Phylogenetic and functional analyses of stress-responsive bacterial transmembrane signal transducing systems

Dissertation

zur Erlangung des Doktorgrades
der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

vorgelegt von
Anna Staroń
aus Warschau

München 2012

Erstgutachter: Prof. Dr. Thorsten Mascher
Zweitgutachter: Prof. Dr. Heinrich Jung

Tag der mündlichen Prüfung: 24.10.2012

Eidesstattliche Versicherung und Erklärung

Hiermit versichere ich an Eides statt, dass die vorliegende Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt wurde. Zudem wurden keine anderen als die angegebenen Quellen verwendet.

Außerdem versichere ich, dass die Dissertation keiner anderen Prüfungskommission vorgelegt wurde und ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

München, 13.06.2012

_____

Anna Staroń

Content

Abbreviations

1CS	one-component system
2CS	two-component system
ABC	ATP binding cassette
ASD	anti-σ domain
bp	base pairs
CMD	carboxymuconolactone decarboxylase
ECF	extracytoplasmic function
Fig.	figure
HK	histidine kinase(s)
IM-HK	intramembrane-sensing histidine kinase(s)
IPTG	isopropyl-β-D-thiogalactopyranoside
MLS	macrolide-lincosamide-streptogramin B
MRSA	methicillin-resistant *Staphylococcus ureus*
MSD	membrane spanning domain
NBD	nucleotide binding domain
OD	optical density
PCR	polymerase chain reaction
RR	response regulator(s)
X-Gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
ZAS	zinc-binding anti-σ

List of publications

Publications and manuscripts presented in this thesis:

Chapters 1 and 8

Schrecke, K.*, **Staroń, A.***, and Mascher, T. Two-component signaling in the Gram-positive envelope stress response: intramembrane-sensing histidine kinases and accessory membrane proteins. In *Two component systems in bacteria.* Beier, D., and Gross, R. (eds). Horizon Scientific Press, *in press* (* shared first authorship)

Staroń, A., and Mascher, T. (2010) Extracytoplasmic function σ factors come of age. *Microbe* **5**:164-170

Chapter 2

Staroń, A., Sofia, H.J., Dietrich, S., Ulrich, L.E., Liesegang, H., and Mascher, T. (2009) The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Mol Microbiol* **74**:557-581

Chapter 3

Staroń, A., and Mascher, T. (2010) General stress response in α-proteobacteria: PhyR and beyond. *Mol Microbiol* **78**:271-277

Chapter 4

Dintner, S.*, **Staroń, A.***, Berchtold, E., Petri, T., Mascher, T., and Gebhard, S. (2011) Co-evolution of ABC-transporters and two-component regulatory systems as resistance modules against antimicrobial peptides in Firmicutes bacteria. *J Bacteriol* **193**:3851-3862 (* shared first authorship)

Chapter 5

Staroń, A., Finkeisen, D.E., and Mascher, T. (2011) Peptide antibiotic sensing and detoxification modules of *Bacillus subtilis*. *Antimicrob Agents Chemother* **55**:515-525

Publications and manuscripts not presented in this thesis:

Wecke, T., Halang, P., **Staroń, A.**, Dufour, Y.S., Donohue, T., and Mascher, T. (2012) Characterization of the novel group ECF41 σ factors. *MicrobiologyOpen* **1**:194-213

Author contributions

Staroń, A., Sofia, H.J., Dietrich, S., Ulrich, L.E., Liesegang, H., and Mascher, T. (2009) The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Mol Microbiol* **74**:557-581

Anna Staroń performed the bioinformatic analyses including sequence retrieval, classification of σ factors, as well as genomic context and promoter analyses. Heidi Sofia performer the bioinformatic analyses regarding anti-σ factors. Heiko Liesegang and Sascha Dietrich constructed and tested the Hidden Markov Models and designed the ECF*finder*. Luke Ulrich implemented the Hidden Markov Models in the MiST database. Anna Staroń, Heidi Sofia, Heiko Liesegang and Thorsten Mascher designed the analyses. Anna Staroń and Thorsten Mascher wrote the manuscript.

Staroń, A., and Mascher, T. (2010) Extracytoplasmic function σ factors come of age. *Microbe* **5**:164-170

Anna Staroń and Thorsten Mascher performed the literature search and wrote the manuscript.

Staroń, A., and Mascher, T. (2010) General stress response in α-proteobacteria: PhyR and beyond. *Mol Microbiol* **78**:271-277

Anna Staroń performed the bioinformatic analysis. Anna Staroń and Thorsten Mascher wrote the manuscript.

Staroń, A., Finkeisen, D.E., and Mascher, T. (2011) Peptide antibiotic sensing and detoxification modules of *Bacillus subtilis*. *Antimicrob Agents Chemother* **55**:515-525

Anna Staroń performed the experiments and the meta-analysis of expression profiles. Dora Finkeisen constructed strains and performed promoter activity measurements. Anna Staroń and Thorsten Mascher designed the experiments and wrote the manuscript.

Dintner, S.*, **Staroń, A.***, Berchtold, E., Petri, T., Mascher, T., and Gebhard, S. (2011) Co-evolution of ABC-transporters and two-component regulatory systems as resistance modules against antimicrobial peptides in Firmicutes bacteria. *J Bacteriol* **193**:3851-3862 (* shared first authorship)

Anna Staroń performed the bioinformatic analyses including initial data acquisition, phylogenetic analysis of histidine kinases, analysis of response regulator binding sites, and assignment of orphan transporters. Sebastian Dintner perfomed the bioinformatic analyses including sequence retrieval, phylogenetic analysis of permeases and secondary structure predictions. Tobias Petri and Evi Berchtold performer the analysis of coevolution. Susanne Gebhard, Tobias Petri and Thorsten Mascher designed the experiments. All authors contributed to the writing of the manuscript.

Schrecke, K.*, **Staroń, A.***, and Mascher, T. Two-component signaling in the Gram-positive envelope stress response: intramembrane-sensing histidine kinases and accessory membrane proteins. In *Two component systems in bacteria.* Beier, D., and Gross, R. (eds). Horizon Scientific Press, *in press* (* shared first authorship)

Karen Schrecke, Anna Staroń and Thorsten Mascher performed the literature search and wrote the manuscript.

For reasons of consistency, the layout of the publications presented in this thesis has been changed.

Summary

Bacteria are constantly exposed to many changes in the surrounding environment. In order to cope with the fluctuating conditions microorganisms need to detect these changes and respond to them. Signal transducing systems enable bacteria to perceive the changes in the relevant parameters and react to them by mounting an appriopriate cellular response. A variety of signal transducing mechanisms orchestrate such responses, with the majority belonging to one of three families: 1CS, 2CS and ECF σ factors.

In the first part of this thesis, we investigated the principles governing the ECF σ factors. ECF σ factors are signal transducing systems widely distributed in bacteria. Comparative genomics analysis revealed common themes present in ECF subfamilies, and led to development of a classification system. This classification allows predictions of novel mechanisms of ECF-dependent signal transduction and potential target promoters. We followed this classification with an in-depth study of a single group of ECF σ factors orchestrating the general stress response of α-proteobacteria. This study demonstrates the potential of comparative genomics analyses and predicts HWE histidine kinases as interaction partners taking part in the ECF-dependent signaling cascade.

In the second part of this thesis, we characterized the family of Bce-like 2CS and associated ABC transporters. Comparative genomics analysis revealed that Bce-like 2CS are restricted to Firmicutes bacteria and appear to have evolved as detoxification mechanisms against antimicrobial peptides. Subsequently, we investigated the inducer spectrum of Bce-like 2CS in *Bacillus subtilis*. The three 2CS, BceRS, PsdRS and YxdJK form stand-alone detoxification modules and respond to a number of peptide antibiotics. All three systems show high level of sequence similarity but do not show significant cross-talk. The insulation at the promoter level seems to be orchestrated by multiple response regulator binding sites. We also present the first data on specificity determination between the histidine kinases and response regulators in these systems.

Zusammenfassung

Bakterien sind ständig wechselnden Umweltbedingungen ausgesezt. Um in den sich verändernden Bedingungen zu bestehen, müssen Mikroorganismen die Veränderungen detektieren und sich an sie anpassen. Signaltransduktionsmechanismen ermöglichen es, die Änderungen wahrzunehmen und auf sie zu reagieren. Ubiquitäre Signaltransduktionsmechanismen umfassen drei Proteinfamilien: Einkomponentensysteme, Zweikomponentensysteme und *extracytoplasmic function* (ECF) σ Faktoren.

Im ersten Teil dieser Arbeit wurden die Grundprinzipien der ECF σ Faktor-abhängigen Signaltransduktion untersucht. ECF σ Faktoren sind bei Bakterien weit verbreitet. Eine vergleichende Genomanalyse zeigte, welche Mechanismen der ECF σ Faktor-abhängigen Signaltransduktion unterliegen. Basierend auf der Analyse haben wir eine ECF Klassifizierung entwickelt, die es uns erlaubt, neue ECF-abhängige Mechanismen vorzuhersagen. Mittels der Klassifizierung können auch potentielle Zielpromotoren vorhergesagt werden. Auf die Analyse folgt eine detaillierte Studie einer Untergruppe der ECF σ Faktoren. Diese Proteinfamilie ist an der generellen Stressantwort in α-Proteobakterien beteiligt. Unsere Analyse veranschaulicht das große Potential der vergleichende Genomanalyse als Ansatz für die Vorhersagen neuer Signaltransduktionsmechanismen.

Im zweiten Teil der Arbeit wurden die Bce-ähnlichen Zweikomponentensysteme untersucht und charakterisiert. Eine vergleichende Genomanalyse zeigte, dass die Bce-ähnlichen Zweikomponentensysteme fast ausschließlich im Phylum Firmicutes vorkommen und wahrscheinlich als Detoxifikationsmodule enstanden sind. Anschließend haben wir das Induktorenspektrum von Bce-ähnlichen Zweikomponentensystemen in *Bacillus subtilis* untersucht. Die drei Systeme, BceRS, PsdRS und YxdJK sind eigenständige Detoxifikationsmodule und reagieren auf mehrere Peptidantibiotika. Alle drei Systeme zeigen einen hohen Grad von Sequenzähnlichkeit aber keine Kreuzregulation. Auf der Promotorebene kann dies durch die Anwesenheit von mehreren Antwortregulatorbindestellen erklärt werden. In dieser Arbeit wurden auch die Spezifitätdeterminanten in den Zweikomponentensystemen untersucht.

CHAPTER 1

Introduction

Parts of this chapter have been adapted from:

A. Staroń, and T. Mascher

Microbe (2010) 5(4):164-170

and

K. Schrecke*, A. Staroń*, T. Mascher

* *contributed equally*

Two component systems in bacteria. D. Beier, and R. Gross (eds.).

Horizon Scientific Press, *in press*

Chapter 1
Introduction

Bacteria are constantly adapting to a changing environment. Their ability to sense and respond to stimuli is critical for survival in various environmental conditions. The challenges bacteria face are very diverse, ranging from fluctuating temperature, changes in concentration of nutrients and oxygen availability in the soil to exposure to antimicrobial peptides secreted by the animal immune system. In order to cope with all the challenges bacteria need to detect changes in the relevant parameters and respond to them. Microorganisms employ different signal transduction mechanisms by which they can perceive a variety of input signals and subsequently react to them by changing this piece of information into an appropriate cellular response.

1.1. Principles of bacterial signal transduction

Different systems take part in the transformation of the signal from the stimulus detection to the cellular response. The stimulus can be detected intra- or extracellularly by a designated protein. Subsequently, the cellular response is induced, either by the same protein that sensed the stimulus or by its interaction partner. The most widely distributed bacterial signalling systems are (i) one-component systems (1CS), where the input- and output domains are located on one polypeptide chain (Ulrich *et al.*, 2005), (ii) two-component systems (2CS), consisting (in the simplest and most widespread set-up) of a sensor kinase and a response regulator (RR) (Mascher *et al.*, 2006), and (iii) ECF (extracytoplasmic function) σ factors, alternative polymerase σ subunits, usually regulated by another accessory protein, the anti-σ factor (Staroń *et al.*, 2009). All of these three signal transduction systems mediate their cellular output mainly through differential gene expression, acting at the level of transcription initiation (Browning and Busby, 2004).

1.1.1. One-component signal transduction systems

In the simplest forms of bacterial signal transduction, two functions are fused in a single polypeptide chain (Ulrich *et al.*, 2005). Its sensor domain perceives a signal and then modulates the activity of an effector domain, which orchestrates the cellular response (Fig. 1.1). Input and output functions can be located in a single domain or can form two separate domains. The umbrella term that has been coined for this large group of phylogenetically

unrelated proteins that are comprised of these two domains only is “One-Component Systems”. Luke Ulrich and Igor Zhulin (University of Tennessee) thereby contrasted the already established term “two-component systems”, which will be addressed later (Fig. 1.1).

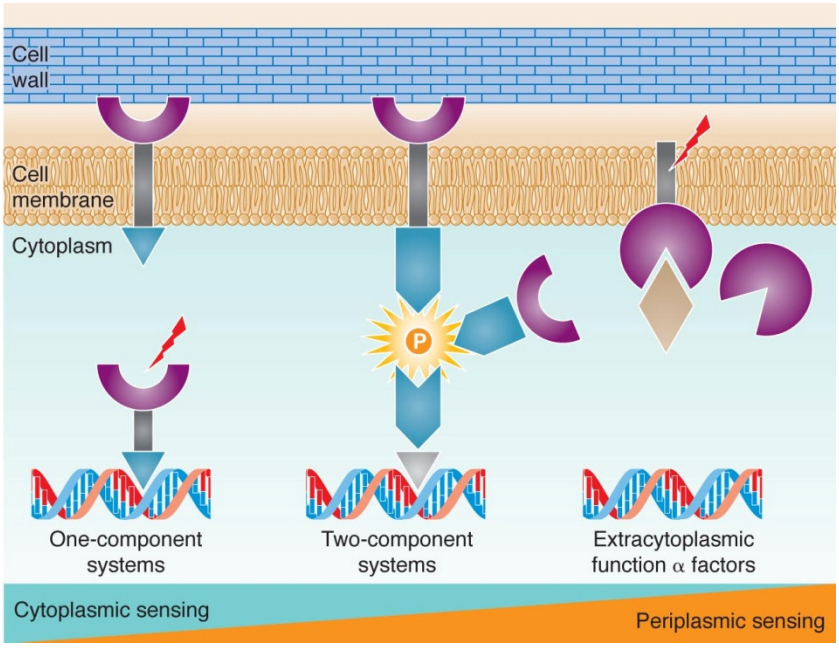

Figure 1.1. Concept of 1CS, 2CS, and ECF σ factors.

The examples from this group include the most well-known textbook classics of bacterial gene regulation, such as the *lac* repressor LacI of *Escherichia coli* (Wilson *et al.*, 2007). In response to the input signal of extracellular lactose, this protein releases its inhibiting grip on the *lac* operator, allowing expression of the *lac* operon, which encodes functions for using this sugar molecule (cellular output). LacI, like other bacterial signal transduction effectors, is a transcriptional regulator. In this and other cases, the cellular output consists of differential gene expression by specific genes. Very often, these regulators affect the activity of only a single transcript - here, the *lacZYA* operon - that encodes proteins devoted to a single purpose - in this case, use of lactose.

Other regulators control a larger number, in some cases even hundreds of target genes, as in global regulators, such as RelA or CRP in *E. coli*, which adjust the overall gene expression pattern in response to the second messengers ppGpp and cAMP, respectively (Green *et al.*, 2001; Wick and Egli, 2004). While many 1CS function as transcriptional regulators, others regulate a variety of other output domains (Ulrich *et al.*, 2005), including enzymes that make and break cyclic di-GMP, a second messenger that has recently gained a lot of attention (Hengge, 2009a).

The relatively simple and straightforward design of 1CS is ideally suited to provide a very direct connection between countless input and output devices. Not surprisingly, 1CS are an evolutionary success story without equal, and the primary, most widely distributed pillar of bacterial signal transduction (Table 1.1 (Ulrich *et al.*, 2005)). However, combining the input and the output domain on a single protein also has disadvantages, imposing restrictions on signal transduction. While it is easily conceivable how 1CS provide an ideal means for virtually any type of intracellular signal, it is much harder to imagine how a single protein can facilitate an efficient connecting between an extracellular signal and an intracellular response. While some examples of membrane-anchored extracellular-sensing 1CS do exist (i.e. ToxR, CadC, or BcrR), most 1CS are indeed soluble proteins that respond to intracellular cues (Gebhard *et al.*, 2009; Krukonis *et al.*, 2000; Tetsch and Jung, 2009).

Table 1.1. Distribution of 1CS, 2CS and ECF σ factors in selected organisms.

Organism	Phylum	1CS	2CS	ECF
Solibacter usitatus	Acidobacteria	296	100	69
Mycobacterium tuberculosis	Actinobacteria	157	13	10
Streptomyces coelicolor	Actinobacteria	666	87	50
Bacteroides thetaiotamicron	Bacteroidetes	108	30	46
Anabaena variabilis	Cyanobacteria	140	74	1
Bacillus anthracis	Firmicutes	245	47	16
Bacillus subtilis	Firmicutes	198	33	7
Agrobacterium tumefaciens	Proteobacteria (α)	371	52	12
Bradyrhizobium japonicum	Proteobacteria (α)	462	90	17
Caulobacter crescentus	Proteobacteria (α)	154	42	15
Rhodobacter sphaeroides	Proteobacteria (α)	188	48	8
Escherichia coli	Proteobacteria (γ)	239	31	2
Pseudomonas aeruginosa	Proteobacteria (γ)	404	67	18
Salmonella enteric	Proteobacteria (γ)	231	31	1
Myxococcus xanthus	Proteobacteria (δ)	283	125	38
Sorangium cellulosum	Proteobacteria (δ)	625	116	84

1.1.2. Two-component signal transduction systems

In contrast, many extracellular stimuli that require transmembrane signalling involve the other two categories of bacterial signal transduction, 2CS and ECF σ factors (Fig. 1.1.). Remarkably, 1CS and 2CS often share their respective input and output domains, again emphasizing the modularity of bacterial signal transduction (Wuichet *et al.*, 2010). But in contrast to 1CS, these domains are located on two different proteins in the case of 2CS.

A typical 2CS consists of a histidine kinase (HK), which functions as a sensor protein, and an effector protein that is a RR (Hoch and Silhavy, 1995; Inouye and Dutta, 2003). After detecting a specific signal, the HK autophosphorylates and, in turn, activates its cognate RR. This effector protein typically is a transcriptional regulator that modulates expression

of its target genes. However, 2CS RR have recruited various other output domains, including RNA/protein binding domains or others with enzymatic activities such as the diguanylate cyclase and phosphodiesterase domains involved in cyclic di-GMP signalling (Galperin, 2006).

Many 2CS use simple and linear pathways from one sensor kinase to one RR to control their respective outputs. In contrast to 1CS that predominantly respond to intracellular cues, more than 50% of known sensor kinases respond to extracellular stimuli (Fig. 1.1.). However, the design of 2CS not only simplifies transmembrane signalling, but it also enables many modifications, including amplification and integration of different signals as well as branching of the pathway (Stock *et al.*, 2000). Integrating pathways, where more than one HK control the phosphorylation state of a single RR can be found e.g. in quorum sensing of *Vibrio cholerae* (Yildiz and Visick, 2009). Moreover, so called phosphorelays – pathways where the phosphate group is shuttled along a longer chain of proteins – are also based on 2CS building blocks (Zhang and Shi, 2005). Such phosphorylation cascades are involved in the control of complex differentiation processes, where many different signals integrate into a single output. Examples include the initiation of endospore formation in *B. subtilis* and cell cycle control in *C. crescentus* (Biondi *et al.*, 2006; Piggot and Hilbert, 2004).

Thus, 2CS is a very flexible signalling principle based on a highly modular design that is adapted to many cellular needs. Not surprisingly, it is widely distributed in bacteria, archaea, and some lower eukaryotes, with the number of 2CS per genome exceeding 100 in some cases (Table 1.1.).

1.1.3. ECF σ Factors

Alternative σ factors of the ECF family are another means by which bacteria direct differential gene expression in response to extracellular cues. In general, σ factors are essential components of RNA polymerase that determine promoter specificity and thereby rates of transcription initiation (Helmann and Chamberlin, 1988). In addition to the primary, or housekeeping, σ factors found in all bacteria, most genomes - especially in species with complex life styles - encode alternative σ factors (Helmann, 2002). These proteins redirect RNA polymerase to initiate transcription from alternative promoters after substituting for primary σ factors. ECF σ factors represent the largest and most diverse group of alternative σ factors in bacteria. This term was coined by Lonetto *et al.* 1994 after several lines of evidence indicated that a small group of loosely related proteins, which

regulate some aspects of the cell surface or transport processes, function as alternative σ factors (Lonetto *et al.*, 1994).

Although the 2CS and ECF σ factors are phylogenetically and biochemically unrelated, they share several mechanistic analogies. For instance, both require two proteins for signal transduction: a typically membrane-anchored sensor protein (HK or anti-σ factor) and a cytoplasmic transcriptional regulator (RR or σ factor, respectively). In both cases the corresponding genes are usually cotranscribed (Staroń *et al.*, 2009). However, 2CS and ECF σ factors differ in the way the two proteins communicate with each other and thereby the mechanism of signal transduction. For 2CS, signal transduction is mediated by intramolecular conformational changes, based on transient cycles of phosphorylation and dephosphorylation of both proteins in the presence or absence of a suitable trigger (Stock *et al.*, 1995). In contrast, the communication between anti-σ factor and σ-factor is normally based on stable protein-protein interactions in the absence of a stimulus. Thus, the anti-σ factor keeps its partner inactive by titrating it from the pool of freely available σ factors (Helmann, 1999).

Once a signal is perceived, the anti-σ factor is inactivated by one of a number of mechanisms, thereby releasing the ECF σ factor from anti-σ and activating expression of its target genes (Helmann, 2002). An alternative promoter, which only the corresponding ECF σ factor recognizes, precedes these genes. Because of the nature of this regulation, ECF-dependent signalling always upregulates its target genes, in contrast to 2CS, where the output varies and may entail positive and/or negative regulation.

In most cases, the operon encoding the σ/anti-σ pair is also positively autoregulated (i.e. at least partially under control of the ECF σ factor it encodes) (Helmann, 2002). Accordingly, both the σ/anti-σ pair and the target proteins are produced under inducing conditions. As long as the stimulus is still present in the environment, the anti-σ factor gets continuously inactivated and the cellular output remains high. Once the inducing conditions cease, the anti-σ factor no longer gets inactivated and hence ultimately inactivates its partner σ factor by protein-protein interaction. Thus, ECF-dependent gene expression is shut down again.

1.2. Bce-like detoxification modules

BceRS-like 2CS constitute part of antimicrobial peptide detoxification modules, which are wide-spread among Firmicutes bacteria (Dintner *et al.*, 2011). They are named after the paradigm example for this class of 2CS, BceRS from *B. subtilis*. A general and unique

feature of these systems is their composition, as they are composed of four subunits, all of which are indispensable for sensing and transducing of the signal (Fig. 1.2) (Bernard *et al.*, 2007; Rietkötter *et al.*, 2008). Two of these proteins constitute a 2CS, with the HK BceS and its cognate RR BceR. Another two form an ATP-binding cassette (ABC) transporter, consisting of a nucleotide-binding domain (NBD) subunit BceA and a membrane-spanning domain (MSD) subunit BceB.

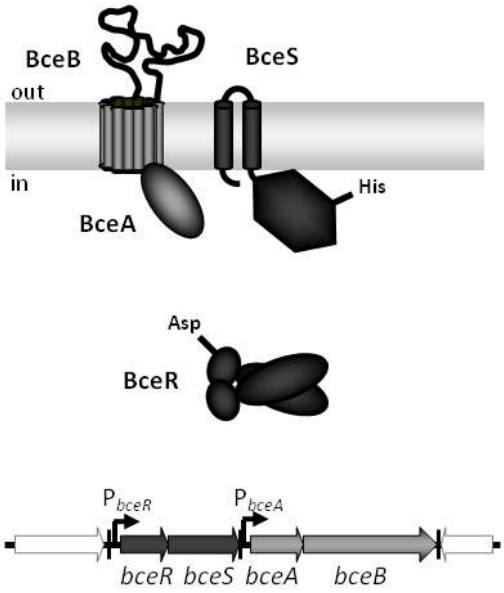

Figure 1.2. Schematic representation of the genes and proteins of the BceRSAB detoxification system in *B. subtilis*. Genes and proteins of the BceRSAB system are marked in dark gray (2CS) and medium gray (subunits of the ABC transporter). Genes flanking the *bce* operon are white. Promoters are marked with bent arrows, terminators are represented by vertical bars.

BceS-like HK are EnvZ-like proteins that belong to a family of cell envelope stress sensors termed intramembrane-sensing HK (IM-HK) (Mascher, 2006). Similarly to other members of this class, they do not harbour any extracytoplasmic sensory domains. Instead, the transmembrane helices are connected by only a few extracellular amino acids. BceS-like proteins, in contrast to LiaS-like HK described later, lack any additional cytoplasmic domains for intramolecular signal transfer, like HAMP or PAS domain (Aravind and Ponting, 1997; Ponting and Aravind, 1997). Therefore, they are indeed minimalistic HK, with only a conserved DHp (dimerization histidine phosphotransfer) and catalytic ATPase domains. In response to a stimulus, they specifically phosphorylate the cognate BceR-like RR, which belongs to the OmpR family and contains a winged helix-turn-helix DNA-binding output domain.

BceAB-like ABC transporters are classified as Peptide-7-Exporter (Pep7E) in the Transport Classification Database (TCDB) (Saier *et al.*, 2009). The MSD contains ten TMRs and a large extracytoplasmic loop of approximately 200 amino acids between helices 7 and 8 (Fig. 1.2). This domain architecture is highly conserved among all BceB-like proteins (Dintner *et al.*, 2011). Interestingly, the ABC transporter has a dual role, as it is both a resistance determinant (output) and a part of the stimulus sensing complex (input of the system). Therefore, BceAB-like ABC transporters basically regulate their own

expression in response to their extracellular substrate by using a 2CS as mediator of signal transduction and gene regulation. In these unique modules, both subunits of the ABC transporter are absolutely required for stimulus sensing and subsequent activation of the kinase. Although the exact mechanism of both stimulus detection and antibiotic removal remains to be elucidated, some experimental data is slowly emerging and will be described in more detail in a later section (see "role of BceAB-like ABC transporters in sensing and detoxification").

1.2.1. Experimentally characterised Bce-like systems

A number of Bce-like systems have been characterised (Table 1.2), with the best studied example being the bacitracin resistance module BceRSAB from B. subtilis. BceRSAB was initially described as a part of the bacitracin stimulon in B. subtilis (hence the name: bacitracin efflux genes) and represent the most effective bacitracin resistance determinant in this organism (Mascher *et al.*, 2003; Ohki *et al.*, 2003a). It is also responsible for conferring resistance to other peptide antibiotics, such as mersacidin and actagardine (Staroń *et al.*, 2011). Moreover, it was reported to respond to a fungal defensin, plectasin (Schneider *et al.*, 2010).

In addition, *B. subtilis* harbours two other paralogous systems named PsdRSAB and YxdJKLM-YxeA. The PsdRSAB (peptide antibiotic sensing and detoxification) module is induced by a broad range of peptide antibiotics, including enduracidin, actagardine, gallidermin, nisin, and subtilin (Staroń *et al.*, 2011). It mediates resistance against most of its inducers, but not actagardine. The third system, YxdJKLM-YxeA is the least understood of the three *B. subtilis* systems. It is induced by an antimicrobial peptide LL-37 (Pietiäinen *et al.*, 2005), and YxdJK 2CS controls expression of *yxdLM-yxeA* and *dltABCDE* operons (Joseph *et al.*, 2004). Interestingly, the expression of the YxdLM ABC transporter is induced more strongly by LL-37 in a *dlt* mutant compared to the wild type (Hyyryläinen *et al.*, 2007). Another unique feature of this system is the fact that the encoding gene locus harbours five genes instead of four. The additional gene encodes a small (115 aa) protein that is predicted to be a transmembrane protein with a DUF1093 domain. Its function is not known.

A close relative of *B. subtilis*, the bacitracin producer *B. licheniformis* also harbours three Bce-like systems. Its genes encode one homologue of BceRS, two homologues of YxdJK, but no direct homologue of PsdRS (Dintner *et al.*, 2011). YtsABCD, similarly to its direct homologue BceRSAB, was also reported to be induced by bacitracin (Wecke *et al.*, 2006).

Two other systems (encoded by *yxdKLM-yxeA* and *Bli04272-yxdJK2L2M2*) are highly similar to the YxdJKLM-YxeA system of *B. subtilis*, with both systems encoding a small transmembrane protein harbouring a DUF1093 domain (in *B. licheniformis yxeA* and *Bli04272*). The second system (*Bli04272-yxdJK2L2M2*) was shown to be induced by bacitracin, while the first one (*yxdKLM-yxeA*) is not. But this lack of induction may be due to a mutation in the gene encoding the YxdJ-like RR.

Staphylococcus aureus has two complete 2CS/ABC transporter modules, termed GraRS-VraFG and BraRSAB (other names used in the literature for these systems are listed in Table 1.2). Additionally, it harbours a third Pep7E-type ABC transporter, VraDE, which is not directly associated with a 2CS. GraRS was initially described as conferring resistance against glycopeptide antibiotics, specifically vancomycin (Cui *et al.*, 2005), hence the name: glycopeptide resistance associated. It is located in an operon directly upstream of genes encoding the VraFG transporter and regulates their expression (Li *et al.*, 2007a). The VraFG transporter is both necessary for sensing the peptide compounds, such as colistin or indolicidin, and for mediating resistance against them (Li *et al.*, 2007a; Meehl *et al.*, 2007). Moreover, GraRS 2CS plays an important role in virulence of *S. aureus* (Kraus *et al.*, 2008; Li *et al.*, 2007a). It also influences autolytic activity (Meehl *et al.*, 2007) and biofilm formation (Shanks *et al.*, 2008).

Expression of the second Bce-like system in *S. aureus*, BraRSAB, is induced by bacitracin and nisin (Hiron *et al.*, 2011). BraRS 2CS controls the expression of two ABC transporters, BraAB and VraDE. Interestingly, the BraAB ABC transporter is responsible solely for sensing of the signal, whereas VraDE confers resistance against the inducing compounds, bacitracin and nisin (Hiron *et al.*, 2011). This complex interplay between the two transporters and the BraRS 2CS will be described in detail in a later section (see Chapter 8).

S. epidermidis, similarly to *S. aureus*, also harbours two Bce-like detoxification systems and one additional ABC transporter. However, experimental data from this organism is restricted only to the GraRS system (here named Aps). It is induced by different antimicrobial peptides, including nisin, LL-37 and hBD3 (see Table 1.2) (Li *et al.*, 2007b). It also confers resistance against hBD3 and contributes to survival after uptake by human neutrophils (Cheung *et al.*, 2010).

Table 1.2. Bce-like detoxification modules in Firmicutes bacteria.

Organism	Resistance module		Inducing substrates	Phenotype of ABC mutant	Additional regulon	References
	2CS	ABC				
B. subtilis	BceRS	BceAB	bacitracin, plectasin, actagardine, mersacidin	sensitivity to bacitracin, actagardine, mersacidin	none	(Bernard *et al.*, 2007; Ohki *et al.*, 2003a; Rietkötter *et al.*, 2008; Staroń *et al.*, 2011)
B. subtilis	PsdRS/ YvcPQ	PsdAB/ YvcRS	enduracidin, actagardine, gallidermin, nisin, subtilin	sensitivity to enduracidin, gallidermin, nisin, subtilin	none	(Rukmana *et al.*, 2009; Staroń *et al.*, 2011)
B. subtilis	YxdJK	YxdLM	LL-37	not determined	*dltABCDE*, *ywaA*	(Joseph *et al.*, 2004; Pietiäinen *et al.*, 2005)
B. licheniformis	YtsAB	YtsCD	Bacitracin	not determined	unknown	(Wecke *et al.*, 2006)
B. licheniformis	YxdJK2	YxdL2M2	Bacitracin	not determined	unknown	(Wecke *et al.*, 2006)
L. lactis	LlrG-KinG	YsaBC	Nisin	sensitivity to nisin	unknown	(Kramer *et al.*, 2006)
L. monocytogenes	VirRS	AnrAB		defects in cell entry, adhesion, virulence, sensitivity to bacitracin, nisin and β-lactams	*dltABCD, mprF*	(Camejo *et al.*, 2009; Collins *et al.*, 2010; Mandin *et al.*, 2005)
S. aureus	GraRS/ ApsRS	VraFG	indolicidin, melittin, nisin, LL-37, mersacidin, vancomycin	increased autolytic activity, reduced infectivity, altered cell surface charge, no biofilm formation after stimulation with citrate, sensitivity to hBD3, nisin, indolicidin, LL-37, vancomycin, polimyxin B, mersacidin, Pep5, LP9, gallidermin, more sensitive to killing by human neutrophil granulocytes	*dltABCD, mprF, vraDE*[a]	(Herbert *et al.*, 2007; Kraus *et al.*, 2008; Li *et al.*, 2007a; Meehl *et al.*, 2007; Sass and Bierbaum, 2009; Sass *et al.*, 2008; Shanks *et al.*, 2008)
S. aureus	BraRS/ NsaRS/ BceRS	BraAB/ NsaAB/ BceAB	hBD3, bacitracin, nisin	sensitivity to nisin and bacitracin	*vraDE*	(Blake *et al.*, 2011; Hiron *et al.*, 2011; Sass *et al.*, 2008; Yoshida *et al.*, 2011)

S. aureus	-	VraDE	mersacidin, hBD3, LL-37, temporin, ovispirin, dermaseptin, bacitracin, teicoplanin	sensitivity to nisin, bacitracin, daptomycin[a], hBD3[a] and Pep5[a]	not applicable	(Hiron *et al.*, 2011; Pietiäinen *et al.*, 2009; Sass *et al.*, 2008)
S. epidermidis	ApsRS	VraFG	hBD3, tachyplesin, nisin, LL-37, magainin, histatin, brevinin	sensitivity to hBD3, lysis in the stationary phase, sensitivity to killing by neutrophils	*dltABCD, mprF, vraDE*	(Cheung *et al.*, 2010; Li *et al.*, 2007b)
S. mutans	MbrCD/ BceRS	MbrAB/ BceAB	bacitracin, defensins	sensitivity to bacitracin and defensins	SMU.862-4, .302 and .1856c	(Kitagawa *et al.*, 2011; Ouyang *et al.*, 2010; Tsuda *et al.*, 2002)
S. pneumoniae	-	Spr0812-3/SP0912-3	nisin and bacitracin	sensitivity to bacitracin, nisin, gramicidin, lincomycin	not applicable	(Becker *et al.*, 2009; Majchrzykiewicz *et al.*, 2010)

[a] contradicting results from different *S. aureus* strains (see text for details)

Streptococcus mutans harbours only one BceRS-like system. MbrABRS was initially characterised as bacitracin resistance determinant in this organism (Tsuda *et al.*, 2002) and therefore termed *mbr* for mutans bacitracin resistance. However, it was recently shown that it also confers resistance to human defensins (Ouyang *et al.*, 2010). MbrRS controls the expression of a number of genes, mostly encoding membrane proteins of unknown functions (Kitagawa *et al.*, 2011; Ouyang *et al.*, 2010). In the closely related species *S. pneumoniae*, the Pep7E-ABC transporter Spr0812-3 is involved in resistance to bacitracin, nisin, gramicidin and lincomycin (Becker *et al.*, 2009; Majchrzykiewicz *et al.*, 2010). However, in this organism no BceRS-like 2CS is located in the neighbourhood of the transporter.

In *Lactococcus lactis* the experimental data is also limited only to the ABC transporter YsaBC, while the role of BceRS-like 2CS has not been elucidated. The YsaBC transporter was initially discovered in a nisin resistant mutant of *L. lactis*. Its expression is induced by nisin and, when overexpressed, it mediates resistance to this lantibiotic (Kramer *et al.*, 2006). Its genes are located upstream of genes encoding a BceRS-like 2CS (*llrG kinG*), separated by a gene encoding a VanZ-like protein. Interestingly, a VanZ protein from *Enterococcus faecium* confers resistance to teicoplanin, a peptide antibiotic interfering with lipid II cycle (Arthur *et al.*, 1995); its role in *L. lactis* is unknown.

In *Listeria monocytogenes* AnrAB transporter mediates resistance to different peptide antibiotics, including bacitracin, gallidermin, and nisin (Collins *et al.*, 2010). It is not located in the direct neighbourhood of any BceRS-like 2CS, but nisin-dependent induction of *anrAB* depends on the VirRS 2CS, located at another locus in the genome (Mandin *et al.*, 2005). VirRS 2CS regulates an array of processes, including adhesion and entry into eukaryotic cells (Mandin *et al.*, 2005). The set of genes under control of VirRS includes the *dltABCD* operon, *mprF* gene and the above mentioned ABC transporter *anrAB* (Camejo et al., 2009) (see "extended Bce regulons and functions").

1.2.2. Common themes in Bce-like systems

1.2.2.1. Stimuli sensed by Bce-like systems. A common trait of all described BceRS-like 2CS is the induction by a comparable set of compounds (Fig. 1.3). To our knowledge, inducers of BceRS-like systems are exclusively short antimicrobial peptides, both of bacterial origin and produced by insects or animals. Although they differ in structure and modifications, they all share the cationic and amphipathic properties and often have similar

modes of action (including membrane damage and blockage of cell wall synthesis) (Yeaman and Yount, 2007). Inducers include peptides like bacitracin or gramicidin, often glycosylated (vancomycin, teicoplanin), with fatty acid modifications (enduracidin, polimyxin B, daptomycin) or lanthionine bridges (lantibiotics including nisin, gallidermin, actagardine, mersacidin, subtilin, Pep5). Other inducers include peptides that play a role in the immune response of higher organisms like cathelicidines (LL-37, indolicidin, ovispirin), and defensins (hBD3, brevinin, plectasin, dermaseptin). Although all inducing molecules belong to the group of small antimicrobial peptides, the exact nature of the signal is puzzling, as the sensors are able to bind compounds of very different structures, but also to distinguish between very similar ones. For example, the Psd system of *B. subtilis* reacts to both nisin, which is a lantibiotic, and enduracidin, a cyclic lipopeptide. However, another lipopeptide, ramoplanin, which is structurally very similar to enduracidin, does not trigger induction of the Psd system (Staroń *et al.*, 2011).

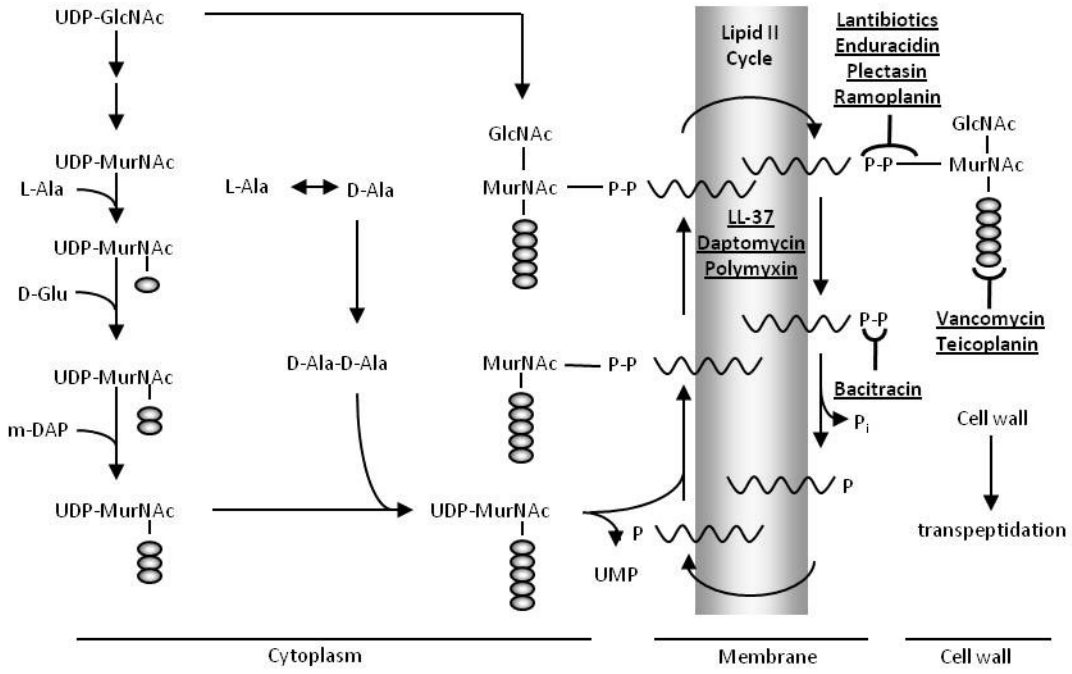

Figure 1.3. Cell wall biosynthesis of Gram-positive bacteria and its inhibition by antibiotics that induce Bce-like systems. Crucial steps in cell wall biosynthesis are schematically indicated, together with their cellular localization. GlcNAc, N-acetyl-glucosamine; MurNAc, N-acetyl-muramic acid. Amino acids are depicted as small grey circles, undecaprenol by the waved line. Antibiotics which induce Bce-like systems are underlined. Lantibiotics is used as a collective term for actagardine, gallidermin, mersacidin, nisin, and subtilin.

1.2.2.2. Role of the BceAB-like ABC transporters in sensing and detoxification. In a classical 2CS, a HK senses the stimulus, often via its extracytoplasmic domain. For

example, PhoQ from *S. enterica* harbours a PDC domain (Zhang and Hendrickson, 2010), which plays an important role in sensing antimicrobial peptides (Bader *et al.*, 2005). However, as mentioned before, BceS-like HK have only a short extracytoplasmic loop that is enough to connect two transmembrane helices, but it is too short to harbour a substrate binding domain. And indeed, for some of the Bce-like systems, including BceRSAB itself (Bernard *et al.*, 2007; Rietkötter *et al.*, 2008), but also for PsdRSAB (Staroń and Mascher, unpublished), MbrABCD (Ouyang *et al.*, 2010), GraRS-VraFG (Meehl *et al.*, 2007), and BraRSAB (Hiron *et al.*, 2011), it has been shown that the kinases require the presence of the ABC transporter for signal transduction. Based on these findings, it can be assumed that this is a general feature of these detoxification systems and that all BceS-like HK are unable to sense the stimulus alone, but instead require the ABC transporter for induction.

While being responsible for signalling, the transporter proteins are also resistance determinants. The involvement of the ABC transporter in resistance against antimicrobial compounds has been shown for many systems including BceAB and PsdAB from *B. subtilis* (Ohki *et al.*, 2003a; Staroń *et al.*, 2011), AnrAB from *L. monocytogenes* (Collins *et al.*, 2010), VraFG, BraAB, and VraDE from *S. aureus* (Hiron *et al.*, 2011; Li *et al.*, 2007a). Although much speculation on the mechanism of signalling and resistance can be found in the literature, there are no conclusive results so far. The available evidence shows that the actual transport of the substrate is important for both sensing of the compound and conferring resistance to it. Mutations in either Walker A or Walker B motifs, leading to defects in ATP binding and hydrolysis, render the BceRSAB (Bernard *et al.*, 2007; Rietkötter *et al.*, 2008), PsdRSAB (Staroń and Mascher, unpublished), BraRSAB and VraDE (Hiron *et al.*, 2011) systems dysfunctional. Moreover, a deletion of the large extracytoplasmic loop, that is a conserved feature of all BceB-like permeases, also abolishes both signalling and resistance (Bernard *et al.*, 2007; Rietkötter *et al.*, 2008). It was suggested that the large extracytoplasmic loop is the actual binding site for the antibiotic compounds (Hiron *et al.*, 2011; Rietkötter *et al.*, 2008). Indeed, in *S. aureus* an exchange of the loop of VraG (responsible for conferring resistance to colistin) for that of VraE (conferring resistance to bacitracin) resulted in a chimeric transporter with a changed specificity that was able to restore bacitracin resistance (Hiron *et al.*, 2011).

1.2.2.3. Signalling versus resistance in Bce-like detoxification modules. BceRS-like systems from *B. subtilis* and *S. aureus* have been extensively characterised with respect to the inducer spectrum and the compounds they confer resistance against. Interestingly, these

two spectra of compounds do not fully overlap, and not every antibiotic that induces a Bce-like system can also be efficiently detoxified of by the same module.

The Psd system from *B. subtilis* is induced by actagardine, nisin, enduracidin and subtilin, but mediates resistance only against the last three compounds (Staroń *et al.*, 2011). Actagardine triggers a strong response, but the elevated expression of the ABC transporter does not lead to any change in the cell's resistance against it. Another very interesting example can be found in *S. aureus*, where the ABC transporter BraAB is responsible solely for signalling, and does not mediate resistance against inducing compounds (nisin and bacitracin) (Hiron *et al.*, 2011). Therefore, it seems that signalling and resistance determination are two distinct functions that can be separated. As mentioned before, not the sole presence of an ABC transporter, but also the ability to transport is important for signalling, as mutations abolishing ATP-hydrolysis and therefore transport also impair signal transduction. One possible explanation of this phenomenon suggests that the rate of AMP transport is not enough to confer resistance against it, but enough to trigger the signalling cascade, the transport rate being the primary stimulus (Gebhard and Mascher, 2011). However, this hypothesis remains to be proven experimentally.

1.2.2.4. Mechanism of stimulus perception in Bce-like systems: different possibilities.

Despite the fast growing amount of experimental data on Bce-like systems, the exact mechanism of stimulus perception in BceRS-like 2CS is still unclear. Since both functional subunits of the ABC transporter, and hence transport, are necessary for signal transduction, three mechanisms of stimulus perception can be proposed (Fig. 1.4). The scenario favoured in the most recent studies on the subject proposes a direct contact between the HK and the ABC transporter as a necessary step in the signalling cascade (Fig. 1.4B). Upon the contact with the inducing peptide, possibly through the extracytoplasmic loop, the transporter changes its conformation, which in turn promotes the interaction with the HK and its subsequent activation. In support of this hypothesis, recent bioinformatic analysis demonstrated a coevolution of BceS-like HK and the cognate BceB-like permeases (Dintner *et al.*, 2011). However, a coevolution of two proteins can be due to other reasons besides a direct interaction, namely participation in the same process or interaction with the same ligand (Pazos and Valencia, 2008). Therefore, this hypothesis needs to be experimentally validated.

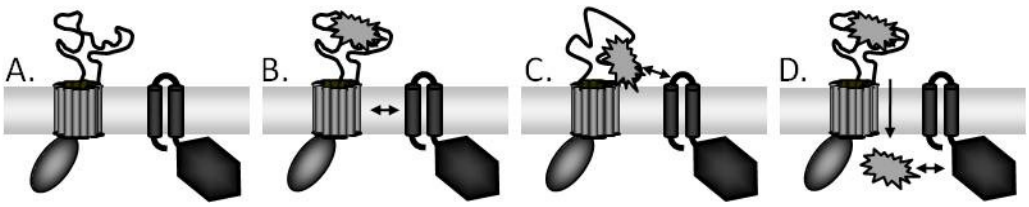

Figure 1.4. Proposed mechanisms of stimulus perception by Bce-like systems. (A) System in uninduced state. (B) Upon the contact with inducing peptide (here: through the extracytoplasmic loop) the transporter changes its conformation and activates the HK. (C) ABC transporter is presenting the inducing peptide to the HK, and the inducing compound is detected in a complex with ABC transporter. (D) BceAB transports the peptide to the cytoplasm where it is detected by BceS-like kinases.

It cannot be ruled out that the ABC transporter is binding the inducing peptide and presenting it to the HK, so that the inducing compound can be detected only as a complex with ABC transporter (Fig. 1.4C). This detection could possibly happen through the short extracytoplasmic loop of the kinase. This hypothesis is supported by evidence obtained in studies on the GraS in *S. aureus* and ApsS in *S. epidermidis*. ApsS responds to hBD3, a compound that does not induce a GraS-dependent expression. An exchange of loops between these two kinases renders the *S. aureus* GraS responsive to hBD3 (Li *et al.*, 2007a). Moreover, antibodies against the short (9 residues) extracellular loop of the *S. epidermidis* HK GraS prevent upregulation of target genes after induction by hDB3 (Li *et al.*, 2007b). The above mentioned results suggest that the extracellular loop of GraS/ApsS is responsible for signal recognition. However, the antibody, being a big molecule, may as well interfere with interaction of the ABC transporter with the HK. Moreover, it is not clear why this manner of stimulus perception should be energy dependent.

Lastly, it is also possible that the ABC transporter is an importer which transports the inducing compounds into the cytoplasm, where the stimulus is detected by the kinase in the cytoplasm (Fig. 1.4D). This hypothesis seems unlikely in the light of the above mentioned data, but cannot be completely excluded at present.

1.2.2.5. Additional BceR-target genes. A few operons for Bce-like detoxification modules are associated with open reading frames encoding additional proteins of unknown function. These genes are located either next to the ABC transporter (*yxeA* in *B. subtilis* and *B. licheniformis*, *spr0811a* in *S. pneumoniae*) or the 2CS (*apsX* and SA2419 in *S. aureus*, *bli04272* in *B. licheniformis*). Usually these are small proteins, ranging from 36 aa (Spr0811a) to 123 aa (Bli04272); only ApsX is significantly larger (307 aa). Although they do not show significant sequence conservation, most of them are transmembrane proteins, and some harbour a DUF1093 domain. Their function in signal perception or

detoxification is not yet known. The only experimental data comes from the ApsRS system of *S. epidermidis* (homologous to GraRS in *S. aureus*), where the deletion of *apsX* leads to inactivation of the whole system (Li *et al.*, 2007b). However, it is not clear whether this observation can be extrapolated to other systems, as ApsX differs significantly from other detoxification module-associated proteins.

1.2.2.6. BceR-like binding sites. BceR-like RR belong to the OmpR/PhoB subfamily, the largest subfamily of RR. This subfamily accounts for half of RR possessing a DNA binding domain (Galperin, 2006), which in this case consists of a winged helix-turn-helix motif. A study of RR binding sites in low-GC Gram positive bacteria predicted an operator for BceR-like proteins to include an inverted repeat (ACA-N_4-TGT) (de Been *et al.*, 2008). A following in-depth study identified motifs recognised by BceR-like RR, and the overall consensus was found to be TNACA-N_4-TGTAA, with an AT-rich spacer (Dintner *et al.*, 2011). All binding sites that were experimentally identified support this consensus, even if the exact position in the identified region and nature of the repeat (inverted or direct) may not have been correctly recognised at the time. Binding sites were identified by promoter deletion analyses, EMSA and primer extension assays for Bce and Yxd systems in *B. subtilis* (Joseph *et al.*, 2004; Ohki *et al.*, 2003a), Vir in *L. monocytogenes* (Mandin *et al.*, 2005), Bce in *S. mutans* (Ouyang *et al.*, 2010), and Bra and Vra systems in *S. aureus* (Hiron *et al.*, 2011).

The family of BceR-like RR shows a high level of sequence conservation, and the same is true for the promoters that these RR recognise. As there is often more than one BceR-like RR in one organism, an interesting question of promoter recognition specificity arises. Recent work from groups of Ann Stock and Michael Laub showed that there is a strong preference for very specific homodimerization of HK (Ashenberg *et al.*, 2011), phosphotransfer to only cognate RR (Skerker *et al.*, 2005), and also a preference for RR homodimerization (Gao *et al.*, 2008). Most probably, there is also a preference for specific promoter recognition sites, especially as no cross-talk on the promoter level has been observed among Bce-like systems. However, it is not yet clear how the RR can recognise only the cognate binding site, given the high similarity of both RR and the binding sites themselves.

1.2.2.7. Extended Bce regulons and functions. Some of the BceRS-like systems control not only the expression of one (or more) ABC transporters, but also the expression of

additional genes, often including the *dltABCD* operon and the *mprF* gene. The *dlt* operon is reported to be regulated by YxdJK system of *B. subtilis* (Joseph *et al.*, 2004), VirRS of *L. monocytogenes* (Camejo et al., 2009), ApsRS of *S. epidermidis* (Li *et al.*, 2007b), GraRS of *S. aureus* (Li et al., 2007a). The *mprF* gene is regulated by the same systems with the exception of YxdJK of *B. subtilis*. Products of both *dlt* and *mprF* genes are involved in changing the cell surface charge by modifying cell wall and cytoplasmic membrane components (Peschel, 2002).

Teichoic acids (TAs) are linear polyglycerol phosphate or polyribitol phosphate chains (Neuhaus and Baddiley, 2003), which are either covalently linked to peptidoglycan via a phosphodiester linkage (wall TAs) or linked to the cytoplasmic membrane via a terminal glycolipid moiety (lipoteichoic acids) (Neuhaus and Baddiley, 2003). TAs are the major source of the negative net charge of the Gram-positive envelope due to intercalated PO_4-groups (Swoboda *et al.*, 2010). One of the modifications lowering the negative charge of TAs is D-alanylation (i.e. replacing the PO_4-groups by D-alanine), which is catalysed by products of genes encoded in the *dltABCD* operon (Peschel, 2002). Mutants in the *dlt* genes are pleiotropic, with phenotypes including increased autolysis (Cao and Helmann, 2002; Nakao *et al.*, 2000; Steen *et al.*, 2005), increased sensitivity to cationic antimicrobial peptides (CAMPs) (Peschel *et al.*, 1999), diminished adherence to phagocytic and non-phagocytic cells (Abachin *et al.*, 2002), and loss of acid tolerance (Boyd *et al.*, 2000). Consequently, many of these phenotypes have been observed in mutants of Bce-like systems, e.g. in *S. aureus* (Li *et al.*, 2007a; Meehl *et al.*, 2007).

Another source of the negative cell envelope net charge are phospholipids. One of the most common phospholipids found in bacterial membranes is the negatively charged phosphatidylglycerol (Ernst and Peschel, 2011). In order to lower the negative charge, the glycerol moiety of phosphatidylglycerol is modified by attaching lysine or alanine groups to it, thereby introducing positive charges into the membrane. These modifications are performed by the MprF protein. Mutants in *mprF* are more sensitive to various antimicrobial peptides, such as nisin, tachyplesin (Peschel *et al.*, 2001) or hBD3 (Nishi *et al.*, 2004) in *S. aureus* or LL-37 in *B. anthracis* (Samant *et al.*, 2009).

Both D-alanylation of TAs and lysylination of phospholipids lower the negative cell surface charge (Peschel, 2002). CAMPs, which by definition carry an overall positive charge, have strong affinity towards negatively charged bacterial membranes. Therefore, introduction of positive charges into the envelope mediates resistance against CAMPs. While the elevated expression of ABC transporters confers a specific resistance against a

small number of compounds, induction of *dlt* and *mprF* genes mediates more general resistance against CAMPs.

1.3. Aims of this work

This thesis aims to investigate the common principles governing two of the subgroups of signalling proteins, the ECF σ factors, and the Bce-like 2CS. To this end, both bioinformatic and experimental methods will be employed.

One aim of this thesis is to analyse and classify ECF σ factors. While there are bioinformatic analyses available for both 1CS and 2CS, there is no comprehensive analysis of ECF σ factors. With extensive amount of genomic data available, we postulate that it is possible to (i) classify ECF σ factors based on the sequence similarity and genomic context conservation, (ii) make predictions of function for novel ECF σ factors based on experimentally investigated examples and/or conserved neighbouring genes, (iii) identify and classify anti-σ factors, and (iv) identify ECF-dependent promoters. The possibilities of the classification should then be tested in detail in the in-depth analysis of one of the interesting novel ECF σ factor groups.

The second aim of this thesis is to perform a bioinformatic analysis for Bce-like 2CS. This analysis should demonstrate whether there is a evolutionary correlation between the permease (BceB-like) and kinase (BceS-like) components of these detoxification modules. This would suggest the postulated, but never experimentally addressed direct interaction between the ABC transporters and the 2CS. This analysis should also enable predictions for novel Bce-like detoxification modules, including potential interactions between kinases and non-cognate permeases, as well as orphan ABC transporters. Promoters recognized by BceR-like RR should also be analysed.

The third aim of this thesis is the investigation of Bce-like 2CS in *B. subtilis*. The three detoxification systems should be analysed in regard to inducer spectrum, both *in vivo* and *in silico*, the latter based on meta-analysis of all previously published stress response microarray data sets. Moreover, the regulation of BceR- and PsdR-dependent promoters should be analysed in detail, with a special focus on specificity determinants in both DNA sequences. Furthermore, the specificity determinants in the 2CS themselves should be experimentally analysed.

CHAPTER 2

The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family

This chapter has been adapted from:

A. Staroń, H. J. Sofia, S. Dietrich, L. E. Ulrich, H. Liesegang, T. Mascher

Mol. Microbiol. (2009) 74(3):557-581

Chapter 2

The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family

The ability of a bacterial cell to monitor and adaptively respond to its environment is crucial for survival. After 1CS and 2CS, ECF σ factors – the largest group of alternative σ factors – represent the third fundamental mechanism of bacterial signal transduction, with about six such regulators on average per bacterial genome. Together with their cognate anti-σ factors, they represent a highly modular design that primarily facilitates transmembrane signal transduction. A comprehensive analysis of the ECF σ factor protein family identified more than 40 distinct major groups of ECF σ factors. The functional relevance of this classification is supported by the sequence similarity and domain architecture of cognate anti-σ factors, genomic context conservation, and potential target promoter motifs. Moreover, this phylogenetic analysis revealed unique features indicating novel mechanisms of ECF-mediated signal transduction. This classification, together with the web tool ECF*finder* and the information stored in the Microbial Signal Transduction (MiST) database, provides a comprehensive resource for the analysis of ECF σ factor-dependent gene regulation.

2.1. Introduction

A hallmark feature of any living cell is its ability to sense and communicate with its environment. For multicellular organisms or microbial communities, such as biofilms, this can be neighbouring cells. For unicellular and free-living organisms, this might be their natural habitat, which they share with other cells of countless species in a constant struggle to survive. The ability to perceive and process numerous environmental parameters enables the cell to adapt its metabolism and behaviour to thrive and prosper in ever-changing habitats, such as the soil, the phyllosphere or aquatic ecosystems.

The genomic era has provided ample evidence that even simple prokaryotes are equipped with large numbers of regulatory devices to facilitate signal transduction in response to environmental cues (Galperin, 2005; Ulrich *et al.*, 2005). The best-known types of such bacterial signaling systems include 1CS and 2CS. Members of the phylogenetically diverse group of 1CS possess input and output domains fused on a single polypeptide chain (Ulrich

et al., 2005), while these domains are located on two different proteins in the case of 2CS. The latter consist of a sensor HK and a cognate RR, which together orchestrate His–Asp phosphorelays (Mascher *et al.*, 2006; Stock *et al.*, 2000). Regardless of the mode of signal transduction, the predominant output response is regulating gene expression, usually at the level of transcription initiation (Browning and Busby, 2004).

Another means by which bacteria can direct differential gene expression at the level of transcription initiation is by using alternative σ factors that are only activated in the presence of suitable triggers. Bacterial σ factors are essential components of the RNA polymerase and determine the promoter specificity (Helmann and Chamberlin, 1988). All bacteria harbour one primary (or housekeeping) σ factor that is responsible for the basal expression level of most genes. Moreover, most bacterial genomes – especially in species with complex lifestyles – encode alternative σ factors, which can redirect RNA polymerases to initiate transcription from alternative promoters after substituting the primary σ factor (Helmann and Chamberlin, 1988). In the absence of a stimulus, most alternative σ factors are kept inactive by a cognate anti-σ factor through direct protein–protein interaction (Brown and Hughes, 1995; Helmann, 1999).

There are two major (and seemingly unrelated) families of bacterial σ factors, the σ^{70} and σ^{54} family (Gross *et al.*, 1998; Gruber and Gross, 2003; Helmann and Chamberlin, 1988). The vast majority of σ factors belong to the σ^{70} family. Based on sequence similarity and protein domain architecture, several groups (or subfamilies) have been identified within this protein family, and these frequently correlate with distinct functions. Essential primary σ factors (group 1) and their non-essential paralogues (group 2), such as *E. coli* σ^{D} and σ^{S}, respectively, contain all four distinct domains typical for σ^{70} proteins (designated regions σ^{1} through σ^{4}) as well as a non-conserved region (NCR) adjacent to region σ^{2} (Fig. 2.1A) (Gruber and Gross, 2003). Group 3 σ^{70} proteins include flagellar, heat shock and sporulation σ factors. These proteins lack both the NCRand region σ^{1}. The largest and most diverse subfamily of σ^{70} proteins is group 4, which is also known as the ECF σ family (Butcher *et al.*, 2008; Helmann, 2002). In their seminal paper, Lonetto *et al.* (1994) coined this term, after several lines of evidence indicated that a small group of loosely related proteins, which regulate some aspects of the cell surface or transport processes, function as alternative σ factors (Lonetto *et al.*, 1994). These small proteins contain only regions σ^{2} and σ^{4}, which are required for both RNA polymerase interaction and recognition of the bipartite sequence motif that forms a typical bacterial promoter (Fig. 2.1B). In contrast to the two classical consensus promoter patterns of the primary σ factors, the -35

('TTGACA') and -10 ('TATAAT') regions (Helmann, 1995; Travers, 1987), many ECF-dependent promoters are characterized by a highly conserved 'AAC' motif in their -35 region and a clustering of 'CGT' tri-nucleotides in the -10 region (Fig. 2.1) (Helmann, 2002; Lane and Darst, 2006).

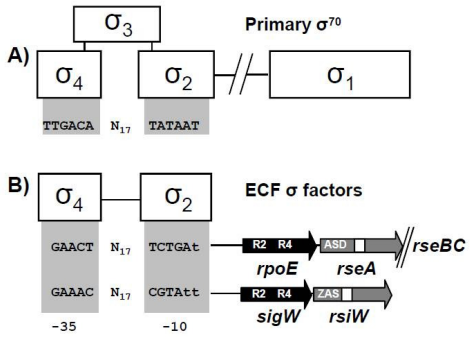

Figure 2.1. Domain architecture and defining features of ECF σ factors. Domain architecture and target promoters of (A) primary σ factors, (B) ECF σ factors are shown. R2 and R4 encode regions σ_2 and σ_4 respectively. Abbreviations: ASD, Anti-Sigma Domain; ZAS, Zinc-binding anti-σ domain. The ECF σ factors are represented by the promoter motif and operons encoding *E. coli* σ^E and *B. subtilis* σ^W.

The best-understood archetypes of ECF σ factors include *E. coli* σ^E and FecI, *B. subtilis* σ^W, *R. sphaeroides* σ^E and *S. coelicolor* σ^R (Ades, 2004; Alba and Gross, 2004; Braun and Mahren, 2005; Campbell *et al.*, 2007; Dufour *et al.*, 2008; Helmann, 2006; Li *et al.*, 2002). These proteins share several common features with respect to the mechanism of ECF-mediated signal transduction and gene regulation. First, ECF σ factors usually autoregulate their own expression (positive feedback loop; FecI-like σ factors are an exception). Second, ECF genes are usually coexpressed with gene(s) that encode cognate anti-σ factor(s). Third, in the absence of a stimulus, the anti-σ factor tightly binds the σ factor, thereby keeping it inactive. Fourth, upon receiving a proper signal, the anti-σ factor is inactivated either by degradation through a cascade of regulated proteolytic steps (σ^W and *E. coli* σ^E), or by conformational changes, as is observed for the redox-sensitive anti-σ switches of *S. coelicolor* σ^R and *R. sphaeroides* σ^E (Ades, 2008; Brooks and Buchanan, 2008; Campbell *et al.*, 2008). Both mechanisms result in the release and consequent activation of the ECF σ factor, which can then redirect gene expression to its target promoters after recruitment by the RNA polymerase core enzyme.

Despite detailed mechanistic insight into the function of these ECF σ factors, the overall knowledge of this fundamental mode of bacterial signal transduction is still very sparse, when compared with the wealth of information on 1CS and 2CS. While in-depth phylogenetic studies of the latter two – based on sequence conservation and domain architecture – resulted in their classification (Galperin, 2006; Grebe and Stock, 1999; Kim and Forst, 2001; Ulrich *et al.*, 2005), no such analysis exists for ECF σ factors.

The classification presented in this article is the result of a comprehensive comparative genomic analysis of the ECF σ factor protein family. ECF σ factors are ubiquitously distributed within and restricted to bacteria, with about six ECF proteins on average per genome. The domain architecture of both σ factors and their putative anti-σ factors, together with the overall genomic context conservation, revealed significant combinatorial complexity and indicates novel mechanisms of ECF-dependent signal transduction. This classification also allows the identification of potential target promoter motifs for individual ECF groups, a prerequisite for *in silico* regulon predictions. Based on our classification system, we have developed a freely available, web-based ECF identification and genome annotation tool, termed ECF*finder* (http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder). Moreover, ECF σ factors, as defined by our classification, have been added to the Microbial Signal Transduction (MiST) reference database (Ulrich and Zhulin, 2007). The ECF classification presented in this article, together with the supplementary material available in *Supporting information* or at http://microbial-stress.iab.kit.edu/87.php, provide a comprehensive resource for identifying and studying ECF-dependent gene regulation, which represents the third pillar of bacterial signal transduction.

2.2. Results and discussion

2.2.1. Identification and grouping of ECF σ factors

The classification in this article is based on a manually curated data set of more than 2700 predicted ECF σ factor protein sequences derived from 369 microbial genomes (see Tables S2.1–S2.4). Our aim in generating the initial data set was to retrieve potential ECF sequences with as little bias as possible. We therefore applied a domain architecture-based definition of ECF σ factors that was derived from the features of all characterized ECF σ factors. A protein can be assumed to be an ECF σ factor, if it harbours two conserved domains, regions σ_2 and σ_4 of the σ^{70} protein family, but lacks a region σ_3 in between, and therefore has a spacing of less than 50 amino acids. Using this definition, we screened the SMART (simple modular architecture research tool) database (Letunic *et al.*, 2006; Schultz *et al.*, 1998) as described in *Experimental procedures*. Initially, we retrieved 4146 sequences, of which 2708 sequences remained as true ECF σ factors in our data set after removal of redundant (950) and false-positive sequences (488). See *Experimental procedures* for criteria, Table S2.1 for the phylum-specific distribution, and Tables S2.2–

S2.4 for the complete data sets; all supplementary material can be found in *Supporting information* or at http://microbial-stress.iab.kit.edu/87.php.

We identified groups of closely related sequences from multiple sequence alignments produced with CLUSTALW (Thompson *et al.*, 1994). The initial grouping was performed manually on the complete multiple sequence alignment, using the output option 'sorted by sequence similarity'. For this purpose, the alignment was divided into sets of roughly 100 sequences, with each set overlapping by at least 10 sequences with the previous set. We then constructed unrooted trees for each set, using the Phylip (Felsenstein, 1989) programs PROTDIST and FITCH, implemented in BioEdit Sequence Alignment Editor (Hall, 1999). By manually analysing the resulting set of unrooted trees and the multiple sequence alignment, we defined groups of closely related ECF sequences.

These initial groups were then used to develop a first generation of group-specific Hidden Markov Models (HMMs) for subsequent evaluation of the ECF classification. After the identification of outliers and refinement of the grouping, an improved set of HMM was developed that was used to define the ECF groups described below (the complete HMM set can be downloaded from the G2L homepage at http://www.g2l.bio.uni-goettingen.de/software/f_software.html, link: 'ECF HMM Models'). Moreover, the HMMs were also used for the development of a web-based ECF-identification tool, termed ECF*finder*, and for the addition of ECF σ factors to the MiST database (Ulrich and Zhulin, 2007).

2.2.2. Major and minor groups of ECF σ factors

The HMM-based analysis of the sequences in our data set allowed the assignment of 1735 out of the 2708 ECF proteins to one of 43 major phylogenetically distinct groups, each consisting of more than 10 proteins, and named ECF01 through ECF43 (see Table S2.2 for the complete data set). A summary of our classification is given in Table 2.1 (ECF σ factors) and Table 2.3 (corresponding anti-σ factors). Important and experimentally addressed ECF σ factors are listed, together with their (putative) function. The phylogenetic distribution of these groups is given in Figs 2.2 and 2.3, while the extended genomic context conservation is illustrated in Fig. 2.4. A detailed description of all 43 major groups can be found later in *Results and discussion*.

Within our classification, the two well-established types of RpoE-like and FecI-like σ factors are the predominant groups, and detailed analyses of the sequence similarity allowed a further subgrouping (RpoE-like: ECF01–04; 272 proteins; FecI-like: ECF05–

ECF10, 254 proteins). In contrast to the RpoE- and FecI-like proteins, the five groups that represent the cytoplasmic-sensing ECF σ factors (ECF11–ECF15; 213 proteins; Table 2.1) do not share a high degree of sequence similarity (Fig. 2.2). Instead, their unifying feature is their physiological role linked to the presence of soluble anti-σ factors (see below), which suggests that they sense stimuli from within the cell, in contrast to most other ECF σ factors that seem to respond to signals from the cell envelope or the environment. The function of some of the other groups (e.g. ECF16–ECF19) can be postulated, based on data from at least one well-investigated member (Table 2.1) and in these cases, its name was used as a descriptive term in Table 2.1.

In addition to these well-established types of ECF σ factors, our analysis also identified a number of major, novel groups with more than 100 members and unique features (e.g. ECF26, ECF41, ECF42). These large novel groups are found in many bacterial phyla, and have yet to be characterized experimentally.

Table 2.1. Summary of the ECF σ factor classification.

Group[a]	No.[c]	Phylum (No.)[d]	Anti-σ factor[e]	Genomic context conservation[f]	Example(s)[g]	(Putative) function	References
RpoE-like[b] (272)[c]							
ECF01	138	Ad (18), At (11), B (30), Cb (3), Cf (10), F (37), Pl (6), Pr (22), T (1)	+ (115)		SigW (Bsu)	Envelope stress response, antimicrobial compounds production and detoxification	(Helmann, 2006)
ECF02	111	Pr (111)	+ (107)		RpoE (Eco), AlgU (Pae)	Envelope stress response stress response, alginate production	(Ades, 2004) (Alba and Gross, 2004)
ECF03	13	B (12), Cf (1)	+ (12)		-		
ECF04	10	B (1), Cf (9)	+ (9)		-		
FecI-like[b] (254)[c]							
ECF05	94	Pr (94)	+ (83)	FecR-like anti-σ factor, outer membrane protein	FecI (Eco), HurI (Bpe)	Iron uptake	(Braun and Mahren, 2005)
ECF06	14	Pr (14)	+ (13)	FecR-like anti-σ factor, outer membrane protein	-		
ECF07	34	Pr (34)	+ (33)	FecR-like anti-σ factor, outer membrane protein	HasI (Sma)	Iron uptake	(Biville *et al.*, 2004)
ECF08	17	Pr (17)	+ (17)		-		
ECF09	20	Pr (20)	-		PbrA (Pfl) PvdS (Pae)	Iron uptake Siderophore biosynthesis	(Sexton *et al.*, 1996) (Leoni *et al.*, 2000)
ECF10	75	B (73), Pr (2)	+ (75)	FecR-like anti-σ factor, outer membrane protein	-	Carbohydrate metabolism	(Xu *et al.*, 2004)

Cytoplasmic-sensing ECF σ factors[b] (213)[c]							
ECF11	58	Pr (58)	+ (56)		RpoE (Rsp)	Oxidative stress response	(Dufour *et al.*, 2008)
ECF12 (SigR-like[b])	47	At (34), B (1), Cb (9), Pl (1), Pr (2)	+ (41)		SigR (Sco) RpoE1 (Mxa)	Redox homeostasis Unknown	(Paget *et al.*, 2001) (Ward *et al.*, 1998)
ECF13	27	Pr (27)	+ (21)		Ecf (Ngo)	Oxidative stress response	(Gunesekere *et al.*, 2006)
ECF14	26	At (26)	+ (25)	*O*-methyltransferase	SigE (Mav)	Stress response	(Wu *et al.*, 1997)
ECF15 (EcfG-like[b])	55	Pr (55)	+ (39)	PhyR-like RR, HK	SigT (Ccr), EcfG1 (Mex), EcfG (Bja), RpoE4 (Ret), RpoE2 (Sme)	General stress response	(Alvarez-Martinez *et al.*, 2007; Francez-Charlot *et al.*, 2009; Gourion *et al.*, 2009; Martinez-Salazar *et al.*, 2009; Sauviac *et al.*, 2007)
ECF16 (SigF-like[b])	36	Pr (36)	+ (34)		SigF (Ccr)	Oxidative stress response in stationary phase	(Alvarez-Martinez *et al.*, 2006)
ECF17 (SigU-like[b])	23	At (23)	+ (22)		SigU (Sco) SigL (Mtu)	Enhanced protein secretion Cell envelope lipid synthesis, secreted protein modification	(Gordon *et al.*, 2008) (Hahn *et al.*, 2005)
ECF18 (RpoT-like[b])	59	Ad (1), Pr (58)	+ (54)		RpoT (Ppu)	Toluene tolerance	(Duque *et al.*, 2007)
ECF19 (SigK-like[b])	26	At (26)	+ (23)		SigK (Mbo)	Antigenic protein expression	(Veyrier *et al.*, 2008)
ECF20	58	Ad (2), Pr (56)	+ (55)		CnrH (Rme)	Cobalt and nickel resistance	(Grosse *et al.*, 2007)
ECF21	22	B (21), Cb (1)	+ (20)		-		
ECF22	35	Ad (3), B (24), Pl (2), Pr (6)	-		-		

ECF23	19	F (19)	+ (19)		-		
ECF24	29	At (3), B (5), F (3), Pl (1), Pr (17)	-		SigZ (Bsu)	Unknown	
ECF25	12	Cy (12)	+ (12)		-		
ECF26	103	Ad (1), Pr (102)	+ (100)	(i) Cytochrome *c* oxidase, metallophosphoesterase (ii) Cytochrome, catalase (iii) Lipoprotein	SigE (Sno)	Regulation of thiosulphate oxidation	(Kappler *et al.*, 2001)
ECF27	20	At (20)	+ (14)		SigT (Sco), SigM (Cgl)	Diverse stress responses	(Nakunst *et al.*, 2007)
ECF28	27	Pr (27)	+ (24)		-		
ECF29	24	Ad (2), B (2), Pr (20)	-	(i) Outer membrane protein, blue copper protein (ii) Exported protein, cytochrome	-		
ECF30	65	At (3), F (62)	+ (63)		SigV (Bsu)	Unknown	(Zellmeier *et al.*, 2005)
ECF31	11	At (1), F (8), S (1), Pr (1)	+ (9)		SigY (Bsu)		(Cao *et al.*, 2003)
ECF32 (HrpL-like[b])	12	Pr (12)	-		HrpL (Eam)	Hypersensitive response regulation	(Wei and Beer, 1995)
ECF33	13	Pr (13)	+ (7)		-		
ECF34	21	Ad (1), At (2), Cf (4), Cy (12), Pr (2)	+ (16)		-		

ECF35	14	Pr (14)	+ (14)	von Willebrand factor domain protein	-		
ECF36	13	At (13)	-		SigC (Mtu)	Virulence regulation	(Karls *et al.*, 2006)
ECF37	24	Pr (24)	+ (22)		-		
ECF38	16	At (16)	+ (14)		-		
ECF39	33	At (33)	+ (22)		SigE (Sco)	Cell wall integrity	(Paget *et al.*, 1999a)
ECF40	19	At (19)	+ (13)		SigD (Mtu)	Expression of subset of ribosomal genes	(Calamita *et al.*, 2005)
ECF41	115	Ad (2), At (64), Cy (1), F (7), Pr (41)	-	Carboxymuconolactone decarboxylase, oxidoreductase	SigJ (Mtu)	Hydrogen peroxide resistance	(Hu *et al.*, 2004)
ECF42	110	Ad (5), At (45), B (2), Cy (2), Pl (2), Pr (54)	-	DGFP protein	-		
ECF43	36	Ad (18), Cy (2), Pl (10), Pr (6)	-	Serine-threonine kinase	-		

[a] Each group represents a phylogenetically distinct set of ECF σ factors, based on sequence similarity.
[b] The group name is either derived from the best-studied example of a (cluster of phylogenetically related) group(s), or – in case of the cytoplasmic-sensing ECF σ factors – is based on a common function, which correlates with the presence of soluble anti-σ factors.
[c] The number given is the total count of ECF σ factors from a given group in our initial data set (see Table S2.2 for details).
[d] Abbreviations used: Ad, Acidobacteria; At, Actinobacteria; B, Bacteroidetes; Cb, Chlorobi; Cf, Chloroflexi; Cy, Cyanobacteria; F, Firmicutes; Pl, Planctomycetes; Pr, Proteobacteria; S, Spirochaetes; T, Thermotogae.
[e] '-' (mostly) absent; '+' means present in (No.) of all cases, relative to the total number of ECF σ factors of this group, which is given in column three. See Table 3 for details.
[f] See Fig. 4 for details.
[g] Abbreviations for species names: Bja, *Bradyrhizobium japonicum*; Bpe, *Bordetella pertussis*; Bsu, *Bacillus subtilis*; Ccr, *Caulobacter crescentus*; Cgl, *Corynebacterium glutamicum*; Eam, *Erwinia amylovora*; Eco, *Escherichia coli*; Mav, *Mycobacterium avium*; Mbo, *Mycobacterium bovis*; Mex, *Methylobacterium extorquens*; Mtu, *Mycobacterium tuberculosis*; Mxa, *Myxococcus xanthus*; Ngo, *Neisseria gonorrhoeae*; Pae, *Pseudomonas aeruginosa*; Pfl, *Pseudomonas fluorescens*; Ppu, *Pseudomonas putida*; Ret, *Rhizobium etli*; Rme, *Ralstonia metallidurans*; Rsp, *Rhodobacter sphaeroides*; Sco, *Streptomyces coelicolor*; Sma, *Serratia marcescens*; Sme, *Sinorhizobium meliloti*.

The taxonomic diversity, phylogenetic relationship and sequence diversity within each group is visualized in the phylogenetic tree shown in Fig. 2.2. For clarity, only the 32 largest groups were included (those that consist of more than 20 proteins and/or show an extended genomic context conservation). Interestingly, the phylogenomic analysis identified one distinct group of ECF-like proteins (ECF43) that fulfill all the criteria for ECF σ factors, but show similar phylogenetic distances to both ECF σ factors and other σ^{70} proteins (Fig. 2.2). Since none of the members of this group has been experimentally studied, the significance of this finding is still unclear.

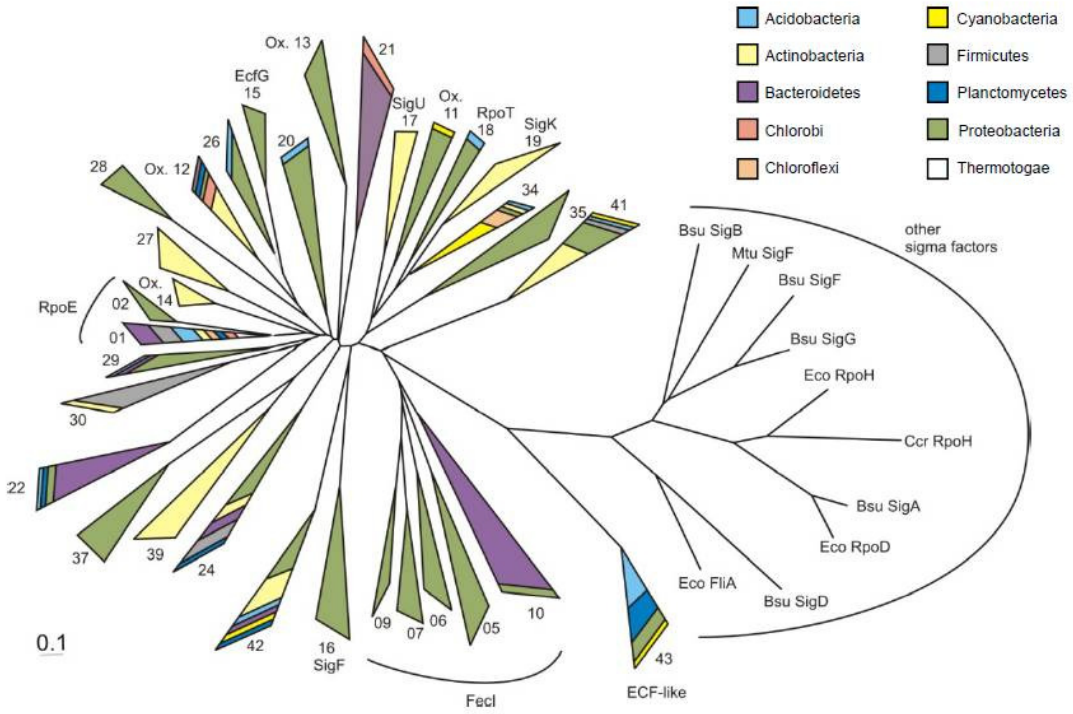

Figure 2.2. Phylogenetic tree of ECF σ factors. The phylogenetic tree of ECF σ factors is based on a gapless multiple sequence alignment of regions σ_2 and σ_4, from the two most distant representatives from within each group. Each triangle represents one major group, colour-coded according to the phylogenetic distribution of the ECF σ factors. Thus, the length of the triangle's edges reflects the overall sequence diversity within the group (the longer the edges, the larger the sequence diversity within this group). A colour-coded multiple sequence alignment, illustrating both the overall and group-specific sequence conservation, is shown in Fig. S2.1. Bsu, *B. subtilis*; Ccr, *C. crescentus*; Eco, *E. coli*; Mtu, *M. tuberculosis*. CLUSTALW (Thompson *et al.*, 1994) was used to construct the multiple sequence alignments. The tree was calculated by the Least Squares method of the Phylip (Felsenstein, 1989) programs PROTDIST and FITCH, which are implemented in the BioEdit Sequence Alignment Editor (Hall, 1999). For reasons of clarity, only the 32 most important groups are shown, which contain at least 20 proteins in our data set and/or show an extended genomic context conservation.

Half of the groups include ECF σ factors from only one phylum, yet the remaining groups are represented in different phyla, with eight of them (ECF01, 12, 22, 24, 34, 41, 42 and 43) showing remarkable taxonomic diversity (Fig. 2.2). For many of these groups, the functional relevance of our classification is underscored by the conservation of additional

features, which includes predicted anti-σ factors, the overall genomic context or putative target promoters, as will be described later.

In addition to the 43 major groups, another 128 ECF σ factors were assigned to one of 24 minor groups (ECF101–124), which by definition contain fewer than 10 proteins (see Table S2.3 for details). A total of 835 proteins did not cluster with any other ECF σ factor in our data set and therefore are regarded as unclassified singletons (Table S2.4). These proteins might have a very unique function or their isolation could be a result of the bias generated by the availability of genome sequences.

2.2.3. Effect of genome sequence availability bias on the ECF classification

The set of publicly available genomes is biased, with some phyla (e.g. Firmicutes, Proteobacteria) or types of organisms (especially pathogens and bacteria of industrial importance) being heavily over-represented. This has a significant impact on both the quality and diversity of the ECF classification.

More than 90% of the 1873 classified ECF sequences belong to only four bacterial phyla – Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria – with 953 σ factors from Proteobacteria alone. In contrast, the remaining 155 classified ECF σ factors are derived from nine bacterial phyla (Table S2.1). These numbers reflect the number of genomes sequenced from the respective phyla. Of the 369 genomes represented in our data set, 200 are proteobacterial, 124 are from the other three major phyla, whereas only 45 genomes represent the remaining nine phyla (Table S2.1).

Because of the phylum specificity of several major groups (Fig. 2.2), these numbers indicate that our classification is far from representative of the microbial world. We expect that other major ECF groups have yet to be identified and that ECFs that may belong to these groups are

currently associated with minor groups or even unclassified ECF σ factors.

This hypothesis is supported by the results shown in Fig. 2.3C. The vast majority of σ factors from species that belong to well-represented phyla can be assigned to one of the major groups. On the other hand, of the 49 ECF σ factors identified in the genome of *Pirrelula baltica*, only five can be assigned to well-described groups (RpoE-like). Ten others belong to poorly understood groups (ECF22 – two; ECF24 – one; ECF42 – two; and ECF43 – five), yet the majority (30) remain unclassified (Fig. 2.3C). This particularly ECF-rich bacterium belongs to the phylum Planctomycetes, which is heavily under-represented with regard to the availability of complete genome sequences, despite its wide

distribution in aquatic ecosystems. Other bacteria that show a remarkable discrepancy between the total number of ECF σ factors and the subset that can be classified include *M. xanthus* and *S. cellulosum* (Fig. 2.3C), again from phyla under-represented in the microbial genome databases.

Because of this unavoidable bias in our current classification, and given the fast increase in genomic sequence information, we believe that novel major and minor groups will emerge in the future. Accordingly, we have designed our nomenclature to easily accommodate additions. Each new group will receive a unique identifier starting with ECF44/ECF125 for major or minor groups respectively. To co-ordinate future additions to this classification, including updates of the ECF*finder* tool and MiST database, we encourage scientists who have identified a novel minor or major group to get in contact with us.

2.2.4. Distribution and phylum-specific patterns of ECF σ factors

ECF σ factors are ubiquitously distributed within and restricted to the bacterial kingdom. The average genome contains about six ECF σ factors. While this number is lower than the average numbers of 1CS and 2CS, it nevertheless underscores the importance of ECF σ factors as a fundamental mechanism of bacterial signal transduction. The number of ECF σ factors generally increases in relation to genome size and the complexity of the organism's environment, a trend also observed for 1CS and 2CS (Ulrich and Zhulin, 2007). But there are noteworthy exceptions to this rule. Cyanobacteria, in particular, possess remarkably few ECF σ factors, especially relative to 1CS and 2CS. The most prominent example is *A. variabilis*, which is one of the most 2CS-rich bacteria sequenced to date (Ulrich and Zhulin, 2007). It harbours 73 putative sensor kinases, 71 RR, 54 hybrid kinases and 145 1CS, but only one ECF σ factor.

No ECF σ factors are found in currently sequenced genomes of the phyla Aquificae, Chlamydiae, nor in the Rickettsiales (phylum Proteobacteria), the genera *Borrelia* (Spirochaetes) and *Mycoplasma* (Firmicutes), all of which are obligate symbionts or pathogens with very small genomes (less than 2 Mb). Conversely, Acidobacteria, Actinobacteria and Bacteroidetes are particularly ECF-rich phyla and contain 40–50 ECF σ factors per genome (Table 2.2). The *S. cellulosum* genome (δ-proteobacteria) contains the most (83) ECF σ factors (Fig. 2.3C).

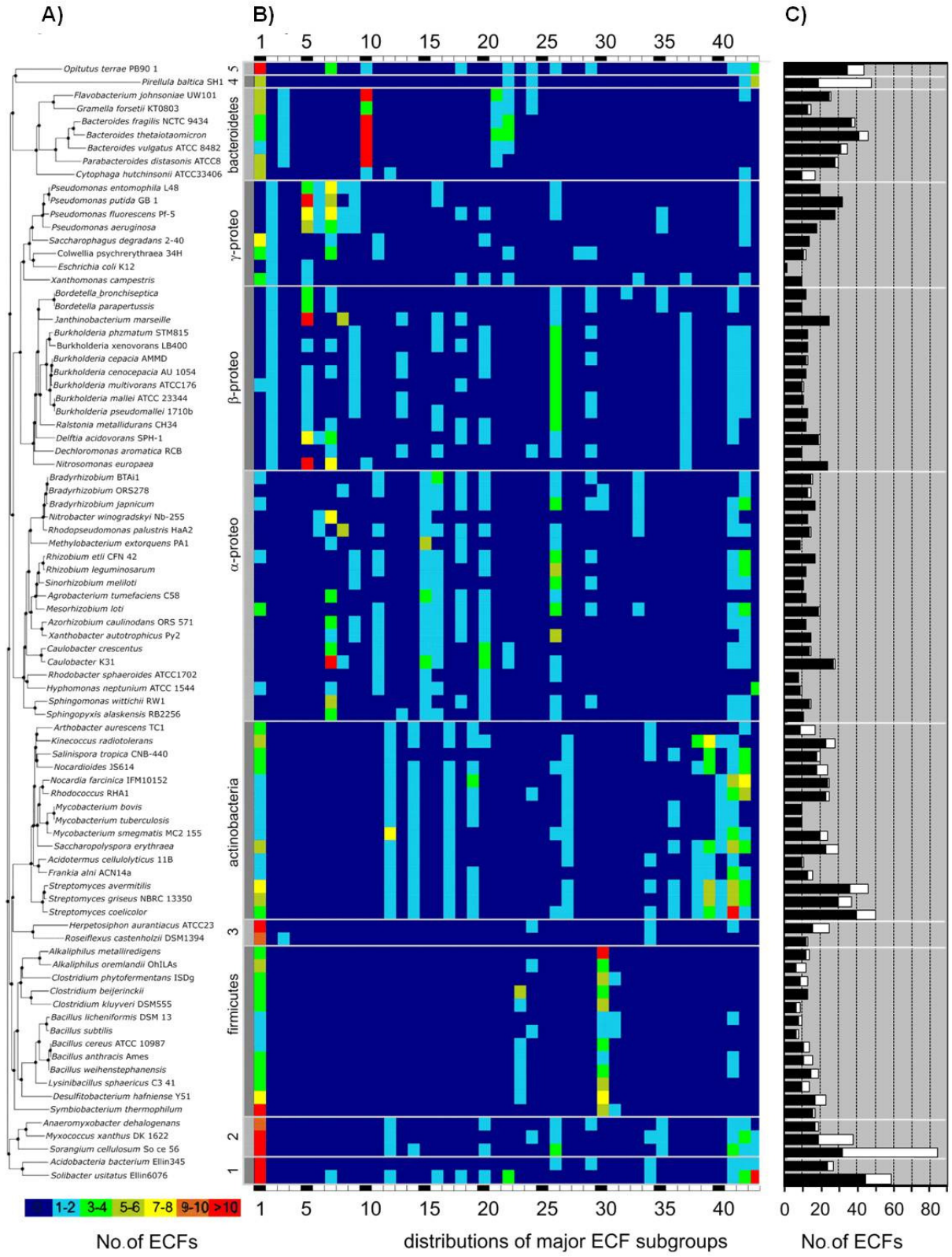

Figure 2.3. Distribution pattern of ECF groups in ECF-rich bacterial species. (A) Phylogenetic tree (based on 16S rRNA) of representative ECF-rich bacterial species. The names of larger phyla is given to the right; less abundant phylogenetic groups are 1 – Acidobacteria, 2 – δ-proteobacteria, 3 – Chloroflexi, 4 – Planctomycetes, and 5 – Verrucomicrobia. The tree was constructed with MEGA4, employing CLUSTALW for the initial alignment. The resulting tree was checked in a bootstrap analysis with 500 replicates. The order of nodes has been clustered by the alignment and was taken directly for the graphic. (B) The corresponding ECF distribution pattern of the 43 major ECF groups is illustrated by a two-dimensional heat map, where the colours indicate the numbers of ECF σ factors from a given group in the genome sequence. The output is

based on our data set and an ECF*finder* analysis of the JGI database. (C) Number of major group (black bars) and minor group/unclassified (white bars) ECF σ factors. Therefore, the absolute length of the bars indicates the total number of ECF σ factors in the given species. The numbers were taken from the data stored in the MiST database.

The distribution pattern of ECF σ factor groups is phylum-specific. Figure 2.3B graphically illustrates the presence of ECF groups in 80 representative bacterial species that contain more than 10 ECF σ factors (including the model organism *S. coelicolor*). We also added other model bacteria for the study of ECF σ factors that harbor fewer than 10 ECF σ factors (i.e. *E. coli*, *B. subtilis* and *R. sphaeroides*).

Table 2.2. Abundance of ECF σ factors in the microbial world.

No. ECFs	No. genomes	Representative examples or bacterial groups[a]
0	239	Aquificae, Chlamydiae, Rickettsiales, Borrelia, *Mycoplasma* spp.
1	98	*Staphylococcus* spp., *A. Variabilis*
2	62	*E. coli*
3-5	85	*C. glutamicum*, *E. faecalis*
6-10	95	*B. subtilis*, *R. sphaeroides*
11-20	89	*B. anthracis*, *C. crescentus*, *P. aeruginosa*
21-30	18	*P. fluorescens*, *P. putida*, *Frankia* sp.
>30	14	*Streptomyces* spp., *Bacteroides* spp., *M. xanthus*

[a] Representative model bacteria are listed as examples; in the context of this table, 'bacterial group' can be anything between genus and kingdom.
The data are based on an ECF*finder* analysis of the JGI microbial genomics database at http://genome.jgi-psf.org/mic_home.html.

2.2.5. Robustness, sensitivity and limitations of the ECF group-specific HMMs

Sixty-seven ECF group-specific HMMs have been constructed and tested against SWISSPROT to determine the robustness in terms of specificity, sensitivity and limitations of the models (see *Experimental procedures*). The SWISSPROT data set was considered a gold standard containing all known ECF σ factors. The protein sequences included in SWISSPROT have been compared with each of the HMMs using the programs HMMSEARCH and HMMPFAM with standard parameters. All SWISSPROT entries of proteins annotated as ECF σ factors produced hits above the trusted cut-offs of the matching HMM. All sequences found in this performance test which were not annotated as ECF σ factors matched the criteria for ECF σ factors and have therefore been classified as novel putative ECF σ factors.

However, even though each ECF protein usually produced exclusively one significant hit, some of the ECF σ factors scored in more than one HMM above the trusted cut-off. This observation might indicate closely related groups of ECF σ factors. For example, this is the case for groups ECF01 and ECF02, as well as the pair ECF20/ECF27 (see Table S2.9 for a summary). The two diverse groups ECF01 and ECF20 (and hence the corresponding

HMMs) were especially prone to such a performance. But so far, the difference in score between the best hits (i.e. group members) and the remaining hits (false positives above trusted score, generated by members of different, closely related ECF groups) were significantly distinct. Moreover, proteins generating such a false-positive hit always scored significantly higher for their specific ECF group, thereby allowing a clear discrimination (the complete scan results of all HMMs against all of the 2708 ECF proteins in our data set can be downloaded at http://www.g2l.bio.uni-goettingen.de/software/f_software.html, link: 'evidence data', for in-depth information). Therefore, we consider the existence of 43 separated models justified. Obviously we cannot rule out the possibility that some of our groups ultimately need to be merged, if additional ECF sequences from future sequencing projects fill the gap between two closely related groups by performing equally well with two closely related HMMs.

Due to the clear discriminative power of the group-specific HMMs, it was not possible to identify ECF σ factors belonging to the singletons with any of the existing models. Therefore, we developed another HMM based on all unclassified ECF sequences. This general HMM is capable of identifying all ECF σ factors, including unclassified, untypical and ECF-like sequences. Unfortunately, it also identifies some non-ECF σ factors. These hits represent proteins highly similar to ECF σ factors, such as the sporulation σ factor σ^H of *B. subtilis* and closely related *Bacillus* spp. In fact, these proteins fulfill all criteria of our ECF definition: they are small proteins containing regions σ_2 and σ_4, but lacking region 3. Phylogenetically, these proteins are more closely related to members of group ECF43 than to any other σ^{70} family protein (data not shown). Therefore, σ^H and its homologues could be considered as potential ECF-like proteins.

So far, it was not possible to construct a single general HMM, which is able to find all true ECF σ factors, while simultaneously discriminating all closely related non-ECF σ factors. Nevertheless, by combining the group-specific HMMs with the singleton-derived general HMM, we maintain a high discriminative power for all classified ECF σ factors, while still being able to gather even more distantly related protein sequences, including those classified as ECF-like.

The fast increase in genome sequence information becoming available (including metagenomic projects, which will presumably also contain numerous so far underrepresented phyla) will hopefully allow us to more specifically classify more and more ECF σ factors that so far belong to the singletons. This would allow the generation of additional group-specific HMMs with a high discriminative power comparable to the

existing models. Moreover, this will also help to clarify the question of discriminating closely related ECF groups, thereby gradually overcoming the current limitations in ECF identification and classification.

2.2.6. Identification and analysis of anti-σ factors specific for ECF groups

Similar to the HK/RR pairs of 2CS, the functional unit of ECF-dependent signalling normally consists of two proteins: the ECF σ factor and its cognate anti-σ factor. Anti-σ factors typically have a small cytoplasmic domain which is often linked by one or sometimes more transmembrane regions (TMR) to a C-terminal periplasmic domain. An anti-σ factor gene is normally located next to and coexpressed with the gene encoding the corresponding σ factor. In contrast to the clearly defined transmitter and receiver domains of 2CS, most anti-σ factors seem poorly conserved and can be difficult to detect by sequence similarity alone because of their overall diversity and the small size of the σ interaction domain (Campbell and Darst, 2005). One exception to this rule are the FecR-like proteins (named after *E. coli* FecR; (Van Hove *et al.*, 1990)), which function as anti-σ factors for FecI-like ECF σ factors (ECF05–10). These proteins harbour a conserved periplasmic FecR domain, which is important for stimulus perception (Enz *et al.*, 2003).

More recently, structural studies of two anti-σ factors, *E. coli* RseA and *R. sphaeroides* ChrR, together with in-depth sequence profiling and cluster analysis of numerous putative ECF anti-σ factors, resulted in the identification of a structural motif, termed ASD (for anti-σ domain), which is found in about one-third of all recognizable anti-σ factors that are genetically linked to ECF σ factors (data in Tables S2.5–S2.7). About 30% of these ASD-containing proteins are so-called zinc-binding anti-σ factors, or ZAS proteins (Campbell *et al.*, 2007; Zdanowski *et al.*, 2006).

We screened the genomic neighbourhood of ECF-encoding genes from our data set for the presence of adjacent, usually downstream located genes encoding candidate anti-σ factors, based on the presence of TMR and regions of group-specific sequence conservation (including ASD and ZAS domains) essentially as described previously (Campbell *et al.*, 2007). The results of these analyses are summarized in Table 2.3, and the complete anti-σ data set is listed in Tables S2.5–S2.7 (major groups, minor groups and singletons respectively).

With the exception of the ASD anti-σ proteins, our analysis verified the difficulty of detecting overall sequence similarity across anti-σ factors derived from different ECF groups, as noted previously (Campbell *et al.*, 2007). Conversely, candidate anti-σ factors

from within a specific ECF group often show a significant degree of sequence similarity in a subpopulation of the proteins, which sometimes extends to an entire group (Table 2.3, and Tables S2.5 and S2.6). In many cases, the C-terminal (periplasmic) domain is unique to a specific ECF group, including already recognized motifs belonging to clusters of orthologous groups of proteins (entries from the COG database at http://www.ncbi.nlm.nih.gov/COG/), such as COG5343 (ECF18/19 and ECF34) and COG579 (ECF33) (Table 2.3 and Table S2.5).

Many sequences lacking ASD domains (typically ~80 amino acids) nevertheless had shorter N-terminal extensions (~50 amino acids) similarly predicted to be located in the cytoplasm. As overall sequence similarity could not be detected in these N-terminal extensions (labeled as CAS, for Candidate Anti-σ Sequence in Table 2.3), they may or may not be structurally related, since the sequence similarity in the ASD is also at the edge of detection (Campbell *et al.*, 2007). Further studies will be required to prove if some of these so far uncharacterized extensions also function as interaction interfaces for σ factor binding.

Sometimes the clusters of anti-σ proteins match up with complete ECF groups and in other cases there is a finer granularity that effectively subdivides the ECF group. Within a category there can be many arrangements of ECF factor, anti-σ N-domain and C-domain, which provides a complex collection of combinations (='combinatorial complexity'). Within the 'diverse' category of C-terminal domains (Table 2.3), there are typically clusters of similarity that can be observed as well as many sequences that appear to be unique. Additional sequence data in the databases may clarify some of the relationships, as well as increasing the numbers of structures. Note that there are two C-terminal domains that are similar to solved three-dimensional structures of unknown function (3FBQ and 3CYG).

Table 2.3. Summary of anti-σ factors.

ECF σ factors[a]	Anti-σ factors				
Group (No.)	**Presence**	**TMR[b]**	**N-domains[c]**	**C-domains[d]**	**Remarks, examples, reference[e]**
RpoE-like (272)					
ECF01 (138)	+ (116)	1 (87), 2 (10), 0 (11), 3+ (8)	ASD (68) [62 ZASD]	Diverse	RsiW (Bsu) (Helmann, 2006)
ECF02 (111)	+ (107)	1	ASD	RseA-C	RseA (Eco) (Ades, 2008)
ECF03 (13)	+	1 (12), 0 (1)	CAS	Short	1 anti-σ with ZAS-TM0 architecture
ECF04 (10)	+ (9)	1	CAS	Short	
FecI-like (254)					
ECF05 (94)	+ (83)	1	ASD	pfam04773	Prototypical systems; FecR (Eco) (Braun and Mahren, 2005)
ECF06 (14)	+ (13)	1	ASD	pfam04773	Functional similarity to ECF05
ECF07 (34)	+ (33)	1	ASD	pfam04773	Functional similarity to ECF05
ECF08 (17)	+	1	ASD	pfam04773	
ECF09 (20)	-	-	-	-	Often near MbtH-like genes (Lautru *et al.*, 2007); 1 FecR-like anti-σ
ECF10 (75)	+	1 (73), 0 (2)	ASD (60), NTE (15)	pfam04773	Bacteroidetes FecIR group
Cytoplasmic-sensing ECF σ factors (213)					
ECF11 (58)	+ (56)	0	ASD (56) [ZASD 51]	Cupin	Similar to ChrR (Rsp) (Campbell *et al.*, 2007)
ECF12 (47) (SigR-like)	+ (43)	0 (40), 1 (3)	ZASD (40), CAS (3)	Short	
ECF13 (27)	+ (23)	0 (23)	ZASD	Short	
ECF14 (26)	+ (25)	0 (15), 1 (10)	ZASD (24)	Diverse	1 anti-σ with CAS-TM1-C architecture
ECF15 (55)	+ (44)	0	Short, NepR-like (no separate N-/C-domains)		Linked to PhyR-like RR (Francez-Charlot *et al.*, 2009)

ECF16 (36) (SigF-like)	+ (34)	6	DUF1109 (no separate N-/C-domains)		
ECF17 (23) (SigU-like)	+ (22)	1	ZASD (21)	Conserved (C-17)	1 anti-σ with CAS
ECF18 (59) (RpoT-like)	+ (54)	1 (52), 0 (2)	ASD (53)	cog5343/DUF2337 (52)	1 ChrR-like anti-σ factor
ECF19 (26) (SigK-like)	+ (23)	1 (22), 0 (1)	ASD (22) [ZASD 6]	cog5343/DUF2337 (22), short (1)	RskA-like anti-σ factors (Veyrier *et al.*, 2008); 1 anti-σ very short
ECF20 (58)	+ (55)	1 (48), 2 (4), 0 (3)	ASD (18), CAS (37)	Diverse	Sometimes near vWA- or small CnrYX/NccYX-like genes
ECF21 (22)	+ (21)	1	CAS	Diverse	
ECF22 (35)	+ (35)	4 (34), 0 (1)	CAS (35)	–	
ECF23 (19)	+ (18)	1 (15), 2 (2), 0 (1)	CAS (18)	3FBQ-like domain (12)	
ECF24 (29)	-	-	-	-	Near arsenate reduction genes
ECF25 (12)	+	1	ASD	Conserved (6)	Remaining C-termini not conserved
ECF26 (103)	+ (100)	1	ASD	cog5662 (91), Conserved (9)	PrtR-like anti-σ factors (Burger *et al.*, 2000)
ECF27 (20)	+ (14)	1	ASD (9) [ZASD 4]	Diverse	Near thioredoxin genes
ECF28 (27)	+ (24)	1	ASD (24)	Conserved (24)	
ECF29 (24)	-	-	-	-	Conserved adjacent genes (Fig. 2.4)
ECF30 (65)	+ (54)	1 (53), 0 (1)	CAS (50), ZASD (4)	Diverse	29 anti-σ factors with C domain similar to 3FBQ or 3CYG proteins; 9 ECF with PadR-like protein instead of anti-σ factors
ECF31 (11)	+ (10)	2 (7), 1 (2), 0 (1)	ASD (2)	Diverse	1 anti-σ with ZASD

ECF32 (12) (HrpL-like)	-	-	-	-	Linked to HrpXY-like 2CS (Merighi *et al.*, 2003)
ECF33 (13)	+ (7)	1	ASD	cog579-C (C-33)	
ECF34 (21)	+ (16)	1 (14), 0 (2)	ASD (16) [ZASD 3]	cog5343 (14), DUF2337 (14)	2 ChrR-like anti-σ factors
ECF35 (14)	+	1	CAS (12)	Diverse	2 anti-σ factors with ZASD
ECF36 (13)	-	-	-	-	
ECF37 (24)	+ (24)	1 (23), 2 (1)	CAS	Diverse	
ECF38 (16)	+ (15)	1	CAS	Diverse	
ECF39 (33)	+ (22)	1	CAS	Diverse	
ECF40 (19)	+ (13)	1	-	Conserved	
ECF41 (115)	-	-	-	-	CMD/oxidoreductase genes adjacent
ECF42 (110)	-	-	-	-	cog04941 proteins (with ferredoxin-like DGPFAETKE domain) adjacent
ECF43 (36)	-	-	-	-	Serine-threonine kinases adjacent

[a] See Table 2.1 for details.
[b] Number of (putative) transmembrane regions (TMR), as derived from TMHMM predictions (Sonnhammer *et al.*, 1998).
[c] N-terminal domains: ASD, anti-σ domain; CAS, Candidate Anti-σ Sequences, diverse non-conserved N-terminal extensions not phylogenetically related to ASD; ZASD, zinc-binding ASD.
[d] C-terminal domains (often separated by TMR from the N-terminal domain). Conserved – group-specific conservation, identifier given as defined in the COG, Pfam or PDB databases.
[e] See Table 2.1 footnotes for species abbreviations. CMD, carboxymuconolactone decarboxylase; vWA, von Willebrand factor A. Identification and classification of anti-σ factors by clustering analysis was performed as described previously (Campbell *et al.*, 2007). Numbers in parentheses give the number of ECF σ factors for the respective category (as a subfraction of the total number of anti-σ factors that were identified, as listed in the first column). No value is given, if the numbers are identical to the those in the next column to the left.

However, not all ECF groups are linked to potential anti-σ factors, even in well-defined groups such as ECF09 (FecI-like σ factors). Other groups lacking well-conserved cognate anti-σ factors include ECF22, ECF24, ECF36, and the two large groups ECF41/ECF42 (Table 2.3). The lack of discernible anti-σ factors suggests alternative mechanisms of controlling σ factor activity. For some of these groups, the ECF σ factor may be regulated at the transcriptional level, as has been described for *S. coelicolor* σ^E. Here, *sigE* transcription is directly controlled by a 2CS, CseBC, that responds to cell wall stress (Hong *et al.*, 2002).

But despite the diversity and combinatorial complexity observed, there is a good correlation between the ECF groups and the conservation of their cognate anti-σ factors throughout our classification (Tables 2.1 and 2.3). Therefore, both the presence and absence of putative anti-σ factors strongly support our ECF classification.

2.2.7. Genomic context conservation within ECF groups

An important lesson learned from the genomic era is the realization of the modularity of microbial genomes. Genes encoding proteins of related functions (e.g. enzymes that catalyse successive steps in a metabolic pathway, regulators and their direct target genes etc.) are often encoded in one or more operon(s) that are located adjacent to each other on the chromosome. This has been demonstrated for numerous 2CS, as well as the FecI-like ECF σ factors (Braun and Mahren, 2005; Joseph *et al.*, 2002). While conserved genes next to regulator-encoding genes often represent direct target genes, they may also encode additional functions important for the mechanism of signal transduction (Jordan *et al.*, 2006).

In some ECF groups, the genes encoding the σ factor are flanked by highly conserved genes encoding altogether novel types of potential anti-σ factors. For example, this is observed in groups ECF15 (NepR-like proteins) and ECF16 (linked to proteins containing the conserved domain of unknown function DUF1109).

Moreover, several ECF groups show a conserved genomic context that extends beyond the classical σ/anti- σ pairs of most ECF groups (Fig. 2.4). Remarkably, this frequently correlates with the absence of a recognizable anti-σ factor (e.g. the novel groups ECF29 and ECF41–43; Tables 2.1 and 3, and Fig. 2.4). This observation raises the potential for a number of alternative routes for ECF activation, including functional links to 2CS (ECF15, ECF32 and potentially ECF39), serine/threonine kinases (ECF43) and proteins with enzymatic activities (ECF41 and ECF42). Where applicable, the group-specific genomic

context conservation therefore serves as an additional parameter supporting our ECF classification (Table 2.1).

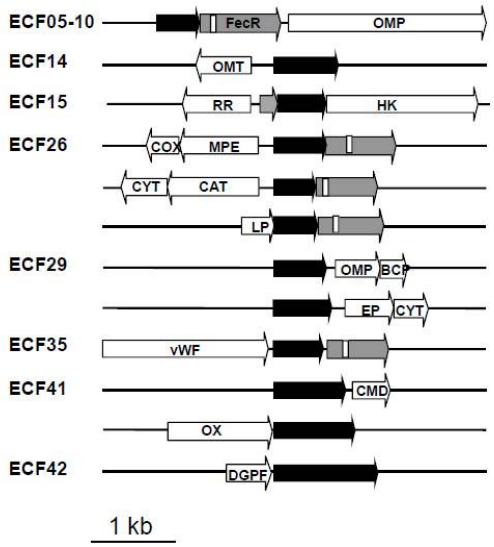

Figure 2.4. Genomic context conservation in selected ECF groups. ECF σ factors are shown in black, anti- σ factors in grey, conserved neighbouring proteins in white. Abbreviations: BCP, blue copper protein; CAT, catalase; CMD, carboxymuconolactone domain protein; COX, cytochrome *c* oxidase; CYT, cytochrome; DGPF, DGPF protein; EP, exported protein; HK, histidine kinase; LP, lipoprotein; MPE, metallophosphoesterase; OMP, outer membrane protein; OMT, *O*-methyltransferase; OX, oxidoreductase; RR, response regulator; vWF, von Willebrand factor domain protein. Genomic context analyses were performed using The SEED at http://theseed.uchicago.edu/FIG/index.cgi (Overbeek *et al.*, 2003), and MicrobesOnline at http://www.microbesonline.org (Alm *et al.*, 2005).

2.2.8. Detailed description of the major ECF groups

In the following paragraphs, we present the major groups of ECF σ factors, as summarized in Tables 2.1 and 2.3, as well as Figs 2.2 and 2.4. We briefly describe the already well-established groups ECF01–10 (i.e. RpoE- and FecI-like proteins), with a special emphasis on those groups that deviate from already described paradigms. Readers are referred to the existing literature, including a number of excellent review articles for detailed information (Ades, 2008; Alba and Gross, 2004; Braun *et al.*, 2006; Helmann, 2006). We specifically focus on novel major groups of ECF σ factors with unique features indicative of novel functions and/or mechanisms of signal transduction. Because there are few experimentally investigated ECF σ factors, some of our hypotheses are solely derived from the available *in silico* data; however, the unique features of novel groups provide useful starting points for guiding future research on the corresponding proteins.

2.2.8.1. ECF01–ECF04 – RpoE-like ECF σ factors. RpoE-like ECF σ factors form a widely distributed group that is found in most bacterial phyla (Table 2.1 and Fig. 2.2). Surprisingly, it is absent in most α-proteobacteria (Fig. 2.3). The RpoE-like proteins fall into four distinct groups, ECF01–04. ECF01 is a very diverse group with proteins from nine different phyla. ECF02 contains the proteobacterial RpoE-like proteins, which show a significantly higher sequence similarity (Fig. 2.2). The remaining two groups are small, but clearly distinct from the other two larger groups (Table 2.1), as judged by sequence similarity and anti-σ factor conservation (Table 2.3).

Experimentally investigated RpoE-like σ factors are involved in diverse (often envelope) stress responses, and include the best-studied ECF σ factors, RpoE of *E. coli* (Ades, 2004; Alba and Gross, 2004; Renzoni et al., 2006) and σ^W from *B. subtilis* (Helmann, 2006).

The corresponding anti-σ factors are usually transmembrane ASD proteins with one TMR, with the ECF01-associated proteins containing the zinc-binding motif in the N-terminal ASD and classified as ZAS proteins. Upon receiving a stimulus, the anti-σ factor is inactivated by regulated proteolysis. This mechanism has been described for RseA and RsiW, regulating the activity of *E. coli* RpoE or *B. subtilis* σ^W respectively (Ades, 2008; Heinrich and Wiegert, 2006; Schöbel *et al.*, 2004; Zellmeier *et al.*, 2006). There is no genomic context conservation beyond the co-transcribed gene(s) encoding the anti-σ factor(s). The Bacteroidetes group of RpoE-like proteins (ECF03) differs from this trend in that it lacks a conserved anti-σ factor.

2.2.8.2. ECF05–ECF10 – FecI-like ECF σ factors. The FecI-like proteins – despite their overall similarity – fall into six clearly definable groups. Most of these are found in Proteobacteria, but one group almost exclusively occurs in the phylum Bacteroidetes (Table 2.1). Based on experimental evidence and genomic context conservation, many ECF σ factors belonging to groups ECF05–ECF09 are presumably involved in regulating iron acquisition in Proteobacteria. While a number of them have been investigated to some degree, most mechanistic insight derives from work on FecI from *E. coli*. This ECF σ factor regulates the transcription of genes encoding outer membrane proteins involved in iron siderophore uptake. For more details, readers are referred to two review articles from the group of Volkmar Braun (Braun *et al.*, 2003; Braun and Mahren, 2005). FecI homologues can be found in many bacterial species, with most of their genes located next to genes encoding the cognate anti-σ factor, which contains a FecR domain (Braun *et al.*, 2003; Braun and Mahren, 2005). In contrast to most ECF σ factors, *fecIR* transcription is not autoregulated.

The majority of ECF σ factors belonging to groups ECF05–ECF07 shows an extended genomic context conservation, with genes encoding outer membrane proteins forming an operon next to (usually downstream) the operon encoding the σ/anti-σ factor pair (see Fig. 2.4). In contrast, there is no extended genomic context conservation in groups ECF08 and ECF09.

While groups ECF05–08 contain clearly defined anti-σ factors with an N-terminal ASD and a C-terminal FecR domain, no anti-σ factor is chromosomally adjacent to genes

encoding FecI-like proteins of group ECF09. Remarkably, FecI-like genes in this group are often flanked by genes encoding MbtH-like proteins (Drake *et al.*, 2007). The function of these small proteins has not been elucidated so far, but a recent publication on *S. coelicolor* indicates that these proteins are involved in regulating peptide antibiotic biosynthesis and siderophore biosynthesis (Lautru *et al.*, 2007). We postulate that these MbtH-like proteins function as a novel type of anti-σ factors for this conserved ECF group. All groups described so far are restricted to Proteobacteria (Table 2.1).

Interestingly, Bacteroidetes ECF σ factors belonging to group ECF10 show substantial similarity to the 'iron acquisition' ECF σ factors, at the sequence and genomic context conservation level. Genes encoding ECF σ factors from this group also form an operon with a *fecR*-like gene and genes encoding outer membrane proteins. The latter are homologous to the polysaccharide-binding outer membrane proteins SusC and SusD (Reeves *et al.*, 1997). Moreover, the genomic context conservation for this group includes downstream genes encoding one or more putative glycosyl hydrolases and other carbohydrate metabolism enzymes. Therefore, these ECF-regulated modules are predicted to enable this human gut symbiont to specifically induce the synthesis of enzymes involved in mobilizing complex polysaccharides from its natural habitat (Xu *et al.*, 2004).

2.2.8.3. ECF11–ECF15 – cytoplasmic-sensing ECF σ factors. In contrast to the RpoE- and FecI-like proteins, the five subgroups ECF11–ECF15 are not closely related to each other (Fig. 2.2). Instead, their unifying feature is their physiological role linked to the presence of soluble anti-σ factors, which suggests they sense cytoplasmic stimuli. The anti-σ factors of groups ECF11–14 harbour N-terminal ZAS domains, and all experimentally investigated members from these groups are involved in some aspect of redox sensing and oxidative stress responses. Group ECF15 is linked to a unique type of NepR-like anti-σ factors and orchestrates the general stress response in α-proteobacteria.

2.2.8.3.1. ECF11. This group is restricted to Proteobacteria and is represented by one of the paradigmatic ECF σ factors, RpoE from *R. sphaeroides*, which is involved in the response to the reactive oxygen species singlet oxygen (Dufour *et al.*, 2008). All ECF σ factors belonging to this group are associated with ChrR-like anti-σ factors (Table 2.3). These anti-σ factors harbour an N-terminal ZAS domain, which binds Zn^{2+} and contacts the σ factor. They also contain a unique C-terminal domain that adopts a cupin-like fold as the site of redox sensing (Campbell *et al.*, 2007).

2.2.8.3.2. ECF12. These SigR-like proteins are predominantly found in actinobacterial species. The eponymous protein from *S. coelicolor*, another ECF model protein, mediates disulphide stress response and mycothiol metabolism (Newton and Fahey, 2008; Paget *et al.*, 2001; Park and Roe, 2008). A second experimentally studied member of this group is SigH, which is involved in heat and oxidative stress response of *C. glutamicum* (Engels et al., 2004; Kim et al., 2005). With three exceptions, all SigR-like proteins are associated with RsrA-like anti-σ factors that harbour a short, conserved C-terminal domain of unknown function, and an N-terminal ZAS domain that responds to oxidative stress by reversible disulphide bond formation (resulting in the loss of the associated Zn ligand). Depending on its state, it tightly binds or releases the cognate σ factor (Li *et al.*, 2002; Zdanowski *et al.*, 2006).

2.2.8.3.3. ECF13. This relatively small group of ECF σ factors (27 proteins) is restricted to Proteobacteria, with only one representative from this group having been mentioned in the literature. Ecf of *N. gonorrhoeae* regulates the methionine sulphoxide reductase MsrAB, indicating a role of this σ factor in oxidative stress response (Gunesekere *et al.*, 2006). The corresponding anti-σ factors are very small proteins with a ZAS domain, but lacking a conserved C-terminal domain (Table 2.3).

2.2.8.3.4. ECF14. This group is restricted to Actinobacteria. It contains SigE, which is responsible for stress response in *M. avium* (Wu et al., 1997). These ECF σ factors group with genes encoding *O*-methyltransferases, which are located directly upstream of, and divergently expressed from the genes encoding the σ/anti-σ factor pair. This genomic context is conserved in 25 out of 26 ECF σ factors belonging to this group (Fig. 2.4).

2.2.8.3.5. ECF15 – EcfG-like. This important and unique ECF group contains 55 ECF σ factors, which are exclusively found in α-proteobacteria (Table 2.1). Recent studies of five members (EcfG1 of *M. extorquens*, EcfG from *B. japonicum*, RpoE2 from *S. meliloti*, SigT from *C. crescentus* and RpoE4 of *R. etli*) strongly suggest that these ECF σ factors represent the long sought-after general stress response of α-proteobacteria (Alvarez-Martinez *et al.*, 2007; Francez-Charlot *et al.*, 2009; Gourion *et al.*, 2009; Martinez-Salazar *et al.*, 2009; Sauviac *et al.*, 2007). The name of this group is derived from the best-

understood example, EcfG1 ('G' indicating its role in *G*eneral stress response) of *M. extorquens* (Francez-Charlot et al., 2009).

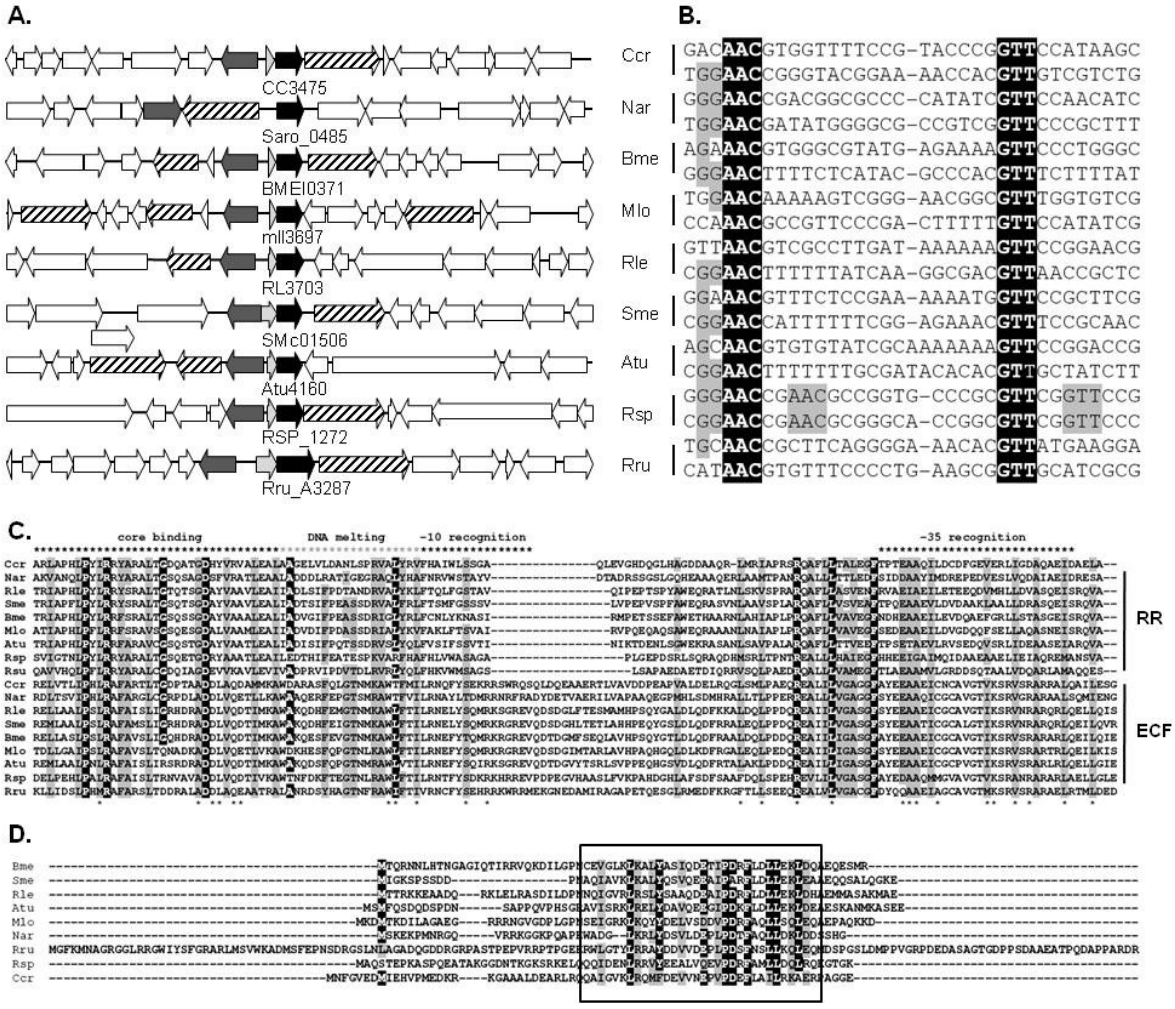

Figure 2.5. Characteristic features of group ECF15. (A) Genomic context conservation. Genes characteristic for this group are marked as coded arrows (black – ECF σ factor; light grey – NepR-like anti-σ factor; dark grey – PhyR-like RR; striped – HK). (B) Alignment of (putative) promoter sequences found upstream from ECF σ factors and the corresponding RR. Conserved motifs are highlighted in black and grey. For each species, the first line represents the promoter upstream of the anti-σ/σ factor operon, and the second line shows the promoter upstream of *phyR* homologues. (C) Multiple sequence alignment of group ECF15 σ factors (bottom) and RR (top). Conserved residues are highlighted in black (identical residues) and grey (similar residues). The regions important for σ factor function are indicated above. Invariant residues of σ factor regions 2 and 4 are indicated below by asterisks. Note that the residues shared by RR and ECF15 σ factors are often not identical to invariant residues in σ^{70} proteins. (D) Multiple sequence alignment of group ECF15 anti-σ factors. Conserved residues are highlighted in black (identical residues) and grey (similar residues); the conserved NepR domain is framed. Abbreviations: Atu, *Agrobacterium tumefaciens*; Bme, *Brucella melitensis*; Ccr, *Caulobacter crescentus*; Mlo, *Mesorhizobium loti*; Nar, *Novosphingobium aromaticivorans*; Rle, *Rhizobium leguminosarum*; Rru, *Rhodospirillum rubrum*; Rsp, *Rhodobacter sphaeroides*; Sme, *Sinorhizobium meliloti*.

The characteristic feature of this group is a conserved genomic context that includes the EcfG-like σ factor, a gene encoding a NepR-like anti-σ factor, a unique type of PhyR-like RR, which functions as the anti-anti-σ factor, and usually sensor HK. Whenever the conserved genomic context is absent, there is at least one additional paralogous ECF σ

factor of the ECF15 group present in the genome that represents the conserved genomic context pattern.

The NepR-like anti-σ factor is encoded by a small gene typically located directly upstream of the ECF σ factor. It is a soluble protein, usually smaller than 50 amino acids (Francez-Charlot *et al.*, 2009; Gourion *et al.*, 2009; Sauviac *et al.*, 2007) that does not bear any sequence similarity to other anti-σ factors. While this ORF is not always annotated, NepR-like anti-σ factors are a conserved feature of the ECF15 group (Fig. 2.5D).

A gene encoding an unusual PhyR-like RR is located upstream of *nepR* and divergently expressed (Fig. 2.5A). This protein harbours the conserved receiver (REC) domain at the C-terminus (instead of the N-terminus, as is the case for almost all other RR; (Galperin, 2006)). Remarkably, the N-terminus shows significant and specific sequence similarity to EcfG-like proteins, as a number of ECF15-specific residues are highly conserved (Fig. 2.5C). However, many invariant core residues in both region σ_2 and σ_4 are not conserved, and the part of region σ_2 responsible for binding the -10 region in the promoter is almost completely absent.

In many cases, gene(s) encoding HK(s) occur in the direct vicinity, although their location relative to the ECF σ factor does not follow a distinct pattern (Fig. 2.5A). In contrast to the clearly conserved NepR and PhyR homologues, these HK are very diverse in sequence and domain architecture, with about 50% of them belonging to the unusual HWE kinases (Karniol and Vierstra, 2004).

The activity of EcfG-like proteins is governed by a phosphorylation-dependent (and therefore reversible) competitive partner-switching mechanism (Francez-Charlot *et al.*, 2009). In the absence of a stimulus, NepR-like proteins bind EcfG-like σ factors, thereby keeping them inactive. After the perception of a signal, possibly via the associated HK, PhyR-like RR become phosphorylated, bind the NepR-like anti-σ factors, thereby ultimately releasing the EcfG-like σ factors to bind RNA polymerase and redirect transcription initiation to the general stress response regulons (Francez-Charlot *et al.*, 2009).

2.2.8.4. Features of the remaining major ECF groups. In addition to the clearly defined groups described above, very little is known about the function of most of the other groups listed in Table 2.1. At best one member has been initially characterized, mostly with regard to the phenotypes and regulons. Six novel ECF groups show remarkable genomic context conservation (Fig. 2.4), and will be described in some detail below. The remaining groups

will only be briefly mentioned (see Tables 2.1 and 2.3, as well as Figs 2.2 and 2.4 for details).

2.2.8.4.1. ECF16 – SigF-like. This group is named after SigF from *C. crescentus*, which is involved in mediating oxidative stress response in stationary phase (Alvarez-Martinez *et al.*, 2006). This ECF group, which is restricted to Proteobacteria, is characterized by the presence of a very unusual potential anti-σ factor, a protein harbouring a membrane-anchored (six TMR) DUF1109 domain of unknown function.

2.2.8.4.2. ECF17 – SigU-like. This actinobacterial ECF group likely participates in protein secretion, as evidenced by studies of *S. coelicolor* SigU and *Mycobacterium tuberculosis* SigL. The exact role of *S. coelicolor* SigU is not known, yet it is involved in the regulation of protein secretion (Gordon *et al.*, 2008). SigL plays an important role in the pathogenesis of *M. tuberculosis*. It is involved in the modification of secreted and membrane proteins, and also regulates the expression of polyketide synthases genes (Hahn *et al.*, 2005). The anti-σ factors of this group harbour an N-terminal ZAS domain, and a unique conserved C-terminal domain potentially involved in stimulus perception.

2.2.8.4.3. ECF18 – RpoT-like. This group is named after the *P. putida* ECF factor, RpoT, which plays a role in the tolerance of toluene and other organic solvents (Duque *et al.*, 2007). This proteobacterial ECF group (58 out of 59 members; one is from Acidobacteria) shares a characteristic anti-σ factor domain architecture with the SigK-like proteins (ECF19). The corresponding putative transmembrane proteins (one TMR) harbour an N-terminal ASD and a C-terminal cog5343 domain of unknown function.

2.2.8.4.4. ECF19 – SigK-like. This σ factor is responsible for the expression of two antigenic proteins in *M. bovis* BCG strains (Charlet *et al.*, 2005). The corresponding ECF group is restricted to Actinobacteria. The cognate anti-σ factors share the same domain architecture with proteins from ECF18, but their overall sequence similarity to RskA (the anti-σ factor for SigK) allows a clear discrimination to those from the previous group.

2.2.8.4.5. ECF20. This ECF group occurs predominantly in Proteobacteria and includes CnrH from *Cupriavidus metallidurans*. This ECF σ factor was investigated in a screen for metal-responsive ECF σ factors and shown to be upregulated in response to Ni(II) and

essential for nickel resistance (Grosse *et al.*, 2007). These ECF σ factors are associated with short membrane-anchored (one TMR) anti-σ factors that harbour either an N-terminal ASD or a CAS domain, but lack a conserved C-terminal domain. In addition, genes encoding CnrXY-like or von Willebrand factor A-containing proteins are also found in direct vicinity to genes encoding ECF20 proteins.

2.2.8.4.6. ECF21–25. None of the members from any of these five phylogenetically diverse groups has been investigated. Two groups (ECF22 and 24) lack an obvious anti-σ factor, while those from the remaining three groups are potentially membrane-anchored proteins (one TMR) that contain an N-terminal ASD (ECF21 and ECF25) or a CAS domain (ECF23) and a conserved C-terminal domain.

2.2.8.4.7. ECF26. This large group, while clearly defined by the sequence similarity of the σ factors, shows a remarkable combinatorial complexity with respect to both genomic context and anti-σ factor conservation. It is restricted to Proteobacteria except for one instance (Tables 2.1 and 2.2). Three possible variations of gene neighbourhood can be observed. Twenty-seven ECF σ factor genes form putative operons with upstream genes encoding lipoproteins. Another 15 ECF σ factors group with a cytochrome *c* oxidase and a metallophosphoesterase (as does SigE from *Starkeya novella*, responsible for regulating thiosulphate oxidation (Kappler *et al.*, 2001)). Fourteen of them group with genes encoding putative catalases and type *c* cytochromes. Although the genomic context conservation seems to indicate three distinct groups of ECF σ factors here, the clear sequence similarity strongly suggests that those ECFs indeed belong to only one large group.

These differences in conserved genomic contexts correlate well with the subclustering of the corresponding anti-σ factors. The typical anti-σ factors contain an N-terminal ASD, one TMR and a conserved C-terminal periplasmic domain of unknown function (COG-identifier: cog5662; see Table S2.5 for details). Clustering analysis allows the discrimination of at least four subtypes. One type of ECF26 anti-σ factors contains ZAS proteins (Zdanowski *et al.*, 2006). The second subtype shows significant overall similarity to PrtR, the anti-σ factor of PrtI, which is involved in the temperature-dependent lipase and protease production in *P. fluorescens* (Burger et al., 2000). The third (and major) subtype contains the remaining anti-σ factors with ASD–TMR–cog5662 domain architectures. Ten ECF σ factors of group ECF26 are linked to an altogether different type of anti-σ factors

that possesses a unique C-terminal periplasmic extension instead of the cog5662 domain (Table S2.5).

Thus, the combinatorial complexity of group ECF26 is based on a correlated anti-σ factor/genomic context clustering within a distinct ECF group, indicative of a functional diversification of one type of ECF σ factors during evolution.

2.2.8.4.8. ECF27. SigM from *C. glutamicum* is involved in the response to various stresses, including heat/cold shock, and disulphide stress in the presence of diamide (Nakunst *et al.*, 2007). ECF σ factors from this group are restricted to Actinobacteria. The cognate anti-σ factors harbor an N-terminal ASD and a group-specific conserved C-terminal domain, separated by a single TMR.

2.2.8.4.9. ECF28. ECF σ factors from this proteobacterial group of unknown function are also linked to membrane-anchored anti-σ factors with an ASD and a group-specific C-terminus.

2.2.8.4.10. ECF29. There are two variants of gene neighbourhood in this group: ECF σ factors group either with genes encoding an outer membrane protein and a blue copper protein (13 ECF σ factors), or with a putative ferritin-like exported protein and cytochrome *c* (seven proteins). None of the ECFs belonging to this group has been experimentally analysed. This group occurs predominantly in Proteobacteria.

2.2.8.4.11. ECF30. This large group of unknown function is almost exclusively found in Firmicutes, and includes *B. subtilis* σ^V (Zellmeier *et al.*, 2005). An anti-σ factor is only present in about 50% of all group members, in which case it contains an N-terminal ASD or CAS domain and a group-specific C-terminus. Nine members are linked to LstR/PadR-like transcriptional regulators, instead of an anti-σ factor.

2.2.8.4.12. ECF31. This small and phylogeneticaly diverse group is of unknown function and includes *B. subtilis* σ^Y (Cao *et al.*, 2003). Most associated anti-σ factors harbour two TMR and an N-terminal ZAS domain.

2.2.8.4.13. ECF32 – HrpL-like. This small but unique ECF group lacks an anti-σ factor. Instead, HrpL-like ECF σ factors are genomically linked to HrpX-like sensor kinases. The

activity of these proteins is regulated at the transcriptional level by the activity of a 2CS (Merighi *et al.*, 2003; Nizan-Koren *et al.*, 2003). This situation is reminiscent of *S. coelicolor* σ^E, which is regulated by the CseBC 2CS (Paget *et al.*, 1999b) and belongs to a different ECF group (ECF39). HrpL is conserved in proteobacterial plant pathogens, including *E. amylovora*, *Pantoea stewartii*, and in *P. syringae*. It is involved in regulating the expression of *hrp* gene cluster encoding a type III secretion system, and is required to elicit the hypersensitive response and to cause disease by this plant pathogen (Ferreira *et al.*, 2006; Lindeberg *et al.*, 2006; Merighi *et al.*, 2003; Nizan-Koren *et al.*, 2003; Wei and Beer, 1995).

2.2.8.4.14. ECF33/34. No example of either of these groups has been investigated so far. While ECF33 is restricted to Proteobacteria, ECF34 occurs in five different phyla, but predominantly in Cyanobacteria. Both groups contain characteristic membrane-anchored anti-σ factors (Table 2.3).

2.2.8.4.15. ECF35. This group consists solely of proteobacterial ECF σ factors. Nine out of 14 ECF σ factors belonging to this group are organized in an operon with a gene encoding a protein that contains 'von Willebrand factor type A' domains. As none of the ECF σ factors has been studied experimentally, and the role of the 'von Willebrand factor type A' domains in bacteria remains elusive (Whittaker and Hynes, 2002), the function of ECF σ factors from this group is unclear.

2.2.8.4.16. ECF36–40. Members from these ECF groups are found solely in Actinobacteria, with the exception of ECF37, which is restricted to Proteobacteria. Only three proteins have been partially characterized. SigC from *M. tuberculosis* (ECF36) has been described as an important regulator of pathogenesis and adaptive survival of this bacterium in the lung and spleen (Karls *et al.*, 2006). SigD from the same organism (ECF40) has also been linked to pathogenesis and regulates a small set of ribosomal genes and ATP transporters (Calamita *et al.*, 2005). σ^E (ECF39) mediates cell envelope stress response in *S. coelicolor*. As with the HrpL-like proteins described above (ECF32), this specific ECF σ factor is regulated at the transcriptional level by a 2CS and a small lipoprotein conserved in Actinobacteria (Hutchings *et al.*, 2006; Paget *et al.*, 1999a; Paget *et al.*, 1999b). While some other members of this ECF group are also genomically associated with genes encoding 2CS, this does not seem to be a common mechanism for

members of this group, since most ECF σ factors are associated with conserved, although uncommon, putative anti-σ factors (Table 2.3). With the exception of ECF36, all these groups contain potential anti-σ factors, which often lack the characteristic domains (see Table 2.3 for details). None of the five groups shows any extended genomic context conservation.

2.2.8.4.17. ECF41. This group consists of over 100 ECF σ factors from five different phyla and none is associated with an anti-σ factor. Instead, ECF genes are found directly upstream or downstream of genes encoding either a CMD protein (carboxy-muconolactone decarboxylase) or an oxidoreductase (Fig. 2.4). The only ECF σ factor from this group that has been experimentally investigated is SigJ from *M. tuberculosis*. This ECF σ factor is proposed to play a role in hydrogen peroxide resistance (Homerova *et al.*, 2008; Hu *et al.*, 2004). All ECF σ factors from this group contain a characteristic and highly conserved C-terminal extension of about 100 amino acids. Due to the observed absence of an anti-σ factor within this group, we speculate that this C-terminal extension might play a role in stimulus perception and/or regulating the activity of ECF σ factors belonging to this group.

2.2.8.4.18. ECF42. This group contains unusual ECF σ factors, which are significantly longer than traditional ECF σ factors (around 400 amino acids) and contain a tetratrico peptide repeat (TPR) domain, which is postulated to be important for protein–protein interactions (Blatch and Lassle, 1999). One hundred and six out of 111 proteins present in this group are found in the direct vicinity of one or more copies of genes encoding so-called DGPF proteins, the function of which is not yet known (see Table S2.10 for a description of protein domains).

2.2.8.4.19. ECF43 – ECF-like proteins. This group contains 36 proteins that, although very similar to ECF σ factors and therefore fulfilling the initial search criteria used to generate the data set underlying this classification, are not 'true' ECF σ factors (see phylogenetic tree, Fig. 2.2). Multiple sequence alignments of these σ factors with sequences of 'true' ECF σ factors reveals the presence of five and three additional residues in the conserved parts of regions σ_2 and σ_4 respectively (Fig. S2.1). This group shows a conserved genomic context conservation, with neighbouring genes encoding Ser/Thr protein kinases.

2.9. Identification of group-specific ECF target promoter motifs

ECF σ factors drive differential gene expression by binding alternative promoters and redirecting transcription initiation after recruiting the RNA polymerase core enzyme. Intuitively, we expect that homologous σ factors (i.e. from within one distinct ECF group) will recognize similar promoter motifs, as has been demonstrated for the proteobacterial RpoE-like proteins (ECF02) (Rhodius *et al.*, 2006), and singlet oxygen responsive ECF σ factors (Dufour *et al.*, 2008) assigned to group ECF11. Since the genes encoding most ECF σ factors are autoregulated, it is possible to identify the corresponding target promoters by screening the upstream regions of ECF operons of a specific ECF group for an over-represented bipartite sequence motif (Fig. 2.1). We constrained our analysis to those groups, for which there was at least one experimentally identified target promoter to serve as a sequence motif for screening. Sixteen ECF groups fulfilled these criteria, and we successfully identified group-specific ECF-target promoters based on the experimentally verified example in the upstream regions of ECF genes for all of them, which again supports the relevance of our classification. Sequence logos generated with the WebLogo tool (Crooks *et al.*, 2004) of the corresponding promoter motifs are shown in Fig. 2.6; the underlying complete data set can be found in Table S2.8.

The hallmark feature of an ECF-type promoter, the 'AAC' motif in the -35 region (Helmann, 2002; Lane and Darst, 2006), exists in roughly two-thirds of the promoter motifs. This motif is typical for RpoE-like proteins, and can also be found in EcfG-, SigF-, SigU- and HrpL-target promoters, but other groups show more or less pronounced alterations from this standard (Fig. 2.6). Nevertheless, there is a clear bias towards GAAC motifs in the -35, while the -10 region contains highly conserved G and T residues. The one exception is the target promoter of SigK-like ECF σ factors (ECF19), which was mapped and predicted in numerous mycobacterial species (Veyrier *et al.*, 2008). This promoter has very little in common with typical ECF-type promoters. Indeed, the motif looks as if -35 and -10 regions have exchanged their positions (Fig. 2.6).

Interestingly, the core sequence motif of the group ECF15 target promoter, which has been experimentally verified for three bacterial species (Alvarez-Martinez *et al.*, 2007; Gourion *et al.*, 2008; Sauviac *et al.*, 2007), represents a perfect inverted repeat, AAC-$N_{17\text{-}18}$-GTT. EcfG-like σ factors orchestrate the expression of two divergently expressed transcriptional units, encoding the anti-σ/σ factor pair and PhyR-like RR respectively. The identified promoter motif, which is usually found upstream of both loci, is very often the identical

sequence motif on the two opposite strands (Fig. 2.5B). This raises the possibility of a bidirectional ECF-type promoter that can be recognized on both DNA strands.

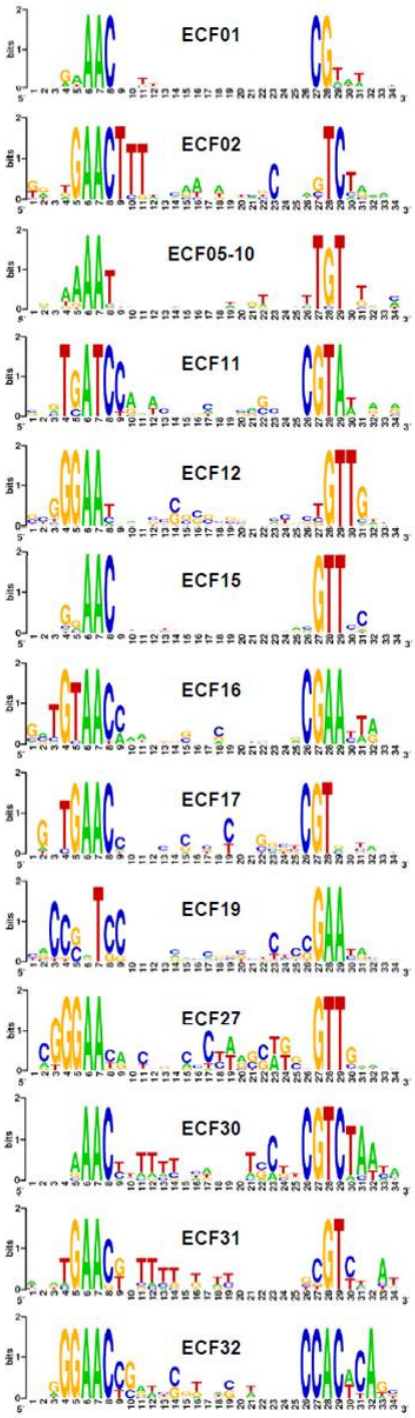

Figure 2.6. Weblogos of selected group-specific ECF target promoters. The weblogos illustrate the degree of sequence conservation for the group-specific target promoter motifs by graphically representing the corresponding position weight matrices. These matrices were derived from a representative selection of mapped promoters from within each group and based on the overall size of the group. Whenever possible, representatives from different phyla (or at least taxonomically distant organisms from within one phylum) were used to ensure sequence diversity in our analysis. Two hundred and fifty base pairs directly upstream of the start codon of the first gene of the ECF/anti-σ factor operons were extracted from the MicrobesOnline database (Alm *et al.*, 2005). Table S2.8 contains the list of all promoter sequences used in this study. For the FecI-like groups, the promoter regions upstream of the *fecA* homologues were chosen for analysis because they are the known target operon of FecI-like ECF σ factors (Braun *et al.*, 2006). The *trxB* promoter region (Nakunst *et al.*, 2007) was chosen for group ECF27. The selected promoter regions of each group were analysed for the presence of putative group-specific target promoters, both manually and with the Virtual Footprint (Münch *et al.*, 2005) algorithm, which is embedded in the Prodoric database (http://prodoric.tu-bs.de/vfp/). These results are based on regular expressions of the sequence motifs derived from the experimental work. The resulting motifs were illustrated based on multiple sequence alignments, using the WebLogo tool (Crooks *et al.*, 2004).

Despite the conserved genomic context and therefore known target region of FecI-like σ factors, the promoter described for FecI of *E. coli* and *P. aeruginosa* could only be identified in one-third of all *fecA* promoter regions. The motif shown in Fig. 2.6 represents the consensus of those target promoters bearing a FecI-like promoter. The successful identification of group-specific promoter motifs – as demonstrated here for 16 ECF groups – provides the basis for genome-wide screens for ECF regulons.

2.2.10. Nomenclature of ECF σ factors

The ubiquitous distribution, abundance and phylogenetic diversity of ECF σ factors in bacterial genomes have also important implications on their future nomenclature.

Classically, genes encoding novel (alternative) σ factors were commonly named by '*rpo*' or '*sig*' designators, according to the 'standards' for Gram-negative and Gram-positive bacteria respectively. In many cases, newly identified ECF σ factors were labelled according to the names of the first described systems (e.g. SigE or RpoE). The limitations of this approach are quite evident when dealing with microbial genomes that contain substantially diverse and abundant ECF σ factors, which can easily outnumber all other σ^{70} proteins in a given genome. The functional and sequence diversity of ECF σ factors, as outlined in this article, argues against a continuing use of a few common names, such as RpoE, for newly described ECF σ factors, irrespective of a true similarity to the specific ECF group.

We propose a nomenclature that uses the information stored in our classification, to choose more appropriate gene/protein names for a novel ECF protein. In this system, the name is based on the paradigm of each group. For example, the name 'RpoE' would only be justified for ECF σ factors belonging to group ECF02 etc. Multiple representatives within a group and present in a single genome should be sequentially numbered (e.g. RpoE1, RpoE2 etc.). With this approach, the name would also serve as an indication of the group's potential function. Novel ECF σ factors, either belonging to uncharacterized groups or unclassifiable, will be differentiated by labelling genes as *ecf″X″* to directly distinguish the gene product from other σ factors. Accordingly, the protein itself would harbour the name Ecf″X″ or $\sigma^{Ecf''X''}$. This nomenclature has already been used in recent publications on novel ECF σ factors in *M. extorquens* and *B. licheniformis* (Francez-Charlot et al., 2009; Wecke et al., 2006).

2.2.11. Summary and outlook: resources to study ECF σ factors, the third pillar of bacterial signal transduction

The data in this article provide ample evidence of the wide distribution, modular design and combinatorial complexity of ECF-dependent signal transduction. Our analysis establishes ECF σ factors as a central signalling mechanism in bacteria, outranked only by the already well-known modes of 1CS and 2CS signalling. While 1CS are the most abundant mechanism of signal transduction in prokaryotes, they are almost exclusively used for intracellular signalling (Ulrich *et al.*, 2005). Approximately half of all sensor kinases belonging to 2CS are membrane-anchored, indicating that they sense extracellular stimuli (Mascher *et al.*, 2006). In contrast, the majority of ECF σ factors seem to respond

to environmental cues via membrane-anchored anti-σ factors. Only a few ECF groups (ECF11–15) are apparently linked to intracellular stress inducers (Tables 2.1 and 2.3).

To facilitate and simplify future studies on ECF σ factors, we have developed a simple, fast and freely available online tool called ECF*finder* (http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder), based on group-specific HMMs derived from our classification. This tool allows the user to identify and classify ECF σ factors in any submitted protein sequences, including complete microbial (meta)proteomes. Moreover, these HMMs were also used to incorporate ECF σ factors into the MiST database (Ulrich and Zhulin, 2007) at http://genomics.ornl.gov/mist/. Our ECF data sets (Tables S2.2–S2.7) and the identified group-specific ECF target promoter motifs (Fig. 2.6 and Table S2.8) provide comprehensive resources to guide future research in studying ECF-dependent signal transduction and gene regulation.

Taken together, our classification of ECF σ factors will be of value to those members of the scientific community interested in studying ECF-dependent gene regulation by providing a comprehensive (and expanding) resource of information to support and direct experimental studies on this third pillar of bacterial signal transduction.

2.3. Experimental procedure

For our analysis, we only used freely available algorithms and databases (see references and links in the following paragraphs) at default settings, unless stated otherwise.

2.3.1. Sequence retrieval

The analysed set of sequences was based on the protein database in 'Normal' SMART (Letunic *et al.*, 2006; Schultz *et al.*, 1998) (http://smart.embl-heidelberg.de/; release No. 5), which contains SWISSPROT, SP-TrEMBL and stable Ensembl proteomes. To identify ECF σ factors, a domain selection-based SMART search on the bacterial subset of the database was performed. Selected sequences lack the Pfam:Sigma70_r3 (PF004539) domain, and harbour at least one of the following domains: Pfam:Sigma70_r2 (PF004542), Pfam:Sigma70_r4 (PF004545), Pfam: Sigma70_r4_2 (PF008281).

The initial output obtained from SMART search was manually corrected with the following criteria: (i) proteins with signal peptides and/or domains other than Pfam: Sigma70_r2 and Pfam:Sigma70_r4/Pfam:Sigma70_r4_2 were excluded from the analysis; and (ii) proteins with more than 50 amino acids between the region σ^2 and region σ^4 (indicative of the presence of a weakly conserved region σ^3 domain, below threshold for graphical representation in SMART) were also classified as negatives.

2.3.2. Classification of ECF σ factors by multiple sequence alignments and genomic context clustering

The initial grouping was performed manually on the complete multiple sequence alignment, based on CLUSTALW (Thompson *et al.*, 1994) from http://www.ebi.ac.uk/clustalw/, using complete protein sequences as input and the output option 'sorted by sequence similarity'. Subsequently, the alignment was

divided into sets of roughly 100 sequences, with each set overlapping by at least 10 sequences with the previous set. Unrooted trees were constructed from these multiple sequence alignments by using the Phylip (Felsenstein, 1989) programs PROTDIST and FITCH, implemented in BioEdit Sequence Alignment Editor (Hall, 1999) (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Phylogenetic trees (as represented by Fig. 2.2) were constructed from gapless multiple sequence alignments of only the conserved regions σ^2 and σ^4, to allow the calculation of meaningful phylogenetic distances between the individual proteins (Baldauf, 2003). Genomic context analysis was performed using ERGO (Overbeek *et al.*, 2003) at http://ergo.integratedgenomics.com/, The SEED (Overbeek *et al.*, 2005) at http://theseed.uchicago.edu/FIG/index.cgi, and MicrobesOnline (Alm *et al.*, 2005) at http://www.microbesonline.org.

2.3.3. Identification and classification of anti-σ factors

This analysis was performed basically as described previously (Campbell *et al.*, 2007). In brief, the protein sequences of genes located next to and presumably co-transcribed with the ECF genes were extracted from the genome databases to generate the anti-σ factor data set (see Tables S2.5–S2.7). An initial prescan, based on the genomic context and multiple sequence alignments of all anti-σ factors, allowed the identification of unusual potential anti-σ factors that contained conserved domains of unknown function, such as FecR-like anti-σ factors or those cognate to groups ECF15 and ECF16. The remaining candidate anti-σ sequences were divided into domains, based on the location of the TMR, when available. The domains of soluble anti-σ factors were identified based on the results previously described (Campbell *et al.*, 2007). The N- and C-terminal domains were then subjected to clustering and multiple sequence alignments to allow subgrouping and the identification of ASD and ZAS domains, as described (Campbell *et al.*, 2007). Finally, a combined clustering analysis of σ factors, anti-σ factors and genomic context was performed using the SimilarityBox tool, which can be freely downloaded from https://www.biopilot.org/docs/Software/SimilarityBox.php.

2.3.4. Construction and evaluation of HMM models

The HMMs have been constructed by using the appropriate programs from HMMER2, version 2.3.2 (available at http://hmmer.janelia.org). The programs have been employed using the default parameters. Initially, the data set of ~2700 ECF sequences described above was sorted by sequence alignments into groups. This initial grouping was used to generate a first set of HMMs based on whole-protein sequence alignments of the group members, generated with CLUSTALW, version 2.0.10 (Larkin *et al.*, 2007). The multiple alignments were converted into HMMs and calibrated using the programs HMMBUILD and HMMCALIBRATE. Subsequently, the HMMs were tested against the initial data set, using HMMSEARCH. Based on this test, the ECF groups have been resorted for proteins which had their best hit to another model instead to the one of their group. Finally, the resorted groups of ECFs have been used to construct an improved set of HMMs. Based on a screen against the initial data, scores for trusted, gathering and noise cut-offs have been defined as follows. (i) A trusted score is considered as the lowest score generated by a true member of the corresponding ECF group. (ii) A gathering score has been defined as the lowest score any other true ECF sequence has produced. (iii) The noise cut-off is the value with the lowest score any known ECF-like sequence has generated with the respective HMM. The improved set of group specific HMMs has been evaluated concerning the discriminative power, their sensitivity and time performance by screening for ECF σ factors in single genomes, the entire SWISSPROT division of the UniProt database (version 14.8, at http://www.uniprot.org/news) (The UniProt Consortium, 2009) and the annotated proteins of the Sargasso sea metagenome (Venter *et al.*, 2004), available at http://www.g2l.bio.uni-goettingen.de/ECFfinder/sargasso.peptide.fa.gz. The complete data set of the Sargasso sea metagenome scan is available for download at http://www.g2l.bio.uni-goettingen.de/ECFfinder/SargassoSeaScanResults.zip.

2.3.5. Implementation of ECF*finder*

ECF*finder* is designed as a WWW tool for the identification and characterization of putative ECF σ factors in any set of protein sequences. It has been written in Java Version 1.6 (http://java.sun.com/). The integration in an apache web-server has been implemented in TOMCAT (http://tomcat.apache.org/). The tool accepts user data in FASTA format and employs ECF-specific HMM models in a three-step procedure to scan for and characterize ECF σ factors. For performance reasons a 'screening ECF HMM' has been constructed based on the typical proteins of the major ECF groups. The model identifies all ECFs of the major groups. We use this model in the ECF*finder* as a filter for major group ECF σ factors, which in a second step are classified by the group-specific models. This two-step procedure reduces the time needed to find and characterize all ECF σ

factors within a typical bacterial genome approximately by a factor of 40. For instance the time to screen the complete protein data set of *B. licheniformis* DSM13 (Veith *et al.*, 2004) has been reduced from 1380 to 30 s by employing the two-step approach instead of a single-step procedure of all proteins with all major group HMMs. However, this general model does not identify most of the singleton ECF σ factors, including some known ECF σ factors. Therefore, a 'general ECF HMM', based on an alignment of all singleton ECF σ factors, has been constructed. The model identified all of the known ECF proteins, in addition to some non-ECF σ factors, like the sporulation σ factor of *Bacillus* spp. The scan with this model for ECF-like sequences is implemented as a third step in ECF*finder* to search for new ECF σ factors and for ECF-like sequences. As a stress test for ECF*finder*, we performed a scan on the entire protein data set of the Sargasso sea metagenome (Venter *et al.*, 2004) which took approximately 20 min. All tests have been run on a computer with two quadcore CPUs and 32 GByte RAM under SUSE-linux. ECF*finder* is open for public access via http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder/ or through the software portal of the Göttingen Genomics Laboratory (G2L) at http://www.g2l.bio.uni-goettingen.de/software/f_software.html.

2.3.6. Identification of group-specific target promoters

At least 10 representative upstream regions of ECF operons from each ECF group (250 bp upstream the start codons; see Table S2.8 for the corresponding data set) were extracted from the MicrobesOnline database (see above) and subsequently subjected to motif searches, by using the MEME suite (Bailey and Elkan, 1994) at http://meme.sdsc.edu/meme4_1/intro.html, and the Virtual footprinting algorithm (Münch *et al.*, 2005) embedded in the Prodoric database at http://prodoric.tu-bs.de/vfp/. The parameter settings for the MEME analysis were: distribution (zero or one), width (minimum 3, maximum 30; this width allows the identification of either the complete promoter or one part of the bipartite target promoter), optional parameters (search given strand only). For the Virtual Footprint, the 'regulon prediction' mode was chosen using a FASTA file containing all promoter regions from a given ECF group as the input DNA sequence. As the input pattern, we used the known target promoters, either as a single or a bipartite pattern, expressed in IUPAC code (allowing one or two mismatches) or as a regular expression. The resulting sequence motifs were illustrated based on Multiple sequence alignments, using the WebLogo tool (Crooks *et al.*, 2004) at http://weblogo.berkeley.edu/.

2.4. Acknowledgements

Work in the authors' labs was supported by grants from the Deutsche Forschungsgemeinschaft (DFG Grant MA2837/2-1, to T.M. and H.L.), the Fonds der Chemischen Industrie (to T.M.), and the BMBF-funded network BiotechGenoMik (to H.L.). The KIT Research Group 11-1 received financial support from the 'Concept for the future' of the Karlsruhe Institute of Technology within the framework of the German Excellence Initiative (to T.M.). The authors would like to thank Tina Wecke and Yann Dufour for critical reading of the manuscript. We would also like to express our sincerest gratitude to the developers of the databases used in this study, for providing the scientific community with versatile, powerful and freely available online tools that allow gaining so much insight from sequences and genomes, yet can even be used by 'normal' biologists without any background in bioinformatics.

CHAPTER 3

General stress response in α-proteobacteria: PhyR and beyond

This chapter has been adapted from:

A. Staroń and T. Mascher, T.

Mol. Microbiol. (2010) 78(2):271-277

Chapter 3
General stress response in α-proteobacteria: PhyR and beyond

In addition to stress-specific responses, most bacteria can mount a general stress response (GSR), which protects the cells against a wide range of unspecific stress conditions. The best-understood examples of GSR are the σ^B-cascade of *B. subtilis* and the RpoS response in *E. coli*. While the latter is conserved in many other Proteobacteria of the β-, γ- and δ-clades, RpoS homologues are absent in α-proteobacteria and their GSR has long been a mystery. Recent publications finally unravelled the core of the GSR in this proteobacterial class, which is mediated by EcfG-like σ-factors. EcfG activity is controlled by NepR-like anti-σ factors and PhyR-like proteins that act as anti-anti-σ factors. These unusual hybrid proteins contain an N-terminal EcfG-like domain that acts as a docking interface for NepR, and a C-terminal receiver domain typical for bacterial RR. Upon phosphorylation, PhyR titrates NepR away from EcfG, thereby releasing the σ-factor to recruit RNA polymerase and initiate transcription of its target genes. In issue 78 of *Molecular Microbiology*, Herrou *et al*. describe the function and three-dimensional structure of PhyR from *C. crescentus*. This structure is key to understanding the mechanism of the reversible, phosphorylation-dependent partner switching module that orchestrates the GSR in α-proteobacteria.

3.1. Introduction

Bacterial life in its natural context stands in stark contrast to the defined laboratory situation, where pure bacterial cultures are grown under fixed and usually optimal conditions. Out in the wild, physicochemical parameters vary greatly, as does the nutrient supply. Moreover, bacteria have to struggle with countless competitors from various species for the available resources and ecological niches. The ability to adapt to the changing conditions present in their natural habitats is a central aspect but also one of the biggest challenges of bacterial life. Adaptation requires the accurate monitoring of critical parameters and a precise and specific information flow in order to mount an adequate response.

Bacteria have evolved a number of basic mechanisms to facilitate signal transduction. 1CS, such as TetR or LacI, combine input and output functions in a single protein and predominantly respond to intracellular signals (Ulrich *et al.*, 2005). In contrast, two major classes of phylogenetically unrelated mechanisms allow bacteria to respond to extracellular

cues. In two-component systems, consisting of a HK that acts as a sensor and a corresponding RR that mediates the output (usually by acting as a transcriptional regulator), signal transduction is based on reversible phosphoryl group transfer (Gao and Stock, 2009). In contrast, the activity of so-called ECF σ factors is controlled by their cognate anti-σ factors through direct protein-protein interactions: in the presence of inducing conditions, the anti-σ releases its grip, thereby allowing the σ-factor to bind alternative target promoters and induce expression of the downstream target genes (Butcher *et al.*, 2008; Helmann, 2002).

Most bacteria, especially those living in complex environments, harbour many (often more than 100) such signal transducing systems to adapt to a plethora of different conditions (Ulrich *et al.*, 2005). Typically, most of these systems respond to a specific stimulus by an adequately specific cellular output. In addition to mounting such specific responses, bacteria also have to adjust their physiology to an overall decline in environmental conditions, caused by hunger, overpopulation or changes in crucial physical parameters that affect many cellular processes simultaneously.

Under laboratory conditions, such global adjustments can, for example, be observed at the onset of stationary phase, when the overall living conditions deteriorate: with cell density reaching critical levels, nutrients start running low, and toxic metabolic by-products begin to accumulate. In the environment, where logarithmic growth rarely occurs, these periods of hardship might be the rule rather than the exception. In bacteria living in complex, quickly changing habitats such as the soil or the phyllosphere, these general protection responses often also facilitate adaption to sudden stress conditions such as heat or oxidative stress that affect the cellular physiology at many different levels and therefore necessitate an overall adjustment of the bacterial physiology. Therefore, these responses are important both for long-term gradual adjustments, but also for quick and transient adaptations to protect the cells. Because of this broad spectrum of inducing conditions and the complexity of the resulting output, perhaps affecting the expression of hundreds of target genes, such adjustments have been termed ‘general stress responses’ (GSR).

From a signalling point of view, GSR requires the computation and integration of numerous different input signals in order to mediate a graded response that adequately protects and stabilizes the cell. Not surprisingly, the underlying regulatory cascades mediating GSR are often very complex, and may involve the activity of dozens of regulatory proteins that perceive stimuli, integrate signals and transduce information to mediate the desired output. Here, we will first give a brief overview of the two best-

understood GSR networks, to highlight the variability but also to extract some common themes involved in bacterial GSR.

3.2. Paradigms of bacterial GSR: *E. coli* RpoS and *B. subtilis* σ^B

One important common theme that emerges from these two well-investigated examples is that the ultimate function of both signalling networks is to control the activity of a single alternative σ-factor. In this sense, 'activity' basically means how much of this specific GSR σ-factor is freely available to compete with the remaining cellular pool of other σ-factors for RNA polymerase core enzyme in order to induce gene expression of its target promoters. But the way that this activation is achieved differs dramatically between these two paradigms.

3.2.1. RpoS of *E. coli*

This alternative σ-factor is closely related to the primary (or vegetative) σ-factor. It is unique in the sense that its activity is not regulated by a typical anti-σ factor, in contrast to most other stress-inducible alternative σ-factors. Instead, its activity is tightly regulated at the level of transcription, translation and especially protein turnover (Hengge, 2010). RpoS is gradually induced upon entry into stationary phase to ensure long-term survival. This fine-tuned mechanism of activation is achieved by a combination of regulation at all the three levels mentioned above. Moreover, RpoS also quickly responds to sudden stresses, such as heat shock, hyperosmotic or acid stress, UV irradiation or sudden starvation. These fast responses are mainly mediated by a rapid inhibition of RpoS proteolysis, leading to a fast increase in the cellular pool of this σ-factor. The central player for this stress-inducible activation is the adaptor protein RssB, an unusual RR, which delivers RpoS to the ClpXP protease under non-inducing condition. This protein therefore represents a functional analogue to a classical anti-σ factor. Its activity is negatively affected by at least three anti-adaptors, which titrate RssB away from RpoS in response to various sudden stresses, such as phosphate starvation (IraP), magnesium starvation (IraM), as well as oxidative stress and DNA damage (IraD). Hence, these proteins represent analogues to the anti-anti-σ factors, which are present in the GSR described below.

RpoS controls the expression of more than 500 genes that are involved in stress-protective functions, alternative metabolic pathways, cell envelope biosynthesis and cell shape, biofilm formation and pathogenicity. RpoS also integrates input signals by cooperating

with various transcription factors, thereby participating as a central player in a complex regulatory network. Moreover, RpoS-dependent gene regulation is also strongly affected by second messengers such as cAMP, ppGpp or c-diGMP. Further details on various aspects of this σ-factor can be found in a number of recent reviews (Hengge, 2008, 2009b, 2010).

3.2.2. σ^B of *B. subtilis*

In contrast, activity of the σ^B GSR regulator from *B. subtilis* (and many other Firmicutes bacteria) is ultimately regulated by a controlled and reversible release from its anti-σ factor RsbW. The major signalling features of this GSR are phosphorylation-dependent partner-switching modules mediated by opposing kinase/phosphatase pairs that integrate input signals from two distinct pathways in response to environmental or energy stress (Fig. 3.1, right side) (Hecker *et al.*, 2007; Price, 2010). Signal integration occurs at the level of the anti-anti-σ factor RsbV, which gets dephosphorylated by the two phosphatases RsbP (energy pathway) and RsbU (environmental pathway) in the presence of stress conditions (Fig. 3.1). Active RsbV than titrates RsbW (anti-σ factor) away from σ^B, thereby ultimately releasing the σ-factor from its inhibitory grip to redirect transcription initiation to the GSR regulon. In the absence of stress, the anti-σ factor RsbW actively phosphorylates (and thereby inactivates) RsbV. This mechanism is also used to reset the system to the inactive pre-stimulus state.

The σ^B response is induced by a plethora of signals, such as entry into stationary phase, addition of uncouplers of the proton motive force (such as CCCP and nitric oxide), starvation for carbon, phosphate or oxygen (energy pathway). Moreover, environmental stresses, such as acid, ethanol, heat or osmotic stress and blue light irradiation also trigger the GSR (environmental pathway). Deletion of *sigB* leads to increased sensitivity to some, but not all inducing conditions. σ^B controls about 150–200 genes, only few of which encode obvious stress-related functions. In similarity to the situation described for RpoS above, many σ^B-dependent genes are co-regulated by other transcription factors (Hecker *et al.*, 2007; Price, 2010).

3.3. GSR in α-proteobacteria: distribution and conservation of the PhyR–NepR–EcfG cascade

Comparative genomic analyses have revealed that homologues to core components of these two different types of GSR networks can be found in many bacterial genomes. On the other hand, some phylogenetic groups such as the α-proteobacteria completely lack any such regulators, and how the GSR was mediated in this bacterial class was unclear for a long time. Recent reports have begun to reveal the nature of the α-proteobacterial GSR. Groundbreaking studies from the group of Julia Vorholt identified the central players of the GSR response in the methylotrophic bacterium *M. extorquens* (Francez-Charlot et al., 2009; Gourion et al., 2006; Gourion et al., 2008). Their data were confirmed by results obtained in other α-proteobacteria, such as *B. japonicum, C. crescentus, R. etli* and *S. meliloti* (Alvarez-Martinez *et al.*, 2007; Bastiat *et al.*, 2010; Gourion *et al.*, 2009; Martinez-Salazar *et al.*, 2009; Sauviac *et al.*, 2007).

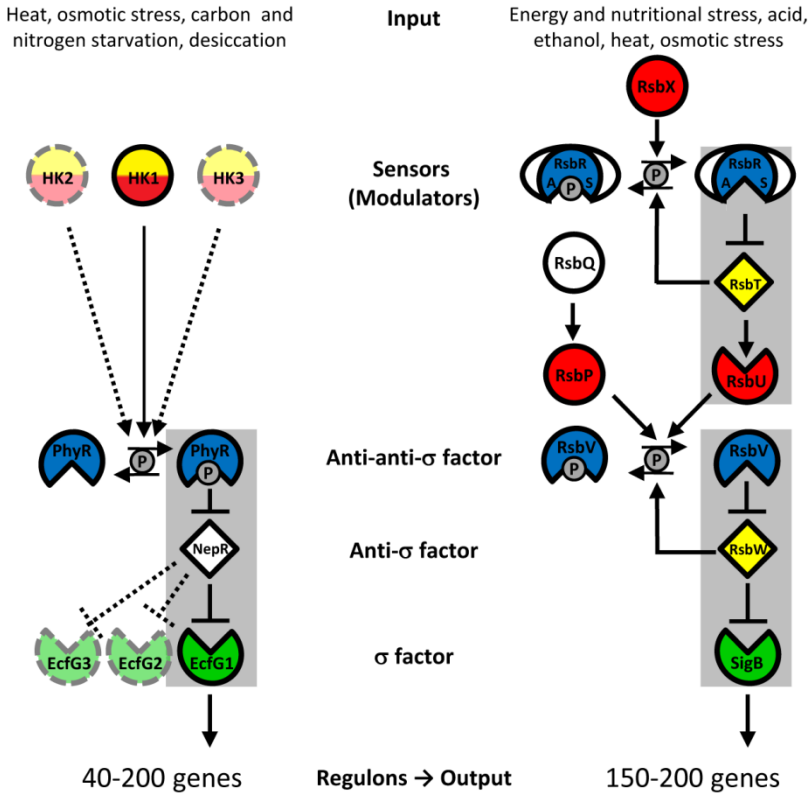

Figure 3.1. Regulatory networks orchestrating the general stress response in α-proteobacteria (left) and *Bacillus subtilis* (right). Colour code: kinases yellow, phosphatases red, phospho-acceptors blue, σ-factors green. Inhibition is indicated by T-shaped lines, phosphoryl groups by grey circles, labelled 'P'. HK = histidine kinase. Partner switching modules are underlain in grey, with the switching proteins represented as tilted squares. See text for detail.

The picture that emerged from these studies is that the regulatory cascade orchestrating the α-proteobacterial GSR shows some similarities with the σ^B-dependent response, but is based on a set of unique proteins not found in any other bacterial stress signalling pathway to date (Fig. 3.1, left side). Instead, it reshuffles and combines domains from two phylogenetically unrelated signalling archetypes, two-component systems and ECF σ-factors. The input section of this GSR is made up of a specialized 2CS consisting of as yet unidentified stress-responsive sensor HK and an unusual RR, PhyR, which will be described in more detail below. The output section is represented by a unique anti-σ/σ-factor pair of NepR- and EcfG-like proteins respectively. In the absence of stress signals, NepR binds to EcfG, thereby keeping this ECF σ-factor inactive. EcfG-dependent gene expression is induced by stress conditions very reminiscent of those described for other GSR, including heat and osmotic stress, carbon and nitrogen starvation, and desiccation (Francez-Charlot *et al.*, 2010). Under such conditions, the postulated sensor kinase(s) phosphorylates their cognate PhyR-like RR, which then binds NepR, thereby titrating it away from EcfG. Phosphorylation PhyR therefore acts as an anti-anti-σ factor. This phosphorylation-dependent partner switch – which is remiscent of the σ^B-cascade in *B. subtilis* (Fig. 3.1) – ultimately releases EcfG, which then recruits RNA polymerase core enzyme to initiate transcription of its target genes. Once the stress is relieved, dephosphorylation of PhyR again reverses the switch, hence shutting off the GSR (Fig. 3.1).

Comparative genomics analyses revealed that all three proteins are conserved within, but also restricted to, the class of α-proteobacteria (Francez-Charlot *et al.*, 2010; Staroń *et al.*, 2009). In this bacterial clade, they are only absent in some degenerated obligate symbionts highly adapted to very constant host conditions, including the genera *Anaplasma, Rickettsia* or *Wolbachia*. The composition and genomic organization of the known GSR components from a number of prominent α-proteobacteria is summarized in Table 3.1.

The 'typical' locus consists of a *phyR* gene, which is divergently transcribed from a shared intergenic region with the *nepR–ecfG* operon. Due to its small size, homologues of *nepR* – while normally present – are not always annotated. In addition, genes encoding sensor kinases are usually in the vicinity and often co-transcribed with either of the two cistrons. With few exceptions, most genomes encode a single copy of PhyR and NepR. In contrast, the number of EcfG-like σ-factors can vary significantly (Table 3.1), indicative for regulatory diversification at the level of gene expression (Fig. 3.1). This represents a remarkable difference compared with all the other GSR described before, which

invariantly co-ordinate the activity of a single σ-factor. The physiological significance of this observation remains to be identified. But one could imagine that more than one output σ-factor either allows for more graded responses or the induction of distinct sub-regulons.

Table 3.1. Distribution of EcfG, PhyR and NepR in selected α-proteobacterial genomes.[a]

Organism	PhyR locus organization[b]	PhyR	NepR[c]	EcfG	HWE[d]	HisKA2[e]
Acetobacter pasteurianus	<H_K-N-E-E< >P>	1	1	2	-	1 (0/1)
Agrobacterium tumefaciens	<E-N< >P-H_W> <H_W<	1	1	3	5 (4/1)	1 (1/0
Bartonella spp.	<H_K-E-N< >P-H_W>	1	1	1	1 (1/0)	1 (0/1)
Bradyrhizobium japonicum	<P< >N-E>	1	1	1	6 (3/3)	3 (3/0)
Bradyrhizobium sp. BTAi1	<P< >N-E> <H_W<	1	1	1	12 (6/6)	3 (2/1)
Brucella spp.	<H_W-X-P< >N-E-H_K>	1	1	1	2 (2/0)	1 (0/1)
Caulobacter crescentus	<H_K-E-N< >P>	1	1	2	7 (5/2)	3 (1/2)
Erythrobacter litoralis	>P><H_W< >N-E>	1	1	1	4 (3/1)	3 (3/0)
Gluconobacter oxydans	<P< >E-E-H_K>	1	0	2	2 (2/0)	2 (1/1)
Hyphomonas neptunium	<H_K-E-N-X< >X-P>// >P> <H_W<	2	1	2	4 (3/1)	1 (0/1)
Methylobacterium extorquens	<N< >P>	1	1	6	11 (9/2)	5 (4/1)
Methylobacterium sp. 4–46	<N< >P>	1	1	10	11 (9/2)	6 (4/2)
Nitrobacter winogradskyi	<P< >N-E> <H_W<	1	1	1	2 (0/2)	0
Paracoccus denitrificans	<H_K-E-N< >P>	1	1	2	1 (0/1)	1 (0/1)
Rhizobium etli	<H_W-P< >N-E>	1	1	2	3 (1/2)	3 (3/0)
Rhodobacter sphaeroides	<H_K-E-N< >P>	1	1	2	4 (4/0)	5 (1/4)
Rhodopseudomonas palustris	<P< >N-E> <H_W<	1	1	1	8 (7/1)	3 (2/1)
Rhodospirillum centenum	P	1	0	0	3 (3/0)	5 (3/2)
Roseobacter denitrificans	<H_K-E< >P>// <H_K-E-N< >P>	2	1	2	0	2 (0/2)
Sinorhizobium medicae	<H_W-E-N< >P>// <P< >N>	2	2	2	5 (4/1)	1 (1/0)
Sinorhizobium meliloti	<H_W-E-N< >P>// P	2	1	2	7 (6/1)	0
Sphingopyxis alaskensis	<E< >X-E> <P< >H_W>	1	0	2	2 (1/1)	0

[a] A complete table containing a detailed list of proteins from all α-proteobacteria is available on request from the authors.
[b] E, EcfG; HK, Histidine kinase harbouring Pfam:HisKA_2 domain; HW, histidine kinase harbouring Pfam:HWE-HK domain; N, NepR; P, PhyR. The arrows indicate the operon organization and orientation.
[c] *nepR* is not always annotated.
[d] Histidine kinases harbouring either HWE or HisKA_2 domains. Values in parentheses: number of (cytoplasmic-/periplasmic-sensing) histidine kinases.

3.4. PhyR: an unusual hybrid protein acting as a phosphorylation-dependent anti-anti-σ factor

As mentioned above, PhyR represents the central regulatory switchboard of the α-proteobacterial GSR. It presumably integrates the information of different sensor kinases (see below) and indirectly regulates the activity of EcfG-like σ-factors that orchestrate the GSR.

PhyR is an unusual protein in a number of respects. It is a RR that carries its receiver domain (the interaction interface with the cognate sensor kinases) at its C-terminus. This

organization, while not unprecedented, is only found in a small minority of other RR, including CheW-like proteins, all of which do not function as transcriptional regulators (Galperin, 2010). The N-terminal output domain of PhyR shows high sequence similarity to ECF σ-factors, at least at first glance. But a closer inspection revealed that this domain lacks a number of residues crucial for DNA-binding, and indeed no such activity has been described for PhyR. Instead, this domain shows specific homology to EcfG-like proteins (Francez-Charlot *et al.*, 2010; Staroń *et al.*, 2009). Biochemical studies revealed that – upon PhyR phosphorylation – this domain serves as a docking module for NepR (Francez-Charlot *et al.*, 2009). Hence, NepR switches its partner under stress conditions, thereby releasing the EcfG to initiate expression of the GSR regulon.

The high-resolution crystal structure of PhyR from *Caulobacter crescentus*, published in issue 78 of *Molecular Microbiology*, offers fascinating insight into how its functional role is achieved by the three-dimensional architecture of this protein (Herrou *et al.*, 2010). Initially, the authors genetically infer that *C. crescentus* PhyR indeed functions as a phosphorylation-dependent anti-anti-σ factor, in line with previous reports on PhyR proteins (Bastiat *et al.*, 2010; Francez-Charlot *et al.*, 2009; Gourion *et al.*, 2009). Subsequently, they solve the crystal structure of this protein at 1.25Å. This structure yields a number of important insights.

In accordance with previous results in solution, the crystal structure is consistent with a model in which PhyR functions as a monomer. In contrast to canonical RR, which dimerize upon phosphorylation, PhyR activity is mediated by phosphorylationdependent changes in intramolecular interactions.

PhyR can be divided into three structural domains that – by themselves – do not offer a lot of surprises. The N-terminal EcfG-like domain contains two sub-domains with folds typical of regions σ^2 and σ^4, which are responsible for binding the -10 and -35 promoter regions, respectively, in all σ^{70} proteins. Hence the σ-like domain of PhyR is structurally homologous to other ECF σ factors, such as RpoE from *E. coli* and *R. sphaeroides* (Campbell et al., 2008). In this unphosphorylated structure of PhyR, the σ-like domain exists in a closed conformation and there is evidence that the loop region linking regions σ^2 and σ^4 is highly flexible. As expected, the C-terminal receiver domain of PhyR adopts a structure typical of classical RR (Bourret, 2010).

What is most remarkable about the PhyR structure is the identification of extensive interaction interfaces between the C-terminal receiver domain and both regions of the EcfG-like domain. Considering that the latter alone is sufficient to perform the anti-anti-σ

activity, the authors suggest that the receiver domain therefore can be viewed as an anti-anti-anti-σ domain that inhibits the EcfG-like anti-anti-σ domain through direct interactions in the absence of phosphorylation. The inhibitory interaction interface differs from the one necessary for the transient interactions between sensor kinase and RR mediating phosphoryl-group transfer. Hence, the former does not get in the way of the latter. The structure offers evidence that, in the absence of inducing signals, the unphosphorylated receiver domain of PhyR tethers the EcfG-like domain in a closed conformation. Phosphorylation of the receiver domains usually induces intramolecular conformational changes that ultimately activate the output domain of a RR (Bourret, 2010). In the case of PhyR, it is attractive to hypothesize that phosphorylation changes its conformation to release the two regions of the EcfG-like domain from its inhibitory grip, thereby allowing this anti-anti-σ domain to interact with NepR-like anti-σ factors (Herrou *et al.*, 2010). But how this interaction is achieved remains to be analysed.

3.5. Outlook: NepR interactions, sensor kinases and beyond

The PhyR structure presented in issue 78 of *Molecular Microbiology* is an important step towards understanding the mechanism underlying the PhyR–NepR–EcfG-mediated signal transduction of the GSR in α-proteobacteria. But it will require further structural studies to truly understand how the partner switching module works.

Specifically, how is the NepR anti-σ factor unlocked from EcfG and titrated to the EcfG-like domain once PhyR gets phosphorylated under stress conditions? For this mechanism to be efficient, one has to postulate that the EcfG-like domain of PhyR – in its phosphorylated state – has a higher affinity for NepR than the corresponding σ-factor itself. NepR does not share any sequence similarity with other anti-σ factors. While structural studies on the RpoE-RseA pair of *E. coli* and RpoE–ChrR of *R. sphaeroides* have revealed the mechanism underlying the inhibition of ECF s-factors by 'classical' anti-σ factors (i.e. those containing a so-called anti-σ domain) (Campbell *et al.*, 2008), it is unclear whether this mode of interaction will also hold true for NepR.

An additional question regards the nature of the HK that sense the stress conditions and – upon autophosphorylation – serve as phospho-donors for PhyR-like RR. Here, the first indications come from comparative genomics. Numerous such analyses have demonstrated that, at least in bacteria, genomic context conservation is usually a reliable indicator of functional links between proteins encoded by genes adjacent to each other on the

chromosome. And indeed, we do find genes encoding sensor kinases in the vicinity of quite a number of *phyR* loci (Table 3.1). But at first glance, these sensor kinases – if present at all – look rather diverse, including both cytoplasmic- and periplasmic-sensing proteins of varying domain architecture (Fig. 3.2).

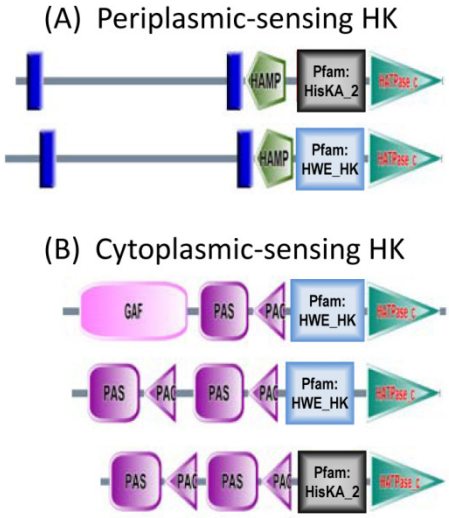

Figure 3.2. Domain architecture of HK associated with PhyR. Proteins are symbolized as grey lines and drawn to scale. Putative transmembrane regions are represented by blue vertical bars. All other domains according to the graphical output of the SMART database (http://smart.embl-heidelberg.de/), with modifications. See text for details.

But bearing in mind that the specificity of a HK for its partner RR is defined by the dimerization and histidine phosphotransfer (DHp) domain (Gao and Stock, 2009), a closer look revealed one important similarity between all kinases genetically linked to PhyR. They all do not contain 'typical' DHp domains (Pfam:HisKA), but instead harbour either a Pfam:HWE-HK or a Pfam:HisKA_2 DHp domain. Remarkably, the HWE and HisKA_2 domains are very similar to each other, but differ significantly from the vast majority of classical sensor kinases (Karniol and Vierstra, 2004). What is even more striking is that both types of sensor kinases are either predominantly (HisKA_2; 55%), or in the case of the HWE kinases almost exclusively (95%) found in genomes of α-proteobacteria. These observations strongly suggest that the only likely phospho-donors for PhyR belong to either of the two kinase groups. This hypothesis, if true, narrows down the range of potential proteins significantly, which should allow their identification even in cases where no candidate gene is found next to *phyR* on the chromosome (Table 3.1). Lastly, the observation that such kinases can be either cytoplasmic- or periplasmic-sensing offers the possibility to integrate both environmental and intracellular signals, again analogous to the situation described for the σ^B-dependent cascade of *B. subtilis*. But experimental evidence will be required to verify these hypotheses on the nature of the sensor kinases able to phosphorylate PhyR-like proteins. Although the study of GSR in the α-proteobacteria is just beginning, the data already available indicate that more surprises are likely in store. And structural studies, like the one of PhyR reported in issue 78 of *Molecular Microbiology*, will be pivotal to ultimately unravel to mechanism behind this novel GSR cascade.

3.6. Acknowledgements

We would like to thank Sean Crosson and John Helmann for critical reading of the manuscript, and Regine Hengge, Chet Price and Julia Vorholt for sharing manuscripts prior to publication. Work in the authors' lab was supported by grants from the Deutsche Forschungsgemeinschaft (MA2837/2-1) and the Fonds der Chemischen Industrie.

CHAPTER 4

Co-evolution of ABC-transporters and two-component regulatory systems as resistance modules against antimicrobial peptides in Firmicutes bacteria

This chapter has been adapted from:

S. Dintner*, A. Staroń*, E. Berchtold, T. Petri, T. Mascher, S. Gebhard

** contributed equally*

J. Bacteriol. (2011) 193(15):3851-3862

Chapter 4

Coevolution of ABC Transporters and Two-Component Regulatory Systems as Resistance Modules against Antimicrobial Peptides in Firmicutes Bacteria

In Firmicutes bacteria, ABC transporters have been recognized as important resistance determinants against antimicrobial peptides. Together with neighbouring 2CS, which regulate their expression, they form specific detoxification modules. Both the transport permease and sensor kinase components show unusual domain architecture: the permeases contain a large extracellular domain, while the sensor kinases lack an obvious input domain. One of the best-characterized examples is the bacitracin resistance module BceRS-BceAB of *B. subtilis*. Strikingly, in this system, the ABC transporter and 2CS have an absolute mutual requirement for each other in both sensing of and resistance to bacitracin, suggesting a novel mode of signal transduction in which the transporter constitutes the actual sensor. We identified over 250 such BceAB-like ABC transporters in the current databases. They occurred almost exclusively in Firmicutes bacteria, and 80% of the transporters were associated with a BceRS-like 2CS. Phylogenetic analyses of the permease and sensor kinase components revealed a tight evolutionary correlation. Our findings suggest a direct regulatory interaction between the ABC transporters and 2CS, mediating communication between both components. Based on their observed coclustering and conservation of RR binding sites, we could identify putative corresponding two-component systems for transporters lacking a regulatory system in their immediate neighbourhood. Taken together, our results show that these types of ABC transporters and 2CS have coevolved to form self-sufficient detoxification modules against antimicrobial peptides, widely distributed among Firmicutes bacteria.

4.1. Introduction

In an era in which resistance against antibiotic compounds is becoming a major health issue worldwide, much attention is being given to the development of new drugs. In recent years, antibiotics that target the lipid II cycle of bacterial cell wall synthesis have been promoted as promising candidates. This group of compounds includes, for example, the glycopeptides vancomycin, lipodepsipeptides like ramoplanin, the cyclic peptide

bacitracin, and most members of the large class of lantibiotics (Breukink and de Kruijff, 2006). The latter two are produced mainly by low-G+C Gram-positive microorganisms (Firmicutes) and target closely related species (Guder *et al.*, 2000; Johnson *et al.*, 1945).

As with most antibiotics, resistance against antimicrobial peptides has been observed, and the major mechanism appears to be via alterations of cell envelope charge and composition (Breukink and de Kruijff, 2006; Cotter *et al.*, 2005). Additionally, producer self-resistance against lantibiotics is often mediated by ABC transporters, collectively termed LanFEG, consisting of two membrane-spanning subunits and one ATPase, which are encoded in the biosynthetic loci for the respective lantibiotic and whose expression is regulated by a 2CS of the same genetic locus (Draper *et al.*, 2008). Over the last decade, several ABC transporters of a different type have been identified as resistance determinants against peptide antibiotics in nonproducing strains (Becker *et al.*, 2009; Collins *et al.*, 2010; Kramer *et al.*, 2006; Majchrzykiewicz *et al.*, 2010; Meehl *et al.*, 2007; Ohki *et al.*, 2003b; Pietiäinen *et al.*, 2009; Tsuda *et al.*, 2002). The permeases of these transporters share unique domain architecture with 10 transmembrane helices and a large extracellular domain (ECD) of about 200 amino acids between helices 7 and 8. They have been classified as the peptide-7 exporter (Pep7E) family in the Transport Classification Database (TCDB) (Saier *et al.*, 2009). Regulation of these transporters generally also occurs via 2CS, which are most commonly encoded in an operon adjacent to that of the transporter (Collins *et al.*, 2010; Joseph *et al.*, 2004; Li *et al.*, 2007a; Meehl *et al.*, 2007; Ohki *et al.*, 2003a; Ouyang *et al.*, 2010). The HK of these 2CS belong to the IM-HK group and are characterized by the possession of two transmembrane helices with a short (2- to 10-aminoacid) extracellular linker and no cytoplasmic domains besides the DHp (dimerization histidine phosphotransfer (Dutta *et al.*, 1999), also referred to as the H-kinase_dim domain in the Pfam database) and catalytic domains (Mascher, 2006). An apparent widespread cooccurrence of these types of transporters and regulatory systems among the Firmicutes bacteria was noted as early as 2002 (Joseph *et al.*, 2002), and this study was recently updated by the same group (Coumes-Florens *et al.*, 2011).

One of the best-understood examples of this resistance mechanism is the BceRS-BceAB module of *B. subtilis* (Fig. 4.1), which confers resistance against bacitracin, actagardine, and mersacidin (Mascher *et al.*, 2003; Ohki *et al.*, 2003a; Staroń *et al.*, 2011). A striking characteristic of this system is the absolute requirement for the ABC transporter (BceAB) for bacitracin-dependent induction of *bceAB* expression, showing that the HK (BceS) alone is unable to detect the presence of the antimicrobial peptide (Bernard *et al.*, 2007;

Rietkötter *et al.*, 2008). The current working model assumes the existence of a sensory complex between the transporter and 2CS. Furthermore, it has also been shown that the presence of the permease's ECD was indispensable for signalling, leading to the proposition that this region might constitute the substrate recognition domain of the module (Rietkötter *et al.*, 2008). Requirement of the ABC transporter for signal transduction was also observed in the bacitracin resistance module MbrABCD of *S. mutans* (Ouyang *et al.*, 2010), and regulation of the vancomycin resistance transporter VraFG of *S. aureus* appears to involve at least the permease component VraG (Meehl *et al.*, 2007). Taken together with their conserved genomic cooccurrence, these findings suggest that a regulatory interplay between the ABC transporter and 2CS is a common theme in these antimicrobial peptide resistance modules.

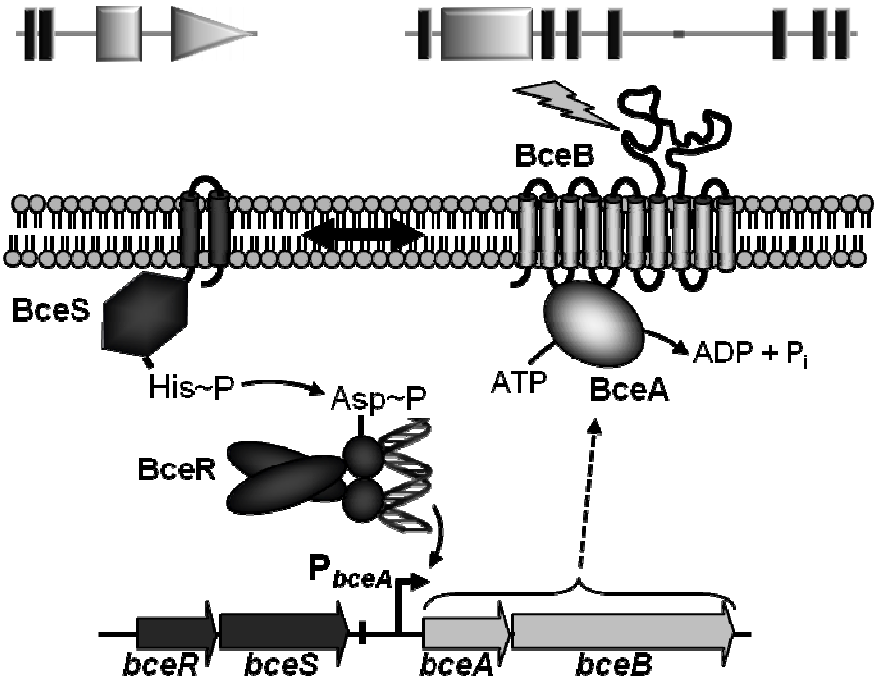

Figure 4.1. Schematic diagram of the BceRS-BceAB bacitracin resistance module of *B. subtilis*. Names of relevant proteins and genes are given. The presence of antimicrobial peptides is indicated by a lightning bolt. The proposed interaction between BceB and BceS is shown by a double-headed arrow. The phospho-relay between BceS and BceR, ATP hydrolysis by BceA, and activation of the *bceA* promoter (P_{bceA}) by BceR are shown by single-headed arrows, and increased expression of *bceAB* is shown by a dashed arrow (diagram adapted from reference (Rietkötter *et al.*, 2008)). The domain architectures of BceS and BceB, as predicted by the SMART database (Letunic *et al.*, 2009), are indicated above. Transmembrane domains are depicted as black rectangles, the DHp and catalytic domains of BceS are shown as a shaded square and triangle, respectively, and the FtsX domain of BceB is shown as a shaded rectangle.

To address this hypothesis, we carried out a comprehensive phylogenetic study and coclustering analysis of these systems and identified a clear coevolutionary relationship between transport permeases and HK. Additionally, we closely investigated the ECDs with

regard to their primary sequences and secondary structures, and we identified conserved putative RR binding sites in the promoter regions of the transporter operons. Furthermore, based on the observed coclustering of HK and transport permeases, we could identify putative corresponding 2CS for transporters lacking regulatory systems in their immediate neighborhood. Our findings suggest the existence of a sensory complex between transporter and HK and thus a novel signaling mechanism involved in the resistance of Firmicutes bacteria against peptide antibiotics.

4.2. Results

4.2.1. Genomic arrangement and phyletic distribution of Pep7E transporters and 2CS

Based on sequence similarity to the three previously described resistance modules of *B. subtilis* (BceRS-AB, PsdRS-AB, and YxdJK-LM) and on the typical domain architecture of Pep7E permeases (ten transmembrane helices, large ECD between helices 7 and 8, FtsX domain encompassing helices 2 to 4) (Fig. 4.1), we identified a total of 266 BceAB-like ABC transporters, as described in Materials and Methods (see Table S4.1 in the supplemental material). Next, we analyzed the genomic context of all permeases in our data set for RR, HK, and ATPases. 2CS were included only if their HK possessed the typical domain architecture of BceS-like IM-HK (Mascher, 2006) (Fig. 4.1). Of the total of 266 permeases, we found 22 orphan permeases lacking an ATPase (Table 4.1). However, 14 of these were found in the immediate neighbourhood of a second, complete Pep7E transporter as discussed below. Moreover, 181 BceS-like HK were identified (Table 4.1). In total, 213 (80%) of the transporters in our data set were associated with a BceRS-like 2CS (Fig. 4.2A; see also Table S4.1 in the supplemental material). All further analyses reported in this study were restricted to the permease and HK components of the modules, which are here referred to as BceB-like and BceS-like, respectively. Beyond the genes described here, no conservation of a genomic context was found for these modules. However, in a few cases, additional genes with a putative role in resistance against cell wall active antibiotics were carried in the same locus, such as undecaprenyl pyrophosphate phosphatase (*uppP*)-like genes in four clostridial species and a *vanZ*-like gene in the Ysa locus of *L. lactis* (not shown). UppP-like proteins are involved in bacitracin resistance (Bernard *et al.*, 2005), and VanZ of *E. faecium* confers resistance to teicoplanin, a glycopeptide antibiotic also interfering with the lipid II cycle of cell wall synthesis (Arthur *et al.*, 1995).

Table 4.2. Phyletic distribution and cooccurrence of ABC transporters with 2CS.

Order	Phylum	No. of:		
		ABC transporters[a]		Associated 2CS[b]
		Complete	Orphan permease	
Bacillales	Firmicutes	116	20	87
Lactobacillales		34		16
Clostridiales		87	2	71
Erysipelotrichales		2		2
Thermoanearobacterales		1		1
Coriobacteriales	Actinobacteria	2		2
Spirochaetales	Spirochaetes	1		1
Unclassified	Unclassified	1		1
Total		244	22	181

[a] Transporters listed as "complete" are encoded by adjacent genes for the ATPase and permease, and transporters listed as "orphan permease" are those for which no ATPase gene was found immediately adjacent to the permease gene.
[b] 2CS listed as "associated" were found in the same genetic locus (within a six-gene distance) as a transport permease.

Among our data set, 10 systems had been described previously (Fig. 4.2A and Table 4.2). The common trait of all these systems was their involvement in resistance against antimicrobial peptides, suggesting that the function of Pep7E transporters is restricted to the detoxification of peptide antibiotics. In good agreement with other studies (Coumes-Florens *et al.*, 2011; Joseph *et al.*, 2002), nearly all modules in our data set were found in the phylum Firmicutes. The vast majority (97%) of these were distributed among the orders *Bacillales*, *Clostridiales*, and *Lactobacillales* (Table 4.1). Only four systems were found in other bacterial phyla: two in the *Actinobacteria*, one in a spirochete, and one in an unclassified bacterium. Most members of the *Lactobacillales* contained only one (e.g., *S. mutans*) or two (e.g., *L. rhamnosus*) copies of such resistance modules. In contrast, members of the *Bacillales* and *Clostridiales* usually contained several copies and, in some cases, up to six modules (e.g., *B. cereus* and *Clostridium sporogenes*) (see Table S4.1 in the supplemental material).

4.2.2. Phylogenetic analysis of the permease and HK components

To analyze the phylogenetic relationship of the transport permeases and HK, we first created multiple sequence alignments with ClustalW2 (Thompson *et al.*, 2002). Both sets of proteins showed good sequence conservation across the data set, with the permeases sharing 25% to 40% pairwise sequence identity and the HK sharing 20% to 55% pairwise sequence identity at the amino acid level. Notably, the region of the permeases containing the large ECD (corresponding to positions 310 to 520 in *B. subtilis* BceB) showed only

very low sequence similarity, with less than 10% pairwise sequence identity observed, resulting in a poor alignment across this region (not shown). We therefore excluded the entire ECD from the phylogenetic analysis. A phylogenetic tree of the transport permeases calculated by the neighbor-joining method showed eight distinct groups (Fig. 4.2A). Seven of these were supported by bootstrap values of above 0.5 (i.e., they occurred in over 50% of 100 random bootstrapping replicates) (Fig. 4.2A, gray dots). We labelled these groups I to VIII. A recently published grouping of this type of transporter reported six distinct phylogenetic groups (Coumes-Florens *et al.*, 2011). For reasons of consistency, we assigned the same group numbers where possible, which required splitting the previously assigned group III into our groups III, VII and VIII. We next calculated a neighbor-joining tree of the HK, which showed the existence of seven distinct groups; however, only three of these were supported by bootstrap values of above 0.5 (Fig. 4.2B). Strikingly, there appeared to be a substantial similarity in the clustering of permeases and HK belonging to the same genetic locus. To analyze this congruence more closely, we chose a layout in which both phylogenetic trees could be juxtaposed in a mirror-image display. Each permease was then linked to the HK from the same genetic locus by a line of the same color as that chosen to highlight its group assignment (Fig. 4.2C). From this, it was clear that, apart from a few exceptions (mainly permeases from group V associating with HK from group VII) (Table 4.2), the HK associated with permeases from one phylogenetic group also formed a distinct group in their own tree, suggesting a coevolution of both protein families (Pazos and Valencia, 2008). Curiously, no permease from group VI was associated with a 2CS, which explained the presence of only seven HK groups rather than eight permease groups and is analyzed in more detail below.

4.2.3. Description and functional analysis of the phylogenetic groups

Following the division of transport permeases into the eight phylogenetic groups, we analyzed these groups for common traits or species distribution to see if it might be possible to assign putative functions to as-yet-uncharacterized systems. The details of this analysis are summarized in Table 4.2. While some groups, such as group II, were comprised of systems from a wide range of genera, others showed a more restricted species distribution, e.g., groups I, IV, VI, and VII. However, we could not detect the opposite correlation, i.e., modules from one species did not cluster into the same phylogenetic group. It was also not possible to detect any correlation between substrate range and phylogenetic group. Systems conferring bacitracin resistance were found in groups II, IV,

and V, and those mediating nisin resistance were found in groups II, III, IV, and V (Table 4.2 and references therein).

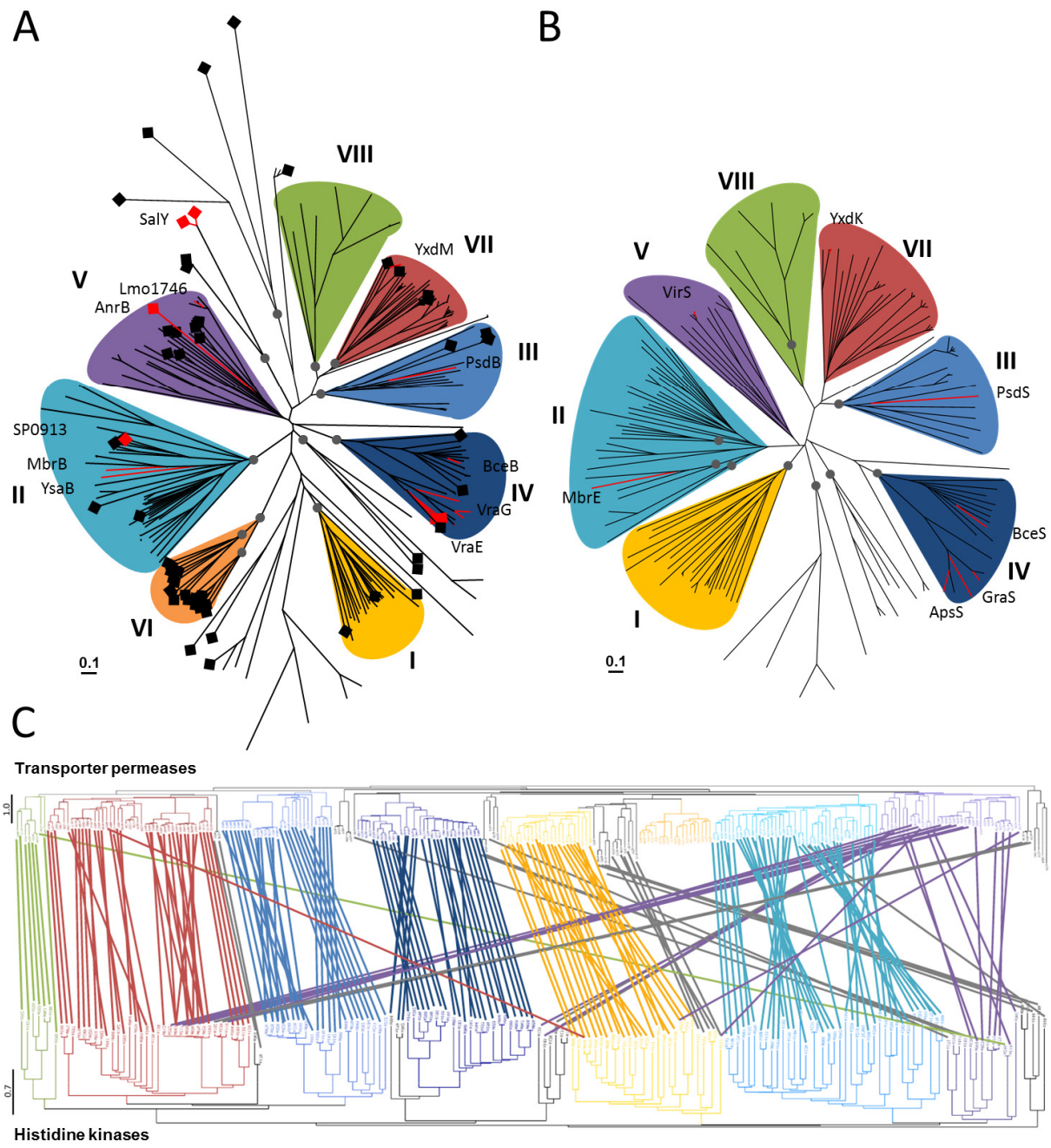

Figure 4.2. Phylogenetic trees and coevolution of Pep7E-type permeases and BceS-like HK. Phylogenetic trees were calculated by the neighbor-joining (NJ) method for both protein families. (A and B) Lines for previously described systems are shown in red, and names of corresponding proteins are given. (A) NJ tree of transport permeases. Permeases lacking a neighbouring 2CS are marked with black or red squares. The assignment of the eight phylogenetic groups (I to VIII) is shown by different colours. (B) NJ tree of HK. Assignment of phylogenetic groups is colour coded to match the corresponding groups shown in panel A. (C) Mirror-image display of the NJ trees of permeases (top) and HK (bottom). Colour coding was done as described for panels A and B. Lines were drawn from each permease to the HK encoded in the same locus, using the colour indicating the permease's group assignment. Scale bars show the number of expected amino acid replacements per site as calculated for each phylogenetic tree. Bootstrap values of above 0.5 for the central branches are indicated by gray dots (shown only in panels A and B).

Table 4.2. In depth analysis of the phylogenetic groups.

Group	No. of permeases	Genera[a]	No. of HK[b]	Example permease(s)	Substrate (inducer) range[c]	Reference(s)
I	24	Clo (14), Bac (2), Eub (2), Ana (1), Ato (1), Bry (1), Col (1), Cop (1), Rum (1)	22 (22)			
II	43	Str (11), Clo (8), Rum (4), Lac (3), Sta (3), Bla (2), Des (2), Dor (2), Cal (1), Cat (1), Eub (1), Hol (1), Lac (1), Leu (1), Ros (1), Tre (1)	39 (38)	MbrB, SP0913, Spr0813, YsaB	Bacitracin, gramicidin, lincomycin, nisin, vancoresmycin	(Becker *et al.*, 2009; Kramer *et al.*, 2006; Majchrzykiewicz *et al.*, 2010; Ouyang *et al.*, 2010; Tsuda *et al.*, 2002)
III	28	Bac (19), Clo (5), Pae (2), Sym (1), The (1)	22 (22)	PsdB	Enduracidin, gallidermin, nisin, subtilin, (actagardine)	(Staroń *et al.*, 2011)
IV	32	Sta (13), Bac (11), Geo (4), Alk (2), Lys (1), Oce (1)	26 (23)	BceB, VraE, VraG	Bacitracin, hBD3, indolicidin, LL-37, nisin, polymyxin B, vancomycin, actagardine, mersacidin	(Bernard *et al.*, 2007; Li *et al.*, 2007a; Meehl *et al.*, 2007; Pietiäinen *et al.*, 2009; Rietkötter *et al.*, 2008; Staroń *et al.*, 2011)
V	37	Bac (13), Clo (13), Lis (8), Alk (2), Eub (1)	22 (8)	AnrB, Lmo1746	Bacitracin, β-lactam antibiotics, gallidermin, nisin	(Collins *et al.*, 2010; Mandin *et al.*, 2005)
VI	19	Bac (9), Lac (8), Ent (1), Ped (1)	0			
VII	44	Bac (34), Clo (4), Geo (4), Bre (1), Ent (1)	26 (25)	YxdL	(LL-37)	(Pietiäinen *et al.*, 2005)
VIII	8	Bac (3), Clo (2), Alk (1), Ana (1), Lys (1)	8 (7)			

[a] Abbreviations: Alk, *Alkaliphilus*; Ana, *Anaerococcus*; Ato, *Atopobium*; Bac, *Bacillus*; Bry, *Bryantella*; Bre, *Brevibacillus*; Bla, *Blautia*; Cal, *Caldicellulosiruptor*; Cat, *Catenibacterium*; Cop, *Coprococcus*; Clo, *Clostridium*; Col, *Collinsella*; Des, *Desulfitobacterium*; Dor, *Dorea*; Ent, *Enterococcus*; Eub, *Eubacterium*; Geo, *Geobacillus*; Hol, *Holdemania*; Lis, *Listeria*; Lys, *Lysinibacillus*; Leu, *Leuconostoc*; Lac, *Lactococcus*; Oce, *Oceanobacillus*; Pae, *Paenibacillus*; Ped, *Pediococcus*; Sta, *Staphylococcus*; Sym, *Symbiobacterium*; Str, *Streptococcus*; The, *Thermobaculum*; Tre, *Treponema*; Ros, *Roseburia*; Rum, *Ruminococcus*. The number of systems from each genus is shown in parentheses.

[b] The number of HK associated with transporters of this group. The number in parentheses shows the number of these HK which belonged to the matching group in the phylogenetic tree of HK.

[c] The substances given in parentheses are antimicrobial peptides, which were shown to induce expression of the transporter genes but against which the transporter did not confer resistance. It should be noted that the compounds listed as substrates may also act as inducers. hBD3, human β-defensin 3.

A unique transport system in our data set was SalXY of *S. salivarius*, which could not be assigned to any of the eight groups (Fig. 4.2A). Regulation of this transporter is mediated by the 2CS SalKR encoded downstream of *salXY* (Upton *et al.*, 2001). However, the HK SalK did not show the typical BceS-like domain architecture and was therefore excluded from our data set. This deviation from the paradigm modules described in this study most likely reflects the fact that SalXY is the only system from our data set that is part of a biosynthetic locus for an antimicrobial peptide, salivaricin A, and is thought to mediate producer self-resistance (Upton *et al.*, 2001).

4.2.4. Alignment correlation supports coevolution of 2CS and ABC transporters

A congruence of two phylogenetic trees, as we found for BceB-like permeases and BceS-like HK, is generally taken to reflect a coevolution of both protein families (Mascher *et al.*, 2003). For example, the coevolution of HK with their cognate RR was deduced from a qualitative comparison of their respective phylogenetic trees (Goh *et al.*, 2000). While phylogenetic trees are useful for visual inspection, alignment correlations as introduced by Goh and colleagues (Goh *et al.*, 2000) bypass the ambiguous tree-building step and instead rely on similarity matrices computed directly from multiple alignments (see Materials and Methods). A correlation coefficient (CC) was, for example, calculated for chemokines and chemokine receptors to show the degree of similarity between the two trees (Goh *et al.*, 2000). However, the interpretation of such results is somewhat problematic in that the similarity of phylogenetic trees of protein families is partly due to the underlying phylogeny of the species. Furthermore, additional effects can lead to an apparent coevolution of two proteins from the same organism, such as the fact that both of them are membrane localized, as is the case for the BceB- and BceS-like proteins. Usually, the significance of the obtained CCs is computed by shuffling the pairwise matrix row assignments, i.e., uncoupling the sequence pairs (Goh *et al.*, 2000). But since such an approach removes all correlation effects on both protein and species levels, it will erroneously detect significant coevolution and therefore does not provide a sufficient control. A more accurate validation of the alignment correlation is to use selected interacting and noninteracting protein families as controls, which provide a direct and easy-to-interpret background.

As described in Materials and Methods, we analyzed the genomes from our data set for a HK and a transport permease that were present in a large number of species but were not functionally associated with each other. These criteria were fulfilled by the HK YycG

(WalK), which is part of the essential YycFG (WalRK) 2CS of the Firmicutes (Dubrac *et al.*, 2008; Szurmant *et al.*, 2007a), and the permease OppB, a component of the widely distributed oligopeptide importer OppABCDF (Detmers *et al.*, 2001). In combination with our data set of BceRS-BceAB homologues, all possible pairs of protein families involving a YycG and/or OppB homologue are assumed not to have coevolved, because there is no functional link between these proteins and Bce-like modules. As shown in Fig. 4.3, gray font, the CCs calculated for these pairs were in the range of approximately 0.4 to 0.6. This degree of correlation can thus be considered the background similarity when comparing any pairs of proteins from our selection of species. The second set of controls consisted of protein pairs known to physically interact, such as the two protein subunits of BceAB-like transporters or of BceRS-like 2CS, which can be assumed to have coevolved. Consistent with this, these pairs all showed high CC values of above 0.9 (Fig. 4.3A, gray shading). Importantly, the pair of proteins under investigation in the present study, i.e., homologues of BceB and BceS, displayed a similarly high CC of 0.97 (Fig. 4.3A, outlined), supporting our hypothesis of their coevolution. The remaining pairs also show high CC values, which is probably due to indirect effects, because they are all components of the same module (Fig. 4.3B), even if a direct interaction between, for example, the HK and the ATPase is unlikely.

To compensate for potential random effects introduced into the ortholog families by single sequences, we devised a subset sampling with subsequent realignment. Each multiple sequence alignment was thus split into 50 alignments of 20 sequences. Using these new alignments, 50 separate CCs were computed against each family. A box plot of the obtained results showed that, apart from a few outliers, all subsets of interacting pairs (positive controls and BceB-BceS pairs) showed CCs of above those calculated for noninteracting pairs, thus providing further validation of our approach (Fig. 4.3C). Taken together, these analyses showed that the degree of coevolution between BceB-like permeases and BceS-like HK is as high as that between proteins known to physically interact and significantly higher than the background correlation caused by speciation effects.

A

	OppB	YycG	BceB	BceA	BceS
BceR	0.37	0.52	0.94 (0.78)	0.91 (0.91)	0.92 (0.84)
BceS	0.41	0.51	0.97 (0.82)	0.9 (0.72)	
BceA	0.43	0.62	0.93 (0.79)		
BceB	0.44	0.55			
YycG	0.64				

B

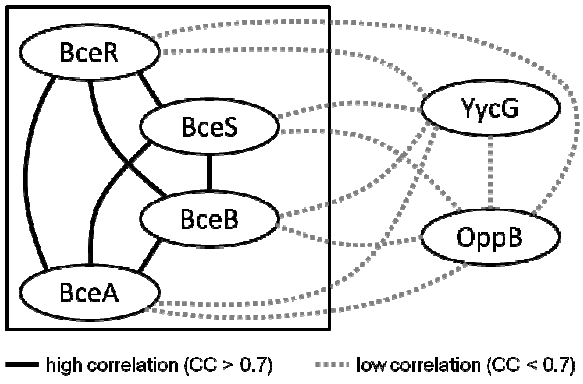

Figure 4.3. Correlation coefficients and coevolved protein partners. (A) CCs calculated from complete core data sets (n = 26) for all combinations of protein family pairs. Results from extended sets (n = 180), where available, are given in parentheses. Values of negative-control pairs are shown in gray, and values of positive-control pairs are shown by gray shading. The field showing the CC of BceB-like permeases with BceS-like HK is outlined in boldface. (B) Schematic presentation of correlation between all six protein families. Gray dotted lines show background correlation levels (CC < 0.7). Solid lines show correlation above the background (CC > 0.7). Components of Bce-like modules are boxed. (C) Box plot of randomly sampled CCs. Sequence subsets (n = 20) were extracted from core (negatives) or extended (all other pairs) data sets. In each column, 25% of values are lower than the bottom edge of the gray box (25% quartile), 50% are lower than the black line (median), and 75% are lower than the upper edge of the gray box (75% quartile). The whiskers show the most extreme data points within the 1.5-fold interquartile range; gray diamonds show data points outside this range. "Negatives" are all pairs involving at least one YycG or OppB homologue, and pairings of other protein families are given.

4.2.5. Primary and secondary structure analysis of the ECD

Our analysis of the phylogenetic groups described above could not identify any definite correlation between group assignment of a permease and its likely substrate range. While no information is available on the transport mechanism of Pep7E-type systems, the ECD of the permeases has been previously proposed to act as a substrate binding domain, based on its size and extracellular location (Rietkötter *et al.*, 2008). Sequence conservation of the ECD across the complete data set was poor, and thus, this region had been excluded from phylogenetic analyses. Following division of the permeases into the eight separate phylogenetic groups, we analyzed the ECD regions for group-specific characteristics, on the levels of both primary sequence and predicted secondary structure (see Fig. S4.1 in the supplemental material). The ECDs of the separate groups were similar in size, ranging from approximately 180 to 230 amino acids. Only group I possessed longer ECDs of 220 to 280 amino acids. Comparison of the secondary structures of all groups revealed a conserved arrangement of α-helices and β-sheets in the order α–β–α–$\beta_{2\text{-}3}$–α–$\beta_{3\text{-}4}$–α–β–$\alpha_{1\text{-}2}$, with only groups I and VI showing slight deviations (Fig. 4.4). We further identified individual conserved amino acid residues for each group (Fig. 4.4), although even within groups, sequence conservation of the ECDs was still moderate (Fig. S4.1). Conserved residues often appeared to cluster in the vicinity of the central α-helix and in the ultimate

or penultimate β-sheet preceding the terminal α–β–$\alpha_{1\text{-}2}$ sequence. Mostly, the conservation applied to hydrophobic amino acids, but group II is characterized by a large number of both positively and negatively charged conserved residues, while groups VII and VIII contain several conserved aromatic side chains. Despite these findings, the sequence of the ECD again was not correlated with substrate range.

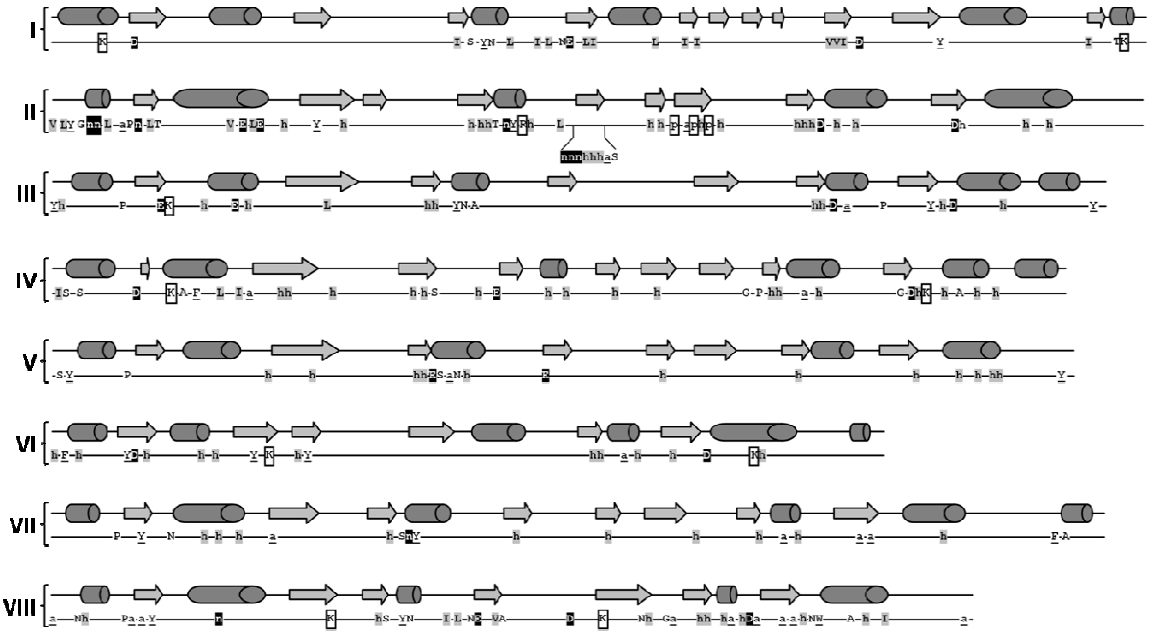

Figure 4.4. Secondary structure and consensus motifs in the extracellular domains of transport permeases. For each of the phylogenetic groups (shown on the left), the consensus secondary structure, based on prediction by JPred3 (Cole *et al.*, 2008) is displayed as α-helices (dark gray barrels) and β-sheets (light gray arrows). Conserved amino acids (threshold of 50%, based on analysis of ClustalW2 alignments in BioEdit (Hall, 1999)) are shown by one-letter codes on the bottom lines. Lowercase letters show conservation of amino acid type, as follows: h, hydrophobic (L, I, V, M); n, negative (D, E); p, positive (K, R); a, aromatic (Y, F). Hydrophobic (gray background), negatively charged (white letter on black background), positively charged (black box), or aromatic (underlined) residues are marked. Drawings are to scale for one representative sequence of each group (I, >106; II, >093; III, >070; IV, >069; V, >264; VI, >233; VII, >080; VIII, >053; [numbers given as GIs in the data set]). Complete alignments are shown in Fig. S4.1 in the supplemental material.

4.2.6. Binding sites for RR

For several BceRS-BceAB-like systems, a RR binding site has been identified in the promoter of the transporter operon (Joseph *et al.*, 2004; Mandin *et al.*, 2005; Ohki *et al.*, 2003a; Ouyang *et al.*, 2010). The core consensus sequence appears to be an inverted repeat around a central ACA-N_4-TGT motif (de Been *et al.*, 2008). Based on the assumption that the ABC transporters identified in the present study are generally regulated by their adjacent 2CS, it should be possible to identify the RR binding sites in the promoter regions of the transporter operons. Moreover, the identification of multiple paralogous Bce-like modules from different groups in a single genome suggests sequence diversification in the

regulator binding sites, which may be group specific, to ensure sufficient regulatory specificity and insulation. We therefore searched the upstream regions of each transport operon in our data set for sequences with similarity to the consensus motif mentioned above. In 70% of all promoter regions analyzed, a complete putative binding site was identified (see Table S4.3 in the supplemental material), and the overall consensus of these 186 sequences was found to be TNACA-N_4-TGTAA, with an AT-rich central 4-nt spacer (Fig. 4.5A). Next, we sorted the putative binding site sequences according to the eight phylogenetic groups and derived a consensus sequence for each group (Fig. 4.5B). The only group that perfectly matched the entire previously proposed consensus of ANCTTACA-N_4-TGTAAGNT (de Been *et al.*, 2008) was group II. For most other groups, the best conserved positions appeared to be TGTNACA-N_4-TGTAAG, with a number of subtle differences between groups, as shown in Fig. 4.5B.

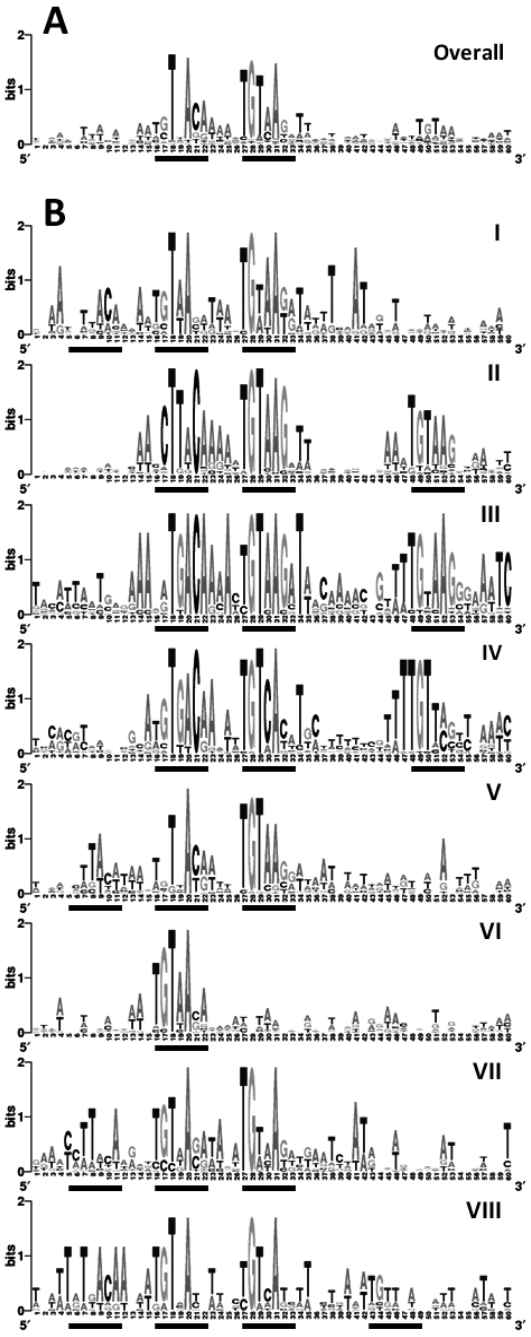

Figure 4.5. Consensus sequences of putative RR binding sites in transport operon promoters. Regions 250 bp upstream of ATPase start codons were scanned for motifs with similarity to previously identified binding sites of BceR-like RR. (A) Overall consensus derived from all 192 identified putative binding sites. (B) Sequence logos showing the identified consensus sequences for each group separately (panels I to VIII). The seven nucleotide repeats are marked by black bars below each logo. Sequence logos were calculated by WebLogo (Crooks *et al.*, 2004), with 20 (group I), 39 (group II), 16 (group III), 31 (group IV), 29 (group V), 7 (group VI), 31 (group VII), or 8 (group VIII) putative binding sites identified.

Strikingly, in all but group VI, a third or even fourth repeat could be identified, located up- and/or downstream of the most highly conserved putative binding site, usually with a spacing of about 13 bp (Fig. 4.5B). A similar situation has been reported for the *mbrAB* operon of *S. mutans*, where this additional repeat did not appear to be required for DNA binding by the RR MbrC (Ouyang *et al.*, 2010). Remarkably, in group VI, whose transporters lack a neighbouring 2CS, most promoters contained

only half of a binding site, raising the question as to how expression of these systems may be regulated. Taken together, each group was found to possess a subtly unique binding consensus, further supporting the regulatory relationship between transporters and 2CS from matching phylogenetic groups.

4.2.7. Loci with multiple ABC transporters

As described above, we identified 20 loci in which a single 2CS was located next to multiple transport permeases. These loci were analyzed in more detail by comparing the group assignments of the HK and all permeases to test if the coclustering also held true in such more complex gene arrangements. Furthermore, we scanned these regions for putative RR binding sites to identify likely regulated promoters. Schematic diagrams of each type of gene arrangement and group combination are shown in Fig. 4.6A.

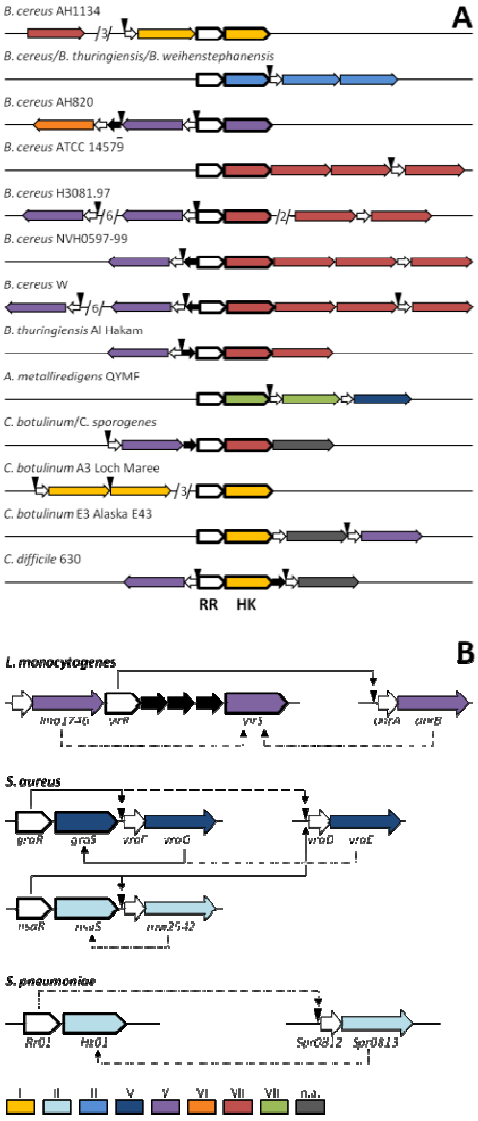

Figure 4.6. (A) Schematic of genetic loci containing multiple transporter genes. The single 2CS of each locus is shown in the center and marked by boldfaced contours. RR and HK components are indicated. HK and permease genes are colored according to their phylogenetic groups, as shown below. The ATPase genes are shown as small white arrows, and unrelated genes are shown as small black arrows. Where several unrelated genes were found in the gene locus, the line is broken and the number of intervening genes is indicated. Putative regulator binding sites are marked as black triangles. (B) Schematic of putative regulatory networks between separate genetic loci in example species. Markings of gene function, phylogenetic classification, and regulator binding sites are as described for panel A. Gene names or locus tags are shown. Experimentally proven regulation of promoters and interaction between transporters and HK is indicated by solid arrows (corresponding references are given in the text). The partial regulation of *vraDE* expression by GraRS is indicated by a dashed line. Hypothetical regulatory interactions are indicated by dotted lines. Drawings not to scale. *B. cereus*, *Bacillus cereus*; *B. thuringiensis*, *Bacillus thuringiensis*; *B. weihenstephanensis*, *Bacillus weihenstephanensis*; *A. metalliredigens*, *Alkaliphilus metalliredigens*; *C. botulinum*, *Clostridium botulinum*; *C. difficile*, *Clostridium difficile*.

In *Bacillus* species, two main patterns were found. The first pattern is comprised of a group III 2CS with several permeases from the same group. The second consists of a group VII 2CS with multiple permeases from the same group and one or two transporters from group V. In other genera, the arrangements were more varied (Fig. 4.6A). Stretches of permeases

from the same group often contained only a single ATPase gene, suggesting that either all permeases form complexes with the same ATPase or only one functional transporter is encoded. Another notable feature was the absence of any regulator binding sites upstream of permeases immediately following the 2CS, which may indicate that these permeases are cotranscribed with the 2CS and not regulated in response to a substrate antimicrobial peptide. In contrast, transporter operons divergently transcribed from the 2CS always contained a binding site, but experimental evidence is needed to determine whether regulation occurs via the neighbouring 2CS.

4.2.8. Putative assignment of regulatory 2CS to orphan transporters

While the vast majority of ABC transporters identified in our study were encoded in a genetic locus together with a BceRS-like 2CS, a number of transporters were lacking a neighbouring 2CS (Fig. 4.2A; see also Table S4.1 in the supplemental material). Based on the coclustering of permeases with HK, we therefore next posed the question if it was possible to identify likely candidate HK in the genomes, matching the classification of these orphan transporters. Of the genomes containing such systems, 25 were available in the Microbial Signal Transduction database MiST2 as of August 2010 (Ulrich and Zhulin, 2010). From these, the protein sequences of all HK were retrieved. For each organism, the HK were aligned with 15 representative sequences from our data set, and phylogenetic trees were calculated using the PROTDIST and FITCH programs of the Phylip package implemented in BioEdit (Felsenstein, 1989; Hall, 1999). Kinases clustering in one branch with the BceS-like reference sequences were identified in eight of the genomes, with each of these genomes containing only a single such sequence.

To determine the group assignment of these newly identified HK, their sequences were added to our data set and the phylogenetic tree was recalculated. Strikingly, for all eight new HK, their group assignment matched that of the respective orphan transporter (Table 4.3), suggesting that they may indeed be responsible for the regulation of the transporter's expression. In eight of the genomes for which no new HK could be identified, a second BceAB-like transporter was present, which belonged to the same phylogenetic group as the orphan transporter and which possessed a matching 2CS (Table 4.3). In these cases, it is conceivable that the 2CS regulates the expression of both transporters. Thus, for a total of 15 orphan transporters, we could assign a putative regulatory system. Furthermore, in 12 of these, a potential regulator binding site was present in the promoter regions, supporting regulation by a BceRS-like 2CS rather than by a different mode of gene regulation. Of the

15 systems for which no assignment of a putative regulatory 2CS was possible, 12 belonged to group VI, and 8 of these did not contain a putative RR binding site in their promoter region. It is unclear if or how expression of this group of transporters is regulated.

Table 4.3. Assignment of a putative regulatory 2CS to orphan transporters.

Organism	Permease		Binding site[c]	Histidine kinase	
	No. in data set[a]	Group[b]		GI or no. in data set[d]	Group[b]
A. metalliredigens	>224	V	-	>001	V
B. cereus	>272	VI	+	NM	
	>015	NA	-	NM	
	>220	VII	+	3002066	VII
B. thuringiensis	>222	VI	+	NM	
	>223	V	+	>076	V
	>226	VII	+	49477785	VII
E. faecalis	>255	VII	+	NM	
	>234	NA	-	29375511	NA
E. faecium	>258	VI	-	NM	
	>230	NA	-	69244814	NA
G. thermodenitrificans	>245	VII	+	-	
L. acidophilus	>233	VI	-	-	
L. brevis	>257	VI	-	-	
L. casei	>242	VI	-	NM	
L. delbrueckii	>251	VI	-	-	
L. gasseri	>235	VI	+	-	
L. johnsonii	>244	VI	+	-	
L. plantarum	>239	VI	-	-	
L. sakei	>259	VI	-	NM	
L. innocua	>264	V	+	>173	V
L. monocytogenes	>256	V	+	>174	V
L. welshimeri	>263	V	+	>176	V
P. pentosaceus	>250	VI	-	-	
S. aureus	>229	IV	+	>194	IV
S. epidermidis	>247	IV	+	>198	IV
S. haemolyticus	>238	IV	+	>202	IV
S. saprophyticus	>254	IV	+	73663370	IV
S. pneumoniae	>253	II	+	15901468	II
S. sanguinis	>266	II	+	125718462	II
S. thermophilum	>211	I	+	NM	

[a] Running number given as an identifier to each system in our data set (see Table S4.1 in the supplemental material).

[b] Group assignments according to our phylogenetic classification. NA, not assigned.

[c] The presence of a putative RR binding site in the region upstream of the ATPase gene of the transporter operon is indicated by "+", and lack of such a site is indicated by "-".

[d] The seven- to nine-digit numbers listed are the GI numbers of HK identified in the genome as part of a candidate 2CS for the regulation of orphan transporters. "-" indicates that no 2CS with a BceS-like HK could be identified in the genome. NM, no match, i.e., a BceRS-like 2CS associated with another Pep7E transporter is present in the genome but did not match the classification of the orphan transporter. Where it did match the classification of the transporter, its identifier in the data set is given preceded by ">".

[e] *E. faecalis*, *Enterococcus faecalis*; *G. thermodenitrificans*, *Geobacillus thermodenitrificans*; *L. acidophilus*, *Lactobacillus acidophilus*; *L. brevis*, *Lactobacillus brevis*; *L. casei*, *Lactobacillus casei*; *L. delbrueckii*, *Lactobacillus delbrueckii*; *L. gasseri*, *Lactobacillus gasseri*; *L. johnsonii*, *Lactobacillus johnsonii*; *L. plantarum*, *Lactobacillus plantarum*; *L. sakei*, *Lactobacillus sakei*; *L. innocua*, *Listeria innocua*; *L. welshimeri*, *Listeria welshimeri*; *P. pentosaceus*, *Pediococcus pentosaccus*; *S. epidermidis*, *Staphylococcus epidermidis*; *S. haemolyticus*, *Staphylococcus haemolyticus*; *S. saprophyticus*, *Staphylococcus saprophyticus*; *S. sanguinis*, *Streptococcus sanguinis*; *S. thermophilum*, *Symbiobacterium thermophilum*.

4.3. Discussion

In this study, we investigated the evolutionary and regulatory relationship between Pep7E-type ABC transporters and 2CS containing a BceS-like IM-HK, which together have been described as detoxification modules against peptide antibiotics (Bernard *et al.*, 2007; Meehl *et al.*, 2007; Ouyang *et al.*, 2010; Rietkötter *et al.*, 2008; Staroń *et al.*, 2011). Our comprehensive data set generated from five independent database searches almost exclusively contained proteins from Firmicutes bacteria, and all previously described systems were involved in resistance against different antimicrobial peptides. Thus, it appears that these modules indeed evolved as a specific detoxification mechanism of Firmicutes bacteria against this class of drugs. Interestingly, the substrate compounds are also produced primarily by Firmicutes species and target species closely related to the producer organism (Cotter *et al.*, 2005; Guder *et al.*, 2000). It is therefore hardly surprising to find a phyletic distribution of the specific resistance genes similar to that of the ability for peptide antibiotic biosynthesis.

4.3.1. Phyletic distribution

The majority of the proteins were found in only three orders, the *Bacillales*, *Lactobacillales*, and *Clostridiales*. Whether this shows the true distribution of such resistance modules or rather reflects the strong bias in genome sequencing projects toward biotechnologically and medically relevant species (about 90% of all currently fully sequenced Firmicutes genomes are from those three orders) cannot be answered at this stage. The presence of multiple paralogous systems in one organism, together with a lack of genomic context conservation beyond the four genes under investigation in this study, suggests a diversification of the resistance modules through horizontal gene transfer and gene duplication. This is further supported by the limited correlation we found between phylogenetic groups and species distribution. Another notable finding was the uneven distribution of the number of systems per organism, as follows: species belonging to the *Bacillales* or *Clostridiales* often possessed many, in some cases up to six, Pep7E transporters, while members of the *Lactobacillales* generally contained only one or two. It therefore appears that resistance to antimicrobial peptides is of particular importance to bacteria that are predominantly found in the soil but only plays a minor role in the more specialized habitats (milk, skin, gut, etc.) of most lactic acid bacteria. Such an increased

prevalence of antibiotic resistance mechanisms in the highly competitive soil environment has been demonstrated previously (D'Costa *et al.*, 2006).

4.3.2. Coevolution

Analysis of the genetic loci of the ABC transporters showed a conserved cooccurrence of the encoding genes with operons encoding BceRS-like 2CS. Furthermore, qualitative and quantitative comparisons of the phylogeny of the permease and HK components of the resistance modules revealed a clear coevolutionary relationship between the two protein families. Coevolution of two protein families can be due to a number of reasons, such as participation of both proteins in the same physiological process, interaction of both proteins with the same ligand, or direct physical contact formation between the two proteins (Pazos and Valencia, 2008). In the case of the latter scenario, the evolutionary process can also be referred to as coadaptation, because changes in one protein partner will necessitate changes in the other partner to ensure their continued ability to interact (Pazos and Valencia, 2008). All three mechanisms could be envisaged to have led to the coevolution of BceB-like permeases and BceS-like HK, based on their joined function in antimicrobial peptide resistance, potential interaction of both proteins with the substrate peptide, and protein-protein interaction between permease and HK. Because of the absolute requirement for the ABC transporter in stimulus detection by the 2CS, such a direct interaction between both components is highly likely but still needs to be experimentally validated.

4.3.3. Substrate specificity

The classification of transport permeases into the eight phylogenetic groups presented here does not allow predictions regarding the function of uncharacterized systems, because we could not find any correlation between group assignment and substrate specificity. The same holds true for ECD sequence and substrate range. It may nevertheless be true that the ECD contains the substrate binding domain of the transporters and that the high degree of variability is required for adaptation to the large variety of substrates recognized by these modules. To date, the mode of substrate recognition by these transporters remains enigmatic, especially since a detailed study of inducer and resistance spectra of the Psd and Bce modules of *B. subtilis* showed their astonishing ability to distinguish between very similar substrates like actagardine and mersacidin while at the same time being able to

recognize such structurally different substrates as nisin and enduracidin (Staroń *et al.*, 2011).

4.3.4. Regulatory networks and genomic rearrangements

Because permeases and HK from one genetic locus are thought to be functionally linked and were found to generally fall into matching phylogenetic groups, we used our classification system to identify putatively corresponding 2CS for ABC transporters whose genes were not directly associated with a 2CS operon. Such candidate systems, in which the HK of the 2CS belonged to the same group as the transport permease, could be identified in about 50% of orphan transporters analyzed. While experimental validation is required for each 2CS-transporter pair to prove their regulatory relationship, such evidence is available in the literature for two of the systems included in our analysis.

In *L. monocytogenes*, one of the ABC transporters (Lmo1747-1746) is encoded in a genetic locus with the VirRS 2CS (Mandin *et al.*, 2005), with both the permease and the HK VirS belonging to group V (Fig. 4.6B). It was recently shown that VirRS does not seem to regulate expression of the Lmo1747-1746 transporter (Mandin *et al.*, 2005), which explains the absence of a putative RR binding site upstream of Lmo1747. However, VirRS (Table 4.3, >174) is responsible for expression of a second orphan Pep7E-type ABC transporter, AnrAB (Table 4.3, >256) (Collins *et al.*, 2010), which also belongs to group V and does possess a putative binding site in its promoter region (Fig. 4.6B). This association of a ABC transporter and 2CS from separate genetic loci was thus correctly identified in our analysis. In further studies, it would be interesting to test if one or both of the two ABC transporters play a role in signaling, together with VirRS.

The second example is *S. aureus*, which possesses three Pep7E-type transporters, one from group II and two from group IV. The group II transporter (MW2543-2542) is associated with its own corresponding 2CS, which was recently shown to be involved in nisin resistance and thus designated NsaRS (Blake *et al.*, 2011). Of the group IV transporters, only one is encoded adjacent to a 2CS (Fig. 4.6B). This is the GraRS(ApsRS)-VraFG module (Table 4.3, >194), where it has been shown that the GraRS 2CS regulates expression of the VraFG transporter (Li *et al.*, 2007a; Meehl *et al.*, 2007). The second group IV transporter is VraDE (Table 4.3, >229), whose expression is partially controlled by GraRS (Li *et al.*, 2007a), again matching our prediction. However, expression of *vraDE* is also dependent on the 2CS NsaRS, as was shown recently in a study on bacitracin resistance of *S. aureus* (Yoshida *et al.*, 2011). Mutants in *nsaRS* (named *bceRS* in the cited

study), *mw2543-2542*, or *vraDE* all displayed increased sensitivity to bacitracin (Yoshida *et al.*, 2011), thus indicating the existence of a more complex regulatory network: while it appears clear that NsaRS regulates the expression of both transporters (Yoshida *et al.*, 2011), it has not been shown if both also mediate the actual resistance or if one transporter merely contributes to the signaling cascade. According to our classification, such a regulatory interaction with the HK would be possible only for the MW2543-2542 transporter, which belongs to the same phylogenetic group as NsaS (Fig. 4.6B).

A further example is the Spr0812-Spr0813 transporter from *S. pneumoniae* (Table 4.3, >253), for which we identified a 2CS as the candidate regulator, based on coclustering of both permease and HK into group II. This is the same 2CS identified previously, albeit without experimental evidence, as the most likely regulator of *spr0812-spr0813* expression (Fig. 4.6B) (Becker *et al.*, 2009). Thus, the phylogenetic classification presented here provides a useful tool for the prediction of likely candidate systems for investigations into the regulatory mechanisms of such orphan transporters. The only group of permeases for which we could never identify a corresponding 2CS was group VI. Since a complete RR binding site was also lacking for most, the regulation of transporters from this group remains unclear.

4.3.5. Functional link between HK and transport permeases

The cooccurrence and coevolution of Pep7E transporters and the BceRS-like 2CS described in this study lead us to hypothesize that a functional link always exists between these systems, in which they cooperate in both signalling and detoxification, as was shown for the bacitracin resistance modules of *B. subtilis* and *S. mutans* (Ouyang *et al.*, 2010; Rietkötter *et al.*, 2008). Based on their lack of an obvious input domain, BceS-like HK have been termed intramembrane-sensing HK and are thought to detect a stimulus at or within the cytoplasmic membrane (Mascher, 2006). Even among the IM-HK, the BceS-like proteins represent a kind of “minimal” system, because they do not even contain any cytoplasmic domains common to other HK besides the DHp and catalytic domains required for autophosphorylation. Taken together with the observation that our database searches did not reveal any BceS homologues in genomes lacking a Pep7E transporter and only few transporters in genomes lacking a BceS homologue, we now propose that HK of the BceS type cannot actually sense any stimulus directly but rather transfer the information received from their associated ABC transporter to their cognate RR.

An interesting scenario is presented in organisms containing multiple Pep7E transporters from the same phylogenetic group, encoded either in the same genetic locus or at different loci on the chromosome (Fig. 4.6). It is conceivable that in such situations, the two roles of the transporter have been split, with one transporter responsible for stimulus detection and the other for detoxification. Such an arrangement would explain the absence of putative regulator binding sites upstream of some transporters (e.g., *lmo1747-1746* in *L. monocytogenes* or transporters immediately downstream of the 2CS in the tandem loci shown in Fig. 4.6A), as a transporter solely responsible for stimulus detection would not necessarily need to be induced in the presence of its substrate antimicrobial peptide.

In summary, the data presented here show the widespread distribution of unique and self-sufficient detoxification modules against antimicrobial peptides among Firmicutes bacteria and suggest a novel signalling and resistance mechanism involving formation of a sensory complex between transport permeases and HK. To date, it has not been shown whether signalling is mediated by direct protein-protein interactions between permeases and HK or whether the function of the transporter is merely the translocation of the substrate to the site at which it is directly sensed by the HK. However, initial results from bacterial two-hybrid assays in *E. coli* suggest that the BceB permease and BceS HK of *B. subtilis* are indeed able to interact with each other (our unpublished results), lending support to the former hypothesis. While a number of key characteristics, such as the nature of such a sensory complex, the basis for substrate specificity, and the mechanism of both signal transduction and detoxification, will have to be addressed experimentally in the future, our study provides an ideal starting point for such investigations, especially regarding the elucidation of regulatory networks within and between these modules.

4.4. Materials and methods

4.4.1. Data acquisition

An initial data set was generated based on homology to BceRSAB (Ohki *et al.*, 2003a), PsdRS-AB (Staroń *et al.*, 2011), and YxdJK-LM (Joseph *et al.*, 2004) of *B. subtilis* by performing a BLASTP search (Altschul *et al.*, 1997) with default parameters (scoring parameters matrix, BLOSUM62; gap opening penalty, 11; gap extension penalty, 1) of all microbial genomes available in the NCBI database as of April 2009, using each of the 12 protein sequences as a query. Of each result list, the top 250 hits were chosen and sorted to match HK, RR, ATPase, and permease components belonging to the same genetic locus. After removal of duplicates, 185 systems were obtained. Because we found BceB-like permeases to contain a PFAM FtsX domain, located in the region encompassing transmembrane helices 2 to 4 (Fig. 4.1), a second search was conducted by querying the SMART 6 database (Letunic *et al.*, 2009) for proteins containing such a domain. Of the 4,492 hits obtained, 274 were found to possess the typical transmembrane architecture of BceB-like proteins, containing 10 transmembrane helices with a large loop of approximately 200 amino acids between helices 7 and 8. These proteins were added to our original data set with removal of duplicates, resulting in a total number of 266 systems. In the Archaeal and Bacterial ABC Systems Database (ABCdb; January 2010

version) (http://www-abcdb.biotoul.fr), BceB-like permeases belong to subfamily M_9. We therefore searched this subfamily for proteins with greater than nine predicted transmembrane sections. The resulting 114 hits were checked for the BceB-like domain architecture, but no sequences not already contained in our data set were found. Additionally, the TCDB (Saier *et al.*, 2009) was searched for Pep7E family transporters, but again, no new BceB-like sequences were identified in this database. Final refinement of our data set was carried out by analyzing the genomic neighborhood of all obtained sequences for genes with similarity to RR, ATPases, and HK using the MicrobesOnline database (Dehal *et al.*, 2010). HK and their corresponding RR were included only if the kinase showed the same domain architecture as BceS of two transmembrane helices separated by approximately 2 to 10 amino acids. The resulting data set is shown in Table S4.1 in the supplemental material.

4.4.2. Phylogenetic analyses

Multiple sequence alignments of permeases and HK were obtained with ClustalW2 (Thompson *et al.*, 2002) using the Gonnet280 substitution matrix and default parameters (gap opening penalty, 10; gap extension penalty, 0.2; gap distance penalty, 5) at the European Bioinformatics Institute website (http://www.ebi.ac.uk). The resulting alignments were manually edited to remove gaps in order to allow calculation of meaningful phylogenetic trees (Baldauf, 2003). In particular, the region containing the ECD of the permeases, corresponding to positions 310 to 520 in *B. subtilis* BceB, was deleted from the alignment followed by realignment of the remaining sequence, because of the very poor conservation observed in this region. Phylogenetic trees were constructed from both of the sequence alignments with the neighbor-joining method implemented in BioNJ (Gascuel, 1997), using the Dayhoff PAM substitution matrix, 100 bootstraps, and otherwise default parameters on the Phylogeny.fr website (Dereeper *et al.*, 2008). Graphical editing of the phylogenetic trees to allow direct comparison by a mirror-image display was performed in TreeDyn 198.3 (Chevenet *et al.*, 2006).

4.4.3. Analysis of coevolution

Coevolving protein families undergo simultaneous sequence changes or conservation (Goh *et al.*, 2000). In addition to our data set of BceRS-BceAB-like modules (see Table S4.1 in the supplemental material), we compiled two new data sets to use as controls (i.e., not coevolving proteins). For this, we analyzed the genomes represented in our original data set for a HK and a transport permease that were present in a large number of species but were not functionally associated with each other. These criteria were fulfilled by homologues of the HK YycG of *B. subtilis* and those of the permease OppB of *B. subtilis*. A total of 26 of the 96 different species represented in our data set were found to contain homologues to both of the proteins (see Table S4.2 in the supplemental material). For further analyses, we used the six obtained sequence families (BceR, BceS, BceA, BceB, YycG, and OppB) in data sets of 26 ("core") and 180 ("extended"; only for Bce homologues) sequences. For each data set, a multiple sequence alignment of all sequences ("complete") was generated using ClustalW2 as described above, consisting of 26 (core set) or 180 (extended set) sequences. Additionally, 50 randomly chosen subsets of 20 sequences were extracted from each core or extended set ("sampled") and realigned.

The sampled subset comparisons estimate alignment stability in terms of single sequence effects. Pairwise calculation of alignment correlations, using either complete or sampled alignments, was then carried out as follows. For a family k containing the sequences $s_i \in S^k, i = 1 \dots |S^k|$ the comparison was done by computing a matrix M of normalized pairwise sequence identities (*id*) $M^k = (m^k)_{ij}$, $i,j \in \{1 \dots |S^k|\}$ where $m_{ij}^k = id(s_i, s_j)/(\overline{|s_i| + |s_j|})$, $s_i, s_j \in S^k, s_i \neq s_j$. Two matrices for families k and l were then compared using the linear correlation coefficient (CC) as introduced by Goh and co-workers (Goh *et al.*,2000) :

$$CC = \frac{\sum_{i=1}^{|n|-1}\sum_{j=i+1}^{|n|}\left(M_{ij}^k - \overline{M^k}\right)\left(M_{ij}^l - \overline{M^l}\right)}{\sqrt{\sum_{i=1}^{|n|-1}\sum_{j=j+1}^{|n|}\left(M_{ij}^k - \overline{M^k}\right)^2}\sqrt{\sum_{i=1}^{|n|-1}\sum_{j=j+1}^{|n|}\left(M_{ij}^l - \overline{M^l}\right)^2}}$$

To assess the significance of the CC for the pair of BceB-like and BceS-like sequences, we used all pairs containing OppB and/or YycG as negative controls and the pairs of BceA-BceB, BceS-BceR as positive controls, computing both core and sampled comparisons. The results served as background distributions of known noninteracting and interacting protein families.

4.4.4. Secondary structure prediction of the large extracellular domain of transport permeases

For a detailed analysis of the ECD of the permeases, the region corresponding to positions 310 to 520 of B. *subtilis* BceB, which had previously been removed from the sequence alignments, was extracted from all sequences. The obtained sequences were sorted into eight sets according to the classification of permeases described in this study. From each group, a representative set was chosen. For this, pairwise identities were calculated for all sequence pairs, and from each cluster sharing over 80% identity, only one randomly chosen sequence was included in further analyses. Multiple sequence alignments for each of the eight sets were generated with ClustalW2 as described above. Additionally, the secondary structure was predicted for several sequences in each set using JPred3 (Cole *et al.*, 2008), and the results were integrated into the sequence alignment. A consensus secondary structure for each of the eight groups was deduced from these alignments.

4.4.5. Identification of RR binding sites

Upstream regions of operons encoding Pep7E ABC transporters were retrieved from the MicrobesOnline database (Dehal *et al.*, 2010). Generally, 250 nucleotides (nt) upstream of the start codons were analyzed, unless the distance to the nearest upstream gene was greater than 250 bp. In this case, the whole intergenic region was analyzed. If the gene encoding the ATPase subunit was not the first in the predicted operon, both 250 bp upstream of the ATPase gene and the region upstream of the preceding gene were analyzed. Retrieved sequences were subjected to conserved motif searches using MEME (Bailey and Elkan, 1994) (http://meme.sdsc.edu/) with the following parameters: distribution, any number of repetitions; width, minimum of 5, maximum of 20. Analysis with MEME was followed by in-depth manual analysis. Sequence motifs were illustrated based on a position weight matrix using the WebLogo tool (Crooks *et al.*, 2004) (http://weblogo.berkeley.edu).

4.5. Acknowledgments

We thank Ralf Zimmer for valuable input regarding calculations of correlation coefficients for multiple sequence alignments.

Work done in our laboratory was supported by grants-in-aid from the Fonds der Chemischen Industrie (to S.G.) and a grant from the Deutsche Forschungsgemeinschaft (to T.M.; grant MA2837/1-3).

CHAPTER 5

Peptide antibiotic sensing and detoxification modules of *Bacillus subtilis*

This chapter has been adapted from:

A. Staroń, D. E. Finkeisen, T. Mascher

Antimicrob. Agents Chemother. (2011) 55(2):515-525

Chapter 5
Peptide antibiotic sensing and detoxification modules of *B. subtilis*

Peptide antibiotics are produced by a wide range of microorganisms. Most of them target the cell envelope, often by inhibiting cell wall synthesis. One of the resistance mechanisms against antimicrobial peptides is a detoxification module consisting of a two-component system and an ABC transporter. Upon the detection of such a compound, the two-component system induces the expression of the ABC transporter, which in turn removes the antibiotic from its site of action, mediating the resistance of the cell. Three such peptide antibiotic-sensing and detoxification modules are present in *B. subtilis*. Here we show that each of these modules responds to a number of peptides and confers resistance against them. BceRS-BceAB (BceRS-AB) responds to bacitracin, plectasin, mersacidin, and actagardine. YxdJK-LM is induced by a cationic antimicrobial peptide, LL-37. The PsdRS-AB (formerly YvcPQ-RS) system responds primarily to lipid II-binding lantibiotics such as nisin and gallidermin. We characterized the *psdRS-AB* operon and defined the regulatory sequences within the P_{psdA} promoter. Mutation analysis demonstrated that P_{psdA} expression is fully PsdR dependent. The features of both the P_{bceA} and P_{psdA} promoters make them promising candidates as novel whole-cell biosensors that can easily be adjusted for high-throughput screening.

5.1. Introduction

Peptide antibiotics are produced by a wide range of organisms and can be synthesized both ribosomally and nonribosomally (Jenssen *et al.*, 2006). Nonribosomally synthesized antimicrobial compounds are produced mainly by bacteria and are often posttranslationally modified (Hancock and Chapple, 1999). They can form linear polypeptides, such as gramicidin (Killian, 1992), or cyclic molecules, such as bacitracin and polymyxins (Landman *et al.*, 2008; Ming, 2003). Glycopeptides (e.g., vancomycin and ramoplanin) consist of a peptide backbone, which is further modified by glycosylation and methylation (Donadio and Sosio, 2008). Ribosomally synthesized peptides, including lantibiotics and defensins, are more widespread and are produced by mammals, amphibians, insects, plants, and bacteria (Hancock and Chapple, 1999). They are often derived from small precursor peptides and are usually small (10 to 50 amino acids), with an overall positive charge and a significant number of hydrophobic residues (Hancock and Sahl, 2006).

Most peptide antibiotics target crucial steps in cell wall biosynthesis. The bacterial cell wall is a vitally important structure that gives the cell its shape, separates it from its environment, and acts as a molecular sieve (Jordan *et al.*, 2008). This makes it an important target for many antimicrobial compounds, which very often act by sequestering lipid II and by blocking transglycosylation and transpeptidation steps (Schneider and Sahl, 2010). Vancomycin, lantibiotics, ramoplanin, and many defensins bind different moieties of lipid II (Jordan *et al.*, 2008; Schmitt *et al.*, 2010; Schneider *et al.*, 2010). Vancomycin binds to the C-terminal Lys-D-Ala-D-Ala of the pentapeptide chain of the cell wall precursor (Breukink and de Kruijff, 2006). Nisin and nisin-like lantibiotics bind the pyrophosphate of lipid II, whereas the binding site of mersacidin and related lantibiotics includes both the MurNAc-GlcNAc sugar moiety and the pyrophosphate (Cudic *et al.*, 2002). Ramoplanin requires the presence of MurNAc-Ala-Glu pyrophosphate in order to bind to lipid II. Bacitracin inhibits a different step of cell wall biosynthesis by binding undecaprenyl pyrophosphate and inhibiting its dephosphorylation, thereby blocking its recycling and, ultimately, cell wall biosynthesis (Rietkötter *et al.*, 2008).

Because the production of peptide antibiotics is widespread, the presence of an appropriate stress response system is necessary both for the producer strains as well as for those bacteria that are exposed to these compounds in their natural habitat. One type of detoxification system against peptide antibiotics found mainly in Gram-positive bacteria is a module consisting of an ABC transporter, which is genetically and functionally linked to a 2CS (Jordan *et al.*, 2008; Joseph *et al.*, 2002; Mascher, 2006). Upon sensing the signal (i.e., the presence of the antibiotic), the HK phosphorylates its cognate RR, which in turn induces the expression of the ABC transporter genes. The transporter facilitates the removal of the antibiotic compound from its active site (Jordan *et al.*, 2008).

While few of these systems have been experimentally characterized to date, all respond and mediate resistance to cell wall peptide antibiotics. In *S. aureus*, the GraRS-VraFG system was previously found to respond to vancomycin, polymyxin B (Meehl *et al.*, 2007), gallidermin (Herbert *et al.*, 2007), and defensins (Kraus *et al.*, 2008). Homologous proteins mediate resistance to nisin in *L. lactis* (Kramer et al., 2006) and to bacitracin in *S. mutans* (Ouyang et al., 2010; Tsuda et al., 2002).

The genome of *B. subtilis* contains three such peptide-sensing and detoxification (PSD) modules consisting of a 2CS and an ABC transporter: BceRS-AB, YxdJK-LM, and YvcPQ-RS (Fig. 5.1 A and B). BceRS-AB (PSD1) was initially identified as a bacitracin-specific detoxification module (Mascher *et al.*, 2003; Ohki *et al.*, 2003a). Recently, it was

also shown to respond to the defensin plectasin (Schneider *et al.*, 2010). The YxdJK-LM system (PSD2) responds to the human antimicrobial peptide LL-37 (Pietiäinen *et al.*, 2005). The third system, YvcPQ-RS (PSD3), was initially described as a part of bacitracin stress response network (Mascher *et al.*, 2003).

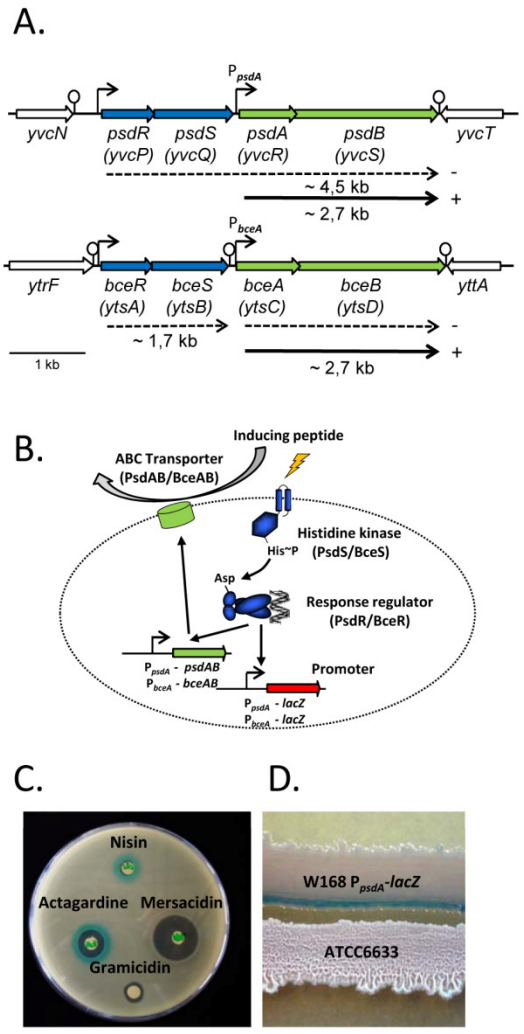

Figure 5.1. Organization of the *psdRS-AB* and *bceRS-AB* loci and induction by peptide antibiotics. (A) Graphic representation of the *psdRSAB* and *bceRSAB* loci. Genes belonging to the *psd* and *bce* loci are shown in blue (two-component system) and green (ABC transporter); the genes flanking both operons are white. Promoters are marked with bent arrows, and putative terminators are represented by vertical bars and a circle. (B) Regulatory principle and genetic setup of the Psd and Bce biosensor strains. The RR (PsdR or BceR), activated by the sensor kinase (PsdS or BceS), binds to its target promoter and induces the expression of the ABC transporter encoding the *psdAB* or *bceAB* operon (detoxification) and *lacZ* (production of β-galactosidase). (C) Example of a qualitative β-galactosidase assay with nisin, actagardine, mersacidin, and gramicidin (disk diffusion assay). The reporter strain carrying a chromosomal P_{psdA}-*lacZ* fusion was used in soft-agar overlays on LB plates containing X-Gal. Bactericidal activity is visualized as the presence of a growth inhibition zone around the filter disk, and P_{psdA}-dependent induction is visualized as a blue ring around the inhibition zone. (D) Qualitative β-galactosidase assay with the subtilin producer strain *B. subtilis* ATCC 6633. P_{psdA}-*lacZ* reporter strain TMB299 was streaked out onto LB-X-Gal plates directly next to the *B. subtilis* ATCC 6633 cultures. The appearance of a blue zone in the reporter strain next to the subtilin-producing strain shows the induction of P_{psdA}.

In this study, we aimed to identify novel inducers for all three PSD modules. Using disk diffusion assays and promoter-*lacZ* fusions, we screened a wide variety of cell envelope-active compounds, including many peptide antibiotics. In addition, we performed a comprehensive meta-analysis of all previously published stress response microarray data sets in order to identify additional inducers of *bceAB*, *yvcRS*, and *yxdLM* expression.

We present evidence that the BceRS-AB system is not only a bacitracin-specific resistance determinant but rather a PSD module that responds to a broader spectrum of compounds, including the lantibiotics mersacidin and actagardine as well as the defensin plectasin. This module also mediates a certain level of resistance to these compounds.

For PSD3, it was recently shown that the weak induction of the *yvcR* promoter by bacitracin is the result of a cross-activation of the RR YvcP by the HK of the paralogous

BceRS-AB system (Rietkötter *et al.*, 2008). In this study, we identified lipid II-binding peptides, mainly lantibiotics but also one lipopeptide, enduracidin, as inducers of *yvcRS* expression. We further demonstrate that the ABC transporter YvcRS confers resistance against its inducers. Based on the primary inducers and the resistance profile, we renamed the *yvcPQRS* locus to *psdRSAB* (for peptide antibiotic sensing and detoxification).

Our data demonstrate that the P_{bceA}- and P_{psdA}-based reporter strains are more sensitive and more specific biosensors for lipid II-binding peptide antibiotics than any of the established cell wall antibiotic biosensors currently available, such as the P_{ypuA}- and P_{liaI}-derived reporter strains (Mascher *et al.*, 2004; Urban *et al.*, 2007). We provide evidence indicating that both biosensors could easily be modified to accommodate high-throughput screens for novel antimicrobial compounds using pure compounds, culture supernatant, or even, directly, the producing strains.

5.2. Results and discussion

5.2.1. Screen for inducers of *bceAB*, *yxdLM*, and *psdAB* expression

In order to identify specific inducers of the three detoxification modules of *B. subtilis*, we used reporter strains carrying a chromosomal pAC6-based transcriptional *lacZ* fusion integrated at the *amyE* locus in the presence of an intact 2CS-ABC system (Fig. 5.1B). Disk diffusion assays were used to screen a variety of peptide antibiotics and other cell-envelope-active compounds for their ability to induce β-galactosidase expression. In this assay, filter plates were placed onto soft-agar overlays on LB plates containing X-Gal. The bactericidal activity of a given antibiotic was visualized as the presence of a growth inhibition zone around the filter disk, and the promoter induction was visualized as a blue ring around the inhibition zone.

To gain a more comprehensive understanding of the inducer spectra, these *in vivo* studies were complemented with an *in silico* meta-analysis of a large panel of genome-wide expression profiles of *B. subtilis* after treatment with inhibitory compounds of different modes of action. This analysis included membrane-active compounds; antibiotics targeting fatty acid biosynthesis, folate biosynthesis, cell wall biosynthesis, translation, DNA topology, and glycosylation; cationic antimicrobial peptides; and metal ions (Cao *et al.*, 2002; Hutter *et al.*, 2004b; Mascher *et al.*, 2003; Moore *et al.*, 2005; Pietiäinen *et al.*, 2005; Wecke *et al.*, 2009) (see footnotes to Table 5.1 for details). Of all the compounds tested, only the peptide antibiotics listed in Table 5.1 acted as inducers of the three PSD modules.

For comparison, the results obtained for an already established peptide antibiotic biosensor, based on P_{liaI} (Mascher *et al.*, 2004), are also listed.

Table 5.1. Inducers of P_{psdA}, P_{bceA}, P_{yxdL}, and P_{liaI} expression[a]

Antibiotic	Class	Charge[b]	Lipid II binding[b,c]	Pore forming[b,c]	Concn[d]	Induction[e] (reference)			
						P_{bceA}	P_{psdA}	P_{yxdL}	P_{liaI}
Bacitracin	Cyclic peptide	+2	-	-	50 µg/ml	+++[g]	+[g]	-	+++[h]
Enduracidin	Cyclic lipopeptide	+4	+	-	0.025 µg/ml	-	+++[i]	-	+++[i]
Ramoplanin	Cyclic lipoglyco-peptide	+2	+	-	5 µg/ml	-	-	-	+++[j]
Vancomycin	Glycopeptide	+2	+	-	2 µg/ml	-	-	-	+++[k]
LL-37	Cathelicidin	+6	-	+	ND	-	-	++[l]	-
Plectasin	Defensin	+2	+	-	ND	+++[m]	-	-	-
Actagardine	Lantibiotic	-1	+	-	3-10 µg/ml	+++	+++	-	-
Duramycin	Lantibiotic	+1	-	-	ND	-	-	-	+++
Gallidermin	Lantibiotic	+2	+	-	100 µg/ml	-	+++	-	+++[n]
Mersacidin	Lantibiotic	-1	+	-	10 µg/ml	+++	-	-	-
Nisin	Lantibiotic	+5	+	+	2 µg/ml	-	+++	-	+++[j]
Subtilin	Lantibiotic	+2	+	+	0.25%[f]	-	+++	-	+++[n]

[a] All inducers are peptide antibiotics. Based on the meta-analysis of transcriptome data sets and our disk diffusion screen, the following compounds do not induce P_{psdA}, P_{bceA}, and P_{yxdL} (Cao *et al.*, 2002; Hutter *et al.*, 2004b; Mascher *et al.*, 2003; Moore *et al.*, 2005; Pietiäinen *et al.*, 2005; Wecke *et al.*, 2009): cell wall biosynthesis inhibitors (amoxicillin, cefalexin, cephalosporin, cefotaxime, cefoxitin, daptomycin, D-cycloserine, friulimicin, oxacillin, penicillin G, Pep5, PG-1, phosphomycin, ristocetin, sublancin 168, and tunicamycin), membrane-active compounds and ionophors (gramicidin A, monensin, nigericin, nitrofurantoin, polymyxin B, poly-L-lysine, and Triton X-114), compounds interfering with DNA topology (ciprofloxacin, coumermycin, moxifloxacin, nalidixic acid, norfloxacin, and novobiocin), fatty acid biosynthesis inhibitors (triclosan and cerulenin), folate biosynthesis inhibitors (dapsone, sulfacetamide, sulfamethizole, and trimethoprim), inhibitors of protein biosynthesis (chloramphenicol, clarithromycin, clindamycin, erythromycin, fusidic acid, neomycin, puromycin, spectinomycin, and tetracycline), metal ions [Ag(I), Cd(II), Cu(II), Ni(II), Zn(II), and As(V)], and miscellaneous compounds (actinonin, azaserine, doxorubicin, ethidium bromide, hexachlorophene, and rifampin).
[b] based on the BACTIBASE database (http://bactibase.pfba-lab-tun.org/) (Hammami *et al.*, 2010).
[c] +, yes; -, no.
[d] Concentration resulting in the highest level of induction in quantitative β-galactosidase assays. ND, not determined.
[e] -, no induction; +, weak inducer; +++, strong inducer.
[f] Percent *B. subtilis* ATCC 6633 supernatant.
[g] (Rietkötter *et al.*, 2008)
[h] (Mascher *et al.*, 2003)
[i] (Rukmana *et al.*, 2009)
[j] (Mascher *et al.*, 2004)
[k] (Cao *et al.*, 2002)
[l] (Pietiäinen *et al.*, 2005)
[m] (Schneider *et al.*, 2010)
[n] (Burkard and Stein, 2008)

5.2.1.1. PSD1. The BceRS-AB system was initially identified as a part of the bacitracin stress response network and is an important bacitracin resistance determinant in *B. subtilis*

(Mascher *et al.*, 2003; Ohki *et al.*, 2003a). Its expression was recently reported to be upregulated after treatment with a fungal defensin, plectasin (Schneider *et al.*, 2010). Using disk diffusion assays, we found that the expression of P_{bceA} is also strongly induced by two lipid II-binding lantibiotics, actagardine (formerly gardimycin) and mersacidin. We did not observe any induction after treatment with other lantibiotics, including nisin, sublancin, and duramycin (Table 5.1). We also did not identify any other compound from the microarray meta-analysis that induced the expression of P_{bceA}.

Bacitracin is a cyclic dodecylpeptide (Fig. 5.2) that binds undecaprenyl pyrophosphate and inhibits cell wall biosynthesis by preventing the recycling of this lipid carrier (Stone and Strominger, 1971). The remaining three inducing compounds, plectasin, mersacidin, and actagardine, also inhibit cell wall biosynthesis but by binding lipid II (Bierbaum and Sahl, 2009; Schneider *et al.*, 2010). Actagardine and the closely related mersacidin belong to the class of lantibiotics with compact globular structures (Willey and van der Donk, 2007) (Fig. 5.2). Both compounds share a conserved structure that is predicted to form the lipid II-binding pocket (Szekat *et al.*, 2003) and bind the MurNAc-GlcNAc pyrophosphate (Cudic *et al.*, 2002). Plectasin is also thought to bind the pyrophosphate moiety of lipid II (Schneider *et al.*, 2010). Surprisingly, ramoplanin, which has been predicted to possess a backbone fold similar to that of actagardine and mersacidin and, therefore, a similar mechanism of action (Cudic *et al.*, 2002), does not induce the expression of the *bceAB* operon (Table 5.1). Therefore, the exact nature of the signal sensed by the PSD1 module BceRS-AB remains to be elucidated.

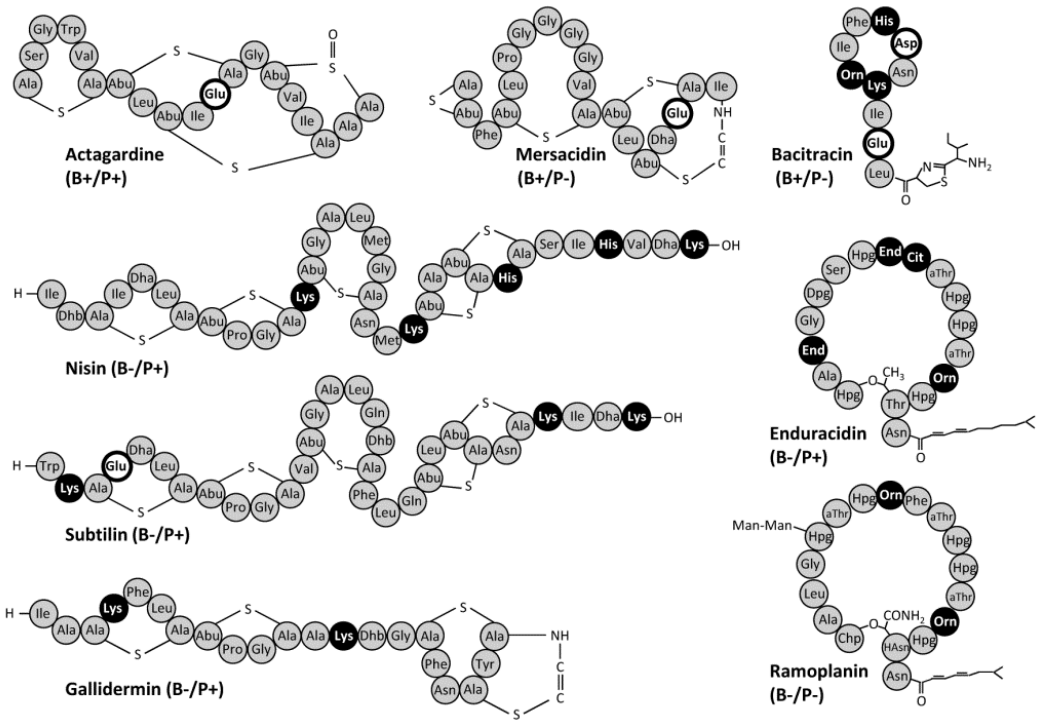

Figure 5.2. Schematic structures of peptide antibiotics inducing the Psd and Bce systems. Amino acids are represented by labeled gray circles. Charged amino acids are highlighted black (positive charge) or white (negative charge). Abu, aminobutyric acid; Chp, 3-chloro-4-hydroxyphenylglycine; Cit, citrulline; Dha, didehydroalanine; Dhb, 2,3-didehydrobutyrine; Dpg, 3,5-dichloro-4-hydroxyphenylglycine; End, enduracididine; HAsn, β-hydroxyasparagine; Hpg, hydroxyphenylglycine; Man, mannose; Orn, ornithine; aThr, allo-threonine. The induction of the Bce/Psd system is indicated by B+/P+.

5.2.1.2. PSD2. For the YxdJK-LM module, we did not identify any novel inducers of P_{yxdL} (Table 5.1). The *yxdL* promoter was previously described to respond to the cationic antimicrobial peptide LL-37 (Pietiäinen *et al.*, 2005). This cathelicidin is produced by human neutrophiles and shows antimicrobial activity against both Gram-positive and Gram-negative bacteria (Turner *et al.*, 1998). LL-37 was shown previously to induce the expression of a homologous system in *S. aureus* (Kraus *et al.*, 2008). As it is unlikely that a soil bacterium like *B. subtilis* responds specifically to a human neutrophil peptide, we suggest that the actual physiological inducer shares some chemical properties with LL-37 but remains to be identified.

5.2.1.3. PSD3. The PsdRS-AB system was initially described to respond to bacitracin (Mascher *et al.*, 2003; Ohki *et al.*, 2003a). We have recently shown that the weak bacitracin induction of the system is a result of the cross-activation of PsdR by the paralogous BceRS-AB system (Rietkötter *et al.*, 2008). Using disk diffusion assays, we found that P_{psdA} is induced by a lipid II-binding lipopeptide, enduracidin, and the lipid II-binding lantibiotics nisin, subtilin, actagardine, and gallidermin. No induction was

observed for other lantibiotics, including mersacidin, Pep5, sublancin, and duramycin, as well as for lipid II-binding antibiotics of other classes, including vancomycin and ramoplanin (Fig. 5.1C and Table 5. 2).

The majority of inducers of P_{psdA} expression are cationic lipid II-binding lantibiotics, namely, nisin, subtilin, and gallidermin (Fig. 5.2). They all share an N-terminal lipid II-binding motif, a so-called pyrophosphate cage (Willey and van der Donk, 2007). Actagardine, another inducing lantibiotic, differs from the above-mentioned compounds, as it lacks a positive charge (Table 5.1). However, it was proposed previously that lantibiotics of this family require Ca^{2+} ions to obtain full activity and that these ions improve the interaction with the bacterial membrane by conferring a positive overall charge (Böttiger *et al.*, 2009). Mersacidin, a lantibiotic closely related to actagardine, does not induce *psdAB* expression. These two lantibiotics are highly similar (Fig. 5.2), but the activity spectrum of actagardine is different from that of mersacidin: while actagardine is most active against streptococci and displays low-level activity against staphylococci, mersacidin is most active against staphylococcal species (Szekat *et al.*, 2003), including methicillin-resistant *S. aureus* (MRSA) (Kruszewska *et al.*, 2004). As actagardine induces P_{psdA} expression whereas mersacidin does not, the PsdRS-AB system has to be able to distinguish between those two closely related compounds.

Enduracidin, a cyclic lipopeptide with a high level of similarity to ramoplanin (McCafferty *et al.*, 2002) (Fig. 5.2), is the only inducer of P_{psdA} expression found in this study that is not a lantibiotic. The induction of *psdAB* expression by enduracidin was recently confirmed in an independent microarray study (Rukmana *et al.*, 2009). The significance of this finding remains unclear, but the differential behaviour of P_{psdA} for two pairs of very closely related compounds (mersacidin-actagardine and ramoplanin-enduracidin) strongly suggests that the Psd system responds to a very specific antimicrobial quality of these related compounds that goes beyond their known structural and/or functional features.

5.2.2. Screen with lantibiotic-producing strains

For the initial screen of inducing antibiotics, we used disk diffusion assays with pure compounds or supernatants of the lantibiotic-producing strains (Fig. 5.1C). Subsequently, we wanted to test if it is also possible to use the reporter strain for the direct screening of lantibiotic producers. Therefore, we streaked out P_{psdA} reporter strain TMB299 on LB-X-Gal plates directly next to *B. subtilis* ATCC 6633, which produces subtilin, a lantibiotic that in the disk diffusion assay induced P_{psdA} expression. The appearance of a blue colour

only on the side of TMB299 adjacent to the producer strain shows that it is possible to visualize the induction not only by pure substances but also directly by producer strains (Fig. 5.1D). Because of their specificity and sensitivity, both the P_{psdA} and P_{bceA} reporter strains are promising candidates for the development of a whole-cell-based biosensor for the identification of novel peptide antibiotics from compound libraries and culture supernatants or even directly from antibiotic-producing colonies (Fig. 5.1).

5.2.3. P_{psdA} and P_{bceA} are induced by cell wall antibiotics in a concentration-dependent manner

To further quantify our data, we analyzed the induction of the *psdA* and *bceA* promoters as a function of the concentration of the inducing compound. To this end, we performed quantitative β-galactosidase assays for the inducing antibiotics (Fig. 5.3A and 5.4A). In the concentration-dependent induction experiments, cultures of the P_{psdA} and P_{bceA} reporter strains were grown to the mid-log growth phase, and antibiotics were added to the cultures. After an induction for 30 min, a sample was taken for β-galactosidase assays. These assays not only confirmed all the compounds identified in the disk diffusion assay as being strong inducers (increase in induction ranging from 100-fold for P_{bceA} after induction with actagardine to 800-fold for P_{psdA} after treatment with the same lantibiotic) but also demonstrate a concentration-dependent induction of both promoters (Fig. 5.3A and 5.4A).

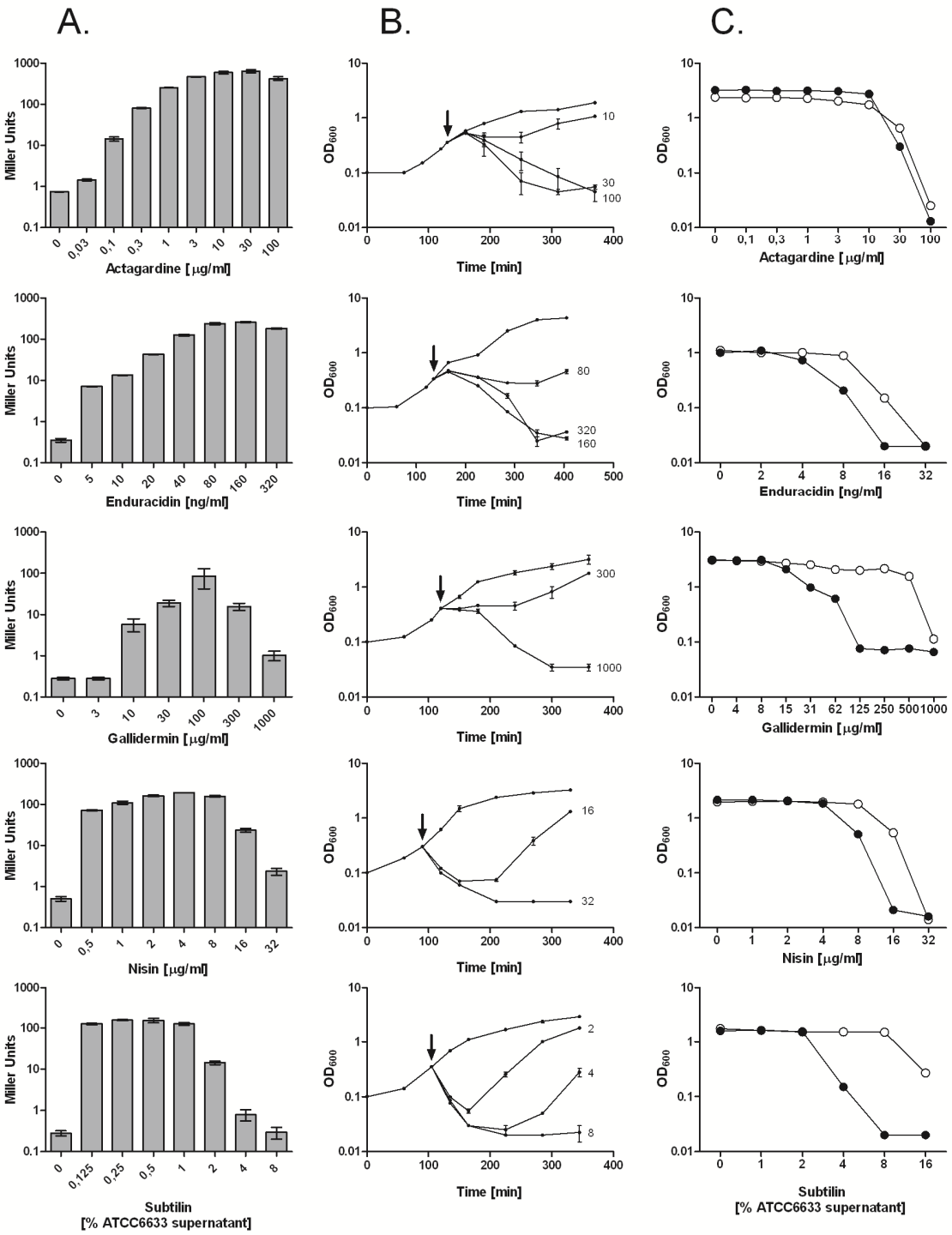

Figure 5.3. Concentration-dependent induction of P_{psdA}, lysis curves, and MICs of *B. subtilis* cultures treated with actagardine, enduracidin, gallidermin, nisin, or subtilin (supernatant of *B. subtilis* strain ATCC 6633). (A) β-Galactosidase activities, expressed as Miller units, of P_{psdA}-*lacZ* reporter strain TMB299 induced with the above-mentioned compounds. A log scale is applied on the *y* axis for reasons of clarity, due to the high dynamic range of β-galactosidase activities. (B) Concentration-dependent killing of TMB299. The times of antibiotic addition are indicated by arrows. The concentrations of actagardine, gallidermin, nisin (in µg/ml), enduracidin (in ng/ml), and subtilin (percent *B. subtilis* ATCC 6633 supernatant) that affect the growth of *B. subtilis* are indicated. (C) MIC assays for *B. subtilis* cultures treated with actagardine, enduracidin, gallidermin, nisin, and subtilin. Wild-type (○) and *psdAB* deletion mutant (•) strains were inoculated to an OD_{600} of 0.05 in Mueller-Hinton medium with different concentrations of antibiotics. Cultures were incubated with agitation at 37°C for 6 h before the determination of the cell density (OD_{600}). The MIC was defined as the lowest concentration of antibiotic that fully inhibited growth.

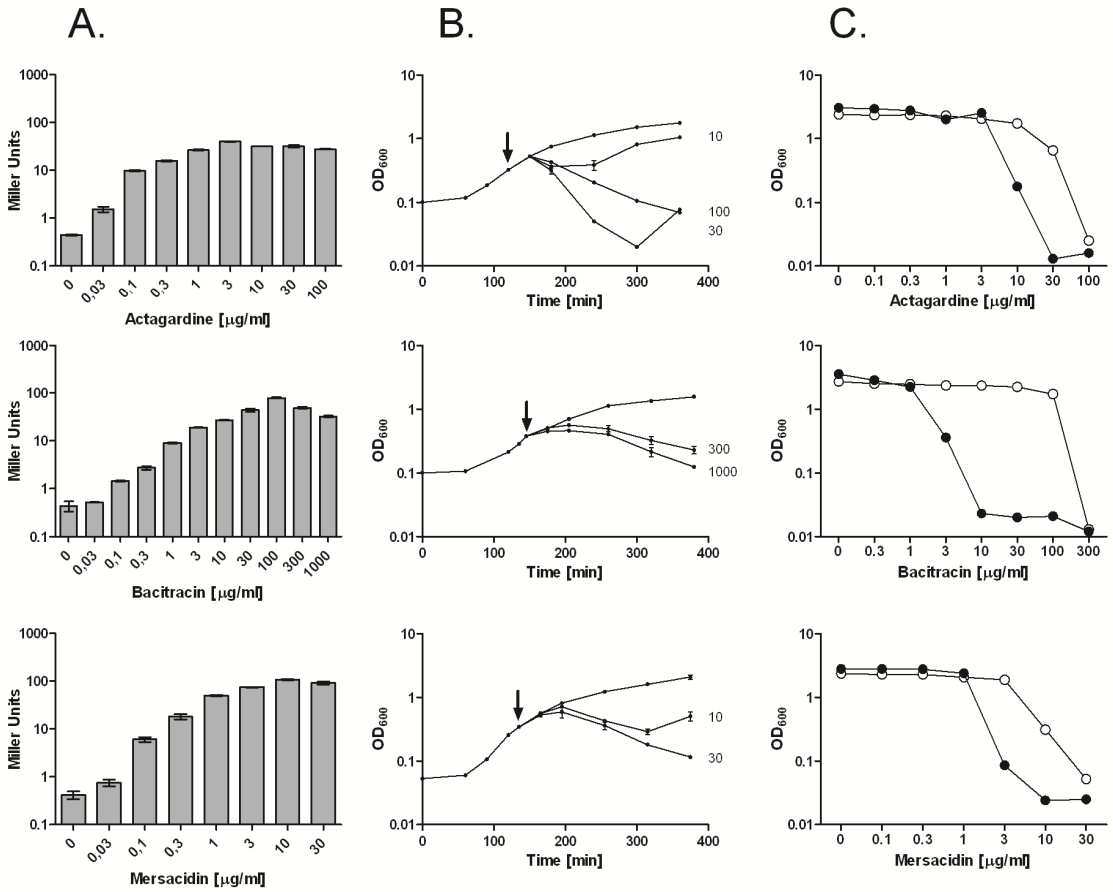

Figure 5.4. Concentration-dependent induction of P_{bceA}, lysis curves, and MICs of *B. subtilis* cultures treated with actagardine, bacitracin, and mersacidin. (A) β-Galactosidase activities, expressed as Miller units, of reporter strain TMB279 induced by actagardine, bacitracin, or mersacidin. A log scale is applied on the y axis for reasons of clarity, due to the high dynamic range of β-galactosidase activities. (B) Concentration-dependent killing of *B. subtilis*. The times of antibiotic addition are indicated by arrows. The concentrations of actagardine, bacitracin, or mersacidin (all in µg/ml) that affect the growth of *B. subtilis* are indicated. (C) MIC assay for the wild type (○) and isogenic *bceAB* deletion mutant (•). See the legend of Fig. 3C for experimental details.

5.2.3.1. PSD3. A strong induction of the P_{psdA} promoter was already observed for 0.5 µg/ml nisin, and it reached its maximum at 4 µg/ml. A similar picture was obtained for subtilin, where 0.125% of the *B. subtilis* ATCC 6633 supernatant strongly induced the P_{psdA} promoter, with the highest induction being observed after the addition of 0.5% supernatant. Induction by gallidermin was visible at a concentration of 10 µg/ml, reaching its maximum at 100 µg/ml. Enduracidin induced P_{psdA} expression at a much lower concentration (5 ng/ml), with the highest induction observed at 160 ng/ml (Fig. 5.3A). Actagardine was the strongest inducer of P_{psdA} expression, with 800-fold induction observed with 10 µg/ml.

5.2.3.2. PSD1. The *bceA* promoter was induced by actagardine at 0.03 µg/ml, with the highest induction observed with 3 µg/ml (Fig. 5.4A). However, the induction was weaker than that observed for the *psdA* promoter, reaching only ~100-fold. Bacitracin and mersacidin induced P_{bceA} to comparable levels, with the induction reaching its maximum at 100 µg/ml and 10 µg/ml, respectively (Fig. 5.4A).

At higher concentrations, a strong decrease in the β-galactosidase activity was observed for all antibiotics tested, indicating cellular damage. Therefore, we also measured the turbidity of the remaining culture of the reporter strain used for the β-galactosidase assay for at least 4 h postinduction to monitor the concentration-dependent effects of the antibiotics on cell growth. A rapid lysis was observed for the concentrations of antibiotics that led to a decrease in the β-galactosidase activity (Fig. 5.3B and 5.4B). These results demonstrate that cellular lysis interfered with the synthesis of β-galactosidase, as was observed previously for other antibiotic reporter strains (Mascher *et al.*, 2004).

5.2.4. The ABC transporters BceAB and PsdAB confer resistance to compounds inducing their expression

As shown above, the *psdA* and *bceA* promoters are strongly induced by peptide antibiotics. The genes under the control of these promoters, *psdAB* and *bceAB*, respectively, encode two subunits of ABC transporters, ATP-binding protein and permease. For the PSD1 module it was previously shown that the ABC transporter BceAB mediates resistance to bacitracin (Bernard *et al.*, 2007; Rietkötter *et al.*, 2008). This prompted the question of whether the same is true for the other inducing compounds and if the paralogous ABC transporter PsdAB also confers resistance to its inducers. Therefore, we determined the corresponding MIC values with Mueller-Hinton medium for the wild type and the isogenic transporter deletion mutants. As indicated by the lysis curves (Fig. 5.3B and 5.4B), the cultures often lyse rapidly after the addition of peptide antibiotics but resume growth after a couple of hours, presumably due to turnover and degradation of the compounds. For MIC determinations, we therefore measured the OD_{600} after 6 h of incubation (see Materials and Methods). This time was sufficient for all antibiotics to develop their full inhibitory effect while simultaneously being short enough to prevent that growth already resumed (data not shown).

The MIC for the *psdAB* mutant was reduced ~2-fold for nisin and subtilin and ~8-fold for gallidermin compared to the wild type (Fig. 5.3C). These results demonstrate that the

PsdAB transporter indeed confers resistance to all of the P_{psdA}-inducing compounds, presumably by acting as an ATP-driven peptide antibiotic-specific resistance pump.

Similar effects were observed for the *bceAB* mutant (Fig. 5.4C). The BceAB transporter confers a high level of bacitracin resistance (the MIC for the *bceAB* mutant is reduced ~30 fold), as was reported previously (Mascher *et al.*, 2003; Ohki *et al.*, 2003a). It also confers a more moderate level of resistance to actagardine and mersacidin (~2- to 4-fold). These results demonstrate that the BceAB transporter, like PsdAB, mediates resistance to a broader spectrum of peptide antibiotics. However, in both cases, the efficiency of removal varied significantly between the different compounds irrespective of the strength of induction (Fig. 5.3 and 5.4).

5.2.5. Identification of the minimal PsdR-dependent promoter region for the *psdAB* operon

In contrast to the BceRS-AB system, the regulation of the paralogous PsdRS-AB module is poorly understood. The *psdRSAB* operon encodes a RR, a HK, and the two subunits of an ABC transporter (ATP-binding protein and a permease), respectively (Fig. 5.1A). Two σ^A-dependent promoters can be identified in this locus, one upstream of *psdR* (RR) and a second weak σ^A-dependent promoter upstream of *psdA* (ATP-binding protein). A potential transcriptional terminator downstream of *psdS* can be predicted albeit with a low $\Delta G°$ value (–6 kcal mol^{-1}) (Joseph *et al.*, 2002). The expression of the *psdAB* genes under inducing conditions was previously verified by Northern blotting (Mascher *et al.*, 2003), and the presence of a longer transcript, *psdRSAB*, was also detected (Joseph *et al.*, 2002; Rasmussen *et al.*, 2009). These results demonstrated a constitutive basal level of expression of the whole *psdRSAB* operon, with a much higher level of expression of the *psdAB* genes under inducing conditions (Fig. 5.1A).

The data presented so far, together with the knowledge gained from the BceRS-AB system, suggest that the RR PsdR is activated by its cognate HK PsdS in the presence of lantibiotics and binds to its operator sequence in the *psdA* promoter region, resulting in a strong induction of *psdAB* expression and, therefore, lantibiotic resistance (Fig. 5.1B). To verify this hypothesis, we analyzed the regulatory elements upstream of the inducible *psdAB* operon in more detail.

A –10 consensus sequence for σ^A can be predicted upstream of the *psdA* gene. No clear –35 sequence can be found at the appropriate position from the –10 sequence (Fig. 5.5A). Such a situation is often found for promoters regulated by transcriptional activators. RR

that act as transcriptional activators usually bind DNA via short binding sites (inverted or direct repeats) a few nucleotides upstream of the –35 promoter element (van Hijum *et al.*, 2009).

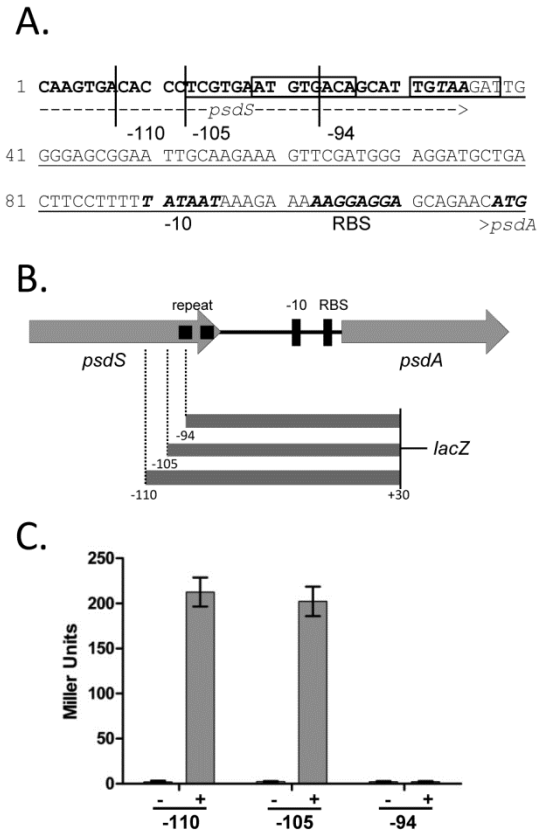

Figure 5.5. Functional analysis of the *psdA* promoter. (A) Intergenic sequence between *psdS* and *psdA*. All features are marked underneath the respective lines of the sequence. The end of *psdS* and the start codon of *psdA* are indicated below the sequence. The –10 P_{psdA} promoter fragment and the putative ribosome-binding site are denoted by –10 and RBS, respectively. The inverted repeat sequence of the putative PsdR-binding site is boxed. The 5' ends of the fragments used for the promoter deletion analysis are labelled according to their positions relative to the *psdA* start codon. The minimal promoter fragment for nisin-dependent induction is underlined. (B) Graphical representation of the intergenic region and the fragments used for the promoter deletion analysis. The features of the region are represented by black boxes and labelled as described above. (C) β-Galactosidase assay for the promoter deletion analysis. Black bars indicate the uninduced control sample for each strain, and gray bars represent the sample induced with nisin.

To localize the binding site of the RR, we used a *lacZ*-based promoter deletion approach. Progressively shorter fragments of P_{psdA} (all ending at position +30 relative to the ATG start codon of *psdA*) (Fig. 5.5B) were used to generate pAC6-based transcriptional *lacZ* reporter fusions integrated at the *amyE* locus. Strains were grown until mid-log phase and treated with nisin (2 µg/ml). All constructs had very low *lacZ* expression levels in the uninduced state (Fig. 5.5C). In cells containing reporter fusions that included at least 105 bp of the upstream *psdA* promoter region, β-galactosidase activity was strongly induced by the addition of nisin to the medium, while a fragment extending to position –94 showed no induction (Fig. 5.5C).

Next, we demonstrated the PsdR dependence of P_{psdA} induction by lantibiotics in a *psdR* deletion mutant by repeating the β-galactosidase assays with strains carrying the P_{psdA} reporter fusion in the presence of an intact PsdRS-AB system (TMB299) and in the *psdR* deletion background (TMB652): nisin-dependent induction was completely abolished in the *psdR* deletion mutant (133 ± 11 and 0.19 ± 0.05 Miller units for the wild type and the *psdR* mutant, respectively). Therefore, the induction of the *psdAB* genes in the presence of lantibiotics is completely PsdR dependent.

Based on the results from the promoter deletion experiments, we identified an 8-nucleotide imperfect (two mismatches) inverted repeat with a 4-nucleotide spacing region (ATGTGACAgcatTGTAAGAT) at positions –99 to –70 (Fig. 5.5A) as a possible site for DNA binding by PsdR. This binding sequence bears similarity to the operator upstream of P_{bceA} in the paralogous BceRS-AB module and has also been predicted by a comprehensive bioinformatic analysis of RR-specific binding sites in low-GC Gram-positive bacteria (de Been *et al.*, 2008). Moreover, the specific binding of PsdR to this DNA region was demonstrated previously by DNase footprint experiments (Giyanto *et al.*, 2003). Taken together, these results demonstrate that PsdR binds to the inverted repeat upstream of the P_{psdA} promoter and activates the expression of the *psdAB* operon in the presence of suitable inducers.

5.3. Conclusions and outlook

The aim of the present study was to identify inducers of the three PSD modules, BceRS-AB (PSD1), YxdJK-LM (PSD2), and PsdRS-AB (PSD3), of *B. subtilis*. By combining an *in silico* meta-analysis of available microarray data sets with disk diffusion assays *in vivo*, we screened a wide range of antimicrobial compounds, including a number of peptide antibiotics. Using a promoter-*lacZ* fusion strain, we identified lipid II-binding lantibiotics as the main group of inducers of P_{psdA} expression. We also identified two closely related lantibiotics, actagardine and mersacidin, as novel inducers of *bceAB* expression, which was previously thought to be a bacitracin-specific resistance pump (Mascher *et al.*, 2003; Ohki *et al.*, 2003a). In contrast, we were not able to identify novel inducers of P_{yxdL}.

The induction of *psdAB* expression is completely dependent on the RR PsdR. We characterized the minimal P_{psdA} promoter and identified an inverted repeat that is necessary for PsdR-dependent P_{psdA} induction and presumably represents the PsdR-binding site. Moreover, we demonstrated that the *psdAB* genes, which encode an ABC transporter, confer resistance to this group of lantibiotics. Therefore, both the PsdRS-AB and the BceRS-AB systems constitute stand-alone PSD modules.

5.3.1. Inducer specificity

The most puzzling result, and an important open question that remains to be answered, concerns the nature of the true stimuli sensed by both PSD modules. Both systems have a more specific inducer range of lipid II-interfering peptide antibiotics than does the P_{liaI}-

based biosensor (Table 5.1). Strikingly, on the one hand, they can discriminate between highly similar compounds, such as the mersacidin-actagardine or enduracidin-ramoplanin pair (PSD1 and PSD3) (Fig. 5.2), while on the other hand, they are also able to respond to compounds even as different as bacitracin and plectasin (PSD1). Nevertheless, all compounds belong to the group of peptide antibiotics that seem to share a cellular target: they all require the pyrophosphate moiety of the lipid carrier of cell wall biosynthesis, undecaprenyl pyrophosphate, as a docking interface to exhibit their antimicrobial activity (Breukink and de Kruijff, 2006; Schneider *et al.*, 2010). While most of the compounds identified in our screen as being inducers of PSD1 and PSD3 activity bind to lipid II, bacitracin directly binds undecaprenyl pyrophosphate, the subsequent intermediate of the lipid II cycle (Stone and Strominger, 1971).

The second surprising result is the apparent lack of a correlation between the strength of induction and the rate of resistance conferred by the induced ABC transporters. Actagardine is the strongest inducer of the PsdRS-AB system. However, the corresponding ABC transporter does not confer any detectable resistance. On the other hand, PsdAB confers significant resistance against gallidermin, despite a 10-fold-lower induction level (Fig. 5.3). Similarly, mersacidin is as potent an inducer of *bceAB* expression as bacitracin. However, the degree of resistance is almost an order of magnitude lower for mersacidin than for bacitracin (Fig. 5.4).

5.3.2. PSD modules as novel biosensors

Bacterial reporter strains based on antibiotic-inducible promoters are an efficient tool for detecting novel bioactive compounds. The well-defined regulatory responses of *B. subtilis* to different types of (antibiotic) stresses and the ease of genetic manipulations make this bacterium a preferred model organism for studying the mode of action of antibiotics by transcriptomics, proteomics, and whole-cell-based biosensors (Bandow *et al.*, 2003; Bandow and Hecker, 2007; Hutter *et al.*, 2004a; Hutter *et al.*, 2004b; Urban *et al.*, 2007; Wecke *et al.*, 2009).

Two comprehensive studies identified sets of *B. subtilis* promoters responding to antibiotics interfering with major biosynthetic pathways (i.e., the biosynthesis of DNA, RNA, proteins, the cell wall, and fatty acids) or specific classes of antibiotics (Hutter *et al.*, 2004a; Urban *et al.*, 2007). Each of these promoters responds to a wide range of compounds that affect the respective pathways. But some of the identified promoters either have a relatively high basal expression level or are only moderately induced (3- to 10-fold).

Therefore, the noise-to-signal ratio (and, hence, robustness) of these biosensors is not always ideal. A biosensor based on the P_{liaI} promoter is both more specific and much more robust (Mascher *et al.*, 2004). It has a very low level of background activity and is induced 50- to 500-fold in the presence of compounds interfering with the lipid II cycle of cell wall biosynthesis, such as bacitracin, ramoplanin, and vancomycin (Table 5.1). This biosensor strain has recently been adapted for high-throughput screens in microtiter plate bioassays (Burkard and Stein, 2008), again demonstrating the potential of whole-cell biosensors for large-scale screens of novel antimicrobial compounds.

Whole-cell biosensors suitable for high-throughput screening need to be compound or pathway specific, robust, and sensitive (Fischer and Freiberg, 2007). Based on the data presented in this study, the P_{psdA} and P_{bceA} reporter strains fulfil these criteria. They are based on an established organism, have very low levels of intrinsic activities and are strongly (more-than-100-fold) induced by a small set of peptide (l)antibiotics that bind lipid II (Fig. 5.2). We have demonstrated that they can be applied to analyze purified compounds, culture supernatants, or the producing strains directly (Fig. 5.1, 5.3, and 5.4 and Table 5.1). Despite our current lack of an exact definition of the nature of the stimuli sensed by the PSD modules, our data indicate that the combined use of the reporter strains derived from PSD1 and PSD3 represents a useful addition to the pool of *B. subtilis* biosensors currently available. They allow the identification of different but related subsets of peptide antibiotics that bind the pyrophosphate moiety of the lipid carrier of cell wall biosynthesis. Such biosensors will be very beneficial for the screening of strain collections and compound libraries, given the great potential of peptide antibiotics as an addition/alternative to the established antibiotics currently in clinical use.

5.4. Materials and methods

5.4.1. Bacterial strains and growth conditions

B. subtilis and *E. coli* cells were routinely grown in Luria-Bertani (LB) medium at 37°C with agitation. For the induction of antibiotic production, *B. subtilis* ATCC 6633 (a subtilin-producing strain) and *B. subtilis* W168 (a sublancin-producing strain) cells were grown in medium A (Banerjee and Hansen, 1988). All strains used in this study are listed in Table 1. Ampicillin (100 µg/ml) was used for the selection of plasmid pAC6 and its derivatives in *E. coli*. Kanamycin (10 µg/ml) and chloramphenicol (5 µg/ml) were used for the selection of the *B. subtilis* strains used in this study.

Table 5.2. Strains, plasmids and oligonucleotides used in this study.

Strain, plasmid, or nucleotide	Characteristic or description[a]	Source or reference
Strains		
E. coli DH5α	F' *endA1 hsdR17*(r_K^- m_K^+) *glnV44 thi-1 recA1 gyrA* (Nal^r) *relA1* Δ(*lacIZYA-argF*)*U169 deoR*[ϕ80 *dlac*Δ(*lacZ*)M15]	Laboratory stock
B. subtilis		
W168	Wild type; *trpC2*	Laboratory stock
ATCC6633	Wild type, subtilin producer	Laboratory stock
TMB035	W168 *bceAB*::kan	(Rietkötter *et al.*, 2008)
TMB279	W168 *amyE*::[*cat* $P_{bceA(-122-82)}$-*lacZ*]	(Rietkötter *et al.*, 2008)
TMB294	W168 *psdAB*::kan	(Rietkötter *et al.*, 2008)
TMB299	W168 *amyE*::[*cat* $P_{psdA(-110-30)}$-*lacZ*]	(Rietkötter *et al.*, 2008)
TMB413	W168 *amyE*::[*cat* $P_{psdA(-105-30)}$-*lacZ*]	This work
TMB414	W168 *amyE*::[*cat* $P_{psdA(-94-30)}$-*lacZ*]	This work
TMB588	W168 *amyE*::[*cat* $P_{yxdL(-194-57)}$-*lacZ*]	This work
TMB652	W168 *amyE*::[*cat* $P_{psdA(-110-30)}$-*lacZ*] *psdR*::kan	This work
Plasmids		
pAC6	*lacZ* fusion vector; integrates at *amyE*; chloramphenicol resistance	(Stülke *et al.*, 1997)
pDF602	pAC6 $P_{psdA(-105-30)}$-*lacZ*	This work
pDF603	pAC6 $P_{psdA(-94-30)}$-*lacZ*	This work
pER605	pAC6 $P_{psdA(-110-30)}$-*lacZ*	(Rietkötter *et al.*, 2008)
pPH601	pAC6 $P_{yxdL(-194-57)}$-*lacZ*	This work
Oligonucleotides[b]		
PpsdA+30	AGTCGGATCCCGATAGGTTCGTTGTTTGCAACACG	
PpsdA-105	AGTCGAATTCGTGAATGTGACAGCATTGTAAG	
PpsdA-94	AGTCGAATTCACAGCATTGTAAGATTGGG	
PyxdL+57	GCATGGATCCAGGACACTTGTCCTTTATAGG	
PyxdL-194	GCATGAATTCTCTCCCGGTGAAGGGACATC	
psdR-up-fwd	CAAAAGAAGAGCTATGGCG	
psdR-up-rev	**CCTATCACCTCAAATGGTTCGCTG**CAAGCAAAATCCGATACACG	
psdR-do-fwd	**CGAGCGCCTACGAGGAATTTGTATCG**CGGAAGGATGAAGCGGAATG	
psdR-do-rev	GAAAACACGATGGTCATCAC	
Kan-fwd	CAGCGAACCATTTGAGGTGATAGG	
Kan-rev	CGATACAAATTCCTCGTAGGCGCTCGG	
Kan-check-fwd	CATCCGCAACTGTCCATACTCTG	
Kan-check-rev	CTGCCTCCTCATCCTCTTCATCC	

[a] The positions of the cloned fragments are given relative to the "A" of the start codon of the corresponding gene.

[b] Sequences are given in the 5'→3' direction. Restriction sites are underlined. Sequences highlighted in boldface type are inverse and complementary to the 5' (up-reverse) and 3' (do-forward) ends of the kanamycin cassette, respectively.

5.4.2. Construction of transcriptional promoter-*lacZ* fusions

All strains, plasmids, and oligonucleotides used in this study are listed in Table 5.2. Ectopic integrations of P_{psdA}-*lacZ* and P_{yxdL}-*lacZ* fusions were constructed based on vector pAC6 (Stülke *et al.*, 1997). Promoter

fragments of increasing lengths were generated by PCR. Standard cloning techniques were applied (Sambrook and Russell, 2001). The inserts were verified by DNA sequencing. The resulting pAC6-derived plasmids (Table 5.2) were linearized with ScaI and used to transform *B. subtilis* with chloramphenicol selection.

5.4.3. Promoter induction assays

Screening for the induction of P_{psdA}, P_{bceA}, and P_{yxdL} was done by disk diffusion assays essentially as described previously (Cao *et al.*, 2002). Briefly, the assays were carried out using soft-agar overlays of the reporter strains on LB plates containing 40 µg/ml X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). Filter paper disks carrying 5 µl of stock solution (antibiotics normally at a concentration of 100 mg/ml; Pep5 and mersacidin at 1 mg/ml, actagardine at 50 mg/ml, duramycin at 10 mg/ml, and subtilin and sublancin as 5 µl of culture supernatants of *B. subtilis* ATCC 6633 and *B. subtilis* W168, respectively, grown overnight) were placed on top of the agar. The plates were incubated at 37°C. After incubation for 24 h, the plates were scored for the appearance of blue rings at or near the edges of the zones of growth inhibition produced by the diffusion of the antibiotics from the filter disks.
For quantitative measurements of β-galactosidase activity, cells were grown in LB medium at 37°C with agitation until they reached an optical density at 600 nm (OD_{600}) of ~0.45. The culture was split, and an inducing substance (at a sublethal concentration) (Table 5.1) was added to one half, leaving the other half untreated (uninduced control). Both cultures were incubated for 30 min at 37°C. Cell pellets were resuspended in 1 ml of working buffer (20 mM β-mercaptoethanol, 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$ [pH 7.0]) and assayed for β-galactosidase activity as described previously, with normalization to cell density (Miller, 1972).

5.4.4. Determination of growth inhibition and the MIC

For concentration-dependent induction and killing experiments, cells were grown in LB medium to the mid-log growth phase (OD_{600} of ~0.45), and antibiotics were added to the cultures as indicated. An uninduced culture was used as a negative control. The cultures were incubated with agitation at 37°C. A sample was taken after 30 min for β-galactosidase assays (see above), and the turbidity of the remaining culture was measured for at least 4 h to monitor the concentration-dependent effects of the antibiotics on cell growth.
MIC assays were performed in Mueller-Hinton medium. Strains W168 (wild type), TMB035 (*bceAB*::Kan), and TMB294 (*psdAB*::Kan) were inoculated to an OD_{600} of 0.05, and different concentrations of antibiotics were added to the medium. Cultures were incubated with agitation at 37°C for 6 h before the determination of cell density. The MIC was defined as the lowest concentration of antibiotic that fully inhibited growth.

5.4.5. Allelic replacement mutagenesis using LFH-PCR

The long-flanking homology PCR (LFH-PCR) technique is derived from a previously reported procedure (Wach, 1996) and was performed as described previously (Wecke *et al.*, 2006). Briefly, a kanamycin resistance cassette was amplified by PCR using vector pDG780 (Guerout-Fleury *et al.*, 1995) as a template. Two primer pairs were designed to amplify ~1,000-bp DNA fragments flanking the region to be deleted at its 5' and 3' ends. The resulting fragments are here called the "up" and "do" fragments. The 3' end of the up fragment as well as the 5' end of the do fragment extended into the gene to be deleted in a way that all expression signals of genes up- and downstream of the targeted genes remained intact. Extensions of ~25 nucleotides were added to the 5' ends of the up-reverse and the do-forward primers that were complementary (opposite strand and inverted sequence) to the 5' and 3' ends of the amplified resistance cassette. A total of 100 to 150 ng of the up and do fragments and 250 to 300 ng of the kanamycin cassette were used together with the specific up-forward and do-reverse primers at standard concentrations in a second PCR. In this reaction the three fragments were joined by the 25-nucleotide overlapping complementary ends and simultaneously amplified by normal primer annealing. The PCR products were used directly to transform *B. subtilis* W168. Transformants were screened by colony PCR using the up-forward primer with a reverse check primer annealing inside the resistance cassette (Table 5.2). The integrity of the regions flanking the integrated resistance cassettes was verified by sequencing of PCR products of ~1,000 bp amplified from the chromosomal DNA of the resulting mutants.

5.5. Acknowledgments

We thank Susanne Gebhard for critical reading of the manuscript. Moreover, we are grateful to H.-G. Sahl for samples of Pep5 and mersacidin and Antony Appleyard (Novacta Biosystems) for the generous gift of purified actagardine.

Work in our laboratory was supported by grants from the Deutsche Forschungsgemeinschaft (MA2837/1-3) and the Fonds der Chemischen Industrie.

CHAPTER 6

Architecture, regulation and specificity determinants of *bceA*-like promoters in *B. subtilis*

Chapter 6
Architecture, regulation and specificity determinants of *bceA*-like promoters in *B. subtilis*

Two-component signal transducing systems are employed by bacteria to sense and react to changes in the environment. They are comprised of a HK that senses a specific stimulus, and a cognate RR that mediates the cellular response, often by binding to certain promoters and influencing gene expression. Bacteria usually harbor numerous 2CS that belong to only few protein families and often show a high sequence similarity. Therefore, preventing unwanted cross-talk is crucial for achieving an efficient signal transmission and requires the maintenance of specific interactions between RR and the cognate promoter to ensure a proper regulatory insulation of orthologous 2CS.

Two closely related 2CS of the EnvZ/OmpR family, BceRS and PsdRS, are encoded in the genome of *B. subtilis*. Together with their cognate ABC transporters, they form detoxification modules that sense and confer resistance to peptide antibiotics (Chapter 5). The corresponding DNA-binding (output) domains of the RR and the target promoters of these two systems show a high level of sequence similarity, which raises the question of how these systems maintain the specificity and insulation. Here, we present results on specificity determinants based on promoter mutagenesis combined with *in vivo* studies of chimeric promoters. The results reveal an unusual architecture and distribution of RR binding sites in the target promoter regions and their unequal contribution to specificity. This mechanism seems to be conserved in homologous detoxification modules of other Firmicutes bacteria.

6.1. Introduction

Bacteria need to react fast and precise even to subtle changes in the surrounding environment. To cope with this constantly present challenge, microorganisms employ various modes that enable them to efficiently regulate the cellular processes. Differential gene expression is one of the main ways to achieve this task. Expression of bacterial genes, organised in a network of interdependent operons, can be switched off or on depending on the current conditions the cell is facing.

Transcription initiation of all genes is controlled by the promoters. These specific sequences in the DNA are recognised by the bacterial RNA polymerase, which binds to certain promoter elements and initiates transcription (reviewed in (Browning and Busby, 2004)). The building blocks of a promoter include the -10 hexamer, -35 hexamer, the extended -10 element, and the UP element (Minchin and Busby, 2009). Every promoter includes some, but not necessarily all of these elements. The two hexamers (with consensus sequences TTGACA and TATAAT for the housekeeping σ factors, (Browning and Busby, 2004)) are located approximately 35 bp and 10 bp upstream from the start point of transcription, respectively, and interact with the regions 4 and 2 of the σ^{70} type subunit of the RNA polymerase (Campbell *et al.*, 2002). The extended -10 element is 3-4 bp motif recognised by region 3 of the σ subunit and is located directly upstream of the -10 hexamer (Murakami and Yamaguchi, 2003). The UP element provides a point of contact for the α subunit of the RNA polymerase, and is composed of approximately 20 bp located upstream of the -35 hexamer (Ross *et al.*, 2001). These four elements contribute in different degrees to the promoter strength, depending on the degree of conservation of each element. They determine the strength of interaction between the RNA polymerase and DNA and provide a static regulation that is not modified in response to changing environmental conditions. The temporal regulation of transcription is achieved with a range of molecular mechanisms, namely interactions with various σ factors, small ligands like ppGpp, folding of the bacterial chromosome, and transcription factors (Browning and Busby, 2004).

Transcription factors are regulatory proteins that bind certain targets on the DNA and thereby influence transcription, either by repressing or activating the gene expression. Activators improve the affinity of the promoter for the RNA polymerase, using one of the three basic mechanisms (Barnard *et al.*, 2004; Browning and Busby, 2004): Transcription activating proteins can bind to a target sequence upstream of the -35 hexamer and directly interact with the smaller carboxy-terminal domain αCTD of the RNA polymerase (class I activation). They can also bind to a target sequence overlapping the -35 hexamer and interact with region 4 of the RNA polymerase σ subunit (class II activation). A DNA conformational change that enables the interaction of the RNA polymerase with -10 and -35 hexamers, which are otherwise not optimally spaced or oriented, can also lead to transcription activation. Repressors on the other hand stop transcription initiation either by binding between or on core promoter elements (steric hindrance) or by blocking of transcription elongation by binding at the start of the coding region (roadblock mechanism). A DNA looping caused by transcription factors binding down- and upstream

from the promoter region can also inhibit transcription initiation in the looped region. These are the basic mechanisms of transcriptional regulation, in more complex cases more than one transcription factor takes part in the activation or repression of transcription initiation (reviewed in detail in (Browning and Busby, 2004)).

Activity of transcription factors is often controlled by ligand binding. A wide variety of ligands is employed by the cell for this purpose. For example the Lac repressor can bind a molecule of allolactose, which lowers its affinity to DNA and changes the expression level of the *lac* operon (Wilson *et al.*, 2007). Very common regulation mechanism is a covalent modification of the transcription factor. Again, various modifications can be employed by the cell. Phosphorylation is extensively used for activation of RR by their cognate HK (Stock *et al.*, 2000). Activity of transcription factors can also be influenced by other mechanisms, such as sequestration by another protein, or simply by changes in their expression level (van Hijum *et al.*, 2009).

RR are small proteins that, once phosphorylated by a cognate HK, usually act as transcription factors. Depending on the position of their binding sites and the nature of the interaction with the RNA polymerase, RR can act as both activators or repressors. A number of RR families has been described in the literature, with the OmpR and NarX being the most abundant, constituting 29.7% and 16.5% of DNA-binding RR, respectively (Galperin, 2010; Gao *et al.*, 2007).

PsdRS-AB and BceRS-AB detoxifications systems belong to a family of peptide antibiotic resistance determinants wide-spread among Firmicutes bacteria (Chapter 5, (Dintner *et al.*, 2011)). They consist of a two-component system and an ABC transporter, which together form a complex responsible for both sensing and detoxification of peptide antibiotics. The BceRS and PsdRS two-component systems belong to the EnvZ/OmpR family of 2CS (de Been *et al.*, 2008). Both the protein sequences of the 2CS, as well as the DNA sequences of the promoters they regulate are highly similar. In this work, we aimed to dissect the specificity determinants in the *psdA* and *bceA* promoters. We determined the minimal promoter regions, and identified multiple RR binding sites. We also analysed the influence of the individual binding sites on the promoter specificity.

6.2. Results

6.2.1. Identification of the minimal *psdA* and *bceA* promoter elements

Both *bceRS-AB* and *psdRS-AB* loci consist of two operons. The first encodes the RR and HK, the second the two subunits of an ABC transporter (ATP-binding protein and permease) (Joseph *et al.*, 2002). While the operons encoding 2CS genes are expressed constitutively, the expression of the ABC transporter is strongly induced by certain antibacterial compounds, such as bacitracin and nisin. The induction of *bceAB* and *psdAB* expression depends on the RR *bceR* and *psdR*, respectively (Ohki *et al.*, 2003a; Staroń *et al.*, 2011). These two proteins belong to the OmpR family of RR (Galperin, 2010).

An approximate identification of the minimal promoter element was performed previously for both *bceA* and *psdA* promoters. The sequence located between -107 and -89 bp upstream of *bceA* start codon (that is -81 and -63 bp upstream of the *bceA* transcription initiation site) is essential for the induction of *bceAB* expression (Ohki *et al.*, 2003a). For the *psdA* promoter this sequence was found to be located between -105 and -94 bp upstream of the *psdA* start codon (Staroń *et al.*, 2011). These two fragments encode a repeat that includes the sequence ACA-N_4-TGT, predicted to be a motif recognised by BceR-like RR (de Been *et al.*, 2008). This is an interesting repeat as it can be seen as both an imperfect inverted and direct repeat, TGTGACG-N_4-TGTCACA in the *bceA* promoter, TGTGACA-N_4-TGTAAGA in the *psdA* promoter. For the *bceA* promoter, a binding of the RR to the region which includes this repeat was demonstrated with DNase I footprinting assay (Ohki *et al.*, 2003a).

A recent analysis of the binding sites upstream of genes encoding BceA-like ATPases demonstrated that in many *bceA*-like promoters a third or even fourth element resembling the repeat can be identified up- and/or downstream of the most highly conserved binding site (Dintner *et al.*, 2011). This is also the case for *psdA* and *bceA* promoters (Fig. 6.1A). In this study, the term "repeat" will be used to denote the full element recognised by the RR (i.e. two 7 bp-long fragments and the intervening spacer), whereas the term "half site" will be used for the 7 bp long halves of the full repeat, excluding the spacer. Both *bceA* and *psdA* promoters encode a half site located downstream of the main binding site (with a 13 bp spacing in P_{bceA} and 14 bp spacer in P_{psdA}), and another half site located upstream (with 4 bp and 10 bp spacers, respectively). From here on, we will refer to them as "downstream" or "upstream half sites".

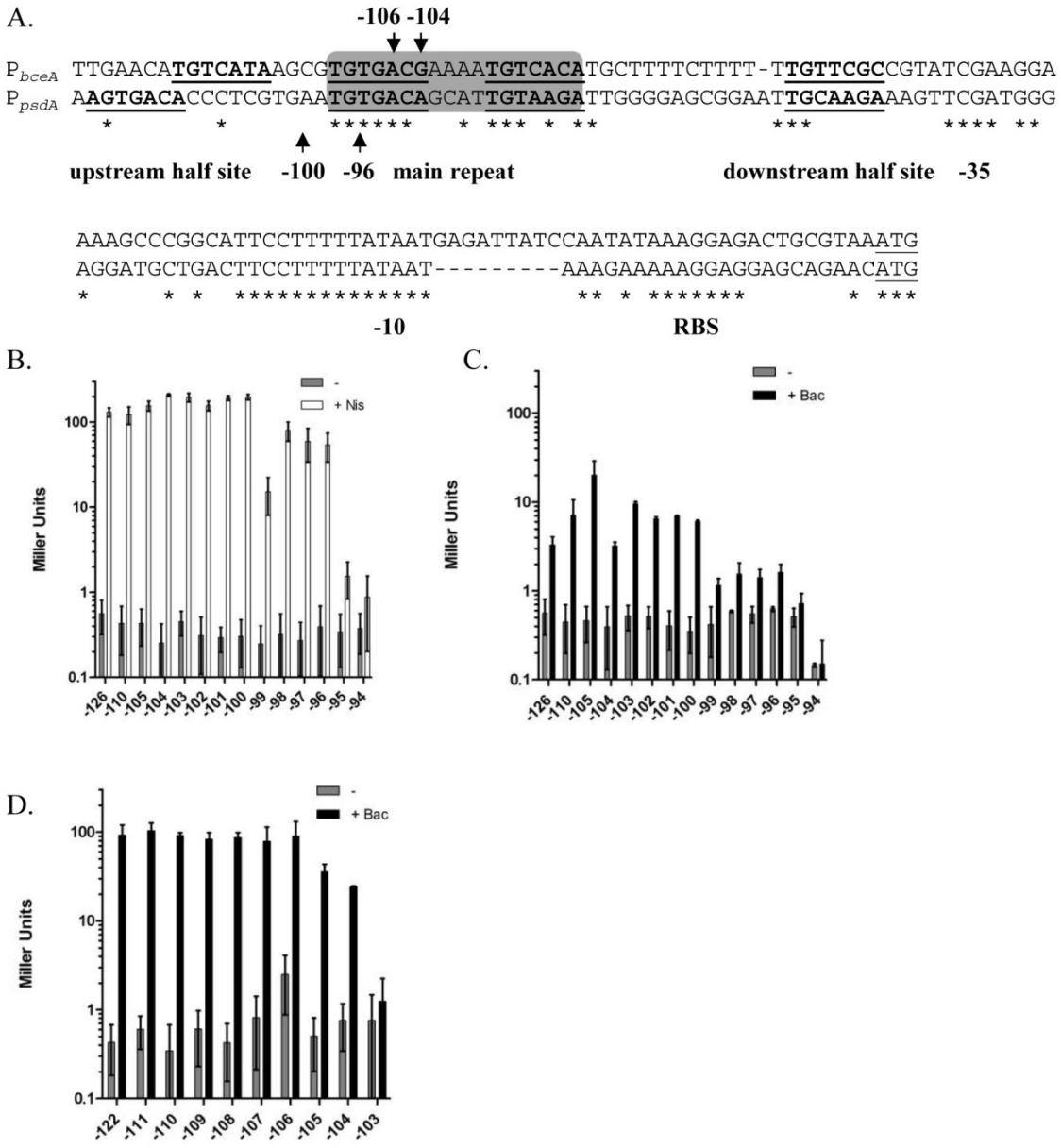

Figure 6.1. . Functional analysis of the *bceA* and *psdA* promoters. (A) DNA sequence of *psdA* and *bceA* promoters. The main RR recognition site is highlighted in grey, the additional half sites are underlined. Chosen postitions in the sequence (relative to the ATG) are marked with arrows. The -35, -10 promoter fragment and the putative ribosome-binding site are denoted by -35, −10 and RBS, respectively. (B-D) β-galactosidase assays for promoter deletion analysis of the *psdA* (B and C) and *bceA* (D) promoters. Grey bars indicate the uninduced control for each strain, white bars represent the samples induced with nisin, and black samples induced with bacitracin.

To precisely localise the minimal promoter elements and investigate the role of the additional half sites, we used a *lacZ*-based promoter deletion approach. Progressively shorter fragments of *bceA* and *psdA* promoters were used to generate transcriptional *lacZ* reporter fusions integrated at the *amyE* locus (based on vector pAC6, see Material and methods section). Strains were grown until mid-log phase and treated with nisin (for P_{psdA} based fusions; 2 μg/ml) or bacitracin (for both P_{bceA} and P_{psdA} fusions; 30 μg/ml). All

promoter fusions show very low *lacZ* expression levels in the uninduced state (Fig. 6.1B-D). In cells harbouring reporter fusions that included at least 100 bp of the *psdA* promoter region, β-galactosidase activity was strongly induced after the addition of both bacitracin and nisin to the medium. The fragments containing fragments with length varying from 99 bp to 96 bp upstream of ATG showed a weaker, but still significant β-galactosidase activity, whereas those containing promoter fragments shorter than 96 bp show no induction (Fig. 6.1.B). The results are the same for induction with nisin and bacitracin, whereby the induction with the latter compound is much weaker (Fig. 6.1.C), consistently with previous observations (Rietkötter *et al.*, 2008). Similar results were obtained for fragments containing the *bceA* promoter. Fragments harbouring reporter fusions that included at least 106 bp of the *bceA* promoter region showed full induction, those containing 105 or 104 bp weaker, but still significant β-galactosidase activity, whereas fragments smaller than 104 bp showed no induction at all (Fig. 6.1.D).

Based on the results from the promoter deletion experiment, we confirmed that the 8-nucleotide repeat in the *bceA* promoter (TGTGACGaaaaTGTCACA) is the possible site for binding by BceR, whereby the deletion of the first two nucleotides of this repeat does not abolish the induction completely, but only weakens it. The minimal fragment necessary for the full induction of the *psdA* promoter is 2 bp longer (aaTGTGACAgcatTGTAAGAtt). However, the length of the minimal fragment both necessary and sufficient for induction corresponds perfectly to the length of *bceA* promoter, and consists of TGACAgcatTGTAAGA.

6.2.2. Mutagenesis of the main binding site

Following the identification of the minimal promoter fragments, we analysed the main binding sites for BceR and PsdR. These repeats consist of two 7 bp-long halves separated by an AT-rich central 4 bp spacer. They differ in only four (out of fourteen) nucleotides within the binding site itself, and in three (out of four) nucleotides within the spacer. As the binding sites are the sequences specifically recognised by RR, it can be assumed they are the specificity determinants in this type of promoters. It can therefore be speculated that either the differences within the binding site, the spacer, or a combination of both are responsible for the specificity in the *bceA* and *psdA* promoters.

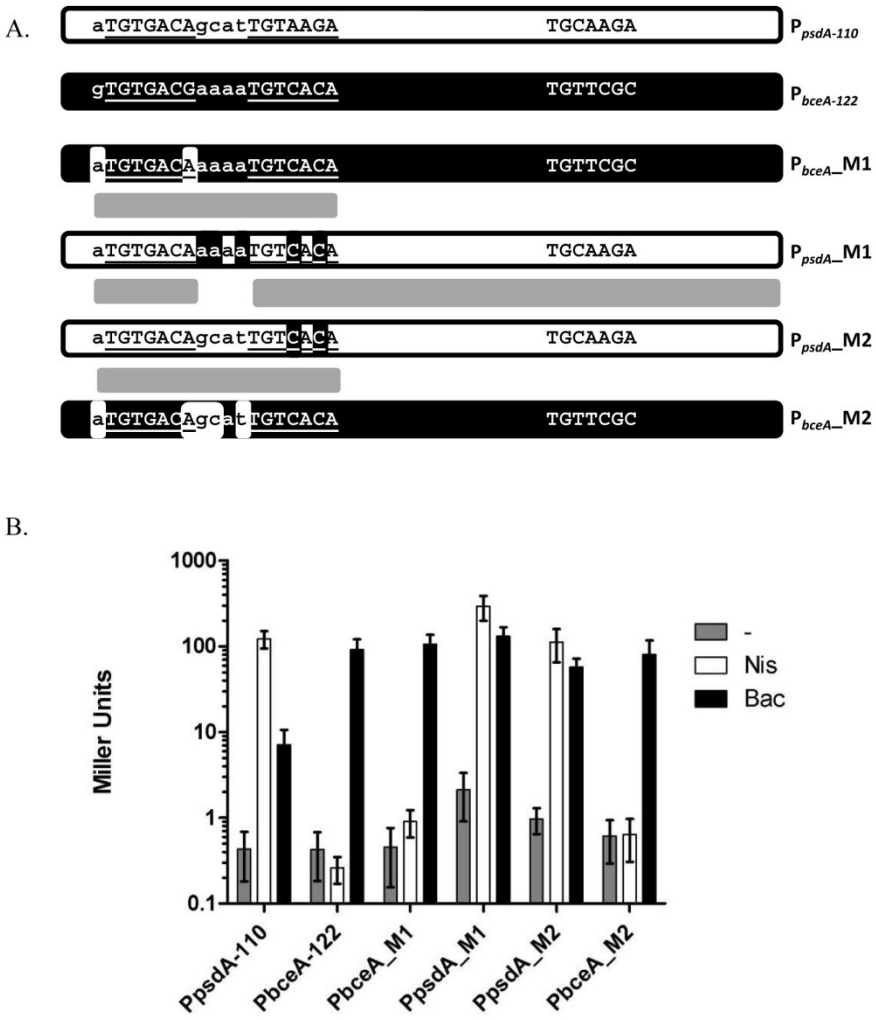

Figure 6.2. . Mutagenesis of the *bceA* and *psdA* promoters. (A) The fragments originating from the *psdA* promoter are shown in white, those origination from the *bceA* promoter in black. Grey bars denote fragments with identical sequences. The main RR recognition site is shown in capital letters. (B) β-galactosidase assays of the strains shown above. Grey bars indicate the uninduced control for each strain, white bars represent the samples induced with nisin, and black samples induced with bacitracin.

To test these assumptions, we performed an extensive mutagenesis of both binding sites. Schematic representation of tested promoter mutants is shown in Fig. 6.2., with mutations introduced within the binding site and/or the spacer. Interestingly, changes introduced both in the binding site and the spacer did not change the specificity of the tested promoters (compare strains $P_{bceA\text{-}122}$, P_{bceA}_M1 and P_{bceA}_M2 for P_{bceA}, and $P_{psdA\text{-}110}$, P_{psdA}_M1, P_{psdA}_M2 for P_{psdA}). The exchange of the spacer influenced the strength of induction (introduction of spacer from P_{bceA} into P_{psdA}, P_{psdA}_M2 vs P_{psdA}_M1), but not the specificity (Fig. 6.2). These results suggest that the specificity of the promoter-RR interaction does not lie within the repeat itself. It has to be kept in mind that the cross-talk between the Bce and Psd systems is observed in the wild-type, hence the weak induction of the *psdA* promoter by bacitracin. In this assay it is impossible to distinguish between the cross-talk

and weak change of specificity. Therefore, it cannot be ruled out that the obviously stronger (compared to the wild-type) bacitracin induction in strains P_{psdA}_M2 and P_{psdA}_M1 comes from a weak shift in specificity. This question has to be addressed by using other experimental methods. However, no significant change of specificity can be observed in these strains.

Indeed, consistent with the observation that the specificity of the RR-promoter interaction lies outside the main recognition sequence, strains with identical binding sites but different downstream regions show completely different induction pattern (compare strains P_{psdA}_M2 and P_{bceA}_M2, Fig. 6.2). Strain P_{psdA}_M2 harbouring a modified PsdR binding site and the region between the binding site and translation start site originating from P_{psdA} is reacting to nisin and bacitracin as the wild type *psdA* promoter does (compare with $P_{psdA\text{-}110}$). Strain P_{bceA}_M2, which harbours a binding site identical to the one in strain P_{psdA}_M2, but a downstream region originating from *bceA* promoter reacts only to bacitracin, in a manner identical to the wild-type *bceA* promoter (compare with $P_{bceA\text{-}122}$). These observations strongly suggest that specificity determinants lie downstream of the main binding site.

6.2.3. Secondary binding sites as specificity determinants

As shown in Fig. 6.1A, both *bceA* and *psdA* promoters harbour two other potential RR binding sites next to the main binding site, defined in the promoter deletion analysis as a repeat consisting of two 7 bp-long halves separated by a 4 bp spacer. One is located upstream of the main repeat (4 bp and 10 bp for *bceA* and *psdA* promoters, respectively), the second one is located 13 bp downstream of the main repeat in P_{bceA} and 14 bp in P_{psdA}. Both these sites are just halves of a repeat, similar to the halves of the main RR binding site.

Promoter deletion analysis demonstrated that the upstream half site is not necessary for induction, as its deletion did not change the observed β-galactosidase activity. In order to investigate the role of the downstream half site, we constructed a series of chimeric promoters, fused them to the promoterless *lacZ* gene, and measured the β-galactosidase activity after induction with nisin or bacitracin.

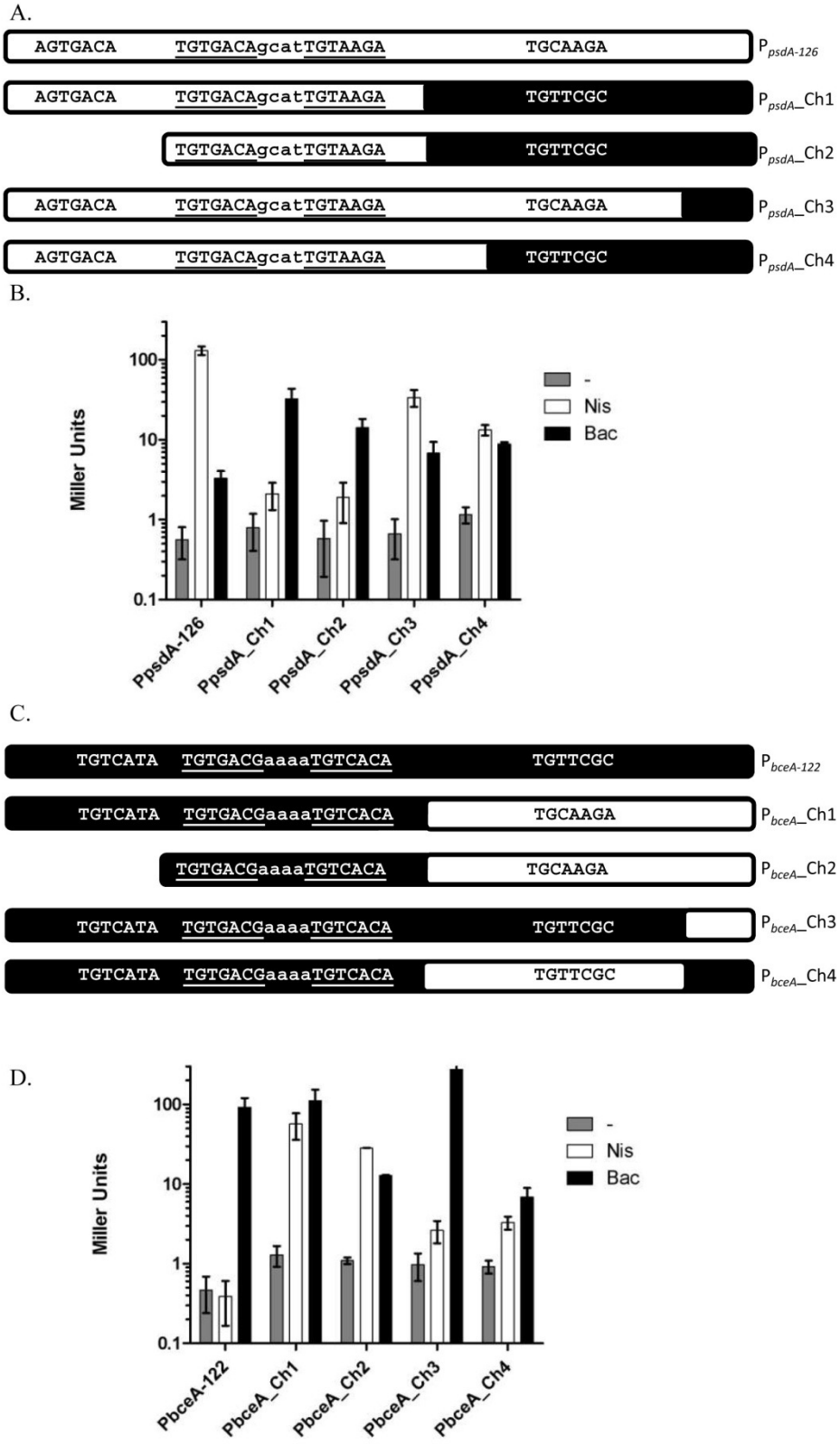

Figure 6.3. Chimeric promoters based on P_{bceA} and P_{psdA}. (A) Chimeric promoters based on P_{psdA}. The fragments originating from the *psdA* promoter are shown in white, those origination from the *bceA* promoter in black. The RR recognition sites are shown in capital letters. (B) β-galactosidase assays of the strains shown above. Grey bars indicate the uninduced control for each strain, white bars represent the samples induced with nisin, and black samples induced with bacitracin. (C) Chimeric promoters based on P_{bceA}. Colour-code as in point (A). (D) β-galactosidase assays of the strains shown above. Colour-code as in point (C).

The wild-type *psdA* promoter shows strong activity after induction with nisin, and much weaker activity after bacitracin treatment. The exchange of the whole region downstream of the main repeat and upstream of the ATG (black bar in Fig. 6.3A, strain P_{psdA}_Ch1) leads to a complete change of specificity, and the promoter is not any more induced by nisin, but by bacitracin (compare strains P_{psdA}_Ch1 and $P_{psdA\text{-}126}$ on Fig. 6.3). The strength of bacitracin induction is approximately four-fold lower than in the case of nisin induction in the wild-type (32.5 MU for bacitracin, 131 MU for nisin). Strain P_{psdA}_Ch1 harbours both the main binding site and the upstream half site originating from the *psdA* promoter, and the downstream half site from the *bceA* promoter. To check the role of the upstream half site in this promoter context, we investigated the response of a similar promoter, but lacking the upstream half site (Fig. 6.3. strain P_{psdA}_Ch2). This strain also responds solely to bacitracin, however, the response is two-fold weaker than in the strain P_{psdA}_Ch1 (14 MU in strain P_{psdA}_Ch2, 32.5 MU in strain P_{psdA}_Ch1). It has to be kept in mind that a certain level of cross-talk in the wild-type promoters exist and the *psdA* promoter responds weakly to bacitracin (approximately 10 fold, up to 5 MU). In this assay it is impossible to distinguish between the effects of cross-talk and the change of specificity. In *psdA* promoter after bacitracin challenge probably additive effects are observed.

The above results show that the specificity determinants in the *psdA* promoter lie in the DNA fragment between the main binding site (but not including it) and the translation start site. This is the fragment where the downstream half site is located. To narrow down the fragment responsible for specificity determination we investigated the induction pattern of another chimeric promoter, in Fig. 6.3 termed P_{psdA}_Ch3. This strain is identical to strain P_{psdA}_Ch1, the only difference being that the fragment between the main binding site and the downstream half site, as well as the downstream half site itself originate from *psdA*, not from *bceA* promoter. As expected, this promoter reacts similarly to the wild-type, albeit with a lower strength. Therefore, the region between the main binding site and the downstream half site, including the latter, harbours a part or the whole specificity determinant recognised by PsdR and BceR. However, an exchange of this fragment only (Fig. 6.3, strain P_{psdA}_Ch4) leads to a promoter that is only slightly inducible by nisin and by bacitracin. Comparison with strain P_{psdA}_Ch1 (Fig. 6.3) shows that the presence of the *psdA* promoter fragment between the downstream half site and the ATG on one hand increases the nisin response (from 2 MU to 13 MU), on the other hand decreases the response to bacitracin (32.5 MU to 9 MU). It seems that the complete fragment shifting the specificity to PsdR is present in the strain P_{psdA}_Ch1, but not in the strains P_{psdA}_Ch3 and

P_{psdA}_Ch4. It could mean that the last two strains harbour only a part of the specificity determining sequence. Interestingly, the putative -35 element is located exactly 4 bp downstream of the downstream half site and bears a certain resemblance to the motif recognised by the BceR-like RR (TCGATGG in the *psdA* promoter, TCGAAGG in the *bceA* promoter). Whether this element plays any role in the promoter specificity has to be investigated experimentally.

Similar, although not identical results were obtained for the *bceA* promoter. Here the results are easier to interpret, as no cross-talk can be observed in the wild type. The *bceA* promoter is induced approximately 250 fold after bacitracin challenge (Fig. 6.3, strain $P_{bceA\text{-}122}$). The exchange of the region downstream of the main repeat and upstream of the translation start site (denoted with a white bar in Fig. 6.3, strain P_{bceA}_Ch1) does not, as it is the case for the *psdA* promoter, lead to a full exchange of specificity. This chimeric promoter is strongly induced by both nisin and bacitracin (56 MU for nisin, 111 MU for bacitracin). A deletion of the upstream half site leads to a dramatic drop in promoter strength (Fig. 6.3, strain P_{bceA}_Ch2, 28 MU after exposure to nisin, 12 MU after bacitracin induction), but the promoter is still induced by both antibiotics. Analogously to the experiments performed for the *psdA* promoter, we also determined the induction pattern of another chimeric promoter, where the fragment originating from P_{bceA} extends to the downstream half site, and the remaining part originates from the *psdA* promoter (strain P_{bceA}_Ch3). As was the case for the chimeric promoter in strain P_{psdA}_Ch3, this promoter also behaves like the wild-type promoter it originates from ($P_{psdA\text{-}126}$), and is induced by bacitracin only (albeit much stronger, up to 274 MU). An exchange of the fragment between the main binding site and the downstream half site (including the repeat) leads to a promoter that is only weakly inducible (strain P_{bceA}_Ch4, 3 fold by nisin, 7 fold by bacitracin).

6.3. Discussion

In this study we investigated the specificity determinants in the *psdA* and *bceA* promoters. These promoters regulate the expression of the genes encoding the ABC transporters PsdAB and BceAB, respectively. The two promoters share similarities in their architecture: in addition to a typical RR binding site, both harbour two half sites similar to the main repeat, one located both upstream and downstream of the main binding site. Using promoter deletion analysis and chimeric promoters fused to a promoterless *lacZ* gene, we

monitored the expression of various combinations of promoter elements, both without induction and after antibiotic challenge.

Using a nucleotide-by-nucleotide approach, we determined the minimal fragments necessary for the induction of both promoters. In both cases the minimal fragment encompasses the core of the main repeat. In the *bceA* promoter it is the sequence TGACGaaaaTGTCACA, in *psdA* promoter TGACAgcatTGTAAGA. It seems that the core binding site TGACRrmawTGTMASA (where the repeats are capitalised, r represents A or G, m represents A or C, w represents A or T, and s represents G or C) plays a similar role in both promoters. However, these shorter versions of both promoters show a slightly lower induction than the full length variants (approximately 3-fold reduction for both promoters). The full length of the promoter necessary for the full induction differs between P_{bceA} and P_{psdA}. The former promoter achieves full induction with the binding site TGTGACGaaaaTGTCACA, the latter needs a two nucleotide longer repeat (aaTGTGACAgcatTGTAAGAtt).

The results clearly show that the main repeat is necessary for induction, the core being the TGACRrmawTGT motif (for abbreviations, see above). The presence of the upstream half site does not influence the antibiotic-dependent promoter induction at all. This is interesting and also surprising, as the presence of such upstream and downstream half sites seems to be a characteristic feature of this family of promoters (Dintner *et al.*, 2011). However, their role can be observed under changed conditions (see below). Such half sites were shown to have a function in other promoters, where it is speculated that they constitute a drafting mechanism to attract transcription factor molecules to the promoter (ArgR in *L. lactis*, (Larsen *et al.*, 2005)). It is possible that under conditions of full induction their role is insignificant or not detectable.

Whether the main repeats can be classified as inverted or direct repeats cannot be answered at this point, as the repeats in both promoters fulfil the criteria of both an imperfect inverted and direct repeat. This family of RR (OmpR/PhoB) is known to bind mainly to direct repeats. However, there are a few examples of RR from this family that bind inverted repeats, such as PmrA from *S. enterica* and RegX3 from *M. smegmatis* (Glover *et al.*, 2007; Wösten and Groisman, 1999).

6.4. Materials and methods

6.4.1. Bacterial strains and growth conditions

B. subtilis and *E. coli* strains were routinely grown in Luria Bertani (LB) medium at 37°C with agitation. All strains used in this study are listed in Table 6.1. Ampicilin (100 µg/ml) was used for the selection of plasmid pAC6 and its derivatives in *E. coli*. Chloramphenicol (5 µg/ml) resistance were used for the selection of the *B. subtilis* strains used in this study. Transformation was carried out as described (Harwood and Cutting, 1990).

6.4.2. Construction of transcriptional promoter-*lacZ* fusions

Strains and plasmids used in this study are listed in Tables 6.1 and 6.2. Ectopic integrations of promoter *lacZ* fusions were constructed based on vector pAC6 (Stülke *et al.*, 1997). Promoter fragments were generated by PCR, standard cloning techniques were applied (Sambrook and Russell, 2001). The inserts were verified by DNA sequencing. The resulting pAC6 derivates were linearised by ScaI and used to transform *B. subtilis* with chloramphenicol selection.

6.4.3. Promoter induction assays

Cells were inoculated from fresh overnight cultures and grown in LB medium at 37°C with agitation until they reached an optical density at 600 nm (OD_{600}) of ~0.45. The culture was split, and an inducing substance (bacitracin at 30 µg/ml or nisin at 2µg/ml) was added to one half, leaving the other half untreated (uninduced control). Both cultures were incubated for 30 min at 37°C with agitation. Cell pellets were resuspended in 1 ml of buffer (20 mM β-mercaptoethanol, 60 mM Na_2HPO_4, 10 mM KCl, 1 mM $MgSO_4$ [pH 7.0] and assayed for β-galactosidase activity as described previously, with normalization to cell density (Miller, 1972).

Table 6.1. Bacterial strains used in this study.

Strain	Genotype or characteristic	Reference or source
E. coli		
DH5α	*recA1 endA1 gyrA96 thi hsdR17rK- mK +relA1 supE44 Φ80 DlacZDM15 D(lacZYA-argF)U169*	Laboratory stock
B. subtilis		
TMB279	W168 *amyE*::pER603 ($P_{bceA-122}$)	(Rietkötter *et al.*, 2008)
TMB299	W168 *amyE*::pER605 ($P_{psdA-110}$)	(Rietkötter *et al.*, 2008)
TMB412	W168 *amyE*::pDF601 ($P_{psdA-126}$)	This study
TMB413	W168 *amyE*::pDF602 ($P_{psdA-103}$)	(Staroń *et al.*, 2011)
TMB414	W168 *amyE*::pDF603 ($P_{psdA-94}$)	(Staroń *et al.*, 2011)
TMB605	W168 *amyE*::pJS603 (P_{psdA}_M2)	This study
TMB607	W168 *amyE*::pJS605 ($P_{bceA-111}$)	This study
TMB610	W168 *amyE*::pJS607 (P_{bceA}_M1)	This study
TMB755	W168 *amyE*::pJS615 (P_{psdA}_Ch1)	This study
TMB756	W168 *amyE*::pJS616 (P_{psdA}_Ch3)	This study
TMB758	W168 *amyE*::pJS618 (P_{bceA}_Ch1)	This study
TMB759	W168 *amyE*::pJS619 (P_{bceA}_Ch3)	This study
TMB761	W168 *amyE*::pJS621 (P_{psdA}_M1)	This study
TMB762	W168 *amyE*::pJS622 (P_{bceA}_M2)	This study
TMB805	W168 *amyE*::pAS601 ($P_{psdA-99}$)	This study
TMB806	W168 *amyE*::pAS602 ($P_{psdA-97}$)	This study
TMB960	W168 *amyE*::pAS603 ($P_{psdA-104}$)	This study
TMB961	W168 *amyE*::pAS604 ($P_{psdA-103}$)	This study
TMB962	W168 *ameE*::pAS605 ($P_{psdA-102}$)	This study
TMB963	W168 *ameE*::pAS606 ($P_{psdA-101}$)	This study
TMB964	W168 *amyE*::pAS607 ($P_{psdA-100}$)	This study
TMB965	W168 *amyE*::pAS608 ($P_{psdA-98}$)	This study
TMB966	W168 *amyE*::pAS609 ($P_{psdA-96}$)	This study
TMB967	W168 *amyE*::pAS610 ($P_{psdA-95}$)	This study
TMB1047	W168 *amyE*::pAS613 ($P_{bceA-110}$)	This study
TMB1048	W168 *amyE*::pAS614 ($P_{bceA-109}$)	This study
TMB1049	W168 *amyE*::pAS615 ($P_{bceA-108}$)	This study
TMB1050	W168 *amyE*::pAS616 ($P_{bceA-107}$)	This study
TMB1051	W168 *amyE*::pAS617 ($P_{bceA-106}$)	This study
TMB1052	W168 *amyE*::pAS618 ($P_{bceA-105}$)	This study
TMB1053	W168 *amyE*::pAS619 ($P_{bceA-104}$)	This study
TMB1054	W168 *amyE*::pAS620 ($P_{bceA-103}$)	This study
TMB1055	W168 *amyE*::pAS621 (P_{bceA}_Ch2)	This study
TMB1059	W168 *amyE*::pAS625 (P_{psdA}_Ch2)	This study
TMB1319	W168 *amyE*::pSK601 (P_{psdA}_Ch4)	This study
TMB1320	W168 *amyE*::pSK602 (P_{bceA}_Ch4)	This study

Table 6.2. Vectors and plasmids used in this study.

Plasmid	Genotype or characteristic features[a]	Primers used for cloning	Reference or source
pAC6	Transcriptional *lacZ* fusion vector, integrates at *amyE*, chloramphenicol resistance		(Stülke *et al.*, 1997)
pAS601	pAC6 $P_{psdA(-99-30)}$-*lacZ*	1591/0600	This study
pAS602	pAC6 $P_{psdA(-97-30)}$-*lacZ*	1592/0600	This study
pAS603	pAC6 $P_{psdA(-104-30)}$-*lacZ*	1688/0600	This study
pAS604	pAC6 $P_{psdA(-103-30)}$-*lacZ*	1687/0600	This study
pAS605	pAC6 $P_{psdA(-102-30)}$-*lacZ*	1686/0600	This study
pAS606	pAC6 $P_{psdA(-101-30)}$-*lacZ*	1685/0600	This study
pAS607	pAC6 $P_{psdA(-100-30)}$-*lacZ*	1684/0600	This study
pAS608	pAC6 $P_{psdA(-98-30)}$-*lacZ*	1683/0600	This study
pAS609	pAC6 $P_{psdA(-96-30)}$-*lacZ*	1682/0600	This study
pAS610	pAC6 $P_{psdA(-95-30)}$-*lacZ*	1681/0600	This study
pAS613	pAC6 $P_{bceA(-110-82)}$-*lacZ*	1869/0555	This study
pAS614	pAC6 $P_{bceA(-109-82)}$-*lacZ*	1870/0555	This study
pAS615	pAC6 $P_{bceA(-108-82)}$-*lacZ*	1871/0555	This study
pAS616	pAC6 $P_{bceA(-107-82)}$-*lacZ*	1872/0555	This study
pAS617	pAC6 $P_{bceA(-106-82)}$-*lacZ*	1873/0555	This study
pAS618	pAC6 $P_{bceA(-105-82)}$-*lacZ*	1874/0555	This study
pAS619	pAC6 $P_{bceA(-104-82)}$-*lacZ*	1875/0555	This study
pAS620	pAC6 $P_{bceA(-103-82)}$-*lacZ*	1876/0555	This study
pAS621	pAC6 $P_{bceA(-111:-88)}$-$P_{psdA(-79:30)}$-*lacZ*	1857/0600	This study
pAS625	pAC6 $P_{psdA(-105:-80)}$-$P_{bceA(-87:82)}$-*lacZ*	1863/0555	This study
pDF601	pAC6 $P_{psdA(-126-30)}$-*lacZ*	0674/0600	This study
pDF602	pAC6 $P_{psdA(-105-30)}$-*lacZ*	0675/0600	(Staroń *et al.*, 2011)
pDF603	pAC6 $P_{psdA(-94-30)}$-*lacZ*	0676/0600	(Staroń *et al.*, 2011)
pER603	pAC6 $P_{bceA(-122-82)}$-*lacZ*	0554/0555	(Rietkötter *et al.*, 2008)
pER605	pAC6 $P_{psdA(-110-30)}$-*lacZ*	0599/0600	(Rietkötter *et al.*, 2008)
pJS603	pAC6 $P_{psdA(-105:30)\ A84C\ G82C}$-*lacZ*	1305/0600	This study
pJS605	pAC6 $P_{bceA(-111-82)}$-*lacZ*	1307/0555	This study
pJS607	pAC6 $P_{bceA(-111:82)\ G107A\ G100A}$-*lacZ*	1309/0555	This study
pJS615	pAC6 $P_{psdA(-202:-80)}$-$P_{bceA(-87:82)}$-*lacZ*	1497/1498 and 1499/0555	This study
pJS616	pAC6 $P_{psdA(-202:-56)}$-$P_{bceA(-64:82)}$-*lacZ*	1497/1500 and 1501/0555	This study
pJS618	pAC6 $P_{bceA(-210:-88)}$-$P_{psdA(-79:110)}$-*lacZ*	1504/1510 and 1511/1507	This study
pJS619	pAC6 $P_{bceA(-210:-65)}$-$P_{psdA(-55:110)}$-*lacZ*	1504/1508 and 1509/1507	This study
pJS621	pAC6 $P_{psdA(-105:30)\ G91A\ C90A\ T88A\ A84C\ G82C}$-*lacZ*	1518/0600	This study
pJS622	pAC6 $P_{bceA(-111:82)\ G107A\ G100A\ A99G\ A98C\ A96T}$-*lacZ*	1519/0555	This study
pSK601	pAC6 $P_{psdA(-202:-81)}$-$P_{bceA(-88:-65)}$-$P_{psdA(-55:110)}$	2096/146, 1499/10 and 1507/1497	This study
pSK602	pAC6 $P_{bceA(-210:-89)}$-$P_{psdA(-80:-56)}$-$P_{bceA(-64:82)}$	1506/10, 2098/146 and 1504/555	This study

[a] The positions of the cloned fragments are given relative to the "A" of the start codon of the corresponding gene

Table 6.3. Primers used in this study.

Nr	Sequence[a]
0010	CTTCGCTATTACGCCAGCTGG
0146	GTCTGCTTTCTTCATTAGAATCAATCC
0554	GATC**GAATTC**GAACATGTCATAAGCGTGTGACG
0555	GATC**GGATCC**TATCGATGCCCTTCAGCACTTCC
0599	AGTC**GAATTC**CACCCTCGTGAATGTGACAGC
0600	AGTC**GGATCC**CGATAGGTTCGTTGTTTGCAACACG
0674	AGTC**GAATTC**TCGTGTTTTCAAGTGACACC
0675	AGTC**GAATTC**GTGAATGTGACAGCATTGTAAG
0676	AGTC**GAATTC**ACAGCATTGTAAGATTGGG
1305	AGTC**GAATTC**GTGAATGTGACAGCATTGTCACATTGGGGAGCG
1307	GATC**GAATTC**AAGCGTGTGACGAAAATGTCACAT
1309	GATC**GAATTC**AAGCATGTGACAAAAATGTCACATGCTTTTCTTTTTTGTTCGC
1497	GATC**GAATTC**GCAAAACAGGTGTGCAGCCGTCTCG
1498	ATCTTACAATGCTGTCACATTCAC
1499	ATGTGACAGCATTGTAAGATGCTTTTCTTTTTTGTTCGC
1500	ACTTTCTTGCAATTCCGCTC
1501	GCGGAATTGCAAGAAAGTTCGAAGGAAAAGCCCGGCATTCC
1504	GATC**GAATTC**GGCGCCGTTATTGATCCATATTG
1506	GTGTGACGAAAATGTCACATTGGGGAGCGGAATTGCAAGAAAGTTCG
1507	AGTC**GGATCC**AATTCTCCTTTTGACACACTG
1508	TACGGCGAACAAAAAAGAAAAGCATGTGAC
1509	CTTTTTTGTTCGCCGTATCGATGGGAGGATGCTGACTTCC
1510	TGCCGGGCTTTTCCTTCGATACGG
1511	CGAAGGAAAAGCCCGGCATTCCTTTTTATAATAAAGAAAAAGGAGG
1518	AGTC**GAATTC**GTGAATGTGACAAAAATGTCACATTGGGGAGCG
1519	GATC**GAATTC**AAGCATGTGACAGCATTGTCACATGCTTTTCTTTTTTGTTCGC
1591	AGTC**GAATTC**ATGTGACAGCATTGTAAGATTGG
1592	AGTC**GAATTC**GTGACAGCATTGTAAGATTGG
1681	AGTC**GAATTC**ACGACAGCATTGTAAGATTGG
1682	AGTC**GAATTC**ATGACAGCATTGTAAGATTGG
1683	AGTC**GAATTC**TGTGACAGCATTGTAAGATTGG
1684	AGTC**GAATTC**TAATGTGACAGCATTGTAAG
1685	AGTC**GAATTC**GAATGTGACAGCATTGTAAG
1686	AGTC**GAATTC**TGAATGTGACAGCATTGTAAG
1687	AGTC**GAATTC**AGTGAATGTGACAGCATTGTAAG
1688	AGTC**GAATTC**CGTGAATGTGACAGCATTGTAAG
1857	ATGC**GAATTC**AAGCGTGTGACGAAAATG
1863	ATGC**GAATTC**GTGAATGTGACAGCATTG
1869	ATGC**GAATTC**AGCGTGTGACGAAAATG
1870	ATGC**GAATTC**GCGTGTGACGAAAATGTC
1871	ATGC**GAATTC**ACGTGTGACGAAAATGTC
1872	ATGC**GAATTC**AAGTGTGACGAAAATGTC
1873	ATGC**GAATTC**AAATGTGACGAAAATGTC
1874	ATGC**GAATTC**GTGACGAAAATGTCAC
1875	ATGC**GAATTC**ATGACGAAAATGTCAC
1876	ATGC**GAATTC**AAGACGAAAATGTCAC
2096	GCGAACAAAAAAGAAAAGCATCTTACAATGCTG
2098	CAATTCCGCTCCCCAATGTGACATTTTCG

[a] Restriction sites are highlighted in bold, linker sequences for joining reactions are underlined.

CHAPTER 7

Insulation and specificity determinants in BceRS-like two-component systems in *B. subtilis*

Chapter 7
Insulation and specificity determinants in BceRS-like two component systems in *B. subtilis*

Two-component signal transducing systems consist of a HK that senses a specific environmental stimulus, and a cognate RR that mediates the cellular response. The HK of the cell envelope stress response in *B. subtilis* are characterized by a very short N-terminal input domain, consisting of two membrane-spanning helices but lacking a discernible extracellular domain. Because of this architecture, they were termed intramembrane-sensing histidine kinases (IM-HK). BceS-like HK represent the largest conserved group of IM-HK. The corresponding 2CS are functionally linked to ABC transporters encoded by neighbouring operons. Such units represent antibiotic-specific detoxification modules widely distributed in Firmicutes bacteria.

Two closely related detoxification modules are encoded in the genome of *B. subtilis*, BceRS-AB and PsdRS-AB. These systems require both the HK and the ABC transporter for sensing of antimicrobial peptides, such as bacitracin, nisin and gallidermin. The corresponding modules of the two systems show a high level of sequence similarity and also some degree of regulatory cross-talk. Here, we present preliminary results on the similarities and differences in mechanism of stimulus perception and gene regulation between these two systems, based on studies of chimeric HK and RR.

7.1. Introduction

Fidelity of protein-protein interactions is the basic prerequisite for many cellular processes. Proteins need to interact specifically with their partners in order to avoid accidental interactions, which can be at best unproductive, at worst have deleterious effects. Specificity in protein-protein interactions becomes extremely important when proteins from the same family take part in various, sometimes mutually exclusive cellular processes. Cells employ various mechanisms to ensure the fidelity of interactions (Ubersax and Ferrell, 2007). Spatial separation is found especially in multicellular organism, where proteins can be localised in different cellular compartments, often with help of specific adaptor or scaffold proteins, or are expressed in a tissue-specific manner (Schwartz and Madhani, 2004). Insulation can be also achieved by differential timing of expression.

Molecular recognition is another important means to achieve fidelity of protein-protein interactions, especially in unicellular organisms such as bacteria.

2CS constitute a large family of proteins harbouring a set of very conserved domains, usually with many systems present in a cell at the same time (Stock *et al.*, 2000). They are involved in various processes, including chemotaxis, metabolism of various nutrients, virulence, quorum sensing and cell cycle regulation (Mascher *et al.*, 2006; Paterson *et al.*, 2006; Wadhams and Armitage, 2004; Williams *et al.*, 2007). Typically, 2CS consist of two proteins, a sensory HK and a RR, which is responsible for orchestrating the cellular response to the stimulus (Stock *et al.*, 2000). Most bacteria harbour quite a number of 2CS, with most species containing 20-30 HK-RR pairs (Skerker *et al.*, 2008). However, these numbers can be much higher. The current champion is *Ktedonobacter racemifer*, which contains 189 HK- and 167 RR-encoding genes (MiST2 database as of October 2011, (Ulrich and Zhulin, 2010)). As the domains directly involved in the phosphotransfer process (dimerization and phosphotransfer, in short DHp domain in HK and receiver domain in RR, (Gao and Stock, 2009)) are highly similar, there is a risk of cross talk between distinct 2CS due to a phosphotransfer from a HK to a non-cognate RR.

It is generally assumed that cross-talk between various 2CS would be at best unproductive, if not even deleterious, and accordingly, there are just a few known examples of cross-talk observed *in vivo* (Gao and Stock, 2009; Laub and Goulian, 2007). This raises a question of how signalling fidelity is achieved to ensure the appropriate response to a certain signal? There are different mechanisms of insulation employed by 2CS, including phosphatase activity of HK and excess of RR molecules (reviewed in detail in (Laub and Goulian, 2007)), but the main mechanism which ensures specificity seems to be the molecular recognition between both interacting partners (Capra *et al.*, 2010; Skerker *et al.*, 2005). Although phosphotransfer is catalysed primarily by amino acids in the RR, the specificity of the reaction requires pairing of residues on both HK and the RR (Laub and Goulian, 2007).

Various methods have been employed to identify specificity determinants in both interacting partners in 2CS. Domain swapping gave first insights into specificity determinants at the molecular level, showing that the cytoplasmic part of the kinase is solely responsible for interaction fidelity (Baumgartner *et al.*, 1994; Perraud *et al.*, 1998; Utsumi *et al.*, 1989). Much more information can be extracted from the crystallographic analysis of co-crystal structures of HK-RR complexes. As of October 2011, there are two such complexes available at high resolution, namely of Spo0B-Spo0F complex of *B.*

subtilis (Zapf *et al.*, 2000), and of HK853-RR468 complex of *Thermotoga maritima* (Casino *et al.*, 2009). Spo0B/Spo0F complex is a part of the *B. subtilis* sporulation phosphorelay, where the kinase Spo0B is a phosphotransferase connecting two RR proteins Spo0F and Spo0A in the pathway (Piggot and Hilbert, 2004). This structure has been used as a general model for HK-RR interactions for many years. However, Spo0B is not membrane linked, has no autokinase activity, and does not harbour a complete catalytic domain (Casino *et al.*, 2009). Therefore, it is not ideal for understanding the specificity of HK-RR pairs. A much more representative example of HK-RR complexes is the HK853-RR468 structure from *T. maritima* (Casino *et al.*, 2009), and many phenomena observed in other 2CS can be explained on its basis. However, knowledge of which residues are in direct contact is not always enough to pinpoint those responsible for specificity, as these two groups of residues do not show a perfect correlation and only a subset of interfacial residues determines interaction selectivity (Skerker *et al.*, 2008).

Computational approaches proved to be another excellent tool for identification of residues important for specific recognition between HK and RR partners. Co-variance analysis is based on the assumption that if a residue in one protein changes over evolution, a compensatory change in the partner protein has to follow for the specific interaction to be maintained (Szurmant and Hoch, 2010). This method can be improved by further statistical analysis and incorporation of structure-derived data ((Schug *et al.*, 2009), reviewed in detail in (Szurmant and Hoch, 2010)).

With this wide array of methods available, we set out to identify specificity determinants in Bce-like detoxification modules. These modules are peptide antibiotic resistance determinants wide-spread among Firmicutes bacteria (Dintner *et al.*, 2011). They consist of a 2CS and an ABC transporter, which together form a complex responsible for both sensing and detoxification of the antibiotic. As there is often more than one Bce-like module in one organism (for detailed overview, see (Dintner *et al.*, 2011)), it is important to identify and understand specificity determinants in these systems, which are composed of four instead of just two proteins. In this study, we focused on two systems, PsdRS-AB and BceRS-AB from *B. subtilis*. The Psd system responds to and confers resistance against various lantibiotics, including nisin. The Bce system is one of the main bacitracin resistance determinants in *B. subtilis*, and also responds to a number of lantibiotics ((Mascher *et al.*, 2003; Staroń *et al.*, 2011), Chapter 5). In the two systems, we focused on specificity of interactions (i) in stimulus sensing, (ii) interaction between HK and permeases, and (iii) interaction between HK and RR.

7.2. Results

7.2.1. Role of the extracellular loop in substrate recognition

In a typical 2CS, the HK senses the stimulus via its extracellular domain. However, BceS-like proteins belong to the family of intramembrane sensing (IM-) HK (Mascher, 2006) and therefore do not harbour any extracellular domains. In PsdS, the transmembrane helices are predicted to be connected by only 14 amino acids, in BceS by 9 (Fig. 7.1A). This raises a question of how such kinases sense the stimulus. For a homologous Bce-like system, Aps in *S. aureus* and *S. epidermidis*, it has been reported that the short extracellular loop of the HK is indeed responsible for the substrate recognition, despite its size (Li *et al.*, 2007a; Li *et al.*, 2007b). In particular, the negatively charged residues were implied in the recognition of cationic AMPs by ApsS. Therefore, we decided to investigate if this also holds true for the PsdS kinase from *B. subtilis*. Transmembrane helices of PsdS are connected by only 14 amino acids, which include three negatively charged glutamate residues (Fig. 7.1A). In order to investigate the role of these amino acids, we exchanged them to alanine and performed quantitative β-galactosidase assays after induction with the PsdS-inducing antibiotic, nisin. We also tested the response to bacitracin, another peptide antibiotic which specifically induces a homologous system BceS. We could observe no difference to the wild type neither after induction with nisin, nor with bacitracin (Fig. 7.1B). We subsequently exchanged the extracellular loops between PsdS and BceS. In contrast to the results obtained for the Aps system from *S. aureus* and *S. epidermidis* species, the chimeric PsdS kinase with the extracellular loop from the BceS kinase showed exactly the same behaviour as the wild-type PsdS (Fig. 7.1B). Moreover, the placement of three glutamate residues into the BceS loop of the PsdS-BceS chimera (loop of wild-type BceS harbours only one negatively charged residue, aspartate; see Fig.7.1A) did not alter the behaviour of the kinase. All of the tested chimeric kinases show the same behaviour as the wild type, suggesting that in PsdS neither the extracellular loop, nor its the negatively charged residues represent the determinants responsible for substrate specificity. Moreover, the loop composition also does not seem to influence the strength of induction.

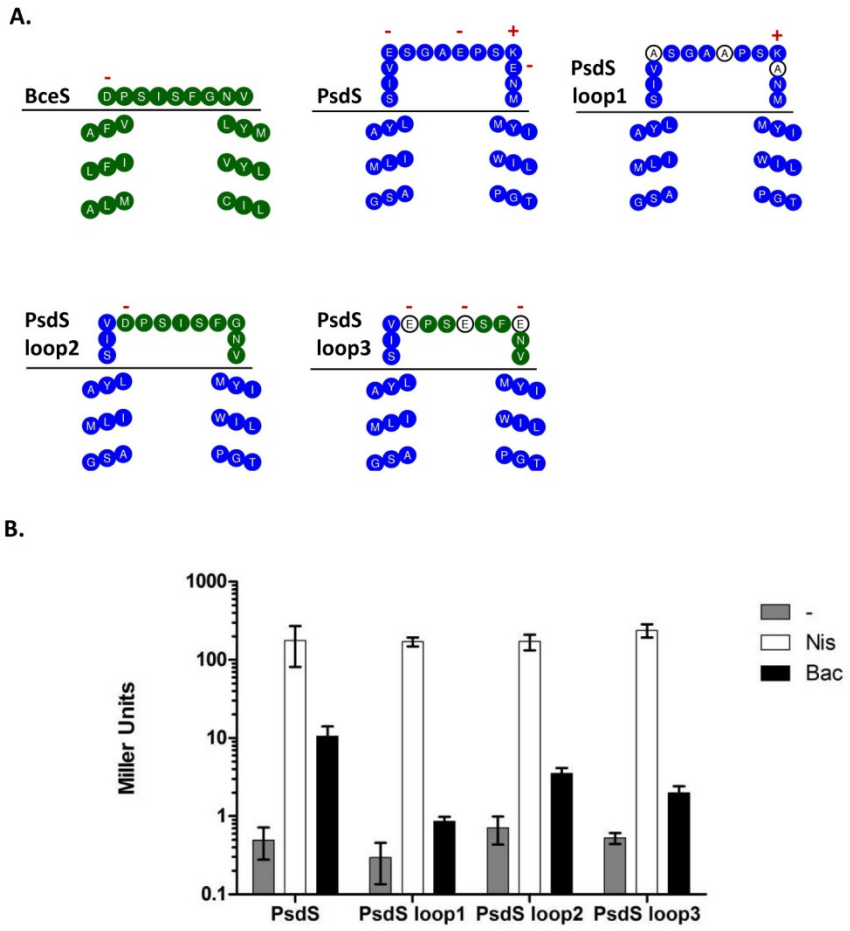

Figure 7.1. Role of the extracellular loop in substrate recognition. (A) The extracellular loop and the flanking regions of the wild type BceS and PsdS proteins, as well as PsdS with three glutamate to alanine exchanges (PsdS_loop1), PsdS with the extracellular loop of BceS (PsdS_loop2), and PsdS with the extracellular loop of BceS with three intoduced glutamates (PsdS_loop3) is shown. Amino acids are depicted as small circles, outside border of the cellular membrane is marked by a horizontal line. Residues originating from BceS are marked in green, from PsdS – in blue. Amino acids not present in any of the wild type kinases are marked in white. (B) β-galactosidase activities, expressed as Miller Units, of strain carrying a P_{psdA}-*lacZ* fusion, insertional mutation in *psdS*, and, under xylose-inducible promoter, different chimeric kinases shown in point (A).

7.2.2. PsdS and BceS HK interact with their cognate ABC transporters

Results shown above clearly demonstrate that the extracytoplasmic loop of BceS-like kinases from *B. subtilis* is not responsible for detecting the stimulus. This remains in contrast with the findings reported for the Aps system in *S. aureus* and *S. epidermidis* (Li *et al.*, 2007a; Li *et al.*, 2007b). However, the data on other Bce-like systems emerging in the last few years revealed that the HK alone is not the sensor, but rather the kinase in complex with the ABC transporter (for review, see (Gebhard and Mascher, 2011)). According to this hypothesis, the ABC transporter changes its conformation upon the contact with the inducing peptide, which in turn promotes the interaction with the HK and its activation. To investigate the potential interaction between the kinase and the permease

in Psd and Bce systems, which is the prerequisite for constitution of the sensory complex, we used a the Bacterial Adenylate Cyclase-based Two Hybrid Assay (BACTH, (Karimova *et al.*, 1998; Karimova *et al.*, 2000)). This system has been previously successfully employed in studies of protein-protein interactions in *B. subtilis*, including interactions of HK with other membrane bound proteins (Szurmant *et al.*, 2007b).

Full length permeases BceB and PsdB, as well as HKs BceS and PsdS were coexpressed as C- and N-terminal fusions to the two subunits (T18 and T25) of the *B. pertussis* adenylate cyclase protein. This assay is based on the fact that the adenylate cyclase is inactive when its domains are expressed individually. However, when the T18 and T25 subunits are fused to interacting proteins, activity of the adenylate cyclase is reconstituted, which in turn results in the expression of β-galactosidase (Karimova *et al.*, 2000). The β-galactosidase activity can be visualised on the X-gal indicator plates.

A representative BACTH assay is shown on Fig. 7.2. Interactions can be observed for both BceS-BceB and PsdS-PsdB pairs, although the strength of interaction, and hence the blue colour of the spotted colonies, seems to differ with different vectors and fusions used. The strongest interaction can be seen when the HK is fused to the N-terminus of the T18 subunit, and the permease to the C-terminus of the T25 subunit. The other combinations of vectors and fusions show varying degree of the strength of interaction. It can be explained by (i) occlusion of interacting fragments by the adenylate cyclase subunits in certain configurations, (ii) toxicity of permease proteins when expressed from high copy vectors (carrying T18 subunits) which would lead to a selective pressure against the expression. False positives (reported to be rare in this assay, (Karimova *et al.*, 2000)) have been eliminated through a set of controls, where each of the fusion constructs was expressed with a complementary vector carrying only the adenylate cyclase subunit (data not shown). While these results have to be confirmed by an independent method, they strongly suggest an interaction between the HK and the permease in Bce and Psd systems in *B. subtilis*.

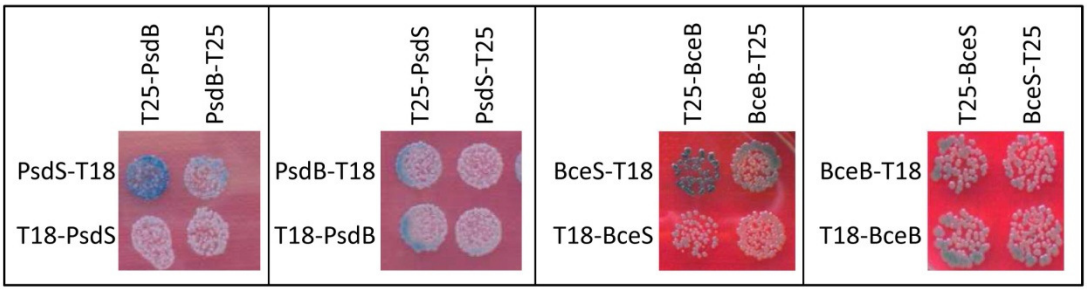

Figure 7.2. Interaction of BceS and PsdS HK with their cognate permeases, BceB and PsdB. The *bceS*, *psdS*, *bceB* and *psdB* genes were cloned into BACTH vectors, generating N- and C-terminal fusions to the T18 or T25 subunits of adenylate cyclase from *B. pertussis*. The bacterial two-hybrid assay was performed as described previously (Karimova *et al.*, 2000). Blue color indicates protein-protein interaction.

7.2.3. Specificity determinants in chimeric HK

All the recently emerging data strongly supports the hypothesis that a direct contact between the HK and the ABC transporter is a necessary step for the signal transduction (Gebhard and Mascher, 2011). Our results of BACTH assay indicate that the above may also hold true for the Bce-like systems in *B. subtilis*. According to this hypothesis, the inducing peptide is bound by the permease, which in turn activates the HK. Therefore, at least in organisms harbouring more than one Bce-like system, the interaction between the permease and the HK has to be specific to prevent unwanted cross-talk. We set out to find these specificity determinants in the HK and asked whether it is possible to switch the interaction specificity and obtain a chimeric HK that reacts to a non-cognate substrate, possibly via an interaction with a non-cognate permease. A series of chimeric kinases was constructed (Fig.7.3), based on bacitracin-sensing BceS and nisin-sensing PsdS kinases from *B. subtilis*. For each chimeric kinase, we tested the induction of the cognate promoter (P_{psdA} or P_{bceA}, the latter only in case of kinase cHK.B) after antibiotic induction (Fig. 7.3.). A very important and surprising finding is that the kinase cHK.A is still functioning similarly to the wild-type PsdS. In this kinase, we substituted the two transmembrane helices and the loop connecting them in PsdS with ones originating from the BceS kinase. When this kinase is expressed in a strain harbouring a P_{psdA}-*lacZ* fusion, β-galactosidase activity can be detected after induction with nisin. The same behaviour can be observed in wild-type PsdS, although the strength of induction in the strain harbouring the chimera is four times weaker than in the wild type. Thus, the signal transfer in PsdS is taking place in the cytoplasmic part of the kinase and is independent of the transmembrane helices and the connecting extracellular loop.

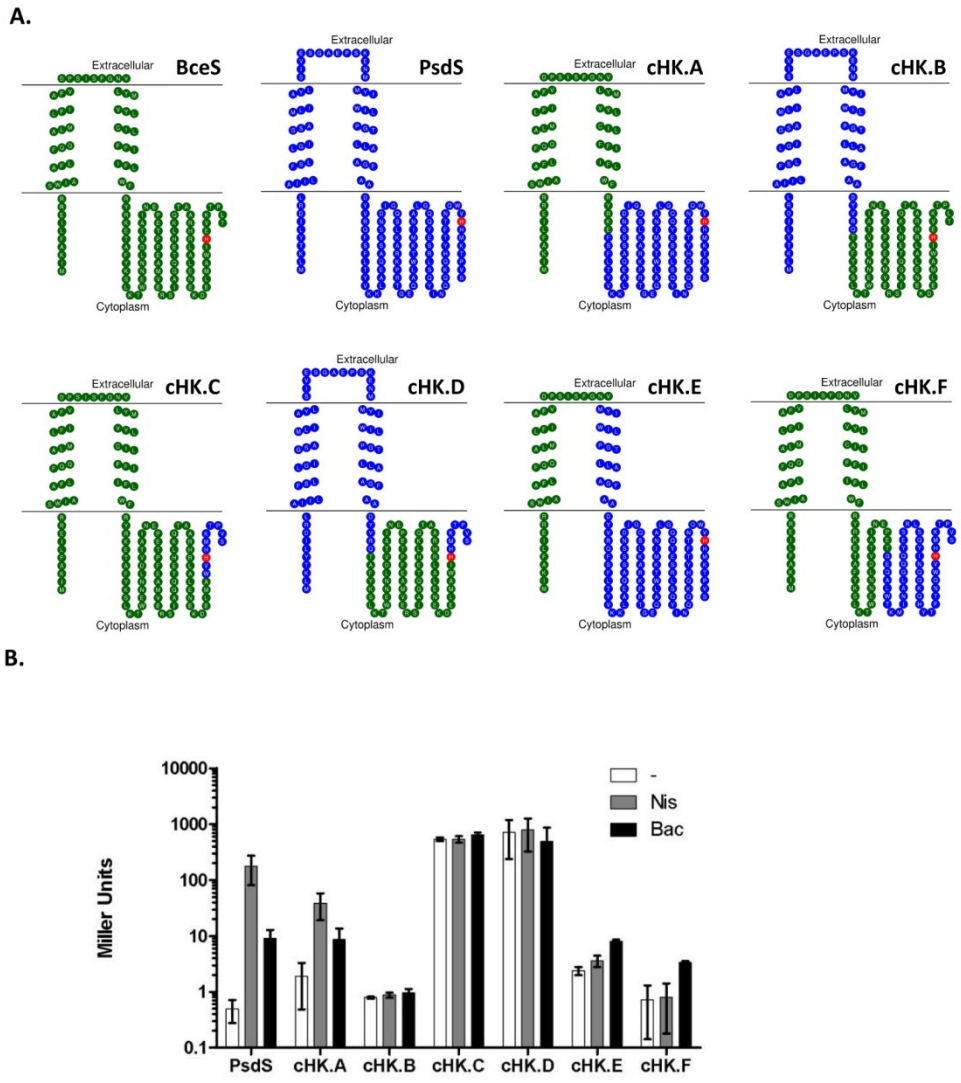

Figure 7.3. Specificity of chimeric kinases derived from BceS and PsdS. (A) Schematic representation of N-terminal part of wild type HK BceS and PsdS until the conserved H-box as well as chimeric kinases containing cHK.A, cHK.B, cHK.D, cHK.E and cHK.F. Amino acids are depicted as small circles, cellular membrane is marked by two horizontal lines. Residues originating from BceS are marked in green, from PsdS – in blue. A conserved histidine residue is marked in red. (B) β-galactosidase activities, expressed as Miller Units, of strain carrying a P_{psdA}-lacZ fusion, insertional mutation in *psdS*, and, under xylose-inducible promoter, different chimeric kinases shown in point (A).

A similar chimeric kinase based on BceS (cHK.B), with the two transmembrane helices and the extracellular loop substituted by ones of PsdS did not show any induction after treatment with nisin or bacitracin. This chimera seems to be not functional, which can be due to multiple reasons. As the kinase was not tagged, it cannot be ruled out that it was not expressed in sufficient amounts, or that it was not properly folded. It can also be argued that the BceS kinase, in contrast to PsdS, needs the transmembrane helices for stimulus sensing. However, in the light of high sequence similarity to PsdS, and many confirmed functional similarities between the two kinases, this does not seem very likely.

Two of the chimeric HK show a stimulus-independent constitutive activity, which results in a 5-6 times stronger promoter activity compared to the fully induced wild-type. The fusion points between the BceS-derived and PsdS-derived fragments in both chimeric HK is located in the first helix of the DHp domain. Although the N-terminus of these two kinases differs, the stimulus-independent behaviour is the same (see Fig. 7.3B, kinases cHK.C and cHK.D), suggesting that the disturbance in the first helix of the DHp domain leads to a constitutive kinase activity.

Two other chimeric kinases did not show any activity, neither without, nor after induction with appropriate antibiotics. Kinase cHK.E is the PsdS based chimera, where the first transmembrane helix and the extracellular loop are BceS-derived. The second kinase, cHK.F, is also PsdS based. Here, both transmembrane helices, as well as a part of the cytoplasmic component of the kinase are BceS-derived. Again, as the kinases are not tagged, it cannot be ruled out that problems with proper folding or expression are the reason for lack of observed activity.

7.2.4. Chimeric RR

In order to further extend the search for specificity determinants in Bce-like 2CS, we searched for residues in BceR and PsdR that are responsible for specificity of interaction with the cognate HK. A prediction of residues in the BceR/PsdR RR and BceS/PsdS HK that strongly covary pointed the interaction to the first helix of the RR (H. Szurmant, unpublished) (Fig. 7.4). All five amino acids predicted to be responsible for specificity were located in the first helix of the RR. Therefore, we exchanged this helix between BceR and PsdR, obtaining two new chimeric RR. However, these chimeric RR did not show any detectable β-galactosidase activity (data not shown).

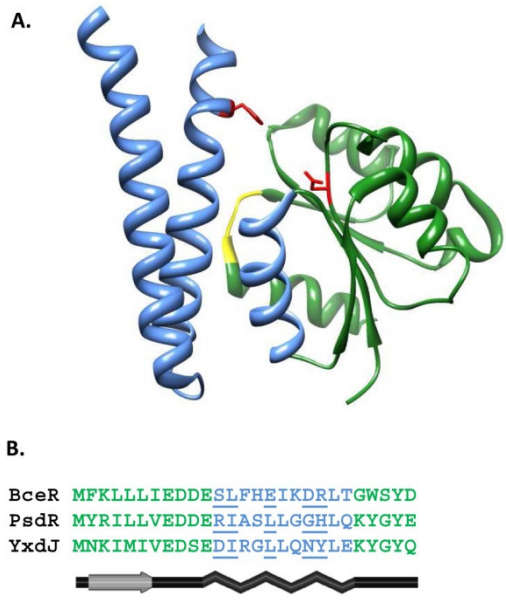

Figure 7.4. Specificity determinants in RR. (A) Exchange of helix 1 in the receiver domain of the RR shown on the co-crystal structure of HK853 and RR468 from *T. maritima* (Casino *et al.*, 2009). Blue ribbons in the receiver domain of the RR represent residues originating from PsdR, green ribbon – residues from BceR. Two helices of the HK responsible for RR recognition are shown in blue. The side chains of the conserved histidine and aspartate residues are shown as red sticks, and the β5-α5 loop is marked in yellow. (B) Alignment of the first 25 amino acids of RR BceR, PsdR and YxdJ. Colour code as in point A. Underlined residues are predicted to be responsible for the interaction with the HK (for details, see text). Schematic representation of the secondary structure is shown under the alignment.

7.2.5. Random mutagenesis of RR

The exchange of the first helix in BceR and PsdR did not lead to the expected switch in specificity. Therefore, we performed a random mutagenesis screen in the search for residues responsible for specificity of interactions between the RR and the HK. Plasmids pSZ701 and pAS727, carrying *bceR* and *yxdJ*, respectively, under control of xylose-inducible promoter were subjected to treatment with hydroxylamine hydrochloride. Subsequently, the plasmids were transformed into strains carrying P_{bce}- or P_{yxdJ}-*lacZ* fusions, lacking *psdR* and *bceR*, and expressing a constitutive ON chimeric BceS-PsdS HK (cHK.C, see Fig. 7.3). This kinase is able to specifically phosphorylate its cognate RR PsdR in the absence of any stimulus which leads to ~1000-fold induction compared to the uninduced wild type. Such strong induction can be visualised in a strain carrying a an appropriate *lacZ* promoter fusion on X-gal plates (Fig. 7.5).

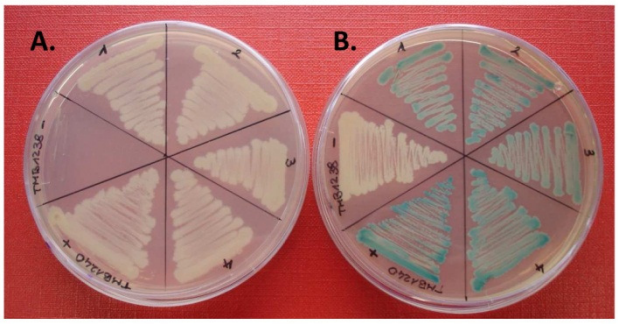

Figure 7.5. Mutants isolated in the random mutagenesis of BceR and YxdJ. Positive control strain (TMB1240, marked here with +), and four mutants (numbered 1-4) are able to grow on agar plates supplemented with spectinomycin (A) and turn blue on plates with X-gal, xylose and IPTG (B). Negative control strain (TMB1238, marked here with -) on the other hand, is not spectinomycin-resistant (A) and does not turn blue on plates with X-gal, xylose and IPTG (B).

In this random screen 12 independent mutants in BceR and one in YxdJ were isolated (Table 7.1, Fig. 7.6). The mutations are located in the receiver and DNA-binding domains, as well as in the flexible linker between them. This screen does not allow the discrimination between mutants that are able to interact with non-cognate HK and so called locked ON mutants. The expression of the latter leads to a constitutive activity of the promoter, independently of the stimulus. Therefore, further experiments are needed to distinguish between the RR variants with changed specificity and those with stimulus-independent activity.

Table 7.1. Mutations isolated in the random mutagenesis screen of BceR and YxdJ.

Position in Fig. 7.6	Strain	Protein	wt codon	codon after mutagenesis	AA exchange	Isolation[a]
1	TMB1402	BceR	**g**ga (G)	**a**ga (R)	G23R[b]	2x
2	TMB1413	BceR	t**g**t (C)	t**a**t (Y)	C48Y[b]	1x
3	TMB1431	BceR	**c**tc (L)	**t**tc (F)	L76F	1x
4	TMB1419	BceR	**g**at (D)	**a**at (N)	D82N	1x
5	TMB1443	YxdJ	**g**at (D)	**a**at (N)	D106N	1x
6	TMB1425	BceR	**g**ag (E)	**a**ag (K)	E125K	1x
7	TMB1401	BceR	**g**ac (D) ac**g** (T)	**a**ac (N) ac**a** (T)	D138N T192T	1x
8	TMB1435	BceR	at**g** (M)	at**a** (I)	M158I	1x
9	TMB1430	BceR	**g**ag (E)	**a**ag (K)	E176K	1x
10	TMB1400	BceR	**g**ac (D)	**a**ac (N)	D205N	2x
11	TMB1433	BceR	**g**gt (G)	**a**gt (S)	G210S	1x
12	TMB1429	BceR	**g**ag (E)	**a**ag (K)	E214K	1x
12	TMB1428	YxdJ	**g**aa (E)	**a**aa (K)	E216K	1x
11 & 12	TMB1418	BceR	**g**gt (G) **g**ag (E)	**a**gt (S) **a**ag (K)	G210S E214K	1x

[a] Number of times a mutation was isolated in independent mutagenesis experiments

[b] Strain carries an additional mutation in P_{xyl}

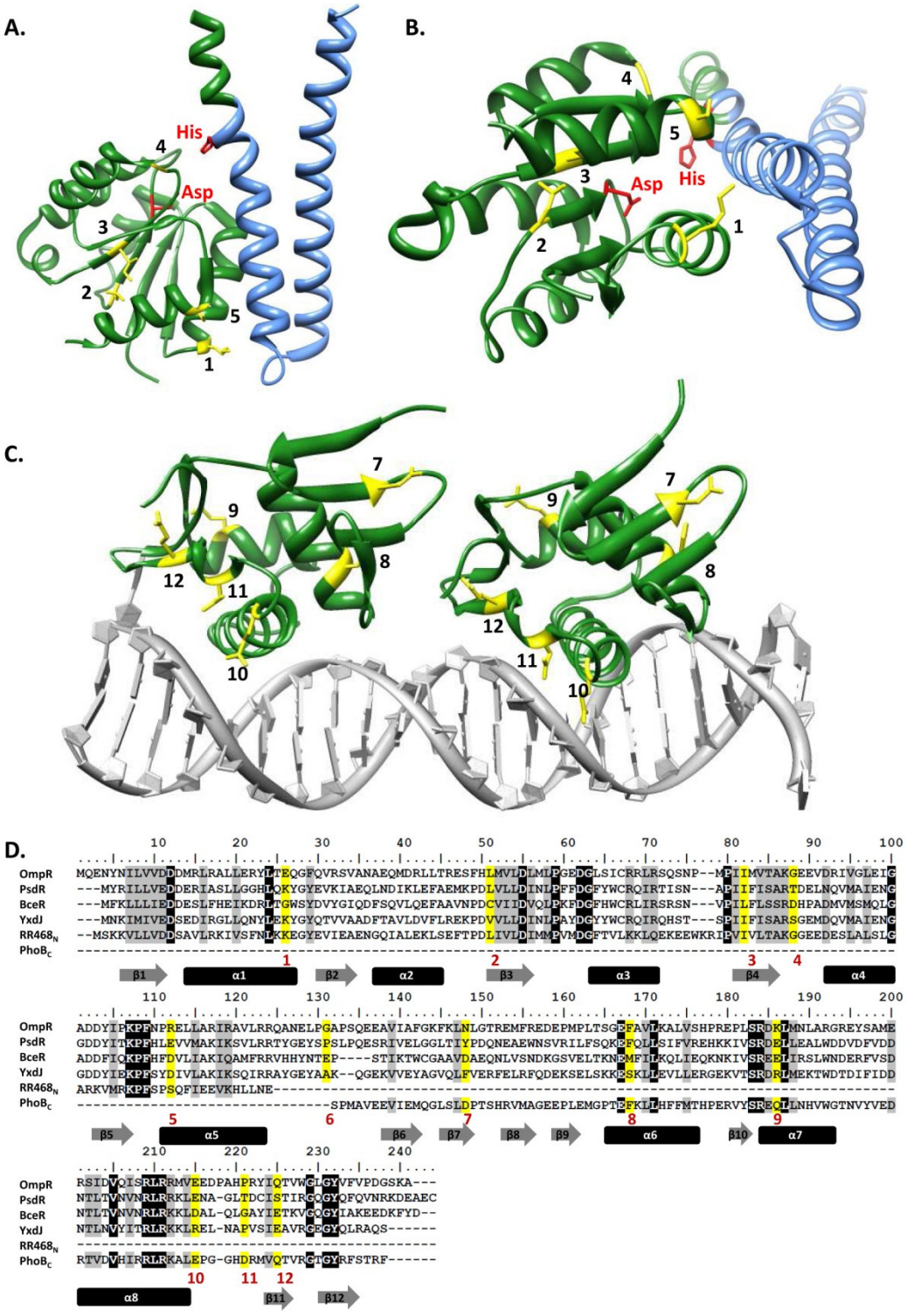

Figure 7.6. Mutations isolated in the random mutagenesis screen of BceR and YxdJ. (A-B) Mutated residues in the receiver domain of the RR shown on the co-crystal structure of HK853 and RR468 from *T. maritima* (Casino *et al.*, 2009). Mutated residues are marked in yellow and numbered according to their order in BceR. Blue ribbon in the structure of the HK represents a part originating from PsdS, green ribbon – part from BceS.The side chains of the conserved histidine and aspartate residues are shown as red sticks. (C) Mutated residues in the DNA binding domain of the RR shown on the co-crystal structure of PhoB-DNA complex from *E. coli* (Blanco *et al.*, 2002). Mutations are marked as described above. (D) Identified mutations shown on a primary sequence alignment of OmpR and PhoB from *E. coli*, PsdR, BceR and YxdJ from *B. subtilis*, and RR468 from *T. maritima*. Residues identical in all RR are highlighted in black, similar in grey. Mutations are highlighted in yellow and numbered as described above. Secondary structure elements,

based on the crystal structures of HK853 and RR468 (Casino *et al.*, 2009), as well as DNA-bound PhoB (Blanco *et al.*, 2002), are shown beneath the sequences.

7.3. Discussion and outlook

In this study, we investigated the specificity determinants in two Bce-like systems from *B. subtilis*, PsdRS-AB and BceRS-AB. We these two systems we searched for specificity determinants involved in (i) substrate recognition, (ii) interactions between HK and permeases, and (iii) interactions HK between and RR.

7.3.1. Loop as substrate recognition determinant

Previous reports on homologous systems from *S. aureus* and *S. epidermidis* suggest that (i) the extracytoplasmic loop in the HK is responsible for substrate recognition, and (ii) the negatively charged amino acids overrepresented in the loop are important for interaction with cationic substrates (Li *et al.*, 2007a; Li *et al.*, 2007b). We could not confirm this observation for the HK PsdS. Chimeric kinase consisting of PsdS core and the BceS extracellular loop show the same behaviour as a wild type PsdS. A substitution of negatively charged amino acids in the loop of PsdS with alanine changes neither the kinase specificity, nor the strength of induction. Introduction of glutamate residues into the chimera consisting of PsdS core and BceS loop also does not influence the kinase activity. It therefore appears that the extracytoplasmic loop is not involved in substrate recognition in *B. subtilis* kinases BceS and PsdS.

7.3.2. Interactions between HK and permease

For a number of Bce-like systems, including BceRSAB itself (Bernard *et al.*, 2007; Rietkötter *et al.*, 2008), but also for PsdRSAB (Staroń and Mascher, unpublished), MbrABCD (Ouyang *et al.*, 2010), GraRS-VraFG (Meehl *et al.*, 2007), and BraRSAB (Hiron *et al.*, 2011), it has been shown that the kinases require the presence of the ABC transporter for signal transduction. But despite the fast growing amount of experimental data on Bce-like systems, the exact mechanism of stimulus perception in BceRS-like 2CS is still unclear. The scenario favoured in the most recent studies on the subject proposes a direct contact between the HK and the ABC transporter as a necessary step in the signalling cascade (Dintner *et al.*, 2011; Hiron *et al.*, 2011).

The bacterial two hybrid assay revealed that PsdS and PsdB, as well as BceS and BceB can interact with each other. These results together with the absolute requirement of the

permease for sensing suggest that the proteins may exist as a sensory complex. To investigate this hypothesis, we constructed a series of chimeric kinases assuming that exchange of the specificity-determining residues will result in a kinase that interacts with non-cognate permease, and therefore responds to a non-cognate inducer. A clear limitation of this method is the assumption that only one region in the kinase is responsible for the contacts with the permease. But, since the BceS-like kinases are very minimalistic and are mostly reduced to core kinase domains (such as DHp and catalytic domains) and two transmembrane helices, this hypothesis seems feasible.

At this point, we can only speculate about the location of permease-specificity determinants in the BceS and PsdS HK. The kinase cHK.A shows the wild-type-like behaviour, which clearly suggests that the interaction does not take place through transmembrane helices or extracellular loop. As other chimeric kinases with progressively longer parts exchanged between PsdS and BceS are either constitutively active or inactive, we cannot yet identify the exact location of the permease-specificity determinant. As the DHp and catalytic domains are strongly conserved in all HK, also including those that do not interact with accessory proteins, it is tempting to speculate that the specificity determinant is indeed located outside these two domains. Since the transmembrane helices most probably do not function as specificity determinants, the prime candidate seems to be the loop connecting the DHp domain with the catalytic domain. It can be proposed that this short fragment lies in the direct vicinity of the inner surface of the cytoplasmic membrane and directly interacts with the permease. But further studies are needed to validate this hypothesis.

7.3.3. Specificity determinants in RR

Based on the covariance analysis, the residues predicted to be responsible for fidelity of interaction with the cognate HK are located in the first helix of the RR. However, an exchange of this helix rendered the RR inactive.

In a recently published paper, Capra and colleagues investigated the specificity determinants in three RR from the same protein family, namely OmpR, RstA and CpxR (Capra *et al.*, 2010). Their covariance analysis also pointed to the corresponding residues in the first helix. Additionally, the authors identified residues in the β5-α5 loop as also potentially participating in the specificity determination. An exchange of the first helix of OmpR for the first helix of CpxR led to a RR that showed weaker phosphotransfer from its cognate kinase EnvZ, as well as a weak phosphotransfer from the cognate kinase of CpxR,

CpxA. The rate of phosphotransfer was in both cases much weaker than in the wild type. Only after the concomitant exchange of the residues in the first helix and the β5-α5 loop (marked yellow in the Fig. 7.4A) were the authors able to completely exchange the specificity, as well as retain the rate of phosphotransfer similar to that of the wild type RR. Based on this data, it can be suggested that the exchange of the first helix was also not sufficient for the exchange of specificity between PsdR and BceR.

In the subsequent random mutagenesis experiment we searched for residues that are critical for interactions with the HK. The experimental set-up included a strain expressing a constitutive ON variant of the PsdS kinase (cHK.C) in the *psdR bceR* deficient background. In this screen we isolated gain-of-function mutants as blue colonies in the presence of IPTG (for the expression of the kinase) and xylose (for the expression of the RR). We were able to distinguish between the mutants in the xylose promoter and the RR mutants by testing on agar plates with xylose gradient.

The isolated mutants are located in the receiver domain, as well as the DNA binding domain and the linker between them (Fig. 7.6D). The mutations influencing interactions with the HK can be expected primarily in the receiver domain, as this is the part of the protein directly contacting the kinase (Casino *et al.*, 2009; Zapf *et al.*, 2000). Especially the first mutation G23R is very promising, as it is located directly at the end of the first helix, implied in the interactions with the HK. Mutations in the DNA binding domain probably influence the DNA binding affinity and give raise to constitutively active RR.

Smith and colleagues (Smith *et al.*, 2004) identified in the literature mutations in the receiver domains that lead to a constitutive activation of different RR. A comparison of 57 mutations identified in that study (located at 36 different amino acid positions) suggests that the activating mutations are located in clearly defined regions of the receiver domain, namely near the phosphorylation site, in the first helix and in the fourth and fifth helices and β strands. Comparison of these mutations with the ones identified in the random mutagenesis of BceR and YxdJ (Fig. 7.7) suggests that especially the first mutation (G23R) can be a candidate mutant with exchanged specificity. Two mutations L76F and D82N on the other hand are exactly in the region where the activating mutations are often located, and therefore it is plausible that these two mutations lead to a constitutive activation of the RR. However, all of these assumptions have to be tested experimentally.

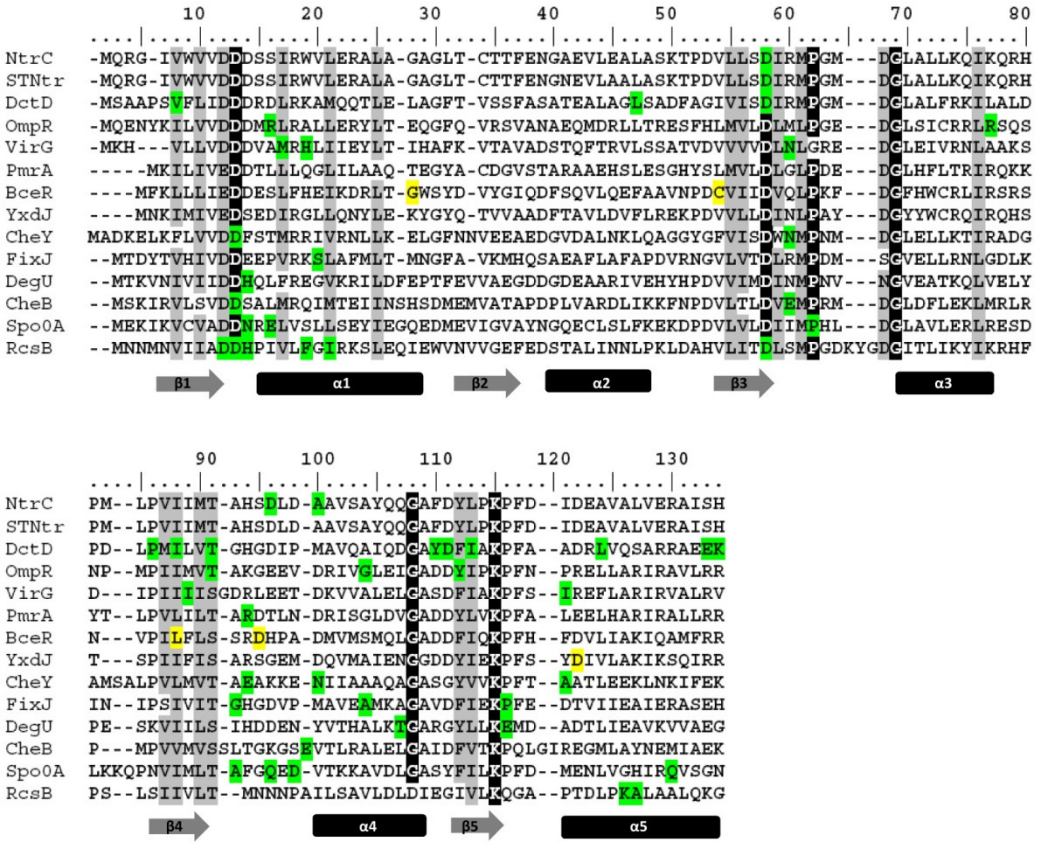

Figure 7.7. Distribution of activating mutations in the receiver domain of RR. RR activating substitutions identified in (Smith *et al.*, 2004) are highlighted in green in the multiple sequence alignment of the receiver domains. Mutations isolated in this study are highlighted in yellow. Secondary structure elements are based on the crystal structures of HK853 and RR468 (Casino *et al.*, 2009) and are shown beneath the sequences.

In summary, the data presented here shows the first insights into the specificity maintenance in BceRS-like detoxification systems in *B. subtilis*. These first results are a basis for future experiments that will focus on (i) interaction of the HK PsdS and BceS with ABC transporters PsdB and BceB, and (ii) interactions of the HK with their cognate RR, and the specificity determinants governing them.

7.4. Materials and methods

7.4.1. Bacterial strains and growth conditions

B. subtilis and *E. coli* strains were routinely grown in Luria Bertani (LB) medium at 37°C with agitation. All strains used in this study are listed in Table 7.2. Ampicilin (100 µg/ml) and kanamycin (50 µg/ml) was used for the in *E. coli*. Chloramphenicol (5 µg/ml), kanamycin (10 µg/ml), spectinomycin (100 µg/ml), and erythromycin (5 µg/ml) with lincomycin (25 µg/ml) for macrolide-lincosamide-streptogramin B (MLS) resistance were used for the selection of the *B. subtilis* strains used in this study. Transformation was carried out as described (Harwood and Cutting, 1990).

7.4.2. Construction of transcriptional promoter-*lacZ* fusions

Strains and plasmids used in this study are listed in Tables 7.2 and 7.3, respectively. Ectopic integrations of promoter *lacZ* fusions were constructed based on vector pAC6 (Stülke *et al.*, 1997). Promoter fragments were generated by PCR (primers are listed in Table 7.4), standard cloning techniques were applied (Sambrook and Russell, 2001). The inserts were verified by DNA sequencing. The resulting pAC6 derivates were linearised by ScaI and used to transform *B. subtilis* with chloramphenicol selection.

7.4.3. Promoter induction assays

Cells were inoculated from fresh overnight cultures and grown in LB medium at 37°C with agitation until they reached an optical density at 600 nm (OD_{600}) of ~0.45. The culture was split, and an inducing substance (bacitracin at 30 µg/ml or nisin at 2µg/ml) was added to one half, leaving the other half untreated (uninduced control). Both cultures were incubated for 30 min at 37°C with agitation. Cell pellets were resuspended in 1 ml of buffer (20 mM β-mercaptoethanol, 60 mM Na_2HPO_4, 10 mM KCl, 1 mM $MgSO_4$ [pH 7.0] and assayed for β-galactosidase activity as described previously, with normalization to cell density (Miller, 1972).

7.4.4. Allelic replacement mutagenesis using LFH-PCR

The long flanking homology PCR (LFH-PCR) technique is derived from a published procedure (Wach, 1996) and was performed as described previously (Mascher *et al.*, 2003). Briefly, a kanamycin resistance cassette was amplified by PCR using the vector pDG780 (Guerout-Fleury *et al.*, 1995) as template. Two primer pairs were designed to amplify ~1,000-bp DNA fragments flanking the region to be deleted at its 5' and 3' ends. The resulting fragments are here called the "up" and "do" fragments. The 3' end of the up fragment as well as the 5' end of the do fragment extended into the gene to be deleted in a way that all expression signals of genes up- and downstream of the targeted genes remained intact. Extensions of ~25 nucleotides were added to the 5' end of the up-reverse and the do-forward primers that were complementary (opposite strand and inverted sequence) to the 5' and 3' ends of the amplified resistance cassette. 100 to 150 ng of the up and do fragments and 250 to 300 ng of the kanamycin cassette were used together with the specific up-forward and do-reverse primers at standard concentrations in a second PCR reaction. In this reaction the three fragments were joined by the 25-nucleotide overlapping complementary ends and simultaneously amplified by normal primer annealing. The PCR products were directly used to transform *B. subtilis* W168 with kanamycin selection. Transformants were screened by colony PCR, using the up-forward primer with a reverse check primer annealing inside the resistance cassette (see Table 1). The integrity of the regions flanking the integrated resistance cassette was verified by sequencing PCR products of ~1,000 bp amplified from chromosomal DNA of the resulting mutants.

7.4.5. Construction of markerless deletion mutants

Clean deletion mutants were constructed using the vector pMAD (Arnaud *et al.*, 2004). Two primer pairs were designed to amplify ~1,000-bp DNA fragments flanking the region to be deleted at its 5' and 3' ends. The resulting fragments are here called the "up" and "do" fragments. As was the case with LFH method, the 3' end of the up fragment as well as the 5' end of the do fragment extended into the gene to be deleted in a way that all expression signals of genes up- and downstream of the targeted genes remained intact. Extension of ~25 nucleotides was added to the 5' end of the do-forward primer so that the do fragment was complementary (opposite strand and inverted sequence) to 3' end of the up fragment. 15 ng of do and up fragments (in equimolar concentrations) were used together with the specific up-forward and do-reverse primers at standard concentrations in a second PCR reaction. The resulting fragment was cloned into pMAD. Generation of the clean deletion followed the established procedure (Arnaud *et al.*, 2004). Briefly, *B. subtilis* W168 was transformed with pMAD derivative and incubated at 30°C with MLS selection on LB agar plates with X-Gal (here and later: 100 µg/ml). Blue colonies were picked and incubated for 6–8 h at 42°C in LB medium with MLS selection, resulting in the integration of plasmid into the chromosome. Again, blue colonies were picked from LB plates (with X-Gal and MLS selection) and incubated for 6 h at 30°C in LB medium without selection. Subsequently, the liquid culture was shifted to 42°C for 3 h, and the cells were then plated on LB (X-Gal) plates, this time without selective pressure. White colonies that had lost the plasmid were picked and checked for MLS sensitivity. Those harbouring a clean deletion (~ 50% of the white clones) were identified by PCR and subsequent sequencing.

7.4.6. Complementation of mutants

An ectopic integration of gene of interest fused to P_{xyl} was constructed based on the vectors pXT or pAL-FLAG (Derre *et al.*, 2000; Schöbel *et al.*, 2004). Vector pXT is a pDG1782 derivative (Guerout-Fleury *et al.*, 1996) that carries a spectinomycin resistance cassette for selection in *B. subtilis* and a xylose-inducible promoter for expression of gene of interest. It integrates into the *thrC* locus by double crossing-over, resulting in a stable integration. pAL-FLAG carrier a MLS cassette for selection in *B. subtilis* and an IPTG-inducible promoter. It integrates into the *lacA* locus. Genes of interest were cloned into pXT and pAL-FLAG using standard cloning methods. For chimeric kinases and RR, separate fragments were fused in sequential PCR reaction, using 15 ng DNA at equimolar concentrations. The resulting derivatives were linearized with ScaI (for pXT) or PvuI (for pAL-FLAG) and used to transform *B. subtilis* strains with appropriate selection. For expression of gene of interest, 0.2% xylose or 0.05 mM IPTG was added to the LB medium.

7.4.7. *In vitro* hydroxylamine mutagenesis

5-10 µg of pAS727 or pSZ701 was incubated for 15 min at 70°C with 1 M hydroxylamine in a total volume of 0.5 ml of 0.05 M sodium phosphate buffer (pH 7.0), 0.1 M NaCl and 2 mM EDTA. Plasmids were purified using an PCR clean-up kit (SLG), linearised with ScaI, and used for transformation of TMB1238 or TMB1423. Transformants were grown overnight on LB-agar plates with spectinomycin selection. Subsequently, transformants were replica plated on LB-agar plates supplemented with X-gal (100 µg/ml), xylose (0.2%), IPTG (0.05 mM) and spectinomycin. Dark blue colonies were transferred to xylose gradient plates (first layer: 25 ml LB with X-gal 100 µg/ml, xylose 0.2% and IPTG 0.05 mM, second layer: 50 ml LB with X-gal 100 µg/ml and IPTG 0.05 mM) in order to distinguish between mutants in RR itself (visible colour change along xylose gradient) and mutants in P_{xyl} (blue colour not dependent on the xylose concentration). PCR products amplified with chromosomal DNA of the mutants as a template were sequenced.

7.4.8. Bacterial two-hybrid assay

The bacterial two-hybrid assay was performed as described previously (Karimova *et al.*, 1998; Karimova *et al.*, 2000). Briefly, adenylate cyclase-deficient *E. coli* strain BTH101 was co-transformed with all combinations of pUT18 and pKT25-derived plasmids and plated as spots on LB agar plates containing X-gal (100 µg/ml), kanamycin (50 µg/ml) and ampicilin (100 µg/ml). Colonies of strains with proteins that do not interact remain white (Lac^- phenotype), while strains coexpressing interacting proteins turn blue (Lac^+ phenotype).

Table 7.2. Bacterial strains used in this study.

Strain	Genotype or characteristic	Reference or source
E. coli		
DH5α	*recA1 endA1 gyrA96 thi hsdR17rK- mK +relA1 supE44 Φ80 DlacZDM15 D(lacZYA-argF)U169*	Laboratory stock
BTH101	F⁻*cya-99 araD139 galE15 galK16 rpsL1* (Str^R) *hsdR2 mcrA1 mcrB1*	(Karimova *et al.*, 1998)
B. subtilis		
TMB551	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS702 (PsdS_loop1)	This study
TMB579	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS703 (cHK.A)	This study
TMB580	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS704 (cHK.F)	This study
TMB581	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS705 (cHK.C)	This study
TMB594	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS707 (PsdS_loop2)	This study
TMB636	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS709 (PsdS_loop3)	This study
TMB646	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS710 (cHK.E)	This study
TMB701	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS714 (cHK.D)	This study
TMB711	W168 *psdR*::kan *amyE*::pER605 *thrC*::pAS712 (cHK.B)	This study
TMB1142	W168 Δ*bceR amyE*::pJS605 *thrC*::pAS721 (BceR-FLAG)	This study
TMB1143	W168 Δ*psdR amyE*::pDF602 *thrC*::pAS720 (PsdR-FLAG)	This study
TMB1183	W168 Δ*psdR amyE*::pDF602 *thrC*::pAS722 (PsdR($\alpha1_{BceR}$)-FLAG)	This study
TMB1184	W168 Δ*bceR amyE*::pJS605 *thrC*::pAS723(BceR($\alpha1_{PsdR}$)-FLAG)	This study
TMB1238	W168 Δ*bceR psdR*::kan *lacA*::pASNFLAG03 *amyE*::pER603	This study
TMB1240	W168 Δ*bceR psdR*::kan *lacA*::pASNFLAG03 *amyE*::pSZ701	This study
TMB1423	W168 Δ*bceR psdR*::kan *lacA*::pASNFLAG03 *amyE*::pPH601	This study

Table 7.3. Vectors and plasmids used in this study.

Plasmid	Genotype or characteristic features[a]	Primers used for cloning	Reference or source
pAC6	Transcriptional *lacZ* fusion vector, integrates at *amyE*, chloramphenicol resistance		(Stülke *et al.*, 1997)
pAL-FLAG	Vector for IPTG-inducible gene expression and N-terminal FLAG-tagging, integrates at *lacA*, MLS resistance		(Schöbel *et al.*, 2004)
pMAD	Vector for introduction of markerless deletions, MLS resistance		(Arnaud *et al.*, 2004)
pXT	Vector for xylose-inducible gene expression, intergrates at *thrC*, spectinomycin resistance		(Derre *et al.*, 2000)
pAS702	pXT P_{xyl}-*psdS* E33A E37A E41A	0768/1026 and 1027/0769	This study
pAS703	pXT P_{xyl}-*bceS*(M_1-E_{60})-*psdS*(F_{66}-L_{356})	0650/1104 and 1105/1103	This study
pAS704	pXT P_{xyl}-*bceS*(M_1-T_{84})-*psdS*(G_{97}-L_{356})	0650/1121 and 1122/1103	This study
pAS705	pXT P_{xyl}-*bceS*(M_1-A_{121})-*psdS*(W_{134}-L_{356})	0650/1123 and 1124/1103	This study
pAS707	pXT P_{xyl}-*psdS*(M_1-V_{32})-*bceS*(D_{30}-V_{38})-*psdS*(M_{44}-L_{356})	0768/1201 and 1202/0769	This study
pAS709	pXT P_{xyl}-*psdS*(M_1-V_{32})-*bceS*(D_{30}-V_{38} D30E I33E G36E)-*psdS*(M_{44}-L_{356})	0768/1270 and 1271/769	This study
pAS710	pXT P_{xyl}-*bceS*(M_1-V_{38})-*psdS*(M_{44}-L_{356})	0650/1269 and 1202/0769	This study
pAS712	pXT P_{xyl}-*psdS*(M_1-Q_{65})-*bceS*(T_{61}-V_{334})	1410/1401 and 1402/0651	This study
pAS714	pXT P_{xyl}-*psdS*(M_1-Q_{65})-*bceS*(T_{61}-I_{123})-*psdS*(H_{136}-L_{356})	1400/1401, 1402/1403 and 1404/1409	This study
pAS715	pXT P_{xyl}-*psdS*(M_1-Q_{65})-*bceS*(T_{61}-I_{123})-*psdS*(H_{136}-	1400/1401,	This study

	R_{180})-*bceS*(I_{169}-T_{230})-*psdS*(N_{251}-L_{356})	1402/1403, 1404/1405, 1406/1407 and 1408/1409	
pAS720	pXT P_{xyl}-*psdR-FLAG*	1777/1833 and 0960/1161	This study
pAS721	pXT P_{xyl}-*bceR-FLAG*	1779/1834 and 0960/1203	This study
pAS722	pXT P_{xyl}-*psdR*(M_1-E_{11})-*bceR*(S_{12}-T_{22})-*psdR*(K_{23}-C_{237})-*FLAG*	0037/2024 and 2025/0038	This study
pAS723	pXT P_{xyl}-*bceR*(M_1-E_{11})-*psdR*(R_{12}-Q_{22})-*bceR*(G_{23}-D_{231})-*FLAG*	0037/2026 and 2027/0038	This study
pAS727	pXT P_{xyl}-*yxdJ*	2269/2270	This study
pAS1801	pUT18-*psdS*	1253/1255	This study
pAS1802	pUT18-*psdB*	1259/1261	This study
pAS1803	pUT18-*bceS*	1355/1356	This study
pAS1805	pUT18-*bceB*	1359/1360	This study
pAS18C01	pUT18C-*psdS*	1253/1254	This study
pAS18C02	pUT18C-*psdB*	1259/1260	This study
pAS18C03	pUT18C-*bceS*	1355/1356	This study
pAS18C05	pUT18C-*bceB*	1359/1360	This study
pAS2501	pKT25-*psdS*	1253/1254	This study
pAS2502	pKT25-*psdB*	1259/1260	This study
pAS2503	pKT25-*bceS*	1355/1356	This study
pAS2505	pKT25-*bceB*	1359/1360	This study
pAS25N01	pKT25N-*psdS*	1253/1255	This study
pAS25N02	pKT25N-*psdB*	1259/1261	This study
pAS25N03	pKT25N-*bceS*	1355/1356	This study
pAS25N05	pKT25N-*bceB*	1359/1360	This study
pASNFLAG03	pAL P_{spac}-*bceS*(M_1-A_{121})-*psdS*(W_{134}-L_{356})	2049/2050	This study
pDF602	pAC6 $P_{psdA(-105-30)}$-*lacZ*	0675/0600	(Staroń *et al.*, 2011)
pER603	pAC6 $P_{bceA(-122-82)}$-*lacZ*	0554/0555	(Rietkötter *et al.*, 2008)
pER605	pAC6 $P_{psdA(-110-30)}$-*lacZ*	0599/0600	(Rietkötter *et al.*, 2008)
pJS605	pAC6 $P_{bceA(-111-82)}$-*lacZ*	1307/0555	This study
pPH601	pAC6 $P_{yxdL(-194-57)}$-*lacZ*	1159/1160	(Staroń *et al.*, 2011)
pSZ701	pXT P_{xyl}-*psdR*	1777/1778	This study
pSZ702	pXT P_{xyl}-*bceR*	1779/1780	This study

[a] The positions of the cloned fragments are given relative to the "A" of the start codon of the corresponding gene

Table 7.3. Primers used in this study.

Nr	Sequence[a]
0037	CCTTACCGCATTGAAGGCC
0038	GTATTCACGAACGAAAATCGCC
0554	GATC**GAATTC**GAACATGTCATAAGCGTGTGACG
0555	GATC**GGATCC**TATCGATGCCCTTCAGCACTTCC
0599	AGTC**GAATTC**CACCCTCGTGAATGTGACAGC
0600	AGTC**GGATCC**CGATAGGTTCGTTGTTTGCAACACG
0650	AGTC**GGTCTCGGATCC**TACATCGCGAAGGAAGAGG
0651	AGTC**GGTCTCGAATTC**GGGCTTTTCCTTCGATACGG
0675	AGTC**GAATTC**GTGAATGTGACAGCATTGTAAG
0768	ACG**GGATCC**CGGAAGGATGAAGCGGAATG
0769	AGCT**GAATTC**CATCGAACTTTCTTGCAATTCC
0960	GATTATAAGGATCATGATGGTG
1026	CATATTTGCTTTGGAAGGTGCTGCGCCGCTCGCTACAATGCTTAAATACGCAATCAGC
1027	GCATTGTAGCGAGCGGCGCAGCACCTTCCAAAGCAAATATGATGTACATATGGATTTTG C
1103	AGTC**GGTCTCGAATTC**CATCGAACTTTCTTGCAATTCC
1104	TTCTTTCCGATAGCGGAACCAAAGG
1105	GGTTCCGCTATCGGAAAGAATTTTCCTTTCTCACATATGTTAAAAAGC
1121	CGTTTCCGGTTCATTTATGG
1122	CCATAAATGAACCGGAAACGGGAGAGCAAGCGCTTTGG
1123	TGCCATCAGCTCATCTTTTTC
1124	GAAAAAGATGAGCTGATGGCATGGGTCCACCATATGAAAACACC
1159	GCAT**GAATTC**TCTCCCGGTGAAGGGACATC
1160	GCAT**GGATCC**AGGACACTTGTCCTTTATAGG
1161	ACG**AAGCTT**TCACTTGTCGTCATCGTCTTTGTAG
1201	TACATTCCCAAATGATATGGAAGGATCTACAATGCTTAAATACG
1202	GATCCTTCCATATCATTTGGGAATGTAATGTACATATGGATTTTGC
1203	TACG**GAATTC**TCACTTGTCGTCATCGTCTTTGTAG
1253	GCTA**GGATCC**GATGTTAAAGACTTATCTCATCG
1254	GCTA**GAATTC**TTACAATGCTGTCACATTCACG
1255	GCTA**GAATTC**GACAATGCTGTCACATTCACG
1259	GCTA**TCTAGA**GATGAATTTGAGAACTATCG
1260	GCTA**GGTACC**TTACATACGCTGAAGAACAGC
1261	GCTA**GGTACC**ATCATACGCTGAAGAACAGC
1269	TACATTCCCAAATGATATGGAAGGATC
1270	TACATTCTCAAATGACTCGGAAGGCTCTACAATGCTTAAATACG
1271	GAGCCTTCCGAGTCATTTGAGAATGTAATGTACATATGGATTTTGC
1307	GATC**GAATTC**AAGCGTGTGACGAAAATGTCACAT
1355	GTCA**TCTAGA**GATGATTAAAGCATTCCTTATCG
1356	GTCA**GGTACC**TGCACGCTTATGACATGTTC
1359	GTCA**TCTAGA**GATGAACATTAATCAGCTCATCC
1360	GTCA**GGTACC**TGCAACGACGATTTAATGACC
1400	GTAC**GAAGACAGGATCC**CGGAAGGATGAAGCGGAATG
1401	TTGCCGAACATAATCAGCAGC
1402	CTGCTGATTATGTTCGGCAAACAGCGTTTTATAAAAGCTTG
1403	GATCCATGCCATCAGCTCATC
1404	TGAGCTGATGGCATGGATCCACCATATGAAAACACCGGTATC
1405	CCGTGCCGTTTGCAGCATC
1406	GATGCTGCAAACGGCACGGATTTCATTTATTGAAAACGACC
1407	TGTCAGCAGCTGCCTGATGATAAACG
1408	TCATCAGGCAGCTGCTGACAAACGCCTTAAAATATTCAAAGC
1409	GTAC**GAAGACAGAATTC**CATCGAACTTTCTTGCAATTCC
1410	GTAC**GGTCTCGGATCC**CGGAAGGATGAAGCGGAATG
1777	GCAT**GGATCC**TCGTGATGTACATATAAGGG
1778	GCAT**AAGCTT**AGCGGTCGATGAGATAAGTC
1779	GCAT**GGATCC**GGAAAAAGTGCGGTGAGTCG
1780	GCAT**GAATTC**CGCCTTTCGATAAGGAATGC
1833	CACCATCATGATCCTTATAATCACATTCCGCTTCATCCTTCC

1834	CACCATCATGATCCTTATAATCATCATAGAACTTGTCCTCTTCC
2024	CGATCCTTGATTTCATGAAACAGCGACTCATCATCTTCCACAAGC
2025	TTTCATGAAATCAAGGATCGTTTAACGAAATACGGATATGAAGTG
2026	TGACCGCCCAGCAAAGAAGCAATCCGTTCATCATCTTCAATCAGC
2027	CTTCTTTGCTGGGCGGTCATCTTCAAGGATGGTCCTATGATGTATACG
2049	GATC**GGATCC**ATGATTAAAGCATTC
2050	GATC**GCATGC**ATCGAACTTTCTTGC
2269	GCAT**GGATCC**TGTAAGAAACGGGAGTGGTTC
2270	GCAT**GAATTC**TCATGACTGCGCCCTCAGCTGG

[a] Restriction sites are highlighted in bold, linker sequences for joining reactions are underlined.

CHAPTER 8

Discussion

Parts of this chapter have been adapted from:

A. Staroń, and T. Mascher

Microbe (2010) 5(4):164-170

and

K. Schrecke*, A. Staroń*, T. Mascher

** contributed equally*

Two component systems in bacteria. D. Beier, and R. Gross (eds.). Horizon Scientific Press, *in press*

Chapter 8
Discussion

Signal transducing mechanisms are means that allow the microorganisms to sense the changing relevant parameters and react to them in a precise and timely manner. Bacteria employ a variety of such mechanisms, most of which belong to one of the three principles: 1CS, 2CS and ECF σ factors (Staroń and Mascher, 2010b).

This thesis aimed at gaining an insight into the common themes and principles governing two of the subfamilies of signalling proteins, namely the ECF σ factors, and the Bce-like 2CS. We analysed the ECF σ factors in all sequenced genomes, and developed a classification system that provides predictions based on known ECF σ factor-dependent mechanisms, and suggests novel modes of ECF-dependent signal transduction (Chapter 2). The classification is followed by a in-depth study of a single group of ECF σ factors orchestrating the general stress responses of α-proteobacteria, where we discuss the unique features of these proteins, as well as other players involved in the signalling cascade (Chapter 3). In Chapter 4, we again apply the methods of comparative genomics to analyse the family of Bce-like 2CS and associated ABC transporters. This analysis is followed by a case study of these systems in *B. subtilis*, where we investigate the inducer and resistance spectra, as well as the regulation in Bce-like systems (Chapter 5). The following in-depth studies investigate the specificity determinants in promoters recognized by the Bce-like systems in *B. subtilis* (Chapter 6), as well as specificity determinants in the 2CS themselves (Chapter 7). Because this thesis deals with different families of signal transducing proteins, the findings will be discussed in three separate sections, one addressing the results on ECF σ factors, the second the classification of Bce-like 2CS, while the third part deals with the studies of Bce-like 2CS in *B. subtilis*. As the detailed findings are discussed in the short sections at the end of each chapter, this general discussion aims to primarily summarise the major results in order to put them in the broader context of the recent literature.

8.1. Classification of ECF σ factors: systematisation and predictions

ECF σ factors are small proteins that together with other subunits form the complex of RNA polymerase. Since the first paper by Lonetto and colleagues (Lonetto *et al.*, 1994) it has become clear that this wide class of proteins is playing an important role in the signal

transduction, mainly, although not exclusively, in the responses to extracytoplasmic signals. A number of ECF-dependent systems have been investigated, both in the model organisms such as *E. coli* and *B. subtilis*, as well as in pathogens (*M. tuberculosis*) or microorganisms with complex differentiation cycles (*S. coelicolor* and *C. crescentus*). But it requires comparative genomic approaches in order to extract unifying or distinguishing features of ECF dependent regulation, as has previously been performed for 1CS and 2CS (Grebe and Stock, 1999; Ulrich *et al.*, 2005).

The field of comparative genomics came to life in 1995, when the first two bacterial genomes were sequenced, providing first insights into genetic organisation of *Haemophilus influenzae* (Fleischmann *et al.*, 1995) and *M. genitalium* (Fraser *et al.*, 1995). As of February 2012 there are more than 3500 fully sequenced bacterial genomes, and more than 6000 in progress, which provides the field with enormous opportunities and allowed many general analyses of structure and evolution of bacterial genomes (Koonin and Wolf, 2008). These possibilities have been also extensively exploited by the field of signal transduction.

The most intensive research has been done in the field of 2CS. Analyses of 2CS in different species like *M. xanthus* (Shi *et al.*, 2008) and *P. syringae* (Lavin *et al.*, 2007), or certain phyla, like Cyanobacteria (Ashby and Houmard, 2006) were performed, demonstrating the variety of signalling systems in various organisms. A comprehensive study of mechanisms involved in cell envelope stress in Gram positive bacteria was also published (Jordan *et al.*, 2008). More importantly, a classification of 2CS was published in 1999 (Grebe and Stock, 1999). This study allows for comparisons inside the HK family, and for the identification of direct homologues. Therefore, it also enables the predictions of function for unknown kinases based on the sequence information alone. Wide genomic analyses are also available for 1CS (Ulrich *et al.*, 2005). However, a comparable analysis of ECF σ factors has been long overdue.

8.1.1. Variety of mechanisms in ECF-dependent signal transduction

Compared to the wealth of information on 1CS and 2CS, the overall knowledge of this third fundamental mode of bacterial signal transduction is still very sparse. However, recent analyses start to shed light on the variety of mechanisms employed in ECF-dependent signal transduction throughout the microbial world (Chapter 2, Table 7.1.). Our in-depth phylogenomic analysis of ECF σ factors identified more than 40 distinct subtypes of σ/anti-σ pairs (Chapter 2, (Staroń *et al.*, 2009)). This study not only revealed the wide distribution of ECF-dependent signal transduction in the microbial world, but also

demonstrated the strict modularity of functional domains observed for 1CS and 2CS. While most ECF σ factors have identical domain compositions, the corresponding anti-σ factors embody surprising variety and combinatorial complexity in terms of their conserved modules. A widely distributed and highly diversified anti-σ domain (ASD) is present in about one-third of all anti-σ factors associated with ECF-dependent signal transduction. Other anti-σ factors harbour a variety of conserved domains of unknown function. Because of this sequence diversity, it is difficult to recognize anti-σ factors on the basis of sequence analysis.

Further, there is significant — and underappreciated — mechanistic diversity in how anti-σ factors perceive signals and how corresponding σ factors are activated. Nevertheless, potential anti-σ factors from within a conserved subtype of ECF σ factors are usually homologous to each other. This finding suggests that σ-/anti-σ pairs from a given conserved subgroup will have the same mechanism of signal transduction.

By combining the in-depth mechanistic knowledge from the ECF paradigms with the results from the comprehensive *in silico* analysis, a number of prominent and conserved mechanisms of ECF-dependent signal transduction emerge. The most widely used mechanisms (based on current knowledge) are illustrated/summarized in Fig. 7.1/Table 7.1.

Table 7.1. Paradigms of ECF σ factors according to their mode of activation (see text for details)

Mechanism of activation	Organism	ECF σ factor	Anti-σ factor	Anti-σ domains	Accessory proteins	Physiological role
Regulated anti-σ proteolysis	*E. coli*	RpoE	RseA	ZAS	RseBC?	Envelope stress response
	B. subtilis	σ^W	RsiW	ASD	n.a.	Envelope stress response
Conformational change of anti-σ factor	*S. coelicolor*	SigR	RsrA	ZAS	n.a.	Oxidative stress response
	R. sphaeroides	RpoE	ChrR	ZAS	n.a.	Oxidative stress response
Protein interaction cascade	*E. coli*	FecI	FecR	ASD	FecA (porin)	Iron acquisition
Transcriptional activation	*S. coelicolor*	SigE	n.a.[a]	n.a.	CseBC (2CS)	Envelope stress response
Partner switching mechanism	*M. extorquens*	EcfG	NepR	Unique	HK/PhyR (2CS)	General stress response

[a] n.a., not available

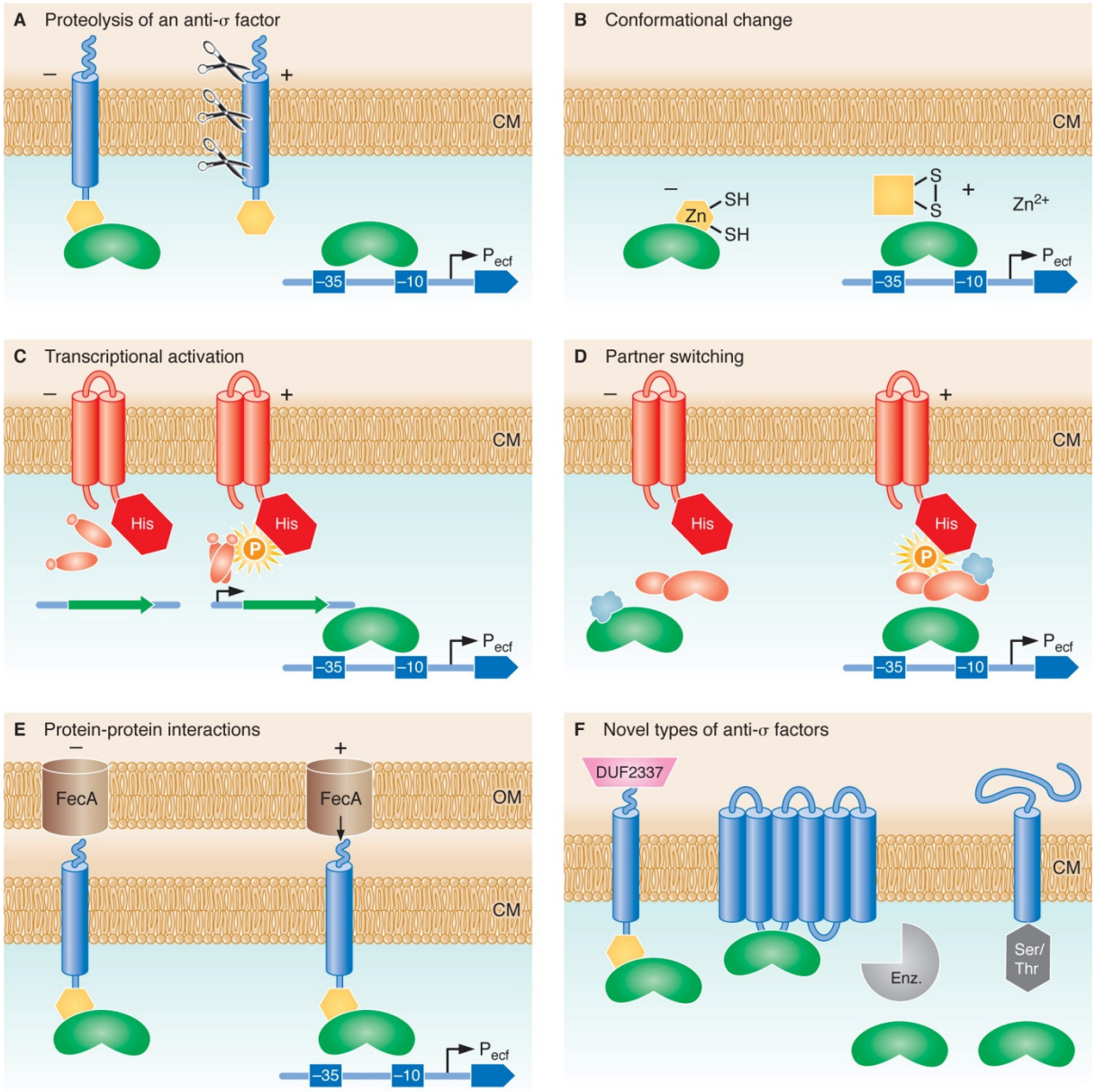

Figure 7.1. Examples and mechanisms of ECF-dependent signal transduction.

8.1.1.1. Regulated (Intramembrane) Proteolysis of the Anti-σ Factor (Fig. 7.1A). Most ECF σ factors interact with membrane-spanning anti-σ factors that harbour a single transmembrane helix. In the absence of a stimulus, the anti-σ factor binds the σ factor, rendering it inactive. After detecting a signal, presumably by a surface-exposed protease, the anti-σ factor is degraded in a stepwise manner, eventually freeing the σ factor to interact with RNA polymerase. The two best-understood ECF σ/anti-σ pairs, *E. coli* RpoE-RseA and *B. subtilis* σ^W-RsiW, belong to this group (Ades, 2008; Helmann, 2006). Here, the cytoplasmic part of the anti-σ factor harbours an ASD that becomes the interface between anti-σ factor and ECF σ factor and defines partner specificity.

8.1.1.2. Reversible Conformational Changes of the Anti-σ Factor (Fig. 7.1B). About one-third of known ASD-containing anti-σ factors coordinate a zinc ion. Such a cytoplasmic domain is called a zinc-binding anti-σ (ZAS) domain (Campbell *et al.*, 2007; Zdanowski *et al.*, 2006). It plays a crucial role in sensing oxidative stress and mounting an adequate stress response. The two best-understood examples are the SigR-RsrA pair from *S. coelicolor* and RpoE-ChrR from *R. sphaeroides* (Greenwell *et al.*, 2011; Li *et al.*, 2002). Again, the σ factor is kept in its inactive state by being bound to its cognate anti-σ factor. Under non-inducing conditions, the zinc cofactor is coordinated by cysteine residues. A change in the cytoplasmic redox potential, indicative of oxidative stress, leads to the formation of a disulfide bond, releasing the zinc cofactor. Thus, the anti-σ factor changes its conformation, releasing its cognate σ factor, which then interacts with RNA polymerase. This mechanism, in contrast to proteolytic degradation of anti-σ factors, is reversible. Hence, the anti-σ factor can be recycled.

8.1.1.3. Protein Interaction Cascades (Fig. 7.1E). Another ECF-dependent signal transduction example is the FecI-FecR pair, which regulates iron acquisition in *E. coli* and other Proteobacteria (Braun *et al.*, 2003; Braun and Mahren, 2005). At first glance, this system contains familiar components, including an anti-σ factor with one transmembrane helix that binds the σ factor FecI via its ASD in the absence of ferric citrate, thereby keeping it inactive. However, σ factor activation is quite different. In this cases, the anti-σ factor remains intact after receiving the signal through protein-protein interactions with a ferric citrate-specific outer membrane porin, FecA. Moreover, FecR is necessary for FecI interactions with RNA polymerase. Thus, it acts as both an anti- σ factor and a mediator of σ factor activity.

8.1.1.4. Transcriptonal Control of ECF Gene Expression (Fig. 7.1C). Not all ECF σ factors are regulated by anti-σ factors. Sometimes, 2CS components combine with ECF σ factors, forming a variety of signaling pathways across the bacterial kingdom. For instance, the SigE-CseCB module from *S. coelicolor* forms a simple yet effective connection between a 2CS and an ECF σ factor (Hutchings *et al.*, 2006; Paget *et al.*, 1999a; Paget *et al.*, 1999b). CseC is a HK localized within the cytoplasmic membrane. After envelope stress is perceived, CseC phosphorylates CseB, its cognate RR, and it, in turn, induces transcription of *sigE*, encoding the ECF σ factor. SigE, which lacks a cognate anti-σ factor, thus enters the cellular σ factor pool and redirects expression to target genes.

8.1.1.5. Partner-Switching Mechanism (Fig. 7.1D). The σ-proteobacteria orchestrate their general stress response with an ingenious mechanism that combines 2CS- and ECF-dependent signal transduction (reviewed in (Staroń and Mascher, 2010a), Chapter 3). In the absence of stress, the σ factor (EcfG) is kept inactive, bound to soluble NepR-like anti-σ factors, which are unlike other anti-σ factors. Once stress is perceived, it is transduced by a 2CS that contains an unusual PhyR-like RR that harbours an output domain that is homologous to EcfG-like σ factors. When phosphorylated by its cognate HK, this protein acts as an anti-anti-σ factor: The anti-σ factor has a higher affinity for phosphorylated PhyR than for its corresponding σ factor. A partner switch releases the σ factor, activating transcription of general stress response genes.

8.1.1.6. Novel Types of Anti-σ Factors (Fig. 7.1F). ECF classification efforts identified numerous other alternative modules that could function as input domains in altogether different signalling mechanisms. Some novel groups are associated with anti-σ factors that carry conserved domains of unknown function. Other ECF σ factors interact with larger proteins that span the cytoplasmic membrane with six helices, rather than the canonical single helix. Some conserved groups of ECF σ factors have long C-terminal extensions that may play a role in signal transduction, especially since these proteins are often not linked to obvious anti-σ factors (Wecke *et al.*, 2012). Yet other ECF σ factors are associated with completely unrelated proteins, such as putative serine/threonine kinases or other enzymes. These modules are conserved within individual ECF subgroups, but functional links between these novel protein pairs remain to be discovered.

8.1.2. Where Do We Stand Now-and Where Do We Go from Here?

All the experimentally characterized ECF σ factors are relatively simple signalling units that hardly go beyond single σ/anti-σ pairs mediating direct input- output connections. Presumably this trend, which also fits 2CS, will hold true for the majority of ECF σ factors because the mechanistic features of the modules built into ECF σ/anti-σ pairs restrict the potential of such units. On the other hand, these σ factors offer some options for more complex modes of signal transduction and gene regulation.

Basically, each communicating interface of the cascade can serve as a switchboard to integrate or diversify the signal(s). Two ways to reach higher signalling complexity are based on the σ/anti-σ interaction. Other paralogous anti-σ factors could divide the labour of

regulating the cellular pool of a single σ factor. Based on different input signals and binding affinities, subpopulations of the σ factor could then be gradually released. Likewise, a single anti-σ factor species could inactivate paralogous σ factors, again based on different binding affinities and the strengths of σ/anti-σ interactions.

The third general mechanism to facilitate a graded response is based on the target promoters. Depending on the degree of sequence conservation within the promoter, high-affinity promoters will be induced at lower σ factor concentrations, while less well-conserved promoters will be activated when higher amounts of σ factor molecules are released. Another way-at least partially-to overcome ECF σ/anti-σ pair signalling limitations occurs when they combine with 2CS proteins. All these mechanisms could combine to form an even more complex stress response network.

ECF σ factors are a widespread and conserved principle of bacterial signal transduction. Moreover, like 1CS and 2CS, ECF σ factors embody a highly modular design, which allows functional domains to shuffle and exchange as separate building blocks across the bacterial kingdom. In the course of evolution, archetypical 1CS, 2CS, and ECF σ/anti-σ pairs succeeded as basic signalling mechanisms in bacteria, presumably because they represent optimized domain combinations. Nevertheless, other combinations persist, as the partner- switching mechanism between 2CS and σ/anti-σ pairs illustrates. More generally, it seems that bacterial signal transduction is based on a universal pool of numerous signalling domains that perform one of three functions: stimulus perception, intra-/intercomponent communication, or cellular output. In theory, evolution or synthetic biologists may choose such domains at random, as long they are put in a functional order.

8.2. Comparative genomics of Bce-like 2CS

Bce-like 2CS form a part of detoxification modules and together with cognate ABC transporters are responsible for mediating resistance to antimicrobial peptides (Dintner *et al.*, 2011). These detoxification modules are formed by a HK, a RR, an ATPase and an associated permease, all of which are indispensable for sensing and transducing of the signal (Bernard *et al.*, 2007; Rietkötter *et al.*, 2008). Bce-like systems are widely spread among many Gram-positive bacteria, also including pathogens like *S. aureus* and *S. epidermidis* (Li *et al.*, 2007a; Li *et al.*, 2007b). In the light of antibiotic resistance becoming an important health issue, studying molecular mechanisms behind resistance determinants becomes increasingly important. Secondly, BceS-like kinases are a unique

example of minimalistic HKs. They harbour no extracytoplasmic sensory domain, only a conserved dimerisation and catalytic ATPase domains, and need an accessory protein (in this case an ABC transporter) to sense the appropriate stimulus. As such they are an interesting subject to study the molecular mechanisms responsible for the signal transduction.

A comparative genomic analysis of all Bce-like system is presented in Chapter 4 (Dintner *et al.*, 2011). This study revealed the phyletic distribution of Bce-like 2CS. The detoxification systems seem to be more prevalent in competitive environments, such as soil, and less common in more specialized habitats like the gut, milk or skin. The analysis presents also some arguments in favour of diversification of the detoxification modules through gene duplication and horizontal gene transfer. This hypothesis seems plausible in the light of the fact that there are often multiple resistance modules in one organism, but the genomic context conservation does not extend beyond the four genes of the 2CS and ABC transporter. There is also limited correlation between phylogenetic groups of detoxification modules and species distribution.

Since BceS-like HK, unlike the classical kinases such as PhyQ, do not harbour any extracytoplasmic domains, a question of how these proteins sense the stimulus has always been posed. Experimental data from many organisms (Bernard *et al.*, 2007; Hiron *et al.*, 2011; Meehl *et al.*, 2007; Ouyang *et al.*, 2010; Rietkötter *et al.*, 2008) show that the kinases require the presence of the ABC transporter for the signal transduction. Our bioinformatic analysis further strengthens this hypothesis, as it shows a tight evolutionary correlation between the permease (BceB-like) and kinase (BceS-like) components. This suggests a direct interaction between the ABC transporters and the 2CS. A preliminary confirmation of this hypothesis is presented in Chapter 7 (Bacterial Two-Hybrid Assay, see below), but the direct protein-protein interaction between the BceS-like kinase and BceB-permease has yet to be demonstrated experimentally.

Moreover, this analysis of Bce-like detoxification modules enables the identification of putative corresponding 2CS for ABC transporters, whose genes were not directly associated with a 2CS operon, as well as binding sites for RR upstream of the ABC transporter genes. These two tools provide the opportunity to predict the regulatory relationship between paralogous systems in one organism.

A good example of such a complex system can be found in *L. monocytogenes* (Fig. 7.2A). This organism harbours the AnrAB ABC transporter, which mediates resistance against a range of peptide antibiotics (Collins *et al.*, 2010). Its expression depends on the VirRS

system (Mandin *et al.*, 2005). Consistent with this fact, there is a BceR-like binding site upstream of the *anrAB* operon. This situation is reminiscent of the typical Bce-like system, even though its components are located at different loci in the genome. However, the set-up is complicated by the fact that there is the second BceAB-like transporter (Lmo1747-6) located upstream of the VirRS 2CS. It is not preceded by a BceR-like binding site, and, consistent with that, its expression is not regulated by the VirRS 2CS (Camejo *et al.*, 2009; Mandin *et al.*, 2005), at least not to the level that is detectable in a microarray experiment. In our classification, both Lmo1746 and AnrB, as well as the HK VirS belong to one group (group V). Therefore, it can be proposed that VirS interacts with both permeases Lmo1746 and AnrB. As there is a putative RR binding site upstream of *anrAB* operon, and such a binding site is missing from the *lmo1747-6* promoter region, it seems that VirS is able to interact with both Lmo14747 and AnrB, but the VirRS 2CS regulates only the expression of *anrAB* operon. It is tempting to imagine that Lmo1747-6 is responsible for sensing and AnrAB for conferring resistance against AMPs, mediated by the VirRS 2CS.

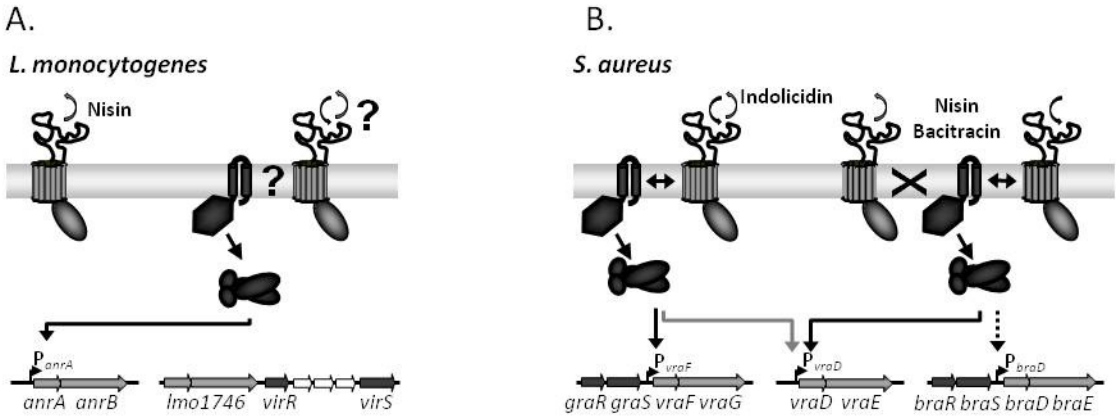

Figure 7.2.. Schematic representation of Bce-like resistance modules of *L. monocytogenes* (A) and *S. aureus* (B). Genes and proteins are marked in dark grey (2CS) and light grey (ABC transporters). Genes of unknown function are white. Inducible promoters are indicated by bent arrows. Substrate recognition by the ABC transporter is indicated by a curved downward arrow, detoxification by a curved upward arrow. Example substrates (if known) are shown above each ABC transporter. Phosphotransfer as well as strong transcriptional activation are shown by black arrows; moderate induction is indicated by a dotted arrow. Regulation of *vraDE* expression by GraR is shown by a light grey arrow, as contradictory results have been reported (see text for details).

A very interesting and complex regulatory system can be found also in *S. aureus* (Fig. 7.2B). This organism harbours two complete Bce-like systems GraRS-VraFG and BraRSAB, and an additional Pep7E-ABC transporter, VraDE. While GraRS-VraFG seems to be a stand-alone detoxification module, albeit with an extended regulon and an additional membrane protein, BraRSAB and VraDE show a surprising interplay. The ABC

transporter BraAB is responsible only for stimulus detection, and the resulting induction of the BraRS system leads only to a slight increase in its expression (Hiron *et al.*, 2011). At the same time, it mediates a strong upregulation of *vraDE* expression. VraDE then confers resistance to the inducing AMPs (nisin and bacitracin), but plays no role in signal transduction.

Other types of complex arrangements can be found in many Firmicutes species, especially among bacilli and clostridia (Dintner *et al.*, 2011). Many genomes contain more BceAB-like ABC transporters than BceRS-like 2CS, sometimes located in one, sometimes in different loci, raising the question of cross-talk and division of labour between different ABC transporters. It seems that the modular design of these detoxification systems allows a certain degree of regulatory flexibility. Such subtle differences in transcriptional control in Bce-like systems can be observed in different strains of *S. aureus*. An orphan ABC transporter, VraDE, is responsible for conferring resistance against bacitracin and nisin. Its expression is controlled solely by the BraRS 2CS in strain HG001 (Hiron *et al.*, 2011), but also by GraRS in strain MW2 (Li *et al.*, 2007a). It is tempting to imagine that a duplication and subsequent specialisation of an ABC transporter, coupled with the modulation of the transporter's expression level by changes in the respective RR-binding site gives bacteria an instrument for evolving adaption to altering environmental conditions.

8.3. Bce-like systems in *B. subtilis*: a case study

According to our classification, Bce-like systems in *B. subtilis* belong to three separate groups. All three systems, BceRS-AB, PsdRS-AB, and YxdJK-LM seem to form stand-alone detoxification modules. So far, only a low level of cross-talk has been reported for Bce and Psd systems, most probably with the HK BceS phosphorylating PsdR RR (Rietkötter *et al.*, 2008). However, this cross-talk is weak (induction of the *psdA* promoter is approximately 30 fold weaker than the *bceA* promoter) and it is not clear if it plays any physiological role.

BceRS-AB, PsdRS-AB and YxdJK-LM respond to and mediate resistance against various peptide antibiotics, similarly to other Bce-like systems. In our study presented in Chapter 5, we aimed to identify novel inducers for all three detoxification modules present in *B. subtilis*. We screened a wide variety of cell envelope-active compounds, including many peptide antibiotics. We also performed an *in silico* meta-analysis of all previously published stress response microarray data sets and searched for additional inducers of these

three systems. In this study, we identified lipid II-binding lantibiotics as the main group of inducers of P_{psdA} expression. P_{psdA} is induced by lantibiotics nisin, subtilin, actagardine and gallidermin, but also a lipid II-binding lipopeptide, enduracidin. Induction by another peptide antibiotic, bacitracin, is a result of cross-talk, as described by Rietkötter and colleagues (Rietkötter *et al.*, 2008). We also identified two lantibiotics, actagardine and mersacidin, as novel inducers of *bceAB* expression. This is an interesting finding, as BceAB was previously thought to be a bacitracin-specific resistance pump (Mascher *et al.*, 2003; Ohki *et al.*, 2003a). We were not able to identify novel inducers of P_{yxdL}, which is until now known to respond only to the human peptide LL-37 (Pietiäinen *et al.*, 2005).

Results from our study presented in Chapter 5 clearly show that the three detoxification systems are independent modules with only very slightly overlapping inducer spectra, in contrast to other organisms like *L. monocytogenes* and *S. aureus*. Interestingly, the three systems show a high level of sequence similarity in all protein components, as well as in their promoter regions. These observations led us to investigate the specificity determinants in Bce-like systems in *B. subtilis*. In chapter 6 we report the first results on the specificity determinants in the BceR and PsdR dependent promoters. Chapter 7 summarizes the results we obtained from investigating the specificity determinants governing the interactions between the kinases and RR, kinases and ABC transporters, and also specificity determinants in stimulus sensing.

8.3.1. Specificity determinants in BceR and PsdR dependent promoters

Promoters controlled by the BceR-like RR show an unusual architecture (Dintner *et al.*, 2011). Similarly to other 2CS-regulated promoters, these promoters harbour a repeat that is necessary for a RR-dependent induction. Whether this is an inverted or direct repeat cannot be distinguiseh at this point, as both repeats in the *psdA* and *bceA* promoters fulfill the criteria of either an imperfect inverted or a direct repeat. Our results clearly show that the core of this repeat is necessary for the induction of the promoter. However, the characteristic feature of the *bceA*-like promoters is the presence of additional half sites, both upstream and downstream of the main repeat. Results presented in Chapter 6 suggest that these half sites play an important role in regulating the expression of *bceA*-like promoters in *B. subtilis*. The upstream half site seems to influence the strength of the promoters, whereas the downstream half site may be responsible for promoter specificity. Based on the first results from the promoter induction assays is conceivable that the the *bceA* and *psdA* promoters are subject to a homocooperative activation, where each of the

recognition sites binds the RR with a different affinity. While this hypothesis will have to be validated by independent *in vitro* and *in vivo* experiments, this study is a good starting point providing the first insights into the specificity determination in *bceA*-like promoters.

8.3.1.1. Flexibility ("fuzziness") of the main binding site. The mutagenesis studies of the main binding site show that although the main binding site is indispensable for induction, its sequence can be modified to some extent. Our experiments clearly demonstrate that the exact composition of the binding sites does not play an important role in promoter specificity, as long as the overall consensus is not changed. However, it influences the strength of the response.

Motif "fuzziness", or differences to the optimal consensus sequence can be observed for binding sites of many transcription factors. So called "fuzzy" binding sites differ at various position from the perfect palindrome, and also have a lower affinity towards the transcription factor then the perfect palindrome would have (Francke *et al.*, 2008). The fuzziness of the binding site may be a consequence of a balance between mutation and selection, since the action of the transcription factor may be insensitive to subtle changes in binding affinity, provided they are above a certain threshold (van Hijum *et al.*, 2009). Our results suggest that this may be true for BceR and PsdR. "Fuzziness" of the transcription factor binding sites was also suggested to play a role in cooperative transcription activation and repression ((Hermsen *et al.*, 2006), see below). It is also possible that the binding sites can show a lower level of conservation (and higher level of "fuzziness") when there are multiple binding sites located in a direct vicinity, as this will cause the local concentration of the transcription factor to be higher than normal (van Hijum *et al.*, 2009).

8.3.1.2. The decisive influence of the downstream half site – homocooperative activation? The first results obtained from the analysis of chimeric promoters show that the fragment located between the main binding site and the translation start site is the one that determines the promoter specificity. Its exchange between *psdA* and *bceA* promoters leads to either full or partial exchange of specificity. This fragment harbours the downstream half site which is most probably one of the specificity determinants in *bceA*-like promoters. However, attempts to further narrow down the region that determines the specificity were as yet unsuccessful. In the context of the chimeric promoter, the upstream half site seems to influence the strength of induction. The similarity of the -35 promoter to

the recognition sequences is intriguing, but can be purely accidental and has to be addressed experimentally.

Taken together, our results show that (i) the full repeat is necessary for the induction, but (ii) some substitutions in its sequence that do not change the specificity of the promoter are possible ("fuzziness"). Moreover, (iii) the upstream half site plays a role mainly in modulating the strength of induction, visible only under certain conditions. The most important finding is that (iv) the specificity determinant is located between the main repeat and the translation start site, and possibly involves the downstream half site. Therefore, it seems that these two half sites and the main repeat all play an important, although different roles in the promoter activation.

It is conceivable that these two promoters are subject to a homocooperative activation. In homocooperative activation multiple binding sites can be identified in the promoter region, each binding the transcription factor with different affinity. The sites which bind the transcription factor molecule interacting directly with the RNA polymerase are called primary sites. In case of transcriptional activators, these sites are usually located next to the -35 promoter element (Hermsen *et al.*, 2006). The remaining sites are termed secondary sites. The latter are expected to be more conserved and bind the transcription factor with higher affinity than the primary site in order to maximise the steepness of response to the transcription factor concentration. An interesting scenario is conceivable where the downstream half site (possibly together with the region annotated as the -35 promoter element?) constitutes the primary site, binding the RR with lower affinity, but higher specificity. The main binding site and the upstream half site would play the accessory role and bind the RR with higher affinity. This hypothesis can be supported by the fact that for the *bceA* promoter a binding of the RR to the region which includes the main repeat as well as the upstream half site was demonstrated with DNase I footprinting assay (Ohki *et al.*, 2003a). The first gel retardation assays with *psdA* promoter and PsdR RR show different molecular weight complexes, indicative of binding of the RR to more than one recognition site (data not shown). However, until now no cooperative transcription activation was shown for these two promoters and this hypothesis requires further studies.

However intriguing the results discussed above are, it is important to bear in mind that the genetic analysis of regulatory sequences has its limitation, and can rather suggest than prove a model. It has to be also kept in mind that the sequence surrounding the binding site can also have an effect on both the binding affinity and the half-life of RR binding. The higher the affinity of the neighbouring sequence to the RR, the longer the RR will take to

diffuse along the DNA towards its binding site, and the shorter it will stay bound to it (van Hijum *et al.*, 2009). This effect has been demonstrated for the CcpA protein of *B. subtilis*, which is able to bind with higher affinity to *cre* boxes located in an AT-rich context that to those located in GC-rich context (Zalieckas *et al.*, 1998). Some regulatory regions – and it cannot be ruled out that this is the case for P_{bceA} and P_{psdA} – are complex and their action involves many factors, some with small effects. Therefore, it will be necessary to test the hypotheses using independent biochemical methods.

8.3.2. Specificity determinants in BceRS-AB and PsdRS-AB 2CS

In Chapter 7, we present a complementary study where we investigate the specificity determinants in BceR-like RR. We set out to identify the residues which interact with the important promoter elements and lead to the observed insulation of the three signalling pathways. In the random mutagenesis screen we identified 12 independent mutations in BceR and one in YxdJ. These mutations are located in the receiver and DNA-binding domains, as well as in the flexible linker between them. These results provide a perfect starting point for further studies and functional analyses of the isolated mutants. We also established a random screen that can be easily modified and allows for searching of various other gain-of-function mutations, e.g. in the *bceA*-like promoters.

In Chapter 7 we also present the first results on stimulus perception in BceRS-like 2CS and their cognate ABC transporters. The question of stimulus perception in BceRS-AB-like systems is an interesting one due to a specific architecture of BceS-like HK. These kinases belong to a family of proteins that has been termed IM-HK, since they lack an obvious input domain and are thought to sense the stimulus in the membrane interface (Mascher, 2006). Homologous HK in in *S. aureus* and *S. epidermidis* were suggested to sense the inducing peptides via the short extracytoplasmic loop of the kinase (Li *et al.*, 2007a; Li *et al.*, 2007b). In our study we demonstrate that this is not the case for the BceS-like HK in *B. subtilis*. We propose that the stimulus (i.e. peptide antibiotic molecule) is sensed by the ABC transporter, which in turn interacts with the HK. This hypothesis is supported by the initial results from bacterial two-hybrid assays in *E. coli*, which suggest that BceS HK and the BceB permease, as well as PsdS HK and PsdB permease are able to interact. Our findings are well in agreement with the results presented in Chapter 4, where we showed the cooccurrence and coevolution of BceB-like transporters and the BceRS-like 2CS.

8.4. Conclusions

The aim of this thesis was to investigate the common principles governing subgroups of signalling proteins, the ECF σ factors, and the Bce-like 2CS, by employing both bioinformatic and experimental methods. The starting point of the studies of both ECF σ factors and Bce-like 2CS was a comparative genomic analysis of the steadily growing wealth of information obtained from sequenced bacterial genomes. Both the classification of ECF σ factors (Chapter 2) and the analysis of Bce-like systems in Firmicutes bacteria (Chapter 4) were followed by in-depth studies. In case of ECF σ factors we focused on one family of these signalling proteins and broadened the analysis to the ECF015-associated proteins (Chapter 3). The analysis of Bce-like systems gave rise to further studies on these systems in *B. subtilis*, including their in-depth characterisation (Chapter 5), as well as the first investigation of the molecular specificity determinants in these systems (Chapters 6 and 7). All these studies demonstrate the great potential of combining the *in silico* methods with classical laboratory experiments in elucidating novel and complex mechanisms in bacterial signal transduction.

Supplementary material

Supplementary material provided on CD

The following supplementary material can be found on the enclosed CD:

Chapter 2

Table S2.1. Statistics/numbers of the generated dataset.

Table S2.2-S2.4. Excel tables of the complete ECF □ factor dataset used in this study, split up into: major groups (S2.2), minor groups (S2.3) and unclassified (S2.4)

Table S2.5-S2.7. Excel tables of the complete anti-□ factor dataset used in this study, split up into: major groups (S2.5), minor groups (S2.6) and unclassified (S2.7)

Table S2.8. List of intergenic regions chosen for promoter analysis and generation of the Weblogos shown in Fig. 2.6.

Table S2.9. Group-specific HMM, including cut-off values + overlap.

Table S2.10. Term definitions of protein domains.

Figure S2.1. Multiple sequence alignment of conserved ECF groups (A) Region σ_2, (B) region σ_4.

Chapter 4

Figure S4.1. Multiple-sequence alignments of permease ECDs

Table S4.1. Dataset for phylogenetic analysis of BceRS-BceAB-like resistance module.

Table S4.2. Dataset of YycG- and OppB homologs from 26 genomes containing BceRS-BceAB-like module.

Table S4.3. Identification of putative response regulator binding sites in promoters of BceAB-like transport operons.

References

Abachin, E., Poyart, C., Pellegrini, E., Milohanic, E., Fiedler, F., Berche, P., and Trieu-Cuot, P. (2002) Formation of D-alanyl-lipoteichoic acid is required for adhesion and virulence of *Listeria monocytogenes*. *Mol Microbiol* **43**: 1-14.

Ades, S.E. (2004) Control of the alternative sigma factor s^E in *Escherichia coli*. *Curr Opin Microbiol* **7**: 157-162.

Ades, S.E. (2008) Regulation by destruction: design of the s^E envelope stress response. *Curr Opin Microbiol* **11**: 535-540.

Alba, B.M., and Gross, C.A. (2004) Regulation of the *Escherichia coli* sigma-dependent envelope stress response. *Mol Microbiol* **52**: 613-619.

Alm, E.J., Huang, K.H., Price, M.N., Koche, R.P., Keller, K., Dubchak, I.L., and Arkin, A.P. (2005) The MicrobesOnline web site for comparative genomics. *Genome Res* **15**: 1015-1022.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., 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.

Alvarez-Martinez, C.E., Baldini, R.L., and Gomes, S.L. (2006) A *Caulobacter crescentus* extracytoplasmic function sigma factor mediating the response to oxidative stress in stationary phase. *J Bacteriol* **188**: 1835-1846.

Alvarez-Martinez, C.E., Lourenco, R.F., Baldini, R.L., Laub, M.T., and Gomes, S.L. (2007) The ECF sigma factor sigma(T) is involved in osmotic and oxidative stress responses in *Caulobacter crescentus*. *Mol Microbiol* **66**: 1240-1255.

Aravind, L., and Ponting, C.P. (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem Sci* **22**: 458-459.

Arnaud, M., Chastanet, A., and Debarbouille, M. (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl Environ Microbiol* **70**: 6887-6891.

Arthur, M., Depardieu, F., Molinas, C., Reynolds, P., and Courvalin, P. (1995) The vanZ gene of Tn1546 from Enterococcus faecium BM4147 confers resistance to teicoplanin. *Gene* **154**: 87-92.

Ashby, M.K., and Houmard, J. (2006) Cyanobacterial two-component proteins: structure, diversity, distribution, and evolution. *Microbiol. Mol. Biol. Rev.* **70**: 472-509.

Ashenberg, O., Rozen-Gagnon, K., Laub, M.T., and Keating, A.E. (2011) Determinants of homodimerization specificity in histidine kinases. *J Mol Biol* **413**: 222-235.

Bader, M.W., Sanowar, S., Daley, M.E., Schneider, A.R., Cho, U., Xu, W., Klevit, R.E., Le Moual, H., and Miller, S.I. (2005) Recognition of antimicrobial peptides by a bacterial sensor kinase. *Cell* **122**: 461-472.

Bailey, T.L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**: 28-36.

Baldauf, S.L. (2003) Phylogeny for the faint of heart: a tutorial. *Trends Genet* **19**: 345-351.

Bandow, J.E., Brotz, H., Leichert, L.I., Labischinski, H., and Hecker, M. (2003) Proteomic approach to understanding antibiotic action. *Antimicrob Agents Chemother* **47**: 948-955.

Bandow, J.E., and Hecker, M. (2007) Proteomic profiling of cellular stresses in *Bacillus subtilis* reveals cellular networks and assists in elucidating antibiotic mechanisms of action. *Prog Drug Res* **64**: 79, 81-101.

Banerjee, S., and Hansen, J.N. (1988) Structure and expression of a gene encoding the precursor of subtilin, a small protein antibiotic. *J Biol Chem* **263**: 9508-9514.

Barnard, A., Wolfe, A., and Busby, S. (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr Opin Microbiol* **7**: 102-108.

Bastiat, B., Sauviac, L., and Bruand, C. (2010) Dual control of *Sinorhizobium meliloti* RpoE2 sigma factor activity by two PhyR-type two-component response regulators. *J Bacteriol* **192**: 2255-2265.

Baumgartner, J.W., Kim, C., Brissette, R.E., Inouye, M., Park, C., and Hazelbauer, G.L. (1994) Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ. *J Bacteriol* **176**: 1157-1163.

Becker, P., Hakenbeck, R., and Henrich, B. (2009) An ABC transporter of Streptococcus pneumoniae involved in susceptibility to vancoresmycin and bacitracin. *Antimicrob Agents Chemother* **53**: 2034-2041.

Bernard, R., El Ghachi, M., Mengin-Lecreulx, D., Chippaux, M., and Denizot, F. (2005) BcrC from *Bacillus subtilis* acts as an undecaprenyl pyrophosphate phosphatase in bacitracin resistance. *J. Biol. Chem.* **280**: 28852-28857.

Bernard, R., Guiseppi, A., Chippaux, M., Foglino, M., and Denizot, F. (2007) Resistance to bacitracin in *Bacillus subtilis*: Unexpected requirement of the BceAB ABC

transporter in the control of expression of its own structural genes. *J Bacteriol* **189**: 8636-8642.

Bierbaum, G., and Sahl, H.G. (2009) Lantibiotics: mode of action, biosynthesis and bioengineering. *Curr Pharm Biotechnol* **10**: 2-18.

Biondi, E.G., Skerker, J.M., Arif, M., Prasol, M.S., Perchuk, B.S., and Laub, M.T. (2006) A phosphorelay system controls stalk biogenesis during cell cycle progression in *Caulobacter crescentus*. *Mol Microbiol* **59**: 386-401.

Biville, F., Cwerman, H., Letoffe, S., Rossi, M.S., Drouet, V., Ghigo, J.M., and Wandersman, C. (2004) Haemophore-mediated signalling in *Serratia marcescens*: a new mode of regulation for an extra cytoplasmic function (ECF) sigma factor involved in haem acquisition. *Mol Microbiol* **53**: 1267-1277.

Blake, K.L., Randall, C.P., and O'Neill, A.J. (2011) In vitro studies indicate a high resistance potential for the lantibiotic nisin in Staphylococcus aureus and define a genetic basis for nisin resistance. *Antimicrob Agents Chemother* **55**: 2362-2368.

Blanco, A.G., Sola, M., Gomis-Ruth, F.X., and Coll, M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* **10**: 701-713.

Blatch, G.L., and Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**: 932-939.

Böttiger, T., Schneider, T., Martinez, B., Sahl, H.G., and Wiedemann, I. (2009) Influence of Ca^{2+} ions on the activity of lantibiotics containing a mersacidin-like lipid II binding motif. *Appl Env Microbiol* **75**: 4427-4434.

Bourret, R.B. (2010) Receiver domain structure and function in response regulator proteins. *Curr Opin Microbiol* **13**: 142-149.

Boyd, D.A., Cvitkovitch, D.G., Bleiweis, A.S., Kiriukhin, M.Y., Debabov, D.V., Neuhaus, F.C., and Hamilton, I.R. (2000) Defects in D-alanyl-lipoteichoic acid synthesis in *Streptococcus mutans* results in acid sensitivity. *J Bacteriol* **182**: 6055-6065.

Braun, V., Mahren, S., and Ogierman, M. (2003) Regulation of the FecI-type ECF sigma factor by transmembrane signalling. *Curr Opin Microbiol* **6**: 173-180.

Braun, V., and Mahren, S. (2005) Transmembrane transcriptional control (surface signalling) of the *Escherichia coli* Fec type. *FEMS Microbiol Rev* **29**: 673-684.

Braun, V., Mahren, S., and Sauter, A. (2006) Gene regulation by transmembrane signaling. *Biometals* **19**: 103-113.

Breukink, E., and de Kruijff, B. (2006) Lipid II as a target for antibiotics. *Nat Rev Drug Discov* **5**: 321-332.

Brooks, B.E., and Buchanan, S.K. (2008) Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. *Biochim Biophys Acta* **1778**: 1930-1945.

Brown, K.L., and Hughes, K.T. (1995) The role of anti-sigma factors in gene regulation. *Mol Microbiol* **16**: 397-404.

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

Burger, M., Woods, R.G., McCarthy, C., and Beacham, I.R. (2000) Temperature regulation of protease in *Pseudomonas fluorescens* LS107d2 by an ECF sigma factor and a transmembrane activator. *Microbiology* **146** 3149-3155.

Burkard, M., and Stein, T. (2008) Microtiter plate bioassay to monitor the interference of antibiotics with the lipid II cycle essential for peptidoglycan biosynthesis. *J Microbiol Methods* **75**: 70-74.

Butcher, B.G., Mascher, T., and Helmann, J.D. (2008) Environmental sensing and the role of extracytoplasmic function (ECF) sigma factors. In *Bacterial Physiology - a Molecular Approach*. El-Sharoud, W.M. (ed). Berlin, Heidelberg: Springer-Verlag GmbH, pp. 233-261.

Calamita, H., Ko, C., Tyagi, S., Yoshimatsu, T., Morrison, N.E., and Bishai, W.R. (2005) The *Mycobacterium tuberculosis* SigD sigma factor controls the expression of ribosome-associated gene products in stationary phase and is required for full virulence. *Cell Microbiol* **7**: 233-244.

Camejo, A., Buchrieser, C., Couve, E., Carvalho, F., Reis, O., Ferreira, P., Sousa, S., Cossart, P., and Cabanes, D. (2009) In vivo transcriptional profiling of Listeria monocytogenes and mutagenesis identify new virulence factors involved in infection. *PLoS Pathog* **5**: e1000449.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002) Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. *Molecular Cell* **9**: 527-539.

Campbell, E.A., and Darst, S.A. (2005) Regulation of bacterial transcription by anti-s factors. In *Structural Biology of Bacterial Pathogenesis*. Waksman, G., Caparon, M. and Hultgren, S. (eds). Washington, DC: American Society for Microbiology Press, pp. 1-15.

Campbell, E.A., Greenwell, R., Anthony, J.R., Wang, S., Lim, L., Das, K., Sofia, H.J., Donohue, T.J., and Darst, S.A. (2007) A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria. *Molecular Cell* **27**: 793-805.

Campbell, E.A., Westblade, L.F., and Darst, S.A. (2008) Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. *Curr Opin Microbiol* **11**: 121-127.

Cao, M., and Helmann, J.D. (2002) Regulation of the *Bacillus subtilis bcrC* bacitracin resistance gene by two extracytoplasmic function s factors. *J Bacteriol* **184**: 6123-6129.

Cao, M., Wang, T., Ye, R., and Helmann, J.D. (2002) Antibiotics that inhibit cell wall biosynthesis induce expression of the *Bacillus subtilis* s^W and s^M regulons. *Mol Microbiol* **45**: 1267-1276.

Cao, M., Salzberg, L., Tsai, C.S., Mascher, T., Bonilla, C., Wang, T., Ye, R.W., Marquez-Magana, L., and Helmann, J.D. (2003) Regulation of the *Bacillus subtilis* extracytoplasmic function protein s^Y and its target promoters. *J Bacteriol* **185**: 4883-4890.

Capra, E.J., Perchuk, B.S., Lubin, E.A., Ashenberg, O., Skerker, J.M., and Laub, M.T. (2010) Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. *PLoS genetics* **6**: e1001220.

Casino, P., Rubio, V., and Marina, A. (2009) Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* **139**: 325-336.

Charlet, D., Mostowy, S., Alexander, D., Sit, L., Wiker, H.G., and Behr, M.A. (2005) Reduced expression of antigenic proteins MPB70 and MPB83 in *Mycobacterium bovis* BCG strains due to a start codon mutation in *sigK*. *Mol Microbiol* **56**: 1302-1313.

Cheung, G.Y., Rigby, K., Wang, R., Queck, S.Y., Braughton, K.R., Whitney, A.R., Teintze, M., DeLeo, F.R., and Otto, M. (2010) Staphylococcus epidermidis strategies to avoid killing by human neutrophils. *PLoS Pathog* **6**.

Chevenet, F., Brun, C., Banuls, A.L., Jacq, B., and Christen, R. (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**: 439.

Cole, C., Barber, J.D., and Barton, G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36**: W197-W201.

Collins, B., Curtis, N., Cotter, P.D., Hill, C., and Ross, R.P. (2010) The ABC transporter AnrAB contributes to the innate resistance of *Listeria monocytogenes* to nisin, bacitracin, and various b-lactam antibiotics. *Antimicrob Agents Chemother* **54**: 4416-4423.

Cotter, P.D., Hill, C., and Ross, R.P. (2005) Bacteriocins: developing innate immunity for food. *Nat Rev Micro* **3**: 777-788.

Coumes-Florens, S., Brochier-Armanet, C., Guiseppi, A., Denizot, F., and Foglino, M. (2011) A new highly conserved antibiotic sensing/resistance pathway in Firmicutes involves an ABC transporter interplaying with a signal transduction system. *PLoS ONE* **6**: e15951.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-1190.

Cudic, P., Kranz, J.K., Behenna, D.C., Kruger, R.G., Tadesse, H., Wand, A.J., Veklich, Y.I., Weisel, J.W., and McCafferty, D.G. (2002) Complexation of peptidoglycan intermediates by the lipoglycodepsipeptide antibiotic ramoplanin: minimal structural requirements for intermolecular complexation and fibril formation. *Proc Natl Acad Sci U S A* **99**: 7384-7389.

Cui, L., Lian, J.-Q., Neoh, H.-m., Reyes, E., and Hiramatsu, K. (2005) DNA microarray-based identification of genes associated with glycopeptide resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **49**: 3404-3413.

D'Costa, V.M., McGrann, K.M., Hughes, D.W., and Wright, G.D. (2006) Sampling the antibiotic resistome. *Science* **311**: 374-377.

de Been, M., Bart, M.J., Abee, T., Siezen, R.J., and Francke, C. (2008) The identification of response regulator-specific binding sites reveals new roles of two-component systems in *Bacillus cereus* and closely related low-GC Gram-positives. *Environ Microbiol* **10**: 2796-2809.

Dehal, P.S., Joachimiak, M.P., Price, M.N., Bates, J.T., Baumohl, J.K., Chivian, D., Friedland, G.D., Huang, K.H., Keller, K., Novichkov, P.S., Dubchak, I.L., Alm, E.J., and Arkin, A.P. (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic acids research* **38**: D396-400.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., and Gascuel, O. (2008)

Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**: W465-469.

Derre, I., Rapoport, G., and Msadek, T. (2000) The CtsR regulator of stress response is active as a dimer and specifically degraded in vivo at 37°C. *Mol Microbiol* **38**: 335-347.

Detmers, F.J., Lanfermeijer, F.C., and Poolman, B. (2001) Peptides and ATP binding cassette peptide transporters. *Res Microbiol* **152**: 245-258.

Dintner, S., Staron, A., Berchtold, E., Petri, T., Mascher, T., and Gebhard, S. (2011) Co-evolution of ABC-transporters and two-component regulatory systems as resistance modules against antimicrobial peptides in Firmicutes bacteria. *J. Bacteriol.* **193**: 3851-3862.

Donadio, S., and Sosio, M. (2008) Biosynthesis of glycopeptides: prospects for improved antibacterials. *Curr Top Med Chem* **8**: 654-666.

Drake, E.J., Cao, J., Qu, J., Shah, M.B., Straubinger, R.M., and Gulick, A.M. (2007) The 1.8 A crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*. *J Biol Chem* **282**: 20425-20434.

Draper, L.A., Ross, R.P., Hill, C., and Cotter, P.D. (2008) Lantibiotic immunity. *Curr Protein Pept Sci* **9**: 39-49.

Dubrac, S., Bisicchia, P., Devine, K.M., and Msadek, T. (2008) A matter of life and death: cell wall homeostasis and the WalKR (YycGF) essential signal transduction pathway. *Mol Microbiol* **70**: 1307-1322.

Dufour, Y.S., Landick, R., and Donohue, T.J. (2008) Organization and evolution of the biological response to singlet oxygen stress. *J Mol Biol* **383**: 713-730.

Duque, E., Rodriguez-Herva, J.J., de la Torre, J., Dominguez-Cuevas, P., Munoz-Rojas, J., and Ramos, J.L. (2007) The RpoT regulon of *Pseudomonas putida* DOT-T1E and its role in stress endurance against solvents. *J Bacteriol* **189**: 207-219.

Dutta, R., Qin, L., and Inouye, M. (1999) Histidine kinases: diversity of domain organization. *Mol Microbiol* **34**: 633-640.

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 s^H. *Mol Microbiol* **52**: 285-302.

Enz, S., Brand, H., Orellana, C., Mahren, S., and Braun, V. (2003) Sites of interaction between the FecA and FecR signal transduction proteins of ferric citrate transport in *Escherichia coli* K-12. *J Bacteriol* **185**: 3745-3752.

Ernst, C.M., and Peschel, A. (2011) Broad-spectrum antimicrobial peptide resistance by MprF-mediated aminoacylation and flipping of phospholipids. *Mol Microbiol* **80**: 290-299.

Felsenstein, J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**: 164-166.

Ferreira, A.O., Myers, C.R., Gordon, J.S., Martin, G.B., Vencato, M., Collmer, A., Wehling, M.D., Alfano, J.R., Moreno-Hagelsieb, G., Lamboy, W.F., DeClerck, G., Schneider, D.J., and Cartinhour, S.W. (2006) Whole-genome expression profiling defines the HrpL regulon of *Pseudomonas syringae* pv. *tomato* DC3000, allows de novo reconstruction of the Hrp *cis* clement, and identifies novel coregulated genes. *Mol Plant Microbe Interact* **19**: 1167-1179.

Fischer, H.P., and Freiberg, C. (2007) Applications of transcriptional profiling in antibiotics discovery and development. *Prog Drug Res* **64**: 21, 23-47.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., and et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**: 496-512.

Francez-Charlot, A., Frunzke, J., Reichen, C., Ebneter, J.Z., Gourion, B., and Vorholt, J.A. (2009) Sigma factor mimicry involved in regulation of general stress response. *Proc Natl Acad Sci U S A* **106**: 3467-3472.

Francez-Charlot, A., Frunzke, J., and Vorholt, J.A. (2010) The general stress response in alphaproteobacteria. In *Bacterial Stress Responses*. Storz, G. and Hengge, R. (eds). Washington D.C.: ASM press, pp. in press.

Francke, C., Kerkhoven, R., Wels, M., and Siezen, R.J. (2008) A generic approach to identify Transcription Factor-specific operator motifs; Inferences for LacI-family mediated regulation in *Lactobacillus plantarum* WCFS1. *BMC Genomics* **9**: 145.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A., 3rd, and

Venter, J.C. (1995) The minimal gene complement of Mycoplasma genitalium. *Science* **270**: 397-403.

Galperin, M.Y. (2005) A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol* **5**: 35.

Galperin, M.Y. (2006) Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol* **188**: 4169-4182.

Galperin, M.Y. (2010) Diversity of structure and function of response regulator output domains. *Curr Opin Microbiol* **13**: 150-159.

Gao, R., Mack, T.R., and Stock, A.M. (2007) Bacterial response regulators: versatile regulatory strategies from common domains. *Trends Biochem Sci* **32**: 225-234.

Gao, R., Tao, Y., and Stock, A.M. (2008) System-level mapping of *Escherichia coli* response regulator dimerization with FRET hybrids. *Mol Microbiol* **69**: 1358-1372.

Gao, R., and Stock, A.M. (2009) Biological insights from structures of two-component proteins. *Annu Rev Microbiol* **63**: 133-154.

Gascuel, O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* **14**: 685-695.

Gebhard, S., Gaballa, A., Helmann, J.D., and Cook, G.M. (2009) Direct stimulus perception and transcription activation by a membrane-bound DNA binding protein. *Mol Microbiol* **73**: 482-491.

Gebhard, S., and Mascher, T. (2011) Antimicrobial peptide sensing and detoxification modules: unraveling the regulatory circuitry of *Staphylococcus aureus*. *Mol Microbiol* **81**: 581-587.

Giyanto, Kobayashi, K., and Ogasawara, N. (2003) The *Bacillus subtilis* YvcRS transporter is induced by a peptide antibiotic through the action of the YvcPQ two component regulatory system. In *Functional Genomics of Gram-positive Microorganisms* Baveno, Italy.

Glover, R.T., Kriakov, J., Garforth, S.J., Baughn, A.D., and Jacobs, W.R., Jr. (2007) The two-component regulatory system senX3-regX3 regulates phosphate-dependent gene expression in Mycobacterium smegmatis. *J Bacteriol* **189**: 5495-5503.

Goh, C.S., Bogan, A.A., Joachimiak, M., Walther, D., and Cohen, F.E. (2000) Co-evolution of proteins with their interaction partners. *J Mol Biol* **299**: 283-293.

Gordon, N.D., Ottaviano, G.L., Connell, S.E., Tobkin, G.V., Son, C.H., Shterental, S., and Gehring, A.M. (2008) Secreted-protein response to sigmaU activity in *Streptomyces coelicolor*. *J Bacteriol* **190**: 894-904.

Gourion, B., Rossignol, M., and Vorholt, J.A. (2006) A proteomic study of *Methylobacterium extorquens* reveals a response regulator essential for epiphytic growth. *PNAS* **103**: 13186-13191.

Gourion, B., Francez-Charlot, A., and Vorholt, J.A. (2008) PhyR is involved in the general stress response of Methylobacterium extorquens AM1. *J Bacteriol* **190**: 1027-1035.

Gourion, B., Sulser, S., Frunzke, J., Francez-Charlot, A., Stiefel, P., Pessi, G., Vorholt, J.A., and Fischer, H.-M. (2009) The PhyR-s^{EcfG} signalling cascade is involved in stress response and symbiotic efficiency in *Bradyrhizobium japonicum*. *Mol Microbiol* **73**: 291-305.

Grebe, T.W., and Stock, J.B. (1999) The histidine protein kinase superfamily. *Adv Microb Physiol* **41**: 139-227.

Green, J., Scott, C., and Guest, J.R. (2001) Functional versatility in the CRP-FNR superfamily of transcription factors: FNR and FLP. *Adv Microb Physiol* **44**: 1-34.

Greenwell, R., Nam, T.-W., and Donohue, T.J. (2011) Features of *Rhodobacter sphaeroides* ChrR Required for Stimuli to Promote the Dissociation of σ^E/ChrR Complexes. *J Mol Biol* **407**: 477-491.

Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol* **63**: 141-155.

Grosse, C., Friedrich, S., and Nies, D.H. (2007) Contribution of extracytoplasmic function sigma factors to transition metal homeostasis in *Cupriavidus metallidurans* strain CH34. *J Mol Microbiol Biotechnol* **12**: 227-240.

Gruber, T.M., and Gross, C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**: 441-466.

Guder, A., Wiedemann, I., and Sahl, H.G. (2000) Posttranslationally modified bacteriocins--the lantibiotics. *Biopolymers* **55**: 62-73.

Guerout-Fleury, A.M., Shazand, K., Frandsen, N., and Stragier, P. (1995) Antibiotic-resistance cassettes for *Bacillus subtilis*. *Gene* **167**: 335-336.

Guerout-Fleury, A.M., Frandsen, N., and Stragier, P. (1996) Plasmids for ectopic integration in *Bacillus subtilis*. *Gene* **180**: 57-61.

Gunesekere, I.C., Kahler, C.M., Ryan, C.S., Snyder, L.A., Saunders, N.J., Rood, J.I., and Davies, J.K. (2006) Ecf, an alternative sigma factor from *Neisseria gonorrhoeae*, controls expression of *msrAB*, which encodes methionine sulfoxide reductase. *J Bacteriol* **188**: 3463-3469.

Hahn, M.Y., Raman, S., Anaya, M., and Husson, R.N. (2005) The *Mycobacterium tuberculosis* extracytoplasmic-function sigma factor SigL regulates polyketide synthases and secreted or membrane proteins and is required for virulence. *J Bacteriol* **187**: 7062-7071.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**: 95-98.

Hammami, R., Zouhir, A., Lay, C.L., Ben Hamida, J., and Fliss, I. (2010) BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiol* **10**: 22.

Hancock, R.E., and Chapple, D.S. (1999) Peptide antibiotics. *Antimicrob Agents Chemother* **43**: 1317-1323.

Hancock, R.E., and Sahl, H.G. (2006) Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat Biotechnol* **24**: 1551-1557.

Harwood, C.R., and Cutting, S.M. (1990) *Molecular Biological Methods for Bacillus.* Chichester: John Wiley & Sons.

Hecker, M., Pane-Farre, J., and Volker, U. (2007) SigB-dependent general stress response in *Bacillus subtilis* and related Gram-positive bacteria. *Annu Rev Microbiol* **61**: 215-236.

Heinrich, J., and Wiegert, T. (2006) YpdC determines site-1 degradation in regulated intramembrane proteolysis of the RsiW anti-sigma factor of *Bacillus subtilis*. *Mol Microbiol* **62**: 566-579.

Helmann, J.D., and Chamberlin, M.J. (1988) Structure and function of bacterial sigma factors. *Annu Rev Biochem* **57**: 839-872.

Helmann, J.D. (1995) Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res* **23**: 2351-2360.

Helmann, J.D. (1999) Anti-sigma factors. *Curr Opin Microbiol* **2**: 135-141.

Helmann, J.D. (2002) The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* **46**: 47-110.

Helmann, J.D. (2006) Deciphering a complex genetic regulatory network: the *Bacillus subtilis* s^W protein and intrinsic resistance to antimicrobial compounds. *Sci Prog* **89**: 243-266.

Hengge, R. (2008) The two-component network and the general stress sigma factor RpoS (sigma S) in *Escherichia coli*. *Adv Exp Med Biol* **631**: 40-53.

Hengge, R. (2009a) Principles of c-di-GMP signalling in bacteria. *Nat Rev Micro* **7**: 263-273.

Hengge, R. (2009b) Proteolysis of sigmaS (RpoS) and the general stress response in *Escherichia coli*. *Res Microbiol* **160**: 667-676.

Hengge, R. (2010) The general stress response in Gram-negative bacteria. In *Bacterial Stress Responses*. Storz, G. and Hengge, R. (eds). Washington, D.C.: ASM press, pp. in press.

Herbert, S., Bera, A., Nerz, C., Kraus, D., Peschel, A., Goerke, C., Meehl, M., Cheung, A., and Götz, F. (2007) Molecular basis of resistance to muramidase and cationic antimicrobial peptide activity of lysozyme in staphylococci. *PLoS Pathog* **3**: e102.

Hermsen, R., Tans, S., and ten Wolde, P.R. (2006) Transcriptional regulation by competing transcription factor modules. *PLoS Computational Biology* **2**: e164.

Herrou, J., Foreman, R., Fiebig, A., and Crosson, S. (2010) A structural model of anti-anti-s inhibition by a two-component receiver domain: the PhyR stress response regulator. *Mol. Microbiol.* **78**: 290-304.

Hiron, A., Falord, M., Valle, J., Debarbouille, M., and Msadek, T. (2011) Bacitracin and nisin resistance in *Staphylococcus aureus*: a novel pathway involving the BraS/BraR two-component system (SA2417/SA2418) and both the BraD/BraE and VraD/VraE ABC transporters. *Molecular microbiology*.

Hoch, J.A., and Silhavy, T.J. (1995) *Two-component signal transduction.* Washington, D.C.: ASM press.

Homerova, D., Halgasova, L., and Kormanec, J. (2008) Cascade of extracytoplasmic function sigma factors in *Mycobacterium tuberculosis*: identification of a s^J-dependent promoter upstream of *sigI*. *FEMS Microbiol Lett* **280**: 120-126.

Hong, H.J., Paget, M.S., and Buttner, M.J. (2002) A signal transduction system in Streptomyces coelicolor that activates the expression of a putative cell wall glycan operon in response to vancomycin and other cell wall-specific antibiotics. *Mol Microbiol* **44**: 1199-1211.

Hu, Y., Kendall, S., Stoker, N.G., and Coates, A.R. (2004) The *Mycobacterium tuberculosis sigJ* gene controls sensitivity of the bacterium to hydrogen peroxide. *FEMS Microbiol Lett* **237**: 415-423.

Hutchings, M.I., Hong, H.-J., Leibovitz, E., Sutcliffe, I.C., and Buttner, M.J. (2006) The s^E cell envelope stress response of *Streptomyces coelicolor* is influenced by a novel lipoprotein, CseA. *J. Bacteriol.* **188**: 7222-7229.

Hutter, B., Fischer, C., Jacobi, A., Schaab, C., and Loferer, H. (2004a) Panel of *Bacillus subtilis* reporter strains indicative of various modes of action. *Antimicrob. Agents Chemother.* **48**: 2588-2594.

Hutter, B., Schaab, C., Albrecht, S., Borgmann, M., Brunner, N.A., Freiberg, C., Ziegelbauer, K., Rock, C.O., Ivanov, I., and Loferer, H. (2004b) Prediction of mechanisms of action of antibacterial compounds by gene expression profiling. *Antimicrob. Agents Chemother.* **48**: 2838-2844.

Hyyryläinen, H.-L., Pietiäinen, M., Lunden, T., Ekman, A., Gardemeister, M., Murtomaki-Repo, S., Antelmann, H., Hecker, M., Valmu, L., Sarvas, M., and Kontinen, V.P. (2007) The density of negative charge in the cell wall influences two-component signal transduction in *Bacillus subtilis*. *Microbiology* **153**: 2126-2136.

Inouye, M., and Dutta, R. (2003) *Histidine Kinases in Signal Transduction.* San Diego, CA: Academic Press.

Jenssen, H., Hamill, P., and Hancock, R.E. (2006) Peptide antimicrobial agents. *Clin Microbiol Rev* **19**: 491-511.

Johnson, B.A., Anker, H., and Meleney, F.L. (1945) Bacitracin: a new antibiotic produced by a member of the *B. subtilis* group. *Science* **102**: 376-377.

Jordan, S., Junker, A., Helmann, J.D., and Mascher, T. (2006) Regulation of LiaRS-dependent gene expression in *Bacillus subtilis*: Identification of inhibitor proteins, regulator binding sites and target genes of a conserved cell envelope stress-sensing two-component system. *J Bacteriol* **188**: 5153-5166.

Jordan, S., Hutchings, M.I., and Mascher, T. (2008) Cell envelope stress response in Gram-positive bacteria. *FEMS Microbiol Rev* **32**: 107-146.

Joseph, P., Fichant, G., Quentin, Y., and Denizot, F. (2002) Regulatory relationship of two-component and ABC transport systems and clustering of their genes in the *Bacillus/Clostridium* group, suggest a functional link between them. *J Mol Microbiol Biotechnol* **4**: 503-513.

Joseph, P., Guiseppi, A., Sorokin, A., and Denizot, F. (2004) Characterization of the *Bacillus subtilis* YxdJ response regulator as the inducer of expression for the cognate ABC transporter YxdLM. *Microbiology* **150**: 2609-2617.

Kappler, U., Friedrich, C.G., Truper, H.G., and Dahl, C. (2001) Evidence for two pathways of thiosulfate oxidation in *Starkeya novella* (formerly *Thiobacillus novellus*). *Arch Microbiol* **175**: 102-111.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* **95**: 5752-5756.

Karimova, G., Ullmann, A., and Ladant, D. (2000) A bacterial two-hybrid system that exploits a cAMP signaling cascade in *Escherichia coli*. *Methods Enzymol* **328**: 59-73.

Karls, R.K., Guarner, J., McMurray, D.N., Birkness, K.A., and Quinn, F.D. (2006) Examination of *Mycobacterium tuberculosis* sigma factor mutants using low-dose aerosol infection of guinea pigs suggests a role for SigC in pathogenesis. *Microbiology* **152**: 1591-1600.

Karniol, B., and Vierstra, R.D. (2004) The HWE histidine kinases, a new family of bacterial two-component sensor kinases with potentially diverse roles in environmental signaling. *J. Bacteriol.* **186**: 445-453.

Killian, J.A. (1992) Gramicidin and gramicidin-lipid interactions. *Biochim Biophys Acta* **1113**: 391-425.

Kim, D., and Forst, S. (2001) Genomic analysis of the histidine kinase family in bacteria and archaea. *Microbiology* **147**: 1197-1212.

Kim, T.H., Kim, H.J., Park, J.S., Kim, Y., Kim, P., and Lee, H.S. (2005) Functional analysis of *sigH* expression in *Corynebacterium glutamicum*. *Biochem Biophys Res Commun* **331**: 1542-1547.

Kitagawa, N., Shiota, S., Shibata, Y., Takeshita, T., and Yamashita, Y. (2011) Characterization of MbrC involved in bacitracin resistance in Streptococcus mutans. *FEMS Microbiol Lett* **318**: 61-67.

Koonin, E.V., and Wolf, Y.I. (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36**: 6688-6719.

Kramer, N.E., van Hijum, S.A.F.T., Knol, J., Kok, J., and Kuipers, O.P. (2006) Transcriptome analysis reveals mechanisms by which *Lactococcus lactis* acquires nisin resistance. *Antimicrob. Agents Chemother.* **50**: 1753-1761.

Kraus, D., Herbert, S., Kristian, S.A., Khosravi, A., Nizet, V., Götz, F., and Peschel, A. (2008) The GraRS regulatory system controls *Staphylococcus aureus* susceptibility to antimicrobial host defenses. *BMC Microbiol* **8**: 85.

Krukonis, E.S., Yu, R.R., and Dirita, V.J. (2000) The Vibrio cholerae ToxR/TcpP/ToxT virulence cascade: distinct roles for two membrane-localized transcriptional activators on a single promoter. *Mol Microbiol* **38**: 67-84.

Kruszewska, D., Sahl, H.G., Bierbaum, G., Pag, U., Hynes, S.O., and Ljungh, A. (2004) Mersacidin eradicates methicillin-resistant Staphylococcus aureus (MRSA) in a mouse rhinitis model. *J Antimicrob Chemother* **54**: 648-653.

Landman, D., Georgescu, C., Martin, D.A., and Quale, J. (2008) Polymyxins revisited. *Clin Microbiol Rev* **21**: 449-465.

Lane, W.J., and Darst, S.A. (2006) The structural basis for promoter -35 element recognition by the group IV sigma factors. *PLoS Biol* **4**: e269.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Larsen, R., Kok, J., and Kuipers, O.P. (2005) Interaction between ArgR and AhrC controls regulation of arginine metabolism in Lactococcus lactis. *J Biol Chem* **280**: 19319-19330.

Laub, M.T., and Goulian, M. (2007) Specificity in two-component signal transduction pathways. *Annu Rev Genet* **41**: 121-145.

Lautru, S., Oves-Costales, D., Pernodet, J.L., and Challis, G.L. (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology* **153**: 1405-1412.

Lavin, J.L., Kiil, K., Resano, O., Ussery, D.W., and Oguiza, J.A. (2007) Comparative genomic analysis of two-component regulatory proteins in Pseudomonas syringae. *BMC Genomics* **8**: 397.

Leoni, L., Orsi, N., de Lorenzo, V., and Visca, P. (2000) Functional analysis of PvdS, an iron starvation sigma factor of *Pseudomonas aeruginosa*. *J Bacteriol* **182**: 1481-1491.

Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J., and Bork, P. (2006) SMART 5: domains in the context of genomes and networks. *Nucl. Acids Res.* **34**: D257-260.

Letunic, I., Doerks, T., and Bork, P. (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res* **37**: D229-232.

Li, M., Cha, D.J., Lai, Y., Villaruz, A.E., Sturdevant, D.E., and Otto, M. (2007a) The antimicrobial peptide-sensing system *aps* of *Staphylococcus aureus*. *Mol Microbiol* **66**: 1136-1147.

Li, M., Lai, Y., Villaruz, A.E., Cha, D.J., Sturdevant, D.E., and Otto, M. (2007b) Gram-positive three-component antimicrobial peptide-sensing system. *Proc Natl Acad Sci U S A* **104**: 9469-9474.

Li, W., Stevenson, C.E., Burton, N., Jakimowicz, P., Paget, M.S., Buttner, M.J., Lawson, D.M., and Kleanthous, C. (2002) Identification and structure of the anti-sigma factor-binding domain of the disulphide-stress regulated sigma factor s^R from *Streptomyces coelicolor*. *J Mol Biol* **323**: 225-236.

Lindeberg, M., Cartinhour, S., Myers, C.R., Schechter, L.M., Schneider, D.J., and Collmer, A. (2006) Closing the circle on the discovery of genes encoding Hrp regulon members and type III secretion system effectors in the genomes of three model *Pseudomonas syringae* strains. *Mol Plant Microbe Interact* **19**: 1151-1158.

Lonetto, M.A., Brown, K.L., Rudd, K.E., and Buttner, M.J. (1994) Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A* **91**: 7573-7577.

Majchrzykiewicz, J.A., Kuipers, O.P., and Bijlsma, J.J.E. (2010) Generic and specific adaptive responses of *Streptococcus pneumoniae* to challenge with three distinct antimicrobial peptides, bacitracin, LL-37, and nisin. *Antimicrob Agents Chemother* **54**: 440-451.

Mandin, P., Fsihi, H., Dussurget, O., Vergassola, M., Milohanic, E., Toledo-Arana, A., Lasa, I., Johansson, J., and Cossart, P. (2005) VirR, a response regulator critical for *Listeria monocytogenes* virulence. *Mol Microbiol* **57**: 1367-1380.

Martinez-Salazar, J.M., Salazar, E., Encarnacion, S., Ramirez-Romero, M.A., and Rivera, J. (2009) Role of the extracytoplasmic function sigma factor RpoE4 in oxidative and osmotic stress responses in *Rhizobium etli*. *J Bacteriol* **191**: 4122-4132.

Mascher, T., Margulis, N.G., Wang, T., Ye, R.W., and Helmann, J.D. (2003) Cell wall stress responses in *Bacillus subtilis*: the regulatory network of the bacitracin stimulon. *Mol Microbiol* **50**: 1591-1604.

Mascher, T., Zimmer, S.L., Smith, T.A., and Helmann, J.D. (2004) Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of *Bacillus subtilis*. *Antimicrob Agents Chemother* **48**: 2888-2896.

Mascher, T. (2006) Intramembrane-sensing histidine kinases: a new family of cell envelope stress sensors in Firmicutes bacteria. *FEMS Microbiol Lett* **264**: 133-144.

Mascher, T., Helmann, J.D., and Unden, G. (2006) Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev* **90**: 910-938.

McCafferty, D.G., Cudic, P., Frankel, B.A., Barkallah, S., Kruger, R.G., and Li, W. (2002) Chemistry and biology of the ramoplanin family of peptide antibiotics. *Biopolymers* **66**: 261-284.

Meehl, M., Herbert, S., Götz, F., and Cheung, A. (2007) Interaction of the GraRS two-component system with the VraFG ABC transporter to support vancomycin-intermediate resistance in *Staphylococcus aureus*. *Antimicrob Agents Chemother* **51**: 2679-2689.

Merighi, M., Majerczak, D.R., Stover, E.H., and Coplin, D.L. (2003) The HrpX/HrpY two-component system activates *hrpS* expression, the first step in the regulatory cascade controlling the Hrp regulon in *Pantoea stewartii* subsp. *stewartii*. *Mol Plant Microbe Interact* **16**: 238-248.

Miller, J.H. (1972) *Experiments in molecular genetics.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Minchin, S.D., and Busby, S.J. (2009) Analysis of mechanisms of activation and repression at bacterial promoters. *Methods* **47**: 6-12.

Ming, L.J. (2003) Structure and function of "metalloantibiotics". *Med Res Rev* **23**: 697-762.

Moore, C.M., Gaballa, A., Hui, M., Ye, R.W., and Helmann, J.D. (2005) Genetic and physiological responses of *Bacillus subtilis* to metal ion stress. *Molecular Microbiology* **57**: 27-40.

Münch, R., Hiller, K., Grote, A., Scheer, M., Klein, J., Schobert, M., and Jahn, D. (2005) Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics* **21**: 4187-4189.

Murakami, S., and Yamaguchi, A. (2003) Multidrug-exporting secondary transporters. *Curr Opin Struct Biol* **13**: 443-452.

Nakao, A., Imai, S., and Takano, T. (2000) Transposon-mediated insertional mutagenesis of the D-alanyl-lipoteichoic acid (dlt) operon raises methicillin resistance in Staphylococcus aureus. *Res Microbiol* **151**: 823-829.

Nakunst, D., Larisch, C., Huser, A.T., Tauch, A., Puhler, A., and Kalinowski, J. (2007) The extracytoplasmic function-type sigma factor SigM of *Corynebacterium glutamicum* ATCC 13032 is involved in transcription of disulfide stress-related genes. *J Bacteriol* **189**: 4696-4707.

Neuhaus, F.C., and Baddiley, J. (2003) A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in Gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **67**: 686-723.

Newton, G.L., and Fahey, R.C. (2008) Regulation of mycothiol metabolism by sigma(R) and the thiol redox sensor anti-sigma factor RsrA. *Mol Microbiol* **68**: 805-809.

Nishi, H., Komatsuzawa, H., Fujiwara, T., McCallum, N., and Sugai, M. (2004) Reduced content of lysyl-phosphatidylglycerol in the cytoplasmic membrane affects susceptibility to moenomycin, as well as vancomycin, gentamicin, and antimicrobial peptides, in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **48**: 4800-4807.

Nizan-Koren, R., Manulis, S., Mor, H., Iraki, N.M., and Barash, I. (2003) The regulatory cascade that activates the Hrp regulon in *Erwinia herbicola* pv. *gypsophilae*. *Mol Plant Microbe Interact* **16**: 249-260.

Ohki, R., Giyanto, Tateno, K., Masuyama, W., Moriya, S., Kobayashi, K., and Ogasawara, N. (2003a) The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in *Bacillus subtilis*. *Mol Microbiol* **49**: 1135-1144.

Ohki, R., Tateno, K., Okada, Y., Okajima, H., Asai, K., Sadaie, Y., Murata, M., and Aiso, T. (2003b) A bacitracin-resistant *Bacillus subtilis* gene encodes a homologue of the membrane-spanning subunit of the *Bacillus licheniformis* ABC transporter. *J. Bacteriol.* **185**: 51-59.

Ouyang, J., Tian, X.-L., Versey, J., Wishart, A., and Li, Y.-H. (2010) The BceABRS four-component system regulates the bacitracin-induced cell envelope stress response in *Streptococcus mutans*. *Antimicrob Agents Chemother* **54**: 3895-3906.

Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr, Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., Bhattacharyya, A., Burd, H., Gardner, W., Hanke, P., Kapatral, V., Mikhailova, N., Vasieva, O., Osterman, A., Vonstein, V., Fonstein, M., Ivanova, N., and Kyrpides, N. (2003) The ERGO^{TM} genome analysis and discovery system. *Nucl. Acids. Res.* **31**: 164-171.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye,

Y., Zagnitko, O., and Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691-5702.

Paget, M.S., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. (1999a) Evidence that the extracytoplasmic function sigma factor s^E is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* **181**: 204-211.

Paget, M.S., Leibovitz, E., and Buttner, M.J. (1999b) A putative two-component signal transduction system regulates s^E, a sigma factor required for normal cell wall integrity in *Streptomyces coelicolor* A3(2). *Mol Microbiol* **33**: 97-107.

Paget, M.S., Molle, V., Cohen, G., Aharonowitz, Y., and Buttner, M.J. (2001) Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the s^R regulon. *Mol Microbiol* **42**: 1007-1020.

Park, J.-H., and Roe, J.-H. (2008) Mycothiol regulates and is regulated by a thiol-specific antisigma factor RsrA and s^R in *Streptomyces coelicolor*. *Mol Microbiol* **68**: 861-870.

Paterson, G.K., Blue, C.E., and Mitchell, T.J. (2006) Role of two-component systems in the virulence of *Streptococcus pneumoniae*. *J Med Microbiol* **55 (Pt4)**: 355-336.

Pazos, F., and Valencia, A. (2008) Protein co-evolution, co-adaptation and interactions. *Embo J* **27**: 2648-2655.

Perraud, A.-L., Kimmel, B., Weiss, V., and Gross, R. (1998) Specificity of the BvgAS and EvgAS phosphorelay is mediated by the C-terminal HPt domains of the sensor proteins. *Mol Microbiol* **27**: 875-887.

Peschel, A., Otto, M., Jack, R.W., Kalbacher, H., Jung, G., and Götz, F. (1999) Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. *J. Biol. Chem.* **274**: 8405-8410.

Peschel, A., Jack, R.W., Otto, M., Collins, L.V., Staubitz, P., Nicholson, G., Kalbacher, H., Nieuwenhuizen, W.F., Jung, G., Tarkowski, A., van Kessel, K.P., and van Strijp, J.A. (2001) Staphylococcus aureus resistance to human defensins and evasion of neutrophil killing via the novel virulence factor MprF is based on modification of membrane lipids with l-lysine. *J Exp Med* **193**: 1067-1076.

Peschel, A. (2002) How do bacteria resist human antimicrobial peptides? *Trends Microbiol* **10**: 179-186.

Pietiäinen, M., Gardemeister, M., Mecklin, M., Leskela, S., Sarvas, M., and Kontinen, V.P. (2005) Cationic antimicrobial peptides elicit a complex stress response in *Bacillus subtilis* that involves ECF-type sigma factors and two-component signal transduction systems. *Microbiology* **151**: 1577-1592.

Pietiäinen, M., Francois, P., Hyyryläinen, H.-L., Tangomo, M., Sass, V., Sahl, H.-G., Schrenzel, J., and Kontinen, V.P. (2009) Transcriptome analysis of the responses of *Staphylococcus aureus* to antimicrobial peptides and characterization of the roles of *vraDE* and *vraSR* in antimicrobial resistance. *BMC Genomics* **10**: 429.

Piggot, P.J., and Hilbert, D.W. (2004) Sporulation of *Bacillus subtilis*. *Curr Opin Microbiol* **7**: 579-586.

Ponting, C.P., and Aravind, L. (1997) PAS: a multifunctional domain family comes to light. *Curr Biol* **7**: R674-677.

Price, C.W. (2010) General stress response in *Bacillus subtilis* and related Gram-positive bacteria. In *Bacterial Stress Responses*. Storz, G. and Hengge, R. (eds). Washington D.C.: ASM press, pp. in press.

Rasmussen, S., Nielsen, H.B., and Jarmer, H. (2009) The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol Microbiol* **73**: 1043-1057.

Reeves, A.R., Wang, G.R., and Salyers, A.A. (1997) Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. *J Bacteriol* **179**: 643-649.

Renzoni, A., Barras, C., Francois, P., Charbonnier, Y., Huggler, E., Garzoni, C., Kelley, W.L., Majcherczyk, P., Schrenzel, J., Lew, D.P., and Vaudaux, P. (2006) Transcriptomic and functional analysis of an autolysis-deficient, teicoplanin-resistant derivative of methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **50**: 3048-3061.

Rhodius, V.A., Suh, W.C., Nonaka, G., West, J., and Gross, C.A. (2006) Conserved and variable functions of the s^E stress response in related genomes. *PLoS Biol* **4**: e2.

Rietkötter, E., Hoyer, D., and Mascher, T. (2008) Bacitracin sensing in *Bacillus subtilis*. *Mol Microbiol* **68**: 768-785.

Ross, W., Ernst, A., and Gourse, R.L. (2001) Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev* **15**: 491-506.

Rukmana, A., Morimoto, T., Takahashi, H., Giyanto, and Ogasawara, N. (2009) Assessment of transcriptional responses of *Bacillus subtilis* cells to the antibiotic

enduracidin, which interferes with cell wall synthesis, using a high-density tiling chip. *Genes Genet Syst* **84**: 253-267.

Saier, M.H., Jr., Yen, M.R., Noto, K., Tamang, D.G., and Elkan, C. (2009) The Transporter Classification Database: recent advances. *Nucleic Acids Res* **37**: D274-278.

Samant, S., Hsu, F.F., Neyfakh, A.A., and Lee, H. (2009) The Bacillus anthracis protein MprF is required for synthesis of lysylphosphatidylglycerols and for resistance to cationic antimicrobial peptides. *J Bacteriol* **191**: 1311-1319.

Sambrook, J., and Russell, D.W. (2001) *Molecular Cloning - a laboratory manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sass, P., and Bierbaum, G. (2009) Native graS mutation supports the susceptibility of Staphylococcus aureus strain SG511 to antimicrobial peptides. *Int J Med Microbiol* **299**: 313-322.

Sass, V., Pag, U., Tossi, A., Bierbaum, G., and Sahl, H.G. (2008) Mode of action of human beta-defensin 3 against *Staphylococcus aureus* and transcriptional analysis of responses to defensin challenge. *Int J Med Microbiol* **298**: 619-633.

Sauviac, L., Philippe, H., Phok, K., and Bruand, C. (2007) An extracytoplasmic function sigma factor acts as a general stress response regulator in *Sinorhizobium meliloti*. *J Bacteriol* **189**: 4204-4216.

Schmitt, P., Wilmes, M., Pugniere, M., Aumelas, A., Bachere, E., Sahl, H.G., Schneider, T., and Destoumieux-Garzon, D. (2010) Insight into invertebrate defensin mechanism of action: oyster defensins inhibit peptidoglycan biosynthesis by binding to lipid II. *J Biol Chem*.

Schneider, T., Kruse, T., Wimmer, R., Wiedemann, I., Sass, V., Pag, U., Jansen, A., Nielsen, A.K., Mygind, P.H., Raventos, D.S., Neve, S., Ravn, B., Bonvin, A.M.J.J., De Maria, L., Andersen, A.S., Gammelgaard, L.K., Sahl, H.-G., and Kristensen, H.-H. (2010) Plectasin, a fungal defensin, targets the bacterial cell wall precursor lipid II. *Science* **328**: 1168-1172.

Schneider, T., and Sahl, H.-G. (2010) An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway. *International Journal of Medical Microbiology* **300**: 161-169.

Schöbel, S., Zellmeier, S., Schumann, W., and Wiegert, T. (2004) The *Bacillus subtilis* σ^W anti-sigma factor RsiW is degraded by intramembrane proteolysis through YluC. *Mol Microbiol* **52**: 1091-1105.

Schug, A., Weigt, M., Onuchic, J.N., Hwa, T., and Szurmant, H. (2009) High-resolution protein complexes from integrating genomic information with molecular simulation. *Proc Natl Acad Sci U S A* **106**: 22124-22129.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**: 5857-5864.

Schwartz, M.A., and Madhani, H.D. (2004) Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. *Annu Rev Genet* **38**: 725-748.

Sexton, R., Gill, P.R., Jr., Dowling, D.N., and O'Gara, F. (1996) Transcriptional regulation of the iron-responsive sigma factor gene *pbrA*. *Mol Gen Genet* **250**: 50-58.

Shanks, R.M., Meehl, M.A., Brothers, K.M., Martinez, R.M., Donegan, N.P., Graber, M.L., Cheung, A.L., and O'Toole, G.A. (2008) Genetic evidence for an alternative citrate-dependent biofilm formation pathway in Staphylococcus aureus that is dependent on fibronectin binding proteins and the GraRS two-component regulatory system. *Infect Immun* **76**: 2469-2477.

Shi, X., Wegener-Feldbrugge, S., Huntley, S., Hamann, N., Hedderich, R., and Sogaard-Andersen, L. (2008) Bioinformatics and experimental analysis of proteins of two-component systems in Myxococcus xanthus. *J Bacteriol* **190**: 613-624.

Skerker, J.M., Prasol, M.S., Perchuk, B.S., Biondi, E.G., and Laub, M.T. (2005) Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol* **3**: e334.

Skerker, J.M., Perchuk, B.S., Siryaporn, A., Lubin, E.A., Ashenberg, O., Goulian, M., and Laub, M.T. (2008) Rewiring the specificity of two-component signal transduction systems. *Cell* **133**: 1043-1054.

Smith, J.G., Latiolais, J.A., Guanga, G.P., Pennington, J.D., Silversmith, R.E., and Bourret, R.B. (2004) A search for amino acid substitutions that universally activate response regulators. *Mol Microbiol* **51**: 887-901.

Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175-182.

Staroń, A., Sofia, H.J., Dietrich, S., Ulrich, L.E., Liesegang, H., and Mascher, T. (2009) The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) s factor protein family. *Mol Microbiol* **74**: 557-581.

Staroń, A., and Mascher, T. (2010a) General stress response in α-proteobacteria: PhyR and beyond. *Mol Microbiol* **78**: 271-277.

Staroń, A., and Mascher, T. (2010b) Extracytoplasmic function σ factors come of age. *Microbe* **5**: 164-170.

Staroń, A., Finkeisen, D.E., and Mascher, T. (2011) Peptide antibiotic sensing and detoxification modules of *Bacillus subtilis*. *Antimicrob Agents Chemother* **55**: 515-525.

Steen, A., Palumbo, E., Deghorain, M., Cocconcelli, P.S., Delcour, J., Kuipers, O.P., Kok, J., Buist, G., and Hols, P. (2005) Autolysis of *Lactococcus lactis* is increased upon D-alanine depletion of peptidoglycan and lipoteichoic acids. *J. Bacteriol.* **187**: 114-124.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochem* **69**: 183-215.

Stock, J.B., Surette, M.G., Levit, M., and Park, P. (1995) Two-component signal transduction systems: structure-function relationships and mechanisms of catalysis. In *Two-Component Signal Transduction*. Hoch, J.A. and Silhavy, T.J. (eds). Washington, D.C.: ASM press, pp. 25-52.

Stone, K.J., and Strominger, J.L. (1971) Mechanism of action of bacitracin: complexation with metal ion and C 55 -isoprenyl pyrophosphate. *Proc Natl Acad Sci U S A* **68**: 3223-3227.

Stülke, J., Martin-Verstraete, I., Zagorec, M., Rose, M., Klier, A., and Rapoport, G. (1997) Induction of the *Bacillus subtilis ptsGHI* operon by glucose is controlled by a novel antiterminator, GlcT. *Mol Microbiol* **25**: 65-78.

Swoboda, J.G., Campbell, J., Meredith, T.C., and Walker, S. (2010) Wall teichoic acid function, biosynthesis, and inhibition. *Chembiochem* **11**: 35-45.

Szekat, C., Jack, R.W., Skutlarek, D., Farber, H., and Bierbaum, G. (2003) Construction of an expression system for site-directed mutagenesis of the lantibiotic mersacidin. *Appl Environ Microbiol* **69**: 3777-3783.

Szurmant, H., Fukushima, T., and Hoch, J.A. (2007a) The essential YycFG two-component system of *Bacillus subtilis*. *Methods Enzymol* **422**: 396-417.

Szurmant, H., Mohan, M.A., Imus, P.M., and Hoch, J.A. (2007b) YycH and YycI interact to regulate the essential YycFG two-component system in *Bacillus subtilis*. *J Bacteriol* **189**: 3280-3289.

Szurmant, H., and Hoch, J.A. (2010) Interaction fidelity in two-component signaling. *Curr Opin Microbiol* **13**: 190-197.

Tetsch, L., and Jung, K. (2009) The regulatory interplay between membrane-integrated sensors and transport proteins in bacteria. *Mol Microbiol* **73**: 982-991.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.

Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **Chapter 2**: Unit 2 3.

Travers, A.A. (1987) Structure and function of E. coli promoter DNA. *CRC Crit Rev Biochem* **22**: 181-219.

Tsuda, H., Yamashita, Y., Shibata, Y., Nakano, Y., and Koga, T. (2002) Genes involved in bacitracin resistance in *Streptococcus mutans*. *Antimicrob. Agents Chemother.* **46**: 3756-3764.

Turner, J., Cho, Y., Dinh, N.-N., Waring, A.J., and Lehrer, R.I. (1998) Activities of LL-37, a Cathelin-Associated Antimicrobial Peptide of Human Neutrophils. *Antimicrob. Agents Chemother.* **42**: 2206-2214.

Ubersax, J.A., and Ferrell, J.E., Jr. (2007) Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* **8**: 530-541.

Ulrich, L.E., Koonin, E.V., and Zhulin, I.B. (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol* **13**: 52-56.

Ulrich, L.E., and Zhulin, I.B. (2007) MiST: a microbial signal transduction database. *Nucleic Acids Res* **35**: D386-390.

Ulrich, L.E., and Zhulin, I.B. (2010) The MiST2 database: a comprehensive genomics resource on microbial signal transduction. *Nucleic Acids Res* **38**: D401-407.

Upton, M., Tagg, J.R., Wescombe, P., and Jenkinson, H.F. (2001) Intra- and interspecies signaling between Streptococcus salivarius and Streptococcus pyogenes mediated by SalA and SalA1 lantibiotic peptides. *J Bacteriol* **183**: 3931-3938.

Urban, A., Eckermann, S., Fast, B., Metzger, S., Gehling, M., Ziegelbauer, K., Rübsamen-Waigmann, H., and Freiberg, C. (2007) Novel whole-cell antibiotic biosensors for compound discovery. *Appl Environ Microbiol* **73**: 6436-6443.

Utsumi, R., Brissette, R.E., Rampersaud, A., Forst, S.A., Oosawa, K., and Inouye, M. (1989) Activation of bacterial porin gene expression by a chimeric signal transducer in response to aspartate. *Science* **245**: 1246-1249.

van Hijum, S.A., Medema, M.H., and Kuipers, O.P. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol Mol Biol Rev* **73**: 481-509, Table of Contents.

Van Hove, B., Staudenmaier, H., and Braun, V. (1990) Novel two-component transmembrane transcription control: regulation of iron dicitrate transport in *Escherichia coli* K-12. *J Bacteriol* **172**: 6749-6758.

Veith, B., Herzberg, C., Steckel, S., Feesche, J., Maurer, K.H., Ehrenreich, P., Baumer, S., Henne, A., Liesegang, H., Merkl, R., Ehrenreich, A., and Gottschalk, G. (2004) The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential. *J Mol Microbiol Biotechnol* **7**: 204-211.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., and Smith, H.O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66-74.

Veyrier, F., Said-Salim, B., and Behr, M.A. (2008) Evolution of the mycobacterial SigK regulon. *J Bacteriol* **190**: 1891-1899.

Wach, A. (1996) PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast* **12**: 259-265.

Wadhams, G.H., and Armitage, J.P. (2004) Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* **5**: 1024-1037.

Ward, M.J., Lew, H., Treuner-Lange, A., and Zusman, D.R. (1998) Regulation of motility behavior in *Myxococcus xanthus* may require an extracytoplasmic-function sigma factor. *J Bacteriol* **180**: 5668-5675.

Wecke, T., Veith, B., Ehrenreich, A., and Mascher, T. (2006) Cell envelope stress response in *Bacillus licheniformis*: Integrating comparative genomics, transcriptional profiling, and regulon mining to decipher a complex regulatory network. *J. Bacteriol.* **188**: 7500-7511.

Wecke, T., Zühlke, D., Mäder, U., Jordan, S., Voigt, B., Pelzer, S., Labischinski, H., Homuth, G., Hecker, M., and Mascher, T. (2009) Daptomycin versus friulimicin B:

In-depth profiling of *Bacillus subtilis* cell envelope stress responses. *Antimicrob Agents Chemother* **53**: 1619-1623.

Wecke, T., Halang, P., Staroń, A., Dufour, Y.S., Donohue, T.J., and Mascher, T. (2012) Extracytoplasmic function σ factors of the widely distributed group ECF41 contain a fused regulatory domain. *MicrobiologyOpen* **1**: 194-213.

Wei, Z.M., and Beer, S.V. (1995) *hrpL* activates *Erwinia amylovora hrp* gene transcription and is a member of the ECF subfamily of sigma factors. *J Bacteriol* **177**: 6201-6210.

Whittaker, C.A., and Hynes, R.O. (2002) Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell* **13**: 3369-3387.

Wick, L.M., and Egli, T. (2004) Molecular components of physiological stress responses in Escherichia coli. *Adv Biochem Eng Biotechnol* **89**: 1-45.

Willey, J.M., and van der Donk, W.A. (2007) Lantibiotics: peptides of diverse structure and function. *Annu Rev Microbiol* **61**: 477-501.

Williams, P., Winzer, K., Chan, W.C., and Cámara, M. (2007) Look who's talking: communication and quorum sensing in the bacterial world. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**: 1119-1134.

Wilson, C.J., Zhan, H., Swint-Kruse, L., and Matthews, K.S. (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci* **64**: 3-16.

Wösten, M.M.S.M., and Groisman, E.A. (1999) Molecular characterization of the PmrA regulon. *J. Biol. Chem.* **274**: 27185-27190.

Wu, Q.L., Kong, D., Lam, K., and Husson, R.N. (1997) A mycobacterial extracytoplasmic function sigma factor involved in survival following stress. *J Bacteriol* **179**: 2922-2929.

Wuichet, K., Cantwell, B.J., and Zhulin, I.B. (2010) Evolution and phyletic distribution of two-component signal transduction systems. *Curr Opin Microbiol* **13**: 219-225.

Xu, J., Chiang, H.C., Bjursell, M.K., and Gordon, J.I. (2004) Message from a human gut symbiont: sensitivity is a prerequisite for sharing. *Trends Microbiol* **12**: 21-28.

Yeaman, M.R., and Yount, N.Y. (2007) Unifying themes in host defence effector polypeptides. *Nat Rev Microbiol* **5**: 727-740.

Yildiz, F.H., and Visick, K.L. (2009) Vibrio biofilms: so much the same yet so different. *Trends Microbiol* **17**: 109-118.

Yoshida, Y., Matsuo, M., Oogai, Y., Kato, F., Nakamura, N., Sugai, M., and Komatsuzawa, H. (2011) Bacitracin sensing and resistance in Staphylococcus aureus. *FEMS Microbiol Lett* **320**: 33-39.

Zalieckas, J.M., Wray, L.V., Jr., and Fisher, S.H. (1998) Expression of the Bacillus subtilis acsA gene: position and sequence context affect cre-mediated carbon catabolite repression. *J Bacteriol* **180**: 6649-6654.

Zapf, J., Sen, U., Madhusudan, Hoch, J.A., and Varughese, K.I. (2000) A transient interaction between two phosphorelay proteins trapped in a crystal lattice reveals the mechanism of molecular recognition and phosphotransfer in signal transduction. *Structure* **8**: 851-862.

Zdanowski, K., Doughty, P., Jakimowicz, P., O'Hara, L., Buttner, M.J., Paget, M.S.B., and Kleanthous, C. (2006) Assignment of the zinc ligands in RsrA, a redox-sensing ZAS protein from *Streptomyces coelicolor*. *Biochemistry* **45**: 8294-8300.

Zellmeier, S., Hofmann, C., Thomas, S., Wiegert, T., and Schumann, W. (2005) Identification of s^V-dependent genes of *Bacillus subtilis*. *FEMS Microbiol Lett* **253**: 221-229.

Zellmeier, S., Schumann, W., and Wiegert, T. (2006) Involvement of Clp protease activity in modulating the *Bacillus subtilis* s^W stress response. *Molecular Microbiology* **61**: 1569-1582.

Zhang, W., and Shi, L. (2005) Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. *Microbiology* **151**: 2159-2173.

Zhang, Z., and Hendrickson, W.A. (2010) Structural characterization of the predominant family of histidine kinase sensor domains. *J Mol Biol* **400**: 335-353.

Acknowledgements

At this point, I would like to say a big thank you to everyone who helped me during the last four years and made this thesis possible.

First of all, I would like to thank Thorsten for his enthusiasm, encouragement, and all the good advice he always had for me. Thank you for your trust to give me a free hand to explore my interests and the guidance along the way from Göttingen, over Karlsruhe to Munich.

At this point I would also like to thank the members of my thesis committee for their time and input.

Another big thank you goes to all current and former members of the Mascher Lab! You created the great atmosphere in the lab and helped me stay sane and happy also in the "real world". It would not be the same without the many breaks we spent together over the best-coffee-on-the-campus and delicious cakes that gave us the strength necessary for the long hours of lab meetings. I will always remember our hiking trips in the Alps (thunderstorms included), sneaking into the tents at Oktoberfest, kidnapping of a certain pink bicycle, and many, many more events that made this time a lot of fun. Ainhoa, Diana, Karen, Schorsch, Sebastian, Susanne, Tina – thank you! I would also like to thank the students I had the pleasure to supervise – Jessi, Petra, Simon and Steffi. Thank you for your valuable input into this thesis, as well as the great time in the lab.

On more private note, I would like to thank my parents for their constant support and encouragement. I have a certain suspicion that giving me a pipette as a child toy somewhow got me started on this scientific adventure (though probably it was not the purpose). Finally, I would like to thank Péter for always being there for me, despite the geographical distance. Thank you for travelling probably many times around the globe, and for constantly surprising me with a new and fresh perspective.