent prophage induction (manuscript submitted, BIORXIV/2019/544056).

An alternative route for counter-silencing lies in the interference between XS proteins belonging to the same protein family. An interesting example has been provided for the unusual H-NS paralog Ler, which functions as a regulator of pathogenicity islands (locus of enterocyte effacement, LEE) in enteropathogenic (EPEC) and enterohemorrhagic *E. coli* strains [65–68]. Structural analysis emphasized that its function as a counter-silencer lies in differences in protein oligomerization as both H-NS and Ler bind to AT-rich regions [69]. Ler shows two different modes of DNA interaction: At low concentrations, Ler

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

is able to increase DNA folding and wraps DNA; otherwise, with increasing concentration, Ler binds DNA in an unwrapped mode where Ler increases the rigidity of DNA similarly to the nucleoprotein filament formed by H-NS [69]. At these high concentrations, Ler displaces H-NS from the bound DNA and therefore overcomes the silencing of target regions. A further interesting example is provided with H-NST, a truncated derivative of H-NS lacking the DNA-binding domain. This XS protein was found to antagonize H-NS in EPEC and uropathogenic *E. coli* by interfering with its oligomerization domain [70,71]. Remarkably, this principle of silencer interference can be harnessed to study the function of essential XS proteins. Overproduction of the N-terminal oligomerization domain of the Lsr2-type silencer CgpS was used to counteract CgpS silencing *in vivo* in the Actinobacterium *C. glutamicum* [9]. Interference between XS proteins was further demonstrated in this study as the expression of other mycobacterial Lsr2 genes as well as introduction of *E. coli* H-NS led to XS interference at AT-rich regions, resulting in prophage induction. These findings are also supported by a bioinformatics analysis showing that different classes of silencers do not occur in the same species [72].

Altogether, these examples provide important insights how silencing and counter-silencing facilitate the expansion of regulatory networks in bacteria [21]. In the following, we will focus on mechanisms employed by phages to counteract XS proteins in the ongoing arms race between the phage and the host. A few studies, discussed below, already suggest a variety of different mechanisms used by phages to gain control. One example is provided by the 5.5 protein of the *E. coli* phage T7, which is able to antagonize H-NS function upon phage infection [73]. By interfering with the central oligomerization domain of H-NS [37], the 5.5 protein blocks H-NS from forming high-order oligomers, leading to counter-silencing of H-NS-silenced genes [74]. Another example has been reported with the Mip protein (*M*vaT *i*nhibiting *p*rotein), encoded by the LUZ24 phage of *P. aeruginosa* [75]. In 2015, Wagemans *et al.* showed that Mip is able to inhibit the binding of the nucleoid-structuring silencer MvaT to DNA and that Mip and MvaT coprecipitate in pulldown assays. However, the exact mechanism of MvaT inhibition by Mip is not completely understood, yet.

In the case of *E. coli* T4 phage, two different proteins were reported to interfere with H-NS silencing. The protein MotB is a DNA-binding protein that co-purifies with H-NS as well as with the H-NS homolog StpA. Deletion of the *motB* gene led to a decreased burst size [76]. The T4 protein Arn represents an interesting example of a phage-encoded DNA mimic protein and was shown to directly interact with *E. coli* H-NS [77]. While Arn was originally described as an inhibitor of the McrBC restriction enzyme, structural analysis revealed that the shape of the protein mimics the shape and charge of double-stranded DNA, and the authors highlight this DNA mimicry as a mechanistic basis for interfering with the function of DNA-binding proteins, like H-NS [77]. Interestingly, the DNA mimic proteins Ocr of the phage T7 and ArdA of the plasmid CollB-P9 were also reported to antagonize H-NS in a similar way [78]. An overview on described counteracting proteins is provided by Table 2.

Finally, the direct interference with TFs or other proteins likely does not represent the only way to fight off XS. In a very recent study, Kronheim *et al.* [79] highlighted the important role of small molecules secreted by bacterial hosts as weapons against phage infection. A link between these compounds and XS proteins does not necessarily exist, but a few examples suggest that small natural compounds—other than proteins—may also counteract XS. One class of compounds is represented by polyamides containing a biaryl motif. These polyamides especially target the minor groove of AT-rich DNA sequences and manipulate their topology [80]. In their study, Brucoli *et al.* therefore suggest an effect of these compounds on XS. A further example is the antiasthma drug zafirlukast, which was shown to inhibit the DNA-binding ability of Lsr2 in *M. tuberculosis* and *M. smegmatis* [81]. This compound was found to inhibit the growth of both mycobacterial strains and led to clarification of the bacterial cultures after three days. A direct interaction between zafirlukast and Lsr2 was revealed by MALDI-TOF analysis. However, we suggest that such interactions are specific for the particular protein since, in our hands, zafirlukast does not counteract the silencing mediated by the Lsr2-like protein CgpS (unpublished data).

Silencers in Actinobacteriophages

With the Lsr2-like silencer CgpS, we recently provided the first example of a prophage-encoded XS protein [9]. We showed that this protein is crucial for silencing of phage gene expression to maintain the lysogenic state of the large cryptic prophage CGP3 on which it is encoded. Thus far, in the current literature, only one publication, which was based on metagenomics, has reported the presence of an H-NS-like gene in a phage genome [84]. To evaluate how commonly XS-encoding genes are found in phage genomes, we screened phage databases for these genes. Using the actinobacteriophage database PhagesDB [85], we obtained >300 hits (blastp, *e*-value < 0.005) for Lsr2-like proteins. No phages encoding the other types of silencers, H-NS, MvaT or Rok, were found in the genomes of phages infecting actinobacteria, which is not surprising, as members of this bacterial class harbor only Lsr2-like proteins. Thus, we extended the screening to the Virus-Host

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

Table 2. Counteracting XS by phages or other MGEs

Counter-actor	Mechanism	Host	Host XS	Phage or MGE	Source
H-NST	H-NST represents a truncated version of H-NS consisting of the oligomerization domain. H-NST interferes with the correct oligomerization of the native H-NS protein and therefore the correct function.	*E. coli*; uro-pathogenic strain CFT073 and entero-pathogenic strain E2348/69	H-NS	CFT073: UPEC-specific island inserted at *serU*, E2348/69: EPEC-specific island at *asnW*	[70]
T7 protein 5.5	Protein 5.5 is able to interact with the central oligomerization domain and hinders H-NS from forming high-order oligomers.	*E. coli*, e.g., BL21 (DE3)	H-NS	Phage T7 (virulent)	[73,74]
Ler	Ler is a DNA-binding H-NS homologue that is able to increase DNA rigidity similar to the nucleoprotein filament formed by H-NS. This binding leads to a displacement of H-NS and counter-silencing.	*E. coli* (EPEC strain E2348/69)	H-NS	Horizontally acquired pathogenicity island (*LEE1*)	[65–67,69]
Mip	The MvaT-inhibiting protein (Mip) coprecipitates with MvaT and was shown to inhibit the DNA-binding of MvaT.	*P. aeruginosa* PAO1	MvaT	Phage LUZ24 (virulent)	[75]
MotB	MotB copurifies with H-NS and StpA. Deletion of the *motB* gene leads to decreased burst size of T4. Hence, a counter-acting ability against H-NS was suggested.	*E. coli*	H-NS	Phage T4 (virulent)	[76]
Arn	Arn acts as DNA mimicking protein (mimicking the charge and structure of dsDNA) and is able to bind H-NS. Arn binding could be shown to interfere with the binding of H-NS to target regions.	*E. coli*, e.g., BL21 (DE3)	H-NS	Phage T4 (virulent)	[77,78]
Ocr	Ocr is a DNA mimicking protein that mimics B-form DNA. It was shown that Ocr is able to counter-silence H-NS-silenced promoters.	*E. coli* (*i.e.,* C600)	H-NS	Phage T7 (virulent)	[78,82]
ArdA	The crystallization of the ArdA dimer led to the assumption that ArdA is a DNA mimic proteins. Furthermore, it could be shown that increased amounts of ArdA leads to a counter-silencing of H-NS silenced promoters *in vivo*.	Multiple organisms	H-NS	Plasmid CoIIb-P9	[78,83]

database (>2500 bacteriophages) [86]. Strikingly, we could not obtain a single hit for H-NS-, MvaT- or Rok-like proteins in >1000 Proteobacterium phages and in >600 phages that infect Firmicutes. These findings illustrate that the function of Lsr2-like proteins has clearly been adopted by actinobacteriophages and that genes are commonly transferred by phages between GC-rich Actinobacteria. In an initial evaluation, we used PHACTS [87] to allocate the >2600 actinophages into temperate (>800) and virulent phages (>1800) and found Lsr2-like proteins in both groups (Fig. 2a, Table S1). Taking into account the respective group sizes (virulent > temperate), Lsr2-like proteins are approximately three times more frequent in temperate than virulent phages (Fig. 2a). An overview of their respective hosts evinced four different genera, *Gordonia*, *Microbacterium*, *Mycobacterium* and *Streptomyces*, in which the mycobacteriophages represent the largest group of Lsr2-encoding phages with 141 members (Fig. 2b). However, mycobacteriophages are strongly overrepresented in PhagesDB (>1600). When considering the overall group sizes of the respective hosts, *lsr2* genes are most likely to be found in genomes of *Streptomyces* phages (18.6%; 32 of 172) (Fig. 2b). Remarkably, comparisons of the GC contents and the genome sizes of Lsr2-encoding phages showed that the genomes were significantly more AT-rich and larger, especially for phages with a putative temperate lifestyle (Fig. 2c). In addition, within the Lsr2 group, genomes of virulent phages exhibit a higher variation with respect to GC content and genome size (Fig. 2c). Furthermore, we performed secondary structure predictions to compare phage- and host-encoded silencer proteins (within the respective group) by their global pairwise identity. While bacterial Lsr2 proteins are highly conserved, we identified a strong variability in terms of the predicted secondary structure within phage-encoded Lsr2-like proteins (Fig. 2d). Taken together, these findings suggest that phage-encoded Lsr2-like proteins have different functions, presumably based on the phage lifestyle.

A hypothesis derived from stealth plasmids (discussed in the section below) and emphasized by the example of CgpS [9] is that XS proteins are involved in maintaining the lysogenic state, thereby presenting a mechanism of mutual adaptation and

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

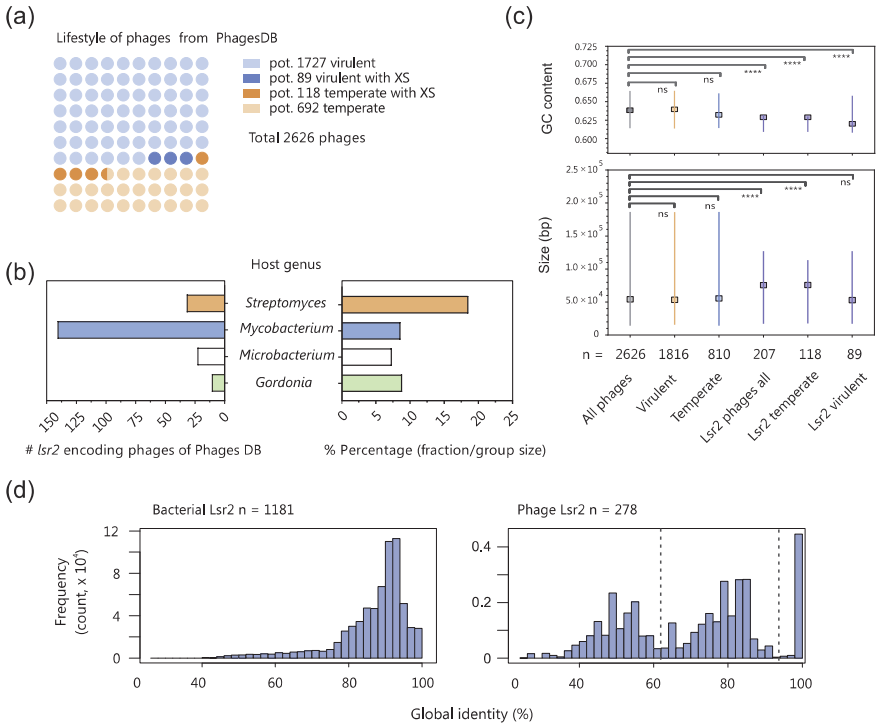


Fig. 2. Lsr2-like proteins are encoded on actinobacteriophage genomes. (a) Distribution of Lsr2-encoding phages is shown among temperate and virulent actinophages. Based on the genomes downloaded from the actinobacteriophage database PhagesDB [85], coding sequences were predicted by prodigal [88] and the lifestyles of phages were predicted using PHACTS [87]. The temperate lifestyle was assigned if the mean minus the standard deviation of the calculated probability was >0.5. Otherwise, a virulent lifestyle was assumed. By this approach, 1816 (1727 corresponding to light blue and 89 corresponding to blue balls) were predicted to be virulent and 810 (692 light orange and 118 corresponding to orange balls) to be temperate (out of 2626 phages, downloaded 19.10.2018). A blastp search (default parameters, *e*-value < 0.005) revealed 207 phages encoding Lsr2-like proteins, of which 89 (blue balls) are virulent phages and 118 are temperate phages (orange balls). Phages containing more than one gene encoding an Lsr2-like protein were counted only once; hits found in draft genomes were excluded. (b) Overview of the host genus of Lsr2-encoding phages. On the left side, the absolute numbers are indicated. The right side of plot shows the proportion of Lsr2-encoding phages among all phages for the respective host genus. (c) GC contents and sizes of temperate, virulent and silencer encoding phages were compared with reference group (all phages) by the Kruskal–Wallis test. The medians are indicated with boxes. In the GC-content plot, the lines represent the interquartile ranges, whereas in the size plot, the lines indicate the range of the minimal and maximal values. (d) Global pairwise secondary structure identity between Lsr2 sequences encoded in bacterial (left side) and phage genomes (right side). The identity of Lsr2 structure within bacterial genomes is relative high (mean ~87%) compared to the phage encoded Lsr2 sequences (mean ~70%). Furthermore, the distribution of the phage encoded Lsr2 structure identity evinces a higher diversity by pointing to the existence of particular pairs with high identity to each other and low to the rest of the data set.

integration into host regulatory networks (Fig. 3). In this scenario, XS proteins would work in conjunction with the endogenous phage repression system, stabilizing lysogeny and minimizing the costs for harboring the temperate phage. Furthermore, it is conceivable that XS proteins may influence the lytic-lysogenic decision during phage infection, but this has not yet been addressed experimentally. In virulent phages, silencer proteins may inherit completely different roles. In addition to affecting multiple other targets, XS proteins were also found to repress genes encoding different phage defense

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

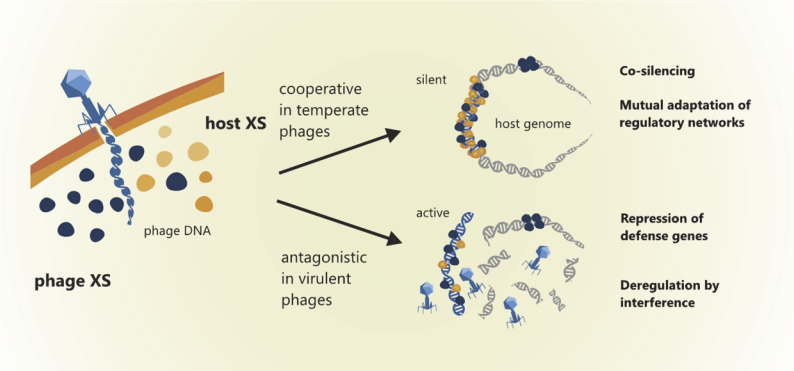


Fig. 3. Model for the functions of phage-encoded silencers depending on the phage lifestyle. Examples like the Lsr2-like silencer CgpS demonstrated that XS proteins may play an important role in maintaining the lysogenic state. Depending on the XS repertoire or the particular host strain, (pro-)phage-encoded XS proteins may also cooperate with the host-encoded protein(s) to form heteromeric complexes. The function of phage-encoded silencers has not been studied experimentally and therefore remains subject to speculation. Nevertheless, it can be postulated that virulent phages might employ XS-like proteins as a weapon to interfere with host XS proteins or to repress other host defense mechanisms.

systems, including CRISPR–Cas genes [89] and an RM system [9]. Furthermore, interference of XS proteins, which are not capable of forming functional heteromeric complexes, may lead to a deregulation of phage genes and counteracting the silencing of phage genes by the host XS. We tested this hypothesis in our previous study where expression of *E. coli hns* as well as other genes encoding mycobacterial Lsr2 proteins led to activation of the CgpS-silenced prophage [9]. Hence, we suggest that virulent phages might employ XS homologs as a weapon to interfere with host defense systems (Fig. 3).

In addition, XS proteins or, in general, NAPs are also used to organize and structure the genome during replication cycles. Hence, it is likely that phages, especially with "larger" genomes, will benefit from efficient DNA packaging proteins that facilitate optimized phage replication and production. Moreover, it is conceivable that phage-encoded silencers may contribute to compaction of the host genome. Here, an analogous example is given by the eukaryotic dinoflagellates. It is suggested that virus-like proteins, termed dinoflagellate/viral nucleoproteins (DVNPs), fulfill the functions of histones by packaging the genomes. This is supported by a recent study in which it was shown that heterologous produced DVNPs outcompete histones in *Saccharomyces cerevisiae* thereby causing toxic effects [90]. Although the dinoflagellate genomes encode histone proteins, the histone expression is strongly reduced, and it is assumed that they provide only regulatory functions [91]. Moreover, based on phylogenetics, it is hypothesized that the histone depletion occurred simultaneously with the acquisition of DVNPs from large-genome viruses (genome size up to 560 kb) and with massive genome expansion [92].

Stealth Silencers Encoded on Plasmids

Usually, the introduction of new MGEs, such as plasmids, imposes high fitness costs on the host organism [93]. The magnitude of the costs depends on many factors, such as plasmid-specific characteristics (replication, plasmid reception, integration and conjugation, encoded traits), expression level of plasmid-borne genes and the genetic background of the host organism [93]. Interestingly, it is assumed that the main burden arises from the expression of plasmid-encoded genes that comes from transcription, translation, or the interactions between plasmid- and host-encoded proteins [94]. One way to reduce the cost is to use "stealth genes," particularly genes encoding silencer proteins. The H-NS-like protein Sfh was one of the first characterized plasmid encoded stealth proteins. Doyle *et al.* [95] investigated the costs of the AT-rich pSf-R27 plasmid and evinced a significant biosynthetic burden in the absence of *sfh*. Therefore, the authors concluded that these proteins are quite useful in infiltrating a new host by reducing the metabolic burden to a minimum. Strikingly, transcriptome and genome-wide binding analysis revealed that the regulons from the chromosomally and plasmid-encoded H-

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

NS-like silencers are completely different [96–98], although the proteins are closely related. While plasmid-encoded variants typically exhibit a specific and narrow target spectrum (mostly with a focus on HGT-acquired regions), the host H-NS is known to act as a global regulator modulating the expression of both HGT-acquired and core genes [97,98]. The basis for this selectivity was addressed by chimeric protein fusions, and the experimental results suggest a correlation between higher flexibility of the linker-domain (connecting the N-terminal oligomerization part to the C-terminal DNA-binding domain) and a decrease in selectivity due to stable binding to broader ranges of DNA geometries [99]. Astonishingly, in a more recent study, a known H-NS target, the *gadAB* operon, was examined in *Shigella flexneri*. Here, reduced expression was observed in the presence of the pSf-R27 plasmid, leading to reduced acid resistance and showing that a plasmid- and host-encoded silencer also co-regulate core genes [100].

Bioinformatics analyses revealed that genes encoding H-NS- and MvaT-like silencers are overrepresented on large plasmids. In addition, H-NS-encoding plasmids are found to be more AT-rich than other NAP-harboring plasmids [101]. To also evaluate the distribution of Lsr2- and Rok-like silencers, we performed a BLASTp search (*e*-value < 0.005) using a plasmid database that we retrieved from the NCBI nucleotide database (>24000 plasmid sequences, source database RefSeq, Fig. 4a). Here, we identified 408 hits for H-NS-like proteins, 35 for MvaT-like proteins, 63 for Lsr2-like proteins and 18 for Rok-like proteins. A comparison of the average sizes revealed that silencer-encoding plasmids are at least five times larger than the average plasmid sequence deposited in the database (Fig. 4a). Furthermore, H-NS-encoding *E. coli* plasmids displayed a significantly higher AT content (96 of the 408 H-NS plasmids) (Fig. 4b). This trend was, however, not observed for GC-rich *Streptomycetes* (Fig. 4b) or *P. aeruginosa* and *B. subtilis* plasmids (data not shown), but the sequence data in these databases feature a strong bias, and only a few sequences are available for some species (e.g., n = 13 for *Streptomycetes* plasmids). In line with previous studies, our findings emphasize that especially large plasmids appear to benefit from DNA-organizing proteins and that self-silencing may represent a strategy for host infiltration harnessed by phages and plasmids alike.

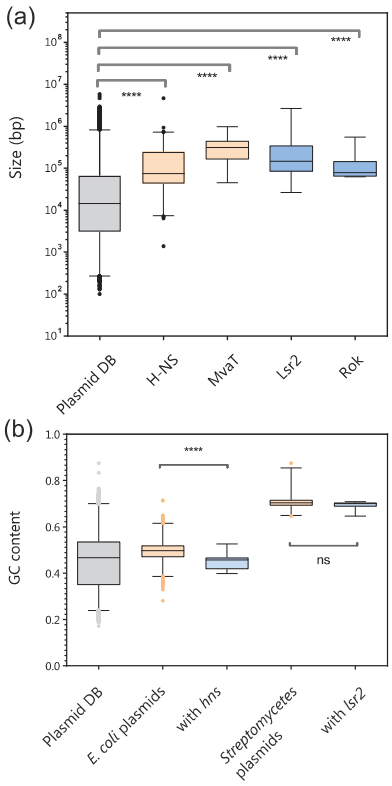


Fig. 4. Distribution of silencer-encoding genes on plasmids. (a) Overall, 24197 sequences for plasmids were retrieved from the NCBI nucleotide database using RefSeq as the source data base (filter criteria: bacteria, genomic DNA, plasmid, RefSeq on 29.10.2018). Via a local blastp search (*e*-value < 0.005) with the amino acid sequences of H-NS (WP_001287378.1), MvaT (WP_003093888.1), Lsr2 (WP_003419513.1) and Rok (WP_003232378.1), approximately 408, 35, 63 and 18 hits, respectively, were found. The sizes were compared in a boxplot with ranges from 1% to 99% and evaluated by Kruskal–Wallis tests. (b) GC content of the plasmids, including all sequences, *E. coli* plasmids (n = 2600), with *hns* (n = 95) and *Streptomycetes* plasmids (n = 147) with *lsr2* (13) were compared in a boxplot (range, 1%–99%; evaluated by Kruskal–Wallis tests).

Future Perspectives

In conclusion, several recent studies highlight the important role of XS proteins in phage–host interactions, for example, by silencing expression of genomically integrated prophages or phage remnants. The first examples, like the T7 5.5 protein or Mip [74,75] encoded by a *Pseudomonas* phage, demonstrate that counteracting XS represents an important aspect of lytic infection. As illustrated by the examples of Lsr2-like proteins encoded by various actinophages, the function of XS proteins apparently has been adopted by phages as well.

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

Nevertheless, many gaps remain in the prokaryotic and phage sequence space as for the majority of prokaryotic phyla, no XS protein has been identified so far. Furthermore, it is striking that while homologs of H-NS, MvaT, Rok and Lsr2 are encoded by plasmids, only *lsr2* homologs were identified in phage genomes. Considering the high sequence variability of phage-encoded silencers, these proteins most likely perform many different functions depending on the particular lifestyle of the phage, which needs to be addressed in future studies.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.02.011.

Acknowledgments

For financial support, we thank the German Research Foundation (SPP 1617, Grant FR2759/2-2) and the European Research Council (ERC-StG-2017, Grant 757563).

We thank Larissa Kever (Institute for Bio- and Geosciences 1, IBG-1) for support in graphic design of the figures and Anna Schumann for the assistance in data management.

Received 19 December 2018;
Received in revised form 7 February 2019;
Accepted 11 February 2019
Available online xxxx

Keywords:
xenogeneic silencer;
H-NS;
Lsr2;
Actinobacteria;
phage–host interaction

Abbreviations used:
HGT, horizontal gene transfer; SPI, spontaneous prophage induction; RM, restriction modification; XS, xenogeneic silencing; NAP, nucleoid-associated protein; PTM, posttranslational modification; MGE, mobile genetic element; TF, transcription factor; EPEC, enteropathogenic; DVNP, dinoflagellate/viral nucleoprotein.

References

[1] S. Casjens, Prophages and bacterial genomics: what have we learned so far? Mol. Microbiol. 49 (2003) 277–300.
[2] C. Canchaya, G. Fournous, H. Brüssow, The impact of prophages on bacterial chromosomes, Mol. Microbiol. 53 (2004) 9–18.
[3] G.F. Hatfull, R.W. Hendrix, Bacteriophages and their genomes, Curr. Opin. Virol. 1 (2011) 298–303.
[4] J. Bondy-Denomy, A.R. Davidson, When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness, J. Microbiol. 52 (2014) 235–242.
[5] J.M. Pennington, S.M. Rosenberg, Spontaneous DNA breakage in single living *Escherichia coli* cells, Nat. Genet. 39 (2007) 797–802.
[6] A.M. Nanda, K. Thormann, J. Frunzke, Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions, J. Bacteriol. 197 (2015) 410–419.
[7] A. Lwoff, Lysogeny, Bacteriol. Rev. 17 (1953) 269–337.
[8] A.M. Nanda, A. Heyer, C. Kramer, A. Grunberger, D. Kohlheyer, J. Frunzke, Analysis of SOS-induced spontaneous prophage induction in *Corynebacterium glutamicum* at the single-cell level, J. Bacteriol. 196 (2014) 180–188.
[9] E. Pfeifer, M. Hünnefeld, O. Popa, T. Polen, D. Kohlheyer, M. Baumgart, et al., Silencing of cryptic prophages in *Corynebacterium glutamicum*, Nucleic Acids Res. 44 (2016) 10117–10131.
[10] F. de la Cruz, J. Davies, Horizontal gene transfer and the origin of species: lessons from bacteria, Trends Microbiol. 8 (2000) 128–133.
[11] H. Ochman, J.G. Lawrence, E.A. Groisman, Lateral gene transfer and the nature of bacterial innovation, Nature. 405 (2000) 299–304.
[12] E. Darmon, D.R.F. Leach, Bacterial genome instability, Microbiol. Mol. Biol. Rev. 78 (2014) 1–39.
[13] L.E. Lamberte, G. Baniulyte, S.S. Singh, A.M. Stringer, R.P. Bonocora, M. Stracy, et al., Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase, Nat. Microbiol. 2 (2017) 16249.
[14] F. Rohwer, M. Youle, H. Maughan, N. Hisakawa, L.L. Pantéa, B. Darby, Life in Our Phage World: A Centennial Field Guide to the Earth's Most Diverse Inhabitants, 2015.
[15] S.J. Labrie, J.E. Samson, S. Moineau, Bacteriophage resistance mechanisms, Nat. Rev. Microbiol. 8 (2010) 317–327.
[16] S. Doron, S. Melamed, G. Ofir, A. Leavitt, A. Lopatina, M. Keren, et al., Systematic discovery of antiphage defense systems in the microbial pangenome, Science 359 (2018).
[17] A. Stern, R. Sorek, The phage–host arms race: shaping the evolution of microbes, BioEssays 33 (2011) 43–51.
[18] M.E. Dupuis, M. Villion, A.H. Magadan, S. Moineau, CRISPR–Cas and restriction–modification systems are compatible and increase phage resistance, Nat. Commun. 4 (2013) 2087.
[19] W.W. Navarre, The impact of gene silencing on horizontal gene transfer and bacterial evolution, Adv. Microb. Physiol. 69 (2016) 157–186.
[20] K. Singh, J.N. Milstein, W.W. Navarre, Xenogeneic silencing and its impact on bacterial genomes, Annu. Rev. Microbiol. 70 (2016) 199–213.
[21] W.R. Will, D.H. Bale, P.J. Reid, S.J. Libby, F.C. Fang, Evolutionary expansion of a regulatory network by counter-silencing, Nat. Commun. 5 (2014) 5270.
[22] W.W. Navarre, M. McClelland, S.J. Libby, F.C. Fang, Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA, Genes Dev. 21 (2007) 1456–1471.
[23] C. Tendeng, O.A. Soutourina, A. Danchin, P.N. Bertin, MvaT proteins in *Pseudomonas* spp.: a novel class of H-NS-like proteins, Microbiology 149 (2003) 3047–3050.
[24] W.K. Smits, A.D. Grossman, The transcriptional regulator Rok binds A + T-rich DNA and is involved in repression of a mobile genetic element in *Bacillus subtilis*, PLoS Genet. 6 (2010), e1001207.

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

[25] B.R. Gordon, R. Imperial, L. Wang, W.W. Navarre, J. Liu, Lsr2 of *Mycobacterium* represents a novel class of H-NS-like proteins, J. Bacteriol. 190 (2008) 7052–7059.
[26] C.J. Dorman, H-NS-like nucleoid-associated proteins, mobile genetic elements and horizontal gene transfer in bacteria, Plasmid 75 (2014) 1–11.
[27] B. Duan, P. Ding, T.R. Hughes, W.W. Navarre, J. Liu, B. Xia, How bacterial xenogeneic silencer rok distinguishes foreign from self DNA in its resident genome, Nucleic Acids Res. 46 (2018) 10514–10529.
[28] E.P. Rocha, A. Danchin, Base composition bias might result from competition for metabolic resources, Trends Genet. 18 (2002) 291–294.
[29] H. Nishida, Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids, Int. J. Evol. Biol. 2012 (2012) 342482.
[30] S.S. Ali, B. Xia, J. Liu, W.W. Navarre, Silencing of foreign DNA in bacteria, Curr. Opin. Microbiol. 15 (2012) 175–181.
[31] P. Ding, K.A. McFarland, S. Jin, G. Tong, B. Duan, A. Yang, et al., A novel AT-Rich DNA recognition mechanism for bacterial xenogeneic silencer MvaT, PLoS Pathog. 11 (2015), e1004967.
[32] B. Lang, N. Blot, E. Bouffartigues, M. Buckle, M. Geertz, C. O. Gualerzi, et al., High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes, Nucleic Acids Res. 35 (2007) 6330–6337.
[33] R. Gulvady, Y. Gao, L.J. Kenney, J. Yan, A single molecule analysis of H-NS uncouples DNA binding affinity from DNA specificity, Nucleic Acids Res. 46 (2018) 10216–10224.
[34] D.C. Grainger, Structure and function of bacterial H-NS protein, Biochem. Soc. Trans. 44 (2016) 1561–1569.
[35] M.V. Kotlajich, D.R. Hron, B.A. Boudreau, Z. Sun, Y.L. Lyubchenko, R. Landick, Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria, Elife. 4 (2015).
[36] R.S. Winardhi, W. Fu, S. Castang, Y. Li, S.L. Dove, J. Yan, Higher order oligomerization is required for H-NS family member MvaT to form gene-silencing nucleoprotein filament, Nucleic Acids Res. 40 (2012) 8942–8952.
[37] S.T. Arold, P.G. Leonard, G.N. Parkinson, J.E. Ladbury, H-NS forms a superhelical protein scaffold for DNA condensation, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 15728–15732.
[38] B.R. Gordon, Y. Li, L. Wang, A. Sintsova, H. van Bakel, S. Tian, et al., Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 5154–5159.
[39] R.S. Winardhi, J. Yan, L.J. Kenney, H-NS regulates gene expression and compacts the nucleoid: insights from single-molecule experiments, Biophys. J. 109 (2015) 1321–1329.
[40] W.R. Will, P.J. Whitham, P.J. Reid, F.C. Fang, Modulation of H-NS transcriptional silencing by magnesium, Nucleic Acids Res. 46 (2018) 5717–5725.
[41] Y. Liu, H. Chen, L.J. Kenney, J. Yan, A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes, Genes Dev. 24 (2010) 339–344.
[42] R.T. Dame, M.S. Luijsterburg, E. Krin, P.N. Bertin, R. Wagner, G.J. Wuite, DNA bridging: a property shared among H-NS-like proteins, J. Bacteriol. 187 (2005) 1845–1848.
[43] Y. Qu, C.J. Lim, Y.R. Whang, J. Liu, J. Yan, Mechanism of DNA organization by *Mycobacterium tuberculosis* protein Lsr2, Nucleic Acids Res. 41 (2013) 5263–5272.
[44] S. Paytubi, C. Madrid, N. Forns, J.M. Nieto, C. Balsalobre, B.E. Uhlin, et al., YdgT, the Hha paralogue in *Escherichia coli*, forms heteromeric complexes with H-NS and StpA, Mol. Microbiol. 54 (2004) 251–263.
[45] T. Ueda, H. Takahashi, E. Uyar, S. Ishikawa, N. Ogasawara, T. Oshima, Functions of the Hha and YdgT proteins in transcriptional silencing by the nucleoid proteins, H-NS and StpA, in *Escherichia coli*, DNA Res. 20 (2013) 263–271.
[46] M.S. Kim, S.H. Bae, S.H. Yun, H.J. Lee, S.C. Ji, J.H. Lee, et al., Cnu, a novel *oriC*-binding protein of *Escherichia coli*, J. Bacteriol. 187 (2005) 6998–7008.
[47] C.A. Seid, J.L. Smith, A.D. Grossman, Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in *Bacillus subtilis*, Mol. Microbiol. 103 (2017) 798–817.
[48] B.A. Boudreau, D.R. Hron, L. Qin, R.A. van der Valk, M.V. Kotlajich, R.T. Dame, et al., StpA and Hha stimulate pausing by RNA polymerase by promoting DNA–DNA bridging of H-NS filaments, Nucleic Acids Res. 46 (2018) 5525–5546.
[49] S. Castang, S.L. Dove, High-order oligomerization is required for the function of the H-NS family member MvaT in *Pseudomonas aeruginosa*, Mol. Microbiol. 78 (2010) 916–931.
[50] Z. Sun, D. Vasileva, C. Suzuki-Minakuchi, K. Okada, F. Luo, Y. Igarashi, et al., Differential protein–protein binding affinities of H-NS family proteins encoded on the chromosome of *Pseudomonas putida* KT2440 and IncP-7 plasmid pCAR1, Biosci. Biotechnol. Biochem. 82 (2018) 1640–1646.
[51] C.S. Yun, Y. Takahashi, M. Shintani, T. Takeda, C. Suzuki-Minakuchi, K. Okada, et al., MvaT family proteins encoded on IncP-7 plasmid pCAR1 and the host chromosome regulate the host transcriptome cooperatively but differently, Appl. Environ. Microbiol. 82 (2016) 832–842.
[52] C.S. Yun, C. Suzuki, K. Naito, T. Takeda, Y. Takahashi, F. Sai, et al., Pmr, a histone-like protein H1 (H-NS) family protein encoded by the IncP-7 plasmid pCAR1, is a key global regulator that alters host function, J. Bacteriol. 192 (2010) 4720–4731.
[53] I.W. Dilweg, R.T. Dame, Post-translational modification of nucleoid-associated proteins: an extra layer of functional modulation in bacteria? Biochem. Soc. Trans. 46 (2018) 1381–1392.
[54] S.H. Hong, X. Wang, T.K. Wood, Controlling biofilm formation, prophage excision and cell death by rewiring global regulator H-NS of *Escherichia coli*, Microb. Biotechnol. 3 (2010) 344–356.
[55] Z. Zeng, X. Liu, J. Yao, Y. Guo, B. Li, Y. Li, et al., Cold adaptation regulated by cryptic prophage excision in *Shewanella oneidensis*, ISME J. 10 (2016) 2787–2800.
[56] S. Castang, S.L. Dove, Basis for the essentiality of H-NS family members in *Pseudomonas aeruginosa*, J. Bacteriol. 194 (2012) 5101–5109.
[57] C. Li, H. Wally, S.J. Miller, C.D. Lu, The multifaceted proteins MvaT and MvaU, members of the H-NS family, control arginine metabolism, pyocyanin synthesis, and prophage activation in *Pseudomonas aeruginosa* PAO1, J. Bacteriol. 191 (2009) 6211–6218.
[58] R. Freiherr von Boeselager, E. Pfeifer, J. Frunzke, Cytometry meets next-generation sequencing—RNA-Seq of sorted subpopulations reveals regional replication and iron-triggered prophage induction in *Corynebacterium glutamicum*, Sci. Rep. 8 (2018) 14856.
[59] M. Albano, W.K. Smits, L.T. Ho, B. Kraigher, I. Mandic-Mulec, O.P. Kuipers, et al., The Rok protein of *Bacillus*

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

subtilis represses genes for cell surface and extracellular functions, J. Bacteriol. 187 (2005) 2010–2019.
[60] A. Caramel, K. Schnetz, Lac and lambda repressors relieve silencing of the *Escherichia coli* bgl promoter. Activation by alteration of a repressing nucleoprotein complex, J. Mol. Biol. 284 (1998) 875–883.
[61] R.R. Yu, V.J. DiRita, Regulation of gene expression in *Vibrio cholerae* by ToxT involves both antirepression and RNA polymerase stimulation, Mol. Microbiol. 43 (2002) 119–134.
[62] M.A. De la Cruz, M. Fernandez-Mora, C. Guadarrama, M.A. Flores-Valdez, V.H. Bustamante, A. Vazquez, et al., LeuO antagonizes H-NS and StpA-dependent repression in *Salmonella enterica ompS1*, Mol. Microbiol. 66 (2007) 727–743.
[63] W.W. Navarre, T.A. Halsey, D. Walthers, J. Frye, M. McClelland, J.L. Potter, et al., Co-regulation of *Salmonella enterica* genes required for virulence and resistance to antimicrobial peptides by SlyA and PhoP/PhoQ, Mol. Microbiol. 56 (2005) 492–508.
[64] A.K. Heroven, G. Nagel, H.J. Tran, S. Parr, P. Dersch, RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and rovA expression in *Yersinia pseudotuberculosis*, Mol. Microbiol. 53 (2004) 871–888.
[65] V.H. Bustamante, F.J. Santana, E. Calva, J.L. Puente, Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression, Mol. Microbiol. 39 (2001) 664–678.
[66] J.L. Mellies, G. Benison, W. McNitt, D. Mavor, C. Boniface, F.J. Larabee, Ler of pathogenic *Escherichia coli* forms toroidal protein–DNA complexes, Microbiology 157 (2011) 1123–1133.
[67] J. Garcia, T.N. Cordeiro, M.J. Prieto, M. Pons, Oligomerization and DNA binding of Ler, a master regulator of pathogenicity of enterohemorrhagic and enteropathogenic *Escherichia coli*, Nucleic Acids Res. 40 (2012) 10254–10262.
[68] T.K. McDaniel, K.G. Jarvis, M.S. Donnenberg, J.B. Kaper, A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 1664–1668.
[69] R.S. Winardhi, R. Gulvady, J.L. Mellies, J. Yan, Locus of enterocyte effacement-encoded regulator (Ler) of pathogenic *Escherichia coli* competes off histone-like nucleoid-structuring protein (H-NS) through noncooperative DNA binding, J. Biol. Chem. 289 (2014) 13739–13750.
[70] H.S. Williamson, A. Free, A truncated H-NS-like protein from enteropathogenic *Escherichia coli* acts as an H-NS antagonist, Mol. Microbiol. 55 (2005) 808–827.
[71] J.A. Levine, A.M. Hansen, J.M. Michalski, T.H. Hazen, D.A. Rasko, J.B. Kaper, H-NST induces LEE expression and the formation of attaching and effacing lesions in enterohemorrhagic *Escherichia coli*, PLoS One 9 (2014), e86618.
[72] E. Perez-Rueda, J.A. Ibarra, Distribution of putative xenogeneic silencers in prokaryote genomes, Comput. Biol. Chem. 58 (2015) 167–172.
[73] Q. Liu, C.C. Richardson, Gene 5.5 protein of bacteriophage T7 inhibits the nucleoid protein H-NS of *Escherichia coli*, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 1761–1765.
[74] S.S. Ali, E. Beckett, S.J. Bae, W.W. Navarre, The 5.5 protein of phage T7 inhibits H-NS through interactions with the central oligomerization domain, J. Bacteriol. 193 (2011) 4881–4892.
[75] J. Wagemans, A.S. Delattre, B. Uytterhoeven, J. De Smet, W. Cenens, A. Aertsen, et al., Antibacterial phage ORFans of *Pseudomonas aeruginosa* phage LUZ24 reveal a novel MvaT inhibiting protein, Front. Microbiol. 6 (2015) 1242.
[76] J. Patterson-West, M. Arroyo-Mendoza, M.L. Hsieh, D. Harrison, M.M. Walker, L. Knipling, et al., The bacteriophage T4 MotB protein, a DNA-binding protein, improves phage fitness, Viruses 10 (2018).
[77] C.H. Ho, H.C. Wang, T.P. Ko, Y.C. Chang, A.H. Wang, The T4 phage DNA mimic protein Arn inhibits the DNA binding activity of the bacterial histone-like protein H-NS, J. Biol. Chem. 289 (2014) 27046–27054.
[78] O.E. Melkina, I.I. Goryanin, G.B. Zavilgelsky, The DNA-mimic antirestriction proteins ArdA CollB-P9, Arn T4, and Ocr T7 as activators of H-NS-dependent gene transcription, Microbiol. Res. 192 (2016) 283–291.
[79] S. Kronheim, M. Daniel-Ivad, Z. Duan, S. Hwang, A.I. Wong, I. Mantel, et al., A chemical defence against phage infection, Nature 564 (7735) (2018) 283–286.
[80] F. Brucoli, J.D. Guzman, A. Maitra, C.H. James, K.R. Fox, S. Bhakta, Synthesis, anti-mycobacterial activity and DNA sequence-selectivity of a library of biaryl-motifs containing polyamides, Bioorg. Med. Chem. 23 (2015) 3705–3711.
[81] L. Pinault, J.S. Han, C.M. Kang, J. Franco, D.R. Ronning, Zafirlukast inhibits complexation of Lsr2 with DNA and growth of *Mycobacterium tuberculosis*, Antimicrob. Agents Chemother. 57 (2013) 2134–2140.
[82] M.D. Walkinshaw, P. Taylor, S.S. Sturrock, C. Atanasiu, T. Berge, R.M. Henderson, et al., Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA, Mol. Cell 9 (2002) 187–194.
[83] S.A. McMahon, G.A. Roberts, K.A. Johnson, L.P. Cooper, H. Liu, J.H. White, et al., Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance, Nucleic Acids Res. 37 (2009) 4887–4897.
[84] C.T. Skennerton, F.E. Angly, M. Breitbart, L. Bragg, S. He, K.D. McMahon, et al., Phage encoded H-NS: a potential achilles heel in the bacterial defence system, PLoS One 6 (2011), e20095.
[85] D.A. Russell, G.F. Hatfull, PhagesDB: the actinobacteriophage database, Bioinformatics. 33 (2017) 784–786.
[86] T. Mihara, Y. Nishimura, Y. Shimizu, H. Nishiyama, G. Yoshikawa, H. Uehara, et al., Linking virus genomes with host taxonomy, Viruses. 8 (2016) 66.
[87] K. McNair, B.A. Bailey, R.A. Edwards, PHACTS, a computational approach to classifying the lifestyle of phages, Bioinformatics. 28 (2012) 614–618.
[88] D. Hyatt, G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification, BMC Bioinf. 11 (2010) 119.
[89] U. Pul, R. Wurm, Z. Arslan, R. Geissen, N. Hofmann, R. Wagner, Identification and characterization of *E. coli* CRISPR–Cas promoters and their silencing by H-NS, Mol. Microbiol. 75 (2010) 1495–1512.
[90] N.A.T. Irwin, B.J.E. Martin, B.P. Young, M.J.G. Browne, A. Flaus, C.J.R. Loewen, et al., Viral proteins as a potential driver of histone depletion in dinoflagellates, Nat. Commun. 9 (2018) 1535.
[91] S.G. Gornik, K.L. Ford, T.D. Mulhern, A. Bacic, G.I. McFadden, R.F. Waller, Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates, Curr. Biol. 22 (2012) 2303–2312.
[92] P.B. Talbert, S. Henikoff, Chromatin: packaging without nucleosomes, Curr. Biol. 22 (2012) R1040–R1043.

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

[93] A.C. Carroll, A. Wong, Plasmid persistence: costs, benefits, and the plasmid paradox, Can. J. Microbiol. 64 (2018) 293–304.
[94] A. San Millan, R.C. MacLean, Fitness costs of plasmids: a limit to plasmid transmission, Microbiol Spectr. 5 (2017).
[95] M. Doyle, M. Fookes, A. Ivens, M.W. Mangan, J. Wain, C.J. Dorman, An H-NS-like stealth protein aids horizontal DNA transmission in bacteria, Science. 315 (2007) 251–252.
[96] R.C. Banos, A. Vivero, S. Aznar, J. Garcia, M. Pons, C. Madrid, et al., Differential regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS, PLoS Genet. 5 (2009), e1000513. .
[97] S.C. Dillon, A.D. Cameron, K. Hokamp, S. Lucchini, J.C. Hinton, C.J. Dorman, Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella* Typhimurium identifies a plasmid-encoded transcription silencing mechanism, Mol. Microbiol. 76 (2010) 1250–1265.
[98] R.C. Banos, S. Aznar, C. Madrid, A. Juarez, Differential functional properties of chromosomal- and plasmid-encoded H-NS proteins, Res. Microbiol. 162 (2011) 382–385.
[99] C. Fernandez-de-Alba, N.S. Berrow, R. Garcia-Castellanos, J. Garcia, M. Pons, On the origin of the selectivity of plasmidic H-NS towards horizontally acquired DNA: linking H-NS oligomerization and cooperative DNA binding, J. Mol. Biol. 425 (2013) 2347–2358.
[100] C. Niu, D. Wang, X. Liu, H. Liu, X. Liu, E. Feng, et al., An H-NS family protein, Sfh, regulates acid resistance by inhibition of glutamate decarboxylase expression in *Shigella flexneri* 2457T, Front. Microbiol. 8 (2017) 1923.
[101] T. Takeda, C.S. Yun, M. Shintani, H. Yamane, H. Nojiri, Distribution of genes encoding nucleoid-associated protein homologs in plasmids, Int. J. Evol. Biol. 2011 (2011) 685015.

Please cite this article as: E. Pfeifer, M. Hünnefeld, O. Popa, et al., Impact of Xenogeneic Silencing on Phage–Host Interactions, Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.02.011

3.3 Spatiotemporal binding dynamics of the xenogeneic silencer CgpS during prophage induction in *Corynebacterium glutamicum*

Hünnefeld M., Filipchyk, A., Polen, T., and Frunzke J.

Part of this thesis; to be submitted

Contributor Role	Contributor
Conceptualization	MH (50 %), JF (40 %), TP (5 %), AF (5 %)
Formal Analysis	MH (80 %), AF (15 %), TP (5 %)
Investigation/Experiments	MH (100 %)
Methodology	MH (80 %), JF (15 %), AF (3 %), TP (2 %)
Project Administration	MH (70 %), JF (30 %)
Software	AF (80 %), MH (20 %)
Supervision	MH (60 %), JF (40 %)
Visualization	MH (90 %), AF (10 %)
Writing – Original Draft Preparation	MH (90 %), JF (10 %)
Writing – Review & Editing	MH (80 %), JF (20 %)

Overall contribution MH: 80 %

All of the presented experimental work was done by MH. The only exception is the analysis of the proteome data (Fig. 3A), which was done in a cooperation with Bianca Klein from the research group of Dr. Stephan Noack from the Forschungszentrum Jülich. The bioinformatical analysis of the genome and the binding data was conducted by MH, AF and TP. AF wrote all scripts used for this analysis. The visualization of the earned data was performed by MH. The manuscript text is part of this doctoral thesis and was written by MH.

Spatiotemporal binding dynamics of the xenogeneic silencer CgpS during prophage induction in *Corynebacterium glutamicum*

Max Hünnefeld, Andrei Filipchyk, Tino Polen, and Julia Frunzke*

Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, Jülich, Germany

*Address correspondence to j.frunzke@fz-juelich.de

Abstract

Virus-derived DNA represents the predominant cause for strain-specific differences within bacterial species. Due to potentially toxic gene products, the expression of viral genes requires a stringent regulation. Recently, the function of the Lsr2-type protein CgpS encoded by the CGP3 prophage was characterized as a prophage silencer with an essential role in maintaining the lysogenic state of the prophage in the Actinobacterium *Corynebacterium glutamicum*.

In the following, we analyzed the genome-wide binding dynamics of CgpS during prophage induction by performing a time-resolved ChAP-Seq analysis combined with proteome and transcriptome analyses. Our data revealed that CgpS shows a dynamic binding behavior under inducing conditions, with a first slight decrease in total binding followed by an increase of binding to the CGP3 region. Subsequently, binding inside the prophage area decreased, while binding to different host regions increased.

CgpS did not fully dissociate from the CGP3 region during induction, but rather accomplishes a remodeling of the silencer nucleoprotein complex at the prophage region together with a partial reallocation to further targets in the host genome. Among these host targets, we found i.a. genes encoding different global regulators (e.g. *dtxR*, *glxR*, *gntR1*), genes encoding proteins involved in cell envelope biosynthesis (e.g. *murA*, *murB*, *wzx*), and genes encoding key players of the DNA replication and repair (e.g. *dnaN*, *recF*, *ruvB*).

Interestingly, under prophage-inducing conditions another CgpS target is the origin of replication of *C. glutamicum*. Altogether, these data provide comprehensive insights into genome wide CgpS binding dynamics during the process of prophage induction and highlight the executive role of Lsr2-type proteins in the coordination of phage life cycles.

Introduction

The activity of bacteriophages, viruses that prey on bacteria, represents an important driver of microbial evolution. A central element shaping this evolution is represented by virus-derived DNA transduced to bacterial cells. Temperate bacteriophages are infectious particles that are able to inject their DNA into host cells and either show lytic or lysogenic behavior. During the lytic life cycle, the phage-DNA is replicated inside the host cells and packaged into newly synthesized phage particles. Subsequently, the host cells are lysed, and the progeny virions are released to the environment, where they can infect further host cells. However, during the lysogenic life cycle, the phage-DNA is integrated into the host cells and remains there in a dormant state as a so-called "prophage" (Little, 2014). Prophage elements are ubiquitously found in bacterial genomes, that may contain up to 20% of viral DNA (Casjens, 2003). On the one hand, these genomically integrated prophage elements can be very beneficial for the host cells, because they can modulate the host's physiology and thus improve e.g. antibiotic resistance, stress tolerance, biofilm formation, defense against further infection or even the whole adaptability towards changing environmental conditions (Hargreaves et al., 2014; Nanda et al., 2015; Wang et al., 2010). On the other hand, harboring a prophage can also be very detrimental for bacteria: prophages can encode toxic genes, highly active phage promoters might lead to high energy costs for the host, or even cause cell death upon induction (Lamberte et al., 2017; Pfeifer et al., 2016; Young, 2014). While the acquisition of new genetic material provides the host with the opportunity to gain new and potentially beneficial traits, it always bears a high risk. Therefore, a very tight regulation of these elements is mandatory.

Many different phage defense systems have been described to play key roles in the molecular arms-race between phages and their hosts, with restriction modification and CRISPR-Cas as the most prominent examples (Doron et al., 2018; Labrie et al., 2010; Stern and Sorek, 2011). Restriction modification systems as well as CRISPR-Cas inherit nuclease-based mechanisms to destroy foreign DNA after entering the host cell in order to enable resistance against phage infection (Dupuis et al., 2013). In contrast to these restriction modification systems and CRISPR-Cas, the only option allowing usage of beneficial prophage-encoded material is xenogeneic silencing (Pfeifer et al., 2019). Xenogeneic silencing describes the silencing of horizontally acquired genes or elements by specific silencing proteins. Four classes of these xenogeneic silencers (XSs) are currently known: H-NS and H-NS-like proteins in different proteobacteria like

Escherichia and *Salmonella* (Navarre et al., 2006; Oshima et al., 2006), MvaT/U in different *Pseudomonas* strains (Tendeng et al., 2003), Lsr2 in Actinobacteria (Ali et al., 2012; Pfeifer et al., 2016) and Rok in *Bacillus subtilis* (Smits and Grossman, 2010). Remarkably, these proteins represent an example of convergent evolution. The structural as well as functional analysis revealed a similar mode of action for these XS proteins. Horizontally acquired DNA usually features a higher content of adenosine and thymine than the host chromosome (Moran, 2004). XSs specifically bind to these AT-rich DNA stretches and form a tight nucleoprotein complex, which leads to a silencing of gene expression at these AT-rich regions (Gordon et al., 2011).

The present study focusses on the Lsr2-like xenogeneic silencing protein CgpS from *C. glutamicum* ATCC 13032. The genome of this *C. glutamicum* strain contains four (cryptic) prophage elements: CGP1-4 (Ikeda and Nakagawa, 2003). CGP1 and CGP2 are highly degenerated. However, the biggest prophage CGP3 is inducible in an SOS-dependent as well as in an SOS-independent manner and spans with more than 200 kb around 7 % of the entire genome and also includes CGP4 (Helfrich et al., 2015; Nanda et al., 2014; Pfeifer et al., 2016). Upon induction, the prophage excises out of the genome, circularizes and is replicated autonomously inside the host cell (Frunzke et al., 2008). Recently, we discovered the XS CgpS as an important key player involved in the regulation of prophage induction and characterized this protein (Pfeifer et al., 2016). CgpS is a small nucleoid-associated protein that shows Lsr2-like properties. It contains a C-terminal DNA-binding domain as well as an N-terminal oligomerization domain. This N-terminal oligomerization domain is crucial for an effective prophage silencing and interference with the DNA binding of CgpS leads to prophage induction (Pfeifer et al., 2016).

Previous studies revealed the important role of XS proteins in maintaining the lysogenic state of prophages, but they only provided a snap shot view on the binding of the particular XS protein under non-inducing conditions. Besides CgpS, a binding to prophage regions and mobile genetic elements was also shown for Lsr2 from *Mycobacterium tuberculosis* (Gordon et al., 2010). Another study demonstrated that the *B. subtilis* XS, Rok, is involved in the regulation of genes inside the SPβ-phage and represses the excision of a mobile genetic element (Albano et al., 2005; Smits and Grossman, 2010). Furthermore, an important role in the silencing of prophages was published for the *E. coli* XS H-NS as well as for the H-NS homologues MvaT/U from *P. aeruginosa*. For both strains, deletion of the silencers is described to cause prophage induction (Castang and Dove, 2012; Hong et al., 2010; Li et al., 2009).

In this study, we performed a time-resolved ChAP-Seq experiment to study the dynamics of CgpS DNA binding in the course of prophage induction. Upon prophage induction, triggered by the addition of the DNA-damaging agent mitomycin C (MMC), we observed a rearrangement of CgpS-binding from the CGP3 region towards >300 targets within the host genome. These results suggest an important role of CgpS in coordinating the phage life cycle.

Results

Mitomycin C causes prophage induction and affects genome replication

Previous studies already revealed that CGP3, the largest prophage of *C. glutamicum* ATCC 13032, can be induced by the addition of the DNA-damaging antibiotic MMC, which is commonly used to trigger the cellular SOS-response (Nanda et al., 2014). Aim of this study was to analyze the DNA-binding dynamics of the xenogeneic silencer CgpS during SOS-triggered prophage induction. Therefore, a first step was the analysis of effects of MMC addition on the *C. glutamicum* chromosome. Figure 1 shows the whole genome sequencing results before MMC addition in comparison to a sample six hours after the addition of MMC. Clearly visible for both samples is the increased genomic coverage at both flanks of the plot, where the *oriC* of the strain is located. In the uninduced sample, the region close to this origin of replication (ori) shows an approximately 2-fold higher in comparison to the middle of the chromosome where the replication terminus (ter) is located. This increased amount of DNA close to the ori is a result of multifork chromosome replication of growing bacterial cells (Cooper and Helmstetter, 1968; Couturier and Rocha, 2006). The increase in genomic coverage of the area close to the ori is even more pronounced in the MMC-induced sample. Here, values are up to 6-fold higher compared to the terminus region. Further, up to 4-fold increase of the genomic coverage of the CGP3 prophage area (~ 1.8 - 2.0 Mb) was found. This illustrates that the prophage CGP3 is induced upon an MMC-triggered SOS-response and is able to replicate inside the *C. glutamicum* cells. These data are important for further analysis of CgpS binding during prophage induction as the changes in the genomic coverage need to be taken into account. To normalize CgpS binding for spatial overabundance of certain genomic regions (e.g. ori and CGP3 region), the information of this genome sequencing experiment served as ‘input control’. For every experiment, the signal obtained by the ChAP-seq analyses was normalized to the input

control obtained from the genome sequencing at the respective time points. An example of this normalization strategy is shown in Figure S1.

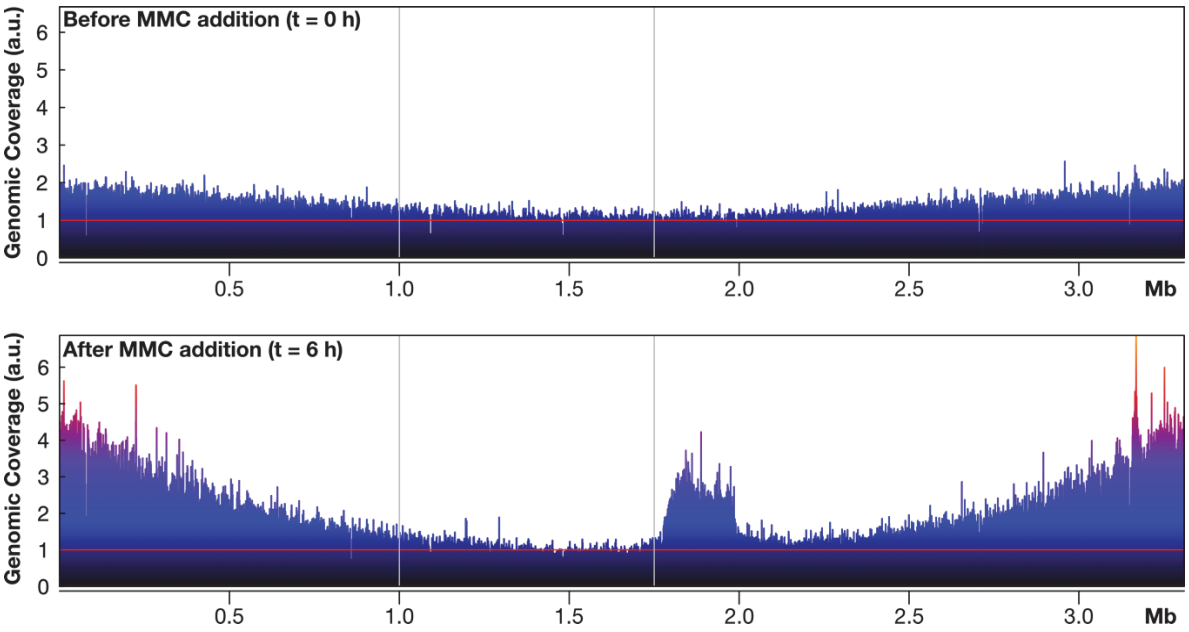


Figure 1: Comparison of sequencing reads from *C. glutamicum* cells with and without MMC. To visualize the effects of MMC addition to the average chromosome copy number of the *C. glutamicum* ATCC 13032::*cgpS-strep* cell culture, cells were grown in CGXII medium containing 2% (w/v) glucose and harvested before and 6 h after addition of mitomycin C. Subsequently, cells were disrupted and sonicated as described in the ChAP-seq section of the materials and methods part. Total DNA was prepared with Roti®-Phenol/Chloroform/Isoamylalcohol (Carl Roth GmbH, Karlsruhe, Germany) following the manufacturer's instructions, and the DNA was sequenced using and Illumina MiSeq system (Illumina, San Diego, California, USA). The reads were mapped against the genome NC_003450.3 and the coverage was normalized to the area between both grey lines. This defines an area close to the replication terminus, which was set as the area were the copy number is in average one per cell (red horizontal line). Further, the color of the genomic coverage graph symbolizes the height of the graph (ranging from black over blue towards red and yellow, increasingly with each color representing 20 %).

CgpS-binding during SOS-dependent prophage induction

The xenogeneic silencer CgpS was shown to silence prophage activity in *C. glutamicum* (Pfeifer et al., 2016). In a small fraction of cells this prophage silencing is spontaneously inhibited as a result of an SOS-response (Helfrich et al., 2015). In order to get a more detailed insight into the changes in silencer binding during this SOS-dependent prophage induction, we performed time-resolved ChAP-seq experiments with an uninduced as well as with MMC-induced *C. glutamicum* samples (Figure 1). For this purpose, the strain *C. glutamicum* ATCC 13032::*cgpS-strep*, encoding a C-terminally Strep–tagged variant of CgpS, was grown in glucose minimal medium and samples for ChAP-Seq analysis were taken before (0 h) and up to 24 h after induction with MMC. Figure 2 shows a genome-wide overview of the normalized binding patterns of CgpS during the different time points.

The uninduced sample shows the highest genomic coverage and thereby the primary CgpS binding inside of the CGP3 area. Further regions targeted by CgpS are other horizontally acquired elements (e.g. CGP1) or in general AT-rich regions inside the chromosome. These findings match to the already published data concerning CgpS binding (Pfeifer et al., 2016). Comparing the uninduced with the MMC-induced samples, the most obvious difference is the high decrease in CGP3 binding six hours after MMC addition. Simultaneously, the peaks outside of the CGP3 region highly increased in genomic coverage. Furthermore, new CgpS binding peaks appear outside of the CGP3 region by addition of MMC (Figure 2). Especially the low GC area around 400,000 bp shows a high increase of CgpS binding. This area contains several putatively horizontally acquired elements involved in cell envelope biogenesis and modification (Kalinowski et al., 2003). In total, the representation of the genomic coverage of the induced sample indicates a rearrangement of CgpS during prophage induction. Nevertheless, no complete dissociation of CgpS from the CGP3 area can be detected under prophage inducing conditions.

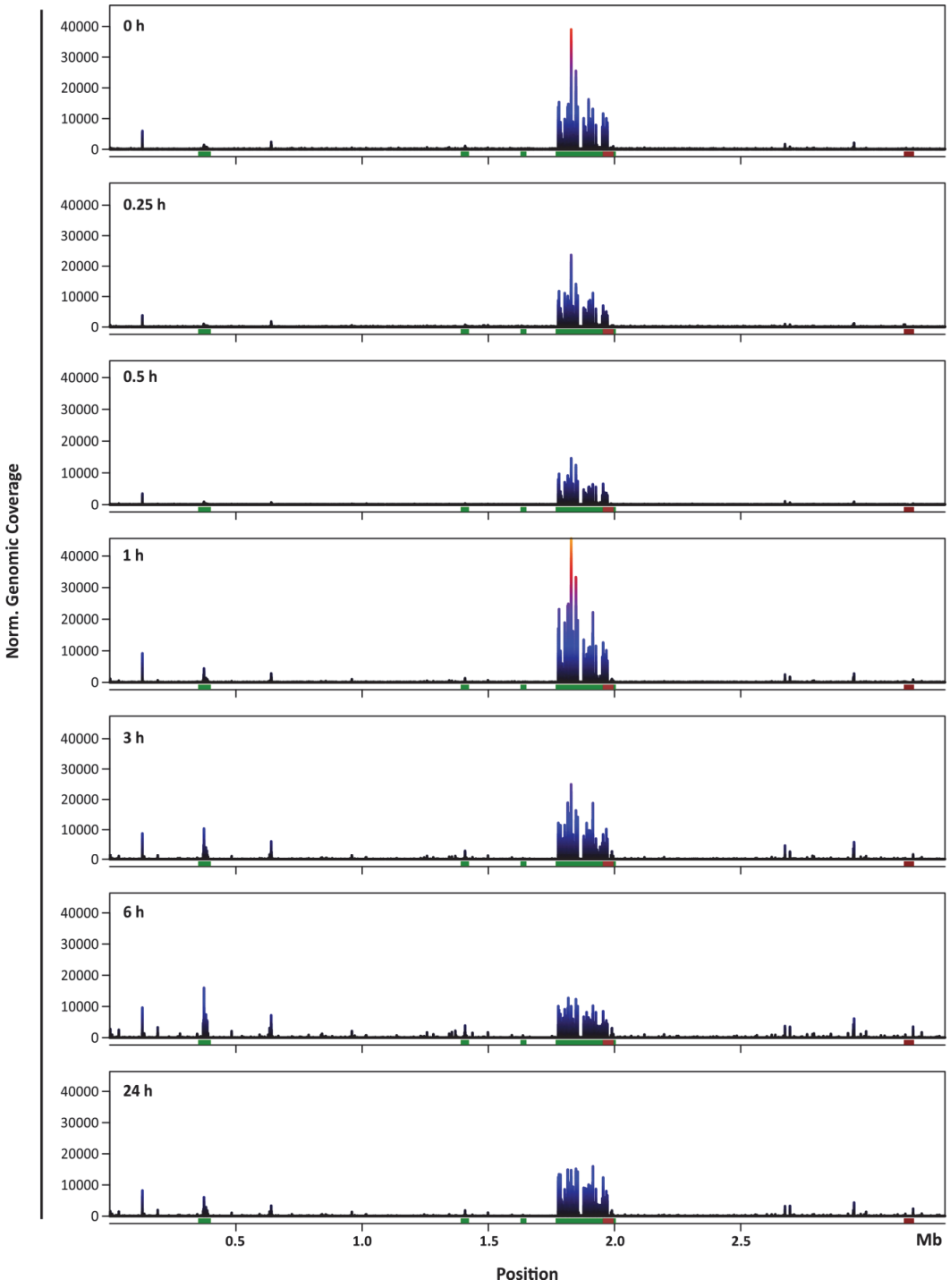


Figure 2: Normalized time series of CgpS binding in *C. glutamicum* under SOS-dependent prophage induction using ChAP-seq analysis. For these ChAP-seq analyses *C. glutamicum* ATCC 13032::*cgpS-strep* cells were grown in CGXII medium containing 2 % (w/v) glucose. The CGP3 prophage was induced by the addition of 600 nM mitomycin C. Samples were taken before (0 h) as well as 0.25 h, 0.5 h, 1 h, 3 h, 6 h and 24 h after induction with MMC. CgpS-bound DNA was purified and sequenced. The presented data were normalized with regard to the average coverage to enable comparisons between the different experiments (exemplarily shown in Figure S1). The coloring of the graphs represents the genomic coverage at a certain point and is split into 20 % sections in the following order (from lowest to highest values): black, blue, purple, red, orange. The green lines represent regions of low GC content and prophage regions (LGC1, CGP1-4), the dark red line represents a high GC region and the brown inclusion inside the big green CGP3 region marks the CGP4 area included in CGP3.

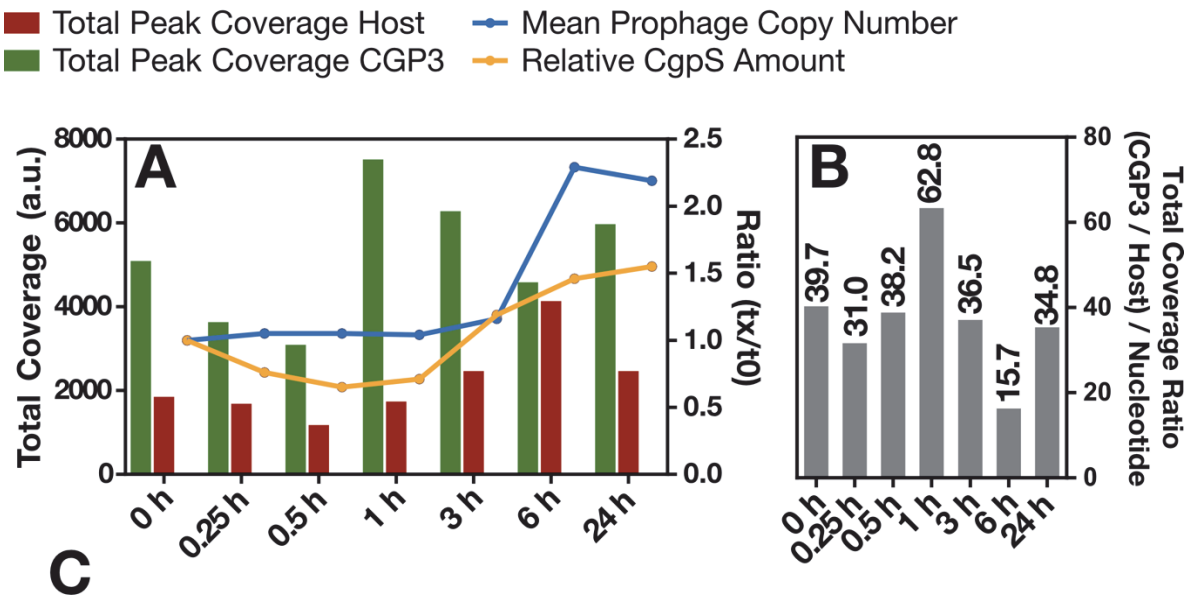
Dynamic rearrangement of CgpS binding enables prophage induction

In addition to the ChAP-seq analyses, we performed whole-genome sequencing (input control) and proteome analysis using LC-MS to determine the CgpS level as well as the prophage copy number at each time point and to earn a broader insight into the cellular changes due to SOS-dependent prophage induction (Figure 3). Both, total peak coverage of CgpS binding towards the host and towards the CGP3 region, show a highly dynamic pattern in the course of prophage induction (Figure 3A). During the first 30 minutes after induction, the total CgpS binding decreases. Additionally, the protein level of CgpS decreases 1.5-fold and reaches its minimum at around 30 minutes after induction with MMC. This decreased amount of the CgpS protein level was in line with an overall decrease in genomic coverage of CgpS peaks. The initial reduction of CgpS binding might be an important requirement to enable CGP3 induction. One hour after MMC addition, the total relative peak coverage increased again (CGP3: 2.5-fold, host: 1.5-fold) as shown in Figure 3A and 3C. This increase is mainly (81.6%, Figure 3C) shaped by an elevated binding inside the CGP3 region. Figure 3B displays further that one hour after MMC addition the ratio of CGP3-associated CgpS to host-associated CgpS reaches its highest values with a 62.8-fold excess. Simultaneously, the relative CgpS amount inside the cell increases, resulting in the 0.71-fold amount of CgpS before induction with MMC. Following the course of the time series, at three hours after MMC addition, the relative peak volume of the prophage peaks decreased slightly, although the relative amount of CgpS raises up to 1.19-fold. On the contrary, the volume of the binding peaks within the host genome significantly increased when compared to the start of the experiment (Figure 3A and 3C). Moreover, at this time point the mean of the prophage copy number started to increase (~1.16-fold).

Six hours after MMC-addition, the mean prophage copy number shows the highest value of the presented data set (~2.29, Figure 3A and 3C). At the same time the CgpS amount is only slightly further increased compared to the previous measurement. However, the CgpS per prophage ratio is with 0.64 at a similar level like measured at t = 0.5 h (0.62; see Figure 3C). Simultaneously, the binding of CgpS outside of the CGP3 area is higher, which led to a ratio of 52.6% CGP3-bound to 47.4% host-bound CgpS (Figure 3C). This is further reflected by the relative CGP3 to host binding ratio (Figure 3B). This value reaches the minimum value during the time series with a 15.7-fold excess of binding inside CGP3. Interestingly, the increase in prophage copy number was higher than the increase in CgpS protein level. A combination of

these results with the rearrangement of the CgpS binding provides hints on how the CGP3 induction is enabled.

A final sample of the presented time series was taken 24 hours after addition of MMC. Here, the mean prophage copy number is slightly decreased compared to the six hours value, which is probably due to cell lysis. Additionally, the relative amount of CgpS featured a further slight increase (Figure 3A and 3C). Together with the decreased prophage copy number, this leads to a CgpS per CGP3 ratio of 0.71 (Figure 3C). This ratio is comparable to the value of t = 0.25 h (0.72). Nevertheless, in this case the percentage of CGP3/host-associated CgpS is with 71.1% to 28.9% very similar to the values at time point t = 0 h (Figure 3C). This is accompanied by a decrease of the total peak coverage inside the host genome, whereas the total peak coverage of CgpS binding increased slightly inside the CGP3 area. Figure 3B shows that the binding of CgpS at 24 h is very similar to t = 0 h. In total, these values suggest that the population reaches an equilibrium again and that the observed signal is caused by cells containing the CGP3 prophage still in its lysogenic state.



Time (h)	Rel. CgpS Amount	Mean CGP3 Copy Number	CgpS / CGP3	Total Coverage CGP3	Total Coverage Host	Rel. Coverage CGP3	Rel. Coverage Host
0	1.00	1.00	1.00	5031	1793	73.7	26.3
0.25	0.76	1.05	0.72	3572	1626	68.7	31.3
0.5	0.65	1.05	0.62	3033	1122	73.0	27.0
1	0.71	1.04	0.68	7453	1678	81.6	18.4
3	1.19	1.16	1.03	6221	2409	72.1	27.9
6	1.46	2.29	0.64	4528	4074	52.6	47.4
24	1.55	2.19	0.71	5913	2405	71.1	28.9

Figure 3: Overview on the intracellular dynamics of the CgpS silencer with regards to CgpS binding, protein level and prophage copy number. (A) The total coverage obtained by the ChAP-seq experiment is represented in two categories: 1) inside the CGP3 region (all peaks between 1776613 – 1995294 bp, green) and outside of the CGP3 region (= Host, red). Additionally, the mean prophage copy number was calculated using the genomic input control (Figure 1), normalized to a minimum of one copy at t_0 (blue line). The orange line indicates the relative amount of CgpS normalized to the t_0 value as determined by LC-MS analysis. Values corresponding to this table are displayed in C. **(B)** The total coverage ratio normalized to the number of nucleotides spanning the chosen region was calculated. The figure shows the binding of CgpS to the CGP3 element compared to the binding of CgpS to the host genome calculated per nucleotide. **(C)** Overview of specific values regarding CgpS binding, CgpS amounts and CGP3 copy number derived from ChAP-seq, proteomics and whole genome sequencing approaches.

CgpS binds to several promoter regions of genes encoding global regulators

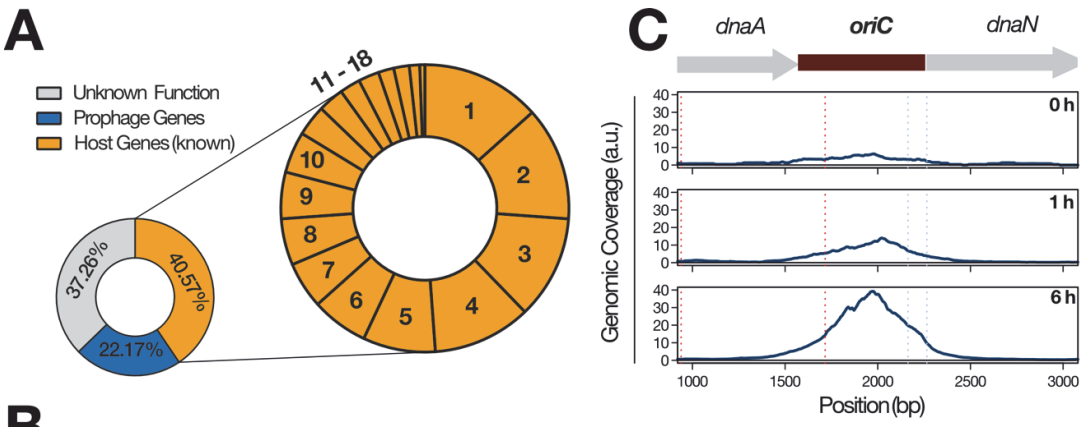
The overview of the binding dynamics as well as the visualization of the ChAP-seq time series (Figure 2, Figure 3) shows a clear shift of CgpS binding from the prophage area towards the host genome. This observation gave rise to the question: "What exactly are the targets of CgpS inside the host genome and how is the timing of CgpS binding towards these targets?" To answer this question, we first created a heatmap with an extract of host targets of CgpS from the data set presented in Table S1. This heatmap, however, only demonstrates high dynamics inside the binding peaks, but does not reflect a clear trend towards different regulations of

different types of bound promoters or any clusters that seem to be functionally related (Figure S2). Thus, a first step to gain further information about the impact of CgpS on different genes was to analyze its binding on a peak-level. A first interesting example of host manipulating binding of CgpS under prophage inducing conditions is given by the association of CgpS towards the ori of *C. glutamicum*. This example is shown in Figure 4C. Here, the binding pattern of CgpS covering the ori at different time points (0 h, 1 h and 6 h) is plotted. This figure demonstrates that already before MMC addition a very slight binding of CgpS to this ori region is detectable. Nevertheless, in the following this binding highly increases. Remarkably, the structure of the peak contains a small bump close to the two + strand TSSs which could hint on both: a specific binding to the ori region as well as a binding to the promoter region of *dnaN*, which is located upstream of this peak area.

For further investigations regarding the physiological meaning of CgpS binding, we filtered for peaks, that bind upstream of a gene and most likely influence their expression. Here, a threshold for peaks located up to 700 bp upstream of a gene was chosen. These genes were then classified into three different categories (unknown function, prophage genes and host genes with known function; Figure 4A). Figure 4A further provides a more detailed zoom into the functional categories of these genes. More details concerning the number of genes as well as additional information about gene categories with special examples can be found in the corresponding Figure 4B. In summary, this figure shows that within the host genes with known function the highest number of binding peaks was detected upstream to genes coding for proteins involved in signal transduction. These proteins include e.g. transcriptional regulators, two-component systems and other proteins involved in cellular signaling. The second largest group is represented by genes with a putative function in DNA replication, recombination or repair, followed by the third group containing genes coding for proteins involved in cell envelope biosynthesis (Figure 4B). These results suggested that the prophage-encoded silencing protein CgpS seems to hijack important components of the cell to enable efficient induction and proliferation.

Figure 4D represents the maximal CgpS peak coverage of each binding peak at the different time points of the ChAP-seq series exemplified for some known transcriptional regulators. Although the single binding patters are highly varying, there are some similarities between those. Four genes that show a similar CgpS binding behavior are *glxR*, *qorR, sugR* and *mtrB*. These display a low start and a maximum binding peak at six hours after MMC-addition. The

binding behavior of CgpS to these four genes is similar to those to the average host genome (Figure 4D). A second group showing similar binding behaviors are the three genes *cadR*, *dtxR* and *znr*. This second group shows similarities with both, the binding behavior of CgpS to average host and to average CGP3 areas. Accordingly, one peak occurs one hour after MMC-addition followed by a second (higher) peak at six hours after MMC-addition. In contrast, the further selected genes show individual binding patterns: *arnR* and *gntR*1 both show a peak at three hours, but subsequently divide in two different patterns. *divS* shows a maximum peak 0.25 h, followed by a continuous decrease of binding, whereas, *acnR* is not bound in the beginning and only shows an increase in binding 24 hours after MMC-addition.



B

No.	Functional Category	Examples	Amount	%
1	Signal transduction mechanisms	*dtxR, glxR, gntR1, divS*	23	13.4
2	DNA replication, recombination, repair, degeneration	*dnaN, dnaB, recF, ruvB*	22	12.8
3	Cell wall/membrane/envelope biogenesis	*murA, murB, ufaA, wzx*	20	11.6
4	Amino acid transport and metabolism	*thrE, dapA, gapA, ilvN*	19	11.0
5	Inorganic ion transport, metabolism and storage	*znuA, znuB, pitA, thtR*	14	8.1
6	Carbon source transport and metabolism	*soxA, ptsI, malQ, benE*	11	6.4
7	Coenzyme transport and metabolism	*panB, thiX, pncA, mog*	9	5.2
8	Translation, ribosomal structure and biogenesis	*fusA, argS, rbfA, hflX*	9	5.2
9	Transport and metabolism of further metabolites	*ntaA, ureD, crtYf, ggtB*	9	5.2
10	Nucleotide transport and metabolism	*purC, pyrH, udgA1, rbsK1*	8	4.7
11	Protein turnover and chaperones	*smpB, pepD, ctaB, clpC*	6	3.5
12	Transcription (incl. sigma-factors, RNA processing and modification)	*hrpB, cspA, sigD, cspB*	5	2.9
13	Post-translational modification	*pup, arc (mpa), phoD, cgtS6*	4	2.3
14	Respiration and oxidative phosphorylation	*cydA, cydB, etfB (fixA), qor3*	4	2.3
15	Central carbon metabolism	*gltA, fumC(fum), malE*	3	1.7
16	Protein secretion	*secA1, secD*	3	1.7
17	Cell division and chromosome partitioning	*divIVA, ftsZ*	2	1.2
18	Lipid transport and metabolism	*accBC*	1	0.6

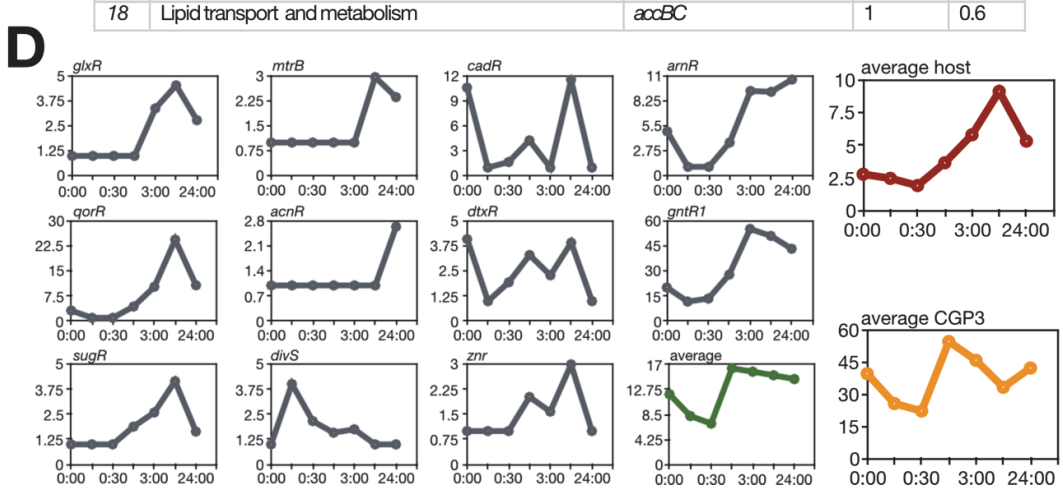


Figure 4: CgpS redistributes towards targets within the host genome upon prophage induction (A) To analyze the distribution of binding peaks of CgpS with regard to the functional category of the targeted genes, in a first step the complete datasets (Figure 2) were sorted for peaks that showed at one of the time points at least a z-score of 5. Further, the remaining peaks were filtered for their proximity to the next upstream gene start (≤ 700 bp). The left-hand circle shows the allocation of those filtered peaks into three different categories (peaks probably related to 1) genes with unknown function (grey), 2) prophage genes (blue) and 3) host genes with a known functional category (orange)). The right-hand circle shows a more detailed zoom into the CgpS-targeted host genes with known functional category. Here, each number represents a special category. The explanation of those numbers, the exact number of peaks and examples of these genes are shown in Figure 4B. **(B)** Functional categories of CgpS-

bound genes. Binding peaks are filtered to be ≤ 700 bp away from the gene starts (compare Table S1 for the complete filtered data set). This table corresponds with (A). **(C)** A zoom into the time points 0 h, 1 h and 6 h from the ChAP-seq time series. Here, the focus lies on the oriC region of *C. glutamicum*, which is bound by CgpS. The grey dotted vertical lines represent transcriptional start sites (TSS) on the + strand, whereas the red dotted vertical line marks a TSS on the - strand. **(D)** The dynamic behavior of the CgpS-binding to the promoter regions of different transcriptional regulators. Plotted on the x-axis is the corresponding timepoint from the ChAP-seq time series (each time point as one intercept, not a linear time-scale; 0:00, 0:15, 0:30, 1:00, 3:00, 6:00, 24:00). The y-axis shows each maximum coverage value of the current peak.

Discussion

In the presented study, we analyzed the spatiotemporal dynamics of the Lsr2-like protein CgpS in *C. glutamicum* upon induction of an SOS-response. A time-resolved profiling of genome-wide CgpS binding, as well as whole genome sequencing and protein quantification, shed light on the CgpS binding pattern allowing an induction of the CGP3 prophage.

The SOS-dependent induction of the CGP3 prophage was triggered using the DNA damaging agent MMC (Szybalski and Iyer, 1964; Tomasz, 1995). This led to changes in the DNA replication as well as to an increase of DNA of the prophage region. To take the multifarious changes inside the cell caused by the SOS-response into consideration we analyzed the effects of the MMC-triggered DNA crosslinking on the DNA replication in *C. glutamicum* (Figure 1). Besides the CGP3 replication, the genome areas close to the origin of replication were also highly overrepresented upon MMC induction. This could be an effect caused by the crosslinking of the complementary strands caused by MMC. For the replication of the chromosome it is mandatory to separate both strains at the ori region in order to form the replication fork, which is conducted using the origin binding AAA^+-family protein DnaA (Kaguni, 2011). For eukaryotic cells it was shown in 2009, that addition of MMC can inhibit the replication fork elongation (Al-Minawi et al., 2009). An inhibition of the replication fork elongation leading to a high increase of multiplied DNA close to the ori could also explain the results of our experiments.

A ChAP-seq time series, performed to gain insights into the dynamics of CgpS binding, revealed that the CgpS binding is highly dynamic during SOS-dependent prophage induction. The CGP3 area was partly released and simultaneously new and higher host peaks accumulated. Nevertheless, even when CGP3 is highly induced and in its circularized replicated state inside the cell (Figure 1), the silencer CgpS never completely dissociates from the prophage region (Figure 2 and Figure 3). It is important to note that the prophage induction (both, spontaneous and triggered) in *C. glutamicum* never happens homogeneously in all cells at the same time

resulting in a heterogeneity within the population (Frunzke et al., 2008; Helfrich et al., 2015). In order to circumvent analysis of heterogeneous populations, a recent approach used fluorescence associated cell sorting (FACS) to allow a targeted analysis of a certain subpopulation (Freiherr von Boeselager et al., 2018).

The dynamics between CgpS level, prophage copy number and binding of CgpS reveal interesting changes that hint on further not yet discovered involved components. Although CgpS levels are still below starting levels, the binding to CGP3 increases very strongly and reaches a top value at t = 1 h. This hints on the involvement of another factor, e.g. a post-translational modification (PTM) of CgpS triggering the binding upon a certain stimulus, or a possible protein-protein interaction, modifying CgpS affinity. An intracellular equilibrium of bound and unbound CgpS probably allows the fast reaction towards different stimuli due to affinity changes. Previous studies with an additional copy of *cgpS* inside the *C. glutamicum* genome showed highly decreased inducibility of CGP3 (Pfeifer, 2017), probably derived from a perturbation of this equilibrium. A recent study investigated in total 101 unique PTMs of the nucleoid-associated proteins H-NS, HU, IHF and FIS in *E. coli* (Dilweg and Dame, 2018). These modifications led either to a perturbation of nucleoid binding or to a changed binding mode.

Furthermore, the coverage ratio of CgpS-bound CGP3 and host peaks appears to be very stable (Figure 3B). Exceptions were only detectable at t = 1 h (maximal CGP3 binding) and at t = 6 h (minimal CGP3 binding). This minimal CGP3 association of CgpS occurs despite of elevated proteins levels. This could be an additional hint on a change of the CgpS affinity upon a possible modification.

To get an insight into how CgpS influences host physiology under inducing conditions, we focused in the following on binding sites in the host genome. A notable peak, located in the low GC area at around 400,000 (Figure 2), is assumed to be acquired horizontally (Kalinowski et al., 2003). This observation leads to the hypothesis, that – upon induction – the affinity of CgpS towards CGP3 is changed. Thus, CgpS attaches to other AT-rich horizontally acquired DNA regions. However, the binding of CgpS to other horizontally acquired elements (CGP1, CGP2) increases only slightly in course of the time series (Figure 2). Possibly another factor displaces CgpS from CGP3. Another hypothesis is that the genes within the low-GC-area are relevant for the induction of CGP3. The low-GC-area encodes ISCg17a transposase fragments and multiple genes encoding proteins involved in cell envelope biogenesis or modification: *murA* and *murB*,

which are genes important for peptidoglycan synthesis (Burkovski, 2013), wzx and wzy, which are involved in O-antigen synthesis and placement (Islam and Lam, 2014), and a putative glycosyl transferase.

The benefit of the CGP3 prophage to target genes involved in cell envelope composition can be a weakening of the peptidoglycan layer via interference with precursor supply. Because of high stability of the mycobacterial membrane, lysis of mycobacteria is connected to multiple cell envelope destroying mechanisms, involving i.a. mycolic acid cleavage (Lysin B), hydrolysis of peptidoglycan (endolysins) or precise lysis timing (holins) (Payne and Hatfull, 2012). This weakening of the stable mycobacterial membrane would facilitate phage-induced cell lysis. Although nothing is known about regulatory influence of phages on transcription of cell envelope biosynthesis genes to date, Chamakura et al. demonstrated an interference of the viral protein LysM directly with peptidoglycan synthesis in *E. coli* (Chamakura et al., 2017).

Other interesting host-targets of CgpS were unveiled by a filtering for functional categories. Different targets with partly similar binding patterns were shown inside the group of signal transduction mechanisms (Figure 4B). The transcriptional regulators *cadR*, *dtxR* and *znr*, for example, exhibited a very specific binding pattern (Figure 4D). Beyond that, these regulators share further similarities: The regulators are all involved in metal homeostasis, like Znr for zinc (Smith et al., 2009), DtxR for iron (Brune et al., 2006; Wennerhold and Bott, 2006) and CadR putatively for cadmium. Additionally, each of the genes encoding these regulators is part of a small operon with another gene (Kalinowski et al., 2003). For *dtxR*, Wennerhold et al. demonstrated that a *C. glutamicum* strain lacking this gene shows upregulation of more than 50 CGP3 genes. Nevertheless, to decipher a direct connection between CgpS and metal homeostasis, further investigations are required.

A further interesting host-target of CgpS is *divS*, which is bound directly after MMC addition Figure 4D. DivS is published as an SOS-inducible suppressor of cell division in *C. glutamicum* (Ogino et al., 2008). Ogino et al. could show that DivS is responsible for the elongated cell growth, appearing as a result of an SOS-response. Interestingly, opposite to the *divS* gene, the gene *lexA* is encoded. This gene encodes LexA, one of the key players inside the cellular SOS-response (Jochmann et al., 2009). The CgpS binding site between both of these genes could influence their transcription. This represents a participation of CgpS in the regulation of SOS-related components.

A further noteworthy target of CgpS is the origin of replication of *C. glutamicum*. Figure 4C shows an increase of the genomic coverage in the course of the time series inside this region. The ori region shows high AT contents and a high density of DnaA boxes (consensus for *C. glutamicum*: 5'-TTATCCACA-3'). These boxes are recognized by the replication initiation protein DnaA (Luo and Gao, 2019; Mott and Berger, 2007). It is already known for diverse species, that other proteins than DnaA can bind to the ori region to modify the replication initiation (Wolanski et al., 2015). Wolanski et al. showed that i.a. histone-like nucleoid-associated proteins like H-NS, IHF and HU from *E. coli* bind inside the ori region. This binding can stabilize the DnaA binding towards the ori region and thus enhance replication. Furthermore, they can block the replication initiation, or they exhibit a not yet deciphered mode of action (Wolanski et al., 2015). Another example of a XS protein associated to the *oriC* was published by Seid et al. in 2017. By using ChIP-seq analysis they could show that DnaA and the *B. subtilis* XS protein Rok share similar binding patterns along the whole genome (Seid et al., 2017). One of these was detected in the ori region. However, tested null mutants of *rok* did not show an effect on replication. Because the ori is a region of high AT-content and XS proteins preferentially bind to those regions, binding of those silencers could also be a secondary effect caused by binding site similarities.

In addition to the promoter association of CgpS, its binding could exhibit structural impact on the DNA topology or the chromosome organization in *C. glutamicum*. For several other XS it is a common feature that they can influence the DNA structure. These structural changes are necessary for the silencing abilities. DNA can be bridged, coated or stiffed. For H-NS, Ler and MvaT these DNA structuring properties could be shown with atomic force microscopy (Dame et al., 2005; Winardhi et al., 2014) as well as with a magnetic-tweezer approach, (Winardhi and Yan, 2017). Because we found regions flanking the CGP3 region that never show CgpS binding, our data also hint on structural properties of CgpS. Maybe these specific gaps arise because of a specific structuring of the chromosome with the help of CgpS. To clarify possible structural properties formed by CgpS further experiments have to be conducted.

In the future, transcriptome analysis would enable a better insight into the regulatory relevance of the CgpS binding. However, during the writing process of this thesis, these data sets were not available.

This manuscript provides for the first time a time-resolved binding profiling of a XS protein during prophage induction combined with proteome and transcriptome analyses. Our data revealed that under inducing conditions, CgpS never fully dissociates from the prophage region. However, the CgpS-DNA nucleoprotein complex appeared to be loosened to enable prophage induction and CgpS was partly redistributed to multiple host gene targets. In total, we demonstrated a dynamic role of the Lsr2-like XS protein CgpS in the coordination of the CGP3 life cycle.

Materials and Methods

Bacterial strain and growth conditions

The bacterial strain *Corynebacterium glutamicum* ATCC 13032 with a chromosomal exchange of *cgpS* with *cgpS-strep* (coding for CgpS with a C-terminal Strep-tag) was used for all experiments in this study (Pfeifer et al., 2016). This strain was cultivated in brain heart infusion medium (BHI, Difco Laboratories, Detroit, MI, USA) or in CGXII minimal medium containing 2% (w/v) glucose (Keilhauer et al., 1993). For the cultivation, a first pre-culture was inoculated with single colonies from agar plates, either directly after transformation or after a streak-out of glycerol cultures. These pre-cultures were grown in 4.5 ml BHI medium in test tubes at 30°C for 8 h. Subsequently, cells were used for inoculating a second overnight pre-culture in CGXII medium containing 2% (w/v) glucose. This CGXII culture was used to inoculate a 500 ml main culture in the same medium to an OD_{600} of 1.5. During this main-culture, *C. glutamicum* ATCC 13032::*cgpS-strep* was pre-grown in CGXII medium containing 2% (w/v) glucose for 1 h in 2 l shaking flasks at 30°C and 120 rpm shaking frequency. Subsequently, the first sample (t = 0 h) was harvested and 600 nM MMC was added to the remaining samples, which were then further incubated for different times (0.25 h, 0.5 h, 1 h, 3 h, 6 h, 24 h). The amount of harvested culture was calculated to always contain 1375 'OD-units' (t = 0 h contained 500 ml culture with an OD_{600} of 2,75), to earn a comparable sample size. This culture method was used to harvest samples for all further described methods.

Chromatin Affinity Purification and next generation sequencing (ChAP-Seq)

Sample Preparation

Different amounts of cells (as described above) were harvested by centrifugation (10 min at 11,325 x g and 4°C). Subsequently, the cells were washed once with CGXII medium without MOPS and resuspended in the same medium containing 1 % (v/v) formaldehyde as a fixation agent and incubated at room temperature for 20 min. To stop this, glycine was added to a final concentration of 125 mM, followed by an additional 5 min incubation step at room temperature. To remove the remaining formaldehyde, the cells were washed twice with buffer A (100 mM Tris-HCl, pH 8.0, 1 mM EDTA), and afterwards resuspended in 25 ml buffer A, containing *cOmplete* Protease Inhibitor (Roche, Basel, Switzerland) and 2.5 mg RNase A. The subsequent preparation of ChAP-seq samples (cell disruption via French press, sonification and ultracentrifugation) was conducted as described previously (Pfeifer et al., 2016) . Afterwards, the supernatant was then purified using a 2-ml bed volume Strep-Tactin-Sepharose column (IBA, Göttingen, Germany), following the manufacturer's protocol. The elution fractions were pooled and incubated over night at 65°C, followed by a treatment with proteinase K (final concentration 400 mg·ml−1) for 3 h at 55°C. In a last step, the DNA of the samples was purified by phenol–chloroform extraction using Roti®-phenol/chloroform/isoamyl alcohol (Carl Roth, Karlsruhe, Germany), precipitated with ethanol and 0.3 M sodium acetate, washed with 70% (v/v) ethanol, dried and resuspended in 100 µl H_2O.

Sequencing, Bioinformatic Analyses and Normalization to Input Control

The obtained DNA fragments of each sample (up to 2 µg) were used for library preparation and indexing using the TruSeq DNA PCR-free sample preparation kit according to the manufacturer's instruction, yet omitting the DNA size selection steps (Illumina, Chesterford, UK). The resulting libraries were quantified using the KAPA library quant kit (Peqlab, Bonn, Germany) and normalized for pooling. Sequencing of pooled libraries was performed on a MiSeq (Illumina) using paired-end sequencing with a read-length of 2 x 150 bases. Data analysis and base calling were accomplished with the Illumina instrument software and stored as fastq output files. The sequencing data obtained for each sample were were collapsed to remove PCR amplification artifacts. The processed fastq files were mapped to accession NC_003450.3 as *C. glutamicum* reference genome with Bowtie2 using the following parameters: --ignore-

quals --local --very-sensitive-local --rfg 9,5 --rdg 9,5 --score-min L,40,1.2 -k 8 --no-unal --no-mixed --threads 8 -I 40 -X 800 (Langmead et al., 2019; Langmead and Salzberg, 2012). Genomic coverage was normalized to chromosome coverage values of the input control and then convoluted with second order Gaussian kernel. The kernel was truncated at 4 sigmas (that is all kernel values positioned further then 4 sigmas from the center were set to zero) and expanded to the "expected peak width". The expected peak width was estimated *via* the following procedure: 1) all the peaks higher than 3 mean coverage were detected 2) Points at which their coverage dropped below ½ of the maximal peak height were found and the distance between them was considered as a peak width 3) The "estimated peak width" was set equal to the median peak width. The convolution profile was scanned in order to find points were first derivative changes its sign from positive to negative. Each such point was considered as a potential peak and was assigned with a convolution score (that is convolution with second order Gaussian kernel centered at the peak position). Furthermore, we explored the distribution of the convolution scores. It appeared to resemble normal distribution, but with a heavy right tail. We assumed that this distribution is indeed bimodal of normal distribution (relatively low scores) representing 'noise' and a distribution of 'signal' (relatively high scores). We fit the Gaussian curve to the whole distribution (via optimize.fit function from SciPy package (Jones et al., 2001)) and set a score thresholds equal mean + 4 sigmas of the fitted distribution. Filtered peaks were normalized to allow inter-sample comparisons. Sum of coverages of the detected peaks was negated from the total genomic coverage. The resulting difference was used as normalization coefficient by which peak intensities were divided.

Whole Genome Sequencing

In order to get ideally comparable samples, for the whole genome sequencing of the input controls for the normalization of the ChAP-seq binding peak, samples were treated exactly as described for the ChAP-seq analysis, leaving out the protein purification step using the Strep-Tactin columns.

Proteome Analysis

The cell pellets were suspended in lysis buffer containing 50 mM potassium phosphate buffer (pH 8.0), 2 mM EDTA, 2 mM DTT and supplemented with complete protease inhibitor cocktail (1697498, Roche Applied Science, Basel, Switzerland). Cell suspensions were disrupted in a Precellys System (Bertin Instruments, Montigny-le-Bretonneux FRANCE) using 0.1 – 0.2 mm glass beads and two glass beads of 5mm for 3× 30 s at maximum frequency. The supernatant containing protein fractions were collected and frozen at -20 °C until analysis. Concentrations of proteins in crude extracts were measured using a Bradford assay (B6916, Sigma Aldrich, USA) with BSA as standard. The resulting crude extracts were then applied for untargeted LC-MS/MS measurements according to previously described methods (Voges and Noack, 2012) using an Infinity 1260 HPLC (Agilent Technologies) coupled to a Q-ToF 6600 mass spectrometer (Sciex, Darmstadt, Germany). Peptides were separated on an Ascentis column (Sigma Aldrich, Schnelldorf, Germany). The HPLC methods employed 0.1 % formic acid and 0.1 % formic acid in acetonitrile (Biosolve BV, Valkenswaard, The Netherlands) as buffers A and B, respectively, and consisted of the following gradient:

Total time (min)	Flow (l min $^{-1}$)	Rate (%B)
-12	200.00	3.0
0	200.00	3.0
70	200.00	40.0
78	200.00	40.0
79	200.00	60.0
89	200.00	60.0
90	200.00	3.0

Autosampler setting were ±6°C and injection volume was 10µL. The TripleTOF6600 was operated with CUR: 35, GS1: 30, GS2: 30, IS: 5500, TEM: 300, DP:120. During the elution under the previously specified parameters the variable width Q1 windows were monitored in a non-scheduled manner. SWATH window width was calculated with SWATH Variable Window Calculator_V1.0 (AB Sciex, Darmstadt, Germany). An information-dependent acquisition was performed on an injection from these samples using the described HPLC method. During the information-dependent acquisition all ions with *m/z* greater than 300, charge state 2-4 and above intensity of 150 were selected for fragmentation. Using ProteinPilot a library of

confidently identified peptides was generated for each sample. All peptides were assembled in a meta-library of *C. glutamicum* covering 1727 proteins. This library incorporates the peptide confidence after identification, peptide (precursor) intensity in the MS1 scan from the information-dependent acquisition, fragment ion intensities and the observed peptide retention time. For SWATH processing, the ProteinPilot library was imported into the MS/MS^{ALL} with SWATH^{TM} Acquisition MicroApp within PeakView. For 12 fragments each of the ten most intense peptides (according to the MS1 precursor ion intensity from the IDA acquisition) the XIC were extracted from the SWATH spectra within a time interval of 8 minutes around the expected peptide retention time from the IDA-acquisition. Each ion trace was integrated and scored as implemented in the SWATH^{TM} Acquisition MicroApp. By extracting a decoy sequence of each peptide and scoring the obtained fragment ion chromatograms analogously, a false discovery rate (FDR) could be calculated for each peptide. For the processing of the dataset presented in the Results and Discussion section, the false discovery rate threshold was set to 0.1 %. From all extracted 120 mass traces per protein only those with a false discovery rate less than the threshold value were exported and subjected to further analysis.

References

Al-Minawi, A. Z., Lee, Y. F., Håkansson, D., Johansson, F., Lundin, C., Saleh-Gohari, N., et al. (2009). The ERCC1/XPF endonuclease is required for completion of homologous recombination at DNA replication forks stalled by inter-strand cross-links. *Nucleic Acids Res.* 37, 6400–6413. doi:10.1093/nar/gkp705.

Albano, M., Smits, W. K., Ho, L. T. Y., Kraigher, B., Mandic-Mulec, I., Kuipers, O. P., et al. (2005). The rok protein of *Bacillus subtilis* represses genes for cell surface and extracellular functions. *J. Bacteriol.* 187, 2010–2019. doi:10.1128/JB.187.6.2010-2019.2005.

Ali, S. S., Xia, B., Liu, J., and Navarre, W. W. (2012). Silencing of foreign DNA in bacteria. *Curr Opin Microbiol* 15, 175–181. doi:10.1016/j.mib.2011.12.014.

Brune, I., Werner, H., Huser, A. T., Kalinowski, J., Puhler, A., and Tauch, A. (2006). The DtxR protein acting as dual transcriptional regulator directs a global regulatory network involved in iron metabolism of *Corynebacterium glutamicum*. *BMC Genomics* 7, 21. doi:10.1186/1471-2164-7-21.

Burkovski, A. (2013). Cell envelope of corynebacteria: structure and influence on pathogenicity. *ISRN Microbiol* 2013, 935736. doi:10.1155/2013/935736.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 49, 277–300. doi:10.1046/j.1365-2958.2003.03580.x.

Castang, S., and Dove, S. L. (2012). Basis for the essentiality of H-NS family members in *Pseudomonas aeruginosa*. *J. Bacteriol.* 194, 5101–9. doi:10.1128/JB.00932-12.

Chamakura, K. R., Sham, L.-T., Davis, R. M., Min, L., Cho, H., Ruiz, N., et al. (2017). A viral protein antibiotic inhibits lipid II flippase activity. *Nat. Microbiol.* 2, 1480–1484. doi:10.1038/s41564-017-0023-4.

Cooper, S., and Helmstetter, C. E. (1968). Chromosome replication and the division cycle of *Escherichia coli* Br. *J. Mol. Biol.* 31, 519–540. doi:10.1016/0022-2836(68)90425-7.

Couturier, E., and Rocha, E. P. C. (2006). Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* 59, 1506–1518. doi:10.1111/j.1365-2958.2006.05046.x.

Dame, R. T., Luijsterburg, M. S., Krin, E., Bertin, P. N., Wagner, R., and Wuite, G. J. L. (2005).

DNA bridging: a property shared among H-NS-like proteins. *J Bacteriol* 187, 1845–1848. doi:10.1128/jb.187.5.1845-1848.2005.

Dilweg, I. W., and Dame, R. T. (2018). Post-translational modification of nucleoid-associated proteins: an extra layer of functional modulation in bacteria? *Biochem. Soc. Trans.* 46, 1381–1392. doi:10.1042/bst20180488.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., et al. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359. doi:10.1126/science.aar4120.

Dupuis, M.-È., Villion, M., Magadán, A. H., and Moineau, S. (2013). CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. *Nat. Commun.* 4, 2087. doi:10.1038/ncomms3087.

Freiherr von Boeselager, R., Pfeifer, E., and Frunzke, J. (2018). Cytometry meets next-generation sequencing – RNA-Seq of sorted subpopulations reveals regional replication and iron-triggered prophage induction in *Corynebacterium glutamicum*. *Sci. Rep.* 8, 14856. doi:10.1038/s41598-018-32997-9.

Frunzke, J., Bramkamp, M., Schweitzer, J. E., and Bott, M. (2008). Population Heterogeneity in *Corynebacterium glutamicum* ATCC 13032 caused by prophage CGP3. *J Bacteriol* 190, 5111–5119. doi:10.1128/jb.00310-08.

Gordon, B. R. G., Li, Y., Cote, A., Weirauch, M. T., Ding, P., Hughes, T. R., et al. (2011). Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proc Natl Acad Sci U S A* 108, 10690–10695. doi:10.1073/pnas.1102544108.

Gordon, B. R. G., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., et al. (2010). Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* 107, 5154–5159. doi:10.1073/pnas.0913551107.

Hargreaves, K. R., Kropinski, A. M., and Clokie, M. R. (2014). Bacteriophage behavioral ecology: How phages alter their bacterial host's habits. *Bacteriophage* 4, e29866. doi:10.4161/bact.29866.

Helfrich, S., Pfeifer, E., Krämer, C., Sachs, C. C., Wiechert, W., Kohlheyer, D., et al. (2015). Live cell imaging of SOS and prophage dynamics in isogenic bacterial populations. *Mol*

Microbiol 98, 636–650. doi:10.1111/mmi.13147.

Hong, S. H., Wang, X., and Wood, T. K. (2010). Controlling biofilm formation, prophage excision and cell death by rewiring global regulator H-NS of *Escherichia coli*. *Microb. Biotechnol.* 3, 344–356. doi:10.1111/j.1751-7915.2010.00164.x.

Ikeda, M., and Nakagawa, S. (2003). The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl. Microbiol. Biotechnol.* 62, 99–109. doi:10.1007/s00253-003-1328-1.

Islam, S. T., and Lam, J. S. (2014). Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. *Can. J. Microbiol.* 60, 697–716. doi:10.1139/cjm-2014-0595.

Jochmann, N., Kurze, A.-K., Czaja, L. F., Brinkrolf, K., Brune, I., Hüser, A. T., et al. (2009). Genetic makeup of the *Corynebacterium glutamicum* LexA regulon deduced from comparative transcriptomics and in vitro DNA band shift assays. *Microbiology* 155, 1459–1477. doi:10.1099/mic.0.025841-0.

Jones, E., Oliphant, T., Peterson, P., and Al., E. (2001). SciPy: Open source scientific tools for Python, http://www.scipy.org. Available at: http://www.scipy.org [Accessed June 5, 2019].

Kaguni, J. M. (2011). Replication initiation at the *Escherichia coli* chromosomal origin. *Curr. Opin. Chem. Biol.* 15, 606–13. doi:10.1016/j.cbpa.2011.07.016.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., et al. (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol* 104, 5–25.

Keilhauer, C., Eggeling, L., and Sahm, H. (1993). Isoleucine synthesis in *Corynebacterium glutamicum*: Molecular Analysis of the *ilvB-ilvN-ilvC* Operon. *J. Bacteriol.* 175, 5595–5603. doi: .

Labrie, S. J., Samson, J. E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327. doi:10.1038/nrmicro2315.

Lamberte, L. E., Baniulyte, G., Singh, S. S., Stringer, A. M., Bonocora, R. P., Stracy, M., et al. (2017). Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat. Microbiol.* 2, 16249. doi:10.1038/nmicrobiol.2016.249.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–9. doi:10.1038/nmeth.1923.

Langmead, B., Wilks, C., Antonescu, V., and Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 35, 421–432. doi:10.1093/bioinformatics/bty648.

Li, C., Wally, H., Miller, S. J., and Lu, C.-D. (2009). The multifaceted proteins MvaT and MvaU, members of the H-NS family, control arginine metabolism, pyocyanin synthesis, and prophage activation in *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* 191, 6211–8. doi:10.1128/JB.00888-09.

Little, J. W. (2014). "Lysogeny, Prophage Induction, and Lysogenic Conversion," in *Phages*, 37–54. doi:10.1128/9781555816506.ch3.

Luo, H., and Gao, F. (2019). DoriC 10.0: an updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res.* 47, D74–D77. doi:10.1093/nar/gky1014.

Moran, N. A. (2004). Microbial Minimalism. *Cell* 108, 583–586. doi:10.1016/s0092-8674(02)00665-7.

Mott, M. L., and Berger, J. M. (2007). DNA replication initiation: Mechanisms and regulation in bacteria. *Nat. Rev. Microbiol.* 5, 343–354. doi:10.1038/nrmicro1640.

Nanda, A. M., Heyer, A., Krämer, C., Grünberger, A., Kohlheyer, D., and Frunzke, J. (2014). Analysis of SOS-Induced Spontaneous Prophage Induction in *Corynebacterium glutamicum* at the Single-Cell Level. *J. Bacteriol.* 196, 180–188. doi:10.1128/JB.01018-13.

Nanda, A. M., Thormann, K., and Frunzke, J. (2015). Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J Bacteriol* 197, 410–419. doi:10.1128/jb.02230-14.

Navarre, W. W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S. J., et al. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science (80-.).* 313, 236–238. doi:10.1126/science.1128794.

Ogino, H., Teramoto, H., Inui, M., and Yukawa, H. (2008). DivS, a novel SOS-inducible cell-division suppressor in *Corynebacterium glutamicum*. *Mol. Microbiol.* 67, 597–608. doi:10.1111/j.1365-2958.2007.06069.x.

Oshima, T., Ishikawa, S., Kurokawa, K., Aiba, H., and Ogasawara, N. (2006). *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res* 13, 141–153. doi:10.1093/dnares/dsl009.

Payne, K. M., and Hatfull, G. F. (2012). Mycobacteriophage endolysins: Diverse and modular enzymes with multiple catalytic activities. *PLoS One* 7, e34052. doi:10.1371/journal.pone.0034052.

Pfeifer, E. (2017). Impact and Regulatory Control of the CGP3 Prophage in *Corynebacterium glutamicum*.

Pfeifer, E., Hünnefeld, M., Popa, O., and Frunzke, J. (2019). Impact of Xenogeneic Silencing on Phage–Host Interactions. Academic Press doi:10.1016/j.jmb.2019.02.011.

Pfeifer, E., Hünnefeld, M., Popa, O., Polen, T., Kohlheyer, D., Baumgart, M., et al. (2016). Silencing of cryptic prophages in *Corynebacterium glutamicum*. *Nucleic Acids Res.* 44, 10117–10131. doi:10.1093/nar/gkw692.

Seid, C. A., Smith, J. L., and Grossman, A. D. (2017). Genetic and biochemical interactions between the bacterial replication initiator DnaA and the nucleoid-associated protein Rok in *Bacillus subtilis*. *Mol Microbiol* 103, 798–817. doi:10.1111/mmi.13590.

Smith, K. F., Bibb, L. A., Schmitt, M. P., and Oram, D. M. (2009). Regulation and Activity of a Zinc Uptake Regulator, Zur, in *Corynebacterium diphtheriae*. *J. Bacteriol.* 191, 1595–1603. doi:10.1128/JB.01392-08.

Smits, W. K., and Grossman, A. D. (2010). The transcriptional regulator Rok binds A+T-rich DNA and is involved in repression of a mobile genetic element in *Bacillus subtilis*. *PLoS Genet.* 6, e1001207. doi:10.1371/journal.pgen.1001207.

Stern, A., and Sorek, R. (2011). The phage-host arms race: Shaping the evolution of microbes. *BioEssays* 33, 43–51. doi:10.1002/bies.201000071.

Szybalski, W., and Iyer, V. N. (1964). Crosslinking of DNA By Enzymatically or Chemically Activated Mitomycins and Porfiromycins, Bifunctionally “Alkylating” Antibiotics. *Fed. Proc.* 23, 946–57.

Tendeng, C., Soutourina, O. A., Danchin, A., and Bertin, P. N. (2003). MvaT proteins in *Pseudomonas* spp.: a novel class of H-NS-like proteins. *Microbiology* 149, 3047–3050. doi:10.1099/mic.0.C0125-0.

Tomasz, M. (1995). Mitomycin C: small, fast and deadly (but very selective). *Chem. Biol.* 2, 575–579. doi:10.1016/1074-5521(95)90120-5.

Voges, R., and Noack, S. (2012). Quantification of proteome dynamics in *Corynebacterium glutamicum* by 15N-labeling and selected reaction monitoring. *J. Proteomics* 75, 2660–2669. doi:10.1016/J.JPROT.2012.03.020.

Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., et al. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.* 1, 147. doi:10.1038/ncomms1146.

Wennerhold, J., and Bott, M. (2006). The DtxR regulon of *Corynebacterium glutamicum*. *J Bacteriol* 188, 2907–2918. doi:10.1128/jb.188.8.2907-2918.2006.

Winardhi, R. S., Gulvady, R., Mellies, J. L., and Yan, J. (2014). Locus of enterocyte effacement-encoded regulator (Ler) of pathogenic *Escherichia coli* competes off histone-like nucleoid-structuring protein (H-NS) through noncooperative DNA binding. *J. Biol. Chem.* 289, 13739–13750. doi:10.1074/jbc.M113.545954.

Winardhi, R. S., and Yan, J. (2017). Applications of Magnetic Tweezers to Studies of NAPs. *Methods Mol Biol* 1624, 173–191. doi:10.1007/978-1-4939-7098-8_14.

Wolanski, M., Donczew, R., Zawilak-Pawlik, A., and Zakrzewska-Czerwinska, J. (2015). oriC-encoded instructions for the initiation of bacterial chromosome replication. *Front. Microbiol.* 6, 735. doi:10.3389/fmicb.2014.00735.

Young, R. (2014). Phage lysis: three steps, three choices, one outcome. *J. Microbiol.* 52, 243–58. doi:10.1007/s12275-014-4087-z.

3.4 The MarR-Type Regulator MalR Is Involved in Stress-Responsive Cell Envelope Remodeling in *Corynebacterium glutamicum*

Hünnefeld M., Persicke M., Kalinowski J., and Frunzke J.

Published in Frontiers in Microbiology, 2019

Contributor Role	Contributor
Conceptualization	MH (50 %), JF (50 %)
Formal Analysis	MH (95 %), MP (5 %)
Investigation/Experiments	MH (95 %), MP (5 %)
Methodology	MH (60 %), JF (30 %), MP (5 %), JK (5 %)
Project Administration	MH (50 %), JF (50 %)
Software	MH (100 %)
Supervision	MH (50 %), JF (50 %)
Visualization	MH (90 %), MP (10 %)
Writing – Original Draft Preparation	MH (90 %), JF (10 %)
Writing – Review & Editing	MH (50 %), JF (30 %), MP (10 %), JK (10 %)

Overall contribution MH: 90 %

Except of the phenotypic microarrays using an OmnilLog System (done by MP; Fig. 6), all of the presented experimental work and data analysis was done by MH. The analysis of cell size distribution using a Coulter Counter (Fig. 3D) was performed with the help of Dr. Holger Morschett from the Forschungszentrum Jülich. Data visualization was performed by MH, except of Fig. S6, which was visualized by MP. The revision and editing process was done by all authors together, with a large part (50 %) being taken over by MH. Furthermore, the writing of the original draft was mainly done by MH.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 21 May 2019
doi: 10.3389/fmicb.2019.01039

The MarR-Type Regulator MalR Is Involved in Stress-Responsive Cell Envelope Remodeling in *Corynebacterium glutamicum*

*Max Hünnefeld[1], Marcus Persicke[2], Jörn Kalinowski[2] and Julia Frunzke[1]**

[1] *Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich, Jülich, Germany,* [2] *Center for Biotechnology, Bielefeld University, Bielefeld, Germany*

It is the enormous adaptive capacity of microorganisms, which is key to their competitive success in nature, but also challenges antibiotic treatment of human diseases. To deal with a diverse set of stresses, bacteria are able to reprogram gene expression using a wide variety of transcription factors. Here, we focused on the MarR-type regulator MalR conserved in the *Corynebacterineae*, including the prominent pathogens *Corynebacterium diphtheriae* and *Mycobacterium tuberculosis*. In several corynebacterial species, the *malR* gene forms an operon with a gene encoding a universal stress protein (*uspA*). Chromatin affinity purification and sequencing (ChAP-Seq) analysis revealed that MalR binds more than 60 target promoters in the *C. glutamicum* genome as well as in the large cryptic prophage CGP3. Overproduction of MalR caused severe growth defects and an elongated cell morphology. ChAP-Seq data combined with a global transcriptome analysis of the *malR* overexpression strain emphasized a central role of MalR in cell envelope remodeling in response to environmental stresses. For example, prominent MalR targets are involved in peptidoglycan biosynthesis and synthesis of branched-chain fatty acids. Phenotypic microarrays suggested an altered sensitivity of a Δ*malR* mutant toward several β-lactam antibiotics. Furthermore, we revealed MalR as a repressor of several prophage genes, suggesting that MalR may be involved in the control of stress-responsive induction of the large CGP3 element. In conclusion, our results emphasize MalR as a regulator involved in stress-responsive remodeling of the cell envelope of *C. glutamicum* and suggest a link between cell envelope stress and the control of phage gene expression.

Keywords: MarR-type regulator, *C. glutamicum*, cell envelope, stress response, antibiotics, cell wall

OPEN ACCESS

Edited by:
Hari S. Misra,
Bhabha Atomic Research Centre, India

Reviewed by:
Masayuki Inui,
Research Institute of Innovative Technology for the Earth, Japan
Alicja Węgrzyn,
Polish Academy of Sciences, Poland

****Correspondence:***
Julia Frunzke
j.frunzke@fz-juelich.de

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *18 February 2019*
Accepted: *25 April 2019*
Published: *21 May 2019*

Citation:
Hünnefeld M, Persicke M, Kalinowski J and Frunzke J (2019) The MarR-Type Regulator MalR Is Involved in Stress-Responsive Cell Envelope Remodeling in Corynebacterium glutamicum. Front. Microbiol. 10:1039. doi: 10.3389/fmicb.2019.01039

INTRODUCTION

In almost every natural habitat, a high number of microbial species coexist and compete for space and nutrients. Consequently, the exposure to bacteriostatic or bactericidal compounds (antibiotics) represents a routine challenge, which bacteria are facing in various ecological niches, and particularly during infection of a specific host (Chao and Levin, 1981; Peschel, 2002; Stubbendieck and Straight, 2016). MarR-type transcriptional regulators constitute a prominent

family of transcription factors involved in the reprogramming of gene expression in response to stress conditions (Wilkinson and Grove, 2006; Deochand and Grove, 2017). Already decades ago, clinical isolates of *Escherichia coli* displaying a multiple antibiotic resistance phenotype where found to carry mutations in the *marR* locus (George and Levy, 1983) and subsequently drew considerable attention to this ubiquitously found class of regulators. Following studies then showed that *E. coli* MarR is a transcriptional repressor of genes conferring resistance toward different antibiotics, organic solvents and lipophilic, mainly phenolic compounds (Alekshun and Levy, 1999). In further studies, it was shown that MarR-type regulators are widely distributed among bacteria and archaea, likely representing an ancient regulator family which emerged before the evolutionary split of these domains more than three billion years ago (Pérez-Rueda and Collado-Vides, 2001; Pérez-Rueda et al., 2004). Overall, the regulatory responses modulated by MarR-type regulators were grouped into three general categories (Wilkinson and Grove, 2006), including (i) environmental stress responses (e.g., triggered by antibiotics) (Poole et al., 1996; Srikumar et al., 2000; Spory et al., 2002), (ii) regulation of virulence genes (Lee et al., 2003; Rouanet et al., 2004), and (iii) degradation of lipophilic (often aromatic) compounds (Providenti and Wyndham, 2001; Galán et al., 2003). The DNA-binding domain of MarR-family regulators is typically comprised of a winged helix-turn-helix domain, recognizing palindromes, or inverted repeats (Grove, 2013). In the classical scenario, the dissociation of the MarR dimer from its genetic target is triggered by ligand binding [e.g., antibiotics, salicylates, and lipophilic compounds (Kumarevel, 2012)], but examples also exist where the binding of ligands fosters the association to DNA targets (Egland and Harwood, 1999; Providenti and Wyndham, 2001).

The suborder of the *Corynebacterineae* covers several prominent pathogenic species, such as *Corynebacterium diphtheriae*, *Mycobacterium tuberculosis*, and *Mycobacterium leprae*, causing millions of deaths every year. Species of this suborder share a very similar and unique cell wall composition hampering antibiotic treatment (Ortalo-Magne et al., 1995; Daffé and Draper, 1997; Zuber et al., 2008; Mishra et al., 2011; Marrakchi et al., 2014). In addition to the peptidoglycan, cells are surrounded by an arabinogalactan zone topped by a lipid bilayer composed of long-chain α-alkyl, β-hydroxy fatty acids – the mycolic acids (Eggeling et al., 2008).

In this study, we have characterized the function of the MarR-type regulator MalR (Cg3315) of the non-pathogenic, Gram-positive model organism *C. glutamicum* (Kalinowski et al., 2003), which – in total – harbors nine MarR-type regulators (Brune et al., 2005). Further, the genome of *C. glutamicum* contains a large prophage element (CGP3), which was shown to be inducible by the cellular SOS response (Nanda et al., 2014), or excises spontaneously in a small fraction of wild type cells (Frunzke et al., 2008; Helfrich et al., 2015).

Corynebacterium glutamicum MalR was previously reported as a repressor of the *malE* gene, encoding the malic enzyme (Krause et al., 2012). Here, we performed a genome-wide profiling of MalR targets by combining ChAP-Seq and a comparative transcriptomics approach. As revealed by phenotypic microarrays, a mutant lacking the *malR* gene displayed an impaired resistance toward different β-lactam antibiotics. The majority of former studies focused on a very distinct operon or small regulon controlled by MarR-type regulators. The present study provides – for the first time – a comprehensive insight into the complex regulon of MalR, which is involved in the remodeling of the cell envelope in response to stress conditions. Interestingly, our data also suggest a role of MalR in the control of the large cryptic prophage element CGP3 and thereby demonstrate a complex regulatory interaction between the host and horizontally acquired elements.

RESULTS

The MarR-Type Regulator MalR Is Conserved in *Corynebacteria* and Mycobacteria

The MarR-type regulator MalR (Cg3315) was previously described as a repressor of the malic enzyme gene in *C. glutamicum* (Krause et al., 2012). Sequence analysis revealed that MalR is conserved in several coryne- and mycobacterial species, also including the prominent pathogens *C. diphtheriae* (57% sequence identity) and *M. leprae* (40% sequence identity). Simulated secondary structures of MalR using Phyre[2] disclosed a high similarity to the secondary structure of MarR from *E. coli* consisting of six α-helices surrounding two β-sheets (Alekshun et al., 2001), although the amino acid sequence identity is only 22% (**Supplementary Figure S1**).

In the genome of *C. glutamicum* ATCC 13032, *malR* is organized in an operon with a gene encoding a universal stress protein (*uspA*) and is divergently located to a small hypothetical protein, followed by an operon coding for a penicillin-binding protein and two putative membrane proteins (**Figure 1**; Pfeifer-Sancar et al., 2013). This genomic ensemble emphasizes a role of MalR in global stress responses and potentially cell envelope-related functions. The genome-wide analysis of MalR target genes and its physiological impact is the aim of the present study.

Genome-Wide Profiling of MalR Target Genes

In order to identify target genes of MalR, ChAP-Seq analysis was performed and selected target promoters were subsequently verified using electrophoretic mobility shift assays (EMSA) (**Figure 2**). To produce MalR at physiological levels, a gene fusion was integrated at the *malR* locus into the *C. glutamicum* ATCC 13032 chromosome, encoding a C-terminally strep-tagged variant of MalR. Cells were grown in CGXII minimal medium with 2% glucose and harvested in the mid-exponential phase. The sequencing of DNA bound to MalR under the chosen conditions revealed 66 binding regions in total (**Figure 2A** and **Supplementary Table S1**). Remarkably, 13 target regions of MalR were found inside the cryptic prophage element CGP3, showing a local maximum in the region cg1895–cg1950 (8 peaks). Besides several genes of unknown function, MalR bound to promoter regions of genes involved in cell envelope biosynthesis,

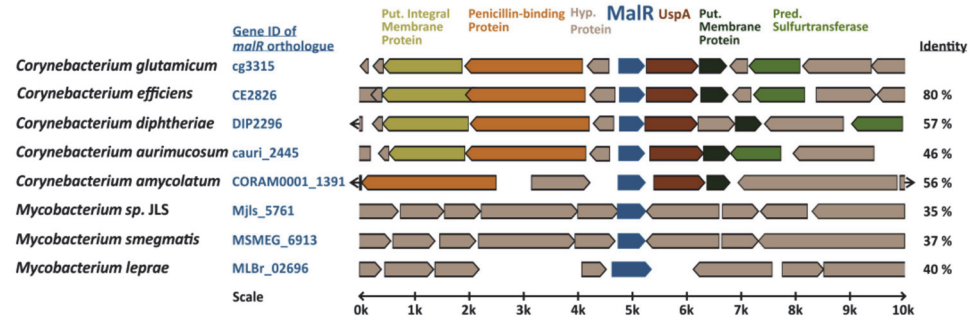


FIGURE 1 | Genomic organization of the *malR* gene in coryne- and mycobacterial species. Amino acid sequence identity to the *C. glutamicum* MalR ortholog is given in the right column. The genomic context of *C. glutamicum malR* was extracted from microbesonline (http://microbesonline.org).

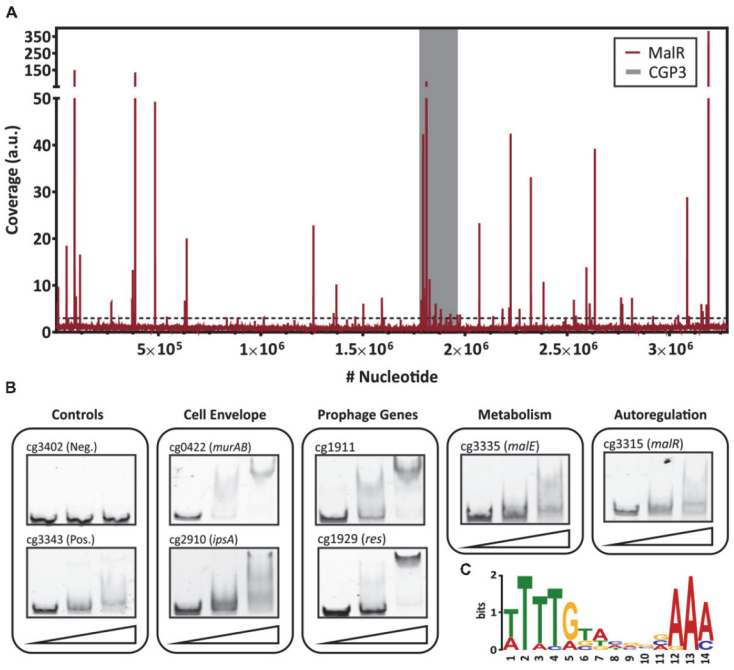


FIGURE 2 | Genome-wide binding profile and *in vitro* verification of binding sites of MalR. **(A)** ChAP-Seq experiments revealed 66 distinct binding peaks of MalR in the genome of *C. glutamicum* ATCC 13032. For ChAP-Seq analysis, a strain containing a genomically encoded Strep-tagged variant of MalR was grown in CGXII medium containing 2% (w/v) glucose for 5 h and treated as described previously (Pfeifer et al., 2016). The coverage of MalR-bound DNA regions (red) was normalized using a rolling mean with a step size of 10 bp and a window size of 50 bp. The gray bar marks the CGP3 prophage region inside of the genome of *C. glutamicum*. Overall, 66 peaks with coverages higher than threefold mean coverage were detected. **(B)** Electrophoretic mobility shift assays (EMSAs) were performed to verify binding of MalR to promoter regions identified via ChAP-Seq. Therefore, 100 bp DNA fragments (50 bp up- and downstream the peak maximum) were incubated without protein (first lane), with 3 molar excess (228 nM, second lane), and 10 molar excess of purified MalR (760 nM, third lane). A complete overview on all tested fragment is shown in **Supplementary Figure S2**. **(C)** DNA sequences of 16 binding sites that were verified using EMSAs were used to deduce a possible binding motif of MalR with the online tool MEME-ChIP (Bailey et al., 2009). A motif based on all peaks obtained by ChAP-Seq, as well as the distribution of the motifs within the uploaded sequences, are shown in **Supplementary Figure S3**.

including *embC* (Alderwick et al., 2005), the *murAB* operon (Alderwick et al., 2015) and *ipsA* (Baumgart et al., 2013). It further associates to promoter regions of genes encoding proteins involved in transport mechanisms, such as *oppA*, cg1454, cg2256, and cg2340.

Conspicuously, 16 peaks were detected in the promoter region of genes coding for (putative) secreted proteins. Consistent with the report of Krause et al. (2012), also binding to the promoter region of *malE* was confirmed by our study. Furthermore, a significant binding peak was observed in the own promoter region of the *malR-uspA* operon, indicating an auto regulation of *malR* expression. In summary, this ChAP-Seq analysis revealed a global role of MalR in the regulation of genes involved in cell envelope-related functions and suggested a regulatory interaction of MalR with the large prophage CGP3.

To validate the obtained binding profile of MalR, EMSAs were performed using different promoter regions identified by ChAP-seq analysis (**Figure 2B**). Except one potential target promoter (cg2962), every tested candidate could be verified using this *in vitro* approach (**Supplementary Figure S2**). As a negative control, the promoter region of cg3402 (a putative copper chaperone) was used. Here, no shift was detectable. *In vitro*, different migration patterns were observed for the tested MalR targets, which likely reflect differences in binding affinities and/or the presence of multiple DNA motifs. Furthermore, in some cases, additional factors may contribute to *in vivo* MalR-DNA association (e.g., in the case of cg2962).

Using the 66 MalR peak sequences, a putative binding motif of MalR was deduced using the online tool MEME-ChIP (Bailey et al., 2009). This tool predicted a very AT-rich palindromic binding motif found in all peaks (motif and distribution in **Supplementary Figure S3**), which is very similar to the motif found in MalR targets verified with EMSAs (**Figure 2C**).

Overproduction of MalR Causes Severe Growth Defects

In a next step, we compared the growth of the *C. glutamicum* wild type with a *malR* deletion strain (Δ*malR*) and a strain overexpressing the *malR* gene under control of the IPTG-inducible P_{tac} promoter (**Figure 3**). In fact, overexpression of *malR* caused a severe growth defect of *C. glutamicum* grown on CGXII minimal medium with 2% glucose (**Figure 3A**), whereas the deletion mutant had only a minor impact on the growth rate compared to the wild type strain under the tested conditions (**Figure 3B**).

Fluorescence microscopy of cells stained with NileRed (lipid components) and Hoechst 33342 (DNA) revealed a heterogeneous morphology of cells overexpressing the *malR* gene. Among cells with wild type cell shape, several cells displayed a significantly elongated cell morphology upon *malR* overexpression (**Figure 3C**). The deletion mutant, however, was indistinguishable from the wild type strain (**Figure 3C**). Furthermore, overexpression of *malR* resulted in an uneven distribution of the lipid fraction as revealed by Nile red staining. The cloudy and heterogeneous distribution of the Hoechst strain also pointed toward problems regarding nucleoid condensation and segregation in the strain overexpressing *malR*. In order to quantify the observed heterogeneous morphology of cells, culture samples were further analyzed using a MultiSizer 3 particle counter (**Figure 3D**). These data show a clear shift of the cells toward an increased cell volume.

The Impact of Altered MalR Levels on Transcription

The multitude of MalR-bound regions identified by ChAP-Seq analysis and the severe morphological changes caused by overexpression of *malR* already suggest a significant impact of MalR on the transcriptomic landscape of *C. glutamicum*. In the following research, we performed a comparative transcriptome analysis of the wild type, containing the empty vector pEKEx2, and the *malR* overexpressing strain, using DNA microarrays. For this purpose, both strains were grown in CGXII minimal medium and harvested in the early exponential phase. Additionally, we verified the obtained data using qRT-PCR with some selected samples (**Supplementary Figure S4**). As illustrated in **Figure 4**, *malR* overexpression resulted in massive changes in the global transcriptome when compared to the wild type. Overall, 170 genes showed a more than fivefold altered mRNA level (p-value < 0.05). A complete overview on the transcriptome analysis is provided in **Supplementary Table S2**. In contrast, deletion of the *malR* gene had only a minor impact under the tested conditions (**Supplementary Table S2**).

Considering the impact of increased MalR levels on growth and cell morphology, a majority of these effects are likely due to the result of secondary effects. To focus on primary targets of MalR, we analyzed the impact on the expression of genes whose promoter was directly bound by MalR as found *via* ChAP-Seq analysis (selection shown in **Table 1**; for a complete overview see **Supplementary Tables S1**, **S2**). In fact, many of the direct target genes of MalR revealed an altered mRNA level due to the overexpression of *malR*. In several cases, however, the effect was rather minor, which may be due to the chosen growth conditions lacking a specific effector molecule of the regulator. Furthermore, the majority of target genes is likely controlled by several regulatory systems affecting the transcriptional output.

Remarkably, many genes encoding proteins involved in cell envelope biosynthesis or remodeling were affected by *malR* overexpression. For example, the *ipsA* gene, encoding a LacI-type regulator, showed about a fivefold downregulation in the *malR* overexpression strain. IpsA was previously described as an important regulator modulating the synthesis of inositol-derived lipids in the cell wall of *C. glutamicum* (Baumgart et al., 2013). Cells lacking *ipsA* revealed an elongated cell morphology with an affected growth. These findings are in line with the phenotype of the MalR overproducing strain (**Figure 3C**). Among the genes repressed by MalR is the *embC* gene, encoding an arabinosyltransferase involved in arabinan biosynthesis (Belanger et al., 1996; Alderwick et al., 2005), and several (secreted) membrane proteins of unknown function (cg0623, cg0636, cg0879, cg0952, cg1578, cg1910, and cg3322). The most distinct downregulation was observed for

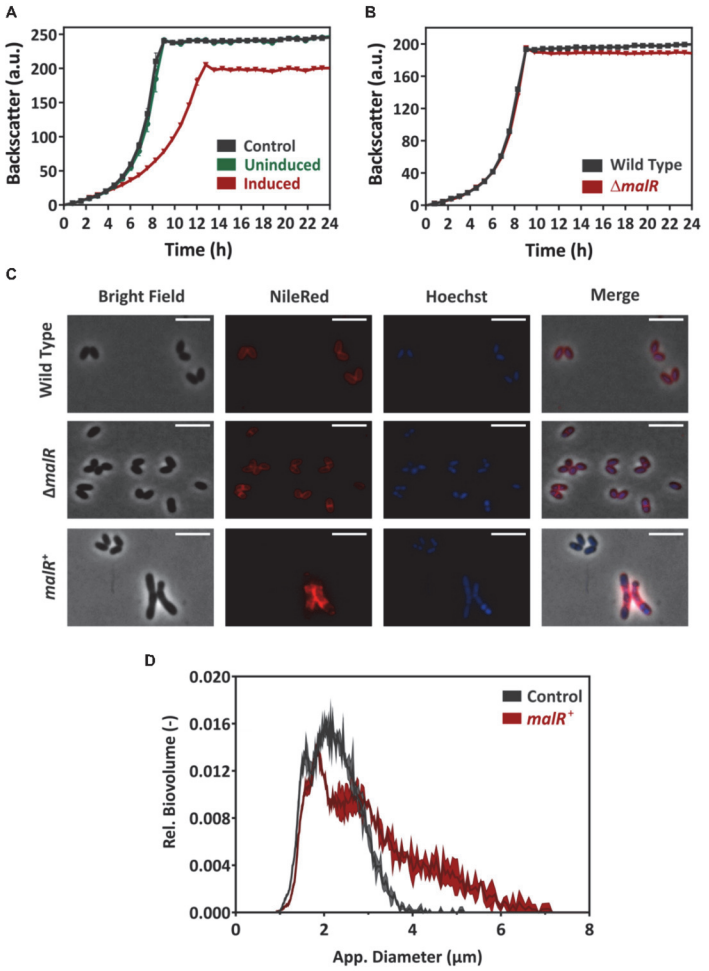


FIGURE 3 | MalR overproduction causes severe growth defects of *C. glutamicum*. **(A)** Comparative growth experiment of *C. glutamicum* ATCC 13032 carrying the empty vector pEKEx2 and strain *C. glutamicum*/pEKEx2-*malR* overexpressing the *malR* gene. Cells were cultivated in CGXII minimal medium containing 2% (w/v) glucose with ("induced") and without ("uninduced") 25 µM IPTG using a microbioreactor cultivation system. **(B)** Growth of *C. glutamicum* ATCC 13032 in comparison with the strain lacking the *malR* gene. **(C)** For microscopic analysis, cells were grown in CGXII medium for 24 h at 30°C. Shown are wild-type *C. glutamicum* ATCC 13032 cells (row 1), *C. glutamicum* Δ*malR* cells (row 2), and *C. glutamicum* ATCC 13032 cells carrying the over expression vector pEKEx2-*malR* (row 3). The expression of *malR* is induced by the addition of 100 µM IPTG. Lipid components of the cell membrane were stained with Nile red (red); DNA was stained with Hoechst 33342 (blue). The scale bars represent 5 µm. Further microscopic pictures of cells overproducing MalR are shown in **Supplementary Figure S5**. **(D)** To quantify the number of cells with an altered morphology, cell counts and biovolume were analyzed using a MultiSizer 3 particle counter (Beckman Coulter, Brea, CA, United States) equipped with a 30 µm capillary in volumentric control mode. Cells were grown as described for **(C)**.

the *malE* gene, encoding the malic enzyme which catalyzes the decarboxylation of malate to pyruvate. Pyruvate itself is a precursor for acetyl-CoA synthesis, which is also required for fatty acid synthesis.

Among the genes showing a slightly increased mRNA level in response to *malR* expression, we also found the *ilvA* gene, encoding a threonine-dehydratase that is necessary for the production of isoleucine (Sharma et al., 2016). Isoleucine is

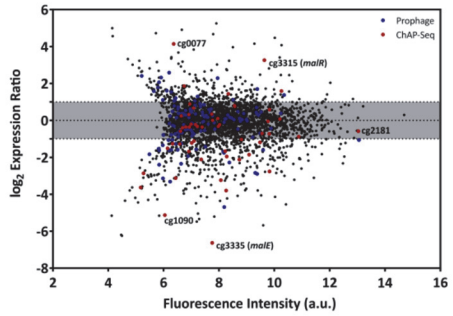


FIGURE 4 | Overexpression of the *malR* gene causes global changes in the *C. glutamicum* transcriptome. Comparative transcriptome analysis of *C. glutamicum* ATCC 13032 cells carrying the overexpression vector pEKEx2-*malR* and a strain carrying the empty plasmid. Cells were cultivated in CGXII glucose minimal medium and harvested in the early exponential growth phase. Shown is an MA-plot where the log2 of the expression ratio is plotted against the fluorescence intensity of the single spots. Red dots indicate genes that were bound by MalR in the ChAP-Seq experiment (**Figure 2**).

a branched chain amino acid and, together with acetyl-CoA, an important precursor for the generation of branched chain fatty acids, which are part of the bacterial cell membrane. Furthermore, the *oppA* gene was slightly upregulated, which codes for an oligopeptide permease required for the modulation of cell-wall associated lipids as well as mycolic acids (Flores-Valdez et al., 2009). The expression level of methyltransferase *mraW* was about threefold increased. In *E. coli,* MraW was described to play an important role during cell division (Carrión et al., 1999).

Among the direct targets of MalR, we also found the gene *murB*, which is involved in the synthesis of peptidoglycan building blocks by converting UDP-N-acetylglucosamine partially to UDP-N-acetylmuramic acid (Burkovski, 2013). However, its mRNA level was almost unaffected by *malR* overexpression, suggesting that further regulatory components are involved in the control of the *murAB* operon. Altogether, ChAP-Seq analysis, the impact of MalR on cell morphology and this transcriptomic study strongly emphasize an important role of MalR in the remodeling of the cell envelope.

MalR Affects the Cell Surface Structure of *C. glutamicum*

Considering the impact of MalR on cell morphology (**Figure 3C**), we analyzed cells overproducing MalR using transmission electron microscopy (TEM) and scanning electron microscopy (SEM) (**Figure 5**). Both approaches suggested differences in the cell surface structure. While wild type cells show a rather homogenous distribution in size, the strain overexpressing *malR* displayed an elongated cell morphology and significant heterogeneity, with regard to cell size (**Figures 3C**, **5**). Moreover, the overall cell surface structure appeared smoother. The fuzzy structure observed by TEM is also typical for the outer layer of the mycobacterial envelope (Zuber et al., 2008). This electron-dense layer is supposed to consist of a protein-carbohydrate matrix with only a few lipids (Daffé and Draper, 1997).

MalR Confers Increased Resistance Toward Cephalosporin Antibiotics

For a better understanding of the physiological impact of MalR, we performed phenotype microarrays of the wild type and the Δ*malR* mutant using a Biolog system. Here, we focused on the plates PM1 and PM2A (carbon sources), PM4 (phosphorus and sulfur sources), PM9 (osmolytes), PM10 (pH), and PM11-PM13 (antibiotics). The only additives that led to a different behavior between the wild type and the *malR* deletion strain were different antibiotics. To be precise, different β-lactams, tetracyclines and other examples of different substance classes revealed an altered metabolic activity of the Δ*malR* mutant (**Supplementary Table S4**). **Figure 6** shows two examples, emphasizing a significantly increased sensitivity of the mutant toward different cephalosporins. Compared to the wild type, which was able to tolerate moderate levels of the antibiotics cefazolin (0.58 μg/ml) and cephalothin (6 μg/ml), the Δ*malR* strain was significantly affected, and did not restore metabolic activity within 40 h under the tested conditions. This phenotype was successfully complemented by plasmid-encoded MalR (*C. glutamicum* Δ*malR*/pEKEx2-*malR*) when the strain was compared to the empty vector control (*C. glutamicum*/pEKEx2). An overview of all tested plates is provided in **Supplementary Figure S6**. Remarkably, some changes in the antibiotic tolerance profile could also be attributed to the presence of kanamycin used as selection marker of the respective plasmid. An additional growth experiment with the *malR* deficient strain, harboring either the empty vector or the plasmid pEKEx2-*malR*, was performed to verify the complementation (**Supplementary Figure S7**).

MalR Counteracts SOS-Dependent Induction of the CGP3 Prophage

Due to several binding sites inside the CGP3 region, an impact of MalR on the inducibility of this large cryptic prophage was the focus of further experiments. For this purpose, the reporter strain *C. glutamicum* ATCC 13032::P_{lys}-*eyfp* carrying the *malR* overexpression plasmid pEKEx2-*malR* was used. The P_{lys}-*eyfp* reporter enables the visualization of prophage induction within single cells by the production of the fluorescent protein eYFP (Helfrich et al., 2015). In the following, we triggered an induction of the cellular SOS response by the addition of the DNA-damaging agent mitomycin C (MMC) and monitored its impact on CGP3 induction. The MalR level was modulated by adding increasing amounts of IPTG (10, 25, and 50 μM). Remarkably, the fraction of CGP3 induced cells significantly declined in response to *malR* overexpression (**Figure 7B**). Also, the growth of the strains was severely affected upon addition of MMC and IPTG (**Figure 7A**). The dose responsive behavior of prophage induction in response to *malR* overexpression suggested that MalR counteracts prophage excision under the tested conditions.

TABLE 1 | Genes bound by MalR with altered expression due to an overexpression of *malR*.

Gene locus	Gene name	Annotation	*malR*$^+$	Binding site
cg3315	*malR*	Transcriptional regulator, repressor of the malic enzyme gene *malE*, and MarR-family	9.75	−90
cg3344		Putative nitroreductase	5.47	87
cg2377	*mraW*	S-adenosyl-methyltransferase	2.92	95
cg2903		Putative protein, conserved	2.75	−179
cg2430		Hypothetical protein	2.21	43
cg0420 (IG)		Putative glycosyltransferase, horizontally transferred	1.72	−733
cg3304	*dnaB*	Replicative DNA helicase, maybe involved in folate or lipopolysaccharide biosynthesis, starch, and sucrose metabolism	1.55	−111
cg0304		Putative membrane protein	1.36	
cg1978		Hypothetical protein, CGP3 region	1.34	
cg0423	*murB*	UDP-N-acetylenolpyruvoylglucosamine reductase, horizontally transferred	1.33	
cg1912		Hypothetical protein, CGP3 region	0.72	
cg0111		Hypothetical protein	0.68	
cg3057		Putative secreted protein	0.64	−45
cg0712		Putative secreted protein	0.54	
cg0235	*embC*	Arabinosyltransferase	0.48	20
cg1909		Hypothetical protein, CGP3 region	0.32	
cg0009		Putative membrane protein	0.32	
cg0722 (IG)		Putative drug exporter, RND superfamily, and terpenoid synthesis	0.31	
cg1967		Hypothetical protein, CGP3 region	0.29	
cg2500	*znr*	Putative transcriptional regulator, ArsR-family	0.27	−26
cg0120		Putative esterase/lipase/thioesterase-family protein, and hydrolase	0.26	
cg1456		Putative signal-transduction protein containing cAMP-binding and CBS domain, conserved	0.24	85
cg2910	*ipsA*	Inositol-phosphate-synthase activator, LacI-family	0.23	8
cg2033		Putative secreted protein, CGP3 region	0.26	
cg2256		Putative ABC-type multidrug/daunorubicin transport system, ATPase component	0.20	3
cg1929	*res*	Resolvase,-family recombinase, CGP3 region	0.18	
cg1905		Hypothetical protein, CGP3 region	0.16	83
cg3219	*ldhA*	NAD-dependent L-lactate dehydrogenase	0.11	−65
cg1911		Putative secreted protein, CGP3 region	0.08	−92
cg2610		Putative ABC-type dipeptide/oligopeptide/nickel transport system, secreted component	0.07	
cg1577		Putative secreted hydrolase	0.07	−78
cg1090	*ggtB*	γ-glutamyltranspeptidase precursor PR	0.02	−50
cg3335	*malE*	Malic enzyme (NADP+)	0.01	

Extraction of data shown in ***Figures 2***, ***4****. The binding site is calculated from TSS (Pfeifer-Sancar et al., 2013) and the maximum peak position. malR*$^+$ *indicates the fold-change of the mRNA caused by an overexpression of malR (p-values < 0.05). IG, intragenic binding.*

DISCUSSION

With this study, we provided comprehensive insights into the complex regulon of the MarR-type regulator MalR in *C. glutamicum*. In the last few decades, members of this regulator family were rewarded with considerable attention as some MarR proteins were shown to contribute to a so-called multiple antibiotic resistance phenotype (Alekshun and Levy, 1999; Deochand and Grove, 2017). In several cases, MarR-type regulators were described to control a small set of target genes, often located in the same operon or in divergent orientation to the regulator gene on the chromosome (Alekshun and Levy, 1999). The resulting phenotype of increased antibiotic resistance was previously proposed to be a result of decreased influx and increased efflux of the toxic compound. In this study, we now provided a genome scale profiling of MalR binding and identified more than 60 promoter regions bound by this regulator. A combination of ChAP-Seq analysis and comparative transcriptomics emphasizes MalR as a global regulator of stress-responsive remodeling of the cell envelope. Remarkably, MalR is conserved in several species of the genera *Corynebacterium* and *Mycobacterium*, also including prominent pathogens like *C. diphtheriae* and *M. leprae*. Conspicuously, the genomic organization of the *malR* locus in *C. diphtheriae* is almost identical to *C. glutamicum*, where *malR* forms an operon with a gene (*uspA*) encoding *uspA* (Kalinowski et al., 2003). The superfamily of Usp proteins comprises a large group of conserved proteins that can be found in all domains of life (Kvint et al., 2003). In *M. bovis* BCG, the tuberculosis vaccine strain, overexpression of a particular Usp led to an increased susceptibility of the cells toward the anti-tuberculosis drug isoniazid (Hu et al., 2015). This

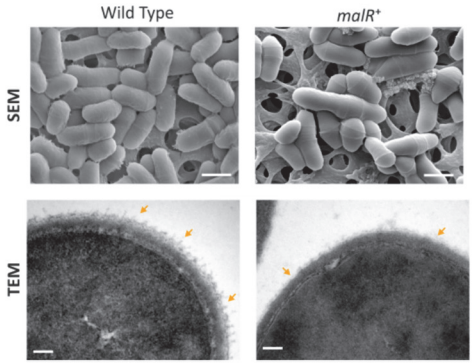


FIGURE 5 | *C. glutamicum* cells overproducing MalR show an altered cell surface structure. Shown are SEM and TEM microscopy pictures of wild type cells and cells overproducing the MalR protein. SEM pictures are 15000 x magnified; TEM pictures 167000×. For microscopic analysis, cells were cultivated in CGXII minimal medium containing 2% (w/v) glucose and *malR* expression was induced by adding 100 μM IPTG. Cells were fixed using 3% Glutaraldehyde in Sörensen phosphate buffer.

conserved genomic organization of the *malR* locus in these species is in favor with a similar role of *C. diphtheriae* MalR in cell envelope remodeling and antibiotic resistance in this important human pathogen.

The effector molecule of MalR has not yet been identified, but based on our findings, we can speculate that MalR binding is affected by one or several antibiotics and/or lipophilic compounds causing cell envelope stress. Interestingly, the results of our phenotypic microarrays revealed a clear difference in respiration of the wild type and the *malR* deficient strain when cephalosporins of the first generation (Cephalothin, Cefazolin) or the second generation (Cefuroxime) were added to the medium. A link between cephalosporins and MalR is provided by cg3313, encoding a penicillin-binding protein, which is located in divergent orientation to the *malR-uspA* operon (**Figure 1**). Penicillin-binding proteins are involved in the cell wall synthesis, to be more precise in the peptidoglycan synthesis (Valbuena et al., 2007), and are in general targets of different β-lactam antibiotics including cephalosporins (Fontana et al., 2000). Upon overexpression of *malR*, four (out of nine in total) genes encoding penicillin-binding proteins showed a more than twofold change in transcription (**Supplementary Table S2**). Nevertheless, only the first and second generation cephalosporins showed an antibiotic impact on the tested strains. Another antibiotic of the third generation of cephalosporins (Ceftriaxone) did not have any effect on the respiration of both strains. Considering the history of cephalosporin development, the described effect can be elucidated: while first and second-generation cephalosporins were mainly active against Gram-positive bacteria, with the third generation an R1 methoxy substitution changed the specificity more toward Gram-negative bacteria, which simultaneously reduced the effect of those cephalosporins toward Gram-positive penicillin-binding proteins (Craig and Andes, 2015).

The genome of *C. glutamicum* comprises in total nine MarR-type transcriptional regulators; four of which were already characterized in former studies. Except for MalR, no impact on antibiotic resistance phenotype for any of the previously studied examples has been reported so far. RosR, which constitutes a hydrogen peroxide sensitive regulator, was shown to play an important role in the oxidative stress response of *C. glutamicum* (Bussmann et al., 2010). The MarR-type regulator PhdR was shown to act as a repressor of the *phd* gene cluster important for phenylpropanoid utilization in *C. glutamicum*. Here, phenylpropanoids or their degradation intermediates were shown to cause dissociation of PhdR and de-repression of the respective target operon (Kallscheuer et al., 2016).

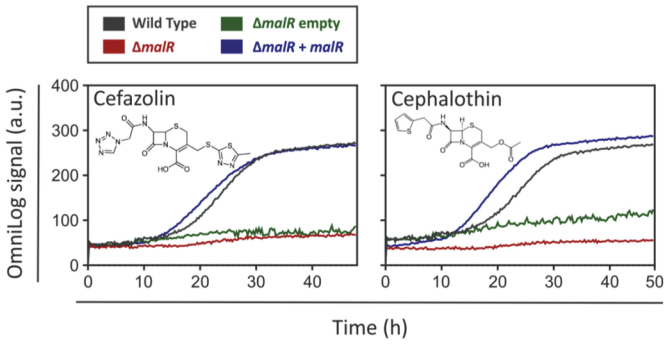


FIGURE 6 | The Δ*malR* mutant strain shows a higher sensitivity toward cephalosporins in phenotypic microarrays. An OmniLog System from Biolog (Hayward, CA, United States) was used to perform phenotypic microarrays with wild type *C. glutamicum* ATCC 13032 as well as the *malR* deletion strain. The experiments were conducted like described in the protocol of the manufacturer (Bochner et al., 2001). One group of compounds that led to a difference in respiration of wild type cells and Δ*malR* cells were the cephalosporins (Cefazolin: 0.58 μg/ml; Cephalothin: 6 μg/ml).

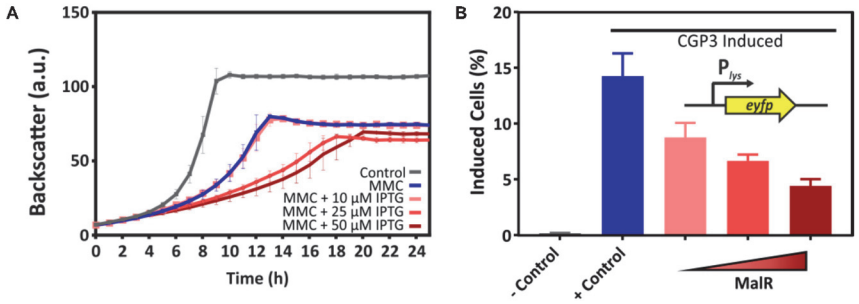


FIGURE 7 | Overexpression of *malR* counteracted SOS-dependent prophage induction. *C. glutamicum* ATCC 13032::P_{lys}-*eyfp* cells containing the plasmid pEKEx2-*malR* were cultivated in CGXII minimal medium containing 2% (w/v) glucose (+25 µM Kanamycin) using a microbioreactor cultivation system. CGP3 induction was triggered by addition of 600 nM mitomycin C (MMC). For the negative control, no MMC was added. Additionally, different concentrations of IPTG (0–50 µM) were used to induce expression of *malR*. Both, MMC and IPTG, were added directly at the beginning of the cultivation (0 h). Under the applied conditions, we analyzed the strain with regard to growth **(A)**. Furthermore, the fraction of CGP3-induced cells was assessed in the stationary phase (after ~25 h of cultivation) using an Accuri C6 flow cytometer **(B)**.

Finally, the isoprenoid pyrophosphate-dependent regulator CrtR was recently described as being involved in the regulation of carotenoid biosynthesis and thus represents an example for a rather specialized MarR-type regulator (Henke et al., 2017). MalR itself was firstly reported by Krause et al. as a repressor of the *malE* gene, encoding an $NADP^+$ - dependent malic enzyme in *C. glutamicum* (Krause et al., 2012). This role is supported by our study, where *malE* was among the genes most affected by MalR overexpression. MalE catalyzes the decarboxylation of malate to pyruvate while generating NADPH (Gourdon et al., 2000). Pyruvate, a precursor for acetyl-CoA synthesis, as well as NADPH as a reducing agent, are required for fatty acid biosynthesis (Cronan and Thomas, 2009). For different oleaginous microorganisms, it is known that malic enzymes play a crucial role in lipid generation (Li et al., 2013; Zhang et al., 2013; Ratledge, 2014). For example, two malic enzymes were recently shown to be important for triacylglycerol and antibiotic production in *Streptomyces coelicolor* (Rodriguez et al., 2012).

Genome-wide profiling of MalR-bound DNA using ChAP-Seq analysis unraveled more than 60 direct target genes and operons in addition to *malE*. Thus, the global impact of MalR ranges from peptidoglycan biosynthesis [*murA-murB* (Burkovski, 2013)] to the synthesis of arabinogalactan [*embC* (Alderwick et al., 2005)], and cell wall associated lipids and mycolic acids (e.g., via *oppA* and *ipsA*). For example, the *oppA* gene codes for an oligopeptide permease, which was further characterized in *M. tuberculosis* (Flores-Valdez et al., 2009). Flores-Valdez and others could show that OppA is required for the modulation of cell-wall associated lipids as well as mycolic acids. The LacI-type regulator IpsA, which is itself repressed by MalR, was previously shown to trigger the expression of the *ino1* gene encoding the inositol phosphate synthase. Deletion of *ipsA* resulted in a severe decrease of inositol-derived lipids and an abolished mycothiol biosynthesis (Baumgart et al., 2013). Remarkably, a Δ*ipsA* mutant features a similar cell morphology as *malR* overexpression, suggesting that some phenotypic effects may be indirectly resulting from reduced IpsA levels (**Figure 3C**).

A role in cell envelope remodeling appears to be a common theme in the family of MarR regulators. For example, the MarR-type regulator SlyA was shown to control several targets impacting cell envelope composition, some of which have direct implications on the resistance of *Salmonella enterica* toward antibiotics (e.g., polymyxins) or virulence (Navarre et al., 2005). The regulator Rv1404 from *M. tuberculosis* was also shown to contribute to an adaptation of the cells to the host environment by enhancing the acid-tolerance of this bacterium (Healy et al., 2016). Altogether, these examples emphasize different roles of MarR-type regulators. Whereas some proteins appear to conduct very distinct regulatory functions, like the control of a certain catabolic gene cluster, others (like MalR or SlyA) act as global regulators orchestrating complex adaptive strategies in response to environmental stresses.

A further prominent target of MalR appeared to be the cryptic prophage element CGP3. MalR bound to 13 regions within the CGP3 element overall and overexpression of the *malR* gene counteracted CGP3 induction upon addition of the SOS inducing antibiotic mitomycin C. So far, little is known about the role of MarR-type regulators in the control of horizontally acquired elements. In *Bacillus subtilis,* a *pamR* deficient strain displayed altered expression of prophage genes, however the precise regulatory impact remained unclear (De San Eustaquio-Campillo et al., 2017). Another example is depicted by RovA, which is a MarR-type transcription factor in *E. coli*, *Yersinia pseudotuberculosis* and its homolog SlyA from *Salmonella* (Navarre et al., 2007). Former studies revealed that RovA and SlyA act as countersilencers of H-NS target promoters controlling genes that impact virulence in *Yersinia and Salmonella species* (Heroven et al., 2004; Navarre et al.,

2005, 2006; Ellison and Miller, 2006). Remarkably, this regulatory plasticity was also the subject of a recent study by Will et al., highlighting the ability of MarR-type regulators to act as counter-silencers of horizontally acquired genes. This feature of this regulator family has significantly contributed to the evolution of *Enterobacteriaceae* by horizontal gene transfer (Will et al., 2019). We previously reported on the xenogeneic silencer CgpS, which plays a crucial role in the silencing of the CGP3 island by binding to AT-rich DNA (Pfeifer et al., 2016, 2019). MalR itself binds to an AT-rich palindromic motif whose composition, of course, increases the likelihood of an overrepresentation in horizontally acquired regions. However, the precise regulatory impact of MalR on CGP3 activation and a potential role in counter-silencing CgpS activity remains unclear. The observed reduction of CGP3 induction in response to *malR* overexpression speaks against a counter-silencing mechanism as reported for RovA. In contrast, the majority of phage targets appeared to be repressed by MalR. In physiological terms, a repressive function of MalR toward CGP3 genes could be overcome by the presence of an effector molecule leading to a dissociation of a putative MalR dimer. Therefore, we could speculate on a role of MalR in stress-responsive induction of the CGP3 prophage, which could literally leave the "sinking ship" when the life of its host is threatened by harsh environmental conditions.

With this study, we provide a comprehensive overview on the many targets controlled by MalR and suggest an important function of this MarR-type regulator in the coordinated control of genes with an impact on cell envelope composition. The relevance of this global response is reflected by the severely increased sensitivity of a *malR* mutant to several β-lactam antibiotics and is further supported by several other cases where members of this family contributed to enhanced antibiotic resistance (Poole et al., 1996; Alekshun and Levy, 1999; Lee et al., 2003; Navarre et al., 2005). We have gained a first glimpse on a complex adaptive response. However, many – if not most – of the MalR targets encode proteins of unknown function. For several others, only very limited data are available. So, many more studies are needed to understand the molecular principles behind this adaptive response. With the multitude of targets identified in this study, we provide a starting point for further studies aiming to enhance our understanding of bacterial adaptation to stress.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Growth Conditions

All bacterial strains and plasmids used in this work are listed in **Table 2**. For cloning and plasmid construction, the strain *E. coli* DH5α was used, whereas the strain *E. coli* BL21 was used for protein production. These strains were – unless stated otherwise – cultivated in lysogeny broth [LB, (Sambrook and Russell, 2001)] containing 50 μg/ml Kanamycin. *Corynebacterium glutamicum* ATCC 13032 was used as a wild type strain (Kalinowski et al., 2003). All derived *C. glutamicum* strains were cultivated in brain heart infusion medium (BHI, Difco Laboratories, Detroit, MI, United States) or in CGXII minimal medium containing 2% (w/v) glucose (Keilhauer et al., 1993); if necessary 25 μg/ml Kanamycin was added. For the cultivation of *C. glutamicum,* a first pre-culture was inoculated with single colonies from agar plates, either directly after transformation or after a streak-out of glycerol cultures. These pre-cultures were conducted in 4.5 ml BHI medium in test tubes or – depending on the purpose – in 1 ml BHI medium in 96-well deep well plates (DWPs) at 30°C for 8 h. Afterward, cells were used for inoculating a second overnight pre-culture in CGXII medium containing 2% (w/v) glucose. This CGXII culture was used to inoculate a main culture in the same medium to an OD_{600} of 1.

Growth experiments were conducted in the BioLector microbioreactor system (m2p labs, Baesweiler, Germany) (Kensy et al., 2009). Therefore, 750 μl CGXII medium containing 2% (w/v) glucose and the particular stated additives [e.g., Isopropyl β–D-1-thiogalactopyranoside (IPTG)] were inoculated to an OD_{600} of 1 in 48-well microtiter plates (Flowerplates, m2p labs) and cultivated for at least 24 h at 30°C and 1200 rpm shaking frequency. Fluorescence, as well as backscatter measurements, were taken every 15 min.

Recombinant DNA Work and Construction of Chromosomal Insertions or Deletions

All standard laboratory methods (PCR, DNA restriction, Gibson Assembly) were performed according to standard protocols and manufacturer's instructions (Sambrook and Russell, 2001; Gibson et al., 2009). The used oligonucleotides, as well as details regarding the plasmid construction are provided in the **Supplementary Tables S3A,B**.

For chromosomal integration or deletion, a two-step homologous recombination system based on the suicide vector pk19*mobsacB* was used (Schäfer et al., 1994; Niebisch and Bott, 2001). This vector contained 500 bps of each site flanking the targeted sequence in the genome of *C. glutamicum*.

Chromatin Affinity Purification and Next Generation Sequencing (ChAP-Seq)

Pre-cultures for ChAP-Seq were conducted as described above. As a main-culture, *C. glutamicum* ATCC 13032::*malR-strep* was grown in CGXII medium containing 2% (w/v) glucose for 5 h at 30°C and 120 rpm shaking frequency. Cells were then harvested by centrifugation (10 min at 11,325 × *g* and 4°C). Subsequently, the cells were washed once with CGXII medium without MOPS and resuspended in the same medium containing 1% (v/v) formaldehyde as a fixation agent and incubated at room temperature for 20 min. To stop this, glycine was added to a final concentration of 125 mM, followed by an additional 5 min incubation step at room temperature. To remove the remaining formaldehyde, the cells were washed twice with buffer A (100 mM Tris–HCl, pH 8.0, 1 mM EDTA),

TABLE 2 | Strains and plasmids used in this work.

Strain or plasmid	Relevant characteristics	Source
E. coli		
DH5α	*supE44 ΔlacU169* (φ80*lacZ*DM15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Invitrogen
BL21 (DE3)	F- *ompT hsdS*$_B$ (r$_B$-, m$_B$-) *gal dcm* (DE3)	Studier and Moffatt, 1986
C. glutamicum		
ATCC 13032	Biotin-auxotrophic wild type	Kinoshita et al., 2004
ATCC 13032 Δ*malR*	ATCC 13032 with an in-frame deletion of *malR* (cg3315)	This work
ATCC 13032::*malR-strep*	ATCC 13032 coding for a Strep-Tag fused C-terminally to MalR; integrated at the native locus	This work
ATCC 13032::P$_{lys}$-*eyfp*	ATCC 13032 containing the prophage reporter P$_{lys}$-*eyfp* integrated into the intergenic region of cg1121-cg1122	Helfrich et al., 2015
Plasmids		
pEKEx2	Kan^R; expression vector with *lacI*^{Q}, P$_{tac}$ and pUC18 multiple cloning site	Eikmanns et al., 1991
pEKEx2-*malR*	Kan^R; P$_{tac}$, overexpression vector for *malR* (cg3315)	This study
pJC1	Kan^R, Amp^R, and *C. glutamicum* shuttle vector	Cremer et al., 1990
pJC1-P$_{lys}$-lys-venus	Kan^R, Amp^R, and prophage reporter plasmid with the gene for the fluorescent protein Venus under control of the promoter of cg1974 (putative phage lysine)	(Pfeifer, unpublished)
pK19 *mob sacB*	Kan^R; plasmid for allelic exchange in *C. glutamicum*; (pK18 *ori*V$_{E.c.}$, *sacB*, *lacZα*)	Schäfer et al., 1994
pK19-*malR*-C-*strep*	Kan^R; plasmid for tagging *malR* (cg3315) genomically with a C-terminal Strep-tag	This study
pK19-Δ*malR*	Kan^R; plasmid for the in-frame deletion of *malR* (cg3315)	This study
pET24b	Kan^R; vector for overexpression of genes in E. coli, with optional C-terminal hexahistidine affinity tag (pBR322 oriVE.c. PT7 lacI)	Novagen
pET24b-*malR*-C-*strep*	Kan^R; plasmid for overexpression of C-terminal Strep-tagged *malR*	This study

and afterward resuspended in 25 ml buffer A, containing *cOmplete* Protease Inhibitor (Roche, Basel, Switzerland) and 2.5 mg RNase A. The subsequent preparation of ChAP-seq samples (cell disruption via French press, sonification, and ultracentrifugation) was conducted as described by Pfeifer et al. (2016). Afterward, the supernatant was then purified using a 2-ml bed volume Strep-Tactin-Sepharose column (IBA, Göttingen, Germany), following the manufacturer's protocol. The elution fractions were pooled, SDS was added to a final concentration of 1% (w/v), and incubated over night at 65°C, followed by a treatment with proteinase K (final concentration 400 mg · ml^{-1}) for 3 h at 55°C. In a last step, the DNA of the samples was purified by phenol–chloroform extraction using Roti® -phenol/chloroform/isoamyl alcohol (Carl Roth, Karlsruhe, Germany), precipitated with ethanol and 0.3 M sodium acetate, washed with 70% (v/v) ethanol, dried and resuspended in 100 μl H_2O. Further steps for the analysis of the gained DNA (sequencing, trimming, data analysis) were conducted as described by Pfeifer et al. (2016).

Protein Purification

MalR with a C-terminal Strep-tag was heterologously produced in *E. coli* BL21 (DE3), transformed with pET24b-*malR*-C-strep. Cells were grown to an OD_{600} of 0.6 at 37°C as described above. Subsequently, the protein production was induced using 100 μM IPTG and the cultivation was continued for 5 h at 16°C. Cells were harvested by centrifugation for 15 min at 5300 × *g* and 4°C and the pellets were snap-frozen using liquid nitrogen. For cell disruption, the pellets were thawed on ice and resuspended in buffer A (50 mM Tris–HCl, 1mM EDTA, pH 8.0) containing *cOmplete* Protease Inhibitor (Roche, Basel, Switzerland). This cell suspension was then treated with a French pressure cell for three passages at 172 MPa. Cell debris was removed by centrifugation for 30 min at 5300 × *g* and 4°C and a subsequent ultracentrifugation for 1 h at 150,000 × *g* and 4°C. The tagged protein was then purified with a 1-ml bed volume Strep-Tactin-Sepharose column (IBA, Göttingen, Germany), following the manufacturer's protocol.

Electrophoretic Mobility Shift Assays

As part of the investigation of the binding properties of MalR and as an *in vitro* verification of the results obtained by ChAP-seq analysis, EMSAs were performed. Therefore, 100 bp DNA fragments, centering the peak maximum of each particular promoter region, were amplified using PCR (oligonucleotide sequences are given in the **Supplementary Table S3C**) and analyzed and purified using an agarose gel with subsequent gel extraction with the "PCR clean-up and Gel extraction" Kit from Macherey-Nagel (Düren, Germany). A total of 90 ng of DNA per lane was incubated with different molar excesses of purified MalR protein (threefold and 10-fold molar excess) for 30 min in bandshift-buffer [50 mM Tris–HCl, 5 mM $MgCl_2$, 40 mM KCl, 5% (v/v) glycerol, pH 7.5]. Subsequently, samples were separated using a 10% native polyacrylamide gel electrophoresis as described previously (Pfeifer et al., 2016).

Transcriptome Analysis Using DNA Microarrays

To analyze the global transcriptomic alterations triggered by an overexpression of *malR C. glutamicum* ATCC 13032 cells, harboring either the empty vector pEKEx2 or the overexpression plasmid pEKEx2-*malR*, were cultivated in CGXII medium containing 2% (w/v) glucose, and 100 μM IPTG as described previously. Subsequently, total RNA of these cultures was

prepared using the RNeasy Mini Kit (QIAGEN, Hilden, Germany) following the manufacturers protocol. The cDNA labeling, and the DNA microarray analysis was performed as described previously (Donovan et al., 2015). The array data have been deposited in the GEO database[1] under the accession number: GSE116655.

Verification of the Transcriptomic Data Measuring mRNA Levels by Quantitative Real-Time PCR (qRT-PCR)

Preparation of total RNA from *C. glutamicum* cultures was carried out as described above. Measurement of differential gene expression was conducted using a qTower 2.2 (Analytik Jena, Jena, Germany) with the Luna® Universal One-Step RT-qPCR Kit (New England Biolabs, Ipswitch, United States). Primer pairs used for the analysis are listed in **Supplementary Table S3D**. For all samples, 100 ng of total RNA were used as a template and all measurements were performed in biological as well as in technical duplicates. The Ct-values of the samples were obtained using qPCR-soft 3.1 (Analytik Jena). Subsequently, the relative transcriptional changes were calculated using the following equation:

$$\Delta Ct = Ct\ (\text{sample}) - Ct\ (\text{control})$$

$$\text{Relative transcriptional change} = 2^{-\Delta Ct}$$

Fluorescence Microscopy and Staining

For microscopic analysis, cells were cultivated in CGXII medium containing 2% (w/v) glucose (as described above) and grown for 24 h at 30°C. Lipids were stained with Nile red and DNA was stained with Hoechst 33342 (Sigma-Aldrich, Munich, Germany). Therefore, 10 µl of the cell suspensions were centrifuged for 5 min at 8,000 × *g* and the pellet was resuspended in 500 µl PBS containing 100 ng/ml Hoechst 33342 and 250 ng/ml Nile red. The cells were incubated for 30 min at room temperature, and subsequently analyzed microscopically using an AxioImager M2 (Zeiss, Oberkochen, Germany) with a Zeiss AxioCam MRm camera and a Plan-Apochromat 100×, 1.40 Oil phase contrast oil-immersion objective. Fluorescence was measured using the 63 HE filter for Nile red fluorescence and the filter set 49 for Hoechst fluorescence.

The optimal exposure time for the different fluorescence images was determined with the automatic measurement option of the AxioVision Rel. 4.8 software (Carl Zeiss MicroImaging GmbH) and the pictures were analyzed with the same software.

Cell Counting and Determination of the Cell Size

Cell counts and biovolume were analyzed using a MultiSizer 3 particle counter (Graham, 2003) equipped with a 30 µm capillary in volumetric control mode. For the measurement, cells were grown for 24 h as described above and afterward diluted to an $OD_{600} \leq 0.025$ in CASYton buffer (Schärfe Systems, Reutlingen, Germany). Only particles sizing from 0.633 to 18 µm were analyzed.

Phenotypic Analysis of *C. glutamicum* Δ*malR* and *C. glutamicum* ATCC 13032

Phenotypical characterization of the strains *C. glutamicum* Δ*malR* and *C. glutamicum* ATCC 13032 was performed using the Phenotypic MicroArrays from BIOLOG (Biolog Inc., Hayward, CA, United States). Both strains were compared regarding their respiratory activity in the presence of various carbon sources (PM1 and PM2), phosphorus and sulfur sources (PM4), different osmolytes (PM9), pH-values (PM10), and antibiotics (PM11, PM12, and PM13). Experimental setup was carried out as described in the BIOLOG protocol for analysis of *Bacillus subtilis* and other Gram-positive bacteria (Biolog Inc., Hayward, CA, United States). In short, both strains were grown overnight at 30°C on blood agar plates. Cells were inoculated in the different PM-media containing 1% (v/v) of the redox dye (dye mix F) to a turbidity of 81% transmittance. Afterward, each well of the PM-plates was filled with 100 µl of the corresponding inoculation medium. Phenotypic MicroArrays were analyzed using the OmniLog incubator (Biolog Inc., Hayward, CA, United States) at 30°C for 48 h with a measuring rate of 15 min. Data visualization was performed by the BIOLOG software OM_Pl_Par 1.20.02 for parametric analyses. For selected examples, GraphPad Prism 7 was used for visualization. An overview of all results is shown in **Supplementary Figure S6**.

Flow Cytometric Analysis

The CGP3 prophage induction was assessed by flow cytometric analysis of a *C. glutamicum* strain, harboring a genomically integrated prophage reporter (ATCC 13032::P_{lys}-*eyfp*), using a BD Accuri C6 flow cytometer (BD biosciences, Heidelberg, Germany). Cells were cultivated in CGXII medium containing 2% (w/v) glucose (as described above) and grown for 25 h at 30°C. As appropriate, different concentrations of IPTG (10, 25, and 50 µM), as well as 600 nM Mitomycin C, were used. Flow cytometric analysis was performed using a 488 nm laser and a 530/30 nm filter for measuring eYFP fluorescence. In total, 100,000 events were analyzed per sample and data was analyzed using BD Accuri C6 software and visualized using GraphPad Prism 7. The gating was performed according to the uninduced negative control.

Scanning Electron Microscopy

For SEM, bacteria were fixed with 3% (v/v) glutaraldehyde (Agar Scientific, Wetzlar, Germany) in PBS for at least 4 h, washed in 0.1 M Soerensen's phosphate buffer (Merck, Darmstadt, Germany) for 15 min, and dehydrated by incubating consecutively in an ascending acetone series (30, 50, 70, 90, and 100%) for 10 min each and the last step thrice. The samples were critical point dried in liquid CO_2 and then sputter coated (Sputter Coater EM SCD500; Leica, Wetzlar, Germany) with a 10-nm gold/palladium layer. Samples were analyzed using an environmental scanning electron microscope (ESEM XL 30 FEG, FEI, Philips, Eindhoven, Netherlands) with a 10-kV acceleration voltage in a high-vacuum environment.

[1] http://ncbi.nlm.nih.gov/geo

Transmission Electron Microscopy

In preparation for TEM analysis, bacteria were fixed with 3% (vol/vol) glutaraldehyde (Agar Scientific, Wetzlar, Germany) in PBS for at least 4 h, washed in 0.1 M Soerensen's phosphate buffer (Merck, Darmstadt, Germany), and postfixed in 1% OsO4 in 17% sucrose buffer. After fixation, bacteria were embedded in 2.5% agarose (Sigma, Steinheim, Germany), then rinsed in 17% sucrose buffer and deionized water and dehydrated by ascending ethanol series (30, 50, 70, 90 and 100%) for 10 min each. The last step was repeated 3 times. Dehydrated specimens were incubated in propylene oxide (Serva, Heidelberg, Germany) for 30 min, in a mixture of Epon resin (Serva, Heidelberg, Germany) and propylene oxide (1:1) for 1 h and finally in pure Epon for 1 h. Samples were embedded in pure Epon. Epon polymerization was performed at 90°C for 2 h. Ultrathin sections (70–100 nm) were cut by ultramicrotome (Reichert Ultracut S, Leica, Wetzlar, Germany) and picked up on Cu/Rh grids (HR23 Maxtaform, Plano, Wetzlar, Germany). Contrast was enhanced by staining with 0.5% uranyl acetate and 1% lead citrate (both EMS, Munich, Germany). Samples were viewed at an acceleration voltage of 60 kV using a Zeiss Leo 906 (Carl Zeiss, Oberkochen, Germany) transmission electron microscope.

DATA AVAILABILITY

The microarray data are available in NCBI via GEO record GSE116655.

AUTHOR CONTRIBUTIONS

MH and JF conceived the study. MH and MP performed the experiments and analyzed the data. MH and JF wrote the manuscript. All authors reviewed and edited the manuscript.

FUNDING

We thank the German Research Foundation (SPP 1617, grant no. FR2759/2-2) and the European Research Council (ERC-StG-2017, grant no. 757563) for their financial support.

ACKNOWLEDGMENTS

We thank Tino Polen and Doris Rittmann for their assistance in sequencing and comparative transcriptome analysis, Holger Morschett for help with the Coulter Counter, Cornelia Gätgens for technical support, and Eugen Pfeifer and Marc Keppel for their fruitful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01039/full#supplementary-material

REFERENCES

Alderwick, L. J., Harrison, J., Lloyd, G. S., and Birch, H. L. (2015). The mycobacterial cell wall-peptidoglycan and arabinogalactan. *Cold Spring Harb. Perspect. Med.* 5, 1–16. doi: 10.1101/cshperspect.a021113

Alderwick, L. J., Radmacher, E., Seidel, M., Gande, R., Hitchen, P. G., Morris, H. R., et al. (2005). Deletion of Cg-emb in corynebacterianeae leads to a novel truncated cell wall arabinogalactan, whereas inactivation of Cg-ubiA results in an arabinan-deficient mutant with a cell wall galactan core. *J. Biol. Chem.* 280, 32362–32371. doi: 10.1074/jbc.M506339200

Alekshun, M. N., and Levy, S. B. (1999). The mar regulon: multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol.* 7, 410–413. doi: 10.1016/S0966-842X(99)01589-9

Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A., and Head, J. F. (2001). The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3A resolution. *Nat. Struct. Biol.* 8, 710–714. doi: 10.1038/90429

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Baumgart, M., Luder, K., Grover, S., Gätgens, C., Besra, G. S., and Frunzke, J. (2013). IpsA, a novel LacI-type regulator, is required for inositol-derived lipid formation in *Corynebacteria* and *Mycobacteria*. *BMC Biol.* 11:122. doi: 10.1186/1741-7007-11-122

Belanger, A. E., Besra, G. S., Ford, M. E., Mikusová, K., Belisle, J. T., Brennan, P. J., et al. (1996). The embAB genes of *Mycobacterium avium* encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. *Proc. Natl. Acad. Sci. U.S.A.* 93, 11919–11924. doi: 10.1073/pnas.93.21.11919

Bochner, B. R., Gadzinski, P., and Panomitros, E. (2001). Phenotype Microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* 11, 1246–1255. doi: 10.1101/gr.186501

Brune, I., Brinkrolf, K., Kalinowski, J., Pühler, A., and Tauch, A. (2005). The individual and common repertoire of DNA-binding transcriptional regulators of *Corynebacterium glutamicum, Corynebacterium efficiens, Corynebacterium diphtheriae* and *Corynebacterium jeikeium* deduced from the complete genome sequences. *BMC Genomics* 6:86. doi: 10.1186/1471-2164-6-86

Burkovski, A. (2013). Cell envelope of corynebacteria: structure and influence on pathogenicity. *ISRN Microbiol.* 2013:935736. doi: 10.1155/2013/935736

Bussmann, M., Baumgart, M., and Bott, M. (2010). RosR (Cg1324), a hydrogen peroxide-sensitive MarR-type transcriptional regulator of *Corynebacterium glutamicum*. *J. Biol. Chem.* 285, 29305–29318. doi: 10.1074/jbc.M110.156372

Carrión, M., Gómez, M. J., Merchante-Schubert, R., Dongarrá, S., and Ayala, J. A. (1999). mraW, an essential gene at the dcw cluster of *Escherichia coli* codes for a cytoplasmic protein with methyltransferase activity. *Biochimie* 81, 879–888. doi: 10.1016/S0300-9084(99)00208-4

Chao, L., and Levin, B. R. (1981). Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 78, 6324–6328. doi: 10.1073/pnas.78.10.6324

Craig, W. A., and Andes, D. R. (2015). "Cephalosporins," in *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, eds J. E. Bennett, R. Dolin, and M. J. Blaser (Hoboken, NJ: Wiley), 278.e5–292.e292. doi: 10.1016/B978-1-4557-4801-3.00021-7

Cremer, J., Eggeling, L., and Sahm, H. (1990). Cloning the dapA dapB cluster of the lysine-secreting bacterium *Corynebacterium glutamicum*. *Mol. Gen. Genet.* 20, 478–480. doi: 10.1007/BF00391757

Cronan, J. E., and Thomas, J. (2009). Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Methods Enzymol.* 459, 395–433. doi: 10.1016/S0076-6879(09)04617-5

Daffé, M., and Draper, P. (1997). *The Envelope Layers of Mycobacteria with Reference to their Pathogenicity*. Cambridge, MA: Academic Press.

De San Eustaquio-Campillo, A., Cornilleau, C., Guerin, C., Carballido-Lopez, R., and Chastanet, A. (2017). PamR, a new MarR-like regulator affecting prophages

and metabolic genes expression in *Bacillus subtilis*. *PLoS One* 12:e0189694. doi: 10.1371/journal.pone.0189694

Deochand, D. K., and Grove, A. (2017). MarR family transcription factors: dynamic variations on a common scaffold. *Crit. Rev. Biochem. Mol. Biol.* 52, 595–613. doi: 10.1080/10409238.2017.1344612

Donovan, C., Heyer, A., Pfeifer, E., Polen, T., Wittmann, A., Krämer, R., et al. (2015). A prophage-encoded actin-like protein required for efficient viral DNA replication in bacteria. *Nucleic Acids Res.* 43, 5002–5016. doi: 10.1093/nar/gkv374

Eggeling, L., Besra, G. S., and Alderwick, L. J. (2008). "Structure and synthesis of the cell wall," in *Corynebacteria: Genomics and Molecular Biology*, ed. A. Burkovski (Norfolk: Caister Academic Press), 267–294.

Egland, P. G., and Harwood, C. S. (1999). BadR, a new MarR family member, regulates anaerobic benzoate degradation by *Rhodopseudomonas palustris* in concert with aadR, an Fnr family member. *J. Bacteriol.* 181, 2102–2109.

Eikmanns, B. J., Kleinertz, E., Liebl, W., and Sahm, H. (1991). A family of *Corynebacterium glutamicum*/*Escherichia coli* shuttle vectors for cloning, controlled gene expression, and promoter probing. *Gene* 102, 93–98. doi: 10.1016/0378-1119(91)90545-M

Ellison, D. W., and Miller, V. L. (2006). H-NS represses inv transcription in *Yersinia enterocolitica* through competition with RovA and interaction with YmoA. *J. Bacteriol.* 188, 5101–5112. doi: 10.1128/JB.00862-05

Flores-Valdez, M. A., Morris, R. P., Laval, F., Daffé, M., and Schoolnik, G. K. (2009). *Mycobacterium tuberculosis* modulates its cell surface via an oligopeptide permease (Opp) transport system. *FASEB J.* 23, 4091–4104. doi: 10.1096/fj.09-132407

Fontana, R., Cornaglia, G., Ligozzi, M., and Mazzariol, A. (2000). The final goal: penicillin-binding proteins and the target of cephalosporins. *Clin. Microbiol. Infect.* 6, 34–40. doi: 10.1111/j.1469-0691.2000.tb02038.x

Frunzke, J., Bramkamp, M., Schweitzer, J. E., and Bott, M. (2008). Population heterogeneity in *Corynebacterium glutamicum* ATCC 13032 caused by prophage CGP3. *J. Bacteriol.* 190, 5111–5119. doi: 10.1128/jb.00310-08

Galán, B., Kolb, A., Sanz, J. M., García, J. L., and Prieto, M. A. (2003). Molecular determinants of the hpa regulatory system of *Escherichia coli*: the HpaR repressor. *Nucleic Acids Res.* 31, 6598–6609. doi: 10.1093/nar/gkg851

George, A. M., and Levy, S. B. (1983). Gene in the major cotransduction gap of the *Escherichia coli* K-12 linkage map required for the expression of chromosomal resistance to tetracycline and other antibiotics. *J. Bacteriol.* 155, 541–548.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. III, Smith, H. O., et al. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343–345. doi: 10.1038/nmeth.1318

Gourdon, P., Baucher, M.-F. F., Lindley, N. D., and Guyonvarch, A. (2000). Cloning of the malic enzyme gene from *Corynebacterium glutamicum* and role of the enzyme in lactate metabolism. *Appl. Environ. Microbiol.* 66, 2981–2987. doi: 10.1128/AEM.66.7.2981-2987.2000

Graham, M. D. (2003). The coulter principle: foundation of an industry. *J. Assoc. Lab. Autom.* 8, 72–81. doi: 10.1016/S1535-5535(03)00023-6

Grove, A. (2013). MarR family transcription factors. *Curr. Biol.* 23, R142–R143. doi: 10.1016/j.cub.2013.01.013

Healy, C., Golby, P., MacHugh, D. E., and Gordon, S. V. (2016). The MarR family transcription factor Rv1404 coordinates adaptation of *Mycobacterium tuberculosis* to acid stress via controlled expression of Rv1405c, a virulence-associated methyltransferase. *Tuberculosis* 97, 154–162. doi: 10.1016/J.TUBE.2015.10.003

Helfrich, S., Pfeifer, E., Krämer, C., Sachs, C. C., Wiechert, W., Kohlheyer, D., et al. (2015). Live cell imaging of SOS and prophage dynamics in isogenic bacterial populations. *Mol. Microbiol.* 98, 636–650. doi: 10.1111/mmi.13147

Henke, N. A., Heider, S. A. E. E., Hannibal, S., Wendisch, V. F., and Peters-Wendisch, P. (2017). Isoprenoid pyrophosphate-dependent transcriptional regulation of carotenogenesis in *Corynebacterium glutamicum*. *Front. Microbiol.* 8:633. doi: 10.3389/fmicb.2017.00633

Heroven, A. K., Nagel, G., Tran, H. J., Parr, S., and Dersch, P. (2004). RovA is autoregulated and antagonizes H-NS-mediated silencing of invasin and rovA expression in *Yersinia pseudotuberculosis*. *Mol. Microbiol.* 53, 871–888. doi: 10.1111/j.1365-2958.2004.04162.x

Hu, X., Li, X., Huang, L., Chan, J., Chen, Y., Deng, H., et al. (2015). Quantitative proteomics reveals novel insights into isoniazid susceptibility in mycobacteria mediated by a universal stress protein. *J. Proteome Res.* 14, 1445–1454. doi: 10.1021/pr5011058

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., et al. (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J. Biotechnol.* 104, 5–25.

Kallscheuer, N., Vogt, M., Kappelmann, J., Krumbach, K., Noack, S., Bott, M., et al. (2016). Identification of the phd gene cluster responsible for phenylpropanoid utilization in *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* 100, 1871–1881. doi: 10.1007/s00253-015-7165-1

Keilhauer, C., Eggeling, L., and Sahm, H. (1993). Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the ilvB-ilvN-ilvC operon. *J. Bacteriol.* 175, 5595–5603.

Kensy, F., Zang, E., Faulhammer, C., Tan, R.-K. K., Büchs, J., and Buchs, J. (2009). Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. *Microb. Cell Fact.* 8:31. doi: 10.1186/1475-2859-8-31

Kinoshita, S., Udaka, S., and Shimono, M. (2004). Studies on the amino acid fermentation. Part 1. Production of L-glutamic acid by various microorganisms. *J. Gen. Appl. Microbiol.* 50, 331–343.

Krause, J. P., Polen, T., Youn, J.-W., Emer, D., Eikmanns, B. J., and Wendisch, V. F. (2012). Regulation of the malic enzyme gene malE by the transcriptional regulator MalR in *Corynebacterium glutamicum*. *J. Biotechnol.* 159, 204–215. doi: 10.1016/j.jbiotec.2012.01.003

Kumarevel, T. (2012). "The MarR family of transcriptional regulators-a structural perspective," in *Antibiotic Resistant Bacteria – A Continuous Challenge in the New Millennium* ed. M. Pana (London: Intech), 403–418. doi: 10.5772/28565 Available at: https://www.intechopen.com/books/antibiotic-resistant-bacteria-a-continuous-challenge-in-the-new-millennium/the-marr-family-of-transcriptional-regulators-a-structural-perspective

Kvint, K., Nachin, L., Diez, A., and Nyström, T. (2003). The bacterial universal stress protein: function and regulation. *Curr. Opin Microbiol.* 6, 140–145. doi: 10.1016/S1369-5274(03)00025-0

Lee, E. H., Rouquette-Loughlin, C., Folster, J. P., and Shafer, W. M. (2003). FarR regulates the farAB-encoded efflux pump of *Neisseria gonorrhoeae* via an MtrR regulatory mechanism. *J. Bacteriol.* 185, 7145–7152. doi: 10.1128/JB.185.24.7145-7152.2003

Li, Z., Sun, H., Mo, X., Li, X., Xu, B., and Tian, P. (2013). Overexpression of malic enzyme (ME) of *Mucor circinelloides* improved lipid accumulation in engineered *Rhodotorula glutinis*. *Appl. Microbiol. Biotechnol.* 97, 4927–4936. doi: 10.1007/s00253-012-4571-5

Marrakchi, H., Lanéelle, M. A., and Daffé, M. (2014). Mycolic acids: structures, biosynthesis, and beyond. *Chem. Biol.* 21, 67–85. doi: 10.1016/j.chembiol.2013.11.011

Mishra, A. K., Driessen, N. N., Appelmelk, B. J., and Besra, G. S. (2011). Lipoarabinomannan and related glycoconjugates: structure, biogenesis and role in *Mycobacterium tuberculosis* physiology and host-pathogen interaction. *FEMS Microbiol. Rev.* 35, 1126–1157. doi: 10.1111/j.1574-6976.2011.00276.x

Nanda, A. M., Heyer, A., Krämer, C., Grünberger, A., Kohlheyer, D., and Frunzke, J. (2014). Analysis of SOS-induced spontaneous prophage induction in *Corynebacterium glutamicum* at the single-cell level. *J. Bacteriol.* 196, 180–188. doi: 10.1128/JB.01018-13

Navarre, W. W., Halsey, T. A., Walthers, D., Frye, J., McClelland, M., Potter, J. L., et al. (2005). Co-regulation of *Salmonella enterica* genes required for virulence and resistance to antimicrobial peptides by SlyA and PhoP/PhoQ. *Mol. Microbiol.* 56, 492–508. doi: 10.1111/j.1365-2958.2005.04553.x

Navarre, W. W., McClelland, M., Libby, S. J., and Fang, F. C. (2007). Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. *Genes Dev.* 21, 1456–1471. doi: 10.1101/gad.1543107

Navarre, W. W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S. J., et al. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science (80-)* 313, 236–238. doi: 10.1126/science.1128794

Niebisch, A., and Bott, M. (2001). Molecular analysis of the cytochrome bc1-aa3branch of the *Corynebacterium glutamicum* respiratory chain containing an unusual diheme cytochrome c1. *Arch. Microbiol.* 175, 282–294. doi: 10.1007/s002030100262

Ortalo-Magne, A., Dupont, M. A., Lemassu, A., Andersen, A. B., Gounon, P., and Daffe, M. (1995). Molecular composition of the outermost capsular material of the tubercle bacillus. *Microbiology* 141, 1609–1620. doi: 10.1099/13500872-141-7-1609

Pérez-Rueda, E., and Collado-Vides, J. (2001). Common history at the origin of the position – Function correlation in transcriptional regulators in archaea and bacteria. *J. Mol. Evol.* 53, 172–179. doi: 10.1007/s002390010207

Pérez-Rueda, E., Collado-Vides, J., and Segovia, L. (2004). Phylogenetic distribution of DNA-binding transcription factors in bacteria and archaea. *Comput. Biol. Chem.* 28, 341–350. doi: 10.1016/j.compbiolchem.2004.09.004

Peschel, A. (2002). How do bacteria resist human antimicrobial peptides? *Trends Microbiol.* 10, 179–186. doi: 10.1016/S0966-842X(02)02333-8

Pfeifer, E., Hünnefeld, M., Popa, O., and Frunzke, J. (2019). Impact of xenogeneic silencing on phage-host interactions. *J. Mol. Biol.* doi: 10.1016/j.jmb.2019.02.011.

Pfeifer, E., Hünnefeld, M., Popa, O., Polen, T., Kohlheyer, D., Baumgart, M., et al. (2016). Silencing of cryptic prophages in *Corynebacterium glutamicum*. *Nucleic Acids Res.* 44, 10117–10131. doi: 10.1093/nar/gkw692

Pfeifer-Sancar, K., Mentz, A., Rückert, C., and Kalinowski, J. (2013). Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. *BMC Genomics* 14:888. doi: 10.1186/1471-2164-14-888

Poole, K., Tetro, K., Zhao, Q., Neshat, S., Heinrichs, D. E., and Bianco, N. (1996). Expression of the multidrug resistance operon mexA-mexB-oprM in *Pseudomonas aeruginosa*: mexR encodes a regulator of operon expression. *Antimicrob. Agents Chemother.* 40, 2021–2028. doi: 10.1111/j.1574-6968.2000.tb09367.x

Providenti, M. A., and Wyndham, R. C. (2001). Identification and functional characterization of CbaR, a MarR-like modulator of the cbaABC-encoded chlorobenzoate catabolism pathway. *Appl. Environ. Microbiol.* 67, 3530–3541. doi: 10.1128/AEM.67.8.3530-3541.2001

Ratledge, C. (2014). The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: a reappraisal and unsolved problems. *Biotechnol. Lett.* 36, 1557–1568. doi: 10.1007/s10529-014-1532-3

Rodriguez, E., Navone, L., Casati, P., and Gramajo, H. (2012). Impact of malic enzymes on antibiotic and triacylglycerol: production in *Streptomyces coelicolor*. *Appl. Environ. Microbiol.* 78, 4571–4579. doi: 10.1128/AEM.00838-12

Rouanet, C., Reverchon, S., Rodionov, D. A., and Nasser, W. (2004). Definition of a consensus DNA-binding site for PecS, a global regulator of virulence gene expression in *Erwinia chrysanthemi* and identification of new members of the PecS regulon. *J. Biol. Chem.* 279, 30158–30167. doi: 10.1074/jbc.M403343200

Sambrook, J., and Russell, D. W. (2001). Molecular cloning: a laboratory manual/Joseph Sambrook, David W. Russell. *Q. Rev. Biol.* 76, 348–349. doi: 10.1086/394015

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., and Pühler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* 145, 69–73. doi: 10.1016/0378-1119(94)90324-7

Sharma, R., Keshari, D., Singh, K. S., Yadav, S., and Singh, S. K. (2016). MRA_1571 is required for isoleucine biosynthesis and improves *Mycobacterium tuberculosis* H37Ra survival under stress. *Sci. Rep.* 6:27997. doi: 10.1038/srep27997

Spory, A., Spory, A., Bosserhoff, A., Bosserhoff, A., von Rhein, C., von Rhein, C., et al. (2002). Differential regulation of multiple proteins of *Escherichia coli* and *Salmonella enterica* serovar Typhimurium by the transcriptional regulator SlyA. *J. Bacteriol.* 184, 3549–3559. doi: 10.1128/JB.184.13.3549

Srikumar, R., Paul, C. J., and Poole, K. (2000). Influence of mutations in the mexR repressor gene on expression of the MexA-MexB-OprM multidrug efflux system of *Pseudomonas aeruginosa*. *J. Bacteriol.* 182, 1410–1414. doi: 10.1128/JB.182.5.1410-1414.2000

Stubbendieck, R. M., and Straight, P. D. (2016). Multifaceted interfaces of bacterial competition. *J. Bacteriol.* 198, 2145–2155. doi: 10.1128/JB.00275-16

Studier, F. W., and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189, 113–130.

Valbuena, N., Letek, M., Ordóñez, E., Ayala, J., Daniel, R. A., Gil, J. A., et al. (2007). Characterization of HMW-PBPs from the rod-shaped actinomycete *Corynebacterium glutamicum*: peptidoglycan synthesis in cells lacking actin-like cytoskeletal structures. *Mol. Microbiol.* 66, 643–657. doi: 10.1111/j.1365-2958.2007.05943.x

Wilkinson, S. P., and Grove, A. (2006). Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr. Issues Mol. Biol.* 8, 51–62.

Will, W. R., Brzovic, P., Le Trong, I., Stenkamp, R. E., Lawrenz, M. B., Karlinsey, J. E., et al. (2019). The evolution of SlyA/RovA transcription factors from repressors to countersilencers in *Enterobacteriaceae*. *MBio* 10:e00009-19. doi: 10.1128/mbio.00009-19

Zhang, H., Zhang, L., Chen, H., Chen, Y. Q., Ratledge, C., Song, Y., et al. (2013). Regulatory properties of malic enzyme in the oleaginous yeast, *Yarrowia lipolytica*, and its non-involvement in lipid accumulation. *Biotechnol. Lett.* 35, 2091–2098. doi: 10.1007/s10529-013-1302-7

Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., and Daffé, M. (2008). Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J. Bacteriol.* 190, 5672–5680. doi: 10.1128/JB.01919-07

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.5 Identification of regulatory interactions between host proteins and the CGP3 prophage in *Corynebacterium glutamicum*

Introduction

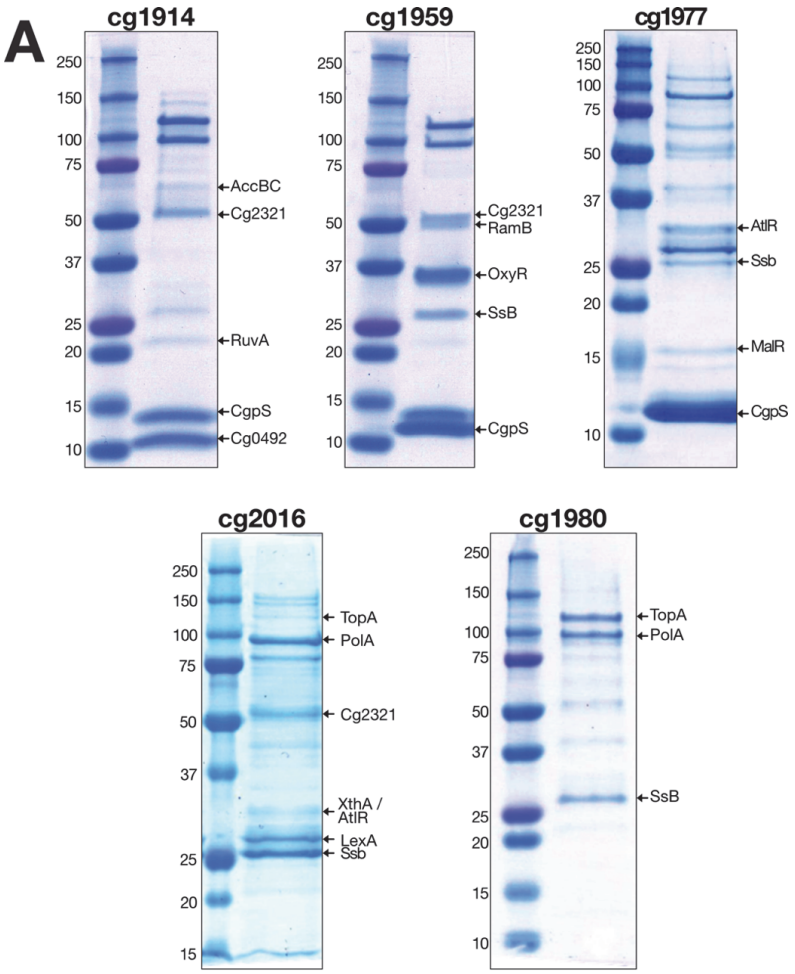
The cryptic prophage CGP3 is the largest prophage of *C. glutamicum* ATCC 13032. Besides the highly degenerated cryptic prophages CGP1 and CGP2, CGP3 is the only inducible prophage element of this bacterial strain. CGP3 spans with ~219 kbp around 7 % of the complete genome of *C. glutamicum* and contains another integrated prophage element - CGP4 (NC_003450.3, Ikeda and Nakagawa, 2003). The studies presented in this thesis revealed the Lsr2-like xenogeneic silencer CgpS as an important regulator involved in the control of CGP3 (Pfeifer et al., 2016). Under prophage-inducing conditions, this silencing protein shows a dynamic redistribution from the CGP3 area towards different host genome areas, which enables an excision and replication of CGP3 inside of *C. glutamicum* (Hünnefeld et al., 2019a). In order to detect further proteins involved in the regulatory mechanism of CGP3 induction, we performed DNA affinity chromatographies with different prophage promoters (cg1914, cg1959, cg1977, cg1980 and cg2016; Figure 1). These analyses resulted in the detection of several further proteins possessing binding sites inside of CGP3. Three of the identified proteins (AtlR, MalR, Cg2904) were further investigated concerning their *in vivo* binding behavior using ChAP-Seq analyses (Figure 2, Table 1 and 2 and Hünnefeld et al., 2019b). The binding of multiple proteins within the CGP3 area hints at a complex regulatory interaction between host regulatory circuits and the horizontally acquired prophage region. To further investigate the regulatory interaction of host regulators and the CGP3 prophage, we harnessed the information stored in our in-house ChAP-Seq database for all transcriptional regulators investigated so far. These comparative analyses focused on the occurrence of prophage binding (Figure 3).

Results and Discussion

DNA-affinity chromatographies were conducted using a variety of CGP3 prophage promoters. In order to have a broad spectrum of different promoter types, the choice was conducted with regard to the SOS-dependent inducibility of these promoters (based on Donovan et al., 2015), the binding behavior of CgpS towards these regions and simultaneously the reaction of the promoter activity to an artificial counter-silencing (based on Pfeifer et al., 2016). Results were also compared to already known regulatory interactions for the promoter regions (based on the internet database 'CoryneRegNet', Baumbach et al., 2006). The proteins bound towards these promoter regions were purified and subsequently analyzed using MALDI-TOF and LC-MS analysis (Figure 1). A summary of the obtained data is presented in Table 2. The two proteins AtlR, Cg2904, and the regulator MalR (Hünnefeld et al., 2019b) were the targets of further investigations.

Table 1: Primers used for the DNA affinity chromatographies (Figure 1).

Primer Name	Sequenz (5'->3')	Product Size
Cg1914_fw	CTTTGATCGAGACTTTCCAGCAG	403 bp
Cg1914_rv_Biotin	GAGGAGTCGTCGATGTGGAGACCGCACCACGACCAACAACATC	
Cg1959_fw	TCATTATGTGGAATTTCCGTCCC	523 bp
Cg1959_rv_Biotin	GAGGAGTCGTCGATGTGGAGACCACTGTAGGGGAGGTTGCTG	
Cg1977_fw	CACACCCCCTCTAGGAGG	503 bp
Cg1977_rv_Biotin	GAGGAGTCGTCGATGTGGAGACCGGCTGAGATCCAGTAGTGACTG	
Cg1980_fw	TTTGCTCTCAAGTATCGACATTG	413 bp
Cg1980_rv_Biotin	GAGGAGTCGTCGATGTGGAGACCTCAAGTTCAACGATGCGAGAC	
Cg2016_fw	CTGCTTAATCGTATATTCATTGCGC	224 bp
Cg2016_rv_Biotin	GAGGAGTCGTCGATGTGGAGACCTTAGACGTGGTGTTGAATCC	



B

	cg1914	cg1959	cg1977	cg1980	cg2016
AtlR			+	+	+
Cg0492	+	+			
RuvA	+			+	
GntR2	+				
GntR1	+			+	
OxyR		+			
Cg2904	+	+			
MalR			+	+	

Figure 1: DNA affinity chromatographies using different prophage promoters. (A) To identify possible regulatory proteins binding inside the CGP3 prophage region, DNA affinity chromatographies using different CGP3 promoter region were conducted. Promoter regions of genes cg1914, cg1959, cg1977, cg1980 and cg2016 were amplified by PCR with biotinylated primers and subsequently coupled to M-280 Streptavidin Dynabeads® (Invitrogen, Carlsbad, CA, USA) as described previously (Pfeifer et al., 2016). These magnetic beads were then incubated with a raw extract of exponentially growing *C. glutamicum* cells (with and without the addition of 600 nM mitomycin C). After elution of the bound proteins, they were separated by an SDS-PAGE and analyzed by MALDI-TOF. 4 – 20 % gradient SDS gels were used for all samples except of cg1977 (15 %) and cg2016 (12 %). The gels shown are an exemplary selection of the clearest gels. The protein bands, which could be successfully analyzed, are marked in the figure. Furthermore, elution fractions were analyzed by LC-MS. A comprehension of all results is presented in Table 2. (B) Comprehensive illustration of all data earned in (A). This table shows a combination of an analysis using SDS-PAGE and LC-MS. Proteins that were either found only or verified additionally with LC-MS are marked in red.

The regulator AtlR (former known as SucR) was first identified in 2010 as a DeoR-type transcriptional regulator involved in the regulation of the expression of a succinyl-CoA-synthetase in *C. glutamicum* ATCC 13032 (Cho et al., 2010). Furthermore, AtlR was published as a regulator of the alcohol dehydrogenase in *C. glutamicum* (Auchter et al., 2011) and as a regulator involved in the arabitol metabolism (Laslo et al., 2012). Laslo and colleagues demonstrated that AtlR functions as a repressor for different genes coding for proteins involved in the carbohydrate metabolism in *C. glutamicum* (i.a. *xylB*, *rbtB*, *mtlD*, *sixA*). This repressor function is disrupted if D-arabitol is present inside the cells (Laslo et al., 2012). In order to verify the results of the DNA-affinity chromatography and to analyze the binding behavior of AtlR, we performed ChAP-seq using a chromosomally integrated Strep-tagged variant of *atlR* (Figure 1 and Table 2). The ChAP-seq binding data set presented in Table 2 was filtered considering the role of AtlR as a transcriptional regulator: only peaks with a minimal peak-score (quality score based on peak size and shape) of 2.5, that were less than 700 bp away from the translation start site are displayed. With this analysis, we were able to verify the published binding site of AtlR inside its own promoter region. Additionally, we could show the binding inside of the CGP3 prophage. In total, we could find 56 binding peaks of AtlR inside the *C. glutamicum* genome, of which six belong to CGP3. Further peaks were e.g. detected inside the promoter regions of genes encoding multiple ribosomal proteins (e.g. *rpsJ*, *rpsH*, *rplM*), inside the promoter of *pyc* (encoding the pyruvate carboxylase), and inside the promoter of the Anti-σE factor *cseE* (Table 2). However, of the three promoters that showed AtlR binding in vitro (Figure 1B), none could be verified with the in vivo approach (Figure 2 and Table 2).

One explanation for this observation could be that the ChAP-seq experiments described here are only insights into one specific time point, which was possibly the right parameter to detect all binding events. Furthermore, there could be a discrepancy between in vitro and in vivo binding, because these conditions also differ in many parameters: the environment of proteins and DNA (either buffers or cytosol), the DNA structure (small fragments for DNA affinity chromatographies or whole chromosome for in vivo studies) and the presence of other binding proteins in vivo could change the outcome of these experiments. Additionally, it has to be taken into account that some proteins also show unspecific binding under in vitro conditions. One example of frequently bound proteins using DNA-affinity chromatographies is e.g. the single-strand binding protein Ssb or the putative RNAP subunit Cg2321 (Figure 1). Interestingly, several in-house conducted DNA-affinity chromatographies imply a frequent occurrence of

AtlR-binding to different DNA fragments. Therefore, it should not be excluded that the binding of AtlR could be a non-specific binding.

The protein Cg2904 is annotated as a hypothetical protein with unknown function (Baumgart et al., 2018). Homology searches and domain comparisons using the online tools InterPro (Mitchell et al., 2015) and BLASTp (Johnson et al., 2008), revealed an N-terminal helix-turn-helix (HTH) domain belonging to the 'lambda repressor-like DNA binding domain'-superfamily. The protein is conserved among different Corynebacteriales and a homolog of *Corynebacterium aurimucosum* ATCC 700975 (cauri_1965, 41 % homology) was shown to be encoded on a prophage element in this strain (Trost et al., 2010). We filtered the ChAP-seq results of Cg2904 using the same approach as previously described for the AtlR ChAP-seq. Here, we found in total 39 binding peak spanning five peaks inside the CGP3 region. Additionally, to these CGP3 binding sites, an interesting example is a high binding peak inside the promoter region of *atlR*. In total, Cg2904 binds inside the promoter region of four transcriptional regulators: *atlR*, *cadR*, *benR* and *ramA*. Furthermore, as already described for AtlR, Cg2904 binds to the promoter regions of different ribosomal proteins. Nevertheless, our in vitro binding studies found Cg2904 associated to the promoter region of cg1914 and cg1959 (Figure 1). Of these in vitro binding sites, similar to the results of AtlR, none could be found in vivo via ChAP-seq analysis.

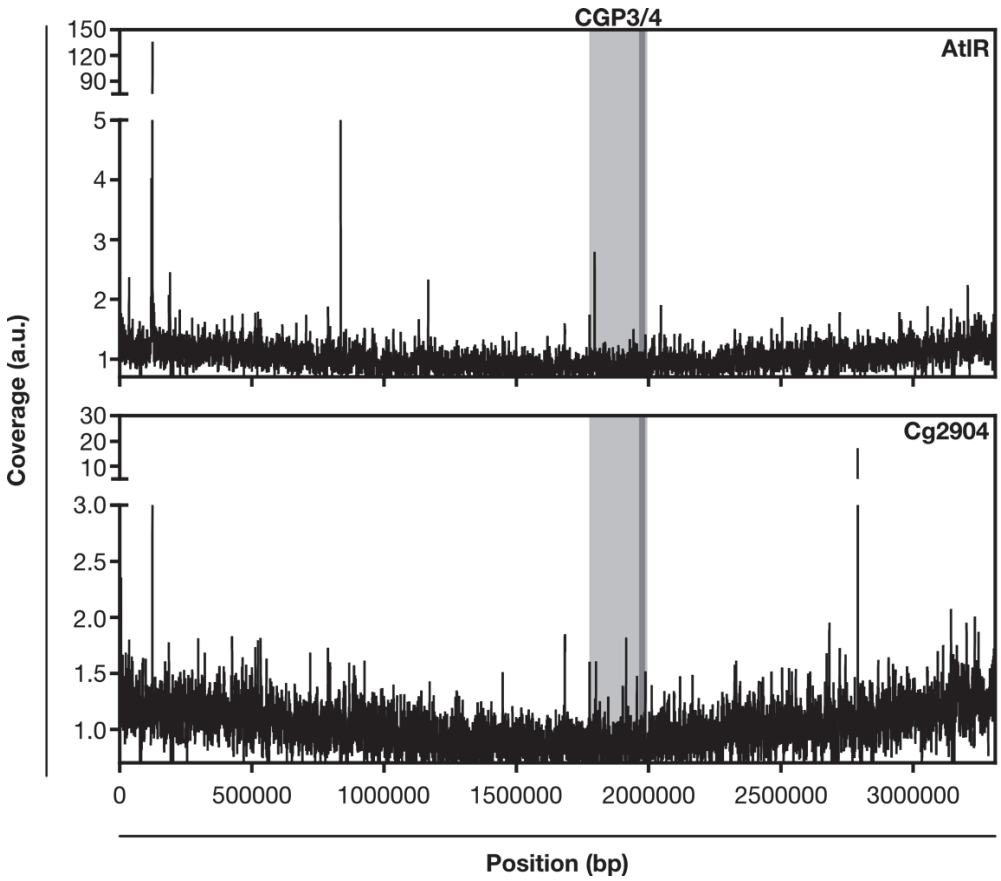


Figure 2: Global binding profiling of the proteins AtlR and Cg2904. For the analysis of the global binding patterns of the *C. glutamicum* regulator AtlR as well as for the CI-like protein Cg2904, *C. glutamicum* ATCC 13032 strains containing genes encoding for Strep-tagged versions of these proteins instead of the native genes were used. The cells were cultivated in CGXII medium containing 2 % (w/v) glucose for 5 h. Subsequently, cells were harvested, the protein-bound DNA was purified, and the data were processed as described previously (Hünnefeld et al., 2019b). The graphs in this figure represent the normalized coverage of the single experiments. Highlighted in light-grey is the CGP3 prophage area with the integrated CGP4 area (dark-grey). Filtered binding peak tables can be found in Table 2 (AtlR) and Table 3 (Cg2904).

Table 2: Filtered dataset of genome wide time resolved AltR binding (ChAP-Seq). This table belongs to Figure 2. The dataset was filtered for peaks less then 700 bp away from ATG and with a peak-score of at least 2.5. For all known TSS, the distance to this TSS is indicated. Annotation data are extracted and modified from Baumgart et al., 2018. (CGP3/4 peaks are marked in yellow).

NCgl-Number	Gene	Cg-Number	Peak max. position	Start	End	Peak-Score	Distance to ATG	Distance to TSS	Annotation
NCgl0046	*fhaA*	cg0064	49647	49607	49688	2.5	280	-	cytoplasmic protein with FHA and DUF domain
NCgl0081		cg0111	89534	89494	89575	2.5	80	-	hypothetical protein
NCgl0104		cg0138	118718	118678	118759	6.6	630	-	putative ATP/GTP-binding protein
NCgl0110	*atlR*	cg0146	124014	123974	124055	967.9	106	484	transcriptional regulator for arabitol metabolism, DeoR-family
NCgl0171	*cspA*	cg0215	186495	186455	186536	3.3	12	168	cold-shock protein A, contains signal peptide for secretion
NCgl0179			192790	192750	192831	3.2	66	-	transposase
NCgl0215	*aroT (pat)*	cg0267	233406	233366	233447	3.8	125	-	aminotransferase, uses aromatic amino acids, AT class I (EC:2.6.1.9)
NCgl0234		cg0291	254270	254230	254311	4.2	452	-	putative 3,4-dioxygenase β subunit
NCgl0250	*sigC*	cg0309	274180	274140	274221	7.2	61	-	RNA polymerase σ factor, ECF-family, control of branched quinol oxidation pathway
NCgl0302		cg0370	322247	322207	322288	2.6	558	-	putative ATP-dependent RNA helicase, DEAD/DEAH box-family
NCgl0468	*rplJ*	cg0572	509671	509631	509712	2.7	326	75	50S ribosomal protein L10
NCgl0471	*rpoB*	cg0576	512852	512812	512893	3.4	78	149	DNA-directed RNA polymerase β subunit (EC:2.7.7.6), essential
NCgl0475		cg0580	522495	522455	522536	3.9	125	-	hypothetical protein, conserved
NCgl0479			526140	526100	526181	3.2	71	-	hypothetical protein
NCgl0486	*rpsJ*	cg0593	532657	532617	532698	4.4	441	20	30S ribosomal protein S10
NCgl0515	*rpsH*	cg0628	555152	555112	555193	2.9	178	54	30S ribosomal protein S8
NCgl0556	*rplM*	cg0673	594825	594785	594866	4.5	113	28	50S ribosomal protein L13
NCgl0626	*cstA*	cg0756	668003	667963	668044	6.0	185	1	carbon starvation protein A
NCgl0659	*pyc*	cg0791	705036	704996	705077	2.8	174	118	pyruvate carboxylase (EC:6.4.1.1)
NCgl0698	*tusK (msiK2)*	cg0835	748607	748567	748648	2.6	582	-	trehalose uptake system, ABC-type, component
NCgl0717		cg0858	787971	787931	788012	3.3	20	-	putative secreted protein
NCgl0780		cg0931	861526	861486	861567	4.4	65	-	putative pyridoxal phosphate aminotransferase, AT class I (EC:2.6.1.1)
NCgl0837	*rpmE*	cg0994	927370	927330	927411	3.8	103	1	50S ribosomal protein L31
NCgl0867	*tnp6c*	cg1030	959740	959700	959781	2.5	340	2	transposase fragment
NCgl0914			1008594	1008554	1008635	5.3	69	1	ABC transporter ATPase
NCgl0947		cg1125	1044935	1044895	1044976	3.3	23	-	hypothetical protein, conserved
NCgl1076	*cseE*	cg1272	1166432	1166392	1166473	8.4	143	-	Anti-σE factor
NCgl1096		cg1292	1192278	1192238	1192319	3.4	119	1081	putative flavin-containing monooxygenase 3 (EC:1.14.13.8)
NCgl1116	*putP*	cg1314	1219694	1219654	1219735	2.6	93	-	proline transport system
NCgl1119		cg1318	1223859	1223819	1223900	2.5	27	-	putative DNA repair exonuclease, conserved
NCgl1324	*infC*	cg1563	1447702	1447662	1447743	4.2	89	117	translation initiation factor IF-3
NCgl1354a			1478930	1478890	1478971	4.8	42	-	hypothetical protein
NCgl1510	*qor*	cg1771	1661327	1661287	1661368	2.6	241	-	probable NADPH:quinone reductase, zeta-crystallin (EC:1.6.5.5)
NCgl1621			1789790	1789750	1789831	2.7	42	-	hypothetical protein
NCgl1625		cg1905	1793522	1793482	1793563	2.5	191	44	hypothetical protein CGP3 region
NCgl1656		cg1942	1818309	1818269	1818350	2.7	150	36	putative secreted protein CGP3 region
NCgl1735		cg2030	1920285	1920245	1920326	2.8	125	-	hypothetical protein CGP3 region
NCgl1808			1988676	1988636	1988717	5.8	371	-	hypothetical protein
NCgl1811			1989828	1989788	1989869	2.9	224	-	hypothetical protein
NCgl1816	*int2*	cg2071	1995684	1995644	1995725	3.3	391	-	putative phage integrase N-terminal fragment, CGP3 region
NCgl1858	*ptsI*	cg2117	2041366	2041326	2041407	2.6	46	-	EI enzyme, general component of PTS (EC:2.7.3.9)
NCgl2217	*malQ*	cg2523	2430255	2430215	2430296	2.5	41	31	4-α-glucanotransferase (EC:2.4.1.25)
NCgl2248	*aceA*	cg2560	2470453	2470413	2470494	4.0	287	176	isocitrate lyase (EC:4.1.3.1), part of glyoxylate shunt
NCgl2280	*rplU*	cg2595	2504856	2504816	2504897	4.1	254	-	50S ribosomal protein L21
NCgl2330		cg2648	2558391	2558351	2558432	2.7	218	-	putative transcriptional regulator, ArsR-family
NCgl2356		cg2683	2586094	2586054	2586135	2.6	119	92	hypothetical protein, conserved
NCgl2446a			2681322	2681282	2681363	2.6	20	-	50S ribosomal protein L36
NCgl2466		cg2822	2709594	2709554	2709635	4.7	284	-	putative sugar phosphate isomerase/epimerase
NCgl2686		cg3084	2969226	2969186	2969267	2.7	608	-	putative flavoprotein involved in K+ transport
NCgl2842	*uspA3*	cg3255	3143416	3143376	3143457	2.5	60	-	universal stress protein no. 3 / protein E
NCgl2848	*rsmP*	cg3264	3155104	3155064	3155145	2.6	338	-	cytoskeletal protein RsmP, regulates rod-shape morphology, conserved
NCgl2881	*rpsF*	cg3308	3185115	3185075	3185156	4.1	130	-	30S ribosomal protein S6
NCgl2938		cg3370	3247047	3247007	3247088	4.7	116	-	putative NADH-dependent flavin oxidoreductase
NCgl2949		cg3382	3255575	3255535	3255616	4.2	228	-	putative dipeptide/tripeptide permease
NCgl2961	*proP*	cg3395	3274332	3274292	3274373	3.2	258	-	proline/ectoine carrier, MFS-type
NCgl2993	*rpmH*	cg3432	3308876	3308836	3308917	2.7	324	-	50S ribosomal protein L34

Table 3: Filtered dataset of genome wide time resolved Cg2904 binding (ChAP-Seq). This table belongs to Figure 2. The dataset was filtered for peaks less then 700 bp away from ATG and with a peak-score of at least 2.5. For all known TSS, the distance to this TSS is indicated. Annotation data are extracted and modified from Baumgart et al., 2018. (CGP3/4 peaks are marked in yellow).

NCgl-Number	Gene	Cg-Number	Peak max. position	Start	End	Peak-Score	Distance to ATG	Distance to TSS	Annotation
NCgl0005	cg0007	*gyrB*	5391	5351	5432	7.9	43	5	DNA topoisomerase/gyrase IV, subunit B (EC:5.99.1.3)
NCgl0110	cg0146	*atlR*	124038	123998	124079	20.0	82	508	transcriptional regulator for arabitol metabolism, DeoR-family
NCgl0171	cg0215	*cspA*	186413	186373	186454	3.2	94	86	cold-shock protein A, contains signal peptide for secretion
NCgl0179			192787	192747	192828	3.4	69	-	transposase
NCgl0234	cg0291		254186	254146	254227	2.9	368	-	putative 3,4-dioxygenase β subunit
NCgl0303	cg0371	*cspA2*	321983	321943	322024	3.5	23	145	cold-shock protein A2
NCgl0348	cg0426	*tnp17a*	377652	377612	377693	2.5	30	-	transposase fragment, putative pseudogene, horizontally transferred
NCgl0471	cg0576	*rpoB*	512727	512687	512768	4.6	203	24	DNA-directed RNA polymerase β subunit (EC:2.7.7.6), essential
NCgl0472	cg0577	*rpoC*	516468	516428	516509	3.1	25	-	DNA-directed RNA polymerase β subunit (EC:2.7.7.6)
NCgl0476	cg0581	*rpsL*	522539	522499	522580	3.2	154	108	30S ribosomal protein S12
NCgl0486	cg0593	*rpsJ*	532736	532696	532777	2.8	362	59	30S ribosomal protein S10
NCgl0556	cg0673	*rplM*	594801	594761	594842	3.4	137	4	50S ribosomal protein L13
NCgl0834	cg0991	*rpmB*	925283	925243	925324	4.9	150	-	50S ribosomal protein L28
NCgl0867	cg1030	*tnp6c*	959732	959692	959773	3.8	348	10	transposase fragment
NCgl0869	cg1032	*cadR*	961123	961083	961164	3.3	173	173	transcriptional regulator, ArsR-family
NCgl1304	cg1531	*rpsA*	1420993	1420953	1421034	3.4	105	43	30S ribosomal protein S1, conserved
NCgl1324	cg1563	*infC*	1447680	1447640	1447721	5.4	111	95	translation initiation factor IF-3
NCgl1354a			1478922	1478882	1478963	4.0	34	-	hypothetical protein
NCgl1496	cg1757	*tnp3b*	1645516	1645476	1645557	2.5	144	91	transposase
NCgl1510	cg1771	*qor*	1661362	1661322	1661403	3.4	206	-	probable NADPH:quinone reductase, zeta-crystallin (EC:1.6.5.5)
NCgl1526	cg1791	*gapA (gap)*	1683714	1683674	1683755	4.7	90	-	glyceraldehyde-3-phosphate dehydrogenase glycolysis EC:1.2.1.12
NCgl1543	cg1809		1702373	1702333	1702414	4.4	58	-	putative DNA-directed RNA polymerase omega subunit, essential
NCgl1632	cg1915		1800490	1800450	1800531	6.5	382	96	hypothetical protein CGP3 region
NCgl1738	cg2033		1927098	1927058	1927139	2.7	262	-	putative secreted protein CGP3 region
NCgl1780			1967310	1967270	1967351	2.5	293	202	hypothetical protein
NCgl1792			1977550	1977510	1977591	2.7	221	173	hypothetical protein
NCgl1874	cg2135	*miaB*	2059451	2059411	2059492	2.7	20	-	tRNA methylthiotransferase
NCgl1977	cg2254		2165511	2165471	2165552	4.0	113	113	putative ankyrin repeat protein
NCgl2244	cg2556		2464235	2464195	2464276	3.0	109	-	putative iron-regulated membrane protein
NCgl2324	cg2641	*benR*	2549721	2549681	2549762	3.0	116	50	transcriptional regulator, LuxR-family
NCgl2446	cg2792	*nadE*	2681485	2681445	2681526	2.8	61	61	NAD+ synthetase, glutamine-hydrolyzing (EC:6.3.5.1)
NCgl2472	cg2831	*ramA*	2721301	2721261	2721342	5.9	74	-	transcriptional regulator, acetate metabolism, LuxR-family
NCgl2493	cg2853		2743967	2743927	2744008	5.8	43	40	putative protein-fragment, conserved
NCgl2532	cg2904		2790107	2790067	2790148	184.1	111	45	hypothetical protein
NCgl2749	cg3153		3033886	3033846	3033927	3.2	295	132	putative membrane protein
NCgl2810	cg3219	*ldhA (ldh)*	3113420	3113380	3113461	3.0	30	-	NAD-dependent L-lactate dehydrogenase
NCgl2848	cg3264	*rsmP*	3155057	3155017	3155098	4.4	291	-	cytoskeletal protein RsmP, regulates rod-shape morphology, conserved
NCgl2898	cg3328	*mutM2*	3201834	3201794	3201875	2.9	66	-	putative formamidopyrimidine-DNA glycosylase protein (EC:3.2.2.23)
NCgl2993	cg3432	*rpmH*	3308858	3308818	3308899	2.7	306	-	50S ribosomal protein L34

In order to get deeper insights into the interactions between the host regulatory circuits and the prophage CGP3, we conducted comparative analyses of all in-house performed ChAP-seq experiments. The focus of our analysis was the degree of prophage association of different transcriptional regulators and other DNA-binding proteins (Figure 3). For this purpose, we completely analyzed the binding patterns of each protein and compared the prophage-associated binding peak sizes with the binding peaks inside of the host genome. Figure 2 reveals the presence of a fraction of prophage binding for both above described proteins (AtlR and Cg2904) as well as for the previously described regulator MalR (Hünnefeld et al., 2019b). The xenogeneic silencer CgpS showed the highest fraction of prophage binding (Figure 3). This result fits to prior studies, which defined prophage regions in *C. glutamicum* as the main targets

of CgpS (Pfeifer et al., 2016). Interestingly, although AtlR and Cg2904 were found with the DNA-affinity chromatographies and the binding profiling revealed binding sites inside of CGP3, these proteins show only a small fraction of prophage binding in comparison to their genome-wide binding pattern. Further proteins present in our in-house ChAP-seq database, that were not part of this thesis, also displayed high fractions of prophage binding. In total, six of these proteins show a higher prophage binding fraction than the expected value.

The regulator with the most considerable fraction of prophage binding (besides CgpS) was DtxR. DtxR is the master regulator of iron-dependent gene regulation in *C. glutamicum,* and thus i.a. involved in the regulation of different iron uptake and iron storage components (Brune et al., 2006; Wennerhold and Bott, 2006). In the study of Wennerhold and Bott from 2006, regulatory effects of DtxR on the prophage CGP3 could already be demonstrated because a *dtxR* mutant strain exhibits increased prophage gene expression. Furthermore, this study revealed the in vitro binding of DtxR towards different prophage areas.

As previously described, prophage induction can be triggered via the SOS-response e.g. caused by DNA-damage (Oppenheim et al., 2005). Especially high intracellular Fe^{2+} levels can be detrimental for bacteria and cause DNA-damage, as this launches the formation of reactive oxygen species (ROS) catalyzed by the Fenton reaction (reviewed by Imlay, 2003). In the presence of iron, DtxR, acts also as a repressor for genes involved in iron acquisition (e.g. siderophore transporters and heme oxygenases) and at the same time as an activator for the expression of iron storage related genes like *dps* and *ftn* (Wennerhold and Bott, 2006). Both genes encode proteins exhibiting iron-sequestering abilities, which are therefore able to prevent ROS generation (Andrews et al., 2003; Wennerhold and Bott, 2006). Therefore, a deletion of *dtxR* leads under iron excess to elevated intracellular Fe^{2+} levels and bears the risk for the formation of ROS. This would explain the increased prophage induction in the *dtxR* mutant. Additional studies verified the effects of DtxR on CGP3 gene expression and that the *dtxR* deletion led to prophage induction (Freiherr von Boeselager et al., 2018; Frunzke et al., 2008).

Interestingly, not only binding of DtxR inside the CGP3 could be shown, but our data further suggest, that under prophage inducing conditions, CgpS binds to the promoter region of *dtxR* (Hünnefeld et al., 2019a). Taken together these findings indicate a complex regulatory

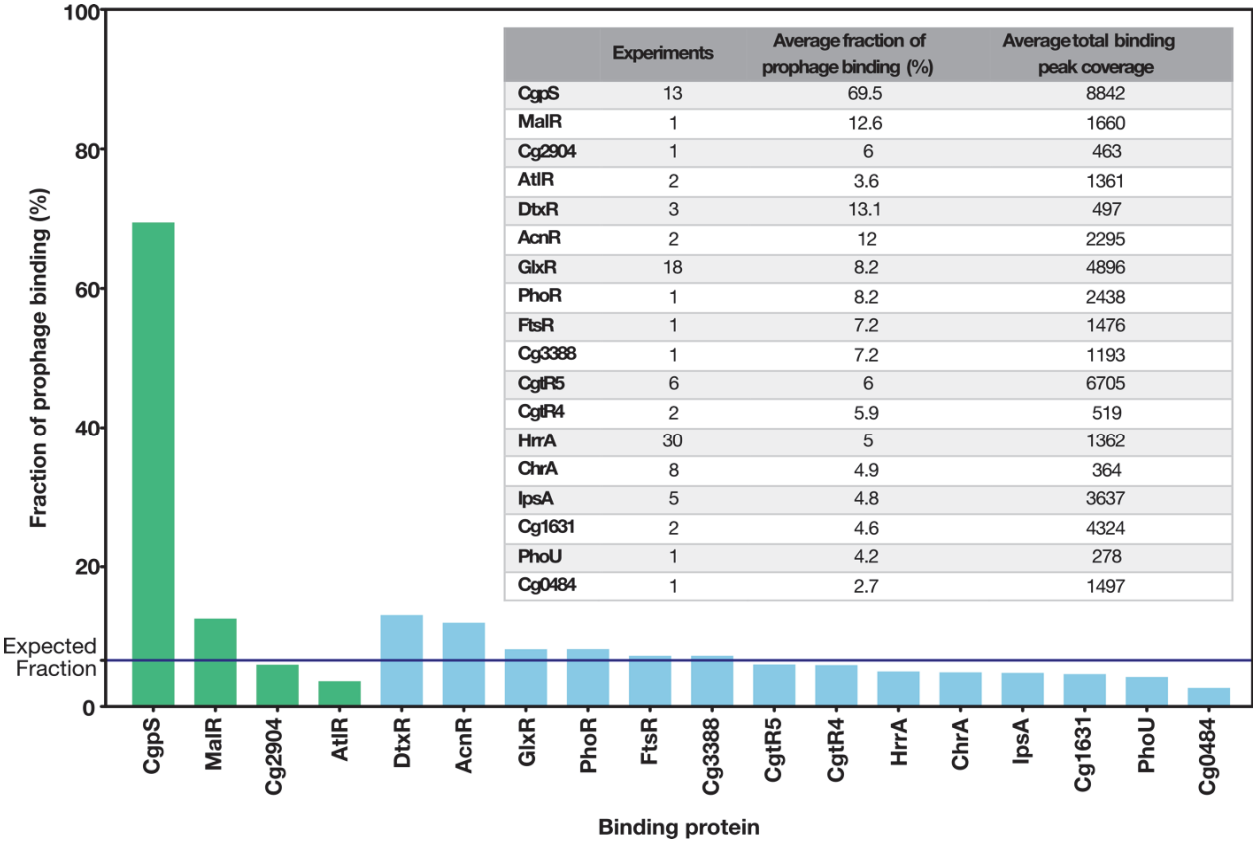
interaction between DtxR and CGP3. This illustrates that this interaction is an extremely interesting target for future investigations.

With 12 % binding peak coverage inside of prophage regions, AcnR represents another highly CGP3-associated protein (Figure 2). This transcriptional regulator was published as a repressor for the aconitase gene in *C. glutamicum* (Krug et al., 2005). The aconitase plays an essential role in the tricarboxylic acid cycle in *C. glutamicum* and requires an iron-sulfur cluster for substrate binding (Beinert et al., 1996). DNA microarray experiments with a strain lacking *acnR* revealed a regulation of CGP3 genes due to this deletion. However, these experiments were conducted using varying iron concentrations, which could trigger an AcnR-independent effect on the expression of prophage genes (as described for DtxR).

It is important to state that the presented analysis does not reflect any reliable quantitative data and does not say anything about the quality of the underlying data sets. Because the number of experiments differs for every protein and the evaluation was performed independently from the experimenters who conducted the sample preparation, there are many sources of errors in these results. However, the results of this analysis hint on a high abundance of proteins binding towards the CGP3 prophage region in *C. glutamicum*, because every analyzed experiment exhibits at least a small fraction of prophage binding. These prophage binding abilities could indicate an old relationship between the cryptic prophage CGP3 and its host because it is very deeply integrated into regulatory circuits.

In summary, the results show, that the level of prophage binding varies highly between the tested regulators. However, with DtxR and AcnR, some interesting candidates for further studies regarding the involvement of host-regulators in prophage regulation could be identified.

In previous studies, Will et al. described transcriptional regulators (SlyA/RovA and PhoP) exhibiting counter-silencing abilities in addition to their regular function (Will et al., 2014, 2019). The authors demonstrated that these dual-functions are products of an evolutionary network expansion of different *Enterobacteriaceae*. Inspired by these studies, the regulators presented in this study provide intriguing examples to investigate the extension of host regulatory networks to horizontally acquired elements. It is possible that the regulators displaying prophage binding also exhibit dual functions and thus are involved in prophage regulation.



	Experiments	Average fraction of prophage binding (%)	Average total binding peak coverage
CgpS	13	69.5	8842
MalR	1	12.6	1660
Cg2904	1	6	463
AtlR	2	3.6	1361
DtxR	3	13.1	497
AcnR	2	12	2295
GlxR	18	8.2	4896
PhoR	1	8.2	2438
FtsR	1	7.2	1476
Cg3388	1	7.2	1193
CgtR5	6	6	6705
CgtR4	2	5.9	519
HrrA	30	5	1362
ChrA	8	4.9	364
IpsA	5	4.8	3637
Cg1631	2	4.6	4324
PhoU	1	4.2	278
Cg0484	1	2.7	1497

Figure 3: Comparative analysis of prophage binding of different *C. glutamicum* regulators. The bar plots represent the fraction of the prophage binding of the different DNA-binding proteins. The green bars show experiments that were conducted in the course of this doctoral thesis. The blue bars show experiments conducted in-house until May 2019. The dark-blue horizontal line marks the expected value (μ) of the average prophage binding fraction. The table inside this figure shows meta data underlying the plots. Here, the number of experiments, average binding of all experiments towards prophage areas and the average total binding peak coverage are presented. The fraction of prophage binding was calculated using the averages of all experiments. For each experiment, the binding peak coverage of the peaks inside the prophage was divided by the total binding peak coverage through the whole genome. Subsequently, the average of these values was plotted.

References

Andrews, S. C., Robinson, A. K., and Rodríguez-Quiñones, F. (2003). Bacterial iron homeostasis. *FEMS Microbiol. Rev.* 27, 215–237. doi:10.1016/S0168-6445(03)00055-X.

Auchter, M., Laslo, T., Fleischer, C., Schiller, L., Arndt, A., Gaigalat, L., et al. (2011). Control of adhA and sucR expression by the SucR regulator in *Corynebacterium glutamicum*. *J Biotechnol* 152, 77–86. doi:10.1016/j.jbiotec.2011.02.003.

Baumbach, J., Brinkrolf, K., Czaja, L. F., Rahmann, S., and Tauch, A. (2006). CoryneRegNet: An ontology-based data warehouse of corynebacterial transcription factors and regulatory networks. *BMC Genomics* 7, 1–12. doi:10.1186/1471-2164-7-24.

Baumgart, M., Unthan, S., Kloß, R., Radek, A., Polen, T., Tenhaef, N., et al. (2018). *Corynebacterium glutamicum* Chassis C1: Building and Testing a Novel Platform Host for Synthetic Biology and Industrial Biotechnology. *ACS Synth. Biol.* 7, 132–144. doi:10.1021/acssynbio.7b00261.

Beinert, H., Kennedy, M. C., and Stout, C. D. (1996). Aconitase as Iron−Sulfur Protein, Enzyme, and Iron-Regulatory Protein. *Chem. Rev.* 96, 2335–2374. doi:10.1021/cr950040z.

Brune, I., Werner, H., Huser, A. T., Kalinowski, J., Puhler, A., and Tauch, A. (2006). The DtxR protein acting as dual transcriptional regulator directs a global regulatory network involved in iron metabolism of *Corynebacterium glutamicum*. *BMC Genomics* 7, 21. doi:10.1186/1471-2164-7-21.

Cho, H. Y., Lee, S. G., Hyeon, J. E., and Han, S. O. (2010). Identification and characterization of a transcriptional regulator, SucR, that influences *sucCD* transcription in *Corynebacterium glutamicum*. *Biochem. Biophys. Res. Commun.* 401, 300–305. doi:10.1016/j.bbrc.2010.09.057.

Donovan, C., Heyer, A., Pfeifer, E., Polen, T., Wittmann, A., Krämer, R., et al. (2015). A prophage-encoded actin-like protein required for efficient viral DNA replication in bacteria. *Nucleic Acids Res* 43, 5002–5016. doi:10.1093/nar/gkv374.

Hünnefeld, M., Filipchyk, A., and Frunzke, J. (2019a). Spatiotemporal binding dynamics of the xenogeneic silencer CgpS during prophage induction in *Corynebacterium glutamicum*. *Inside this PhD Thesis*.

Hünnefeld, M., Persicke, M., Kalinowski, J., and Frunzke, J. (2019b). The MarR-Type Regulator MalR Is Involved in Stress-Responsive Cell Envelope Remodeling in *Corynebacterium glutamicum*. *Front. Microbiol.* 10, 1039. doi:10.3389/fmicb.2019.01039.

Ikeda, M., and Nakagawa, S. (2003). The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl. Microbiol. Biotechnol.* 62, 99–109. doi:10.1007/s00253-003-1328-1.

Imlay, J. A. (2003). Pathways of Oxidative Damage. *Annu. Rev. Microbiol.* 57, 395–418. doi:10.1146/annurev.micro.57.030502.090938.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5-9. doi:10.1093/nar/gkn201.

Krug, A., Wendisch, V. F., and Bott, M. (2005). Identification of AcnR, a TetR-type repressor of the aconitase gene *acn* in *Corynebacterium glutamicum*. *J. Biol. Chem.* 280, 585–595. doi:10.1074/jbc.M408271200.

Laslo, T., von Zaluskowski, P., Gabris, C., Lodd, E., Rückert, C., Dangel, P., et al. (2012). Arabitol metabolism of *Corynebacterium glutamicum* and its regulation by AtlR. *J Bacteriol* 194, 941–955. doi:10.1128/jb.06064-11.

Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., et al. (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* 43, D213-21. doi:10.1093/nar/gku1243.

Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., and Adhya, S. (2005). Switches in Bacteriophage Lambda Development. *Annu. Rev. Genet.* 39, 409–429. doi:10.1146/annurev.genet.39.073003.113656.

Pfeifer, E., Hünnefeld, M., Popa, O., Polen, T., Kohlheyer, D., Baumgart, M., et al. (2016). Silencing of cryptic prophages in *Corynebacterium glutamicum*. *Nucleic Acids Res.* 44, 10117–10131. doi:10.1093/nar/gkw692.

Trost, E., Götker, S., Schneider, J., Schneiker-Bekel, S., Szczepanowski, R., Tilker, A., et al. (2010). Complete genome sequence and lifestyle of black-pigmented *Corynebacterium aurimucosum* ATCC 700975 (formerly C. nigricans CN-1) isolated from a vaginal swab of a woman with spontaneous abortion. *BMC Genomics* 11, 91. doi:10.1186/1471-2164-11-91.

Wennerhold, J., and Bott, M. (2006). The DtxR regulon of *Corynebacterium glutamicum*. *J*

Bacteriol 188, 2907–2918. doi:10.1128/jb.188.8.2907-2918.2006.

Will, W. R., Bale, D. H., Reid, P. J., Libby, S. J., and Fang, F. C. (2014). Evolutionary expansion of a regulatory network by counter-silencing. *Nat Commun* 5, 5270. doi:10.1038/ncomms6270.

Will, W. R., Brzovic, P., Le Trong, I., Stenkamp, R. E., Lawrenz, M. B., Karlinsey, J. E., et al. (2019). The Evolution of SlyA/RovA Transcription Factors from Repressors to Countersilencers in *Enterobacteriaceae*. *MBio* 10, e00009-19. doi:10.1128/mbio.00009-19.

4 Appendix

4.1 Supplementary Information "Silencing of cryptic prophages in *Corynebacterium glutamicum*"

Supplementary Material to

Silencing of cryptic prophages in *Corynebacterium glutamicum*

Eugen Pfeifer[1], Max Hünnefeld[1], Ovidiu Popa[2], Tino Polen[1], Dietrich Kohlheyer[1], Meike Baumgart[1], and Julia Frunzke[1,*]

Supplementary Tables

Table S1. Strains and plasmids used in this study.

Strains	Relevant characteristics	Reference
E. coli		
DH5α	*supE44* Δ*lacU169* (φ80*lacZ*ΔM15) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*, strain used for cloning procedures	Invitrogen
BL21(DE3)	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* BL21(DE3), protein production host	(1)
S3974	Derivate of K-12 (CGSC #6300), F^-, λ^-, *rph*$^+$ *ilvG*$^+$	(2)
T221	S3974 Δ*hns*$_{FRT}$, *E. coli* strain used for complementation studies of Δ*hns* phenotype	(3)
M. tuberculosis		
H37Rv	wild-type laboratory strain, DNA used as PCR template	ATCC 25618
C. diphtheriae		
ATCC 27010	wild-type laboratory strain, DNA used as PCR template	DSM 44123
C. amycolatum		
PAP 272	wild-type, genomic DNA was used for PCR as template	DSM 44737

C. glutamicum		
ATCC 13032	Biotin-auxotrophic wild type	(4)
WT::*cgpS-strep*	Derivative of ATCC 13032 with genomic exchange of the *cgpS* gene to *cgpS-strep*, encoding a C-terminal Strep-tag fusion.	This study
ATCC 13032 ΔCGP3	ATCC 13032 with in-frame deletion of prophage CGP3 (cg1890-cg2071)	(5)
WT::P_{lys} *-eyfp*	Derivative of ATCC 13032 containing the prophage reporter P_{lys}-*eyfp* integrated into the intergenic region of cg1121-cg1122	(6)
Plasmids		
pAN6	*Kan*^{R}*; C. glutamicum*/*E. coli* shuttle vector for gene expression under control of the *tac* promoter; (P_{tac}, *lacI*^{q}, *pBL*$_1$ *ori*$V_{C.g.}$, pUC18 *ori*$V_{E.c.}$)	(7)
pAN6-*cgpS*	Derivative of pAN6 containing the *cgpS* gene	This study
pAN6-*cgpS-Strep*	Derivative of pAN6 containing the *cgpS* gene without stop codon encoding a C-terminal Strep-tag fusion	This study
pAN6-*N-cgpS*	Derivative of pAN6 containing the first 65 amino acids of the *cgpS* gene	This study
pAN6-*N-cgpS-Strep*	Derivative of pAN6 containing the first 65 amino acids of the *cgpS* gene fused C-terminally to a Strep-tag coding region	This study
pAN6-*lsr2-N-M.tub*	Derivative of pAN6 containing the first 58 amino acids of the *lsr2* gene (Rv3597c) of *Mycobacterium tuberculosis* H37Rv	This study
pAN6-c*gpS-N-C.amyc*	Derivative of pAN6 containing the first 66 amino acids of the homologous *cgpS* gene (CORAM0001_2081) of *Corynebacterium amycolatum* DSM 44373	This study
pAN6-*cgpS-N-C.diph*	Derivative of pAN6 containing the first 59 amino acids of the homologous *cgpS* gene (DIP2266) of *Corynebacterium diphtheriae* DSM 44123	This study

2

pAN6-*alpA-eyfp*	Derivative of pAN6 containing a *alpA-eyfp* fusion	This study
pK19*mobsacB*	Kan^R; plasmid for allelic exchange in *C. glutamicum*; (pK18 $oriV_{E.c.}$, *sacB*, *lacZ*α)	(8)
pK19mobsacB-*cgpS-Strep*	Derivative of pK19*mobsacB* containing the *cgpS-Strep* construct for the allelic exchange of the native *cgpS* gene to a C-terminally strep-tagged version in the chromosome of *C. glutamicum*.	This study
pK18*mobsacB*-	Kan^R; plasmid for integration of foreign DNA into the intergenic region between cg1121-cg1122 ($oriV_{E.c.}$, *sacB*, *lacZ*α).	(5)
pEC-XC99E	cat_I, $lacI^q$, P_{trc}, *rrnB* (T1 and T2), $oriV_{E.c}$, *per* and *repA* (pGA1) $_{C.g.}$. *E. coli* – *C. glutamicum* shuttle and expression vector conferring chloramphenicol resistance.	(9)
pEC-XC99E-*cgpS-mcherry*	Derivative of pEC-XC99E containing the *cgpS* gene cloned upstream of the *mcherry* gene under control of the *tac* promoter.	This study

Table S2. Oligonucleotides used in this study for cloning, qPCR and affinity chromatography. Bold sequences represent the overlapping sequences needed for Gibson assembly (10). Restriction sites are underlined.

Application	Oligo-nucleotide	Sequence (5´→ 3´) and properties	Comment
pK19*mobsacB-cgpS-strep*	LF_cgpS_pK19_fw	**CCTGCAGGTCGACTCTAGAG** CTGGTCGTCTGTGTAGCTAC	PCR product contains an overlapping sequence to *Bam*HI-digested pK19*mobsacB* plasmid
	LF_cgpS_rv	GTCCATAGTCCTAACCAATCATGTAA	
	cgpS_strep_fw	**GATTGGTTAGGACTATGGAC** ATGGCCATTATTCAGTCGGTC	PCR product contains an overlapping sequence to the left flank of *cgpS* (PCR product above)
	cgpS_strep_rv	TTACTTCTCGAACTGTGGGTG	
	RF_cgpS_fw	**CACCCACAGTTCGAGAAGTAA** GAGCCCTGTGGAGAATTGTTG	PCR product contains overlapping sequences to *cgpS-strep* and to an *Eco*RI-digested pK19*mobsacB* plasmid
	RF_cgpS_pK19_rv	**AAAACGACGGCCAGTGAATT** ACGCGGCGACCTCATC	
	Cgps_indel-fw	GGACATTATCACCCAACCACAC	Oligonucleotides to verify the correct integration of *cgpS-strep*
	CgpS_indel_rv	CAAGGAATCGTTTACCTATATCGAG	
			Restriction enzyme
pAN6 with the coding regions for the N-terminal parts of the CgpS/Lsr2 homologs	C.a.fw	GCGC CATATG ATGGCACGCCGCGAACTAAT	*Nde*I
	C.a.fw	CGCG CCCGGG ATGGCACGCCGCGAACTAAT	*Sma*I
	C.a.N.rv	GCGC GCTAGC CTATACAACCGTGCTGTGATCAATAG	*Nhe*I
	C.a.rv	GCGC GGATCC CTAGTTAGCGCTCTCGTACTTTTC	*Bam*HI
	C.d.fw	CGCG CATATG ATGGCACGTCGTGAAATC	*Nde*I
	C.d.fw	CGCG CCCGGG ATGGCACGTCGTGAAATC	*Sma*I
	C.d.N.rv	GCGC GCTAGC CTAGTGCGCTTTTTCTATGAAGGG	*Nhe*I
	C.d.rv	GCGC GGATCC TTAGCGCTTGGTGGACTTAAG	*Bam*HI
	M.t.fw	GCGC CATATG ATGGCGAAGAAAGTAACCGTC	*Nde*I
	M.t.fw	CGCG CCCGGG ATGGCGAAGAAAGTAACCGTC	*Sma*I
	M.t.N.rv	GCGC GCTAGC CTAGACGCGACGGCCCG	*Nhe*I
	M.t.rv	GCGC TCTAGA TCAGGTCGCCGCGTG	*Xba*I

4

pAN6 *cgpS*/ *cgpS-strep*/ *cgpS*-N/ *cgpS*-N-*strep*	cgps_fw	CGCGC CATATG ATGGCCATTATTCAGTCGGTCG	*Nde*I
	cgps_strep_rv	CGCGC GCTAGC TTCGAAAGGAATGCCTTCTTTTTC	*Nhe*I
	cgps_rv	CGCGC GAATTC TTA TTCGAAAGGAATGCCTTC	*Eco*RI
	cgpS_n_rv	CGCGC GCTAGC TTA CTGGCGTGCAGATTCCTC	*Nhe*I
	cgpS_n_strep _rv	CGCGC GCTAGC CTGGCGTGCAGATTCCTC	*Nhe*I
pAN6-*alpA-eyfp*	alpA_OL_pAN 6_fw	**TGCAGAAGGAGATATACATA** ATGGCTCAAAAACAGGACACGAC	PCR product contains overlapping sequences to *Nde*I and *Eco*RI-digested pAN6 plasmid
	eYFP-OL_pAN6_rv	**AAAACGACGGCCAGTGAATT** TTATCTAGACTTGTACAGCTCGTCC	
pEC-XC99E-*cgpS-mcherry*	PcgpS-pEC-fw	**GCGGTATTTCACACCGCATATG** CTGGTCGTCTGTGTAGCTAC	PCR product contains overlapping sequences to *Nde*I -digested pEC-XC99E plasmid and to *mcherry*
	cgpS-rv-OL-mcherry	**CTCGCCCTTGCTCACCAT** TTCGAAAGGAATGCCTTCTTTTTCG	
	mcherry_fw	ATGGTGAGCAAGGGCGAG	PCR product contains an overlapping sequence to *PstI* -digested pEC-XC99E plasmid
	mCherry_rv_OL	**AACAGCCAAGCTTGCATGCC** TTACTTGTACAGCTCGTCCATGC	
Application	**Oligo-nucleotide**	**Sequence (5´→ 3´)**	**Comments**
qPCR (circular phage DNA and reference gene)	Phage-LC-for	CCCACGTTCACCCCACAAACG	
	Phage-LC-rev	CTAAAATGAAGCCATCGCGACC	
	ddh-LC-for	ACGTGCTGTTCCTGTGCATGG	
	ddh-LC-rev	GCTCGGCTAAGACTGCCGCT	
Affinity chromatography with P_{alpAC}	PalpAC-Biotin-Tag fw	*GAGGAGTCGTCGATGTGGAGACC* TCGCACTCAATAATGCGGTGG	Asterisks highlight the biotin labelled sequences
	Biotin-oligo	*GAGGAGTCGTCGATGTGGAGACC*	
	PalpAC rv	GCGCATACGCACATTACGC	

5

Table S3. Oligonucleotides used for the generation of DNA fragments for EMSA experiments.

Oligonucleotide	Sequence (5´→ 3´) and properties	Product length (bp)	GC content of product (%)
gntK-Prom-fw	ATGGTGGCGTCATGCTCGGCCG	560	49.3
gntK-Prom-rv	GGATTTGCCGCAGCCAGAAACGC		
cg0150fw	GGGGTAATAAGACAAAACAGTGGG	500	39.6
cg0150rv	TAGAAATCAGCGACAACCATGCTTC		
cg0421fw	GGATACTTTCTGTTTTGGTTGGTC	500	41.5
cg0421rv	GAAATTACCAAGATGCACCACCTC		
cg0432fw	CCTTTTCTAGACAAGACCTGATC	500	42.0
cg0432rv	ACCAACGACGTCGGATTAGG		
cg0718fw	ATAAGTCATGGTTCAACCTCGG	500	44.0
cg0718rv	CCTAAAACGACACCATCTCAAAAG		
cg0726fw	TACCACTTGCCTTTGTAGCGTTC	500	46.0
cg0726rv	ACTTGGAAACCGGCAGCAAG		
cg1028fw	TGGTCAGCGCAGCGAC	500	50.3
cg1028rv	AAGTTGAGTCTTGGGCCGG		
cg1517fw	GTATGACCAAATGGGACGAAGG	500	42.0
cg1517rv	GATAAGCCACTCAACCACCAAAC		
cg2782fw	GACGCTGAGAAGGACTACG	500	49.5
cg2782rv	TTGAAGGTATCTCCGACAGCAAC		
cg2805fw	AAGAAGGCTGAGTTTAGTGGGG	500	44.8
cg2805rv	AGAAGACGTCCAAAATCCCGTC		
cg3060fw	CAAAATCAATGCGAGAGCGAAG	500	44.0
cg3060rv	CTGCAGAGCTGAAATTATCGAC		
cg3304fw	GGATAACTTCCCCACAATTGAC	500	47.7
cg3304rv	AAGCGTGCCATTGTTCTCCC		
cg1951fw	CTCTATTGGCTCTTAATGGTCAATTAC	500	33.4
cg1951rv	GCCTCTTAAAGCACAGTTATTGCG		
cg1966fw	GCTCAGTATCAATGTCGTCACC	500	36.3
cg1966rv	GTCGAAGTGGTGTCGTTATTTAGG		
cg2023fw	GCACCACCAACAAGTGCC	500	40.7
cg2023rv	TGGGAGCATTTCACTGCACG		
cg1977fw	GTTCTAAACATAAGGAACGCGC	500	39.1
cg1977rv	CGATGGTGCAGTGACCATG		
cg1936fw	CATCGCTCATTGTTACTTAATTACCC	500	36.0
cg1936rv	CCTGAAGAATTTGCTCAGCCG		
cg1940fw	CCATAGTCAAGATTCCCAATCAAC	500	39.5
cg1940rv	GATTCAGGTGATGTAGCGCTG		
cg1917fw	CCTGTAGCCTGCGACGTTAA	500	42.2
cg1917rv	GTGCACCGGTAGCCATAATAG		
cg1895fw	TCACGGGTGGAATCGGAG	500	38.3
cg1895rv	GCTTGGATCATCTGAACAGAGTG		
cg2014fw	AGCGTCAATCGGAATCTGCG	500	40.7
cg2014rv	CAGTTGCGCTAGATAAGCGAG		
cg1890fw	GCGACAAACAAATAGATCAGCTG	500	41.8
cg1890rv	GGGGTTTATTACCTGCCTGC		

6

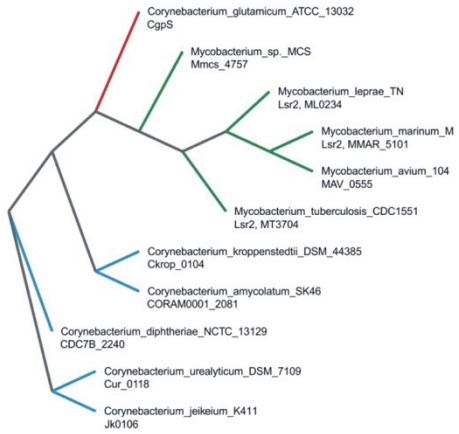
Table S4. Results of the ChAP-Seq experiment. The 90 identified regions are evaluated regarding their peak width, peak maxima and area. Furthermore, the regions are classified into three categories as described in Figure S3. Genes within the CGP3 region are highlighted in green.

Table S5. Impact of CgpS countersilencing on the *C. glutamicum* transcriptome. CGP3 prophage genes are highlighted in green. ORFs exhibiting are more than two-fold altered mRNA ratio (of >2 or < 0.5, *p*-value <0.05) are shown.

Table S6. PSI-BLAST results of CgpS. e-value was set <= 0.005 across several orders of the phylum Actinobacteria and phages as annotated in the NCBI database (http://www.ncbi.nlm.nih.gov/).

Supplementary Figures

A



B

	Corynebacterium glutamicum ATCC 13032	100	27.52	25.66	25.96	27.27	25.45	21.55	23.42	24.76	24.55	21.1
1	*Mycobacterium spec.* MCS	27.52	100	87.93	90.09	87.83	86.09	28.45	27.35	30.77	36.61	34.82
2	*Mycobacterium marinum* M	25.66	87.93	100	93.69	94.87	94.02	26.89	29.91	28.57	40.18	35.71
3	*Mycobacterium avium* 104	25.96	90.09	93.69	100	91.82	91.82	27.27	27.93	30.3	36.45	35.19
4	*Mycobacterium tuberculosis* CDC1551	27.27	87.83	94.87	91.82	100	92.31	28.45	31.03	28.85	40.54	34.23
5	*Mycobacterium leprae* TN	25.45	86.09	94.02	91.82	92.31	100	27.59	28.45	28.85	37.84	35.14
6	*Corynebacterium kroppenstedtii* DSM 44385	21.55	28.45	26.89	27.27	28.45	27.59	100	41.32	39.81	40.35	38.6
7	*Corynebacterium amycolatum* SK46	23.42	27.35	29.91	27.93	31.03	28.45	41.32	100	40.74	41.74	43.48
8	*Corynebacterium diphtheria* NCTC 13129	24.76	30.77	28.57	30.3	28.85	28.85	39.81	40.74	100	44.86	49.06
9	*Corynebacterium urealyticum* DSM 7109	24.55	36.61	40.18	36.45	40.54	37.84	40.35	41.74	44.86	100	53.51
10	*Corynebacterium jeikeium* K411	21.1	34.82	35.71	35.19	34.23	35.14	38.6	43.48	49.06	53.51	100

Figure S1: CgpS orthologs. A. Phylogenetic tree based on the multiple sequence alignments of CgpS/Lsr2 homologs of selected *Corynebacteria* (*C. kroppenstedtii*, *C. amycolatum*, *C. diphteriae*, *C. urealyticum*, *C. jeikeium*), and *Mycobacteria* (*M. tuberculosis*, *M. spec.*, *M. leprae*, *M. marinum*, *M. avium*). Alignments were performed using Clustal Omega (11) with standard configurations. Data for phylogenetic tree were derived from alignments and visualized using tree vector (12). Analysis indicates that CgpS displays a higher sequence identity to mycobacterial Lsr2 proteins than to the corynebacterial orthologs.

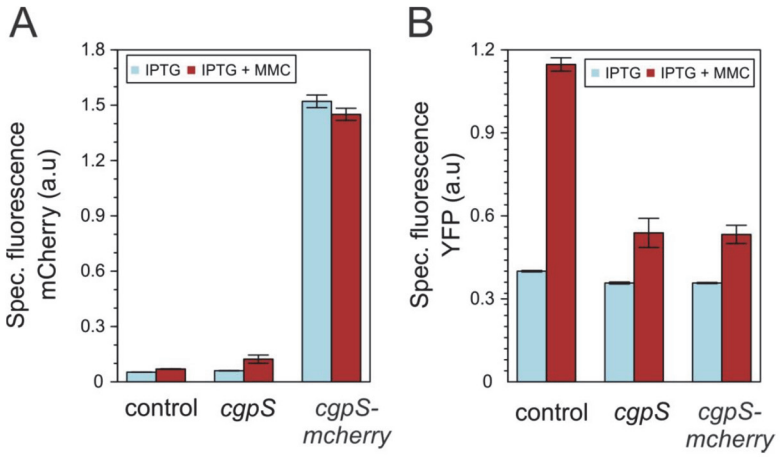


Figure S2: Silencing of CGP3 prophage induction. A and B. Phage reporter cells (WT::P_{lys}-*eyfp*) were transformed with pAN6, pAN6-*cgpS* and pAN6-*cgpS-mcherry* and were cultivated in CGXII with 50 µM IPTG and in the presence or absence of 0.6 µM MMC. The mCherry (A) and eYFP (B) fluorescence as well as backscattered light were measured in the BioLector® microcultivation system and were used to calculate the specific fluorescence. The specific fluorescence after 20 h of cultivation is shown. The data represent average values from three biological replicates including the standard deviation.

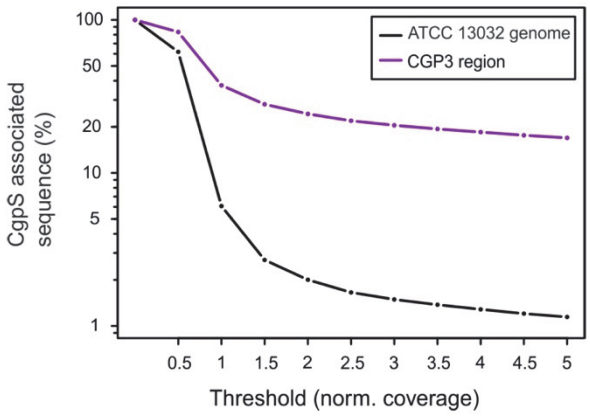


Figure S3: Threshold variation of the CgpS ChAP-Seq data. Based on mean normalized coverage values which were obtained by ChAP-sequencing experiments, thresholds were varied to validate its impact on the estimated binding of CgpS to the CGP3 region and to the entire genome of ATCC 13032. Based on this analysis, bound regions showing a threshold T >3 were considered as CgpS targets in this study (20.46% of CGP3 and 1.49% of the genome).

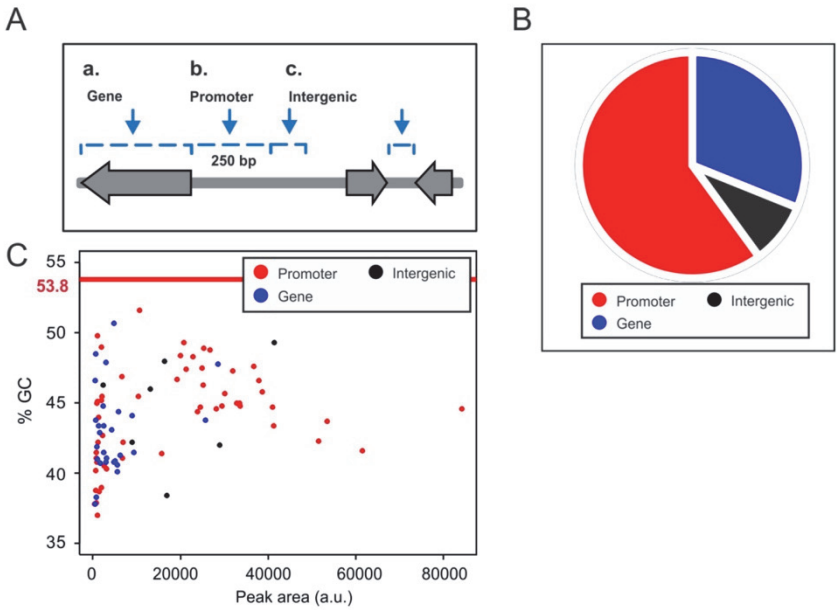


Figure S4: **Genomic distribution of CgpS binding sites within genes, promoters or intergenic regions. A.** The 90 regions bound to CgpS were classified into three categories: i. Binding sites within open reading frames (genes), ii. 250 bp upstream of translational start or according to published transcription start sites (promoter regions), and, iii. intergenic regions. **B**. Distribution of the 90 CgpS-bound genomic regions. Overall, 60% of the peaks are located in promoter regions and 31% within genes. Only 9% are assigned to intergenic regions**. C**. The %GC content of the regions were plotted against peak areas. Red line illustrates average GC content of *C. glutamicum* ATCC 13032, which is about 53.8% (13). Interestingly, a trend to higher peak areas was observed for promoter regions in comparison to intergenic regions or ORFs.

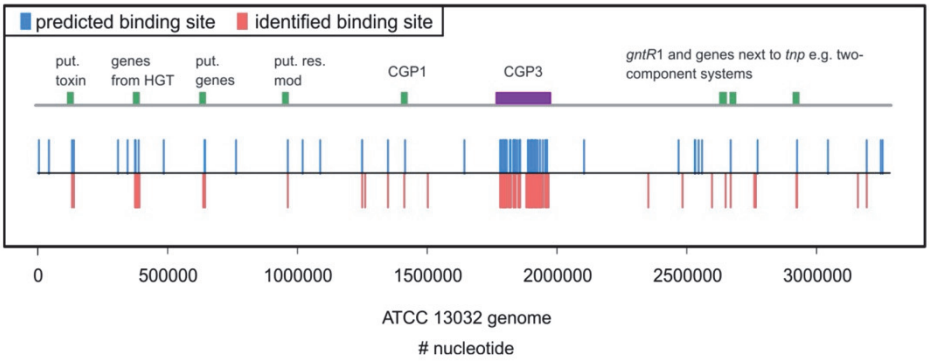
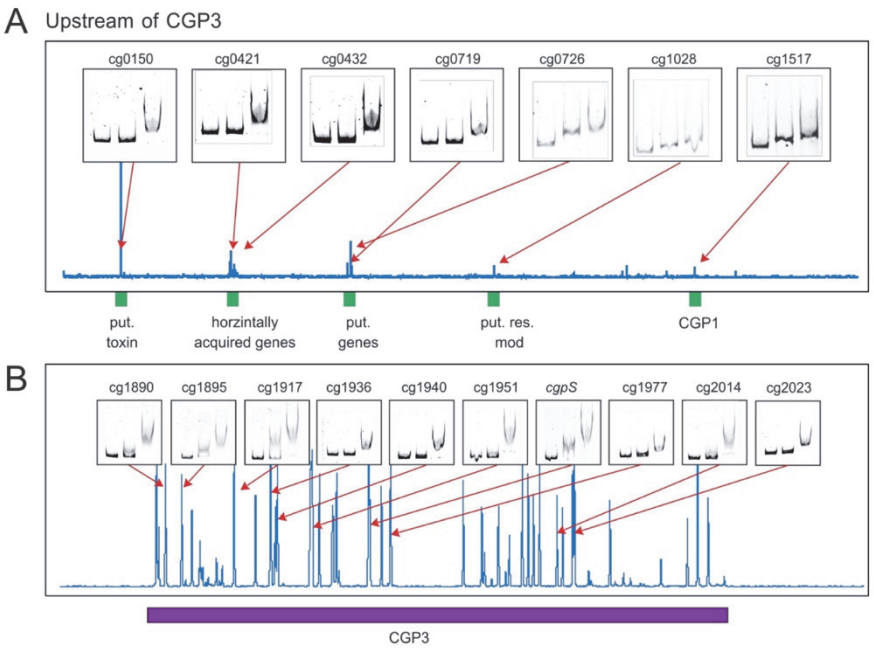


Figure S5: Comparison of predicted and experimentally identified CgpS binding sites. The DNA binding motif derived from ChAP-Seq results (Fig. 3C) was checked for further hits in the genome of ATCC 13032 using FIMO (14). Here, 90 positions exhibiting highest probability (p-Values: $2.7 \cdot 10^{-10}$ – $2.3 \cdot 10^{-6}$) (in blue) were compared with the 90 experimentally identified binding sites acquired by ChAP-Seq binding studies (in red). Potential CgpS site within the CGP3 region (purple boxes) and outside (green boxes) are highlighted. Correlation between experimentally identified and predicted CgpS binding sites ~75 %.



D

cg0150	putative transcriptional regulator, Fic/Doc- family	**cg1936**	putative secreted protein, CGP3 region
cg0421	*wzx*, putative translocase involved in export of a cell surface polysaccaride	**cg1940**	putative secreted protein, CGP3 region
cg0432	putative lipopolysaccharide modification acyltransferase	**cg1951**	*tnp*14a, transposase fragment, CGP3 region
cg0719	*crtYe*, C50 carotenoid cyclase, terpenoid synthesis	**cg1966**	*cgpS*, Lsr2-like protein, CGP3 region
cg0726	putative secreted lipoprotein	**cg1977**	putative secreted protein, CGP3 region
cg1028	putative restriction modification system: methylase (EC:2.1.1.72)	**cg2014**	hypothetical protein, CGP3 region
cg1517	putative secreted protein, CGP1 region	**cg2023**	putative membrane protein, CGP3 region
cg1890	*alpC*, actin like protein, CGP3 region	**cg2783**	*gntR1*, gluconate-responsive repressor
cg1895	putative secreted protein, CGP3 region	**cg2806**	putative membrane protein
cg1917	hypothetical protein, CGP3 region	**cg3060**	*cgtS6*, two-component sensor kinase
		cg3304	*dnaB*, replicative DNA helicase

13

Figure S6: *In vitro* binding studies of CgpS to its putative target sites. Electrophoretic mobility shift assays (EMSAs) were performed with purified CgpS-Strep protein and 21 putative target DNA regions derived from ChAP-Seq data (Fig. 3). Green boxes indicate regions outside of CGP3 and the purple box sites within the CGP3 region. All tested DNA fragments had a size of about 500 bp and were chosen 250 bp up and downstream of the peak maxima, which were detected by the ChAP-Seq analysis. Overall, eleven candidate regions were chosen outside of CGP3 ((**A**) seven upstream and (**C**) four downstream of CGP3) and ten sites within the CGP3 region (**B**). In all lanes 90 ng DNA (12-14 pM) were incubated without (lane 1) or with increasing amounts of CgpS protein (lane 2: 1 µM and lane 3: 2 µM). The promoter region of *gntK* (560 bp) was used as a negative control. Annotations und potential functions of the bound regions are listed in **D**.

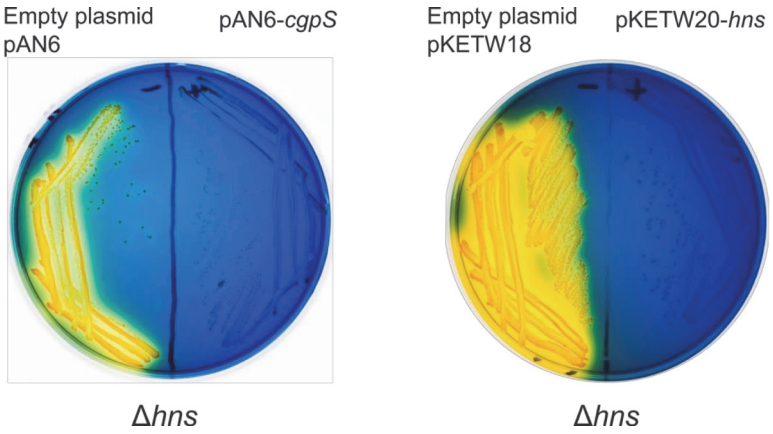


Figure S7: Complementation studies of a *E. coli* K-12 Δ*hns* strain with *cgpS* cloned into the overexpression plasmid pAN6. Cells were grown on bromothymol blue salicin indicator plates as described in Dole et al., 2002 (15). *E. coli* cells lacking *hns* were transformed with the empty plasmid pAN6, pAN6-cgpS or with the empty plasmid pKETW18 or pKETW20 carrying *hns*. Plates were incubated at 37°C overnight. Complementation is based on the utilization of salicin. Salicin can be used as carbon source if the *bgl* operon is expressed. This operon is repressed by H-NS in the wild type situation. Thus, in the absence of H-NS, salicin is metabolized leading to a decrease of the pH resulting in a colour shift from blue to yellow. Complementation of the Δ*hns* phenotype was achieved by expressing either *hns* or *cgpS* suggesting a similar function of both proteins.

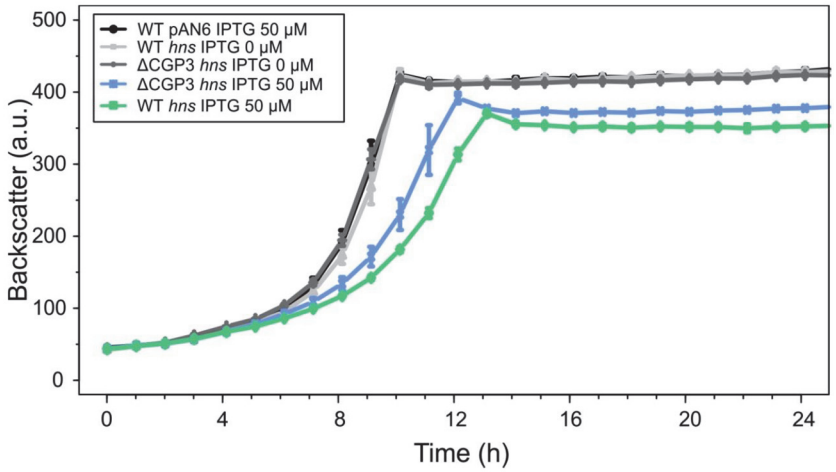


Figure S8: Overexpression of *hns* in *C. glutamicum* strains. H-NS encoding gene located on the overexpression plasmid pAN6 was overexpressed in the prophage reporter strain WT::P_{lys}-*eyfp* and in the ΔCGP3 strain. Cells were cultivated in CGXII minimal medium and *hns* expression was induced with 50 µM IPTG. The data represent average values of three biological replicates including the standard deviation.

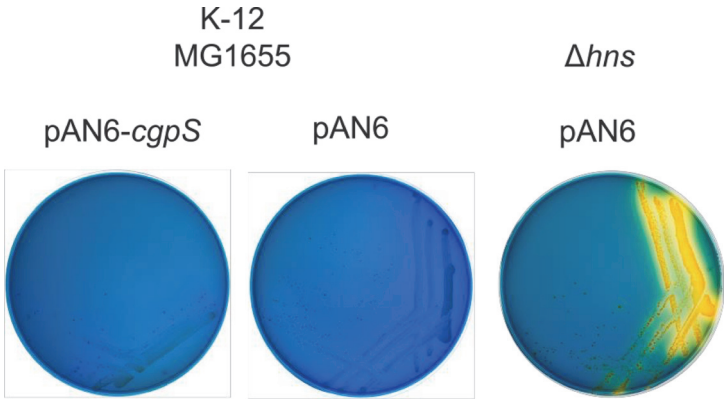


Figure S9: *cgpS* overexpression in *E. coli* wild type cells. To verify whether CgpS is interfering with the function of H-NS in its native host, *E. coli* K-12 MG1655 wild type cells were transformed with the pAN6-*cgpS* plasmid. Cells were streaked on bromothymol blue salicin indicator plates (15) supplemented with 100 µM IPTG. As control, the wild type strain and a Δ*hns* mutant were transformed with the empty plasmid pAN6. The obtained results suggest that heterologous *cgpS* expression is not able to counteract H-NS silencing at the *bgl* promoter when compared to a mutant lacking the *hns* gene. However, it needs to be highlighted that the resulting *E. coli* strain expressing the *cgpS* gene (left plate) showed a significant growth defect in comparison to the empty vector controls (middle and right).

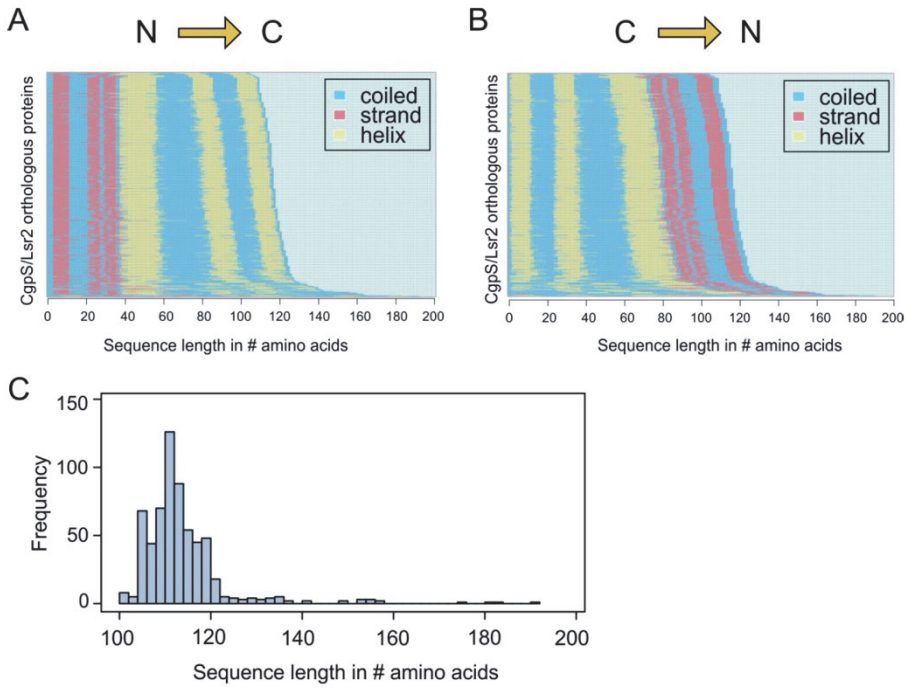


Figure S10. Bioinformatic analysis of CgpS related proteins. A PSI-BLAST search on CgpS homologs with an *e*-value of 0.005 was conducted and achieved 5230 hits (Table S6). 1920 sequence are individual and can be assigned to 863 taxonomical units; 618 of these can be allocated to bacteria or phages. Secondary structure predictions of the 618 sequences are shown in direct comparison in N->C (**A**) and C->N (**B**) orientation. The increasing length of the amino acid sequences entails distortet matches in secondary structure prediction and hence for a better overview the two possibilites are shown. **C.** Histogramm of the 618 sequences ordered according to their amino acid sequence length. The maximum of this distribution is located around 110 amino acids.

Supplementary Videos

Video S1: Time lapse video of a *C. glutamicum* microcolony under standard conditions (without IPTG, control). Cells of the prophage reporter strain ATCC 13032::P_{lys}-*eyfp* carrying the countersilencing plasmid pAN6-N-*cgpS* were cultivated in microfluidic chambers (16) in standard minimal medium (CGXII with 2% (w/v) glucose, 25 µg·ml kanamycin for 20 h without IPTG). The video shows the first 12 h of the cultivation.

Video S2: Time lapse video of the effect of CgpS countersilencing (150 µM IPTG) on prophage activation. The same reporter strain (Video S1) was grown in the presence of 150 µM IPTG inducing the expression of the truncated CgpS protein (aa 1-65) covering its oligomerization domain. The video shows the first 16.5 h of the experiment.

References

1. Studier, F.W. and Moffatt, B.A. (1986) Use of Bacteriophage-T7 Rna-Polymerase to Direct Selective High-Level Expression of Cloned Genes. *J Mol Biol*, **189**, 113-130.
2. Venkatesh, G.R., Koungni, F.C.K., Paukner, A., Stratmann, T., Blissenbach, B. and Schnetz, K. (2010) BglJ-RcsB Heterodimers Relieve Repression of the *Escherichia coli bgl* Operon by H-NS. *J Bacteriol*, **192**, 6456-6464.
3. Stratmann, T., Pul, U., Wurm, R., Wagner, R. and Schnetz, K. (2012) RcsB-BglJ activates the *Escherichia coli* leuO gene, encoding an H-NS antagonist and pleiotropic regulator of virulence determinants. *Mol Microbiol*, **83**, 1109-1123.
4. Kinoshita, S., Udaka, S. and Shimono, M. (1957) Studies on the amino acid fermentation - Part I. Production of L-glutamic acid by various microorganisms. *J Gen Appl Microbiol*, **50**, 331-343.
5. Baumgart, M., Unthan, S., Rückert, C., Sivalingam, J., Grünberger, A., Kalinowski, J., Bott, M., Noack, S. and Frunzke, J. (2013) Construction of a prophage-free variant of *Corynebacterium glutamicum* ATCC 13032 for use as a platform strain for basic research and industrial biotechnology. *Appl Environ Microbiol*, **79**, 6006-6015.
6. Helfrich, S., Pfeifer, E., Krämer, C., Sachs, C.C., Wiechert, W., Kohlheyer, D., Nöh, K. and Frunzke, J. (2015) Live cell imaging of SOS and prophage dynamics in isogenic bacterial populations. *Mol Microbiol*, **98**, 636-650.
7. Frunzke, J., Engels, V., Hasenbein, S., Gätgens, C. and Bott, M. (2008) Co-ordinated regulation of gluconate catabolism and glucose uptake in *Corynebacterium glutamicum* by two functionally equivalent transcriptional regulators, GntR1 and GntR2. *Mol Microbiol*, **67**, 305-322.
8. Schäfer, A., Tauch, A., Jager, W., Kalinowski, J., Thierbach, G. and Pühler, A. (1994) Small Mobilizable Multipurpose Cloning Vectors Derived from the *Escherichia-Coli* Plasmids Pk18 and Pk19 - Selection of Defined Deletions in the Chromosome of *Corynebacterium-Glutamicum*. *Gene*, **145**, 69-73.
9. Kirchner, O. and Tauch, A. (2003) Tools for genetic engineering in the amino acid-producing bacterium Corynebacterium glutamicum. *J Biotechnol*, **104**, 287-299.
10. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343-U341.
11. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**, 539.
12. Pethica, R., Barker, G., Kovacs, T. and Gough, J. (2010) TreeVector: Scalable, Interactive, Phylogenetic Trees for the Web. *Plos One*, **5**.
13. Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B.J., Gaigalat, L. *et al.* (2003) The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol*, **104**, 5-25.
14. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017-1018.
15. Dole, S., Kühn, S. and Schnetz, K. (2002) Post-transcriptional enhancement of *Escherichia coli bgl* operon silencing by limitation of BglG-mediated antitermination at low transcription rates. *Mol Microbiol*, **43**, 217-226.
16. Grünberger, A., Probst, C., Helfrich, S., Nanda, A., Stute, B., Wiechert, W., von Lieres, E., Nöh, K., Frunzke, J. and Kohlheyer, D. (2015) Spatiotemporal Microbial Single-Cell Analysis Using a High-Throughput Microfluidics Cultivation Platform. *Cytometry*, **87A**, 1101-1115.

4.2 Supplementary Information “Impact of Xenogeneic Silencing on Phage-Host Interactions”

Supplementary information of this review consists of two data sheets:

S1: All Actinophages. 2626 actinophages from PhagesDB were allocated into temperate and virulent phages using PHACTS after the genes were predicted by prodigal. Lsr2-encoding phages were marked by the Lsr2 annotation taken from PhagesDB.

S2: Lsr2_hits_draft_0_005. BLAST output of all Lsr2 hits using the amino acid sequence from Lsr2 (WP_003419513.1) on the database PhagesDB with an lower e-value than 0.005. Not shown are hits from draft genomes.

4.3 Supplementary Information "Spatiotemporal binding dynamics of the xenogeneic silencer CgpS during prophage induction in *Corynebacterium glutamicum*"

Table S1: Filtered dataset of genome wide time-resolved CgpS binding (ChAP-Seq). *C. glutamicum* ATCC 13032::*cgpS-strep* was cultivated in CGXII minimal medium supplemented with 2% (w/v) glucose and 600 nM mitomycin C. Cells were harvested at different time points as described in the materials and methods section. The dataset was filtered for peaks less then 700 bp away from ATG. Additionally, a peak was only added to this table, if the peak-score (quality measurement depending on peak height and peak shape) for at least one time point was ≥ 5. Column A, B and C show the locus as NCgl as well as Cg number and the corresponding gene name. Column D and E show the start and end point of each centered peak region. Column F indicates the distance of the peak maximum to the translational start site of the gene in column A/B/C. For all known transcriptional start sites, the distance to the TSS is indicated in column F. Column H - N show ChAP-Seq maximum peak coverages at 0 h, 0.25 h, 0.5 h, 1 h, 3 h, 6 h and 24 h after mitomycin C addition. The annotation is shown in column O and the functional prediction in column P (Annotation data extracted and modified from Baumgart et al., 2018).

NCgl number	Cg number	Gene name	Peak start	Peak end	Distance ATG	Distance to TSS	ChAP-Seq t=0	t= 0.25	t= 0.5	t= 1	t=3	t = 6	t= 24	Annotation	Functional category
NCgl0002	cg0004	*dnaN*	1790	2091	351	222	5.8	6.0	1.0	13.7	21.7	39.2	21.8	DNA polymerase III subunit beta	DNA replication, recombination, repair, and degradation
NCgl0003	cg0005	*recF*	3438	3739	4	63	3.0	1.0	1.0	1.0	1.0	2.1	2.9	recombination protein F	DNA replication, recombination, repair, and degradation
NCgl0026	cg0041	*znuA2*	28971	29272	5	-	1.0	1.0	1.0	2.1	1.0	2.6	1.0	ABC transporter periplasmic component	Inorganic ion transport, metabolism, and storage
NCgl0027	cg0042	*znuB2*	29827	30128	13	-	1.0	1.0	1.0	1.9	3.8	4.3	3.8	ABC transporter permease	Inorganic ion transport, metabolism, and storage
NCgl0036	cg0052	-	36404	36705	690	774	1.0	1.0	1.0	1.6	1.0	5.3	1.0	ABC transporter permease	Inorganic ion transport, metabolism, and storage
NCgl0057	cg0077	-	58871	59172	69	-	1.0	1.0	1.0	1.0	2.4	2.8	2.9	hypothetical protein	Unknown function
NCgl0074	cg0103	*crnT*	83404	83705	13	-	5.0	1.0	1.0	2.2	1.0	1.0	1.0	permease	Amino acid transport and metabolism
NCgl0089	cg0119	*ureD*	95158	95459	208	34	1.0	1.0	1.0	1.0	2.2	2.5	2.1	urease accessory protein UreH	Transport and metabolism of further metabolites
NCgl0100	cg0133	*abgT*	112271	112572	48	4	1.0	1.0	1.0	1.0	1.7	2.2	3.1	p-aminobenzoyl-glutamate transporter	Amino acid transport and metabolism
NCgl0113	cg0149	*panB*	128031	128332	184	-	4.2	1.0	1.0	1.0	7.6	1.0	1.0	3-methyl-2-oxobutanoate hydroxymethyltransferase	Coenzyme transport and metabolism
NCgl0113	cg0149	*panB*	128318	128619	471	-	5.8	1.0	3.7	1.0	10.2	29.3	14.5	3-methyl-2-oxobutanoate hydroxymethyltransferase	Coenzyme transport and metabolism
NCgl0114	cg0150	-	129294	129595	585	-	63.5	39.3	40.4	102.5	105.6	116.1	98.3	hypothetical protein	Signal transduction mechanisms
NCgl0123	cg0160	-	136481	136782	67	-	1.0	2.8	1.0	2.8	11.7	18.4	10.3	hypothetical protein	Unknown function
NCgl0139	cg0178	*hrpB*	155722	156023	260	-	1.0	1.0	1.0	1.0	2.7	2.5	2.8	HrpA-like helicase	Transcription including sigma factors, RNA processing and modification
NCgl0171	cg0215	*cspA*	185770	186071	587	407	1.0	1.0	1.0	1.0	1.0	4.0	3.6	cold shock protein	Transcription including sigma factors, RNA processing and modification
NCgl0172	cg0216	-	186667	186968	59	49	1.0	1.0	1.0	1.0	3.0	3.1	3.7	hypothetical protein	Unknown function
NCgl0176	cg0221	-	189812	190113	227	-	3.8	1.0	1.0	7.8	18.8	35.9	22.9	transcriptional regulator	Signal transduction mechanisms
NCgl0235	-	-	255422	255723	135	-	1.0	1.0	2.7	3.4	1.4	1.4	1.5	transposase	DNA replication, recombination, repair, and degradation
NCgl0246	cg0304	-	269059	269360	86	-	2.4	1.0	1.0	2.3	4.1	6.1	4.5	hypothetical protein	Unknown function
NCgl0249	cg0308	-	271547	271848	63	40	1.0	1.0	1.0	1.0	1.0	2.6	1.8	hypothetical protein	Unknown function
NCgl0256	cg0316	-	278213	278514	87	-	1.0	1.0	1.0	1.0	5.0	16.6	5.3	hypothetical protein	Unknown function
NCgl0286	cg0350	*glxR*	307531	307832	220	-	1.0	1.0	1.0	1.0	3.4	4.5	2.8	cAMP-binding domain-containing protein	Signal transduction mechanisms
NCgl0294	cg0360	-	314400	314701	87	-	1.0	1.0	1.5	1.0	3.1	4.8	4.3	phosphoserine phosphatase	General function prediction only
NCgl0310	cg0384	*rluC1*	331235	331536	166	166	1.0	1.0	1.9	1.5	1.0	2.8	1.0	23S RNA-specific pseudouridylate synthase	Translation, ribosomal structure and biogenesis
NCgl0322	cg0397	*ushA*	343457	343758	28	163	1.0	1.0	1.0	1.8	3.0	3.3	2.7	5'-nucleotidase	DNA replication, recombination, repair, and degradation; Inorganic ion transport, metabolism, and storage
NCgl0323	cg0399	-	346807	347108	498	-	1.0	1.0	1.0	3.1	5.4	13.7	7.8	hypothetical protein	Unknown function
NCgl0334	cg0411	-	360427	360728	262	-	4.3	1.0	1.0	1.0	2.4	6.2	3.6	hypothetical protein	Unknown function
NCgl0343	cg0420	-	370317	370618	424	49	1.0	3.9	3.6	9.5	25.5	58.1	22.6	glycosyltransferase	General function prediction only
NCgl0344	cg0421	*wzx*	371680	371981	98	58	1.0	1.0	1.0	1.0	1.0	35.3	1.0	O-antigen and teichoic acid membrane export protein	Cell wall/membrane/envelope biogenesis
NCgl0345	cg0422	*murA*	373125	373426	224	140	14.1	1.0	1.0	47.4	110.0	172.0	67.9	UDP-N-acetylglucosamine enolpyruvyl transferase	Cell wall/membrane/envelope biogenesis
NCgl0345	cg0422	*murA*	373348	373649	1	48	14.8	8.5	10.1	1.0	1.0	172.0	67.9	UDP-N-acetylglucosamine enolpyruvyl transferase	Cell wall/membrane/envelope biogenesis
NCgl0346	cg0423	*murB*	374332	374633	350	102	10.1	5.3	6.4	17.9	32.1	49.6	11.9	UDP-N-acetylmuramate dehydrogenase	Cell wall/membrane/envelope biogenesis
NCgl0346	cg0423	*murB*	374469	374770	213	35	10.1	5.3	6.4	17.9	32.1	49.6	11.9	UDP-N-acetylmuramate dehydrogenase	Cell wall/membrane/envelope biogenesis
NCgl0347	cg0424	-	375032	375333	659	133	6.5	1.0	1.0	8.2	1.0	1.0	1.0	cell wall biogenesis glycosyltransferase	General function prediction only
NCgl0347	cg0424	-	375269	375570	422	148	6.5	1.0	1.0	1.0	26.9	53.7	16.3	cell wall biogenesis glycosyltransferase	General function prediction only
NCgl0349	cg0431	-	380006	380307	320	-	1.0	1.0	1.0	4.0	6.1	1.0	1.0	hypothetical protein	General function prediction only
NCgl0350	cg0432	-	380745	381046	54	-	8.2	5.9	4.7	18.3	50.5	84.0	42.0	acyltransferase	Cell wall/membrane/envelope biogenesis
NCgl0350	cg0432	-	380887	381188	196	-	8.2	5.9	4.7	18.3	1.0	84.0	1.0	acyltransferase	Cell wall/membrane/envelope biogenesis
NCgl0351	cg0435	*udgA1*	381630	381931	167	79	2.8	3.9	2.7	7.9	24.5	51.8	23.2	UDP-glucose 6-dehydrogenase	Nucleotide transport and metabolism
NCgl0353	cg0438	-	386048	386349	4	-	5.2	5.1	3.4	10.4	20.0	64.1	19.6	cell wall biogenesis glycosyltransferase	General function prediction only
NCgl0355	cg0441	*lpd*	387058	387359	483	36	1.0	1.0	1.0	2.4	1.0	5.7	4.5	dihydrolipoamide dehydrogenase	Coenzyme transport and metabolism
NCgl0356	cg0442	*galU2*	389096	389397	46	1	4.6	4.0	3.0	7.0	12.9	22.1	10.1	UDP-glucose pyrophosphorylase	Carbon source transport and metabolism
NCgl0376	cg0465	-	410423	410724	98	-	1.0	1.0	1.0	2.4	2.0	3.8	3.4	hypothetical protein	Unknown function
NCgl0399	cg0492	-	435712	436013	86	19	1.0	1.0	1.0	2.0	2.4	3.3	2.6	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl0445	cg0545	*pitA*	482761	483062	136	-	1.0	3.8	2.8	4.8	12.6	27.8	14.5	phosphate/sulphate permease	Inorganic ion transport, metabolism, and storage
NCgl0478	cg0583	*fusA*	523649	523950	84	20	1.0	2.8	1.0	1.0	3.5	6.8	4.4	elongation factor G	Translation, ribosomal structure and biogenesis
NCgl0484	cg0591	-	531720	532021	80	-	3.2	1.0	1.0	1.0	1.7	3.2	2.7	ABC transporter permease	Inorganic ion transport, metabolism, and storage; Transport and metabolism of further metabolites
NCgl0522	cg0636	*creB*	558862	559163	44	-	1.0	1.0	1.0	2.1	2.4	4.9	2.6	hypothetical protein	Unknown function
NCgl0535	cg0650	-	572238	572539	41	41	1.0	1.0	1.0	1.0	2.2	3.1	2.3	hypothetical protein	Unknown function
NCgl0551	cg0666	-	588938	589239	65	-	1.0	1.0	1.0	1.0	1.0	2.2	2.8	hypothetical protein	Unknown function
NCgl0554	cg0671	-	593668	593969	116	-	4.3	1.0	1.0	2.5	5.0	12.7	9.0	hypothetical protein	Unknown function
NCgl0559	cg0676	-	597557	597858	184	56	1.0	1.8	1.0	1.0	1.0	2.7	2.0	hypothetical protein	Unknown function
NCgl0565	cg0683	-	602502	602803	158	4	1.0	1.0	1.0	1.0	3.8	5.7	3.0	hypothetical protein	General function prediction only

NCgl0575	cg0696	*sigD*	613018	613319	13	46	1.0	1.0	1.0	1.0	2.3	4.1	2.9	RNA polymerase sigma factor SigD	Transcription including sigma factors, RNA processing and modification
NCgl0590	cg0713	-	627444	627745	168	168	1.0	1.0	1.0	3.6	5.9	13.7	4.7	hypothetical protein	Unknown function
NCgl0595	cg0718	*crtYf*	633196	633497	268	-	1.0	1.0	1.9	6.5	22.9	41.3	18.7	C50 carotenoid epsilon cyclase	Transport and metabolism of further metabolites
NCgl0601	cg0725	-	639423	639724	50	1369	26.5	20.0	8.7	38.0	68.8	102.0	37.9	MarR family transcriptional regulator	signal transduction mechanisms
NCgl0602	cg0726	-	640710	641011	18	-	15.4	11.0	8.3	16.6	31.6	38.2	19.4	lipocalin	Unknown function
NCgl0604	cg0728	*phr*	644413	644714	563	-	1.0	1.0	1.0	2.2	4.4	6.9	5.2	deoxyribodipyrimidine photolyase	DNA replication, recombination, repair, and degradation
NCgl0615	cg0742	-	657919	658220	68	-	1.0	1.0	1.0	1.7	1.8	3.5	1.0	hypothetical protein	Unknown function
NCgl0625	cg0755	*metY*	667759	668060	140	-	3.0	3.1	1.0	1.8	3.9	7.1	4.1	O-acetylhomoserine aminocarboxypropyltransferase	Amino acid transport and metabolism
NCgl0670	cg0802	*accBC*	720254	720555	53	-	1.0	1.0	2.0	1.0	1.0	2.5	1.0	acyl-CoA carboxylase	Lipid transport and metabolism
NCgl0671	cg0803	*thtR*	721407	721708	109	-	5.0	1.0	1.0	1.0	3.1	7.2	7.2	thiosulfate sulfurtransferase	Inorganic ion transport, metabolism, and storage
NCgl0677	cg0811	*dtsR2 (accD2)*	728469	728770	268	-	3.4	1.0	1.0	1.0	1.9	3.9	1.9	detergent sensitivity rescuer dtsR2	Cell wall/membrane/envelope biogenesis
NCgl0687	cg0823	*ntaA*	738772	739073	394	-	1.0	1.0	1.0	1.0	3.5	1.3	3.6	nitrilotriacetate monooxygenase	Transport and metabolism of further metabolites
NCgl0687	cg0823	*ntaA*	738997	739298	619	-	1.0	1.0	14.9	1.0	1.0	1.0	3.6	nitrilotriacetate monooxygenase	Transport and metabolism of further metabolites
NCgl0703	cg0841	-	760742	761043	97	-	1.0	1.0	1.0	1.0	1.5	3.6	2.8	hypothetical protein	Unknown function
NCgl0710	cg0849	*manC (rmlA2)*	779848	780149	87	25	1.0	1.0	1.0	2.3	2.3	4.0	2.8	nucleoside-diphosphate-sugar pyrophosphorylase	Cell wall/membrane/envelope biogenesis
NCgl0718	cg0859	-	787952	788253	93	27	1.0	1.0	1.0	3.3	4.7	12.3	5.0	hypothetical protein	Unknown function
NCgl0722	cg0864	*mtrB*	791301	791602	60	-	1.0	1.0	1.0	1.0	1.0	3.0	2.4	two-component system sensory transduction histidine kinase	Signal transduction mechanisms; Cell wall/membrane/envelope biogenesis
NCgl0729	cg0872	-	800937	801238	1	-	5.0	1.0	1.0	1.0	1.0	1.0	1.0	hypothetical protein	General function prediction only
NCgl0760	cg0905	*psp2*	836558	836859	603	25	3.5	1.0	1.0	4.2	7.7	11.6	7.2	hypothetical protein	Unknown function
NCgl0761	cg0906	-	838295	838596	479	180	42.3	1.0	1.0	4.3	7.1	5.8	2.7	hypothetical protein	Unknown function
NCgl0763	cg0908	-	839458	839759	21	58	1.0	1.7	1.0	8.5	3.5	2.9	3.6	hypothetical protein	Unknown function
NCgl0764	cg0909	-	840893	841194	299	-	1.0	1.0	1.0	2.6	6.4	15.1	7.0	hypothetical protein	Unknown function
NCgl0770	cg0916	*smpB*	845972	846273	14	14	1.0	1.0	2.2	1.0	2.7	6.6	3.6	SsrA-binding protein	Protein turnover and chaperones
NCgl0777	cg0926	-	852372	852673	95	63	1.0	1.0	2.1	1.0	2.7	5.4	2.1	ABC-type cobalamin/Fe3+-siderophore transport system%2C permease	Inorganic ion transport, metabolism, and storage; Transport and metabolism of further metabolites
NCgl0780	cg0931	-	860861	861162	580	-	1.0	1.0	1.0	2.8	21.3	1.0	1.0	aminotransferase	General function prediction only
NCgl0780	cg0931	-	861021	861322	420	-	4.2	1.0	1.0	2.8	1.0	2.5	3.2	aminotransferase	General function prediction only
NCgl0786	cg0938	*cspB*	868759	869060	28	38	1.0	1.0	1.0	1.0	1.8	3.5	2.4	cold shock protein	Transcription including sigma factors, RNA processing and modification
NCgl0795	cg0949	*gltA*	877654	877955	33	161	1.0	1.0	1.0	1.0	1.0	2.3	1.0	type II citrate synthase	Central carbon metabolism
NCgl0798	cg0952	*mctB*	881254	881555	302	-	1.0	2.3	1.0	1.0	4.8	7.5	5.8	hypothetical protein	Unknown function
NCgl0803	cg0958	-	894521	894822	31	6	1.0	1.0	1.0	1.0	2.1	2.3	2.1	hypothetical protein	Unknown function
NCgl0805	cg0961	-	896603	896904	112	-	4.4	1.0	1.0	1.0	1.0	1.7	1.0	hypothetical protein	Amino acid transport and metabolism
NCgl0825	cg0982	-	917775	918076	110	1	1.0	1.0	1.8	1.0	1.7	3.5	1.9	hypothetical protein	Unknown function
NCgl0841	cg0998	*pepD*	929716	930017	459	-	1.0	1.0	1.0	1.0	1.7	2.6	2.0	trypsin-like serine protease	Protein turnover and chaperones
NCgl0843	cg1001	*mscL*	932862	933163	39	-	2.2	1.0	1.0	2.2	1.0	4.4	1.0	large-conductance mechanosensitive channel	Carbon source transport and metabolism; Inorganic ion transport, metabolism, and storage
NCgl0859	cg1019	-	951412	951713	103	-	1.0	1.0	1.0	2.3	3.5	6.8	2.8	hypothetical protein	General function prediction only
NCgl0863	cg1031	*tnp6d*	953904	954205	222	-	1.0	4.6	1.2	1.0	2.0	1.0	1.3	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl0864	cg1024	*tnp7a*	954288	954589	502	326	1.0	23.0	1.0	1.0	1.0	1.0	10.8	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl0864	cg1024	*tnp7a*	954591	954892	199	23	3.5	1.0	1.0	3.1	1.6	2.9	1.0	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl0866	cg1028	-	959071	959372	224	-	6.9	6.3	3.9	13.3	19.2	28.1	20.2	adenine-specific DNA methylase	DNA replication, recombination, repair, and degradation
NCgl0867	cg1030	*tnp6c*	959528	959829	402	41	1.0	4.3	1.0	1.0	1.0	19.0	6.4	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl0867	cg1030	*tnp6c*	959776	960077	154	119	1.0	2.8	1.0	1.7	1.0	15.1	1.0	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl0868	-	-	960095	960396	217	-	1.0	1.0	3.0	1.0	17.7	11.5	1.0	hypothetical protein	Unknown function
NCgl0868	-	-	960293	960594	19	-	1.0	46.1	1.0	1.0	1.0	11.5	1.0	hypothetical protein	Unknown function
NCgl0869	cg1032	*cadR*	960522	960823	624	200	10.6	1.0	1.7	4.3	1.0	11.5	1.0	ArsR family transcriptional regulator	Signal transduction mechanisms
NCgl0869	cg1032	*cadR*	960703	961004	443	19	1.0	1.0	1.0	1.0	1.0	1.0	1.8	ArsR family transcriptional regulator	Signal transduction mechanisms
NCgl0880	cg1046	*ppk2A (ppk2)*	970669	970970	44	22	1.0	1.0	1.0	1.0	2.8	9.1	4.7	hypothetical protein	Inorganic ion transport, metabolism, and storage
NCgl0890	cg1057	-	982559	982860	36	-	1.0	1.0	1.0	2.2	1.0	1.0	1.0	hypothetical protein	Unknown function
NCgl0916	cg1090	*ggtB*	1014216	1014517	600	-	1.0	2.9	1.0	4.2	10.9	15.2	11.8	gamma-glutamyltranspeptidase	Transport and metabolism of further metabolites
NCgl0920	cg1095	-	1015422	1015723	79	719	12.1	4.0	1.0	1.0	8.5	1.0	1.0	hypothetical protein	Unknown function
NCgl0921	cg1096	-	1016181	1016482	225	162	5.1	3.1	1.0	2.7	9.2	12.3	7.2	hypothetical protein	Unknown function
NCgl0955	cg1134	*pabAB (pab)*	1051814	1052115	35	1225	1.0	1.0	1.0	1.0	1.0	3.4	2.7	anthranilate/para-aminobenzoate synthase I	Coenzyme transport and metabolism
NCgl0967	cg1145	*fumC (fum)*	1063494	1063795	28	-	1.9	1.0	2.9	1.0	1.0	1.5	1.0	fumarate hydratase	Central carbon metabolism
NCgl0983	cg1165	-	1079053	1079354	124	-	1.0	1.0	1.0	1.0	1.0	2.4	2.6	hypothetical protein	Carbon source transport and metabolism; Transport and metabolism of further metabolites
NCgl0987	cg1170	*cmt5*	1084075	1084376	43	-	1.0	1.0	1.0	1.0	1.0	3.7	2.5	hypothetical protein	Cell wall/membrane/envelope biogenesis
NCgl0995	cg1181	-	1092575	1092876	501	92	1.0	1.0	14.9	1.0	1.0	1.0	1.0	hypothetical protein	General function prediction only
NCgl1006	cg1195	-	1099754	1100055	12	12	1.0	3.1	1.0	1.0	2.1	4.1	2.8	sulfate permease	General function prediction only
NCgl1009	cg1199	-	1102578	1102879	463	677	1.0	1.0	1.0	1.0	1.7	3.4	3.2	hypothetical protein	Unknown function
NCgl1010	cg1201	-	1103224	1103525	576	-	1.0	1.0	1.0	1.0	1.0	2.9	1.0	hypothetical protein	Unknown function

NCgl1013	cg1204	-	1106772	1107073	580	517	4.0	1.0	1.0	1.0	1.0	1.0	1.0	phosphoglycerate mutase	General function prediction only
NCgl1020	cg1212	-	1112945	1113246	6	39	1.0	1.0	1.0	1.0	2.3	3.6	1.0	major facilitator superfamily permease	General function prediction only
NCgl1042	cg1237	-	1136752	1137053	70	-	1.0	1.0	1.0	1.0	5.8	10.7	3.9	hypothetical protein	Unknown function
NCgl1053	cg1248	-	1143793	1144094	174	96	5.2	3.4	2.0	2.1	2.1	4.2	2.9	stress response membrane GTPase	General function prediction only
NCgl1059	cg1254	-	1150243	1150544	47	12	1.0	1.0	1.0	1.0	1.7	2.8	2.8	hypothetical protein	Unknown function
NCgl1096	cg1292	-	1191846	1192147	401	799	1.0	1.0	1.0	1.0	2.3	3.0	2.0	K+ transport flavoprotein	General function prediction only
NCgl1103	cg1300	*cydB*	1201900	1202201	37	-	1.0	1.0	1.0	1.0	1.0	2.8	3.3	cytochrome bd-type quinol oxidase%2C subunit 2	Respiration and oxidative phosphorylation
NCgl1104	cg1301	*cydA*	1203619	1203920	138	-	1.0	1.0	1.0	2.2	2.3	5.9	3.2	cytochrome bd-type quinol oxidase%2C subunit 1	Respiration and oxidative phosphorylation
NCgl1106	cg1303	-	1207387	1207688	315	-	1.0	1.0	1.0	1.0	1.0	3.4	1.0	mutT-like protein	Nucleotide transport and metabolism
NCgl1116	cg1314	*putP*	1219466	1219767	15	-	1.0	1.0	1.0	1.0	1.7	1.9	1.9	Na+/proline%2C Na+/panthothenate symporter	Amino acid transport and metabolism
NCgl1119	cg1318	-	1223753	1224054	17	-	1.0	1.0	1.9	1.0	1.0	1.7	1.0	DNA repair exonuclease	DNA replication, recombination, repair, and degradation
NCgl1127	cg1327	-	1231426	1231727	234	72	1.0	1.0	1.0	3.3	4.5	11.5	3.2	catabolite gene activator	Signal transduction mechanisms
NCgl1132	cg1333	*argS*	1238067	1238368	56	499	1.0	1.0	1.0	1.0	1.0	2.7	1.0	arginyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
NCgl1135	cg1336	-	1242142	1242443	18	-	1.0	1.0	1.0	2.8	2.8	5.5	3.6	hypothetical protein	Unknown function
NCgl1138	cg1340	*arnR*	1245525	1245826	144	-	4.9	1.0	1.0	3.7	9.3	9.2	10.6	hypothetical protein	Signal transduction mechanisms
NCgl1144	cg1346	*mog*	1253942	1254243	56	121	1.0	1.0	1.0	1.0	1.8	3.5	2.1	molybdopterin biosynthesis enzyme	Coenzyme transport and metabolism
NCgl1145	cg1347	-	1256662	1256963	139	-	7.9	4.5	1.0	6.0	16.3	24.9	4.3	serine protease	Cell wall/membrane/envelope biogenesis
NCgl1149	cg1351	*moeA3*	1261083	1261384	33	-	3.5	1.0	1.0	1.6	3.8	3.6	2.1	molybdopterin biosynthesis enzyme	Coenzyme transport and metabolism
NCgl1150	cg1352	*moaA*	1262753	1263054	86	-	1.0	1.0	1.0	1.0	1.0	2.7	2.8	molybdenum cofactor biosynthesis protein A	Coenzyme transport and metabolism
NCgl1157	cg1360	-	1271065	1271366	36	36	1.0	1.0	1.0	1.0	1.9	2.6	1.0	hypothetical protein	Unknown function
NCgl1172	cg1375	-	1281993	1282294	50	10	2.8	1.0	1.0	5.2	10.1	16.6	6.9	thioredoxin domain-containing protein	General function prediction only
NCgl1180	cg1384	-	1292888	1293189	189	75	1.0	1.0	1.0	1.0	2.1	5.6	3.3	hypothetical protein	General function prediction only
NCgl1182	cg1386	*etfB (fixA)*	1294764	1295065	520	185	1.0	1.0	3.7	1.0	1.0	1.0	1.0	electron transfer flavoprotein beta-subunit	Respiration and oxidative phosphorylation
NCgl1185	-	-	1298527	1298828	25	-	2.4	1.0	1.0	2.1	3.5	7.8	2.7	hypothetical protein	Unknown function
NCgl1218	cg1429	-	1331785	1332086	17	17	1.0	1.0	1.0	1.0	1.7	2.2	2.2	hypothetical protein	Unknown function
NCgl1223	cg1436	*ilvN*	1339790	1340091	84	215	5.6	1.0	1.0	1.0	1.0	1.0	1.0	acetolactate synthase small subunit	Amino acid transport and metabolism
NCgl1232	cg1447	*zrf (czcD)*	1344834	1345135	501	501	7.2	1.0	2.8	8.0	12.3	16.0	14.6	Co/Zn/Cd efflux system component	Inorganic ion transport, metabolism, and storage
NCgl1239	cg1456	-	1355491	1355792	47	35	1.0	1.0	2.3	6.9	12.6	23.2	3.0	signal-transduction protein	Signal transduction mechanisms
NCgl1250	cg1469	-	1365684	1365985	439	-	1.0	3.7	1.0	1.0	2.7	3.3	2.2	hypothetical protein	Unknown function
NCgl1252	cg1475	-	1369086	1369387	314	56	1.0	4.0	2.7	4.6	15.5	27.2	4.4	hypothetical protein	Unknown function
NCgl1254	cg1478	-	1372098	1372399	556	52	1.0	1.0	1.0	1.0	4.7	12.9	4.7	hypothetical protein	Unknown function
NCgl1255	cg1479	*malP (glgP1)*	1375640	1375941	57	-	1.0	1.0	1.0	1.0	3.0	4.9	3.3	glucan phosphorylase	Carbon source transport and metabolism
NCgl1259	cg1484	-	1378807	1379108	16	-	1.0	1.0	1.0	2.2	3.2	5.4	1.8	hypothetical protein	Unknown function
NCgl1278	cg1504	-	1397573	1397874	162	-	1.0	1.0	1.0	1.0	1.0	2.7	1.0	ABC transporter periplasmic component	Amino acid transport and metabolism
NCgl1280	cg1506	-	1400076	1400377	693	-	1.0	1.0	1.0	1.0	3.0	5.8	1.0	hypothetical protein	Unknown function
NCgl1283	cg1509	-	1400990	1401291	192	192	1.0	1.0	1.0	17.0	7.1	12.5	5.4	hypothetical protein	Prophage genes
NCgl1283	cg1509	-	1401129	1401430	53	53	1.0	1.0	3.0	17.0	7.1	12.5	1.0	hypothetical protein	Prophage genes
NCgl1281	cg1507	*int1*	1401271	1401572	482	-	1.0	1.0	3.0	17.0	7.1	1.0	1.0	integrase	Prophage genes, DNA replication, recombination and repair
NCgl1284	cg1510	-	1401787	1402088	334	168	3.0	1.0	1.0	1.4	1.4	4.0	2.9	hypothetical protein	Prophage genes
NCgl1286	cg1512	-	1403276	1403577	570	24	1.0	1.0	1.0	1.0	2.0	7.1	4.3	hypothetical protein	Prophage genes
NCgl1286	cg1512	-	1403792	1404093	54	16	1.0	1.0	1.0	1.0	1.0	6.7	1.0	hypothetical protein	Prophage genes
NCgl1290	cg1516	-	1407399	1407700	15	-	6.3	1.0	2.5	14.2	23.0	15.2	16.1	hypothetical protein	Prophage genes
NCgl1291	cg1517	-	1407887	1408188	165	-	16.2	7.9	5.2	18.4	43.9	48.3	23.1	hypothetical protein	Prophage genes
NCgl1293	cg1519	-	1410288	1410589	637	-	1.0	1.0	1.0	1.8	1.0	4.0	3.5	hypothetical protein	Prophage genes
NCgl1298	cg1524	-	1413131	1413432	366	-	5.3	1.0	1.0	2.1	6.8	12.8	7.0	hypothetical protein	Prophage genes
NCgl1307	cg1540	-	1426116	1426417	10	-	1.0	1.0	1.0	1.0	2.2	3.2	1.8	hypothetical protein	Unknown function
NCgl1311	cg1546	*rbsK1*	1431801	1432102	373	-	1.0	1.0	1.0	1.0	1.0	2.5	1.0	sugar kinase	Nucleotide transport and metabolism
NCgl1317	cg1552	*qorR*	1437139	1437440	41	-	3.2	1.0	1.0	4.4	10.3	24.4	10.8	transcriptional regulator	Signal transduction mechanisms
NCgl1337	cg1577	-	1460794	1461095	212	90	1.0	1.0	1.0	2.2	2.0	4.9	2.3	hypothetical protein	General function prediction only
Cgl1354a	-	-	1479094	1479395	356	-	3.2	15.4	2.4	3.7	5.3	5.0	2.9	hypothetical protein	Unknown function
NCgl1354	cg1597	-	1484034	1484335	541	455	1.0	1.0	1.0	3.4	7.1	1.0	1.0	TPR repeat-containing protein	Unknown function
NCgl1354	cg1597	-	1484283	1484584	292	206	1.8	2.7	1.0	2.2	1.0	7.3	1.0	TPR repeat-containing protein	Unknown function
NCgl1365	cg1609	-	1498175	1498476	2	-	7.3	6.1	3.2	9.1	14.3	26.9	15.4	ABC transporter duplicated ATPase	Transport and metabolism of further metabolites
NCgl1384	cg1629	*secA2*	1516974	1517275	45	13	3.4	1.0	1.3	1.0	1.0	2.4	1.0	accessory Sec system translocase SecA2	Protein secretion
NCgl1388	cg1633	-	1521553	1521854	67	5	3.4	2.9	1.0	2.9	3.1	4.0	3.5	transcriptional regulator	Signal transduction mechanisms
NCgl1410	cg1657	*ufaA*	1544737	1545038	88	6	1.0	1.0	1.0	1.0	1.0	2.9	2.2	cyclopropane fatty acid synthase	Cell wall/membrane/envelope biogenesis
NCgl1423	cg1672	*ppmC (ppm1)*	1557786	1558087	114	-	1.0	1.0	1.0	1.0	1.0	1.8	1.0	glycosyltransferase	Cell wall/membrane/envelope biogenesis
NCgl1438	cg1689	*pup*	1575341	1575642	356	-	1.0	1.0	1.0	2.3	1.0	1.0	1.0	hypothetical protein	Post-translational modification
NCgl1440	cg1691	*arc (mpa)*	1578457	1578758	80	-	1.0	1.4	1.0	2.5	1.0	1.0	1.0	ATPase of the AAA+ class	Post-translational modification; Protein turnover and chaperones
NCgl1444	cg1695	-	1581654	1581955	46	39	1.0	1.0	1.0	1.0	1.9	2.4	1.0	plasmid maintenance system antidote protein	General function prediction only
NCgl1445	cg1696	-	1583407	1583708	77	-	1.0	1.0	1.0	1.0	2.7	3.5	2.9	major facilitator superfamily permease	General function prediction only
NCgl1446	cg1697	*aspA*	1585818	1586119	479	-	5.6	1.0	1.0	1.0	1.0	1.0	1.0	aspartate ammonia-lyase	Amino acid transport and metabolism
NCgl1453	cg1705	*arsB1 (arsC2)*	1592496	1592797	690	55	1.0	1.0	1.0	2.2	4.6	8.2	2.5	arsenite efflux pump ACR3	Inorganic ion transport, metabolism, and storage
NCgl1453	cg1705	*arsB1 (arsC2)*	1593187	1593488	1	636	1.0	1.0	1.0	3.3	8.7	10.0	9.3	arsenite efflux pump ACR3	Inorganic ion transport, metabolism, and storage
NCgl1464	cg1716	*tnp16b*	1603175	1603476	146	221	1.0	1.0	1.2	10.8	0.8	1.0	2.2	transposase	DNA replication, recombination, repair, and degradation
NCgl1464	cg1716	*tnp16b*	1603341	1603642	20	387	1.0	1.0	1.0	1.0	0.8	11.5	2.2	transposase	DNA replication, recombination, repair, and degradation

NCgl1469	cg1722	*act3*	1609664	1609965	421	38	1.0	2.8	1.0	1.0	2.0	3.3	2.6	histone acetyltransferase HPA2 and related acetyltransferase	Amino acid transport and metabolism
NCgl1473	cg1727	-	1616337	1616638	39	-	1.0	1.0	1.0	1.0	1.0	3.0	2.4	hypothetical protein	Unknown function
NCgl1474	cg1728	-	1616949	1617250	298	228	4.6	1.0	1.0	2.6	1.0	1.0	1.0	hypothetical protein	Unknown function
NCgl1483	cg1738	*acnR*	1629118	1629419	29	-	1.0	1.0	1.0	1.0	1.0	1.0	2.6	transcriptional regulator	Signal transduction mechanisms
NCgl1486	cg1742	-	1631112	1631413	397	342	1.0	1.0	1.0	1.0	1.0	2.2	1.0	hypothetical protein	Unknown function
NCgl1490	cg1746	-	1636920	1637221	10	-	1.0	3.4	1.0	3.3	4.8	10.5	6.5	hypothetical protein	Prophage genes
NCgl1491	cg1750	-	1639212	1639513	553	-	1.0	1.0	2.5	1.7	2.7	3.5	1.0	hypothetical protein	Prophage genes
NCgl1497	cg1758	-	1645890	1646191	508	508	1.0	1.0	1.0	25.5	7.1	1.0	1.0	hypothetical protein	Unknown function
NCgl1497	cg1758	-	1646123	1646424	275	275	5.3	1.0	1.0	1.0	1.0	2.2	1.0	hypothetical protein	Unknown function
NCgl1511	cg1773	*ctaB*	1663561	1663862	114	-	4.4	1.0	1.0	1.0	3.4	6.5	3.8	protoheme IX farnesyltransferase	Protein turnover and chaperones
NCgl1518	-	-	1671598	1671899	1	1342	4.9	1.0	1.0	1.0	1.0	2.9	1.0	hypothetical protein	Unknown function
NCgl1517	cg1781	*soxA*	1671937	1672238	411	-	1.0	1.0	7.5	1.0	2.4	1.0	1.0	hypothetical protein	Carbon source transport and metabolism
NCgl1543	cg1809	-	1702394	1702695	229	-	1.0	1.0	3.3	1.0	1.0	1.0	1.0	DNA-directed RNA polymerase subunit omega	Transcription including sigma factors, RNA processing and modification
NCgl1576	cg1844	-	1740521	1740822	103	-	1.0	2.3	1.0	1.0	1.0	1.0	2.3	hypothetical protein	Unknown function
NCgl1583	cg1852	*sdaA*	1744091	1744392	642	601	1.0	1.0	1.0	1.0	1.0	1.0	2.7	L-serine deaminase	Amino acid transport and metabolism
NCgl1589	cg1860	-	1752056	1752357	48	15	1.0	1.0	2.4	3.3	4.3	6.6	3.1	hypothetical protein	Unknown function
NCgl1594	cg1867	*secD*	1760966	1761267	383	-	3.2	1.0	1.0	1.0	1.0	2.8	1.0	preprotein translocase subunit SecD	Protein secretion
NCgl1596	cg1869	*ruvB*	1762812	1763113	456	-	1.0	1.0	1.0	1.0	2.1	4.5	1.8	Holliday junction DNA helicase RuvB	DNA replication, recombination, repair, and degradation
NCgl1600	cg1873	*tesB2*	1765723	1766024	14	-	2.4	1.0	1.0	1.0	1.8	3.1	2.3	acyl-CoA thioesterase II	Carbon source transport and metabolism
NCgl1612	cg1891	*olpA*	1780085	1780386	185	-	151.6	123.6	110.6	248.7	160.5	104.8	189.4	hypothetical protein	Prophage genes
NCgl1616	cg1895	-	1785402	1785703	80	-	90.3	66.9	43.6	108.9	157.4	81.7	156.9	hypothetical protein	Prophage genes
NCgl1617	cg1896	-	1786722	1787023	29	-	11.6	8.1	5.9	13.0	13.9	16.3	12.8	hypothetical protein	Prophage genes
NCgl1619	cg1898	-	1788493	1788794	436	283	1.0	1.0	1.0	31.3	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1618	cg1897	-	1788709	1789010	31	-	57.1	46.1	30.7	56.5	82.9	53.4	95.6	hypothetical protein	Prophage genes
NCgl1619	cg1898	-	1788921	1789222	8	109	1.0	1.0	33.4	29.5	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1623	cg1903	-	1791403	1791704	288	9	37.2	29.5	21.4	66.7	58.5	42.6	52.1	ABC transporter ATPase	Prophage genes
NCgl1624	cg1904	-	1791731	1792032	546	8	36.4	24.8	13.6	49.0	73.0	58.8	50.2	ABC transporter permease	Prophage genes
NCgl1625	cg1905	-	1793056	1793357	507	202	1.0	4.8	3.8	8.3	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1625	cg1905	-	1793332	1793633	231	84	5.8	4.7	1.0	9.5	15.4	18.4	11.9	hypothetical protein	Prophage genes
NCgl1625	cg1905	-	1793521	1793822	42	8	1.0	4.7	1.0	9.5	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1626	cg1907	-	1794925	1795226	515	212	1.0	5.5	3.9	5.8	1.0	24.2	1.0	phosphopantothenoylcysteine synthetase/decarboxylase	Prophage genes
NCgl1626	cg1907	-	1795113	1795414	327	24	1.0	1.0	1.0	5.8	15.7	24.2	13.0	phosphopantothenoylcysteine synthetase/decarboxylase	Prophage genes
NCgl1627	cg1908	-	1796001	1796302	58	118	4.4	3.6	1.0	1.0	1.0	1.0	11.9	hypothetical protein	Prophage genes
NCgl1628	-	-	1796812	1797113	387	15	37.1	24.4	20.6	61.2	86.5	95.5	42.9	hypothetical protein	Prophage genes
NCgl1629	cg1911	-	1798737	1799038	131	-	22.0	21.3	14.3	39.6	79.2	69.4	29.8	hypothetical protein	Prophage genes
NCgl1635	cg1918	-	1802592	1802893	146	781	106.5	112.1	87.2	208.2	163.1	108.8	112.6	hypothetical protein	Prophage genes
NCgl1640	cg1923	-	1804976	1805277	648	-	1.0	1.0	3.8	1.0	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1645	cg1929	*res*	1809617	1809918	7	49	104.8	69.4	71.6	146.5	85.4	98.9	65.4	site-specific recombinase	Prophage genes
NCgl1647	cg1931	-	1810939	1811240	474	117	3.4	1.0	1.0	4.1	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1647	cg1931	-	1811437	1811738	24	66	1.0	1.0	1.0	1.0	2.3	2.7	3.8	hypothetical protein	Prophage genes
NCgl1650	cg1935	*gntR2*	1813666	1813967	37	81	5.4	1.0	1.0	1.0	1.0	1.0	1.0	transcriptional regulator	Prophage genes
NCgl1651	cg1936	-	1814531	1814832	205	157	144.3	112.5	103.2	248.0	226.5	110.3	183.9	hypothetical protein	Prophage genes
NCgl1653	cg1938	-	1816092	1816393	208	473	79.4	57.0	47.8	110.4	83.8	54.1	54.5	hypothetical protein	Prophage genes
NCgl1654	cg1940	-	1816852	1817153	129	48	151.3	94.7	90.9	291.1	177.4	153.8	148.3	hypothetical protein	Prophage genes
NCgl1657	-	-	1818674	1818975	57	22	7.9	3.7	7.1	7.2	12.6	20.5	14.6	hypothetical protein	Prophage genes
NCgl1659	cg1946	-	1822430	1822731	47	47	4.1	1.0	1.0	2.5	1.9	4.0	5.8	hypothetical protein	Prophage genes
NCgl1660	cg1947	-	1823682	1823983	538	409	5.4	1.0	1.0	7.1	6.6	9.5	8.9	hypothetical protein	Prophage genes
NCgl1663	-	-	1825246	1825547	627	62	1.0	1.0	1.0	1.0	4.8	4.5	8.9	hypothetical protein	Prophage genes
NCgl1666	cg1955	-	1830364	1830665	250	96	1.0	1.0	1.0	41.9	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1666	cg1955	-	1830577	1830878	37	117	61.4	37.7	34.3	68.3	73.3	54.0	73.7	hypothetical protein	Prophage genes
NCgl1668	cg1957	-	1834799	1835100	22	-	90.7	65.8	71.4	128.9	111.2	69.9	104.7	hypothetical protein	Prophage genes
NCgl1668	cg1957	-	1835199	1835500	422	-	89.8	67.0	63.5	172.5	92.7	64.5	104.0	hypothetical protein	Prophage genes
NCgl1669	cg1959	*priP*	1835915	1836216	486	290	1.0	1.0	1.0	53.0	1.0	1.0	1.0	ATPase	Prophage genes
NCgl1669	cg1959	*priP*	1836190	1836491	211	15	68.4	46.4	40.9	95.6	71.4	55.2	98.7	ATPase	Prophage genes
NCgl1672	cg1962	-	1842568	1842869	202	141	4.0	1.0	1.0	1.0	3.8	6.3	10.2	hypothetical protein	Prophage genes
NCgl1676	cg1966	*cgpS*	1846697	1846998	168	-	257.1	140.3	142.6	359.1	176.9	142.4	173.9	hypothetical protein	Prophage genes
NCgl1677	cg1967	-	1847147	1847448	29	1	87.0	1.0	53.5	1.0	101.0	77.2	106.0	hypothetical protein	Prophage genes
NCgl1682	cg1974	-	1850951	1851252	53	-	73.3	50.9	42.3	99.2	94.7	72.6	97.8	hypothetical protein	Prophage genes
NCgl1685	cg1977	-	1853850	1854151	281	185	151.1	106.0	84.4	212.8	189.0	119.9	164.8	hypothetical protein	Prophage genes
NCgl1685	cg1977	-	1853990	1854291	141	45	151.1	106.0	84.4	212.8	189.0	119.9	164.8	hypothetical protein	Prophage genes
NCgl1687	cg1980	-	1854958	1855259	423	43	1.0	1.0	1.0	1.0	4.5	5.6	8.7	hypothetical protein	Prophage genes
NCgl1699	cg1992	-	1868722	1869023	22	-	1.0	1.0	2.2	1.0	1.0	2.7	4.0	hypothetical protein	Prophage genes
NCgl1702	cg1995	-	1877791	1878092	56	-	110.4	62.3	59.3	150.4	118.2	88.0	118.1	hypothetical protein	Prophage genes
NCgl1706	cg1999	-	1883905	1884206	66	-	79.6	61.8	46.8	94.9	85.2	69.8	124.3	hypothetical protein	Prophage genes
NCgl1707	cg2000	-	1884780	1885081	43	-	29.3	19.0	12.1	40.7	62.5	62.4	48.2	hypothetical protein	Prophage genes
NCgl1708	cg2001	-	1884966	1885267	113	113	1.0	16.2	7.2	1.0	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1709	cg2002	-	1886613	1886914	641	332	1.0	6.1	1.0	1.0	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1709	cg2002	-	1887076	1887377	178	131	9.7	1.0	5.3	16.8	17.7	18.8	22.8	hypothetical protein	Prophage genes
NCgl1710	cg2003	-	1888057	1888358	206	-	3.9	1.0	1.0	1.0	2.6	1.0	7.2	helicase	Prophage genes
NCgl1710	cg2003	-	1888353	1888654	502	-	1.0	1.0	2.0	3.7	3.5	4.6	8.2	helicase	Prophage genes
NCgl1711	cg2004	-	1889410	1889711	467	-	57.7	52.1	40.0	115.0	143.6	87.0	110.3	hypothetical protein	Prophage genes
NCgl1712	cg2005	-	1891688	1891989	222	-	16.9	7.9	9.4	22.1	28.0	27.7	27.6	hypothetical protein	Prophage genes
NCgl1713	cg2006	-	1892855	1893156	31	-	61.7	25.5	21.3	72.1	94.2	74.6	64.1	hypothetical protein	Prophage genes
NCgl1714	cg2007	-	1894475	1894776	6	-	1.0	3.2	1.0	1.0	1.9	2.4	4.9	hypothetical protein	Prophage genes
NCgl1715	cg2008	-	1897181	1897482	101	-	213.9	90.4	63.2	128.6	110.6	76.2	125.6	hypothetical protein	Prophage genes
NCgl1716	cg2009	-	1899034	1899335	27	-	107.0	54.7	65.1	123.5	114.6	74.7	122.5	ATPase with chaperone activity%2C ATP-binding subunit	Prophage genes
NCgl1718	cg2011	-	1900811	1901112	187	-	81.8	68.6	51.6	117.7	81.1	74.8	101.2	hypothetical protein	Prophage genes
NCgl1719	cg2012	-	1901794	1902095	34	-	1.0	1.0	1.0	1.0	1.0	1.0	3.9	hypothetical protein	Prophage genes
NCgl1720	cg2014	-	1902788	1903089	56	-	101.5	96.2	52.0	118.6	139.1	86.0	120.3	hypothetical protein	Prophage genes
NCgl1721	cg2015	-	1905706	1906007	21	-	1.0	1.0	1.0	2.2	4.3	5.9	9.3	hypothetical protein	Prophage genes
NCgl1724	cg2018	-	1908564	1908865	55	-	85.3	56.8	47.4	76.8	87.2	71.4	102.1	hypothetical protein	Prophage genes
NCgl1726	cg2020	-	1910364	1910665	7	-	79.7	44.1	61.7	101.9	72.2	56.9	106.6	hypothetical protein	Prophage genes
NCgl1727	cg2021	-	1912290	1912591	141	-	1.0	1.0	1.0	3.5	2.5	3.3	5.5	hypothetical protein	Prophage genes
NCgl1728	cg2022	-	1913700	1914001	118	-	142.7	66.3	73.1	162.9	154.5	110.9	139.5	hypothetical protein	Prophage genes

NCgl1729	-	-	1914260	1914561	82	-	133.4	121.4	64.8	229.8	228.2	135.4	201.5	hypothetical protein	Prophage genes
NCgl1730	cg2024	-	1916125	1916426	43	-	5.1	2.9	1.0	3.5	3.0	5.2	13.6	hypothetical protein	Prophage genes
NCgl1732	-	-	1916701	1917002	92	50	1.0	1.0	1.0	3.3	6.6	13.7	18.5	hypothetical protein	Prophage genes
NCgl1735	cg2030	-	1920103	1920404	93	-	1.0	1.0	1.0	1.0	20.6	29.1	24.6	hypothetical protein	Prophage genes
NCgl1735	cg2030	-	1920342	1920643	332	-	1.0	1.0	2.8	6.6	1.0	21.9	1.0	hypothetical protein	Prophage genes
NCgl1737	cg2032	-	1925955	1926256	96	-	97.3	64.4	69.3	128.2	105.8	102.0	113.0	hypothetical protein	Prophage genes
NCgl1738	cg2033	-	1927265	1927566	579	-	6.9	1.0	1.0	1.0	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1740	cg2035	-	1927656	1927957	404	141	7.1	6.6	3.4	13.3	24.3	26.5	37.2	hypothetical protein	Prophage genes
NCgl1741	-	-	1928266	1928567	117	79	9.5	5.6	3.8	4.8	7.6	9.7	14.5	hypothetical protein	Prophage genes
NCgl1743	-	-	1931456	1931757	417	-	3.2	1.0	2.2	1.0	2.6	4.8	7.0	hypothetical protein	Prophage genes
NCgl1746	cg2041	-	1932686	1932987	60	-	8.8	7.1	6.2	19.5	32.6	50.2	52.9	hypothetical protein	Prophage genes
NCgl1748	cg2043	-	1935930	1936231	169	-	9.5	7.4	1.0	6.9	25.1	22.1	28.7	periplasmic serine protease	Prophage genes
NCgl1750	-	-	1936985	1937286	66	17	1.0	1.0	1.0	5.2	9.0	9.2	16.0	hypothetical protein	Prophage genes
NCgl1751	cg2047	-	1937992	1938293	406	-	1.0	1.0	1.0	2.7	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1755	-	-	1940801	1941102	155	-	1.0	1.0	1.0	4.0	4.7	8.1	11.7	hypothetical protein	Prophage genes
NCgl1756	cg2052	-	1942587	1942888	308	-	8.3	8.4	9.0	23.9	49.7	41.1	45.6	hypothetical protein	Prophage genes
NCgl1758	-	-	1942723	1943024	221	50	8.3	8.4	9.0	23.9	49.7	41.1	45.6	hypothetical protein	Prophage genes
NCgl1760	cg2056	-	1943594	1943895	382	289	1.0	1.0	1.0	1.0	2.8	3.8	9.6	hypothetical protein	Prophage genes
NCgl1762	-	-	1946067	1946368	336	-	1.0	1.4	1.0	1.0	2.8	2.9	1.0	hypothetical protein	Prophage genes
NCgl1763	cg2059	-	1946220	1946521	39	-	7.1	1.4	1.0	1.0	2.8	1.0	10.8	hypothetical protein	Prophage genes
NCgl1764	cg4007	-	1946405	1946706	3	-	1.0	1.0	1.0	1.0	1.0	2.9	10.8	hypothetical protein	Prophage genes
NCgl1765	cg2060	-	1947261	1947562	375	-	1.0	1.0	1.0	1.0	7.1	3.5	1.0	hypothetical protein	Prophage genes
NCgl1765	cg2060	-	1947489	1947790	603	-	1.0	6.6	2.1	2.1	9.5	3.5	1.0	hypothetical protein	Prophage genes
NCgl1766	cg2061	*psp3*	1948641	1948942	142	-	2.7	1.0	1.0	1.6	2.8	8.0	6.6	hypothetical protein	Prophage genes
NCgl1767	cg2062	-	1951356	1951657	57	-	93.6	40.4	36.1	103.5	78.8	59.3	82.0	hypothetical protein	Prophage genes
NCgl1769	cg2064	-	1954663	1954964	8	-	114.1	75.4	76.4	134.8	127.2	95.9	166.0	topoisomerase IA	Prophage genes
NCgl1770	cg2065	-	1958190	1958491	63	-	71.5	50.4	49.0	99.3	84.6	58.3	85.2	hypothetical protein	Prophage genes
NCgl1774	cg2069	*psp1*	1962948	1963249	132	-	1.0	1.0	6.7	14.1	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1775	-	-	1963346	1963647	68	-	70.8	27.8	22.6	43.1	60.0	31.4	44.4	hypothetical protein	Prophage genes
NCgl1779	-	-	1966164	1966465	34	16	83.1	49.5	29.6	67.8	97.7	59.4	70.3	micrococcal nuclease-like protein	Prophage genes
NCgl1780	-	-	1967390	1967691	63	57	116.1	56.1	48.2	107.0	130.7	73.5	102.1	hypothetical protein	Prophage genes
NCgl1784	-	-	1971349	1971650	172	155	1.0	1.0	1.0	55.8	92.1	1.0	78.9	hypothetical protein	Prophage genes
NCgl1784	-	-	1971547	1971848	26	-	103.6	32.5	38.9	82.1	92.1	66.2	78.9	hypothetical protein	Prophage genes
NCgl1804	-	-	1984362	1984663	18	48	5.4	1.0	1.0	1.0	1.0	1.0	1.0	hypothetical protein	Prophage genes
NCgl1808	-	-	1988757	1989058	602	-	1.0	1.0	1.0	1.0	1.8	4.7	3.0	hypothetical protein	Prophage genes
NCgl1811	-	-	1989619	1989920	165	-	6.7	5.1	2.7	15.6	37.4	51.4	25.3	hypothetical protein	Prophage genes
NCgl1825	cg2080	-	2001270	2001571	68	5	1.0	1.0	1.0	1.0	1.0	2.4	1.0	hypothetical protein	Unknown function
NCgl1845	cg2103	*dtxR*	2021870	2022171	245	104	4.1	1.0	1.9	3.3	2.3	3.9	1.0	Mn-dependent transcriptional regulator	Signal transduction mechanisms
NCgl1854	cg2113	*divS*	2036287	2036588	251	-	1.0	4.0	2.2	1.6	1.8	1.0	1.0	hypothetical protein	Signal transduction mechanisms
NCgl1856	cg2115	*sugR*	2037601	2037902	63	2	1.0	1.0	1.0	1.9	2.6	4.2	1.6	transcriptional regulator of sugar metabolism	Signal transduction mechanisms
NCgl1858	cg2117	*ptsI*	2041295	2041596	125	-	1.0	1.0	1.8	1.0	6.2	7.9	7.8	phosphoenolpyruvate-protein kinase	Carbon source transport and metabolism
NCgl1865	cg2126	*hflX*	2050042	2050343	86	-	3.0	1.0	1.0	1.0	1.0	2.3	1.0	GTPase	Translation, ribosomal structure and biogenesis
NCgl1875	cg2136	*gluA*	2059627	2059928	28	-	1.0	1.0	1.0	2.2	1.0	2.6	1.0	glutamate ABC transporter ATPase	Amino acid transport and metabolism
NCgl1882	cg2146	-	2066226	2066527	33	-	3.9	1.0	1.0	2.8	5.0	9.4	5.7	hypothetical protein	Unknown function
NCgl1896	cg2161	*dapA*	2080045	2080346	13	-	1.0	1.0	1.0	1.0	1.9	2.5	2.4	4-hydroxy-tetrahydrodipicolinate synthase	Amino acid transport and metabolism
NCgl1909	cg2175	*rbfA*	2093425	2093726	75	-	1.7	1.0	1.0	1.7	4.0	6.9	6.4	ribosome-binding factor A	Translation, ribosomal structure and biogenesis
NCgl1927	cg2193	-	2115415	2115716	186	186	1.0	1.0	1.0	1.0	1.0	2.9	1.6	hypothetical protein	Cell wall/membrane/envelope biogenesis
NCgl1929	cg2195	-	2118410	2118711	175	59	6.3	3.9	3.1	4.4	11.0	14.6	6.1	hypothetical protein	Unknown function
NCgl1931	cg2197	-	2119415	2119716	149	13	1.0	1.0	2.4	1.0	2.4	4.3	2.1	hypothetical protein	Unknown function
NCgl1948	cg2218	*pyrH*	2138692	2138993	179	-	1.0	1.0	2.2	1.0	1.0	2.9	2.1	uridylate kinase	Nucleotide transport and metabolism
NCgl1984	cg2262	*ftsY*	2175373	2175674	20	-	4.2	2.5	1.0	2.3	3.3	5.8	5.2	signal recognition particle GTPase	Translation, ribosomal structure and biogenesis
NCgl1991	cg2270	-	2188051	2188352	523	-	2.0	1.0	1.0	1.8	4.2	5.6	5.0	hypothetical protein	Unknown function
NCgl1999	cg2280	*gdh*	2196167	2196468	236	-	4.6	2.6	3.7	5.8	10.8	16.6	8.6	glutamate dehydrogenase	Amino acid transport and metabolism
NCgl2022	cg2306	-	2218926	2219227	82	20	3.0	1.0	1.0	1.0	2.5	6.6	5.5	hypothetical protein	Unknown function
NCgl2034	cg2320	-	2232413	2232714	107	-	1.0	1.0	1.0	1.0	1.0	3.2	2.1	transcriptional regulator	Signal transduction mechanisms
NCgl2038	cg2324	-	2237170	2237471	11	8	4.0	1.0	1.0	1.0	1.0	1.0	1.0	hypothetical protein	Unknown function
NCgl2047	cg2336	-	2246179	2246480	78	22	1.0	1.0	1.0	2.2	3.4	8.0	3.3	hypothetical protein	Unknown function
NCgl2049	cg2338	*dnaE1*	2252200	2252501	411	-	1.0	1.0	1.0	1.0	2.6	1.0	5.7	DNA polymerase III subunit alpha	DNA replication, recombination, repair, and degradation
NCgl2051	cg2340	-	2252763	2253064	279	77	1.0	1.0	1.0	1.0	1.0	7.6	1.0	hypothetical protein	Amino acid transport and metabolism
NCgl2051	cg2340	-	2252935	2253236	107	249	1.0	1.0	1.0	1.0	2.7	1.0	3.6	hypothetical protein	Amino acid transport and metabolism
NCgl2053	cg2342	-	2255523	2255824	106	-	1.0	1.0	1.0	1.0	2.1	2.4	2.1	dehydrogenase	General function prediction only
NCgl2070	cg2361	*divIVA*	2275827	2276128	116	-	1.0	1.0	1.0	1.0	1.7	2.1	2.2	cell division initiation protein	Cell division, chromosome partitioning
NCgl2075	cg2366	*ftsZ*	2280156	2280457	91	-	1.0	1.0	1.0	1.0	1.0	2.7	3.4	cell division protein FtsZ	Cell division, chromosome partitioning
NCgl2097	cg2390	-	2306078	2306379	86	38	1.0	1.0	1.0	1.0	2.4	3.1	1.9	hypothetical protein	Unknown function
NCgl2102	cg2395	-	2313983	2314284	41	-	1.0	1.0	2.5	1.9	2.6	5.7	3.3	hypothetical protein	Unknown function
NCgl2116	cg2410	*ltsA*	2328311	2328612	55	32	1.0	1.0	1.0	1.0	2.4	5.5	1.0	asparagine synthase	Amino acid transport and metabolism; Cell wall/membrane/envelope biogenesis
NCgl2118	cg2412	-	2330934	2331235	86	778	1.0	1.0	1.0	1.5	1.0	2.5	1.0	hypothetical protein	Unknown function
NCgl2125	cg2420	-	2339467	2339768	477	-	1.0	1.0	1.0	2.5	1.8	2.9	1.8	hypothetical protein	Unknown function
NCgl2134	cg2430	-	2350442	2350743	28	-	5.2	1.0	1.0	3.6	1.0	1.0	1.0	hypothetical protein	Unknown function
NCgl2166	cg2465	-	2379312	2379613	36	-	1.0	1.0	1.0	1.8	2.0	3.3	2.8	hypothetical protein	Unknown function
NCgl2175	cg2474	*nagD (hdpA)*	2387838	2388139	9	9	1.0	1.0	1.0	1.0	1.0	3.4	1.0	HAD family sugar phosphatase	Carbon source transport and metabolism
NCgl2177	cg2477	-	2390958	2391259	249	-	1.0	1.0	2.2	2.4	3.6	6.6	3.8	hypothetical protein	Unknown function
NCgl2185	cg2485	*phoD*	2395001	2395302	65	38	3.5	1.0	1.0	1.0	1.0	3.1	3.4	phosphodiesterase/alkaline phosphatase D	Post-translational modification; Signal transduction mechanisms

NCgl2191	cg2492	*glmS*	2404524	2404825	662	-	1.0	1.0	2.6	1.8	4.1	6.6	4.0	glucosamine--fructose-6-phosphate aminotransferase	Amino acid transport and metabolism; Cell wall/membrane/envelope biogenesis
NCgl2196	cg2497	-	2410358	2410659	244	-	1.0	1.0	1.0	1.0	2.4	3.1	3.1	hypothetical protein	Unknown function
NCgl2199	cg2500	*znr*	2412312	2412613	118	14	1.0	1.0	1.0	2.0	1.6	3.0	1.0	transcriptional regulator	Signal transduction mechanisms
NCgl2217	cg2523	*malQ*	2430059	2430360	87	77	2.1	1.0	1.0	1.0	2.6	3.1	2.7	4-alpha-glucanotransferase	Carbon source transport and metabolism
NCgl2235	cg2545	-	2455550	2455851	6	-	1.0	1.0	1.0	1.0	3.0	2.9	3.9	hypothetical protein	Unknown function
NCgl2249	cg2561	*thiX*	2472052	2472353	48	11	1.0	1.0	1.0	1.0	2.8	6.4	2.2	thiamine biosynthesis protein x	Coenzyme transport and metabolism
NCgl2250	cg2562	-	2473408	2473709	6	-	1.0	1.0	1.0	2.8	3.8	1.0	5.9	hypothetical protein	General function prediction only
NCgl2250	cg2562	-	2473563	2473864	161	-	1.0	1.0	1.0	1.0	1.0	6.8	1.0	hypothetical protein	General function prediction only
NCgl2254	cg2566	-	2477493	2477794	7	-	1.0	1.0	1.0	1.0	1.0	3.0	2.1	hypothetical protein	Unknown function
NCgl2260	cg2572	-	2484389	2484690	122	28	1.0	1.0	1.9	1.0	4.4	10.7	4.4	hypothetical protein	Unknown function
NCgl2285	cg2601	-	2510255	2510556	641	430	1.0	46.1	1.0	1.0	1.0	1.0	10.8	pirin	General function prediction only
NCgl2285	cg2601	-	2510459	2510760	437	226	1.0	1.0	1.5	1.0	3.0	1.4	10.8	pirin	General function prediction only
NCgl2293	cg2609	*valS*	2518226	2518527	31	-	1.0	3.2	2.0	1.0	4.5	6.0	6.3	valyl-tRNA synthetase	Translation, ribosomal structure and biogenesis
NCgl2316	cg2633	-	2542317	2542618	117	-	2.8	1.0	1.0	1.0	1.0	3.3	3.2	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl2326	cg2643	*benE*	2553881	2554182	5	-	1.0	1.0	1.0	1.0	1.0	3.6	3.0	benzoate membrane transport protein	Carbon source transport and metabolism
NCgl2330	cg2648	-	2558406	2558707	53	-	4.1	2.6	1.0	2.8	4.5	6.7	5.9	hypothetical protein	Signal transduction mechanisms
NCgl2333	cg2651	-	2560431	2560732	534	498	1.0	1.0	1.7	3.5	1.0	2.5	2.5	hypothetical protein	Unknown function
NCgl2338	cg2659	-	2565353	2565654	258	-	1.0	1.0	1.7	1.0	1.0	2.6	1.6	hypothetical protein	Unknown function
NCgl2357	cg2684	-	2587183	2587484	643	187	1.0	1.0	1.0	1.7	3.5	1.0	1.0	hypothetical protein	Unknown function
NCgl2390	-	-	2624274	2624575	654	-	1.0	2.3	1.7	4.3	28.4	2.9	1.0	hypothetical protein	Unknown function
NCgl2393	cg2726	-	2624853	2625154	597	360	1.0	11.5	1.0	1.7	7.1	5.8	1.3	hypothetical protein	Unknown function
NCgl2401	cg2734	*pncA*	2631918	2632219	475	439	1.0	1.0	1.0	3.4	1.0	1.0	1.0	amidase	Coenzyme transport and metabolism
NCgl2406	cg2739	-	2636990	2637291	644	-	1.0	1.0	1.0	3.0	5.9	14.5	8.3	major facilitator superfamily permease	General function prediction only
NCgl2412	cg2748	-	2650284	2650585	7	7	1.0	1.0	1.0	1.8	1.9	3.2	2.4	hypothetical protein	Unknown function
NCgl2417	cg2756	-	2654101	2654402	239	-	1.0	4.1	2.0	1.7	5.0	1.0	2.2	hypothetical protein	Unknown function
NCgl2419	cg2758	-	2656618	2656919	568	-	1.0	1.0	1.0	2.1	2.4	1.0	1.0	hypothetical protein	Unknown function
NCgl2421	cg2760	-	2657564	2657865	81	-	1.0	1.0	1.0	1.0	1.0	2.7	1.0	hypothetical protein	Unknown function
NCgl2434	cg2777	-	2668624	2668925	14	-	1.0	1.0	1.7	1.7	4.5	9.0	5.5	hypothetical protein	Unknown function
NCgl2440	cg2783	*gntR1*	2675250	2675551	91	40	20.2	11.8	13.6	27.9	55.3	51.1	43.4	transcriptional regulator	Signal transduction mechanisms
NCgl2459	cg2806	-	2694786	2695087	10	74	9.5	9.2	8.2	23.4	45.6	54.3	45.7	hypothetical protein	Unknown function
NCgl2460	cg2807	*tnp11a*	2696046	2696347	430	-	1.0	1.0	5.0	17.0	1.0	2.9	1.0	transposase	DNA replication, recombination, repair, and degradation
NCgl2465	cg2812	-	2703741	2704042	553	-	3.9	1.0	1.0	1.0	2.8	1.2	3.6	ABC transporter ATPase	General function prediction only
NCgl2466	cg2822	-	2709072	2709373	656	-	5.3	1.0	1.0	3.4	3.5	2.7	9.4	hypothetical protein	Nucleotide transport and metabolism
NCgl2468	cg2824	-	2711360	2711661	340	340	1.0	1.0	1.0	2.4	4.3	8.4	7.0	hypothetical protein	General function prediction only
NCgl2486	cg2846	*pstS*	2737551	2737852	163	-	3.0	1.0	1.0	1.0	2.6	4.1	2.1	ABC transporter periplasmic component	Inorganic ion transport, metabolism, and storage
NCgl2493	cg2853	-	2743675	2743976	185	102	1.0	1.0	1.0	2.2	3.0	6.3	5.9	hypothetical protein	Unknown function
NCgl2508	cg2874	*purC*	2762657	2762958	132	-	6.1	1.0	1.0	1.0	5.9	8.6	2.4	phosphoribosylaminoimidazole-succinocarboxamide synthase	Nucleotide transport and metabolism
NCgl2508	cg2874	*purC*	2763137	2763438	612	-	1.0	1.0	1.0	3.1	7.8	14.0	4.2	phosphoribosylaminoimidazole-succinocarboxamide synthase	Nucleotide transport and metabolism
NCgl2525	cg2896	-	2784320	2784621	1	-	4.6	4.1	1.0	5.4	16.0	15.1	7.4	hypothetical protein	General function prediction only
NCgl2532	cg2904	-	2789983	2790284	85	19	5.8	4.2	3.1	7.4	11.8	17.1	7.1	hypothetical protein	Unknown function
NCgl2533	cg2905	*thrE*	2790471	2790772	361	301	1.0	1.0	2.5	1.0	1.0	8.9	1.0	hypothetical protein	Amino acid transport and metabolism
NCgl2542	cg2914	*tnp5b*	2799462	2799763	20	-	1.0	1.0	1.0	1.0	1.0	2.7	1.0	transposase	DNA replication, recombination, repair, and degradation
NCgl2543	-	-	2800415	2800716	548	236	1.0	1.0	1.0	17.0	1.0	1.0	1.0	hypothetical protein	Unknown function
NCgl2543	-	-	2800602	2800903	361	49	4.7	1.0	2.8	1.0	4.3	1.0	1.0	hypothetical protein	Unknown function
NCgl2551	cg2923	-	2808217	2808518	3	-	1.0	1.8	1.0	2.0	3.3	7.0	3.6	23S rRNA (guanosine(2251)-2'-O)-methyltransferase RlmB	Translation, ribosomal structure and biogenesis
NCgl2553	cg2925	*ptsS*	2811686	2811987	30	-	1.0	1.0	1.0	1.0	2.3	4.3	3.9	phosphotransferase system IIC component	Carbon source transport and metabolism
NCgl2566	cg2941	-	2827378	2827679	693	-	2.8	2.2	1.0	2.2	3.4	7.1	5.1	threonine efflux protein	Amino acid transport and metabolism
NCgl2583	cg2960	-	2841751	2842052	592	410	4.2	1.0	1.0	1.0	2.7	5.7	4.3	hypothetical protein	Unknown function
NCgl2584	cg2962	-	2843719	2844020	147	-	1.0	1.0	1.0	1.0	1.0	2.4	2.4	hypothetical protein	Cell wall/membrane/envelope biogenesis
NCgl2585	cg2963	*clpC*	2846886	2847187	96	-	1.0	1.0	2.4	1.0	1.0	2.4	1.0	ATPase with chaperone activity%2C ATP-binding subunit	Protein turnover and chaperones
NCgl2590	cg2969	-	2853469	2853770	150	13	1.0	1.0	1.0	2.0	3.7	6.2	4.9	hypothetical protein	Unknown function
NCgl2616	cg3000	-	2879698	2879999	117	75	2.3	1.9	1.0	1.0	1.0	2.6	1.8	rhodanese-related sulfurtransferase	General function prediction only
NCgl2647	cg3038	-	2925108	2925409	111	-	1.0	1.0	2.4	2.6	8.7	19.7	11.2	major facilitator superfamily permease	General function prediction only
NCgl2661	cg3052	-	2942481	2942782	164	-	1.0	1.0	1.7	1.0	1.0	4.6	2.1	hypothetical protein	Unknown function
NCgl2666	cg3059	*tnp8a*	2947842	2948143	106	-	1.0	1.0	1.0	1.9	6.2	14.5	6.7	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl2666	cg3059	*tnp8a*	2948033	2948334	297	-	1.0	1.0	1.0	1.0	6.7	1.0	1.0	hypothetical protein	DNA replication, recombination, repair, and degradation
NCgl2667	cg3060	*cgtS6*	2949132	2949433	94	-	24.5	12.9	10.8	30.5	1.0	1.0	55.4	two-component system%2C sensory transduction histidine kinase	Post-translational modification; Signal transduction mechanisms
NCgl2668	cg3061	*cgtR6*	2949769	2950070	37	-	1.0	1.0	2.6	1.0	3.7	8.0	1.0	two-component system%2C response regulator	Signal transduction mechanisms
NCgl2682	cg3079	*clpB*	2966145	2966446	134	-	1.0	1.0	1.0	1.0	1.7	2.7	2.8	ATPase with chaperone activity%2C ATP-binding subunit	Protein turnover and chaperones
NCgl2693	cg3091	-	2975420	2975721	16	165	1.0	2.8	1.0	2.7	10.8	18.5	8.7	hypothetical protein	Unknown function
NCgl2704	cg3102	-	2988684	2988985	63	-	1.0	1.0	1.0	3.4	6.0	7.8	7.4	nucleosidase	Nucleotide transport and metabolism
NCgl2710	-	-	2996821	2997122	180	-	4.6	1.0	2.7	4.1	13.1	23.5	21.2	hypothetical protein	Unknown function
NCgl2741	cg3142	-	3027375	3027676	36	36	1.0	1.0	1.0	1.0	2.7	2.8	3.2	hypothetical protein	Unknown function
NCgl2748	-	-	3032407	3032708	104	1264	1.0	1.0	1.0	6.8	1.3	3.3	1.7	transposase	DNA replication, recombination, repair, and degradation
NCgl2766	cg3170	-	3054406	3054707	197	-	2.2	1.0	1.0	1.0	1.9	3.6	2.5	hypothetical protein	General function prediction only
NCgl2767	cg3172	*trmB*	3055689	3055990	28	27	1.0	4.0	1.0	3.3	7.0	10.5	6.6	tRNA (guanine-N(7)-)-methyltransferase	Translation, ribosomal structure and biogenesis

NCgl2773	cg3178	*pks*	3067766	3068067	136	-	1.0	2.8	1.5	1.0	2.1	3.0	2.6	polyketide synthase	Cell wall/membrane/envelope biogenesis
NCgl2774	cg3179	*fadD2*	3069917	3070218	65	-	1.0	1.0	1.9	3.0	3.5	9.3	3.9	acyl-CoA synthetase	Cell wall/membrane/envelope biogenesis
NCgl2806	cg3214	-	3107286	3107587	514	-	1.0	1.0	1.0	1.6	5.5	7.4	4.7	hypothetical protein	Unknown function
NCgl2841	cg3254	-	3141925	3142226	106	-	1.0	1.0	1.0	2.1	3.2	4.1	2.7	hypothetical protein	Unknown function
NCgl2847	cg3263	-	3152196	3152497	67	20	1.0	1.0	1.0	2.8	5.2	12.7	9.5	hypothetical protein	Unknown function
NCgl2878	cg3304	*dnaB*	3183050	3183351	334	-	1.0	3.3	4.4	10.8	22.7	49.6	24.6	replicative DNA helicase	DNA replication, recombination, repair, and degradation
NCgl2885	cg3314	-	3189542	3189843	7	-	1.0	1.0	1.0	1.0	1.0	2.6	2.3	hypothetical protein	Unknown function
NCgl2902	cg3332	*qor3*	3206247	3206548	165	-	1.0	1.0	2.0	1.0	1.8	4.2	3.4	hypothetical protein	Respiration and oxidative phosphorylation
NCgl2904	cg3335	*malE*	3207991	3208292	138	99	1.0	1.0	1.0	1.8	2.2	3.1	1.0	malic enzyme	Central carbon metabolism
NCgl2911	-	-	3217229	3217530	164	-	1.0	1.0	1.0	4.6	9.3	20.8	10.7	hypothetical protein	Unknown function
NCgl2912	cg3343	-	3217708	3218009	81	-	1.0	1.0	1.0	1.0	6.8	6.5	4.6	hypothetical protein	Unknown function
NCgl2948	cg3381	-	3253391	3253692	19	8	1.0	1.0	1.0	1.0	1.0	2.2	3.1	hypothetical protein	Protein secretion
NCgl2960	cg3394	-	3271146	3271447	168	53	1.0	1.0	1.0	1.0	1.0	2.9	1.0	hypothetical protein	Unknown function

Figure S1

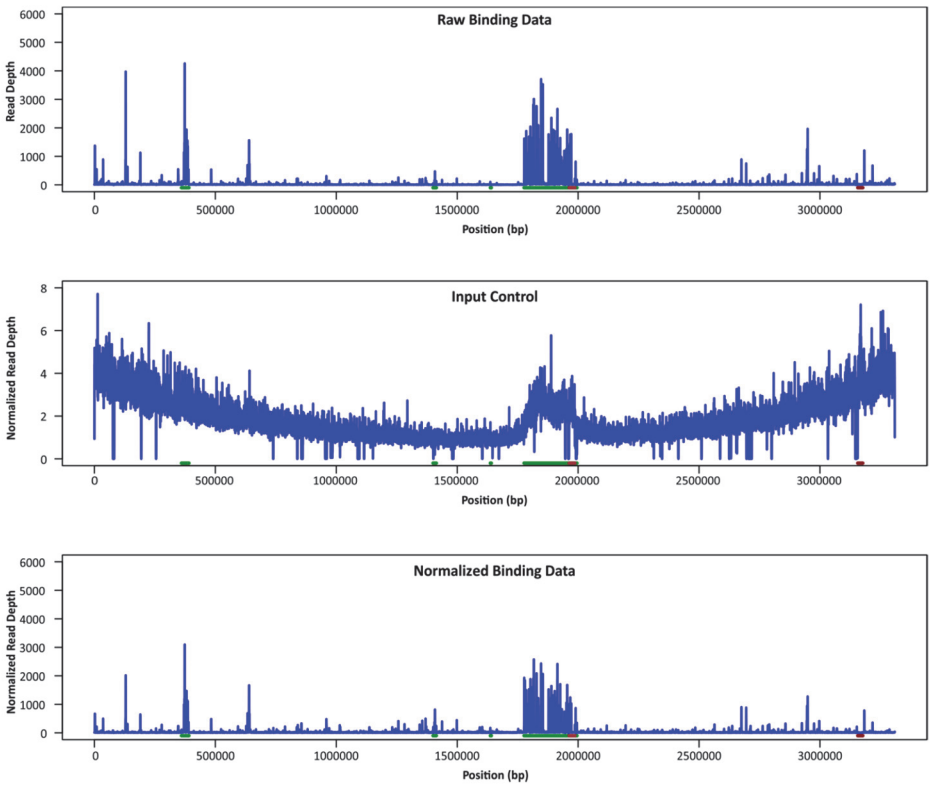


Figure S1: Example normalization strategy used for the analysis of the ChAP-seq time series. The data shown here derives from the time series represented in Figure 4. A normalization for the ChAP-seq analyses was performed using an input control to remove a bias created by an overrepresentation of specific DNA regions. Here, an example for t = 6 h is represented. The top graph shows the raw binding peaks of CgpS, only normalized intrinsically to the average coverage. The second graph visualizes the whole genome sequencing results of the corresponding input control, normalized to its minimum area as 1, for the minimum copy number of the bacterial chromosome. The bottom graph shows a normalization of the top graph towards its input control. The green lines represent regions of low GC content and prophage regions, the dark red line represents a high GC region and the brown inclusion inside the big green CGP3 region marks the CGP4 area that integrated into CGP3.

Figure S2

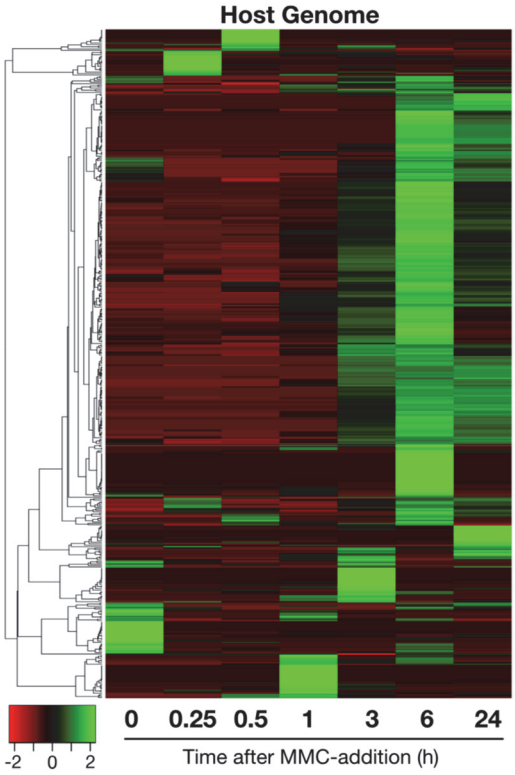


Figure S2: Heatmap of the binding pattern of CgpS towards host targets at different time points. Shown is a heatmap of an extract of the binding data of CgpS (cf. Table S1). Included are only targets outside of the CGP3 area. The heatmap was hierarchical clustered using the Pearson correlation and an average linkage clustering with the online tool heatmapper.ca (Babicki et al., 2016).

Figure S3

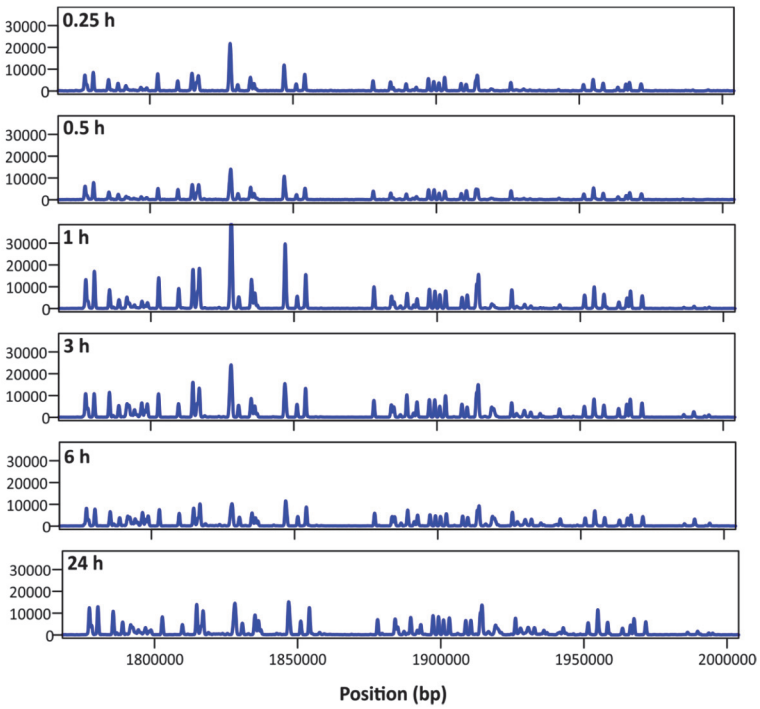


Figure S3: Enlargement of the CgpS binding inside the CGP3 region. This plot is a zoomed-in version of the time series represented in Figure 4. This zoom only shows the CGP3/CGP4 area of the chromosome. For the data generation, *C. glutamicum* ATCC 13032::*cgpS-strep* cells were grown as described previously in CGXII medium containing 2 % (w/v) glucose and 600 nM mitomycin C (all but t = 0 h). Subsequently, cells were harvested at different timepoints (0 h, 0.25 h, 0.5 h, 1 h, 3 h, 6 h, 24 h) and CgpS-bound DNA was purified and sequenced. The presented data was normalized with regard to the average intrinsic coverage for inter-experiment comparison as well as to an genomic input controls for each time point, in order to remove an impact of the prophage copy number as well as of the higher amount of DNA towards the ori region on the binding peaks size after mitomycin C addition (compare Figure S1).

4.4 Supplementary Information "The MarR-Type Regulator MalR Is Involved in Stress-Responsive Cell Envelope Remodeling in *Corynebacterium glutamicum*"

Supplementary Data

Table S1: DNA regions discovered as binding sites from MalR via ChAP-Seq. Listed are all regions bound by MalR, as identified by ChAP-Seq experiments and the corresponding mRNA reaction obtained from comparative transcriptomics of the wild type and a *malR* overexpression strain (*malR*$^{+}$) or the deletion mutant (Δ*malR*). Values represent the fold-change of downregulated genes (<0.5, marked in red) and upregulated genes (>2, marked in green). The CGP3 prophage region is shaded in blue. Genes with a p-value > 0.05 are shown in dark red. (IG) = intragenic binding. Light-grey numbers highlight very broad peak areas, which may affect also the neighbouring genes.

Table S2: Transcriptomic changes triggered by an overexpression or a deletion of *malR*. Shown is the average fold change of triplicates with the corresponding p-value. Marked in red are the downregulated genes (<0.5) and marked in green the upregulated genes (>2). The CGP3 prophage region is shaded in blue. p-values worse than 0.05 are highlighted in red. If no p-value is given, only one replicate was analyzed.

frontiers

Supplementary Material

1 Supplementary Data

2 Supplementary Figures and Tables

2.1 Supplementary Figures

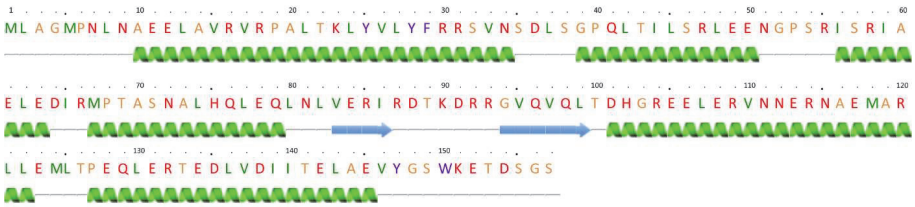


Figure S1: Prediction of the secondary structure of MalR. Using the online tool Phyre[2], the secondary structure of MalR was predicted with 99.9% confidence of 92 % of residues (Kelly L.A., Mezulis S., Yates C., Wass M., Sternberg, M. 2015. Nat. Protoc.10:845–858.).

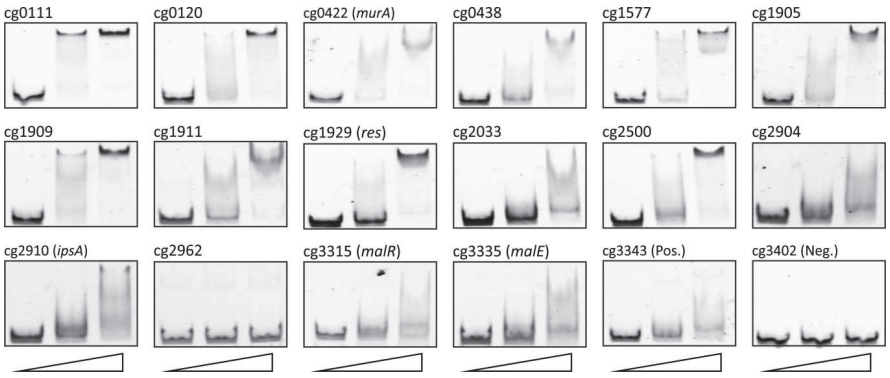


Figure S2: *In vitro* DNA-binding of purified MalR protein. Electrophoretic mobility shift assays (EMSAs) were performed to verify binding of MalR to promoter regions identified via ChAP-Seq. Therefore, MalR was purified with a C-terminal Strep-tag fusion. At total, 90 ng of 100 bp DNA fragments (50 bp up- and downstream the peak maximum) were incubated without protein (first lane), with 3 molar excess (228 nM, second lane) and 10 molar excess of purified MalR (760 nM, third lane) for 30 min in bandshift-buffer (50 mM Tris-HCl, 5 mM MgCl2, 40 mM KCl, 5 % (v/v) glycerol, pH 7.5). Subsequently, samples were separated on a 10% native polyacrylamide gel electrophoresis and gels were stained using SYBR Green I Nucleic Acid Gel Stain (Lonza, Rockland, ME, USA).

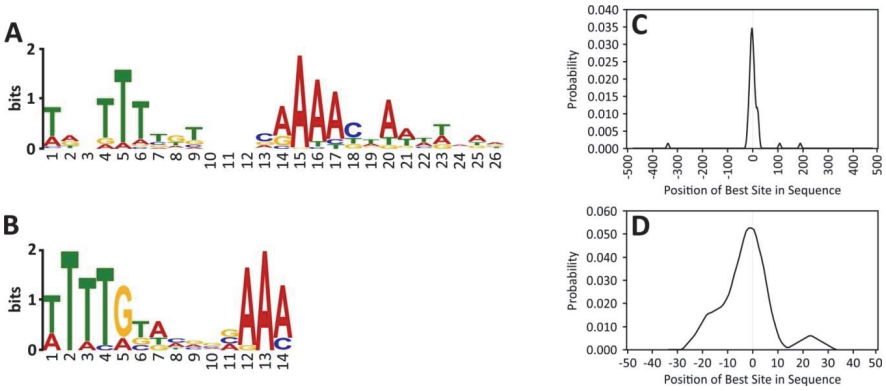


Figure S3: Deduction of a DNA binding motif of MalR The online tool MEME-ChIP (Ma W, Noble WS, Bailey TL. 2014. Nat Protoc 9:1428–1450.) was used to predict MalR binding sites on the basis of all targets sequences (each sequence 1000 bp) identified via ChAP-Seq analysis **(A)** and based on the 16 MalR targets (100 bp each) verified using EMSAs **(B)**. **(C, D)** The distribution of the predicted binding sites throughout the single uploaded DNA sequences was determined using CentriMo (Bailey TL, MacHanick P. 2012. Nucleic Acids Res 40:e128.).

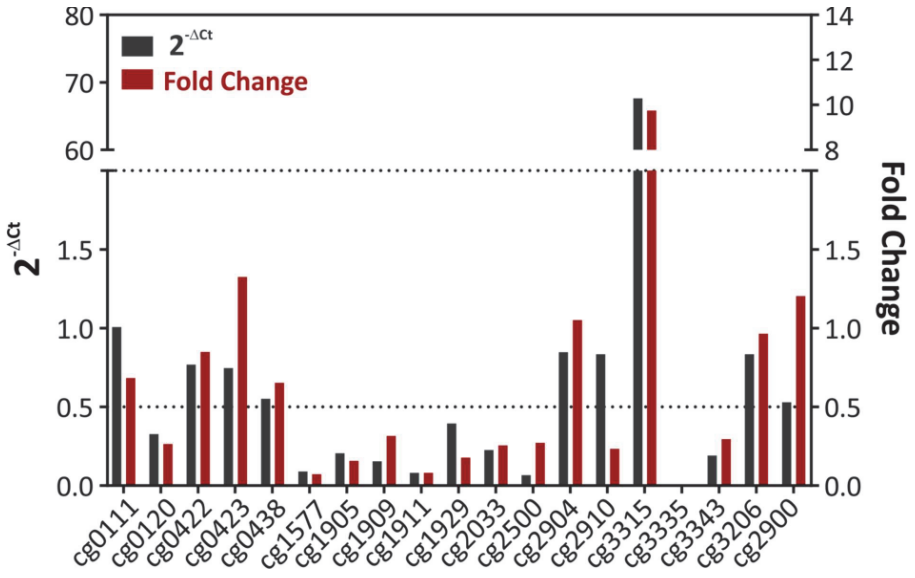


Figure S4: Verification of DNA-microarray data using qRT-PCR. Microarray data were verified with qRT-PCRs (oligonucleotides are listed in Table S3 D). The dark-grey bars represent the relative transcriptional change based on the qRT-PCR data and the red bars show the average fold-change obtained from DNA microarray experiments. The values below 0.5 are classified as "downregulated", the values above 2 are classified as "upregulated". In the case of cg3335, the bars are not visible because of very low values (qRT-PCR: 0.018, DNA-microarray: 0.007).

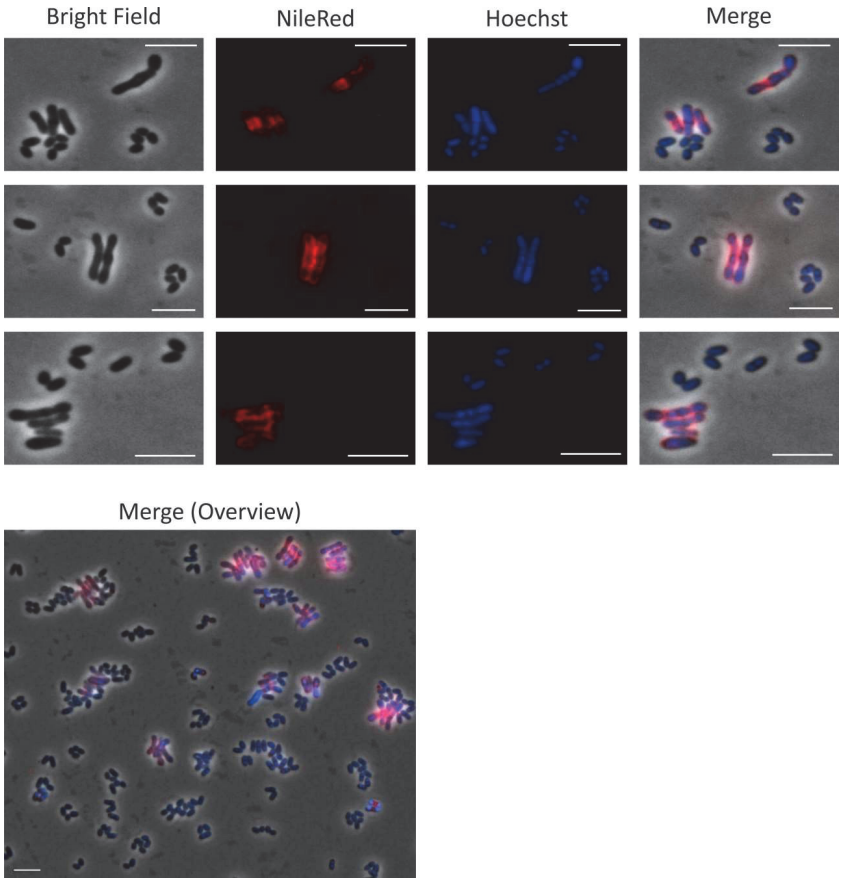


Figure S5: MalR overproduction causes severe growth defects of *C. glutamicum*. For microscopic analysis, cells were grown in CGXII medium for 24 h at 30°C. Shown are *C. glutamicum* ATCC 13032 cells carrying the over expression vector pEKEx2-*malR*. The expression of *malR* is induced by the addition of 100 µM IPTG. Lipid components of the cell membrane were stained with Nile Red (red); DNA was stained with Hoechst 33342 (blue). The white scale bars represent 5 µm. In addition to chosen examples of elongated cells an overview is shown to present the distribution of elongated cells inside the samples.

5

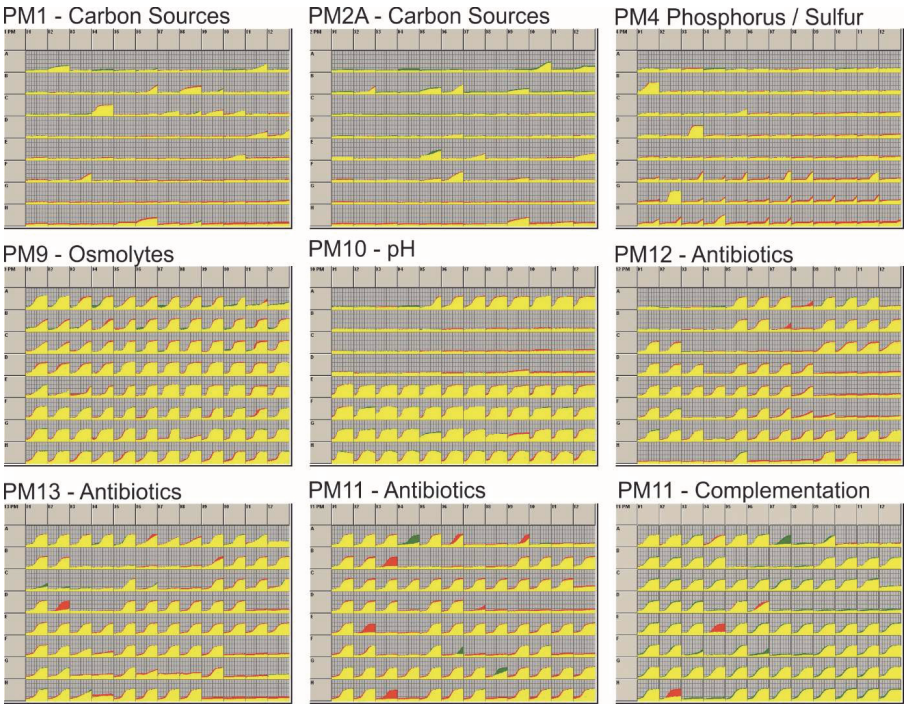


Figure S6: Overview of all tested PM plates for a comparison of *C. glutamicum* Δ*malR* and *C. glutamicum* wild type. An OmniLog System from Biolog (Hayward CA, USA) was used to perform phenotypic microarrays with wild type *C. glutamicum* ATCC 13032 as well as the *malR* deletion strain. The experiments (PM1, PM2A, PM4, PM9, PM10, PM11, PM12, PM13) were conducted like described in the protocol of the manufacturer (Bochner BR, Gadzinski P, Panomitros E. 2001. Genome Res 11:1246–1255.). The wild type strain is displayed in red, whereas the *malR* deficient mutant is green. Yellow means a similar behavior. PM11 was repeated ("PM11 – Complementation") using the *malR* deficient strain with an empty vector (pEKEx2, green) as well as the *malR* deficient strain carrying the vector pEKEx2-*malR* for complementation studies. The expression of *malR* was induced by the addition of 5 µM IPTG.

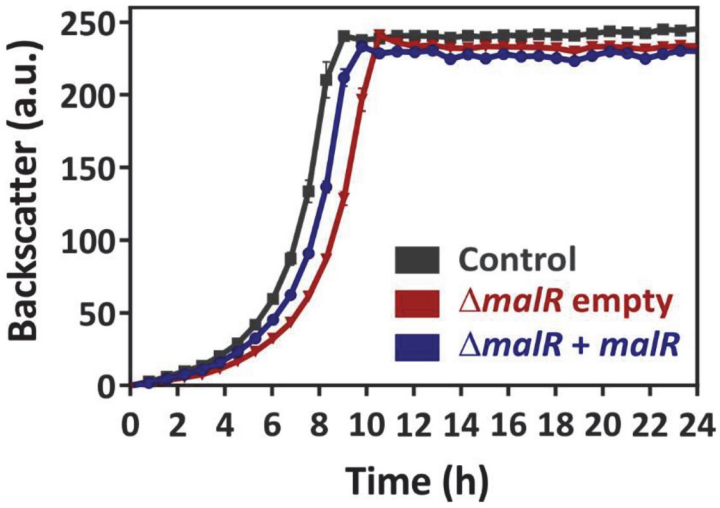


Figure S7: Comparison of the growth of the *C. glutamicum* strains used for the complementation of phenotypes analysed via phenotypic microarrays. To complement the effects of the *malR* deletion, *C. glutamicum* ATCC 13032 wild type cells with the empty vector pEKEx2 (Control) were cultivated in comparison with a strain lacking the *malR* gene either with the empty vector pEKEx2 or with the *malR* encoding vector. The production of MalR was induced using 5 µM IPTG. Remarkably, strain Δ*malR* shows a reduced growth rate under these conditions. This phenotype was successfully complemented by the introduction of plasmid-encoded *malR*. Overall, this finding suggested a slightly increased sensitivity of strain Δ*malR* towards the antibiotic kanamycin.

7

2.2 Supplementary Tables

Table S3 A: Oligonucleotides used in this study

Number	Sequence (5'-3')
1	AGGTCGACTCTAGAGGATCCTCGAACGGAAATTACTTGGCAATAC
2	CTTCTCGAACTGTGGGTGGGACCAGCTAGCAGAACCGCTGTCGGTCT
3	GCTAGCTGGTCCCACCCACAGTTCGAGAAGTAACAGTTTTCTCCATCTCAACTCC
4	AAACGACGGCCAGTGAATTCACAAGTCCTAGTGGGACGG
5	AGGTCGACTCTAGAGGATCCCCGTCCATTCTTTACCAACTGT
6	CCCATCCACTAAACTTAAACAAGCGGTATTGCCAAGTAATTTCC
7	TGTTTAAGTTTAGTGGATGGGCAGTTTTCTCCATCTCAACTCCG
8	AAACGACGGCCAGTGAATTCGGCACAAGTCCTAGTGGGAC
9	TTTAAGAAGGAGATATACATATGCTGGCAGGCATGC
10	AAATACAGGTTCTCGCTAGCAGAACCGCTGTCGGTCTC
11	AAGCTTGCATGCCTGCAGAAGGAGGAGTCGTATGCCTAATTTAAACGCTGAGGAG
12	AAACGACGGCCAGTGAATTCTTAAGAACCGCTGTCGGTCTC
13	GCCAAGGTAAGTTGTACTTTTTCTG
14	GTAGTTGACGCGGGTGAC
15	TAACAATCCCAACTCGAAGCAC
16	TGACAATCCTCTCCACGAAGC
17	GAGTGTGAGAAACATGGGACGTG
18	ACAGGTTACCAGCCAAAGTG
19	TCGATACCCAGAAAGAATTGCATTTG
20	ATCTCCCTGGATAGTGATCTTGC
21	CCGTCCAGAACTAGGACTATTTG
22	TAATCCCAACAATCGCTTATGACG

23	TTGAAATCGTGATCGCCTGTTATTG
24	AGCCATGTTTGCTTCTCCTTTTC
25	TCAGGTTTACTACCGACTGTAGTAG
26	TTTGAAAGTAGAAATTACTGGTAGTGGATTA
27	TTTACCTTCCTCTACCTAATCTCCC
28	ATCTCAGTAGCGTTGCCTCC
29	TTTCTCATTTCGCCACCCCC
30	CTAAGATGAGTATTTAAGCCCTGTTTAT
31	ACGCCCGTATCGTTTCGC
32	ATGTACTTGTTTAAAACAATAGTTGTCAATG
33	GCACCCATGGTTAGCGTACT
34	TTCTACCGGGGGTAGTAGCG
35	TGTTGCTTCCGCTGACC
36	ACCTTCTATATAAACCTTTTATGAGGGAAATG
37	CTGGACCGAGCTCAATGC
38	TGAAAAATTGTTTTAATATGCAACACAACTA
39	TTGTAGGCGTTGGACACTG
40	AGATTGAAATGATTATGGGTAGGAAAC
41	AAGAGCCGTGATGTTAACAAATG
42	TAGTTAGGTTACACTAATGGTGTCC
43	GGAACAGCACAGAATTAAGGC
44	GAAAAAGCAATAATTTGGACAGAAAAAG
45	AAACCACCCCTGTACAAAATTAGC
46	AAAAACACTAGCTAAATCTGTAGCTCAAC
47	ATGTTTCTTACTTTAAGCGCAGTTAATTG

48	GGATTTAAATTTTGTATTGGAAAGCTAATAATT
49	CATAGGGGTATAGCCTTGAG
50	CAGTGTGCGCAGGTCATGCC
51	GTATGGGTTGAGATTCCGACAG
52	TCTTCCCTATCACCTCCAGTT
53	GGAGCTTTCGCTGACTATCTT
54	AATTCTCCTGCGTCGTCTTT
55	GACGTGGATCGTGTATGGAATTA
56	CAGCCTTCTCAAGGTGGATAAA
57	ACTGTGGATCGACATTCCTTTC
58	CCGCGTCTAAGTCCCTTTAATC
59	ATGACTCGATCCCAGGACTAT
60	TGAATGCGGTATGAGCTAAGG
61	GCCCAAACACCACCGATATT
62	CATCAGTCAAGCAGGTCTGAAC
63	CATCTTCCATCCACAGACCTAAT
64	GAAGTCGGACACGATGTAGAAG
65	ATCTCGTGCTTGCTGTGATTA
66	CTGGGTGAATCCTAAAGACCTG
67	CTGATGGGTCAGTCGTAGTTTC
68	CCGAATAGAGCCAGGAACAAT
69	CTCGACAACGTGAAGCTGTTA
70	CAAGCCAGGTCTGTGTGATT
71	GCTGTAGGTAAGGGCTTTGATA
72	TAGCCAACAGCCTGGTAATC

73	GAGCTTCAGAAACTTGCAACAG
74	CATTGAGGGCGAGGATGATT
75	AGCGAAAGTTCCCGAATCTG
76	CGGACGGTCAGTCTTGTTATG
77	CTCGACACCGCTGAAGATATG
78	CAGTAAGCAGCACTCCGATT
79	GGGTCCACAGCTCACTATTT
80	AGCGGTTGGCATACGAATA
81	AGCAATCAAGGAAGATCCAGAG
82	GATATCGCCAAGGCCAAGAA
83	CCGGTTCTTCTGCATCTTCT
84	GGTTACATCGTCGAAGTCCTTA
85	CAGACTCACAACAACGTCAAAC
86	CTAGTTCGTGGCCAACTTCA
87	TTCCTCACAGATCGCTTTCG
88	GAGCAGGTATGGAGCAACTT

Table S3 B: Construction of plasmids used in this study. Numbers represent oligonucleotide pairs used for PCR (see Table S3 A). The restriction enzymes were used for linearization of the vectors and plasmids were assembled using Gibson assembly. Sequencing for verification of the chromosomal modifications was conducted using primers 13 + 14.

Plasmid	Template	Primers	Vector	Restriction Enzymes
pEKEx2-*malR*	*C. glutamicum* chromosome	11 + 12	pEKEx2	*PstI *EcoRI
pK19-*malR*-C-*strep*	*C. glutamicum* chromosome	1 + 2; 3 + 4	pK19*mobsacB*	*BamHI *EcoRI
pK19-Δ*malR*	*C. glutamicum* chromosome	5 + 6; 7 + 8	pK19*mobsacB*	*BamHI *EcoRI
pET24b-*malR*-C-*strep*	*C. glutamicum* chromosome	9 + 10	pET24b	*NheI *NdeI

Table S3 C: Amplification of EMSAs DNA probes. Indicated are the oligonucleotide pairs used for PCR amplification of the 100 bp fragments. The fragments cover the maximum peak position determined by ChAP-Seq and lay inside the promoter regions of each gene.

Fragment	Gene	Oligonucleotides (Table S1)
1	cg0111	15 + 16
2	cg0120	17 + 18
3	cg0423	19 + 20
4	cg0438	21 + 22
5	cg1577	23 + 24
6	cg1905	25 + 26
7	cg1909	27 + 28
8	cg1911	29 + 30
9	cg1929	31 + 32
10	cg2033	33 + 34
11	cg2500	35 + 36
12	cg2904	37 + 38
13	cg2910	39 + 40
14	cg2962	41 + 42
15	cg3315	43 + 44
16	cg3335	45 + 46
17	cg3343 (positive control)	47 + 48
18	cg3402 (negative control)	49 + 50

Table S3 D: Oligonucleotide combinations for qRT-PCR analysis to verify microarray data. Indicated are the oligonucleotide pairs used for qRT-PCR. Each pair comprises approximately 100 bp fragments and shares similar features regarding GC-content and melting temperature.

Fragment	Gene	Oligonucleotides (Table S1)
1	cg0111	51 + 52
2	cg0120	53 + 54
3	cg0422	55 + 56
4	cg0423	57 + 58
5	cg0438	59 + 60
6	cg1577	61 + 62
7	cg1905	63 + 64
8	cg1909	65 + 66
9	cg1911	67 + 68
10	cg1929	69 + 70
11	cg2033	71 + 72
12	cg2500	73 + 74
13	cg2904	75 + 76
14	cg2910	77 + 78
15	cg3315	79 + 80
16	cg3335	81 + 82
17	cg3343	83 + 84
18	cg3206	85 + 86
19	cg2900	87 + 88

Table S4: Antibiotics affecting the growth of wild type cells or Δ*malR* cells in phenotypic microarrays (BioLog). For evaluation of the sensitivity of the wild type ATCC 13032 or the *malR* deletion mutant, the BioLog plates PM11, PM12 PM13 were used. The table shows antibiotics were both strains showed differences in the metabolic activity (see Figure 6 and Figure S6).

Antibiotic	Better Growth	Substance class
Gentamicin	Δ*malR*	Aminoglycoside
Amikacin	Δ*malR*	Aminoglycoside
Lincomycin	wt	Lincosamide
Erythromycin	Δ*malR*	Glycoside
Chlortetracyclin	wt	Tetracycline
Demeclocycline	wt	Tetracycline
Tetracyclin	wt	Tetracycline
Penimepicyclin	wt	Tetracycline
Cefazolin	wt	β-Lactame; Cephalosporine
Cephalothin	wt	β-Lactame; Cephalosporine
Cefuroxime	wt	β-Lactame; Cephalosporine
Amoxicillin	wt	β-Lactame; Penicilline

Acknowledgement

First of all, I would like to express my special thanks to Prof. Dr. Julia Frunzke, for giving me the opportunity to work on this interesting project, for all her help and support, for all fruitful discussions and for her continuous interest in my scientific and personal development.

Furthermore, I would like to thank Prof. Dr. Michael Feldbrügge for agreeing to be my co-supervisor.

A special thanks to all current and former group members of AG Frunzke. Especially, Dr. Eva Davoudi, Dr. Andrei Filipchyk, Cornelia Gätgens, Aël Hardy, Isabel Huber, Dr. Marc Keppel, Larissa Kever, Aileen Krüger, Sophia Lorke, Dr. Eugen Pfeifer, Nadine da Silva, Robert Stella, Ulrike Viets, Johanna Wiechert, and Raphael Freiherr von Boeselager. Thank you for all your help, for nice discussions, for funny lunchbreaks and retreats, and thank you for creating the best working atmosphere one could wish for!

An additional very special thanks to Conni Gätgens and Nadine da Silva for all their help in the lab and to my 'proofreading gang' for their time and valuable input. Especially a very big thank you goes to Dr. Eva Davoudi for her 'pre-submission-weekend-assistance'.

Ein riesiges Danke geht auch an alle meine Freunde, die nicht nur meine stetige Abwesenheit und gelegentlichen schlechten Launen während der Schreibphase ertragen haben, sondern mich auch immer wieder aufgebaut und unterstützt haben.

Ganz besonders möchte ich an dieser Stelle auch meiner Familie danken. Danke an meine Eltern, dass Ihr immer für mich da wart und da seid, wenn ich euch brauche und dass ihr mir ermöglicht habt, den Weg einzuschlagen, den ich gegangen bin. Danke auch an meine Omas für ihre Unterstützung in jeglicher Lebenslage. Ebenfalls ein großes Danke geht an meinen Onkel Frank, auf dessen Hilfe ich immer bauen kann.

Ich danke auch meiner kleinen Tochter Jana, dafür, dass sie es immer schafft, mir, egal wie angespannt ich bin, ein Lächeln aufs Gesicht zu bringen.

Zu guter Letzt möchte ich meiner Frau Sabrina danken. Du hältst mir immer den Rücken frei und bist immer für mich da, wenn ich dich brauche. Besonders während der intensiven Endphase der Promotion hast mich immer aufgemuntert und an mich geglaubt. Ich danke dir für Alles!

Erklärung

Hiermit versichere ich an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Max Hünnefeld