Two-component signal transduction in *Bacillus cereus* and closely related bacteria

Mark W.H.J. de Been

Thesis committee

Thesis supervisors

Prof. Dr. T. Abee Personal chair at the Laboratory of Food Microbiology Wageningen University

Prof. Dr. R.J. Siezen Professor of Bioinformatics of Microorganisms, Centre for Molecular and Biomolecular Informatics, NCMLS UMC St. Radboud, Nijmegen

Thesis co-supervisor

Dr. C. Francke Senior scientist, TI Food and Nutrition Centre for Molecular and Biomolecular Informatics, NCMLS UMC St. Radboud, Nijmegen

Other members

Prof. Dr. M. Kleerebezem, Wageningen University Prof. Dr. K.J. Hellingwerf, University of Amsterdam Prof. Dr. M.A. Huynen, UMC St. Radboud, Nijmegen Prof. Dr. M. Ehling-Schulz, University of Veterinary Medicine Vienna, Austria

This research was conducted under the auspices of the Graduate School VLAG.

Two-component signal transduction in *Bacillus cereus* and closely related bacteria

Mark W.H.J. de Been

Thesis submitted in partial fulfilment of the requirements for the degree of doctor at Wageningen University by the authority of the Rector Magnificus Prof. Dr. M.J. Kropff, in the presence of the Thesis Committee appointed by the Doctorate Board to be defended in public on Monday 14 September 2009 at 1.30 PM in the Aula

Mark W.H.J. de Been Two-component signal transduction in *Bacillus cereus* and closely related bacteria, 176 pages

Thesis. Wageningen University, Wageningen, NL (2009) With references, with summaries in Dutch and English

ISBN 978-90-8585-435-7

Table of Contents

	Abstract	7
Chapter 1	Introduction and outline of the thesis	9
Chapter 2	Comparative analysis of two-component signal transduction systems of <i>Bacillus cereus, Bacillus thuringiensis</i> and <i>Bacillus anthracis</i>	25
Chapter 3	The identification of response regulator specific binding sites reveals new roles of two-component systems in <i>Bacillus cereus</i> and closely related low-GC Gram-positives	47
Chapter 4	Integrating phylogenetic footprinting approaches with transcriptome profiling reveals new roles for two <i>Bacillus cereus</i> two-component signal transduction systems	69
Chapter 5	A novel hybrid kinase is essential for activating the $\sigma^{\text{B}}\text{-}\text{mediated}$ stress response of Bacillus cereus	85
Chapter 6	Novel σ^{B} activation modules of Gram-positive bacteria involve the use of complex hybrid histidine kinases	111
Chapter 7	Summarising discussion	127
	References	139
	Samenvatting voor iedereen	159
	Dankwoord	165
	List of publications	169
	Curriculum Vitae	171
	VLAG graduate school activities	173

Supplementary material: Chapter 2: http://mic.sgmjournals.org Chapter 3: www.cmbi.ru.nl/RRoperators Chapters 5-6: www.cmbi.ru.nl/~mdebeen/Thesis

Abstract

Bacillus cereus is a Gram-positive pathogen that is recognised as an important cause of food-borne disease worldwide. Within the genus Bacillus, B. cereus and its closest relatives form a homogeneous subdivision that has been termed the B. cereus group. This group includes B. anthracis, a pathogen that can cause anthrax in mammals, and B. thuringiensis, an insect pathogen that is used as an insecticide worldwide. Members of the B. cereus group can adapt to a wide range of environmental challenges. In bacteria, these challenges are generally monitored by two-component systems (TCS), which consist of a histidine kinase (HK) and a partner response regulator (RR). Upon sensing a specific environmental stimulus, the HK activates its cognate RR, which in turn controls the expression of genes that are involved in the appropriate response. This thesis describes the functional analysis of TCSs in the B. cereus group. By using in silico techniques, 50-58 HKs and 48-52 RRs were identified in eight different B. cereus group genomes. Biological functions, including the involvement in sporulation, biofilm formation and host-microbe interactions were predicted for these TCS proteins. A phylogenetic footprinting approach was developed and used to identify specific binding sites and target genes for over 50% of the B. cereus group DNA-binding RRs. These predictions allowed relating several RRs to a minimal regulon and thereby to a characteristic transcriptional response. To further support these predictions, the transcriptomes of two B. cereus TCS deletion mutants ($\Delta yvrHG$ and $\Delta yufLM$) were analysed and compared with the transcriptome of wild-type B. cereus. This revealed that the minimal regulon predictions were correct for the two respective TCSs. Furthermore, the predicted biological roles for these TCSs, including roles in antibiotic resistance (YvrHG) and fumarate metabolism (YufLM), were supported by phenotypic tests. Besides the many "classical" HKs and RRs detected in the *B. cereus* group, several atypical TCS proteins were found. These included five RRs without a DNA-binding output domain and two hybrid HKs (HK-RR fusions). Genome analyses revealed that one of the hybrid HK-encoding genes (BC1008) is located in a conserved gene cluster that also encodes the a-typical RR RsbY. In B. cereus, RsbY is known to activate the key stress-responsive sigma factor σ^B . As a partner HK for RsbY was still "missing", the role of BC1008 in the σ^{B} -mediated stress response was tested. Indeed, a *bc1008* deletion strain appeared incapable of inducing σ^{B} and its associated regulon upon stress conditions and appeared impaired in its heat adaptive response. In addition, truncation of the BC1008 fused RR receiver domain indicated that this domain plays a role in fine-tuning BC1008 activity. A comparative genome analysis further indicated that BC1008-type hybrid HKs control σ^{B} -like sigma factors in at least several other Gram-positive bacteria, including Geobacillus, Paenibacillus and actinobacteria. In summary, the research described in this thesis contributes to our understanding of B. cereus adaptive responses through TCSs. This knowledge may be applied for the development of novel intervention strategies for an improved control of B. cereus in food production environments.

Chapter 1

Introduction and outline of the thesis

Bacillus cereus

Bacillus cereus is a Gram-positive, facultative anaerobic bacterium. It was first described in the United Kingdom in 1887, where it was isolated from air in a cow-shed (Frankland and Frankland, 1887). It was given the name "*Bacillus*" (small rod) "*cereus*" (wax-like) because its rod-shaped vegetative cells form easily recognisable, wax-like colonies on agar plates. The first isolated *B. cereus* strain is now known as the type strain ATCC (American Type Culture Collection) 14579. A phase contrast microscopy image and electron micrograph of the vegetative cells of this strain are shown in Fig. 1.1A and 1.1B, respectively. Most *Bacillus* species, including *B. cereus*, are known to produce endospores, which are highly specialised metabolically dormant cell types that are resistant to extreme environmental conditions, including heat, dehydration and other physical stresses (Setlow, 2000) (Fig. 1.1A and 1.1C). Other distinguishing features of *B. cereus* include its motility, haemolysis, its active production of lecithinase (phospholipase) and its inability to ferment mannitol (Johnson, 1984).



Fig. 1.1. Bacillus cereus ATCC 14579 vegetative cells and endospores.

(A) Phase contrast microscopy image of *B. cereus* vegetative cells (lower left) and endospores (upper right).

(B) Scanning electron micrograph of *B. cereus* vegetative cells.

(C) Scanning electron micrograph of B. cereus endospores. The scale bars in B and C represent 1 $\mu m.$

The images were kindly provided by Menno van der Voort (A and C) and Masja Nierop Groot (B).

B. cereus is an important food-borne pathogen

B. cereus is a human pathogen that can cause severe local or systemic infections, such as endophthalmitis and septicaemia. However, *B. cereus* is best known for its ability to cause food-poisoning and food spoilage. Many types of food have been associated with *B. cereus* food-borne disease, including meat, vegetables, puddings, milk, rice, pasta and noodles (Johnson, 1984; Kotiranta *et al.*, 2000). *B. cereus* can cause two distinct types of food-borne disease: the diarrhoeal and the emetic type. Although both types are generally mild and self-containing, more serious and even lethal cases have been reported (Dierick *et*

al., 2005; Fricker et al., 2007; Lund et al., 2000; Mahler et al., 1997). The diarrhoeal disease is often associated with protein-rich foods, such as meat, vegetables, puddings and milk products. It was first described by Steinar Hauge, who investigated several food-borne outbreaks in Norwegian hospitals between 1947-1949 (Hauge, 1950, 1955). The diarrhoeal disease is thought to be caused by vegetative cells (ingested as viable cells or spores) that produce enterotoxins in the small intestine. Typical symptoms include abdominal pains, watery diarrhoea, nausea and vomiting. The incubation time generally ranges between 8-16 hours after ingestion, while the symptoms normally last for 12-24 hours. However, longer incubation times have been observed and the symptoms can last for up to several days (Kotiranta et al., 2000). The emetic disease is often associated with starch-rich foods, such as fried and cooked rice, pasta and noodles. It was first identified in 1974 when B. cereus was linked to several outbreaks caused by eating cooked rice in the United Kingdom in the early 1970s (Mortimer and McCann, 1974). The emetic disease is caused by the B. cereus emetic toxin cereulide, which is produced in foods before ingestion. Symptoms mainly include nausea and vomiting, which occur between 30 minutes to 6 hours after ingestion and which generally last for 6-24 hours (Ehling-Schulz et al., 2004).

B. cereus is recognised as an important cause of food-borne disease worldwide (Granum, 2007). Still, a general consensus about its exact importance as a food-borne pathogen is lacking. This is reflected by the outcome of several surveillance studies, which revealed large differences between the numbers of reported B. cereus outbreaks between different countries. Studies in Europe have shown that especially in The Netherlands and Norway, B. cereus is an important food-borne pathogen. In The Netherlands, between 1999 and 2000, B. cereus was the most important causative agent of food-borne outbreaks, causing 25% of the outbreaks where a causative agent was identified, while in Norway B. cereus even accounted for 35% and 32% of the outbreaks in 1999 and 2000, respectively (Schmidt and Gervelmeyer, 2003). These numbers are comparable with reported numbers in Taiwan, where B. cereus accounted for 18% of food-borne outbreaks between 1986 and 1995 (Pan et al., 1997). However, much lower numbers have been reported in other countries, such as Germany, England, France and Italy, where the numbers ranged between 0.5% and 5% in the period 1999-2000 (Schmidt and Gervelmeyer, 2003). Similarly, B. cereus has been estimated to be responsible for only 0.2% of the food-borne illnesses in the United States (Mead et al., 1999). The high numbers of B. cereus outbreaks reported in The Netherlands, Norway and Taiwan are probably due to high local awareness of B. cereus. Indeed, it is likely that the low numbers of reported cases in other countries reflect a large under-estimation of B. cereus food-borne outbreaks. This under-estimation is thought to be caused by the generally short and mild course of B. cereus food-borne disease, which does not encourage patients to seek medical care. Indeed, B. cereus outbreaks tend to be reported only when many people are affected or when the symptoms are more severe than usual. In addition, B. cereus outbreaks are not always ascribed to B. cereus, as the symptoms of the diarrhoeal disease are similar to the symptoms caused by Clostridium perfringens type A food-poisoning and the symptoms of the emetic disease are similar to those caused by Staphylococcus aureus

intoxication (reviewed in Stenfors Arnesen *et al.*, 2008). Nevertheless, there seems to be a general trend of increasing numbers of (severe) *B. cereus* food-borne outbreaks especially in industrialised countries (Kotiranta *et al.*, 2000). Therefore, it will be interesting to see whether future awareness of *B. cereus* will increase and how this will affect the number of reported outbreaks caused by this food-borne pathogen.

B. cereus enterotoxins

The *B. cereus* diarrhoeal disease is an infection caused by vegetative cells that are thought to produce different enterotoxins in the small intestine. Several pore-forming enterotoxins have been associated with the diarrhoeal disease, of which the best-studied are haemolysin BL (Hbl), non-haemolytic enterotoxin (Nhe) and cytotoxin K (CytK). The genes encoding these enterotoxins are part of a virulence regulon that is controlled by the pleiotropic transcriptional regulator PIcR (Gohar et al., 2008). Hbl and Nhe are homologous three-component enterotoxins that seem to be unique to members of the B. cereus group (see below) (From et al., 2005). Hbl is composed of the protein components L_2 , L_1 and B, which are encoded by the operonic genes hblC, hblD and hblA, respectively (Heinrichs et al., 1993; Ryan et al., 1997), while Nhe is composed of the protein components NheA, NheB and NheC, which are encoded by the operon nheABC (Granum et al., 1999). Hbl has been shown to cause accumulation of fluids in rabbit ileal loops (Beecher et al., 1995) and vascular permeability (Beecher and Wong, 1994) and it displays dermonecrotic and cytotoxic activity (Lund and Granum, 1997) and haemolytic activity (hence its name) towards erythrocytes (Beecher and Wong, 1997; Beecher et al., 2000; Lindbäck et al., 1999). The latter activity probably involves independent binding of the three Hbl components to erythrocytes after which the components assemble into a membrane-attacking (pore-forming) complex that causes lysis via a colloid osmotic mechanism. In contrast to Hbl, Nhe has long been considered to be non-haemolytic and to display cytotoxic properties only (Lindbäck et al., 2004). However, a recent study has shown that Nhe also displays haemolytic activity towards erythrocytes (Fagerlund et al., 2008). The third well-studied enterotoxin, CytK, consists of only one component that belongs to the family of β -barrel pore-forming toxins. CytK was first isolated from a B. cereus strain that caused a severe outbreak of gastro-enteritis, killing three people in France in 1998 (Lund et al., 2000). CytK acts by the formation of pores in planar lipid bilayers. This may infer cytotoxic effects on intestinal epithelial cells, resulting in loss of their barrier function (Hardy et al., 2001). CytK has been found in many food-poisoning B. cereus strains, but is also found in the type strain ATCC 14579, which is non-pathogenic (Fagerlund et al., 2004; Guinebretière et al., 2002).

Several lines of evidence have suggested that Nhe is the most dominant *B. cereus* diarrhoeal enterotoxin. However, the contribution of the different enterotoxins to causing disease may vary between strains and multiple toxins seem to act together (reviewed in Stenfors Arnesen *et al.*, 2008). In addition, *B. cereus* produces several less well-characterised proteins, including enterotoxin FM (Asano *et al.*, 1997; Luxananil *et al.*, 2003), cereolysin O (Kreft *et al.*, 1983),

haemolysin II (Baida *et al.*, 1999), haemolysin III (Baida and Kuzmin, 1995), InhA2 (Fedhila *et al.*, 2003) and three phospholipases C (Kuppe *et al.*, 1989) and the contribution of these proteins to gastro-entiritis is still poorly understood.

B. cereus emetic toxin

In contrast to the *B. cereus* diarrhoeal disease, the emetic disease is caused by the production of only one toxin, the so-called emetic toxin or cereulide. Cereulide is a 1.2 kDa cyclic dodecadepsipeptide with the chemical formula [D-O-Leu-D-Ala-O-Val-L-Val]₃ (Agata et al., 1995; Ehling-Schulz et al., 2004). It is produced by a single non-ribosomal peptide synthetase, which is encoded by the cereulide synthetase (ces) gene cluster that is located on a mega-plasmid related to pXO1 of B. anthracis (see below) (Ehling-Schulz et al., 2005; Ehling-Schulz et al., 2006; Hoton et al., 2005). Cereulide is chemically related to the ionophore valinomycin and, indeed, it also seems to act as a cation ionophore that causes the inhibition of fatty acid oxidation by the mitochondria (Mikkola et al., 1999). The cereulide mode of action has not been completely resolved, but a 5-HT₃ receptor-mediated mechanism appears to be involved (Agata *et al.*, 1995). In contrast to the enterotoxins, cereulide is highly resistant to heat, low pH and proteolytic enzymes of the gastrointestinal tract (Ehling-Schulz et al., 2004). This feature causes problems when food is prepared in advance and needs reheating prior to consumption. Between preparation and reheating, B. cereus spores that survived the initial preparation steps can germinate and produce cereulide. Reheating inactivates the germinated bacteria, but does not inactivate cereulide, which may subsequently cause emetic food-poisoning.

The "*B. cereus* group"

B. cereus is a member of the genus *Bacillus*, which is a highly heterogeneous genus that includes species with large variations in phenotypes, nutritional requirements and other physiological and metabolic characteristics. Within the genus *Bacillus*, *B. cereus* and its closest relatives form a highly homogeneous subdivision, which has been termed the "*B. cereus* group". This group comprises the species *B. cereus*, *B. anthracis*, *B. thuringiensis*, *B. mycoides* and *B. weihenstephanensis*.

B. anthracis is the etiological agent of the acute and fatal disease anthrax in mammals. It is the most notorious member of the *B. cereus* group, especially since it was used in bioterrorist attacks in the USA in 2001 (Jernigan *et al.*, 2002). Three different forms of anthrax are known, including cutaneous, gastrointestinal and inhalation anthrax. The cutaneous form is the most common form and is easily treated with antibiotics. In contrast, the other two forms are much more dangerous as the initial symptoms generally resemble a mild flu or fever that can abruptly develop into an often fatal systemic infection which is resistant to antibiotic treatment. The most important *B. anthracis* virulence factors are the regulator AtxA, a poly- γ -D-glutamate capsule that is used for escaping the host immune response, and two toxins. The toxins are constituted from combinations of protective antigen (PA) with either lethal factor (LF) or oedema factor (EF). PA, LF and EF are encoded by *pagA*, *lef* and *cya*,

respectively, which are located on a large plasmid pXO1 that also harbours *atxA*. The *B. anthracis* capsule is encoded by the five-gene operon *capBCADE*, which is located on another large plasmid pXO2 (Mock and Fouet, 2001). Besides its property to cause anthrax, *B. anthracis* is distinguished from its closest relatives by its non-haemolytic and non-motile nature and its susceptibility to penicillin and γ -phage (Hoffmaster *et al.*, 2004).

B. thuringiensis is generally considered to be an insect pathogen only. The insecticidal properties are associated with the ability of *B. thuringiensis* to produce toxins in the form of parasporal crystal proteins, which are encoded on large plasmids and which are widely used for the biological control of insects in crop protection (Aronson and Shai, 2001; Berry *et al.*, 2002; Schnepf *et al.*, 1998).

The least-studied members of the *B. cereus* group are *B. mycoides* and *B. weihenstephanensis*. *B. mycoides* is differentiated from its closest relatives by its rhizoidal colony morphology (Di Franco *et al.*, 2002), while *B. weihenstephanensis* is a psychrotolerant species that can grow below 7 °C, but not at 43 °C (Lechner *et al.*, 1998).

Historically, members of the *B. cereus* group were assigned to one of the above species on the basis of the associated phenotypic traits. However, species definitions within bacterial systematics are ambiguous as illustrated by genome sequencing data, which showed that B. cereus, B. anthracis and B. thuringiensis are so closely related that they were proposed to belong to one bacterial species (Helgason et al., 2000). Their 16S rDNA sequences share over 99% similarity (Ash et al., 1991) and phylogenetic studies based on chromosomal markers have shown that there is no taxonomic basis for a separate species status for B. cereus and B. thuringiensis (Helgason et al., 2004; Ko et al., 2004; Priest et al., 2004), while B. anthracis can be considered as a clone of B. cereus (Hill et al., 2004; Keim et al., 2000). Furthermore, as described above, many of the distinguishing features between B. cereus, B. anthracis and B. thuringiensis are encoded on plasmids, which are recognised as highly mobile genetic elements. This was clearly illustrated when Hoffmaster and colleagues found a B. cereus strain that caused a severe inhalation anthrax-like illness and that carried a pXO1-like plasmid (Hoffmaster et al., 2004).

Niches and life cycles

Members of the *B. cereus* group are ubiquitously present in soil and can adapt to a wide range of environmental conditions (Abee and Wouters, 1999; Jensen *et al.*, 2003; Kotiranta *et al.*, 2000). The ecological niches and life cycles of *B. anthracis* and *B. thuringiensis* may be more specialised than those of *B. cereus*. *B. anthracis* spores are ingested by herbivores and germinate within the host to produce vegetative cells, which multiply and produce virulence factors, ultimately killing the host. Upon death, large numbers of *B. anthracis* cells are released into the soil, where they sporulate upon contact with air, thus completing the *B. anthracis* life cycle (Mock and Fouet, 2001). The life cycle of *B. thuringiensis* may be compared to that of *B. anthracis*. Upon death of its insect host, *B. thuringiensis* is released into the soil where it is a ubiquitous inhabitant and where it can germinate and grow under favourable conditions.

Although *B. thuringiensis* is generally considered to be an insect pathogen only, the interaction of *B. thuringiensis* with its insect host is thought to be not always lethal to the host (Jensen *et al.*, 2003). The niches and life cycle of *B. cereus* remain more obscure. *B. cereus* spores can be found in many types of soils, sediments, dust and plants (Jensen *et al.*, 2003; Kotiranta *et al.*, 2000). For this reason, *B. cereus* is considered to be mainly a soil inhabitant and, indeed, *B. cereus* can germinate, grow and sporulate in soil, thus demonstrating a saprophytic life cycle (Vilain *et al.*, 2006). However, genome analyses have indicated that *B. cereus* specialises in protein metabolism, pointing towards a symbiotic or parasitic life cycle (Ivanova *et al.*, 2003). This is further supported by the fact that *B. cereus* has been isolated from guts of soil-dwelling arthropods (Margulis *et al.*, 1998) and from stool samples of healthy humans (Jensen *et al.*, 2003; Yea *et al.*, 1994). Furthermore, *B. cereus* has been isolated from surface waters, a niche that may allow this bacterium to easily enter food-processing lines (Østensvik *et al.*, 2004).

Two-component signal transduction

The ubiguitous presence and high adaptive capabilities of *B. cereus* and its closest relatives may imply that these species are equipped with a large amount of signalling pathways that allow rapid and robust responses to fluctuations in their environment. In bacteria, including archaebacteria, the dominant signalling pathways for monitoring environmental cues are the so-called two-component systems (TCS) (Galperin, 2005). Analyses of whole-genome and metagenome sequences have revealed that TCS protein domains are the second most numerous Pfam domains in bacteria, exceeded only by ABC-type transporter domains (Bateman et al., 2004; Whitworth and Cock, 2008, 2009). TCSs are known to monitor a wide variety of conditions, such as nutrient deprivation, cold/heat shock and the presence of antimicrobial compounds (Aguilar et al., 2001; Jordan et al., 2008; Sun et al., 1996), and they can control virtually all types of microbial behaviour, including motility and chemotaxis, sporulation, biofilm formation and quorum sensing (Jiang et al., 2000; López et al., 2009; Lyon and Novick, 2004; Szurmant and Ordal, 2004). TCS variants have also been found in eukaryotes, where they regulate hormone responses and circadian rhythms in plants (Mizuno, 2005) and osmotic adaptation in yeast (Posas et al., 1996). Interestingly, cyanobacterial-type TCSs that control photosynthesis have recently been found in the chloroplasts of algae and plants (Duplessis et al., 2007; Puthiyaveetil et al., 2008; Puthiyaveetil and Allen, 2009). These findings support previous suggestions that TCSs have entered the eukaryotic domain of life through endosymbiotic lateral gene transfer from the evolutionary ancestors of chloroplasts and mitochondria (Koretke et al., 2000). Still, despite the apparent widespread nature of TCSs throughout all domains of life, they do not seem to occur in animals, making these modules an attractive target for novel bioactives against, for example, human pathogenic bacteria (Watanabe et al., 2008).

The two-component signal transduction paradigm

TCSs consist of a sensor histidine kinase (HK) and its cognate response regulator (RR), which are generally encoded by a pair of adjacent genes. All TCSs make use of the same mechanistic principle; that is phosphoryl transfer between the HK and the cognate RR. A classical TCS essentially consists of four domains, including an N-terminal, membrane-associated sensory domain and a C-terminal phosphotransferase domain in the HK and an N-terminal phosphotransferase (receiver) domain and C-terminal DNA-binding output domain in the RR. The phosphotransferase domain of the HK can be further sub-divided into the H-box and the HATPase domain (Fig. 1.2).



Fig. 1.2. Signal transduction in a "classical" TCS.

Upon environmental stimulus perception by the HK sensory domain (dark grey), the downstream phosphotransferase domain (white), which consists of an H-box and HATPase domain, "auto"-phosphorylates on a conserved histidine (H) in the H-box. phosphoryl The histidine-bound is transferred to a conserved aspartate (D) in the receiver domain (dark grey) of the cognate RR. This activates the RR DNAbinding output domain (black), which binds to specific operators on the DNA, thereby activating/repressing genes that are involved in the adaptive response towards the initial stimulus perceived by the HK. Note that HKs and RRs generally form `auto"and that homodimers HK phosphorylation involves phosphoryl crosstransfer from the HATPase domain of one dimer subunit to the H-box of the other dimer subunit. For illustrative purposes, both the HK and the RR are depicted as monomers.

Perception of a specific environmental stimulus by the HK sensory domain triggers the activation of the downstream HK phosphotransferase domain. The activated HATPase domain consumes ATP and cross-transfers the released phosphoryl group to a highly conserved histidine residue that is located within the H-box of the HK dimerising partner. After this "auto"-phosphorylation step, the histidine-bound phosphoryl is transferred to a highly conserved aspartate residue within the receiver domain of the cognate RR (Hoch, 2000; Stock *et al.*, 2000) (Fig. 1.2). This leads to activation of the RR output domain, which subsequently binds to directly or indirectly repeated elements (operators) on the DNA that are located close to the promoter regions of the target genes (de Been *et al.*, 2008). Binding of the RR to these operators activates (or represses) the associated target genes, which function in the adaptive response to the initial signal perceived by the HK. Thus, the TCS provides an elegant tool for bacteria

to accurately and rapidly respond to environmental changes (Hoch, 2000; Stock *et al.*, 2000).

TCS families

Considering the fact that TCSs connect to a wide variety of biological processes, it is not surprising that significant variability occurs between different TCSs. This variability is most apparent for the HK sensory and RR output domains, but is also found for the most conserved TCS domains: the HK and RR phosphotransferases. Phylogenetic analyses on the basis of these domains revealed the existence of 11 major TCS families, which have presumably evolved from a common TCS ancestor (Grebe and Stock, 1999; Koretke et al., 2000). It appears that most TCSs belong to families 1a (OmpR) and 7 (NarL), after the classification scheme of Grebe and Stock (1999). TCSs within one family tend to have similar HK sensory and RR output domains and (especially for specific families) they also tend to control similar cellular processes. For example, TCSs of family 5 are known to specifically function in the uptake and metabolism of tricarboxylic acid cycle intermediates (Asai et al., 2000; Janausch et al., 2002b; Tanaka et al., 2003; Yamamoto et al., 2000). Although the average bacterium contains multiple specimens of each TCS family, some bacteria contain disproportionate amounts of one family, as is exemplified by lactobacilli, which contain a relatively large amount of quorum-sensing TCSs of family 10 (Sturme et al., 2007). As stated above, most variability between TCSs resides in the HK sensory and RR output domains. This becomes most apparent when comparing HKs and RRs that belong to different families. In fact, most TCS families are named after the specifc RR output domains that are generally encountered within the associated RRs (e.g. OmpR, NarL). Still, large variability can be found even within one family, especially when regarding the HK sensory domains (Alm et al., 2006; Stephenson and Hoch, 2002).

HK sensory and RR output domains

In the past few decades, several types of HK sensory and RR output domains have been functionally described. Sensory domains that have been characterised in some detail and that are most commonly associated with HKs include the redox/light/metabolite sensing PAS domain (Ponting and Aravind, 1997; Zhulin et al., 1997), the small ligand (Ca²⁺) sensing Cache domain (Anantharaman and Aravind, 2000) and the small ligand/cyclic GMP/photopigment sensing GAF domain (Aravind and Ponting, 1997; Ho et al., 2000). Although advancements have been made in the past few years, the exact mechanistic principles of stimulus perception by HK sensory domains are still poorly understood. In addition, the stimulating signals themselves are largely unknown for numerous HK sensory domains (Heyl et al., 2007; Mougel and Zhulin, 2001; Pas et al., 2004; Zhulin et al., 2003). Some domains that are frequently associated with HKs do not seem to have a sensory role. The best-studied example is the HAMP domain, which probably functions in transmitting signals from sensory to signalling domains (Aravind and Ponting, 1999; Hulko et al., 2006; Jin and Inouye, 1994).

As compared to what is known about signal input and transmission within HKs, the signalling mechanism within RRs seems to be better understood. Phosphorylation of the RR receiver domain is known to increase the propensity of the RR to form homodimers, which exposes the output domains to perform their DNA-binding task (Kern et al., 1999; Maris et al., 2002). Structural studies have shown that most RR output domains are composed of a DNA-binding helixturn-helix (HTH) fold. Between different RR families, slight structural variations occur in the HTH. An example is the 'winged' HTH for RRs of the OmpR family (Martínez-Hackert and Stock, 1997). Other HTH structural variations have been found for RRs that belong to the NarL, Fis, AraC and Spo0A family (Baikalov et al., 1996; Lewis et al., 2000; Pelton et al., 1999; Rhee et al., 1998). However, not all RRs recognise the DNA through a HTH fold. A recent structural study on AgrA of Staphylococcus aureus, a member of the LytTR family, has shown that AgrA recognises the DNA by means of long loops that insert into successive major grooves (Sidote et al., 2008). Finally, no structural information is available for the CitB and GlnL RR output domain families, but characteristics of their DNA-binding properties have been deduced from in vitro binding studies (Satomura et al., 2005; Yamamoto et al., 2000).

Non-typical TCSs and complex signalling cascades

With the availability of more and more bacterial genome sequences, it has become clear that many variations on the TCS paradigm occur in bacteria. First of all, a recent genome survey has shown that around 40% of all bacterial HKs do not contain any predicted transmembrane helices, suggesting that many of these HKs do not directly perceive extracellular signals (Galperin, 2005). Similarly, around 30% of all bacterial RRs do not contain a DNA-binding output domain, but are single domain RRs or contain an output domain with enzymatic or RNA-, protein-, or ligand-binding properties (Galperin, 2006). Furthermore, molecular studies have shown that some HKs are bi-functional, not only acting as kinases but modulating the activity of their partner RRs by also acting as phosphatases (Igoshin et al., 2008). Finally, not all TCSs involve a single phosphoryl transfer step. More complex signalling cascades are known that involve multistep His-Asp-His-Asp phosphotransfers (Fig. 1.3). This is exemplified by the BvgS-BvgA TCS of Bordetella pertussis. In this signalling cascade, the first three phosphoryl transfer steps occur within the hybrid kinase BvgS. This a-typical HK contains a HK phosphotransferase domain that is followed by a fused RR receiver and Hpt-type phosphotransferase domain. In BvqS, phosphoryl is first shuttled between the H-box, the RR receiver and the Hpt domain (His-Asp-His) prior to its transfer to the RR receiver domain of BvqA (His-Asp) (Uhl and Miller, 1996a, 1996b). Another example of an extended signalling pathway that involves TCS proteins is formed by the sporulation initiation pathway of Bacillus subtilis. Here, the His-Asp-His-Asp phosphorelay is divided over separate proteins, including multiple HKs that can feed phosphoryl into the Spo0F-Spo0B-Spo0A cascade (Burbulys et al., 1991; Jiang et al., 2000). Although this is not the case for the latter pathway, bacterial multistep phosphorelays generally involve complex hybrid HKs, like BvgS (Fig. 1.3) (Appleby et al., 1996). Still, not all hybrid HKs take part in linear, multistep

phosphorelays. This is exemplified by the hybrid HK VirA of *Agrobacterium tumefaciens*, of which the fused RR receiver domain plays a role in fine-tuning the kinase activity of VirA towards its cognate RR VirG (Brencic *et al.*, 2004; Chang and Winans, 1992; Chang *et al.*, 1996). From the above examples, it becomes apparent that TCSs are highly modular in nature. This may at least partly explain why they have evolved as the most prominent signalling devices in bacteria.



Fig. 1.3. Signal transduction in multistep phosphorelays.

Besides the "classical" TCS that involves one single phosphoryl transfer step (top), more complex multistep TCS phosphorelays have been found: for example, in *Bordetella pertussis* (Uhl and Miller, 1996a, 1996b), *Saccharomyces cerevisiae* (Posas *et al.*, 1996) and *Bacillus subtilis* (Burbulys *et al.*, 1991). Multistep phosphorelays often include hybrid HKs that contain HK as well as RR domains (Appleby *et al.*, 1996). However, hybrid HKs do not necessarily act in (linear) multistep phosphorelays as is illustrated by VirA of

Agrobacterium tumefaciens of which the fused RR domain fine-tunes VirA kinase activity towards the cognate RR VirG.

Domain colouring is similar as in Fig. 1.2. The output domain of Ssk1 (light grey) is not involved in DNA-binding, but interacts with downstream signalling proteins. *, Besides KinA, four other HKs (KinB, KinC, KinD and KinE) can feed phosphoryl into the Spo0F-Spo0B-Spo0A cascade (Jiang *et al.*, 2000).

Outline of this thesis

The research presented in this thesis involves the characterisation of TCSmediated responses in *B. cereus*. This study was initiated to obtain a better understanding of the high adaptive capabilities of *B. cereus*, which contribute to the problematic nature of this organism to food processing industries. To assess the biological role of the *B. cereus* TCSs, extensive use was made of the complete genome sequence information available for several members of the *B. cereus* group as well as for a large number of related low-GC Gram-positive bacteria. The combined use of *in silico*, as well as genetic and molecular techniques has allowed us to obtain better insight into the molecular mechanisms underlying the TCS-mediated adaptive properties of *B. cereus*.

Prior to this study, virtually nothing was known about TCS signal transduction in *B. cereus*. Therefore, we set out to identify and analyse the complete arsenal of TCSs in eight completely sequenced members of the *B. cereus* group, including *B. anthracis* and *B. thuringiensis*. The results of this analysis are described in **chapter 2**. A relatively large set of TCS proteins was found in each of the *B. cereus* group members, ranging from 50-58 HKs and 48-52 RRs. Biological functions, including the involvement in sporulation, biofilm formation and hostmicrobe interactions were predicted for these TCS proteins. Interestingly, *B. anthracis* appeared to lack various TCSs and contained many truncated HK and RR genes, possibly reflecting its specialised pathogenic life cycle.

Chapter 3 describes the development of a phylogenetic footprinting/shadowing approach that was used to identify specific operators and cognate target genes for the DNA-binding RRs of the *B. cereus* group. Analysis of the operator sequences revealed characteristic traits for each RR subfamily. Moreover, the identification of the specific operator motifs allowed relating several RRs to a minimal regulon and thereby to a characteristic transcriptional response. In this way, new biological roles were attributed to various TCSs, including roles in cytochrome *c* biogenesis, transport of nutrients and resistance to toxic ions and antibiotics.

To further characterise and support the TCS regulon predictions, the transcriptomes of two *B. cereus* TCS deletion mutants were analysed and compared with the transcriptome of wild-type *B. cereus*. The results of these analyses are described in **chapter 4**. The transcriptome data revealed that our initial *in silico* predictions were correct for the two respective TCSs. The predicted RR-specific operators and the generated transcriptome data were further integrated to extend the TCS regulon predictions. This revealed additional biological roles for the two TCSs, including roles in antibiotic resistance and fumarate metabolism. These roles were supported by phenotypic tests.

In the initial TCS identification analyses described in chapter 2, we found two different hybrid HKs (HK-RR fusions) in members of the *B. cereus* group. The

functional role of one of these hybrid HKs is described in **chapter 5**. Genome sequence analysis revealed that the gene encoding this hybrid HK (BC1008) is located in a conserved gene cluster that also encodes the unique RR RsbY. In *B. cereus*, RsbY is known to activate the key stress-responsive sigma factor σ^B . As a partner HK for RsbY was still "missing", we wondered about a possible role of BC1008 in the σ^B -mediated stress response. Indeed, a *bc1008* deletion strain was incapable of inducing σ^B and its associated regulon upon stress conditions and was impaired in its heat adaptive response. Truncation of the BC1008 fused RR receiver domain indicated that this domain plays a role in fine-tuning BC1008 activity.

The results in chapter 5 indicated that members of the *B. cereus* group use a unique signalling route to activate the stress-responsive sigma factor σ^{B} . To elucidate this potentially unique feature of the *B. cereus* group, we performed a comparative genome analysis focusing on BC1008. The results of this analysis are described in **chapter 6**. We found that BC1008-type hybrid HKs are (i) conserved across a wide variety of bacterial species, (ii) display extremely variable sensory as well as C-terminal RR domains and (iii) may be subject to complex fine-tuning systems. In addition, we found that BC1008-type HKs seem to control σ^{B} -like sigma factors in at least several other low- as well as high-GC Gram-positives, including *Geobacillus*, *Paenibacillus* and actinobacteria.

Chapter 7 includes the general discussion, the conclusions and future perspectives of the research described in this thesis.

Chapter 2

Comparative analysis of two-component signal transduction systems of *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus anthracis*

Mark de Been, Christof Francke, Roy Moezelaar, Tjakko Abee, Roland J. Siezen

Published in: Microbiology, 2006, 152(Pt 10): 3035-3048

Members of the Bacillus cereus group are ubiquitously present in the environment and can adapt to a wide range of environmental fluctuations. In bacteria, these adaptive responses are generally mediated by two-component signal transduction systems (TCS), which consist of a histidine kinase (HK) and its cognate response regulator (RR). With the use of in silico techniques, a complete set of HKs and RRs was recovered from eight completely sequenced B. cereus group genomes. By applying a bidirectional best hits method combined with gene neighbourhood analysis, a footprint of these proteins was made. Around 40 HK-RR gene pairs were detected in each member of the B. cereus group. In addition, each member contained many HK and RR genes not encoded in pairs ("orphans"). Classification of HKs and RRs based on their enzymatic domains together with the analysis of two neighbour-joining (NJ) trees of these domains revealed putative interaction partners for most of the "orphans". Putative biological functions, including involvement in virulence and hostmicrobe interactions, were predicted for the B. cereus group HKs and RRs by comparing them with those of *B. subtilis* and other microorganisms. Remarkably, B. anthracis appeared to lack various HKs and RRs and was found to contain many truncated, putatively non-functional, HK and RR genes. It is hypothesised that specialisation of *B. anthracis* as a pathogen could have reduced the range of environmental stimuli to which it is exposed. This may have rendered some of its TCSs obsolete, ultimately resulting in the deletion of some HK and RR genes.

Introduction

The Bacillus cereus group consists of Gram-positive, spore-forming bacteria. It includes Bacillus cereus, a species often associated with food-borne disease, B. thuringiensis, which is used as a biological pesticide worldwide, and B. anthracis, a pathogen of warm-blooded animals that can cause the often fatal disease anthrax. Members of the B. cereus group form a highly homogeneous subdivision within the genus Bacillus and it has been proposed that B. cereus, B. thuringiensis and B. anthracis are in fact varieties of the same species (Daffonchio et al., 2000; Helgason et al., 2000). However, B. anthracis and B. thuringiensis differ from B. cereus by containing plasmid-encoded specific toxins and capsule (B. anthracis only) (Okinaka et al., 1999b; Schnepf et al., 1998) and recent studies have shown that B. anthracis is rather monomorphic, while there is large phylogenetic diversity within B. cereus and B. thuringiensis (Bavykin et al., 2004; Hill et al., 2004; Priest et al., 2004). Members of the B. cereus group are ubiquitously present in the environment and can adapt to a wide range of environmental conditions (Abee and Wouters, 1999; Jensen et al., 2003; Kotiranta et al., 2000). This raises the question of how these organisms are able to monitor these conditions and respond to them. In bacteria, sensing and adapting to environmental fluctuations is generally mediated by two-component signal transduction systems (TCSs) (Parkinson and Kofoid, 1992; Stock et al., 1989). These systems have been shown to monitor a wide variety of conditions, including nutrient deprivation, cold/heat shock, osmotic stress, low pH and many others (Aguilar et al., 2001; Bearson et al., 1998; Jung and Altendorf, 2002; Sun et al., 1996). In addition, TCSs have been shown to initiate important adaptive responses, such as sporulation, biofilm formation and chemotaxis (Jiang et al., 2000; Lyon and Novick, 2004; Szurmant and Ordal, 2004). TCSs consist of a sensor histidine kinase (HK) and its cognate response regulator (RR), which are often encoded on adjacent genes. A typical HK contains an N-terminal, membrane-associated sensory domain and a Cterminal, cytosolic H-box and HATPase domain. Together, these cytoplasmic domains make up the phosphotransferase domain. A typical RR is a cytosolic protein consisting of an N-terminal receiver domain and a C-terminal DNAbinding domain. Upon sensing specific environmental stimuli the HATPase domain mediates "auto"-phosphorylation of the HK at a conserved histidine residue of the H-box. The histidine-bound phosphoryl group is subsequently transferred to an aspartic acid residue of the RR receiver domain, leading to activation of the RR. The activated RR then binds to specific regions on the DNA, which leads to the activation/repression of genes involved in adaptive responses (Parkinson and Kofoid, 1992; Stock et al., 1989). Besides the prototypical TCSs, in which phosphoryl is transferred to the RR in a single step, more complex signal transduction systems also occur in bacteria. In these so-called phosphorelays, activation of the RR by the HK occurs through a multitude of

phosphoryl transfer steps (Appleby *et al.*, 1996; Burbulys *et al.*, 1991; Posas *et al.*, 1996).

Although the *B. cereus* group has received much attention in the past few years and many *B. cereus* group genomes have recently been sequenced and

published (Han *et al.*, 2006; Ivanova *et al.*, 2003; Rasko *et al.*, 2004; Read *et al.*, 2002; Read *et al.*, 2003), hardly any research has been done on TCSs in this bacterial group. Only recently, a number of HKs has been shown to initiate sporulation in *B. anthracis* (Brunsing *et al.*, 2005) and a RR has been shown to activate the alternative sigma factor σ^{B} in *B. cereus* (van Schaik *et al.*, 2005a). Since so little is known about two-component signal transduction in the *B. cereus* group, we initiated a computational analysis to predict which kind of TCSs are present in each member of this group and, more importantly, to predict the differences between the members of this group regarding these signal transduction systems.

Materials and Methods

Sequence information

Complete genome sequences of the *B. cereus* group (*B. cereus* strains ATCC 14579, ATCC 10987 and ZK, *B. thuringiensis konkukian* and *B. anthracis* strains Ames, Ames 0581, Sterne and A2012) and *B. subtilis* 168 were retrieved from the NCBI (ftp.ncbi.nih.gov/genomes/Bacteria/) on October 5, 2004. Sequence information of *B. cereus* group plasmids was obtained from the NCBI microbial plasmid database (www.ncbi.nlm.nih.gov/genomes/static/eub_p.html) on July 21, 2005. At this date, 1 plasmid of *B. cereus* ATCC 14579 (pBClin15), 1 of *B. cereus* ATCC 10987 (pBc10987), 5 of *B. cereus* ZK, 12 of *B. thuringiensis*, 6 of *B. anthracis* (3x pX01, 3x pX02) and 4 of *B. mycoides* were available.

Sequence analysis

HMMER 2.3.2 (Durbin et al., 1998) was used for hidden Markov model (HMM) searches against amino acid sequences and a DeCypher® hardware-accelerated system (Active Motif, Carlsbad, CA) was used to perform HMMER searches against nucleic acid sequences. Protein domain organisations were determined by running HMMER searches against the Pfam_ls (Bateman et al., 2004) and the SMART (Schultz et al., 1998) HMM databases, using default threshold values, while TMHMM 2.0 (Krogh et al., 2001) was used to detect transmembrane helices. Sequence similarity was detected using the NCBI protein-protein BLAST server (www.ncbi.nlm.nih.gov/BLAST/Blast.cgi) or the NCBI microbial BLAST server (www.ncbi.nlm.nih.gov/sutils/genom table.cgi). The latter server was used to scan the whole genome shotgun sequences of *B. cereus* G9241 and *B.* anthracis strains A1055, Australia 94, CNEVA-9066, Kruger B, Vollum and Western North America USA6153. Multiple sequence alignments were created with MUSCLE 3.51 (Edgar, 2004) and bootstrapped neighbour-joining (NJ) trees were created with CLUSTAL W 1.83 (Thompson et al., 1994). Trees were visualised with Levels of Orthology through Phylogenetic Trees (LOFT) (van der Heijden et al., 2007). DNA patterns were detected using PatScan (Dsouza et al., 1997).

Identification of HKs and RRs

The genome and plasmid sequences of the *B. cereus* group and *B. subtilis* 168 were searched for genes encoding putative HKs and RRs. To detect these genes, HMMER searches were performed against the protein and nucleic acid datasets of the different genomes and plasmids, using the Pfam HATPase_c (Pfam02518) and Response_reg (Pfam00072) HMMs. The HATPase_c HMM was used to scan for the highly conserved HATPase domain of HKs, while the Response_reg HMM was used to scan for the highly conserved phosphoryl-accepting domain of RRs. Recovered sequences were further scrutinised according to the following criteria: (i) the HATPase domain had to be located in the C-terminus (last 2/3) of the encoded protein and (ii) a putative H-box had to precede the HATPase domain. In case no H-box was detected, the H-box was localised by hand. HMMER searches against the *B. cereus* group nucleic acid datasets were performed to detect HK- and RR-encoding genes for which the ORF prediction was erroneous. In these cases, translation start sites were localised by hand. Frameshifts and/or overlap of more than 75 bp with an existing gene were not allowed.

Detection of HK-RR gene pairs and "orphan" HK and RR genes

HKs and their cognate RRs are often encoded on adjacent genes on the DNA. Therefore, all gene clusters containing at least one HK and one RR gene were considered to encode functional HK-RR gene pairs and were thus considered to encode a specific TCS. Single HK and RR genes were categorised as "orphans". The definition of a gene cluster was set as follows: intergenic distances within a cluster had to be less than 300 bp and genes had to lie in the same transcription direction or in a divergent direction on the DNA. A convergent direction was not allowed, since converging genes do not lie in a single operon.

Detection of orthologous TCSs

Potential protein orthologues (and in-paralogues) were automatically detected from pairwise species comparisons using INPARANOID 1.35 (Remm *et al.*, 2001). To identify orthologous HK-RR pairs between members of the *B. cereus* group and *B. subtilis*, the genomic protein datasets of these species were used as input for INPARANOID. TCSs were regarded as orthologous when both the HKs of TCSs A (species 1) and A' (species 2) and the RRs of these systems were detected as orthologues (Fig. 2.1A). When only the HKs of TCSs A and A' but not the RRs of these systems (or vice versa) were detected as orthologues, the TCSs had to share gene context to be regarded as orthologous systems (Fig. 2.1B). The rationale behind this was that gene neighbourhood has been shown to provide strong signals for functional association between gene products within and between species (Dandekar *et al.*, 1998; Overbeek *et al.*, 1999).





Fig. 2.1. Detection of orthologous TCSs using INPARANOID.

(A) TCS A of species 1 and A' of species 2 are regarded as orthologous because both the HKs of TCSs A and A' and the RRs of these systems are detected as orthologues by INPARANOID.

(B) In this situation, only the HKs of TCS A and A' are detected as orthologues. However, systems A and A' are still regarded as orthologous because they share gene context (gene X and gene X' are orthologues).

Results and Discussion

Initial identification of HKs and RRs

The Pfam HMMs HATPase_c and Response_reg were used to recover all TCSs from eight completely sequenced genomes of the *B. cereus* group. The *B. subtilis* genome was scanned in the same way for benchmarking and comparative analysis. As shown in Table 2.1, 50 to 58 putative HKs containing a C-terminal HATPase domain preceded by an H-box and 48 to 52 putative RRs containing a RR receiver domain were detected in the genomes of the *B. cereus* group. In contrast, 35 HKs and 35 RRs were found in *B. subtilis*, which is in agreement with what was found before in this organism (Fabret *et al.*, 1999). Among the total of HK and RR genes detected, sixteen had previously not been annotated due to erroneous ORF predictions (gene coordinates are shown in supplementary materials S2.1, which are included with the online version of this chapter at http://mic.sgmjournals.org). For all HKs and RRs detected, the protein domain organisation was analysed using TMHMM, Pfam and SMART. The results of these analyses are shown in supplementary materials S2.2 (http://mic.sgmjournals.org).

Around 40 HK-RR gene pairs were identified in each genome of the *B. cereus* group, which is about ten more than the number of pairs found in *B. subtilis*. Remarkable is that, in contrast to *B. subtilis*, the members of the *B. cereus*



group contain HK-RR fusion proteins, which have both a HK phosphotransferase domain and a RR phosphoryl-accepting domain. Typically, two fusion proteins were found in each of the three *B. cereus* genomes, whereas only one was found in the *B. thuringiensis* and *B. anthracis* genomes. All HK and RR genes not clustering in HK-RR gene pairs and not encoding fusion proteins were considered "orphans". As many as ten to fourteen "orphan" HKs and seven to eleven "orphan" RRs were found in the members of the *B. cereus* group, compared to six of each in *B. subtilis*. The number of HKs and RRs and their distribution among pairs, fusions and "orphans" was exactly the same for the *B. anthracis* strains, Ames, Ames 0581 and Sterne (Table 2.1). The numbers shown in Table 2.1 correspond with those of a recent, more limited, study in which only the genomes of *B. cereus* ATCC 14579, *B. anthracis* A2012 and the draft genome of *B. thuringiensis israelensis* were scanned for the amount of TCSs (Anderson *et al.*, 2005).

Table 2.1. Number of HK-RR gene pairs, fusions and "orphans" detected in eight *B. cereus* group genomes and *B. subtilis*.

Species	HKs	RRs	HK-RR pairs	HK-RR fusions	"orph	ans"
					HKs	RRs
B. cereus ATCC 14579	55	48	39	2	14	7
B. cereus ATCC 10987	54	49	40	2	12	7
B. cereus ZK	57	52	43	2	12	7
B. thuringiensis konkukian	58	52	44	1	13	7
B. anthracis Ames (0581), Sterne	52	51	41	1	10	9
B. anthracis A2012	50	50	38	1	11	11
B. subtilis 168	35	35	29	0	6	6

Since main differences between members of the *B. cereus* group have been attributed to their plasmids (Okinaka *et al.*, 1999b; Rasko *et al.*, 2005; Schnepf *et al.*, 1998), the DNA of 29 *B. cereus* group plasmids was also scanned for genes encoding TCSs. Surprisingly, only plasmid pBc10987 of *B. cereus* ATCC 10987 appeared to encode a TCS, while one "orphan" RR was found on the megaplasmid pE33L466 of *B. cereus* ZK (results not shown). Apparently, plasmid-encoded features, such as toxin production and host-specificity, are not regulated by specific plasmid-encoded TCSs.

Classification of HKs and RRs

As a basis for assigning biological functions to the HKs and RRs detected, we made a classification of these proteins. To that end, two bootstrapped NJ trees were constructed, a HK tree and a RR tree. The HK tree was constructed with all *B. cereus* group and *B. subtilis* HK phosphotransferase domains, while the RR tree was constructed with all RR receiver domains. In addition to this initial set of sequences, homologous sequences of other bacterial species were included to improve the resolution of both trees (the HK and RR tree are included as supplementary materials S2.3 and S2.4 at http://mic.sgmjournals.org). Information on the nature of the RR output domains, as identified using Pfam and SMART, was also added. Based on the two trees and the RR output domains detected, we were able to classify the *B. cereus* group HKs and RRs into the

subfamilies described by Grebe and Stock (1999), who discerned HK and RR subfamilies on a similar basis. The results of the classification procedure are shown in Table 2.2. Analysis of the two trees showed that the receiver domains of all RRs pairing to a HK of a certain class, generally clustered together in the same branches of the RR tree. Furthermore, their DNA-binding output domains also roughly fell into distinct groups. For example, all RRs pairing with a class 7 HK contained a NarL-like output domain. These results are in agreement with the findings of Grebe and Stock (1999), who suggested that the HK phosphotransferase domains, the cognate receiver domains and the RR output domains have evolved as integral units.

Function prediction: a footprint analysis including B. subtilis

Several classes of TCSs have been shown to function in distinct cellular processes. TCSs consisting of a class 4, 5, 9 and 10 HK are known to be involved in sporulation initiation, C_4 -dicarboxylate metabolism, chemotaxis and quorum-sensing, respectively (Asai *et al.*, 2000; Grebe and Stock, 1999; Jiang *et al.*, 2000; Kaspar and Bott, 2002; Lyon and Novick, 2004; Szurmant and Ordal, 2004; Tanaka *et al.*, 2003; Yamamoto *et al.*, 2000; Zientz *et al.*, 1998). The fact that members of the *B. cereus* group contain HKs of these classes strongly suggests that some of their TCSs mediate the signals that initiate the processes described above.

To get a more specific functional annotation of the *B. cereus* group HKs and RRs, they were compared with those of B. subtilis, which is the model Gram-positive organism and for which relatively much is known about the functionality of its TCSs. We used INPARANOID (Remm et al., 2001) to detect protein orthologues. With this program, the protein datasets of the B. cereus group were compared with each other and with the protein dataset of B. subtilis. From the INPARANOID output, we were able to detect HK-RR pairs, fusions and "orphans" shared between the different B. cereus group genomes and between each B. cereus group genome and B. subtilis. The resulting footprint is shown in Table 2.2. The *B. cereus* group appeared to share as many as twenty orthologous HK-RR pairs and six "orphans" with *B. subtilis*. Not all these HKs and RRs were found in every single *B. cereus* group genome. For example, the wellcharacterised B. subtilis HK CheA is absent from B. anthracis. In contrast, the well-characterised B. subtilis systems ResED, PhoRP, YycGF, YufLM, LiaSR and components of the B. subtilis sporulation initiation phosphorelay were found in all members of the B. cereus group (see Table 2.2, column 11 for biological functions). Interestingly, some well-known *B. subtilis* TCSs appeared to be absent from the *B. cereus* group. Among these were CssSR, BceSR, DesKR and DegSR.

Table 2.2. Classification, footprint analysis and function prediction of the *B. cereus* group HKs and RRs.

Column 1 contains the codes referring to the *B. cereus* group HKs and RRs. A translation of these codes to NCBI codes can be found in supplementary materials S2.5 at http://mic.sgmjournals.org.

Columns 2 and 3 show the classification of HKs and RRs, respectively. The classification into HK and RR subfamilies was based on the classification described by Grebe and Stock (1999). Null: RR does not contain an output domain.

Columns 4 to 9 show the HKs and RRs detected in each genome of the *B. cereus* group. *Bce, B. cereus; Bth, B. thuringiensis; Ban, B. anthracis*



HK-RR fusion protein

Orthologous HK-RR pairs and "orphans" detected in *B. subtilis* are boldly printed in column 10. All other, homologous, HK-RR pairs and "orphans" are in regular print. id%, amino acid identity; *, conserved gene neighbourhood with the corresponding *B. cereus* group HK and RR gene(s). Species name abbreviations: Bli, *Bacillus licheniformis*; Bsu, *Bacillus subtilis*; Cpe, *Clostridium perfringens*; Efa, *Enterococcus faecalis*; Efc, *Enterococcus faecium*; Eco, *Escherichia coli*; Lpl, *Lactobacillus plantarum*; Lmo, *Listeria monocytogenes*; Lla, *Lactococcus pneumoniae*; Spy, *Streptococcus pyogenes*.

Column 11 shows the biological functions predicted for the *B. cereus* group HKs and RRs. References: CesKR (Kallipolitis *et al.*, 2003), CroSR (Comenge *et al.*, 2003), VanSR (Arthur and Quintiliani, 2001), VanSR_b (Evers and Courvalin, 1996), YvrGH (Serizawa *et al.*, 2005), SpaKR (Klein *et al.*, 1993), NisKR (Engelke *et al.*, 1994), BacSR (Neumüller *et al.*, 2001), LisKR (Cotter *et al.*, 1999; Cotter *et al.*, 2002), ArlSR (Liang *et al.*, 2005), CiaHR (Guenzi *et al.*, 1994; Mascher *et al.*, 2003b; Throup *et al.*, 2000), LcoSR (Liu *et al.*, 2002), ResED (Nakano *et al.*, 1996), SrrBA (Yarwood *et al.*, 2001), PhoRP (Sun *et al.*, 1996), YycGF (Fabret and Hoch, 1998), VicKR (Dubrac and Msadek, 2004; Martin *et al.*, 1996), YycGF (Fabret and Hoch, 1998), VicKR (Dubrac and Msadek, 2004; Martin *et al.*, 1999; Mohedano *et al.*, 2005), GInKL (Satomura *et al.*, 2005), YvcQP, YxdKJ, LiaSR (Mascher *et al.*, 2003a; Pietiäinen *et al.*, 2005), KinA, KinB, KinC, KinD, KinE, Spo0A, Spo0F (Burbulys *et al.*, 1991; Jiang *et al.*, 2000; Trach and Hoch, 1993), DctSR (Asai *et al.*, 2003), DcuSR (Zientz *et al.*, 1998), DesKR (Aguilar *et al.*, 2001), ComPA, ComDE, AgrCA (Lyon and Novick, 2004), LamCA (Sturme *et al.*, 2005), YdfHI (Serizawa and Sekiguchi, 2005), VraSR (Kuroda *et al.*, 2000), SalKR (Upton *et al.*, 2001), LytSR (Brunskill and Bayles, 1996), CheAY, CheV (Szurmant and Ordal, 2004), RsbY (van Schaik *et al.*, 2005a).

<i>subtilis</i>) and homologous systems, predicted function	9, Cpe virulence, carbohydrate uptake/meta	o; CroSR, 34/41, Efa; VanSR, 31/41, Efc cell wall stress response, antibiotic r	o; CroSR, 32/49, Efa; VanSR*, 36/45, Efc cell wall stress response, antibiotic r	o; CroSR, 39/57, Efa; VanSR, 36/46, Efc cell wall stress response, antibiotic r	o; CroSR, 37/49, Efa; VanSR, 34/47, Efc cell wall stress response, antibiotic r	o; CroSR, 35/48, Efa; VanSR, 36/47, Efc cell wall stress response, antibiotic r	a cell wall stress response, antibiotic ri	a cell wall stress response, antibiotic ri	unknown	unknown	unknown	unknown	su lantibiotic production/resistance	su; NisKR*, 25/40, Lla lantibiotic production/resistance	su cell envelope maintenance	unknown	unknown	i lantibiotic production/resistance	Bsu; BacSR*, 28/45, Bli lantibiotic production/resistance	unknown	unknown	unknown	unknown	ı; LisKR, 26/41, Lmo; ArlSR, 27/43, Sau cell envelope maintenance, virulence	su; LisKR, 34/56, Lmo; ArISR, 31/48, Sau cell envelope maintenance, virulence
orthologous (<i>B.</i> id% HK/id% RR	CPE0235-6*, 29/4	CesKR, 33/45, Lm	CesKR, 37/47, Lm	CesKR, 40/59, Lm	CesKR, 37/46, Lm	CesKR, 35/44, Lm	VanSR _b , 30/38, Ef	VanSR _b , 25/42, Ef					SpaKR*, 24/38, B	SpaKR*, 31/52, B	YvrGH, 38/57, B			BacSR*, 54/72, B	YcbML*, 58/65,					YkoHG, 30/46, Bs	YkoHG, 28/48, E
2102A n68				Ĺ	Í																				
Ban Ames, Sterne																									
.ληολ ήτα																									
Bce ZK						_																			
28601 əว8																									
Bce 14579																									
RR output	OmpR	OmpR	OmpR	OmpR	OmpR	OmpR	OmpR	OmpR	OmpR	ı	OmpR	OmpR	OmpR	OmpR	OmpR	OmpR	OmpR	OmpR	OmpR	Null	OmpR	OmpR	OmpR	OmpR	OmpR
HK class	1a	la	la	la	1a	1a	la	1a	1a	la	la	la	1a	1a	la	la	1a	1a	1a	1a	1a	1a	1a	la	1a
ref. code	01	02	33	4)5	90	7	8(60	0	E	[2	E	4	5	۲e	[]	8	6]	20	21	22	53	4	25

27	1a	OmpR	LcoSR, 35/53, Lla	metal resistance/uptake/metabolism
28	1a	OmpR	LcoSR, 40/55, Lla	metal resistance/uptake/metabolism
29	1a	OmpR	ResED*, 47/70, Bsu; SrrBA*, 30/65, Sau	aerobic/anaerobic respiration, virulence
30	la	OmpR	ResED, 45/48, Bsu; SrrBA, 39/49, Sau	aerobic/anaerobic respiration, virulence
31	1a	OmpR	PhoRP*, 48/73, Bsu	phosphate uptake/metabolism
32	1a	OmpR	YycGF*, 54/73, Bsu; VicKR*, 46/69, Sau	fatty acid biosynthesis, virulence
33	3с	GInL∼	GlnKL, 41/53, Bsu	amino acid (Gln) uptake/metabolism
34	3c	GInL∼	GlnKL*, 49/54, Bsu	amino acid (Gln) uptake/metabolism
35	Зd	OmpR	YbdKJ*, 50/67, Bsu	unknown
36	3i	OmpR	YvcQP*, 34/60, Bsu	cell wall stress response, antimicrobial resistance
37	3i	OmpR	YxdKJ*, 38/55, Bsu	cell wall stress response, antimicrobial resistance
38	3i	OmpR	YxdKJ, 40/52, Bsu; YvcQP, 32/48, Bsu	cell wall stress response, antimicrobial resistance
39	4	ı	KinA, 33, Bsu	sporulation initiation
40	4	ı	KinC, 36, Bsu	sporulation initiation
41	4	ı	KinC, 33, Bsu	sporulation initiation
42	4	ı	KinE, 40, Bsu	sporulation initiation
43	4	ı	KinE, 34, Bsu	sporulation initiation
44	4		KinE, 34, Bsu	sporulation initiation
45	4	ı	KinE, 35, Bsu	sporulation initiation
46	4		KinC, 27, Bsu; KinE, 34, Bsu	sporulation initiation
47	4	ı	KinB, 36, Bsu	sporulation initiation
48	4	ı	KinD*, 38, Bsu	sporulation initiation
49	4		KinB, 40, Bsu	sporulation initiation
50	4	1	KinE, 37, Bsu	sporulation initiation
51	4	ı	KinB, 32, Bsu	sporulation initiation
52	4		KinB, 26, Bsu	sporulation initiation
53	ß	CitB	DctSR, 31/29, Bsu; CitAB, 32/36, Eco	C ₄ -dicarboxylate (citrate) uptake/metabolism
54	S	CitB	CitST*, 44/48, Bsu	C ₄ -dicarboxylate (citrate) uptake/metabolism
55	ß	CitB	YufLM*, 43/52, Bsu; DcuSR, 38/44, Eco	C4-dicarboxylate (malate) uptake/metabolism
56	7	NarL	YvfTU*, 43/62, Bsu; DesKR, 37/62, Bsu	membrane fatty acid saturation/desaturation
57	7	NarL	YfiJK*, 21/38, Bsu	unknown

			resistance	resistanc	resistance		/autolysis			ance					/autolysis		
			, antimicrobia	, antimicrobia	, antimicrobia	sistance	Irolase activity			ce, cell-adher		virulence			Irolase activit)		response
	npetence		ress response	ress response	ress response	production/re	of murein hyc		S	ensing, virulen		ess response,	initiation	initiation ו	of murein hyc	S	ediated stress
unknown	natural coi	unknown	cell wall st	cell wall st	cell wall st	lantibiotic	regulation	unknown	chemotaxi	duorum-se	unknown	general str	sporulatior	sporulatior	regulation	chemotaxi	sigmaB-m
	L	-	su ; VraSR, 35/60, Sau	u ; LiaSR, 48/46, Bsu	: LiaSR, 33/48, Bsu	~	isu ; LytSR*, 44/42, Sau		Bsu	n; AgrCA, 27/29, Sau; LamCA, 28/26, Lp					3, 25, Sau		
	ComPA, 26/38, Bs	YdfHI, 36/60, Bsi	LiaSR*, 42/64, B	YhcYZ, 48/46, Bs	YhcYZ, 52/51, Bsu;	SalKR*, 24/35, Sp	LytST*, 62/63, B		CheAY*, 42/67, I	ComDE, 26/31, Sp		LisR, 40, Lmo	Spo0F*, 77, Bsu	Spo0A*, 81, Bsu	LytT, 31, Bsu; Lytl	CheV, 48, Bsu	RsbY, 100, Bce
	ComPA, 26/38, Bs	YdfHI, 36/60, Bsi	LiaSR*, 42/64, B	YhcYZ, 48/46, Bs	YhcYZ, 52/51, Bsu;	SalKR*, 24/35, Sp	LytsT*, 62/63, B		CheAY*, 42/67, 1	ComDE, 26/31, Sp		LisR, 40, Lmo	Spo0F*, 77, Bsu	Spo0A*, 81, Bsu	LytT, 31, Bsu; LytI	CheV, 48, Bsu	RsbY, 100, Bce
	ComPA, 26/38, Bs	YdfHI, 36/60, Bs	LiaSR*, 42/64, B	YhcYZ, 48/46, Bs	The second s	SalKR*, 24/35, Sp	LytST*, 62/63, B		CheAY*, 42/67, 1	ComDE, 26/31, Sp		Lisk, 40, Lmo	Spo0F*, 77, Bsu	Spo0A*, 81, Bsu	LytT, 31, Bsu; Lyt	CheV, 48, Bsu	RsbY, 100, Bce
	ComPA, 26/38, Bs	YdfHI, 36/60, Bs	LiaSR*, 42/64, B	YhcYZ, 48/46, Bs	The second s	SalKR*, 24/35, Sp	LytST*, 62/63, B		CheAY*, 42/67, 1	ComDE, 26/31, Sp		Lisk, 40, Lmo	Spo0F*, 77, Bsu	Spo0A*, 81, Bsu	LytT, 31, Bsu; Lyt	CheV, 48, Bsu	RsbY, 100, Bce
NarL	NarL ComPA, 26/38, Bs	NarL YdfHI, 36/60, Bs	NarL LiaSR*, 42/64, B	NarL YhcYZ, 48/46, Bs	NarL MarL MarL NarL	NarL SalKR*, 24/35, Sp	LytTR LytST*, 62/63, B	AraC	null CheAY*, 42/67, 1	LytTR ComDE, 26/31, Sp	OmpR	OmpR LisR, 40, Lmo	null SpoOF*, 77, Bsu	Spo0A~ Spo0A*, 81, Bsu	LytTR LytTR LytT Lyt1 Bsu; Lyt1	CheW CheW CheV, 48, Bsu	PP2Csig RsbY, 100, Bce
7 NarL	7 NarL ComPA, 26/38, Bs	7 NarL YdfHI, 36/60, Bs	7 NarL A 12/64, B	7 NarL 7 Na	7 NarL 7 NarL 7 YhcYZ, 52/51, Bsu	7 NarL SalKR*, 24/35, Sp	8 LytTR 5 LytST*, 62/63, B	8 AraC	9 null CheAY*, 42/67, 1	10 LytTR ComDE, 26/31, Sp	- OmpR	- OmpR - DmpR - Lisk, 40, Lmo	- null Spo0F*, 77, Bsu	- Spo0A~ Spo0A*, 81, Bsu	- LytTR Lyt1 Lyt1	- CheW CheV, 48, Bsu	- PP2Csig RsbY, 100, Bce
Function prediction: TCSs putatively involved in (I)antibiotic resistance/ production and virulence

The *B. cereus* group HKs and RRs were also compared with those of other bacterial species, using the NCBI BLAST server. Maintaining an E-value cut-off of 1×10^{-15} , we found a number of *B. cereus* group TCSs to be similar to systems with a known biological function (Table 2.2). Among the functionally defined systems, many are known to respond to cell wall-acting antibiotics or general cell-envelope stresses, such as CesKR, CroSR, VanSR, VanSR_b and VraSR (Arthur and Quintiliani, 2001; Comenge *et al.*, 2003; Evers and Courvalin, 1996; Kallipolitis *et al.*, 2003; Kuroda *et al.*, 2000) and many are known to function in lantibiotic production and resistance, such as SpaKR, NisKR, BacSR and SalKR (Engelke *et al.*, 1994; Klein *et al.*, 1993; Neumüller *et al.*, 2001; Upton *et al.*, 2001).

We also identified TCSs putatively involved in virulence and host-microbe interactions. Among these were TCSs 24, 25 and 26, which are similar to LisKR of Listeria monocytogenes, ArISR of Staphylococcus aureus and CiaHR of streptococci (Table 2.2). LisKR plays an important role in cellular responses of L. monocytogenes to ethanol, pH, hydrogen peroxide and antimicrobials, but also contributes to the virulence potential of this organism (Cotter et al., 1999; Cotter et al., 2002). ArISR mediates the expression of many genes involved in autolysis, cell division and virulence (Liang et al., 2005) and CiaHR has been suggested to regulate maintenance of the cell envelope (e.g. modifications of peptidoglycan), virulence and repression of competence (Guenzi et al., 1994; Mascher et al., 2003b; Throup et al., 2000). Because of their similarity with LisKR, ArISR and CiaHR, it is conceivable that TCSs 24, 25 and 26 of the B. cereus group also play a role in virulence. However, the TCSs described above influence many different processes, indicating that their primary function is, for example, to maintain the cell envelope, which has great influence on the virulence potential of an organism.

Another virulence-associated system of the B. cereus group might be the class 10 TCS 68. In Gram-positive bacteria, class 10 TCSs are known as guorumsensing systems. They function as intercellular communication modules that use small peptides as signalling molecules. After processing, the peptides are exported and sensed by other cells via the sensory domains of the HK. In this way, distinct cellular processes are generated in a cell density-dependent manner (Lyon and Novick, 2004). A well-known example of a guorum-sensing system is AgrACDB of *S. aureus*. The propeptide AgrD is processed and secreted by AgrB and is then sensed by the HK AgrC. Phosphoryl is transferred to the RR AgrA, which mediates transcription of *agrACDB* and RNAIII from two promoters. RNAIII is an intracellular effector that targets the production of virulence factors (Tegmark et al., 1998). Other known agr-like systems are ComDE of streptococci and FsrCA of Enterococcus faecalis, both involved in virulence (Lyon and Novick, 2004) and LamCA of Lactobacillus plantarum, which mediates the production of surface proteins and cell-adherence (Sturme et al., 2005). Like many agr-like modules, TCS 68 of the B. cereus group might function as a quorum-sensing system, regulating the production of virulence factors and

mediating host-microbe interactions. Analysis of the *B. cereus* genomes did not reveal a putative signalling peptide-encoding gene nor an *agrB*-like gene in the near vicinity of the TCS genes, but it has to be pointed out that, across species, the signalling peptides and the AgrB-like processing enzymes often share low sequence similarity, making it difficult to detect novel ones with *in silico* techniques (Lyon and Novick, 2004).

Other important virulence regulators of the B. cereus group may be ResED, PhoRP and YycGF (TCSs 29, 31 and 32, respectively). They form a group of TCSs that is highly conserved in the low G + C Gram-positives. They were originally identified in B. subtilis and are known as important regulators of respiration, phosphate uptake and maintenance of the cell envelope (Mohedano et al., 2005; Nakano et al., 1996; Sun et al., 1996). Moreover, YycGF (VicKR) has been shown to be essential in a number of organisms (Fabret and Hoch, 1998; Martin et al., 1999; Throup et al., 2000). ResED and YycGF have also been implicated in the regulation of virulence factors in several pathogens. In S. aureus, ResED (SrrBA) represses the production of staphylococcal exotoxin and surface-associated virulence factors under low-oxygen levels (Yarwood et al., 2001), while YycGF has been shown to regulate the production of major staphylococcal surface antigens (Dubrac and Msadek, 2004). Because of the implicated role of the above-mentioned systems in virulence and because of the apparent conservation of their RR binding sites across species (Dubrac and Msadek, 2004), we scanned the B. cereus group genomes with the B. subtilis binding sites for ResD [5'-(A/T)(A/T)T(T/C)TTGT(T/G)A(A/C)-3'], PhoP [5'-TT(A/T/C)ACA-N3 to N7-TT(A/T/C)ACA-3'] and YycF [5'-TGT(A/T)A(A/T/C)-N₅-TGT(A/T)A(A/T/C)-3'] (Howell et al., 2003; Makita et al., 2004). Just as in B. subtilis, the ResD, PhoP and YycF binding sites were detected upstream of genes involved in respiration (e.g. resB), phosphate transport (e.g. pstA, pstC) and cell division (e.g. ftsE, ftsX), respectively (results not shown). Interestingly, we detected putative ResD binding sites 54 bp upstream of the hemolysin IIencoding gene of B. cereus ATCC 14579, 85 bp upstream of the hemolysin Aencoding gene of all B. cereus group genomes and 55 bp upstream of the capsule-encoding gene (capA) of plasmid pXO2. We did not find any putative PhoP or YycF binding sites upstream of genes clearly involved in virulence. The results suggest that ResED might regulate the virulence-associated genes described above. We are currently working on a more extended promoter analysis, which may shed light on the complicated transcriptional network of these RRs.

Another system that is possibly involved in virulence is TCS 01. This TCS is similar to a system of unknown function (CPE0235/CPE0236) of *Clostridium perfringens* 13 (Table 2.2). In both the HK and RR tree, the phosphoryl-transferring domains of these systems clustered closely together in distinct branches, indicating that the TCSs are highly related. Furthermore, the genes encoding the TCSs appeared to share strong gene neighbourhood conservation. Given these data, we conclude that these TCSs are specific for the *B. cereus* group and *C. perfringens* and that the shared genes lie in one operon with the TCS genes. Based on the neighbouring genes, which encode putative (carbohydrate) transport systems, and the fact that *C. perfringens* is a notorious pathogen of humans and animals, these TCSs might be virulence-associated,

functioning in the breakdown of host-tissues and the subsequent import of nutrients.

The predicted functions of the *B. cereus* group TCSs, as revealed by the comparative analyses, are shown in column 11 of Table 2.2. Column 10 and the Table legend give information on detected gene context conservation.

HK-RR fusion proteins

Although many HKs and RRs could be assigned putative biological functions, the function of a large number is still completely unknown. For instance, it is unclear what role the two HK-RR fusion proteins fulfil and whether they interact with other HKs and/or RRs. In general, HK-RR fusion proteins are involved in more complex phosphorelays (Appleby et al., 1996). Fusion protein 20, found in all members of the *B. cereus* group investigated, might function in a phosphorelay similar to the Sln1-Ypd1-Ssk1 phosphorelay of Saccharomyces cerevisiae (Posas et al., 1996). Activation of the protein probably results in phosphoryl transfer from its HK phosphotransferase domain to its own RR receiver domain. Subsequent steps may include phosphoryl transfer to the H-box of a second protein and, finally, to the RR receiver domain of a third protein that carries a RR output domain. Fusion protein 58, which was only found in B. cereus, is probably not involved in such a phosphorelay. The fact that it contains a DNAbinding domain suggests that it functions as a single unit. However, TMHMM predicted the protein to be membrane-bound (supplementary materials S2.2: http://mic.sgmjournals.org), which seems conflicting with its putative role as a transcriptional regulator. Typically, fusion protein 58 does not share any sequence similarity with other HK-RR fusion proteins, indicating that it is unique for B. cereus. To shed light on the biological role of the two HK-RR fusion proteins, we are currently investigating these B. cereus proteins in our lab (see also chapters 5-6).

Matching of "orphans"

In silico detection of HKs and RRs in members of the *B. cereus* group revealed a relatively large number of "orphans". To uncover the signal transduction routes in which these "orphans" are involved, we compared the NJ trees (HK and RR trees described above) of the interacting domains and coupled "orphans" on the basis of cognate clustering within these trees. This method was successfully employed by us before (C. Francke *et al.*, unpublished results) and it has been shown that HKs and RRs that are known to interact, fall into corresponding phylogenetic subfamilies (Grebe and Stock, 1999; Koretke *et al.*, 2000).

For most "orphans", a putative partner HK or RR could be predicted. For example, the distribution of the "orphan" class 1a HK 10 in the HK tree was identical to that of the "orphan" RR 69 in the RR tree, suggesting that HK 10 and RR 69 act together in a TCS (Fig. 2.2A). The fact that RR 69 contains an OmpR output domain strengthens this assignment, as class 1a HKs generally act with RRs containing these DNA-binding domains.

The largest group of "orphans" that were matched to partner proteins was the group of class 4 HKs (Fig. 2.2B). In *B. subtilis*, these HKs have been shown to act in the sporulation initiation phosphorelay, transferring phosphoryl to the

"orphan" RR Spo0A via the "orphan" single domain RR Spo0F and the phosphotransferase Spo0B. The multi-component structure of this transduction route provides for many levels of regulation. In addition, it includes the input of several environmental signals by the different HKs (Burbulys *et al.*, 1991; Jiang *et al.*, 2000). Orthologues of Spo0F, Spo0B and Spo0A were found in all members of the *B. cereus* group, indicating that these species use a similar phosphorelay.



Fig. 2.2. Matching of "orphans".

(A) A part of the HK tree, built with the HK phosphotransferase domains, is shown on the left side. A part of the RR tree, built with the RR receiver domains, is shown on the right side. HKs and RRs known to pair are connected with black lines. Because the "orphan" HK 10 and RR 69 (shown in grey boxes) fall into corresponding clusters in the NJ trees, they are predicted to pair (grey lines).

(B) Using a similar matching procedure, the "orphan" class 4 HKs 39 to 52 are predicted to feed phosphoryl into an extended signal transduction route, including Spo0F, Spo0B and Spo0A. PT, phosphotransferase.

(C) HK 65, which pairs with RR 65, probably also transfers phosphoryl to the "orphan" RR 73.

(D) Just as in e.g. *B. subtilis*, CheA, which pairs with CheY, probably also transfers phosphoryl to the "orphan" RR CheV. Arrows indicate predicted routes of phosphoryl transfer between the encoded proteins.

While *B. subtilis* contains five class 4 HKs (KinA, B, C, D and E), members of the *B. cereus* group contain a larger number of these HKs, suggesting that they contain an even more extended system with more signal inputs. In *B. anthracis*, nine class 4 HKs were detected, while as many as fourteen were detected in *B.*

cereus ATCC 14579. The HK tree shows that all these HKs clustered within or close to branches containing one of the *B. subtilis* sporulation HKs. In fact, they only clustered in branches containing HKs of species known to form endospores. Class 4 HK 39 clustered closest to HKs of non-spore forming bacteria, such as AtoS of *Escherichia coli*. However, overexpression of the HK 39 orthologue in *B. thuringiensis* EG1351 has been shown to bypass sporulation defects and a *spoOF* mutation in different *B. thuringiensis* strains (Malvar *et al.*, 1994). In addition, it has recently been shown that HKs 39, 40, 48 (KinD orthologue), 49 (KinB orthologue) and 50 are capable of inducing sporulation in *B. anthracis* (Brunsing *et al.*, 2005).

In addition to predicting putative partners for the "orphans" described above, putative partners were found for "orphan" RR 73 (LytT homologue) and 74 (CheV orthologue). However, these RRs were not matched to "orphan" HKs, but to HKs already found in HK-RR pairs (Fig. 2.2C and 2.2D). In the RR tree, RR 73 clustered close to a branch containing RR 65 (LytT orthologue). Since RR 73 also contains a LytTR output domain, we hypothesise that the class 8 HK 65 (LytS orthologue) is not only capable of phosphorylating its cognate RR 65, but can also transfer phosphoryl to RR 73. The fact that the RR 73-encoding gene shares gene context with LytST orthologues of other species (e.g. TTE0871/TTE0870 of Thermoanaerobacter tengcongensis MB4) and the fact that it clusters with genes putatively involved in cell envelope maintenance, the confirmed function of LytST (Brunskill and Bayles, 1996), further strengthens this prediction. Similarly, the "orphan" RR 74 (CheV orthologue) of B. cereus and B. thuringiensis was matched to the class 9 HK 67 (CheA orthologue). In the RR tree, RR 74 clustered with CheV of B. subtilis, which is known to accept phosphoryl from the chemotactic signal modulator CheA (Szurmant and Ordal, 2004). Since HK 67 clustered together with CheA in the HK tree, it is likely that phosphoryl transfer from CheA to CheV occurs in B. cereus and B. thuringiensis. In B. anthracis, a frameshift mutation has probably rendered cheA non-functional (Fig. 2.3C), leaving CheV and CheY (the RR that pairs with CheA) as "orphans". In addition to cheA, the cheV gene of B. anthracis also carries a frameshift mutation, encoding a putative CheV protein without a CheW domain. This suggests that the complete chemotaxis system of *B. anthracis* is non-functional. The fact that B. anthracis carries truncations in other genes of the flagellar gene cluster (Read et al., 2003) and the fact that most B. anthracis strains are non-motile (Turnbull, 1999), strengthens this hypothesis.

Besides CheY and CheV in *B. anthracis*, a few other "orphans" could not be matched to putative partners. For example, using the methods described above, we could not find a putative partner for the "orphan" RR RsbY (RR 75), which is responsible for activating the alternative sigma factor σ^{B} in *B. cereus* (van Schaik *et al.*, 2005a).

Differences in TCSs within the *B. cereus* group

As already mentioned, differences were found within the B. cereus group regarding the number of HK-RR fusion proteins, the amount of sporulation HKs and the chemotaxis machinery. In addition, other remarkable differences were found. Strikingly, a number of TCSs appeared to be truncated in all four B. anthracis strains (for examples, see Fig. 2.3). Besides the truncation in CheV, truncations were found in the *B. anthracis* TCSs 02, 34, 38, 43, 53 and 63. These systems were regarded as truncated since their HK sensory or their RR output domains are reduced by at least 50 amino acids as compared to their orthologues in the other B. cereus group genomes. Two other systems (TCSs 09 and 36) were not regarded as truncated in *B. anthracis*, but they differ by having a slightly shorter RR output (TCS 09) or HK sensory domain (TCS 36). Closer analysis of the B. anthracis genome sequences showed that the truncations were not caused by such trivialities as gene annotation errors. Moreover, the fact that the truncations were found in all four B. anthracis strains reduces the chance of sequencing errors as the cause for finding these truncations.

The truncations in the putative genes encoding HKs 02, 34, 38 and 63 and RR 53 presumably render their corresponding proteins non-functional, since no sensory domains are left in the HKs and no output domain is left in the RR. Interestingly, many of the truncated TCSs are similar to systems known to respond to cell wall-acting antibiotics or cell-envelope stresses in general (TCSs 02, 36, 38 and 63). Since a distinguishing feature of *B. anthracis* is its susceptibility to penicillin (Turnbull, 1999), it is possible that one (or more) of these TCSs contributes to penicillin resistance in *B. cereus* and *B. thuringiensis* and that it is indeed non-functional in *B. anthracis*. Recent work has shown that penicillin-susceptible *B. anthracis* strains contain silent β -lactamase genes, while these genes are active in penicillin-resistant members of the *B. cereus* group (Chen *et al.*, 2003b; Chen *et al.*, 2004b). Given these data, it is plausible that one or more of the non-truncated TCSs in *B. cereus* and *B. thuringiensis* provide a route for activation of the β -lactamase genes, while their truncated orthologues in *B. anthracis* are unable to activate these genes.

Although the truncations in the *B. anthracis* TCSs may indicate the inactivity of these systems, it has to be mentioned that next to the genes encoding the truncated HKs, putative genes encoding the "missing" sensory domains were found. For example, we found that the putative gene upstream of the truncated HK 63 gene actually encodes the two "missing" GAF domains (Fig. 2.3B). The presence of putative genes encoding the "missing" sensory domains leaves open the possibility that the truncated HKs are part of functional systems. It is conceivable that these HKs can somehow interact with the proteins containing their "missing" sensory domains, thereby forming three-component systems. An example of such a system might be YycHGF of *B. subtilis*. YycH, which is located external to the cell membrane, has been proposed to function as an extracellular sensor that confers its activity to the HK YycG (Szurmant *et al.*, 2005). Another possibility is that the truncated HKs are relieved from sensory constraints and are therefore more active than their non-truncated orthologues.

In addition to the truncated TCSs, some systems appeared completely absent from the four *B. anthracis* strains. Among these were, as already mentioned, fusion protein 58 (also absent from *B. thuringiensis*) and CheA, but also ComPA (TCS 59), the system that regulates natural competence in *B. subtilis* (Lyon and Novick, 2004), and the two putative sporulation HKs 46 and 47. Fragments of some of the corresponding genes were still found in the *B. anthracis* genomes. *CheA*, for example, is disrupted by a frameshift, separating the H-box- from the HATPase-encoding sequence (Fig. 2.3C).



BAS1542

BAS1543

RR 67

Fig. 2.3. Examples of truncated and degraded HKs in *B. anthracis*.

Upper genes are of *B. cereus* ATCC 14579. Lower genes are corresponding orthologues in *B. anthracis* str. Sterne. (A) The gene encoding the putative sporulation HK 43 is truncated in *B. anthracis*. However, the gene is probably still functional, since the part encoding the two PAS domains and the enzymatic HK

domains is still intact. (B) The truncation in the gene encoding HK 63 of *B. anthracis* has probably rendered this gene non-functional, since the translated HK would have no sensory domains left.

(C) A frameshift between the H-box- and the HATPase-encoding parts of *cheA* (HK 67 in *B. cereus*) has probably rendered this gene non-functional in *B. anthracis*.

To examine the nature of the TCSs described above in other B. cereus group genomes, their HK and RR protein sequences were compared to the whole genome shotgun sequences of B. cereus G9241 and B. anthracis strains A1055, Australia 94, CNEVA-9066, Kruger B, Vollum and Western North America (http:// USA6153. As shown in supplementary materials S2.6 mic.sgmjournals.org), not all the TCS truncations/deletions detected in B. anthracis strains Ames, Ames 0581, Sterne and A2012 were found in the 6 additional B. anthracis strains. Perhaps most remarkable was the detection of a complete cheV gene in all the newly sequenced B. anthracis genomes. However, all new strains (except strain A1055) do contain a disrupted cheA gene, indicating that their chemotaxis machinery is non-functional. Except from the cheV gene, most of the TCS truncations/deletions were found in the new strains,

indicating that some TCSs are generally degraded in or completely absent from *B. anthracis*.

Conclusions

In this paper we describe the results of an *in silico* comparative analysis of the TCSs of the *B. cereus* group. With the use of Pfam HMMs, 50 to 58 HKs and 48 to 52 RRs were detected in each member of the *B. cereus* group. A footprint analysis of these HKs and RRs, including those of *B. subtilis*, revealed which of these proteins are shared between the different members of the *B. cereus* group, which ones are specific for certain members and which ones are shared between the *B. cereus* group and *B. subtilis*. In addition, we were able to assign putative interaction partners for most of the "orphan" HKs and RRs detected by using a congruence-of-trees analysis.

The combination of these in silico techniques revealed interesting differences within the B. cereus group. For example, the fact that B. anthracis contains fewer class 4 "orphan" HKs than B. cereus and B. thuringiensis indicates that its sporulation initiation machinery is somewhat less fine-tuned than this mechanism is in its closest relatives. Besides the reduced number of sporulation HKs, other TCS genes appeared to be absent from or truncated in *B. anthracis*. If the truncated genes are indeed non-functional, the effective number of B. anthracis TCSs would be drastically reduced compared to the number in B. cereus and B. thuringiensis. This would suggest that B. anthracis is less capable of processing extracellular signals than its close relatives, which may proliferate in more fluctuating environments. It has been posed that B. anthracis has evolved as a pathogen of warm-blooded animals early in the B. cereus group evolution, while the other members of this group kept exploiting more fluctuating environments (e.g. invertebrate guts, plant rhizospheres and supplemented soils) (Jensen *et al.*, 2003; Turnbull, 1999). *B. anthracis* might have a more specialised pathogenic lifecycle than the other members of the B. cereus group. It probably survives in the environment mainly in the form of dormant endospores. Upon ingestion by herbivores, spores germinate to form toxin-producing vegetative cells that kill the host. Death of the host results in the release of large numbers of B. anthracis cells into the environment. These cells probably sporulate immediately upon contact with air, completing the B. anthracis life cycle (Jensen et al., 2003; Rasko et al., 2005). Specialisation of B. anthracis as a pathogen could have reduced the range of environmental stimuli to which it is exposed. This might have rendered some TCSs obsolete, ultimately resulting in the inactivation of HK and RR genes. This hypothesis is in agreement with earlier results, which showed that bacteria that inhabit relatively stable host environments generally encode fewer signalling systems than environmental bacteria with the same genome size (Galperin, 2005).

With this work, we provide the first in-depth analysis of the complete TCS arsenal of the *B. cereus* group. By scanning different *B. cereus* group genomes for HK- and RR-encoding genes, we have gained insights into the capacity of these organisms to adapt to changes in their environment. The results presented here provide a basis for future research on signal transduction mechanisms in the *B. cereus* group.

Acknowledgements

The authors would like to thank Richard Notebaart and Maarten Mols for useful discussions and Bernadet Renckens for gene context conservation analyses.

Chapter 3

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

Mark de Been, Marieke J. Bart, Tjakko Abee, Roland J. Siezen and Christof Francke

Published in: Environmental Microbiology, 2008, 10(10):2796-2809

In bacteria, environmental challenges are often translated into a transcriptional response via the cognate response regulators (RR) of specialised two component systems (TCS). A phylogenetic footprinting/shadowing approach was designed and used to identify many novel RR specific operators for species of the Bacillus cereus group and related Gram-positives. Analysis of the operator sequences revealed characteristic traits for each RR sub-family. For instance, operators related to the largest sub-family (OmpR) typically consisted of direct repeats (e.g. TTAAGA-N5-TTAAGA), whereas operators related to the second largest family (NarL) consisted of inverted repeats (e.g. ATGACA-N2-TGTCAT). This difference indicates a fundamentally different organisation of the bound RR dimers between the two sub-families. Moreover, the identification of the specific operator motifs allowed relating several RRs to a minimal regulon and thereby to a characteristic transcriptional response. Mostly, these regulons comprised genes encoding transport systems, suggesting a direct coupling of stimulus perception to the transport of target compounds. New biological roles could be attributed to various TCSs, including roles in cytochrome c biogenesis (HssRS), transport of carbohydrates, peptides and/or amino acids (YkoGH, LytSR), and resistance to toxic ions (LiaSR), antimicrobial peptides (BceRS) and beta-lactam antibiotics (BacRS, YcbLM). As more and more bacterial genome sequences are becoming available, the use of comparative analyses such as the approach applied in this study will further increase our knowledge of bacterial signal transduction mechanisms and provide directions for the assessment of their role in bacterial performance and survival strategies.

Introduction

In bacteria, two-component systems (TCS) monitor a wide range of environmental challenges (Stock *et al.*, 2000), such as nutrient deprivation, cold/heat shock or the presence of antimicrobial compounds (Aguilar *et al.*, 2001; Neumüller *et al.*, 2001; Sun *et al.*, 1996), and they mediate important cellular responses, such as chemotaxis, sporulation and biofilm formation (Jiang *et al.*, 2000; Lyon and Novick, 2004; Szurmant and Ordal, 2004). A typical TCS consists of a sensor histidine kinase (HK) and its cognate response regulator (RR), which are often encoded within the same operon. Upon sensing a specific stimulus, the often membrane-associated sensory domain of the HK induces "auto"-phosphorylation of a conserved histidine residue in the phosphotransferase domain. The histidine-bound phosphoryl is then transferred to a conserved aspartate residue within the N-terminal receiver domain of the RR. This phosphotransfer reaction activates the RR output domain.

A recent study has shown that around two-thirds of all bacterial RRs contain a DNA-binding output domain. The remaining one-third mainly consists of single receiver domain RRs, but also includes RRs with an enzymatic or RNA-, proteinor ligand-binding output domain (Galperin, 2006). Upon phosphorylation, the RR receiver domain undergoes major conformational changes, which increase the propensity of the RR to form homodimers, thereby exposing the output domains to fulfil their DNA binding task (Kern et al., 1999; Maris et al., 2002). Structural studies show that the DNA-binding domain of most RRs is composed of a helixturn-helix (HTH) fold and that slight variations occur in the structural characteristics of the HTH between different RR sub-families. An example is the "winged" HTH for RRs of the OmpR family (Martínez-Hackert and Stock, 1997). Other HTH structural variations have been observed in RRs that belong to the NarL, Fis, AraC and Spo0A family (Baikalov et al., 1996; Lewis et al., 2000; Pelton et al., 1999; Rhee et al., 1998). No detailed structural information is available for the LytTR, CitB and GInL families, although several characteristics of their DNA-binding properties have been deduced from in vitro binding studies (Koenig et al., 2004; Satomura et al., 2005; Yamamoto et al., 2000).

We have recently identified the complete set of TCSs in different members of the *B. cereus* group (de Been *et al.*, 2006). This group of low-GC Gram-positive organisms includes the medically and economically important species *B. cereus* (a food-borne pathogen), *B. thuringiensis* (an insect pathogen that is widely used as an insecticide) and *B. anthracis* (the causative agent of anthrax disease) (Rasko *et al.*, 2005). Members of the *B. cereus* group are ubiquitously present and display a high degree of adaptability (Abee and Wouters, 1999; Jensen *et al.*, 2003). This property may be explained partly by the large number of TCSs encoded on their genomes: 50 to 58 HKs and 48 to 52 RRs. Although general functions were predicted for several TCSs and experimental studies have been done on some of them (Brunsing *et al.*, 2005; Duport *et al.*, 2006; Fagerlund *et al.*, 2007; Malvar *et al.*, 1994; van Schaik *et al.*, 2005a; Vetter and Schlievert, 2007), the biological role of most of these systems remains unknown.

To find a potential role for each of the *B. cereus* group TCSs, we have identified RR specific operators and with these have predicted minimal TCS regulons. For

the detection of RR specific operators, we used a generic phylogenetic footprinting/shadowing approach, which has been formulated by us recently (Francke *et al.*, 2008). The approach is based on the general observation that gene context provides strong signals for functional associations between genes (Dandekar *et al.*, 1998; Overbeek *et al.*, 1999). We show here that: (i) the approach yields authentic operator sequences for a majority of the RRs analysed. In fact, all of the predictions that could be checked with experimental data from literature proved to be correct and (ii) that identified operator motifs could be used to search for additional binding sites on the respective genomes. This allowed the prediction of a more precise biological role for a large number of TCSs.

Materials and Methods

Sequence information

Complete genome sequences of low-GC Gram-positive bacteria (Firmicutes) and the genome sequence of the Gram-negative enterobacterium *Escherichia coli* K12 (MG1655) were retrieved from the NCBI (ftp.ncbi.nih.gov/genomes/Bacteria/). A complete list of the species analysed is provided on our website (www.cmbi.ru.nl/RRoperators/).

Sequence analysis

HMMER 2.3.2 (Durbin *et al.*, 1998) was used to perform hidden Markov model (HMM) searches. Sequence similarities were detected with BLAST 2.2.13 (Altschul *et al.*, 1997), protein domain organisations were analysed using Pfam (Bateman *et al.*, 2004) and SMART (Schultz *et al.*, 1998) and multiple sequence alignments were created with MUSCLE 3.6 (Edgar, 2004). CLUSTAL W 1.83 (Thompson *et al.*, 1994) was used to build bootstrapped neighbour-joining (NJ) trees, which were visualised with LOFT (van der Heijden *et al.*, 2007). Motif detection was performed with MEME 3.0.10 (Bailey and Elkan, 1994). The program was set to screen for 5-to-20 bp motifs that occurred on the direct strand and once (or zero times) per sequence. Alternatively, the options to screen for larger motifs, palindromes and motifs occurring multiple times per sequence were used to manually refine the detected motifs. Relevant motifs were scanned against genome sequences using MAST 3.0.10 (Bailey and Gribskov, 1998).

Identification of *B. cereus* group RR homologues

Recently, we have made a classification of all TCSs found in members of the *B. cereus* group (de Been *et al.*, 2006) following the classification scheme of Grebe and Stock (1999). In the *B. cereus* group, the RRs that contain a DNA-binding output domain belonged to one of the following domain families: OmpR, NarL, CitB, LytTR, GlnL, Spo0A and AraC. Because the DNA-binding properties of these RRs are localised in the C-terminus of the protein, the functional analysis was restricted to this domain. The domain boundaries were identified using the Pfam

Response_reg HMM (Pfam00072) (Bateman *et al.*, 2004). The collected Cterminal sequences were scanned against a non-redundant database of Firmicutes, which included the sequence information of only one strain for each bacterial species. As our main focus was on species of the *Bacillus cereus* group, the sequence information of all members of this group was added to the database. The best three BLAST hits per genome were recovered (E-value $\leq 1e^{-0^2}$). The hits were gathered into seven "RR-family" files according to the domain families mentioned earlier. In addition, the C-terminal sequences of *E. coli* RRs were retrieved from KEGG (Kanehisa *et al.*, 2004) and added. Finally, the sequences were aligned for each family, the alignments were inspected for aberrations and RR-family specific NJ trees were generated. The trees were used to determine groups of orthologous (and paralogous) RRs.

Detection of RR specific operators

To identify RR specific operators, we basically followed the approach developed by Francke et al. (2008). In this approach, putative Groups of Orthologous Functional Equivalents (GOOFEs) are defined on basis of gene context conservation. Our implementation of the approach is depicted in Fig. 3.1 (steps 1-5). Many TCSs regulate the transcriptional activity of genes located in the near vicinity of the TCS genes themselves. To find these target genes, all genes encoded close to (10 kb up- and downstream) the RR genes were collected together with their assigned COGs (Tatusov et al., 2001), as attributed within the Microbial Genome Viewer (Kerkhoven et al., 2004). COGs that were found in the gene neighbourhood of three or more RR orthologues were considered conserved COGs in this neighbourhood and the corresponding genes were considered putative RR targets. For genes with no assigned COG, homologues were retrieved and NJ trees were built. These trees were inspected to determine the conservation of these genes in the vicinity of RR orthologues. Gene context conservation analyses were performed as objectively as possible. Therefore, the orientation of the putative target genes with respect to the TCS genes was not considered. Because of their close evolutionary relationship, members of the B. cereus group were treated as a single species per orthologous branch. This step was performed to avoid unnecessary bias in the motif searching procedure (see below). The upstream regions (250 bp upstream of the translation start) of the conserved genes (or associated operons) were collected for every RR within a GOOFE. Operons were predicted as described previously (Wels et al., 2006). In case of doubt, for example when in a species two genes are considered part of an operon, while in another species the respective orthologues are adjacent, but separated by too many bp, multiple sets of upstream regions were collected. As TCSs are often autoregulatory (Bijlsma and Groisman, 2003), separate sets containing the upstream regions of the TCS operons were also collected. The upstream regions were scanned for the presence of overrepresented stretches of sequence (motifs), which can act as RR operators. The MEME Evalue threshold was set at 2.5e⁻⁰² and motifs were only approved when they were found at the same distance from the translation start (\pm 30 bp) in at least 70% of the upstream regions in a given set. All motifs thus found were scanned manually against the database of B. subtilis promoters and transcription factors

(DBTBS) (Makita *et al.*, 2004). In those cases where no putative RR operators could be detected, an additional branch of the NJ tree was included and the analysis (upstream region selection and motif detection) was repeated. This extension was only applied when the newly incorporated branch had a minimal bootstrap support of 50/1000 with the "source" branch. Finally, the regions surrounding the putative RR operators were aligned, inspected manually and submitted to DBTBS to predict sigma-factor binding sites (see our supporting website: www.cmbi.ru.nl/RRoperators/).



Fig. 3.1. Schematic representation of RR specific operator detection and the retrieval of TCS minimal regulons.

The procedure involves: an initial classification of RRs (step 1), collection of close homologues (step 2), creation of bootstrapped NJ trees for each collection of homologues (step 3), definition of groups of orthologous functional equivalents (GOOFE) based on tree topology and synteny (step 4), and ultimately, identification of RR specific operator motifs in the conserved upstream regions within each GOOFE (step 5). The identified motifs were used in genome-wide searches to obtain minimal regulons (steps 6 and 7).

Minimal regulon predictions

To predict additional RR target genes, the RR specific operator motifs were used to scan the Firmicutes genomes with MAST (Fig. 3.1, steps 6 and 7). Prior to these searches, the motif's Position-Specific Scoring Matrices (PSSM) were recalculated using a "neutral" background model (A, T, C and G background frequencies were set at 0.25). Separate matrices were built for each GOOFE. However, based on tree topology and gene context conservation, some matrices were built from multiple GOOFEs (see website: www.cmbi.ru.nl/RRoperators/). From the MAST results, sequences were retrieved that: (i) did not overlap with a gene, (ii) matched the PSSM with a P-value equal to or less than 1e⁻⁰⁸ or 5x the P-value of the least significant sequence used to build the PSSM (absolute maximum set at 5e⁻⁰⁶), and (iii) contained no mismatches in one of the important (\geq 75% conserved) residues. Important residues were defined on basis of alignments of the sites found in a footprint. When clear internal repeats were found, important residues were deduced from alignments of these repeats. In species where the RR operator found in the footprint contained a deviating residue at a conserved position, both the deviating and the consensus residue were allowed at that position. Finally, all sequences that complied with the given criteria and that were found within the promoter region of a gene and on the same strand as that gene were regarded as RR operators and the cognate genes as members of the TCS regulon.

Results and Discussion

I. RR specific operators: motifs, relative position and occurrence

In a previous study, we found that each *B. cereus* group member contains around 50 RRs, divided over 60 orthologous groups (de Been *et al.*, 2006). Of these, 55 RRs contain a DNA-binding output domain. Based on the classification scheme of Grebe and Stock (1999), these domains were classified into one of the domain families: OmpR, NarL, LytTR, CitB, GlnL, Spo0A or AraC. The main goal of this study was to identify specific operator motifs and target genes for the DNA-binding RRs of the *B. cereus* group.

Initial identification of RR specific operators

To start the identification, we collected the DNA-binding domains (C-termini) of the *B. cereus* group RRs and defined GOOFEs on basis of gene context conservation (see methods and Fig. 3.1). The upstream regions of putative RR target genes within these GOOFEs were used to predict RR specific operator motifs. The recovered motifs and the functions of the connected target genes are summarised in Table 3.1. Additional information can be found on our website (www.cmbi.ru.nl/RRoperators/).

Most of the identified RR operator motifs consisted of direct or inverted repeats with internal spacers of 2 to 11 nucleotides. In the case of OmpR-family RRs, the motifs typically consisted of direct repeats separated by four to five nucleotides, such as GTTCATA-N4-GTTCATA (*B. cereus* RR22/23). In contrast,

for NarL-family RRs, most of the identified operator motifs were inverted repeats separated by two to six nucleotides, such as ATGACA-N2-TGTCAT (RR56). LytTR-family RRs were found associated with direct repeats that were separated by ten to eleven nucleotides (e.g. ACAGTTAAGNA-N10-ACAGTTAAGNA for RR68). The direct repeats in the identified operators suggest that the corresponding RRs bind to the DNA as homodimers. In the case of OmpR- and NarL-family RRs, the length of the operators and their internal spacers imply that the RR dimers interact with two successive major groove regions on the same face of the DNA helix. Structural studies on OmpR- and NarL-family RRs indicate that this indeed is their mode of DNA binding (Blanco *et al.*, 2002; Maris *et al.*, 2002). Although LytTR-family RRs supposedly recognise the DNA in a similar fashion, the length of their operator spacers implies that these RR dimers bind two non-successive major groove regions on the same face of the helix, skipping one complete "unrecognised" major groove.

Validity of the predicted operators

In all cases where we could find appropriate literature data reporting on a RR, the predicted operator and target gene were correct (Table 3.1). For example, the operator motifs predicted for the *B. cereus* RR22/23 (OmpR), 60 (NarL) and 68 (LytTR) and their functional equivalents in other species exactly matched with experimentally verified operators for orthologous RRs (Koenig *et al.*, 2004; Serizawa and Sekiguchi, 2005; Stauff *et al.*, 2007). The operator motifs that were identified for the CitB and Spo0A RR-families (these motifs do not contain clear internal repeats) were also in perfect agreement with experimental findings. In the case of the *B. subtilis* CitB-family regulator YufM and its orthologues (*B. cereus* RR55), the specific operators seemed to be mere AT-rich stretches. Yet, these stretches probably represent genuine operators for the associated CitB-family regulators (Tanaka *et al.*, 2003). Similarly, the operator motif identified for the Spo0A RRs (*B. cereus* RR72) corresponded to the known operator sequence in *B. subtilis* (Liu *et al.*, 2003; Molle *et al.*, 2003; Strauch *et al.*, 1992).

Table 3.1. Specific operator motifs and putative target genes for the *B. cereus* group RRs.

The table is organised on basis of the RR output domain families. The operator motif logos were created with Weblogo (Crooks *et al.*, 2004) on basis of the individual sequences retrieved for each member of a particular group of orthologous functional equivalents (GOOFE). The individual sequences that contribute to each logo can be found on our website (www.cmbi.ru.nl/RRoperators/). Studies reporting on the DNA-binding characteristics of RRs of the particular GOOFE are listed in column 6. In all cases, the operator motifs complied with the experimental studies indicated. The predicted cognate target genes are listed in column 4. Autoregulatory connections are shown in red. Column 5 lists the functionalities of the genes incorporated by the minimal regulons that were defined on basis of genome-wide motif searches. Most of those connections have not been reported before (indicated in italics). Functional annotations were derived from (i) the COG database (Tatusov *et al.*, 2001), (ii) the protein domain organisations defined by Pfam (Bateman *et al.*, 2004) and SMART (Schultz *et al.*, 1998), and/or (iii) the references listed. A visual representation of the footprints can be found on our website (www.cmbi.ru.nl/RRoperators/).

Species abbreviations: BLI, Bacillus licheniformis; BSU, B. subtilis; CHY, Carboxydothermus hydrogenoformans; CAC, Clostridium acetobutylicum; CTT, C. tetani; DHA, Desulfitobacterium hafniense; EFA, Enterococcus faecalis; LSK, Lactobacillus sakei; LLA, Lactococcus lactis; MTE, Moorella thermoacetica; SAU, Staphylococcus aureus SAG, Streptococcus agalactiae; SMU, S. mutans; SPN, S. pneumoniae; THT, Thermoanaerobacter thengcongensis.

RR output family and related TCS	complete module (TCS, target gene and RR operator) conserved in:	operator motif	operators identified upstream of genes / operons encoding:	functionalities encoded in minimal regulons	references and/or studied orthologous TCSs
OmpR TCS05/06	bacilli, clostridia, DHA	ŢĨĨġŴŎŴŀŔĸĴĸŢĨġŀſġġĸĬ	RR, HK	antibiotic resistance (TCS05/06) <i>lipid transport & metabolism</i> (TCS06)	-
OmpR TCS18/19	bacilli, CTT	cel MUAAAT TT MUG	ABC-type multidrug (bacitracin) exporter	bacitracin resistance (TCS18/19) beta-lactam resistance, lipid transport & metabolism (TCS19)	- BacRS (BLI)
OmpR TCS31	lactobacilli, listeria, EFA, SPN	ADAIL AA AAIC	ABC-type phosphate transporter	phosphate transport & metabolism, <i>nucleotide transport</i> & <i>metabolism</i>	(Liu <i>et al.</i> , 1998b), PhoPR (BSU)
OmpR TCS22/23	bacilli, listeria, staphylococci, DHA, SAG		ABC-type heme efflux pump or multidrug transporter	heme-mediated toxicity resistance & cytochrome c biogenesis (TCS22)	(Stauff <i>et al.</i> , 2007), HssRS (SAU)
OmpR TCS29	bacilli		proteins involved in cytochrome c biosynthesis	aerobic/anaerobic respiration	(Geng <i>et al.</i> , 2007), ResDE (BSU)
OmpR TCS30	bacilli	A TTCAAAGT I JEITCHCA	uncharacterised conserved protein, RR , HK	1	
OmpR TCS13/14	bacilli, CTT, SMU		ABC-type multidrug (lantibiotic) transporter, <mark>RR, HK</mark>	lantibiotic resistance (not involved in lantibiotic production)	(Stein <i>et al.</i> , 2003), SpaRK (BSU)
OmpR TCS25	bacilli		secreted PepSY protein	C-3 carboxylate metabolism and/or protein degradation	
OmpR TCS36/37/38	bacilli, clostridia, staphylo- & streptococci, DHA, LSK, LLA	A sladdaa.IG als I	ABC-type cationic antimicrobial peptide (CAMP) exporter	CAMP resistance (TCS36/37/38) cell wall/membrane/envelope biogenesis (TCS37)	(Joseph <i>et al.</i> , 2004; Ohki <i>et al.</i> , 2003), BceRS (BSU)
CitB TCS55	bacilli, streptococci, EFA	œITataTaahaM	Na ⁺ / citrate symporter, malic enzyme	C-4 dicarboxylate (malate) metabolism, <i>general</i> carbohydrate metabolism	(Tanaka <i>et al.</i> , 2003), YufLM (BSU)

RR output family and related TCS	complete module (TCS, target gene and RR operator) conserved in:	operator motif	operators identified upstream of genes / operons encoding:	functionalities encoded in minimal regulons	references and/or studied orthologous TCSs
NarL TCS56	bacilli, staphylo- & streptococci, lactobacilli	ATCHCATTCOCAT La CONCATTCOCAT	ABC-type multidrug transporter, HK, RR	(anaerobic) respiration & toxin recognition/production	1
NarL -	bacilli	ATCHCAA TGTCAT #GalCh4A TGTCAT #GalCh4ggt Tg4gACT	fatty acid desaturase	regulation of membrane fatty acid saturation/desaturation	(Aguilar <i>et al.</i> , 2001; Cybulski <i>et al.</i> , 2004), DesKR (BSU)
NarL TCS57	bacilli, CAC	A GUC_AAAG_SAL	ABC-type multidrug transporter	antibiotic resistance & cell wall/membrane/envelope biogenesis	-
NarL TCS60	bacilli, CAC	OCC. AA. GTAC	predicted drug exporter of the RND superfamily or hypothetical protein	antibiotic and/or heavy metal resistance and/or lipid transport & metabolism	(Serizawa and Sekiguchi, 2005), YdfHI (BSU)
NarL TCS61	bacilli		hypothetical protein, phage shock protein A	antibiotic resistance, <i>toxic anion</i> resistance (tellurite resistance)	(Jordan <i>et al.</i> , 2006; Mascher <i>et al.</i> , 2004), LiaSR (BSU)
LytTR TCS65	bacilli, staphylo- & streptoccocci, EFA	ICOM. I CA CIII I I COMA I EA.	effector of murein hydrolase (2x)	regulation of autolysis, carbohydrate and/or inorganic ion transport & metabolism	- LytSR (SAU)
LytTR TCS68	staphylococci	ACAG TAGGIAT XIAXXXX ARCAG TAGGIA	AIP-modifying membrane protein, AIP, <mark>HK, RR</mark>	quorum-sensing, virulence, cell adherence	(Koenig <i>et al.</i> , 2004), AgrCA (SAU)
Spo0A RR72	bacilli, clostridia, CHY, DHA, MTE, THT	I ILO COMI	RR	sporulation initiation	(Strauch <i>et al.</i> , 1992), Spo0A (BSU)

Identification of additional RR operators

In most of the cases (>75%) where we found conserved gene context between orthologous RR genes, we were able to predict RR operators. Furthermore, we noticed that the operator sequence-based matrices used for the genome-wide detection of specific RR operators could also be used for the identification of operators of closely related RRs. For example, in B. licheniformis, an YvfU specific operator was detected with the search matrix describing the DesR specific operator motif. Therefore, we explored the raw MAST output data for similar sites located in the gene neighbourhood of closely related RRs. In this way, novel operators were identified for the B. cereus group RR02, 15, 16, 24 and 25 (OmpR family). Notably, RR02, 05, 06, 15, 16, 18 and 19 all seem to recognise similar TAAG-like repeats. To obtain a better insight in their overall mutual relationships, we constructed an intra-species tree, containing the complete RR sequences of the *B. cereus* group. In this tree, the "TAAG-binding" RRs clustered strongly together, forming a clear OmpR "sub-branch" (Fig. 3.2). Finally, we used DBTBS (Makita et al., 2004) to further increase the number of predicted *B. cereus* group RR operators. Operators could be identified for RR31, 32 (OmpR family), 33 (GlnL) and 53 (CitB), on basis of data available for orthologous/paralogous systems in B. subtilis: PhoP, YycF, GlnL and DctR/CitT, respectively (de Been et al., 2006). The identified operators are similar to known operators for the related B. subtilis RRs (Asai et al., 2000; Howell et al., 2003; Liu et al., 1998b; Satomura et al., 2005; Yamamoto et al., 2000) (Fig. 3.2). The RR33 and 53 operators were found by inspecting the associated TCS gene neighbourhoods. The specific operator for RR33 is located upstream of a dicistronic operon (BC1739/40) putatively encoding a Na⁺/H⁺-dicarboxylate symporter and an aspartate ammonia-lyase, while the operator for RR53 is located upstream of a putative $H^+/citrate$ symporter (BC0562). The RR31 and 32 operators were not found within the TCS gene neighbourhoods, but upstream of a pstS orthologue (BC4269) and a yocH paralogue (BC5239), respectively. In B. subtilis, PhoP is known to target the promoter region of pstS, where it binds to a TTTACA direct repeat (Liu et al., 1998b), whereas YycF is known to bind to a TGTAA(A/T) direct repeat upstream of yocH (Howell et al., 2003). Both regulatory cascades seem conserved between the *B. cereus* group and *B.* subtilis. Altogether, specific operators were predicted for around 55% of the B. cereus group RRs (Fig. 3.2).



Fig. 3.2. Intra-species NJ tree of the *Bacillus cereus* group DNA-binding RRs.

RR orthologous group numbers are shown to the right of the NCBI codes and were taken from de Been *et al.* (2006). The RR output domain families are listed on the outer right. The distinction into the domain sub-families is clearly supported by the associated bootstrap values: in case the sub-family consists of more than one member, the value connected to the "sub-family branching point" is high, in case there is only one member, the value is low (i.e. it does not clearly belong to any other sub-family). The identified operator sequences are listed next to the intra-species tree in black. Those identified using operator matrices of related RRs are indicated in bold and those found manually after comparative analysis with known *B. subtilis* motifs are shown in grey and between quotation marks. In some cases, the operator motif consensus for an orthologous RR was available (motif in italics). Related RRs are indicated to the right of the operators.

Species abbreviations: BLI, *B. licheniformis*; BSU, *B. subtilis*; EFC, *Enterococcus faecium*; LMO, *Listeria monocytogenes*; SAU, *Staphylococcus aureus*. The codes in the tree are taken from the NCBI database: BC, *B. cereus* ATCC 14579; BCE, *B. cereus* ATCC 10987; BCZK, *B. cereus* E33L; BT9727, *B. thuringiensis* konkukian; BAS, *B. anthracis* Sterne.

Autoregulation

In about one-third of the cases, we found the putative RR operator upstream of the operon containing the TCS genes themselves (Table 3.1), indicating that the associated TCSs regulate their own transcription and constitute auto-regulatory modules. Notably, the auto-regulatory nature of a TCS does not always seem conserved for orthologous TCSs. An example is formed by the *B. cereus* TCS22 (OmpR) and its orthologous systems in listeria, staphylococci and *Streptococcus agalactiae*. In *B. cereus*, listeria and staphylococci, the TCS was predicted to target a di-cistronic ABC-transporter-encoding operon, whereas in *S. agalactiae* the TCS probably targets a four-gene operon, which contains the transporter genes, but also the TCS genes themselves. This illustrates that across species, the organisation of transcription regulation of orthologous TCSs may differ significantly.

Operator positions

Given the initial constraints (see methods) most of the predicted RR operators were found within a range of 20 to 175 bp from the translation start of the target gene (Fig. 3.3). This range complies with the range in which promoter sites have been found upstream of target genes in *B. subtilis* and *E. coli* (Mwangi and Siggia, 2003; Shultzaberger *et al.*, 2007). Accordingly, in all cases where we could predict promoter sites, we found these sites close to and downstream of the associated RR operators (see website: www.cmbi.ru.nl/RRoperators/).



Fig. 3.3. Position of predicted RR operators with respect to the translation start of their cognate target genes.

The position of the operator with respect to the translation start is plotted for all operators that were used to generate the operator motifs. The distance was divided into 10 bp bins starting from the translation start site.

Prediction of TCS minimal regulons

Motifs describing the RR specific operators were scanned against Firmicutes genomes to find additional genes potentially targeted by the RR. To limit the list of predicted operators in a conservative way, we decided to use an additional criterion for approval of new putative RR operators, besides the criteria described in the materials and methods section. The criterion was based on the operator positions described above. As most of the initially identified RR operators are located within a distance of 20 to 175 bp from the translation start of the target gene, we only approved new sites when they fell within this range

from a gene start. All RR operators and their associated target genes thus found are described on our website (www.cmbi.ru.nl/RRoperators/).

We tested the validity of the criteria by comparing the predicted *B. subtilis* Spo0A minimal regulon with available experimental data. The *B. subtilis* Spo0A regulon has been characterised in great detail. Although the established regulon is much larger than the one predicted here, all Spo0A regulon members predicted by us were found to be genuine Spo0A target genes (Liu *et al.*, 2003; Makita *et al.*, 2004; Molle *et al.*, 2003). The complete correspondence between prediction and experimentally established fact strengthens the conservative approach we have followed to ascertain the detection of genuine RR target genes. A release of the strict criteria led to many false positive predictions (data not shown). It is therefore clear that additional constraints will be needed to expand the predictions reliably.

II. Inference of new biological roles for *B. cereus* group TCSs

The biological role of a TCS can be deduced from the relation between the TCS and the molecular functions encoded by its target genes. To be more precise, the functionality encoded by the predicted minimal regulons implies a characteristic biological role of the related TCS. In the following paragraphs, we will discuss several newly predicted/extended connections that we made for TCSs of the *B. cereus* group. A summary of all newly inferred functionalities is given in Table 3.1 (column 6). Although the focus is on *B. cereus* TCSs, relevant orthologous systems of other species are also described.

In many cases, we predicted transporter-encoding genes to be targeted by genomically associated TCSs. The functional annotations most commonly found for these predicted target genes were ATPase or permease components of ABC-type multidrug or antimicrobial peptide transporters. These conserved associations between TCSs and transporters suggest a direct coupling of the perception of environmental stimuli to the import/export of compounds that constitute or cause the presence of these specific stimuli.

BceRS-like modules: TCS-ABC transporter connections involved in cationic antimicrobial peptide resistance

A strong functional link between a subclass of OmpR-family RRs and particular ABC transporters was reported by Joseph *et al.* (2002). They showed that the genomic association of these TCS and transporter genes is highly conserved across bacilli and clostridia and that the transporter genes are targeted by the cognate TCSs in *B. subtilis* (i.e. YxdJK, YvcPQ, BceRS). Our study confirmed this strong association (Fig. 3.4A) as the operator predictions suggested that all the RRs of this OmpR subclass directly target the proximal transporter genes. Remarkably, the RR specific operators consisted of inverted repeats (ANCTTACA-N4-TGTAAGNT), which is unusual for OmpR-type regulators as they commonly recognise direct repeats (Blanco *et al.*, 2002) (Table 3.1). Nevertheless, the predicted operators fully complied with the experimentally determined operators for YxdJ and BceR (Joseph *et al.*, 2004; Ohki *et al.*, 2003) and were highly similar to the motif established for the homologous RR VirR of *Listeria monocytogenes* (CTNACA-N4-TGTNAG) (Mandin *et al.*, 2005). YvcPQ, YxdJK,

BceRS and VirRS, as well as the recently described homologous staphylococcal system ApsRS all mediate resistance of the bacterial cell against cationic antimicrobial peptides (CAMPs), which, in the case of listeria and staphylococci, directly influences the virulence potential of these organisms (Herbert *et al.*, 2007; Li *et al.*, 2007; Mandin *et al.*, 2005; Mascher *et al.*, 2003a; Meehl *et al.*, 2007; Ohki *et al.*, 2003; Pietiäinen *et al.*, 2005). We have previously shown that members of the *B. cereus* group encode three "BceRS-like" modules on their genome (TCS36, 37 and 38) (de Been *et al.*, 2006). Genome-wide searches with the operator motifs revealed a putative extended regulon for TCS37. Interestingly, among its putative targets were genes/operons that may function in the protection against CAMPs, such as a putative peptidoglycan transglycosylase-encoding gene (BC0541) and an operon encoding a putative ABC-type transporter (BC0814-0815) with an efflux component (BC0816) (Fig. 3.4A).

BacRS-like modules and the prediction of extended regulons involved in antimicrobial compound resistance

Another example of coupling between TCSs and ABC transporters is provided by the B. cereus group TCS19 (OmpR family) (Fig. 3.4B). An orthologous system is present in several Firmicutes and included the TCS of unknown function YcbLM of B. subtilis. Some B. cereus group members contain an additional paralogous system (TCS18). We have shown before that TCS19 and 18 are similar to BacRS of B. licheniformis ATCC 10716 (de Been et al., 2006). TCS19 and 18 and the related systems are genomically associated with operons encoding ABCtransport systems. In B. licheniformis ATCC 10716, this transport system (bcrABC) mediates protection against bacitracin and its expression is regulated by BacRS (Neumüller et al., 2001; Podlesek et al., 1995). Within the upstream regions of the transporter-encoding operons, we identified a highly conserved direct repeat ([T/G]TAAG[A/G]-N5-[T/G]TAAG[A/G]), a specific operator that has not been described before. Genome-wide searches with the operator motif revealed putative additional regulon members for TCS19 of the *B. cereus* group and its orthologous system in B. clausii. These candidate regulon members included a putative beta-lactamase- and fatty acid desaturase-encoding gene (B. cereus group), an operon in B. cereus E33L that might be a duplication of the transporter operon downstream of TCS19 and a putative phospholipid phosphatase gene in B. clausii (Fig. 3.4B). Notably, this latter gene is orthologous to the BacR target gene bcrC. The results suggest that TCS19 and the orthologous B. clausii system govern a regulon that functions in protecting the cell against different antibacterial compounds.

YvfTU and DesKR: distinct modules with similar characteristics

The *B. cereus* group TCS56 (NarL family) is another example of a TCS that is genomically associated with an ABC transport system (Fig. 3.4C). Its orthologues in other Firmicutes included the system of unknown function YvfTU of *B. subtilis*. YvfTU is highly similar to the well-characterised *B. subtilis* system DesKR, a TCS that is absent from the *B. cereus* group (de Been *et al.*, 2006). DesKR functions as a thermosensing module that controls membrane fluidity by

targeting the des gene (Aguilar et al., 2001; Cybulski et al., 2004). Within the gene neighbourhood of yvfTU and its orthologues, we found two conserved ABC transporter genes (yvfRS). Interestingly, when we extended the analysis to related TCSs, we found a similar gene arrangement in lactobacilli and streptococci. Scanning the regions upstream of all yvfRSTU(-like) operons revealed the presence of two highly conserved YvfU-specific operator sequences that were separated by six to seven bp. The site furthest away from the translation start consisted of a six bp inverted repeat with a two bp spacer (ATGACA-N2-TGTCAT), while the second one consisted of a non-palindromic sequence, containing the first ATGACA-part, but lacking the TGTCAT-part. In fact, similar sites were detected upstream of all des orthologues. Studies on the des promoter region in B. subtilis have shown that both sites are recognised by DesR dimers, and that the active RR first binds to the palindromic site. Occupancy of this site allows the binding of another DesR dimer to the second site, which may be crucial for productive contacts of DesR with RNA polymerase (Aguilar et al., 2001; Cybulski et al., 2004). The similarity between YvfU and DesR and their respective operator motifs imply that these RRs activate their target genes in a similar fashion. Scanning the Firmicutes genomes for additional YvfU operators revealed several genes with an upstream region containing the palindromic site, but lacking the non-palindromic site. In the B. cereus group, one of these genes encoded a putative transcriptional regulator of the CRP/FNR family. Notably, an orthologue in S. mutans also contained the palindrome in its promoter region, indicating that the regulatory connection is conserved to some extent (Fig. 3.4C). CRP/FNR-type regulators play an important role in (anaerobic) respiration (Körner et al., 2003) and an FNR-type regulator of B. cereus F4430/73 was shown to have strong regulatory impact on enterotoxin production (Zigha et al., 2007). Whether the B. cereus group YvfTU module fulfils a role in respiration and toxin production via the FNR-type regulator remains to be elucidated. Finally, the data presented here suggest that YvfTU and DesKR are two related, but distinct modules of which the former one is conserved across many Firmicutes.

YkoGH, a TCS involved in nutrient uptake?

An example of a regulatory connection between a TCS and a transport system that acts via a different translocation mechanism is provided by TCS25 of the *B. cereus* group (OmpR family) (Fig. 3.4D). This system is found in *B. cereus* ATCC 10987 and *B. anthracis*, and is orthologous to the system of unknown function YkoGH of *B. subtilis* (de Been *et al.*, 2006). Orthologues were also found in *B. halodurans*, *Oceanobacillus iheyensis* and *Desulfitobacterium hafniense*. Two genes, homologous to each other, were found downstream of the TCS genes in the bacilli (except in *B. cereus*). Analysis of their protein sequences indicated they encode secreted proteins with PepSY domains. These domains probably play a role in regulating glycosylase and/or protease activity in the local environment of the cell (Yeats *et al.*, 2004). Putative RR operators (TTCTNAT-N4-TTCTNAT) were found upstream of the PepSY-encoding genes located farthest downstream of the associated TCS genes (Fig. 3.4D). Notably, genomewide searches with the operator motifs revealed additional hits close to the TCS

genes of *O. iheyensis* and *B. anthracis*. In *O. iheyensis*, an additional site was found upstream of the other PepSY-encoding gene, located closest to the TCS genes. In *B. anthracis*, an additional site was found in front of a di-cistronic operon located immediately upstream of the TCS genes (Fig. 3.4D). This operon putatively encodes a glycerate kinase and a sodium/carboxylate symporter. These results suggest a link between PepSY and carboxylate metabolism. At least in *B. anthracis*, YkoGH might monitor breakdown and uptake of C-3 compounds by targeting the PepSY gene and the glycerate kinase/carboxylate symporter operon.



Fig. 3.4. Examples of TCSs, their target genes and the identified RR operators.

(A) BceRS-like modules (putatively) involved in cationic antimicrobial peptide resistance.(B) BacRS-like modules (putatively) involved in resistance to antimicrobial compounds.

(C) YvfTU- and DesKR-like modules. YvfTU of *B. cereus* and *S. mutans* are putatively involved in (anaerobic) respiration and virulence.

(D) YkoGH-like modules putatively involved in protein degradation and carboxylate metabolism (BAN).

In each panel, orthologous genes and/or genes with the same COG assignment are in similar colouring. RR and HK genes are in green and red, respectively. Genes encoding transporter components are in blue. Arrows indicate the location of the identified RR specific operators of which the sequences are displayed on the right. Putative additional members of the TCSs minimal regulons are below the dashed lines.

Abbreviations: BCE, the module is conserved in all *B. cereus* group members considered in this study (genes of *B. cereus* ATCC 14579 are shown by default); BCE^{*}, the HK gene is truncated in *B. anthracis*; BCU, *B. cereus* E33L; BTR, *B. thuringiensis* konkukian; BAN, *B. anthracis* strains Sterne, Ames and Ames 0581; BCA, *B. clausii*; BLI, *B. licheniformis*; BSU, *B. subtilis*; OIH, *Oceanobacillus iheyensis*; SAU, *Staphylococcus aureus*; SMU, *Streptococcus mutans*.

Conclusions

In this paper we describe the results of a generic phylogenetic footprinting/shadowing approach, focusing on the *B. cereus* group DNA-binding RRs. We identified a large number of RR specific operators, typically consisting of short repeated sequences separated by 2 to 11 nucleotides. A comparison of our predictions with experimental data showed that the approach retrieves authentic RR operators. In all cases where we found appropriate experimental data, we had identified the operator(s) and corresponding target gene(s) correctly.

The success of our approach is largely dependent on the occurrence of gene context conservation between orthologous regulator genes. Indeed, in most cases (>75%) where genomic associations between TCS and other genes were found conserved across species, we could retrieve RR specific operators. In various cases, knowledge of the specific operator of an evolutionary related RR could be used to guide the identification of operators for *B. cereus* group-specifc RRs. Altogether, this led to operator predictions for about 55% of the *B. cereus* group DNA-binding RRs. Considering the observed consistency of the genomic associations, we hypothesise that especially for transporter-encoding genes that directly flank a given TCS operon, chances should be high that they are targeted by the associated TCS.

Once a RR operator motif has been identified, our data indicate that it may prove fruitful to inspect the same promoter region for additional but more degenerate (low-affinity) operators. For example, multiple RR operators were found upstream of the putative target genes for DesKR, YvfTU (TCS56), YdfHI (TCS60) and TCS57 (all NarL family). The presence of multiple binding sites upstream of target genes may present a common mechanism to mediate a more tightly regulated induction of target genes. It is clear that low-affinity sites will only be occupied when the local concentration of phosphorylated RR is high enough. Induction of the controlled genes can then only be reached when the HK-activating signal is persistent.

In most documented cases, we observed that our predicted minimal regulons did not include all experimentally established regulon members. This indicates that when studying an individual TCS and its related regulon in more detail, it will be useful to employ also other criteria for the *in silico* detection of potential binding sites. Additionally, RR operator motifs could be made more degenerate by

averaging with newly detected sites. Nevertheless, as more and more bacterial genome sequences are becoming available, the use of comparative analyses such as the approach applied here will further increase our knowledge on bacterial TCS regulation and signal transduction mechanisms in general.

Acknowledgements

C. Francke was supported by the Kluyver Centre for Genomics of Industrial Fermentation, which is part of The Netherlands Genomics Initiative / Netherlands Organisation for Scientific Research.

Chapter 4

Integrating phylogenetic footprinting approaches with transcriptome profiling reveals new roles for two *Bacillus cereus* two-component signal transduction systems

Mark de Been, Julien Brillard, Véronique Brousolle, Tjakko Abee

Submitted for publication

Members of the Bacillus cereus group can adapt to a wide range of environmental challenges. In bacteria, these challenges are often translated into a transcriptional response via the cognate response regulators (RRs) of specialised two-component systems (TCSs). We have previously developed a phylogenetic footprinting approach that was successfully implemented to predict specific binding sites (operators) and target genes for the RRs of B. cereus and related species. In this study, this footprinting approach was integrated with transcriptome analyses of two B. cereus TCS deletion mutants, involving the TCSs YvrHG and YufLM. Comparison of mutant versus wild-type transcriptomes revealed that the respective TCSs were significantly active during the exponential growth phase in rich medium and that the footprinting-based operator and regulon predictions were correct for the two TCSs. Moreover, the predicted specific operators were used in combination with the transcriptome data to guide the identification of more extended TCS regulons. This revealed new roles for the respective TCSs, including the participation in an intricate transcriptional network involved in antibiotic resistance, including the confirmed resistance to oxolinic acid (YvrHG) and the confirmed uptake and metabolism of fumarate and the repression of fermentative pathways (YufLM).

Introduction

The Bacillus cereus group consists of low-GC Gram-positive, facultative anaerobic, spore-forming bacteria. It includes the six species B. cereus, B. anthracis, B. thuringiensis, B. weihenstephanensis, B. mycoides and B. pseudomycoides. The B. cereus group is recognised as a medically and economically important group of bacteria, as B. cereus is often associated with food-borne disease, B. anthracis is a pathogen of warm-blooded animals that can cause the often fatal disease anthrax, B. thuringiensis is an insect pathogen that is widely used as an insecticide, and B. weihenstephanensis is a psychrotolerant species that can spoil and poison foods at refrigerator temperatures (reviewed in Stenfors Arnesen et al., 2008). Members of the B. cereus group form a highly homogeneous subdivision within the genus Bacillus. In fact, genome sequencing data and phylogenetic studies based on chromosomal markers have revealed that B. cereus, B. anthracis and B. thuringiensis are so closely related that there is no taxonomic basis for a separate species status (Ash et al., 1991; Carlson et al., 1994; Daffonchio et al., 2000; Helgason et al., 2000; Hill et al., 2004; Ko et al., 2004; Priest et al., 2004). This is further supported by the fact that the most distinguishing features between B. cereus, B. anthracis and B. thuringiensis are encoded on mobilisable plasmids (Berry et al., 2002; Okinaka et al., 1999a; Okinaka et al., 1999b; Schnepf et al., 1998).

Members of the B. cereus group are ubiquitously present in the environment and display a high degree of adaptive properties (Abee and Wouters, 1999; Jensen et al., 2003; Kotiranta et al., 2000). These features may be at least partly explained by the relatively large number of two-component system (TCS) proteins that were recently found in each of eight B. cereus group genomes (de Been et al., 2006). In bacteria, TCSs are the dominant signalling pathways for monitoring environmental signals (Galperin, 2005), such as nutrient limitations (Sun et al., 1996), cold/heat shock (Aguilar et al., 2001) and the presence of cell envelope damaging agents (Jordan et al., 2008). In addition, TCSs can control virtually all types of microbial behaviour, including motility and chemotaxis (Wadhams and Armitage, 2004), sporulation (Stephenson and Hoch, 2002) and biofilm formation and quorum sensing (Lyon and Novick, 2004; Sturme et al., 2007). A classical TCS consists of a membrane-associated sensor histidine kinase (HK) and its cognate response regulator (RR), which are generally encoded by a pair of adjacent genes. All TCSs make use of the same mechanistic principle; that is phosphoryl transfer between the HK and the partner RR. The perception of a specific stimulus by the HK N-terminal sensory domain leads to "auto"-phosphorylation of a highly conserved histidine residue located within the downstream phosphotransferase domain. The histidine-bound phosphoryl is subsequently shuttled to a highly conserved aspartate residue located within the N-terminal RR receiver domain. This activates the RR Cterminal output domain, which generally displays DNA-binding properties. The activated RR recognises and binds to specific sites (operators) located within the promoter regions of one or more associated target genes. This allows the RR to activate or repress these genes, which function in the adaptive response towards

the initial signal perceived by the HK. Thus, TCSs provide an elegant tool for bacteria to accurately and rapidly respond to their environment (Hoch, 2000). Structural studies on RR output domains have shown that most RRs bind to their specific operators by a helix-turn-helix (HTH) fold. Slight structural variations exist between the HTH folds of RRs that belong to different subfamilies. An example is the "winged" HTH for RRs of the OmpR family (Martínez-Hackert and Stock, 1997). Other HTH structural variations have been found for RRs that belong to the families NarL, Fis, AraC and Spo0A (Baikalov et al., 1996; Lewis et al., 2000; Pelton et al., 1999; Rhee et al., 1998). However, not all RRs seem to use a HTH fold for DNA-binding as was recently demonstrated for the LytTR family RR AgrA of Staphylococcus aureus (Sidote et al., 2008). Although B. cereus group members are such important species, little is known about their TCS-mediated adaptive responses. The only experimentally studied examples include several sporulation HKs (Brunsing et al., 2005; Malvar et al., 1994), the RR RsbY that controls the activity of the alternative sigma factor $\sigma^{\dot{B}}$ (van Schaik et al., 2005a) and the TCSs SctRS (Marraffini and Schneewind, 2006), HssRS (Stauff and Skaar, 2009), ResDE (Duport et al., 2006; Esbelin et al., 2009; Vetter and Schlievert, 2007; Wilson et al., 2008) and YvfTU (Brillard et al., 2008) that are involved in the regulation of cell-surface proteins, resistance to heme toxicity, anaerobic respiration and toxin production, and the regulation of the virulence regulator PIcR, respectively. To shed more light on the role of TCSs in the B. cereus group, we have recently developed a phylogenetic footprinting approach that was used to predict specific operators and target genes for the RRs of the *B. cereus* group (de Been et al., 2008). With this approach, specific operators were predicted for over 50% of the 60 different DNA-binding RRs of B. cereus and its closest relatives. Moreover, scanning the respective genomes for sites similar to the identified operator sequences allowed relating several RRs to a minimal regulon and thereby to a characteristic transcriptional response. To further validate and elaborate on these previous predictions we present here the results of an integrative approach that combines the in silico regulon predictions with transcriptome analyses. Two B. cereus TCSs (YvrHG and YufLM), belonging to different RR families (OmpR and CitB, resp.) and for which specific minimal regulons could be predicted, were selected for genetic deletion. The transcriptome of each deletion mutant was compared to the transcriptome of wild-type B. cereus. This revealed that these TCSs were active during the exponential growth phase in "standard" rich medium and that our regulon predictions were correct for the related TCSs. Moreover, we used the predicted specific operators in combination with the transcriptome data to guide the identification of additional TCS regulon members. This revealed new biological roles for the respective TCSs, including roles in antibiotic resistance (YvrHG) and the uptake and metabolism of fumarate (YufLM).
Materials & Methods

Bacterial strains, culture media and growth conditions

B. cereus ATCC 14579 and its isogenic derivatives (Table 4.1) were cultured in Brain Heart Infusion medium (BHI, Difco, Le Pont de Claix, France) at 30°C with aeration by rotary shaking at 200 rpm. *Escherichia coli* strains were cultured in LB (MERCK, Darmstadt, Germany) at 37°C. *E. coli* TG1 (Stratagene) was used as a general purpose cloning host. To maintain selection, the medium was supplemented with kanamycin monosulfate (Sigma-Aldrich) at 100 μ g/ml, erythromycin at 10 μ g/ml for *B. cereus* and ampicillin at 100 μ g/ml for *E. coli*. The bacterial strains and plasmids used in this study are listed in Table 4.1.

rabie mil baccental beranio ana plasinas asea in enio seaayi
--

Strain or plasmid	Relevant genotype and characteristics ^a	Reference or source
<i>B. cereus</i> strains WT Δ <i>yvrHG</i>	ATCC 14579, wild type WT ^{AyvrHG} , Km ^r	lab stock (Fagerlund <i>et al.</i> , 2007)
ΔyufLM	WT ^{ΔyutLM} , Km ^r	this study
<i>E. coli</i> strains TG1	general purpose cloning host	Stratagene
plasmids		
pHT304-Km	pHT304 derivative harbouring a Km ^r cassette	(Brillard <i>et al.</i> , 2008)
pMAD	Ap ^r , Em ^r shuttle vector, thermosensitive origin of replication	(Arnaud <i>et al.</i> , 2004)
pMADΔ <i>yufLM</i>	pMAD derivative harbouring construct for the allelic replacement of <i>bc0577-78</i> (<i>yufLM</i>), Ap ^r , Em ^r , Km ^r	this study

^a Km, kanamycin; Em, erythromycin; Ap, ampicillin

Genetic methods and the generation of TCS deletion mutants

Chromosomal DNA was purified from B. cereus as previously described (Bouillaut et al., 2005). Plasmid DNA was purified from E. coli using the Wizard SV miniprep purification system (Promega, Madison, Wis., USA). Pwo polymerase (Roche Diagnostics, Meylan, France) was used to generate PCR products for cloning steps. Tag polymerase (Applied Biosystem, Courtaboeuf, France) was used for control PCRs. DNA digestion and ligation reactions were performed according to the manufacturer's instructions. The B. cereus TCS deletion strains were generated by allelic replacement of one of the respective operons (i.e. *bc1119-18* or *yvrHG* and *bc0577-78* or *yufLM*) with a kanamycin resistance (Km^r) cassette. Note that the allelic replacement procedure for the yvrHG deletion mutant ($\Delta yvrHG$) has been described previously (Fagerlund et al., 2007). A similar procedure was used to generate the *yufLM* deletion mutant. In short: the regions up- and downstream of yufLM were PCR amplified using the primer-pairs yuf_5Up-KR/yuf_3Up-KR and yuf_5Dn-KR/yuf_3Dn-KR, respectively (Table 4.2). PCR fragments were digested in accordance to the restriction sites introduced in the primers. The Km^r cassette was digested from pHT304-Km (Table 4.1). The digested DNA fragments and the Km^r cassette were

purified and cloned into the thermosensitive plasmid pMAD (Table 4.1). 10 µg of the recombinant plasmids were used to transform *B. cereus* ATCC 14579 by electroporation. Transformants were subjected to allelic exchange as described previously (Arnaud et al., 2004). Strains that were resistant to kanamycin and sensitive to erythromycin arose through a double cross-over event in which the chromosomal TCS copy was replaced with the Km^r cassette. The chromosomal allelic replacements in the TCS mutant was confirmed by DNA sequencing of PCR fragments amplified using the primer-pairs Km5out/5Up-KR and Km3out/3Dn-KR (Table 4.2). The resulting deletion mutant was designated $\Delta y u f L M$.

Table 4.2. Oligonucleotides used in this study.

oligonucleotides	sequence (5' - 3') ^a
yuf_5Up-KR	GGTA <u>AGATCT</u> ACAAAGTCAAGGTGTTGAAGAA
yuf_3Up-KR	GCTA <u>TCTAGA</u> GACTCCCCTAATTGCTGAATAG
yuf_5Dn-KR	GCTA <u>CTGCAG</u> CAAAAGAAAATAGAGTTTTACGTAT
yuf_3Dn-KR	GCTA <u>GTCGAC</u> TGGCTTCTCTGTATTTTGTTCT
Km5out	CGGTATAATCTTACCTATCACC
Km3out	ТАСТСТБАТБТТТАТАТСТТТТСТАА
allodarlingd puckati	des indicate introduced restriction sites

Underlined nucleotides indicate introduced restriction sites.

Total RNA isolation and cDNA synthesis

RNA was extracted from B. cereus ATCC 14579 and its isogenic derivatives $\Delta yvrHG$ and $\Delta yufLM$. Two independent cultures of each strain were grown to mid-exponential growth phase ($OD_{600} = 0.5-0.6$), after which 20 ml of each culture was transferred into a 50 ml Falcon tube and spun down (13,000 x g, 30 s, 4°C). Cell pellets were immediately resuspended in 1 ml cold TRI-Reagent (Ambion, Huntingdon, United Kingdom), after which the suspensions were snapfrozen in liquid nitrogen. RNA was extracted according to the manufacturer's instructions. Residual DNA was removed using TURBO DNA-free (Ambion) and extracted RNA samples were stored in 70% ethanol, 0.3 M sodiumacetate buffer (pH 5.2) at -80°C. cDNA was prepared as described previously (van Schaik et al., 2007).

Transcriptome analyses

Cy3/Cy5 labeling and purification of cDNAs were performed as described previously (van Schaik et al., 2007). Transcriptome analyses were performed using custom-made Agilent microarrays (see below), which were hybridised with 600 ng labelled cDNA. The microarray experiments were performed to compare the mid-exponential growth phase transcriptome of B. cereus ATCC 14579 against those of the two B. cereus TCS deletion mutants described above. All experiments were performed with two biological duplicates (including Cy3/Cy5 dye-swaps).

The microarrays used in this study were basically the same as those used in a previous study (van Schaik et al., 2007), except that the 18K instead of the 11K Agilent microarray platform was used. In addition, modifications to the probe design resulted in a set of 11,240 spotted probes as compared to the previous set of 10,263 probes. The new probe set represented 5249 (i.e. 98.4%) of the

5334 annotated chromosomal open reading frames of *B. cereus* ATCC 14579 (NCBI accession, NC_004722) and represented all 21 annotated open reading frames of the linear plasmid pBClin15 (NC_004721). After hybridisation at 60°C for 17 hours, the microarrays were washed with 6x SSC-0.005% Triton X-102 (10 min, 20°C), 0.1x SSC-0.005% Triton X-102 (10 min, 20°C) and 0.1x SSC-0.005% Triton X-102 (10 min, 37°C). Slides were scanned with an Agilent microarray scanner (G2565BA) (extended dynamic range scan mode) and data were processed as described previously (van Schaik *et al.*, 2007), which included the use of the web-based VAMPIRE platform (P-value threshold < 0.05) (Hsiao *et al.*, 2005).

Extending minimal regulon predictions using transcriptome data

For the prediction of RR-specific operators and TCS minimal regulons, we followed the same procedures as described previously (de Been et al., 2008; Francke et al., 2008), using an operator distance range of 25-175 bp upstream of the associated target genes. To extend the minimal regulon predictions, the *in* silico predictions and the transcriptome data were integrated by following an "in silico-based" and a "transcriptome-based" approach. For the former approach, all candidate operators with (i) a P-value < 1e⁻⁰⁷, (ii) a maximum of one mismatch in one of the important residues (de Been et al., 2008), (iii) a maximum of four mismatches with the cognate operator initially detected in the footprint and (iv) a significant up- or down-regulation of at least 1.5-fold for one or more of the associated downstream genes, were approved as RR-specific operators. The downstream genes were assigned as TCS target genes. For the "transcriptome-based" approach, the P-value threshold and the maximum number of mismatches with the cognate operator were relieved and set at $< 5e^{-1}$ ⁰⁶ and five, respectively. Conversely, the fold up- or down-regulation was increased to \geq 2-fold. In the above approaches, the distance range of the identified operators was kept at 25-175 bp, but the operator strand was relieved (i.e. allowed to be opposite from the downstream target gene). Newly predicted operators were used, together with the cognate operator, to build a new scoring matrix and to re-determine the most important operator residues. The new scoring matrix was used to scan the B. cereus genome for additional candidate operators. The above steps were repeated until no additional regulon members were detected anymore.

Phenotype analyses

High-throughput phenotypic comparisons between *B. cereus* ATCC 14579 and its isogenic derivative $\Delta yvrHG$ were performed by Biolog Inc. (Hayward, CA, www.biolog.com). Phenotypes assessed by Biolog were further tested as follows: cells were pre-cultured overnight in BHI buffered at pH 7.4 with 0.1 M KH₂PO₄/K₂HPO₄. Cells were subsequently inoculated (OD₆₀₀ = 0.01) in buffered BHI, which contained various concentrations of the substance of interest (see results). Cells were grown in 200 µl wells at 30°C without shaking.

Growth of *B. cereus* ATCC 14579 and $\Delta yufLM$ was tested in minimal medium containing different carbon sources. The minimal medium was derived from the previously described minimal medium for *B. cereus* (Mols *et al.*, 2007) and

contained: 90 mM K₂HPO₄, 105 mM KH₂PO₄, 0.8 mM MgSO₄, 0.04 mM MnCl₂, 0.2 mM NaCl, 0.2 mM CaCl₂, 0.05 mM ZnCl₂, 0.04 mM FeCl₃, 2mM glutamic acid (nitrogen source), 20 mM carbon source and 250 or 500x diluted LB medium. The carbon sources investigated were: glucose, malonic acid, malate, fumarate, succinate and citrate. Overnight cultures (BHI) were washed three times in the above minimal medium (without the carbon/nitrogen sources and LB) before inoculation (OD₆₀₀ = 0.01 or 0.02) and growth in 10 or 20 ml minimal test medium at 30°C, 200 rpm.

Results & Discussion

RR-specific operator and TCS minimal regulon predictions

We have previously identified and classified the complete set of TCSs, including 50-58 HKs and 48-52 RRs in each of eight different members of the B. cereus group (de Been et al., 2006). The majority (~90%) of the RRs was predicted as a "classical" RR with a C-terminal DNA-binding output domain. Based on the classification scheme of Grebe and Stock (1999), these output domains were classified into one of the domain families: OmpR (28-31 per genome), NarL (5-8), CitB (2-3), LytTR (2-3), GlnL (1-2), Spo0A (1) and AraC (0-1). By designing and using a phylogenetic footprinting approach, we predicted specific binding sites (operators) and cognate target genes for around 55% of these RRs (de Been et al., 2008). Genome-wide screening for sequences similar to the predicted operators allowed for relating several RRs to a minimal regulon and thereby to a characteristic transcriptional response. To further confirm and extend the RR operator and target gene predictions, we have performed and report here the transcriptome analyses of the two B. cereus TCS deletion strains $\Delta yvrHG$ and $\Delta yufLM$, which represent members of the RR families OmpR and CitB, respectively. In our previous study, we predicted RR-specific operators and associated target genes for both YvrH and YufM (de Been et al., 2008). These operators and target genes are shown in Fig. 4.1. In the case of YufM, the operator motif (i.e. scoring matrix) could be updated by including newly available genome sequence data. However, scanning the *B. cereus* genome with this newly generated operator motif did not reveal any candidate YufM target genes, except for the cognate target operon *bc0579-80*. Similarly, the predicted minimal regulon governed by YvrHG consisted of one operon, the cognate target operon *bc1120-22*.

A BC1119-18 (YvrHG)
BC1120
BCE HK RR TAAGG-N5-TTAAGG
B BC0577-78 (YufLM)
BC0579
BCE HK RR
BCA HA RE GCTTTTTTTTAATTACAAAA
OB3218
OIH
~5.3 kb BPUM_2827
BPU HK RR ACTTAATTTAGTTTAATAA
~4.0 kb BL02543
BLI HK KK GATTATTTTAATTTAAAAA
CBEI_4008 ~2.7 kb
CBE HK RR AGTATTCTTAATTTACTAA
VEF1207
ЕFA (RR НК АСТТАААТТААТТАААТАА
SAG1920
SAG RR HK GCTTAAATTAAATTAAATAA
LCA RR HK GGTTTAATAAGTTAAATAA

Fig. 4.1. (A) YvrH- and (B) YufMspecific operator predictions.

In both panels, orthologous genes are indicated with the same colourings. RRand HK-encoding genes are indicated in black. Predicted target genes flanking the respective TCS-encoding genes are indicated in grey. The locations of the conserved putative RR-specific operators are indicated by black arrows. The sequence of the detected operator is shown to the right of the gene context image.

Species abbreviations: BCE, Bacillus cereus; BCA, B. clausii; BHA, B. halodurans; BLI, B. licheniformis; BPU, B. pumilus; BSU, B. subtilis; OIH, Oceanobacillus iheyensis; AME, Alkaliphilus metalliredigens; CBE, Clostridium beijerinckii; CBO, C. botulinum; EFA, Enterococcus faecalis; LCA, Lactobacillus casei; SAG, Streptococcus agalactiae; SPY, S. pyogenes.

Confirmation of minimal regulon predictions by transcriptome analyses

To confirm the above minimal regulon predictions, we compared the transcriptomes of $\Delta yvrHG$ and $\Delta yufLM$ against the transcriptome of strain ATCC 14579. Because the specific signals that trigger the activity of the two TCSs are unknown, we chose to compare the transcriptomes of exponentially growing cells in a "standard" rich medium (BHI). When using an expression ratio cut-off of at least 2-fold up- or down-regulated, we found 10 up- and 3 down-regulated genes in WT/ Δ yvrHG and 3 up- and 29 down-regulated genes in WT/ Δ yufLM (excluding the deleted TCS genes themselves). When considering the most highly affected genes, our minimal regulon predictions appeared to be highly accurate (Table 4.3). For both WT/ Δ yvrHG and WT/ Δ yufLM, we found that the predicted cognate target genes bc1120-22 and bc0579-80, respectively, were among the most highly up-regulated genes (3.0-, 4.0-, 2.7-fold and 18.4-, 5.8-fold, respectively). The observed up-regulation suggests that these cognate target genes are positively regulated by the associated TCSs under the given conditions.



Maximising the minimal regulons by an integrative approach

Given the above data, our RR-specific operator predictions appeared to be correct. Therefore, we decided to use the predicted operator sequences in combination with the generated transcriptome data to extend the minimal regulon predictions. For this, the "in silico-based" criteria for approving RRspecific operators were slightly relieved. In other words, candidate target genes were allowed to contain slightly deviating upstream RR-specific operators, as long as they were also significantly up-, or down-regulated in WT/mutant. In this way, we extended the minimal regulons for both YvrHG and YufLM (Table 4.3). For YvrHG, as many as ten new candidate target genes were identified, divided over the four operons bc5049, bc5255-52, bc4825-23 and yvrHG itself. The yvrHG operon was added because yvrH was the most highly down-regulated gene in WT/ $\Delta yvrHG$ (0.18-fold). The yvrH regions that had been spotted on the microarray corresponded to the most upstream part of the vvrH gene, which had not been replaced by the kanamycin resistance cassette in $\Delta yvrHG$. Therefore, transcription of the upstream yvrH gene part could still be accurately measured in $\Delta y v r H G$. The observed down-regulation indicates that Yvr HG is part of a negative feedback loop. A plausible scenario is that, when YvrH binds to its cognate operator for regulation of the divergently transcribed operon bc1120-22, it simultaneously blocks its own transcription (see Fig. 4.1 for operon orientations). For the other three operons, the genes bc5049 and bc5255-52 were among the most highly up-regulated genes (2.7-4.1-fold). The upstream regions of these operons contained putative YvrH-specific operators (TTTAAGG-N4-TTTAAGA and TTTAAGA-N4-TTTAAGG, resp.) that were highly similar to the cognate operator upstream of bc1120 (Fig. 4.1A). These predicted operators were located on the strand opposite from their associated target genes, a scenario that has been found for other RRs, such as PhoP of B. subtilis (Liu et al., 1998a). Notably, further proof for bc5049 as a direct target gene for YvrHG comes from the fact that orthologues of bc5049 directly flank the equivalent yvrHG operons in B. halodurans and Lysinibacillus sphaericus. Indeed, YvrH-like operators were also identified upstream of these respective *bc5049* orthologues, bh3427 (GTAAGG-N5-TTAAGC, opposite strand) and bsph_0757 (TTAAGA-N5-TTAAGG, direct strand). Finally, for the YvrHG predicted target operon bc4825-23, we found one gene (bc4823) that was significantly affected (1.6-fold up) and a putative YvrH-specific operator (TTTAAGA-N4-ATTAAGA) was found upstream of bc4825. For YufLM, we predicted one additional target gene using the integrative approach. This gene (bc4870) was among the most highly upregulated genes in the deletion strain (5.1-fold) and contained an upstream putative YufM-specific operator (ATTTTAATTAATTACTTAA) that displayed similarity to the cognate AT-rich YufM-specific operator (Fig. 4.1B).

Table 4 transcri	l.3. ipton	TCS ne da	target Ita ana	regulons lyses.	derived	from	a a	combination	of in	, silic	8	R-specific	operator	predictions	anc

TCS and	predicted RR-specific	distance	Р-	annotation/predicted function	expression ratio
target gene	operators	to gene (bp)	value ^a		WT/mutant
YvrHG					
bc1118				Two-component histidine kinase YvrG	*
<i>bc1119</i>	TTTAAGG-N4-TTTAAGG	57	5.8e ⁻¹¹	Two-component response regulator YvrH	0.18
<i>bc1120</i>	TTTAAGG-N4-TTTAAGG	59	5.8e ⁻¹¹	Metal-dependent membrane protease	3.0
<i>bc1121</i>	I	ı	ı	Hypothetical protein	4.0
bc1122	ı	ı	ı	Hypothetical protein	2.7
bc5049	TTTAAGG-N4-TTTAAGA	79	3.5e ⁻¹⁰	Hypothetical (membrane) protein (YxaH-like)	2.7
bc5252	I	ı	ı	Hypothetical protein	4.0
bc5253	I	ı	ı	ABC-type antimicrobial peptide transport system, permease	3.9
bc5254	ı	ı	ı	ABC-type antimicrobial peptide transport system, ATPase	4.1
bc5255	TTTAAGA-N4-TTTAAGG	158	1.5e- ⁰⁹	Membrane fusion protein	3.6
bc4823	I	ı	ı	ABC-type antimicrobial peptide transport system, permease	1.6
bc4824	ı	ı	ı	ABC-type antimicrobial peptide transport system, ATPase	1
bc4825	TTTAAGA-N4-ATTAAGA	91	3.5e ⁻¹⁰	ABC-type antimicrobial peptide transport system, ATPase	ı
YufLM					
bc0579	ACTITITIAATTACAAAA	80	1.2e ⁻¹⁰	malate/Na ⁺ symporter (MaeN)	18.4
<i>bc0580</i>	I	ı	I	NAD(P)-dependent malic enzyme	5.8
bc4870	ΑΤΓΓΓΑΑΤΤΑΑΤΤΑCΤΤΑΑ	46	1.2e ⁻¹⁰	Lactate dehydrogenase (Ldh)	0.19
The diver	D-values are the ones o	ohtained w	ith the fin	al croring matrix after the lact iteration ^b - The evere	seion ratio was not

^a The given P-values are the ones obtained with the final scoring matrix, after the last iteration. ^a -, The expression ratio was not significant and is therefore not shown; *, The regions of this gene that had been spotted on the microarray were deleted in the mutant strain. Therefore, expression ratios are not shown.

Using minimal regulons to assign biological roles to TCSs

The biological role of a TCS can be deduced from the functionalities encoded by its predicted target genes. These functionalities also provide indications about the specific stimuli that trigger the associated TCS. Based on the above predicted regulons, new biological roles could be assigned to the TCSs studied here. In the following paragraphs, we describe the newly inferred functionalities for YvrHG and YufLM, which were further supported by different phenotypic comparisons.

YvrHG participates in an intricate transcriptional network of genes putatively involved in the resistance to antimicrobial compounds

In our previous study, we showed that YvrH of *B. cereus* belongs to a group of related RRs that may all be involved in the resistance of the cell to antibiotics or antimicrobial peptides (de Been et al., 2008). Therefore, YvrHG of B. cereus may fulfil a similar role. The YvrHG regulon defined in the previous section further supports this hypothesis as two of the predicted target operons encode putative ABC-type multidrug/antimicrobial peptide transport systems. One of these operons, bc4825-23, is related to the B. subtilis operons ytsCD, yvcRS and yxdLM, which are targeted by the TCSs BceRS, YvcPQ and YxdJK, respectively. These TCSs are relatively unrelated to YvrHG, but are all involved in responses to cationic antimicrobial peptides (CAMP) (Joseph et al., 2002; Joseph et al., 2004; Mascher et al., 2003a; Ohki et al., 2003; Pietiäinen et al., 2005). Interestingly, another link between YvrHG and one of the above *B. subtilis* TCSs comes from the predicted YvrHG target gene *bc5049*. This gene is paralogous to the *B. subtilis* gene *yxaH*, which was shown to be induced upon YvcP overproduction (Kobayashi et al., 2001) and which appears to contain an upstream YvcP-specific operator (AGATGACA-N4-TGTAAGGT). In addition, yxaH is a known target gene of the regulator LmrA, which is involved in multidrug resistance in B. subtilis (Yoshida et al., 2004). To confirm a possible role for YvrHG in the resistance against antimicrobial compounds, strains ATCC 14579 and $\Delta y v r H G$ were compared using phenotypic microarrays (Biolog Inc.). Indeed one compound, oxolinic acid, was found to inhibit the metabolic activity of the yvrHG deletion strain, whereas that of the wild-type strain was less affected (Fig. 4.2A). This phenotype was subsequently confirmed in several independent growth experiments (Fig 4.2B), although the appearance of the indicated phenotype is quite delicate, since the differences in sensitivity appeared not significant in two other independent experiments (results not shown). An explanation for this and for the fact that the phenotypic arrays did not reveal any other clear differences between wild-type and $\Delta yvrHG$ may be that YvrHG acts in an intricate transcriptional network that involves other regulators with functions similar to those of YvrHG. Indications for such a network are supported by our transcriptome data, where some of the most highly induced genes in $\Delta yvrHG$ included bc5285-84, bc4831, bc1357-58 and bc0816-15. These genes all encode putative ABC-type multidrug/antimicrobial peptide transporters. Interestingly, bc4831 and bc0816-15 have been predicted to be targeted by a B. cereus TCS that is orthologous to the above-mentioned YxdKJ of B. subtilis (de

Α [oxolinic acid] □ wt and ∆yvrHG 🔳 wt ■ ∆yvrHG В С 0.6 wt wt 0.6 □ ∆yvrHG $\Box \Delta yufLM$ 0.4 OD600 0.4 0.2 0.2 0 ۵ 0 ng/ml 200 ng/ml 200 mM 200 mM oxolinic acid glucose fumarate

Been *et al.*, 2008). These data suggest that deletion of *yvrHG* in *B. cereus* may be compensated by the activation of other (TCS) modules.

Fig. 4.2. Phenotypes identified for the *B. cereus yvrHG* and *yufLM* deletion strains.

(A) Phenotypic microarrays (Biolog Inc.) showed a significant decrease in metabolic activity of strain $\Delta y v r H G$ as compared to the wild-type strain at the two highest oxolinic acid concentrations after 24 hours of growth. The slightly increased metabolic activity of the mutant strain in the first two the significance was below panels threshold. Oxolinic acid was present in three-fold concentration steps with a midpoint concentration of 0.44 µg/ml (Michael Ziman, personal communication). The results show a consensus of two independent array runs.

(B) *B. cereus* wild-type and $\Delta yvrHG$ were grown in 200 µl wells, containing buffered BHI with no or 200 ng/ml oxolinic acid. After 25 hours (30°C, no shaking), growth of the mutant strain in the presence of oxolinic acid was more affected as compared to the wild-type strain. The results are from two independent experiments. For one experiment, two independent measurements were performed, while for the other one, three independent measurements were performed.

(C) *B. cereus* wild-type and $\Delta yufLM$ were grown in minimal medium with 1/500 diluted LB (30°C, 200 rpm, OD₆₀₀ at inoculation was 0.02) with either glucose or fumarate as the sole carbon source. After 46 hours, less growth was observed on fumarate for the mutant strain as compared to wild-type. The results presented are from a single experiment. Similar effects were observed in three other independent experiments (results not shown).

YufLM plays a role in aerobic growth on fumarate

Previous analyses showed that YufLM of B. cereus is orthologous to MalKR of B. subtilis and DcuSR of E. coli (de Been et al., 2006). MalKR, DcuSR and paralogous TCSs in B. subtilis and E. coli are involved in the uptake and metabolism of tricarboxylic acid (TCA) cycle intermediates (Asai et al., 2000; Kaspar and Bott, 2002; Tanaka et al., 2003; Yamamoto et al., 2000; Zientz et al., 1998). For example, MalKR responds to external malate and mediates growth on malate by targeting different transporter genes (Doan et al., 2003; Tanaka et al., 2003), while DcuSR functions in the uptake and metabolism of (mainly) fumarate (Janausch et al., 2002a; Zientz et al., 1998). The regulon predictions and transcriptome data presented here suggested that YufLM is also involved in the uptake of TCA cycle intermediates, as it appeared to positively regulate bc0579-80. BC0579 is orthologous to MaeN of B. subtilis, which functions as a malate/Na⁺ symporter (Wei et al., 2000), while BC0580 is paralogous to YtsJ of *B. subtilis*, which functions as an NADP⁺-dependent malic enzyme (Lerondel et al., 2006). These findings suggested a specific role for YufLM in malate metabolism. To test this hypothesis, the *B. cereus* wild-type and $\Delta y u f L M$ strains were grown aerobically in minimal salts medium containing

glucose or one of the carbon sources citrate, succinate, fumarate and malate. Both strains grew equally well on all sources (results not shown), except fumarate (Fig. 4.2C), suggesting a key role for YufLM in fumarate instead of malate metabolism. As described above, the predicted additional target gene bc4870 is probably negatively regulated by YufLM. Bc4870 is orthologous to the lactate dehydrogenase-encoding gene *ldh* of *B. subtilis*, which plays an important role in lactate fermentation (Cruz Ramos *et al.*, 2000; Romero *et al.*, 2007). Interestingly, other genes putatively involved in fermentation processes, including bc0612 (lactate permease), bc4996 (an *ldh* paralogue), bc0491(pyruvate-formate lyase) and bc1308 (formate transporter) were among the most highly induced genes in $\Delta yufLM$. These data suggest that YufLM plays a dual role under aerobic conditions, as it stimulates the metabolism of TCA cycle intermediates (i.e. fumarate), while it simultaneously suppresses fermentation processes, either directly (bc4870) or indirectly.

Concluding remarks

In this study, we have combined a previously designed phylogenetic footprinting approach with a transcriptome data analysis of two *B. cereus* TCS deletion mutants. Comparison of the mutant versus the wild-type transcriptomes revealed that, during the exponential growth phase in rich medium (BHI) both TCSs (i.e. YvrHG and YufLM) were significantly active. In other words, the TCS footprint-based predicted cognate target genes were significantly affected in the corresponding mutant strain as compared to wild-type. These data may suggest that detailed knowledge on a specific TCS-activating stimulus is not always required to study the function and associated targeted regulon of a given TCS. Apparently, these specific stimuli can be present in significantly high amounts in nutrient-rich media, like BHI, to trigger the activity of the TCS in exponentially growing cells.

The generated transcriptome data indicate that our phylogenetic footprint-based cognate target gene predictions were correct. This led us to use the predicted RR-specific binding sites, in combination with the generated transcriptome data, to guide the identification of additional TCS regulon members. This resulted in the identification of additional candidate target genes for both YvrHG and YufLM. In a previous study, we showed that YvrH belongs to a specific OmpR subfamily that contains nine other closely related B. cereus group RRs (de Been et al., 2008). The RRs of this specific subfamily are all related to RRs that have been implicated in mediating antibiotic resistance to the bacterial cell. Examples are VanR of Enterococcus faecium (Arthur and Quintiliani, 2001) and BacR of B. licheniformis (Neumüller et al., 2001). For six of the nine related RRs, specific operators were predicted. These operators were all similar to the predicted YvrH-specific operator, consisting of (T/G)TAAG(A/G) direct repeats, separated by five nucleotides. Because the data presented in this study support the validity of the predicted YvrH-specific operator and support a role for YvrHG in oxolinic acid resistance, they further support the existence of the specific OmpR subfamily of RRs. Finally, besides the proposed role for YvrHG, this study provides new strong leads for the functionalities of YufLM in that it does not

respond to malate, but to fumarate. In addition, YufLM of *B. cereus* seems to be involved in the repression of fermentative processes.

Acknowledgements

Julien Brillard thanks the MRI (*mission des relations internationales*) from INRA for a grant allowing him to perform part of his work in Wageningen

Chapter 5

A novel hybrid kinase is essential for activating the σ^{B} -mediated stress response of *Bacillus cereus*

Mark de Been, Marcel H. Tempelaars, Willem van Schaik, Roy Moezelaar, Roland J. Siezen, Tjakko Abee

Modified version submitted for publication

A common bacterial strategy for monitoring environmental challenges is to use two-component signal transduction systems (TCS), which consist of a sensor histidine kinase (HK) and a response regulator (RR). In the food-borne pathogen Bacillus cereus, the alternative sigma factor σ^{B} is activated by the unique RR RsbY. Here we present strong indications that the PP2C-type phosphatase RsbY receives its input from the multi-sensor hybrid kinase BC1008 (renamed RsbK). Genome analyses revealed that *rsbY* is located in a conserved gene cluster including rsbK. An rsbK deletion strain was shown to be incapable of inducing σ^{B} upon stress conditions and was impaired in its heat adaptive response. Comparison of the wild-type and *rsbK* mutant transcriptomes upon heat shock revealed that RsbK was primarily involved in the activation of the σ^{B} -mediated stress response. Truncation of the RsbK RR receiver domain demonstrated the importance of this domain for σ^{B} induction upon stress and indicated that this domain plays a role in fine-tuning RsbK kinase activity. The domain architecture of this unique sensory protein suggests that in the B. cereus group, environmental and intracellular stress signalling routes are combined into one single protein, thus markedly differing from known σ^{B} activation pathways in related low-GC Gram-positives.

Introduction

Bacteria are often exposed to strongly fluctuating environmental conditions, including changes in temperature, osmolarity, pH and the concentration of nutrients and toxic compounds. A common bacterial strategy for monitoring these challenges is to use two-component signal transduction systems (TCS). A classical TCS consists of a membrane-associated histidine kinase (HK) and its partner response regulator (RR). In essence, the TCS module harbours four different functional domains: a sensory and phosphotransferase domain in the HK and a receiver and DNA-binding output domain in the RR. Upon sensing specific environmental stimuli, the sensory domain triggers the HK to "auto"phosphorylate a conserved histidine residue in the H-box of the phosphotransferase domain. Histidine-bound phosphoryl is then transferred to a conserved aspartate residue in the RR receiver domain. This activates the RR output domain and results in binding of the RR to specific DNA target sites, thereby either activating or repressing genes involved in adaptive responses (Hoch, 2000; Stock et al., 2000).

Deviations from the TCS paradigm have been found in bacteria. A recent genome survey revealed that around 40% of all bacterial HKs do not contain any predicted transmembrane helices, suggesting that many HKs do not directly sense environmental signals (Galperin, 2005). Similarly, around 30% of all bacterial response regulators do not contain a DNA-binding output domain, but are single receiver domain RRs or contain an output domain with enzymatic or RNA-, protein-, or ligand-binding properties (Galperin, 2006). Finally, not all TCSs involve a single His-Asp phosphoryl transfer step. More complex signalling cascades are known that involve multistep His-Asp-His-Asp phosphorelays. An example is found in Bordetella pertussis, which uses the TCS BvgS-BvgA to regulate the production of virulence factors. In this TCS, the first three phosphotransfer steps occur within the hybrid kinase BvgS. This protein contains the classical HK domains, but also contains a RR receiver and Hpt-type phosphotransferase domain. Phosphoryl is first transferred from the BvgS H-box to its own RR receiver domain (His-Asp), then to a histidine residue in its Hpt domain (Asp-His), and finally to the RR receiver domain of BvgA (His-Asp) (Uhl and Miller, 1996a, 1996b). Another example is formed by the sporulation initiation pathway of Bacillus subtilis. Here, the four-step phosphorelay is divided over separate proteins, including five sporulation HKs that can feed phosphoryl into the Spo0F-Spo0B-Spo0A cascade (Burbulys et al., 1991; Jiang et al., 2000). Although this is not the case in the latter example, multistep phosphorelays often include hybrid HKs like BvgS (Appleby et al., 1996).

In previous studies, we have identified and analysed the complete set of TCSs in different members of the *Bacillus cereus* group (de Been *et al.*, 2006; de Been *et al.*, 2008). The most prominent members of this group of low-GC Gram-positive bacteria are *B. cereus* (a food-borne pathogen), *B. thuringiensis* (an insect pathogen that is widely used as a biopesticide), *B. anthracis* (the causative agent of anthrax) and *B. weihenstephanensis* (a soil psychrophile) (Rasko *et al.*, 2005). The relatively large number of TCSs that was detected in the different *B. cereus* group members may reflect the ubiquitous presence and high adaptive

capabilities of these bacteria (Jensen *et al.*, 2003). Interestingly, two different hybrid kinases and several a-typical RRs without a DNA-binding output domain were detected in the *B. cereus* group. One of these RRs, RsbY, contains a PP2C-type phosphatase domain and it was shown to be essential for the activation of the alternative sigma factor σ^{B} in *B. cereus* (van Schaik *et al.*, 2005a).

In several Gram-positive bacteria, including the model Gram-positive Bacillus subtilis, the alternative sigma factor σ^{B} plays a prominent role in redirecting gene expression under stress conditions (Kazmierczak et al., 2005; Price, 2002). In *B. cereus*, σ^{B} is also induced under a variety of stress conditions, most notably heat stress, and a *B. cereus sigB* deletion strain displayed an impaired heat adaptive response and disturbed spore properties (de Vries et al., 2005; van Schaik et al., 2004a). In B. anthracis, σ^{B} also plays a role in heat stress and a B. anthracis sigB deletion strain was slightly affected in its virulence (Fouet et al., 2000). In Gram-positive bacteria, the activity of σ^{B} is regulated by a socalled partner-switching mechanism. Under non-inducing conditions, σ^{B} is held in an inactive state by the anti-sigma factor RsbW. Upon stressful conditions, the anti-anti-sigma factor RsbV is activated and binds to RsbW, which releases σ^{B} from its inactive state complex. Released σ^{B} can bind to core RNA polymerase, resulting in the transcription of σ^{B} -dependent genes. Besides its role in sequestering σ^{B} , RsbW also functions as a kinase on RsbV. This implies a feedback mechanism on σ^{B} activation, because only the dephosphorylated form of RsbV can bind RsbW. Under stressful conditions, the main mechanism of σ^{B} activation involves the dephosphorylation of RsbV by one or more specific PP2C phosphatases (reviewed in references Kazmierczak et al., 2005; Price, 2002).

Although the partner-switching mechanism is highly conserved in species that contain σ^{B} , significant variation has been found in the N-terminal regions of PP2C phosphatases. These N-terminal regions receive signals from elements that act further upstream in the signalling route. For example, *B. subtilis* contains two σ^{B} -activating PP2C phosphatases, RsbP and RsbU. Energy stress is signalled through RsbP, which contains an N-terminal PAS sensory domain (Vijay *et al.*, 2000), while environmental stress is signalled through RsbU, which contains an N-terminal domain that interacts with the regulator RsbT. In turn, RsbT interacts with a 1.8-megadalton protein complex, called the stressosome (Chen *et al.*, 2003a; Marles-Wright *et al.*, 2008). In the *B. cereus* group, the activation of σ^{B} by the PP2C phosphatase RsbY seems to be a unique feature (van Schaik and Abee, 2005; van Schaik *et al.*, 2005a). Until now, no proteins acting upstream of RsbY have been identified, but as RsbY contains an N-terminal RR receiver domain, it is plausible that one or more partner HKs act as upstream sensors that control σ^{B} activity.

In this study, we show that one of the two hybrid HK-encoding genes of *B. cereus*, *bc1008*, is located in the direct gene neighbourhood of *rsbY* and that this gene arrangement is conserved across different bacilli outside the *B. cereus* group. We show that BC1008 plays an essential and specific role in the induction of σ^{B} upon stress exposure and that its fused RR receiver domain is essential in this process.

Materials and Methods

Sequence information and analysis

Complete bacterial genome sequences were retrieved from the NCBI (ftp.ncbi.nih.gov/genomes/Bacteria/) and information on bacterial draft genomes was retrieved via the NCBI protein and microbial BLAST servers (http://blast.ncbi.nlm.nih.gov/Blast.cgi and http://www.ncbi.nlm.nih.gov/sutils/ genom_table.cgi, respectively) on September 2008. Protein domain architectures were analysed using Pfam (Bateman *et al.*, 2004) and SMART (Schultz *et al.*, 1998), while Phobius (Käll *et al.*, 2004) was used to detect transmembrane helices. Sequence similarity was detected with BLAST 2.2.13 (Altschul *et al.*, 1997) and multiple sequence alignments were generated using MUSCLE 3.6 (Edgar, 2004). Bootstrapped neighbour-joining (NJ) trees were built with CLUSTAL W 1.83 (Thompson *et al.*, 1994) and visualised for inspection with LOFT (van der Heijden *et al.*, 2007).

Detection of *bc1008* **orthologues**

In a previous study, we established that the hybrid kinase-encoding gene bc1008 of B. cereus ATCC 14579 is conserved in seven other B. cereus group members (de Been et al., 2006). To analyse bc1008 conservation across bacteria, we performed separate BLAST runs with the BC1008 HK phosphotransferase (residues 495 to 733) and RR receiver (776 to 891) domain against all completely sequenced bacterial genomes. The reason for using only these domains was that (i) these domains have been shown to contain enough information-density for using them in classification and evolutionary studies (Fabret et al., 1999; Grebe and Stock, 1999), and (ii) it has been shown that evolutionary related HKs can contain highly different N-terminal regions due to shufflings and duplications of sensory domains (Alm *et al.*, 2006). The best three BLAST hits per genome were retrieved (E-value $\leq 1e^{-10}$), protein sequences were aligned and two bootstrapped NJ trees were built, one for the recovered HK phosphotransferase domains and one for the RR receiver domains. Proteins that clustered with BC1008 (bootstrap \geq 50%) were considered BC1008 orthologues. Finally, the analysis was extended and updated by including information on incompletely sequenced bacterial genomes. For this, we used the NCBI protein BLAST server (blastp). Novel protein hits were added to the NJ trees, which were inspected for new BC1008 orthologues. The E-value cutoff was set at $1e^{05}$ x the E-value of the best scoring sequence that did not cluster with BC1008 in the already available tree. The trees that were generated in this way can be found in supplementary figures S5.1 and S5.2. (http://www.cmbi.ru.nl/ ~mdebeen/Thesis). Draft *B. cereus* group genomes with no available protein annotations were also screened for BC1008 orthologues. For this, we used the NCBI microbial BLAST server (tblastn). For the detection of other orthologies (see results) we used similar approaches as described above.

Bacterial strains, culture media and growth conditions

B. cereus ATCC 14579 and its isogenic derivatives were cultured in Brain Heart Infusion medium (BHI, Difco, Le Pont de Claix, France) at 30°C with aeration by shaking at 200 rpm. *Escherichia coli* strains were cultured in LB (MERCK, Darmstadt, Germany) at 37°C. *E. coli* DH5a (Sambrook *et al.*, 1989) was used as a general purpose cloning host, whereas *E. coli* HB101/pRK24 (Trieu-Cuot *et al.*, 1991) was used as the donor host in conjugation experiments. To maintain selection, the medium was supplemented with antibiotics at the following concentrations: ampicillin (Sigma, Zwijndrecht, The Netherlands) at 50 µg/ml, erythromycin (Sigma) at 150 µg/ml (for *E. coli*) or 5 µg/ml (for *B. cereus*), spectinomycin (Sigma) at 100 µg/ml, chloramphenicol (Sigma) at 5 µg/ml and polymyxin B (VWR, Amsterdam, The Netherlands) at 50 µg/ml. The bacterial strains and plasmids used in this study are listed in Table 5.1.

-		
B. cereus strains		
WT	ATCC 14579, wild type	lab stock
FM1400	$WT^{\Delta sigB}$, Em^r	Van Schaik <i>et al.</i> (2004a)
FM1401	WT ^{∆rsbY} , Em ^r	Van Schaik <i>et al.</i> (2005a)
FM1404	WT ^{∆bc1008} , Em ^r	this study
FM1404ev	FM1404/pHP13, Em ^r , Cm ^r	this study
FM1404compl	FM1404/pHP13-bc1008, Em ^r , Cm ^r	this study
FM1404trunc	FM1404/pHP13-bc1008trunc, Em ^r , Cm ^r	this study
WT-18Z	WT/pHT304-18Z	this study
WT-18Zpfr1 to pfr7	WT/pHT304-18Zpfr1 to pfr7, Ap ^r , Em ^r	this study
E. coli strains		
DH5a	general purpose cloning host	Sambrook <i>et al.</i> (1989)
HB101/pRK24	donor host in conjugation experiments	Trieu-Cuot <i>et al.</i> (1991)
plasmids		
pAT∆S28	tra ⁺ conjugative suicide vector for <i>B. cereus</i> group, Sp ^r	Trieu-Cuot <i>et al.</i> (1990)
pAT∆S28-Em	pAT Δ S28 derivative, harbouring erythromycin resistance cassette, Sp ^r , Em ^r	Van Schaik <i>et al.</i> (2004a)
pAT∆S28-Em-	pATAS28-Em derivative harbouring 1.4 kb upstream	this study
bc1008up	flanking region of <i>bc1008</i> , Sp ^r , Em ^r	-
pAT∆bc1008	$pAT\Delta S28$ -Em-bc1008up derivative harbouring 1.4 kb downstream flanking region of <i>bc1008</i> , Sp ^r , Em ^r	this study
pHP13	complementation vector, Cm ^r , Em ^r	Haima <i>et al.</i> (1987)
pHP13-bc1008full	pHP13 derivative harbouring a 3.2 kb fragment, comprising the full-length <i>bc1008</i> gene and 347 bp of its upstream region, Cm ^r , Em ^r	this study
pHP13-bc1008trunc	pHP13 derivative harbouring a 2.6 kb fragment, comprising a truncated <i>bc1008</i> gene and 347 bp of its upstream region, Cm ^r , Em ^r	this study
pHT304-18Z	cloning vector harbouring a \textit{lacZ} reporter gene, Ap^r, Em^r	Agaisse and Lereclus (1994)
pHT304-18Zpfr1 to pfr7	pHT304-18Z derivatives harbouring <i>bc1009</i> and <i>bc1008</i> promoter fragments upstream of the <i>lacZ</i> gene, Ap ^r , Em ^r	this study

^aEm^r, erythromycin resistance; Cm^r, chloramphenicol resistance; Sp^r, spectinomycin resistance; Ap^r, ampicillin resistance.

Genetic methods and the generation of a *bc1008* deletion mutant

Chromosomal DNA was purified from B. cereus with the Wizard genomic DNA purification kit (Promega, Madison, Wis., USA). Plasmid DNA was purified from *E. coli* with a Qiaprep Spin Miniprep kit (Westburg, Leusden, The Netherlands). Pwo polymerase (Roche Diagnostics, Almere, The Netherlands) was used to generate PCR products for cloning steps. Taq polymerase (Fermentas, Amersfoort, The Netherlands) was used for control PCRs. DNA digestion and ligation reactions were performed according to the manufacturer's instructions. To generate a *B. cereus bc1008* deletion strain, the *bc1008* gene was replaced with an erythromycin resistance cassette. First, a 1.4 kb bc1008 upstream flanking region was amplified from *B. cereus* ATCC 14579 DNA using primers bc1008up_fwd and bc1008up_rev (Table 5.2). The PCR product was digested with XbaI (Fermentas) and cloned into vector pATAS28-Em, which harbours an erythromycin resistance cassette (van Schaik et al., 2004a), to generate pAT Δ S28-Em-bc1008up. As bc1008up could insert into pAT Δ S28-Em in two ways, the proper orientation (i.e. Em-rev-bc1008up-fwd) was confirmed with PCR and restriction digestion analyses. Secondly, a 1.4 kb bc1008 downstream flanking region was amplified using primers bc1008do_fwd and bc1008do_rev (Table 5.2). The PCR product was digested with BamHI and EcoRI (Fermentas) and cloned into pATAS28-Em-bc1008up to generate pATAbc1008. This vector was transformed into E. coli HB101/pRK24 and the resulting strain was used as the donor host in conjugation experiments with *B. cereus* ATCC 14579. Conjugation was performed following the standard protocol for conjugative plasmid transfer from E. coli to Gram-positive bacteria (Bron, 1990). Polymyxin B was used for counterselection against E. coli after conjugation. B. cereus transconjugants were screened for erythromycin resistance and spectinomycin sensitivity. PCR analyses confirmed the deletion of the *bc1008* gene in the selected transconjugants and showed that the erythromycin cassette had recombined into the chromosome via a double-crossover event (results not shown). The resulting bc1008 deletion mutant was designated B. cereus FM1404 (Table 5.1).

Complementation of the *bc1008* deletion mutant and truncation of *bc1008*

To complement *B. cereus* FM1404, the full-length *bc1008* gene, including 347 bp of its upstream region, was amplified using primers bc1008fwd and bc1008full_rev (Table 5.2). The resulting fragment (bc1008full) was digested with PstI and Cfr9I/XmaI (Fermentas) and cloned into the low-copy complementation vector pHP13 (Haima *et al.*, 1987) to generate pHP13-bc1008full (confirmed by DNA sequencing). *B. cereus* FM1404 was transformed with pHP13 (empty vector) and pHP13-bc1008full by electroporation (Bone and Ellar, 1989). Selected chloramphenicol-resistant colonies were designated *B. cereus* FM1404ev and FM1404compl, respectively (Table 5.1). Similarly, a *B. cereus* strain, encoding a truncated BC1008 protein that lacks the C-terminal RR receiver domain was generated. Primers bc1008fwd and bc1008trunc_rev (Table 5.2) were used to amplify a truncated *bc1008* fragment. Note that the latter primer harboured a mismatching adenine that introduced an in-frame stop

codon in the *bc1008* reading frame. Bc1008trunc was cloned into pHP13 and the resulting plasmid (pHP13-bc1008trunc) was introduced into *B. cereus* FM1404 to generate strain FM1404trunc (Table 5.1).

Table 5.2. Oligonucleotides used in this study.

oligonucleotides	Sequence (5' - 3') ^a
bc1008up_fwd	ACGG <u>TCTAGA</u> GCGAATAGTATAAGCACTACAAGTG
bc1008up_rev	ACGG <u>TCTAGA</u> GGCATTCGAAATATAGAGTGAAAC
bc1008do_fwd	TCCG <u>GGATCC</u> GCAACGATAGAAGATGTTATT
bc1008do_rev	GCCA <u>GAATTC</u> GCGCAATCAGATGATATGTG
bc1008fwd	GTGG <u>CTGCAG</u> GTAAACATGGAACACTTACTTCTA
bc1008full_rev	CTGG <u>CCCGGG</u> CGATTACAAATCCTTCTATAAAT
bc1008trunc_rev	CTGG <u>CCCGGG</u> TCTTCAACAACTT A AGCAGGAATAA
bc1008_GSP1	ATCTGTTTTAATTTTTCTTGCTG
bc1008_GSP2	TTCGTAATCCTTTTCCGCAGAGT
pbc1009fwd_12	GCTA <u>AAGCTT</u> CATAGACATCTTACAATTCATCTT
pbc1009rev_1	GCTA <u>GGATCC</u> TTCCATATAGTAAGCATAACAATA
pbc1009Rtrev_2	GACT <u>GGATCC</u> CAAACTAGAAATCCCCCTA
pbc1008fwd_34	GACT <u>AAGCTT</u> ATACAATAGGGGGATTTCTA
pbc1008rev_3	GACT <u>GGATCC</u> CATTAAAAATTCCTACCTTCA
pbc1008fwd_5	GACT <u>AAGCTT</u> AAATAGAAAAGTTTTAATAACCTT
pbc1008fwd_6	GACT <u>AAGCTT</u> TAACCTTGAAGGTAGGAAT
pbc1008fwd_7	GCAT <u>AAGCTT</u> CAAAATTTAGTATTCGCTACA
pbc1008rev_4567	GACT <u>GGATCC</u> AAACAAGTAAGAACAAAATGATA

^aUnderlined nucleotides indicate introduced restriction sites. The bold "A" indicates an inframe introduction of a stop-codon.

Effect of *bc1008* deletion and truncation on σ^{B} induction upon stress exposure

Two independent cultures of *B. cereus* ATCC 14579 and its genetic derivatives FM1404, FM1404ev, FM1404compl and FM1404trunc were grown to midexponential growth phase ($OD_{600} = 0.5$ -0.6) and were exposed to a temperature upshift from 30 to 42°C or to 2.5% (wt/vol) NaCl. Cells were harvested before and at 5, 10 and 30 minutes after stress exposure. Total cellular protein was extracted by bead beating as described previously (Periago *et al.*, 2002). The protein extracts were used in western blotting experiments with anti- σ^{B} (van Schaik *et al.*, 2004a) and anti-BC1008 antiserum. The latter antiserum was obtained by raising anti-peptide polyclonal antibodies in mice (Eurogentec S.A., Seraing, Belgium). The serum against which the antibodies were raised contained the two synthesised peptide stretches FADQGADVGKQSIKM (BC1008 residues 315-329) and QQADGATIRKYGGTG (683-697). Immunoblotting was performed as described previously (van Schaik *et al.*, 2004a). Immunocomplexes were visualised with SuperSignal West Pico Chemiluminescent Substrate (Perbio, Etten-Leur, The Netherlands).

Heat resistance assay for *B. cereus* cells

B. cereus ATCC 14579 and FM1404 cells were grown to mid-exponential growth phase and pre-exposed to mild heat treatment at 42°C for 30 minutes, after which they were exposed to a lethal temperature of 50°C. Survival at this temperature was determined at 0, 10, 20, 30 and 40 minutes after exposure by

plating appropriate dilutions on BHI agar plates followed by overnight incubation at 30°C. For all exposures, three independent experiments were performed and samples were plated in duplicate for each time point.

Total RNA isolation and cDNA synthesis

RNA was extracted from *B. cereus* ATCC 14579 and strain FM1404. Two independent cultures of each strain were exposed to a mild heat shock of 42°C for 0, 5, 10, or 30 minutes as described above, after which 20 ml of each culture was transferred into a 50 ml Falcon tube and spun down (13,000 x g, 30 s, 4°C). Cell pellets were immediately resuspended in 1 ml cold TRI-Reagent (Ambion, Huntingdon, United Kingdom), after which the suspensions were snap-frozen in liquid nitrogen. Within 60 minutes after freezing, RNA was extracted according to the manufacturer's instructions and residual DNA was removed using TURBO DNA-*free* (Ambion). Extracted RNA samples were stored in 70% ethanol, 0.3 M sodiumacetate buffer (pH 5.2) at -80°C. cDNA was prepared as described previously (van Schaik *et al.*, 2007).

Microarray analyses

Cy3 and Cy5 labelling and purification of cDNAs were performed as described previously (van Schaik *et al.*, 2007). For transcriptome analyses, custom-made Agilent microarrays (see below) were hybridised with 600 ng labelled cDNA. Microarray experiments were performed to compare the heat shock (42°C, 10 min) transcriptomes of strains ATCC 14579 and FM1404. Additional, microarray experiments were performed to determine the set of genes that was differentially expressed in *B. cereus* ATCC 14579 upon a temperature upshift from 30°C to 42°C (10 min). This latter experiment had been performed previously (van Schaik *et al.*, 2007) and was re-performed mainly for comparative purposes and because the microarrays used in this study represented a larger part of the annotated *B. cereus* genome than the previous ones (see below). All experiments were performed with two biological duplicates (including Cy3/Cy5 dye-swaps).

The microarrays used in this study were essentially the same as those used in a previous study (van Schaik et al., 2007), except that we used the 18K instead of the 11K Agilent Technologies microarray platform. In addition, some modifications were made to the previous probe design, resulting in a total number of 11,240 spots. These spots represented 5249 (i.e. 98.4%) of the 5334 annotated chromosomal open reading frames (ORFs) of B. cereus ATCC 14579 (see NCBI accession number NC_004722 for ORF-calling). In addition, all 21 annotated ORFs of the linear plasmid pBClin15 (NC_004721) and 82 putative small RNA-encoding genes extracted from Rfam (Griffiths-Jones et al., 2005) were represented on the microarray. All ORFs for which probes could be designed were represented by two probes. For the majority of these ORFs (98.2%) two different non-overlapping probes were designed, while for the remaining set of ORFs only one unique probe could be designed. After hybridisation at 60°C for 17 hours, the microarrays were washed with 6x SSC-0.005% Triton X-102 (10 min, 20°C), 0.1x SSC-0.005% Triton X-102 (10 min, 20°C) and 0.1x SSC-0.005% Triton X-102 (10 min, 37°C), according to the

manufacturer's instructions. Slides were scanned with an Agilent microarray scanner (G2565BA) (extended dynamic range scan mode) and data were processed (including a Lowess normalisation step) and extracted with Agilent feature extraction software (version 9.5.1.1). After removal of the data for the control spots, the normalised data were analysed for statistical significance with the web-based VAMPIRE platform (Hsiao et al., 2005), using a false discovery rate threshold of 0.05. Further data analysis was performed as has been described previously (van Schaik et al., 2007).

5' RACE and promoter deletion analyses

The 5' ends of the bc1008 transcript were mapped using RNA isolated from B. cereus ATCC 14579 cells that had been exposed to a mild heat shock for 0, 5, 10 or 30 minutes. The 5' transcript ends were determined with the 5' RACE System for Rapid Amplification of cDNA Ends (Invitrogen, Breda, The Netherlands) using the GSP primers bc1008_GSP1 and bc1008_GSP2 (Table 5.2), according to the manufacturer's instructions. To further study the bc1008 promoter activity, β-galactosidase assays were performed. For this, different bc1008, but also bc1009 (control) promoter regions were PCR amplified. The following primer combinations were used: pbc1009fwd 12 with pbc1009rev 1 to generate fragment 1, pbc1009fwd_12 with pbc1009RTrev_2 (fragment 2), pbc1008fwd_34 with pbc1008rev_3 (fragment 3), and pbc1008rev_4567 with one of the primers pbc1008fwd_34, pbc1008fwd_5, pbc1008fwd_6, and pbc1008fwd_7 to generate fragments 4, 5, 6 and 7, respectively. The primers used are listed in Table 5.2. The resulting fragments were digested, according to the restriction sites introduced in the primers, and were cloned into the low-copy plasmid pHT304-18Z (Agaisse and Lereclus, 1994) to create lacZ transcriptional fusions. All constructs were confirmed by DNA sequencing. The empty plasmid pHT304-18Z and the resulting recombinant plasmids (pHT304-18Zpfr1 to pfr7, Table 5.1) were introduced into *B. cereus* ATCC 14579 by electroporation (Bone and Ellar, 1989). The resulting B. cereus strains were cultured to midexponential growth phase and were then exposed to a 42°C heat shock (30 min) or grown for another 24 hours at 30°C, after which cells were harvested. β -Galactosidase specific activitities were measured and expressed in Miller units as described before (Agaisse and Lereclus, 1994). For all measurements, two independent experiments were performed.

Microarray data accession number

Microarray data have been submitted to the Gene Expression Omnibus repository (http://www.ncbi.nlm.nih.gov/projects/geo), under accession number GSE14589.

Results

BC1008 is a putative membrane-associated multi-sensor hybrid kinase

In a previous study, we detected 50 to 58 HK- and 48 to 52 RR-encoding genes in each of eight completely sequenced members of the *B. cereus* group (de Been *et al.*, 2006). Two genes encoding putative hybrid kinases, which contain a HK phosphotransferase as well as a RR receiver domain, were also detected. Interestingly, one of these genes (*bc1008* in *B. cereus* ATCC 14579) was found in the gene neighbourhood of the operons encoding the alternative sigma factor σ^{B} and its three regulatory proteins, RsbV, RsbW and the response regulator RsbY (Fig. 5.1A). Considering that (i) gene context provides strong signals for functional associations between genes (Dandekar *et al.*, 1998; Overbeek *et al.*, 1999), (ii) response regulators generally receive their input from a partner histidine kinase, and (iii) *bc1008* has recently been assigned as a member of the σ^{B} regulon (van Schaik *et al.*, 2007), we hypothesised a specific regulatory link between the hybrid kinase BC1008 and σ^{B} via the response regulator RsbY.

With its 896 amino acids, BC1008 is by far the largest HK of the B. cereus group. Protein sequence analysis indicated that it functions as a membraneassociated multi-sensor HK. Within the BC1008 N-terminal region, we detected a putative CHASE3 domain. These domains have been predicted to function as extracellular sensory units, but the environmental signals that trigger them remain unknown (Zhulin et al., 2003). The CHASE3 domain is flanked by two putative transmembrane helices. As predicted by LocateP (Zhou et al., 2008), the N-terminal helix is part of a signalling peptide that ensures Sec-dependent localisation towards the cell membrane. Towards the C-terminus, the helixenclosed CHASE3 domain is followed by a HAMP linker, a GAF sensory domain, the HK phosphotransferase domain and the RR receiver domain (Fig. 5.1B). HAMP domains are thought to link N-terminal sensory domains with intracellular phosphotransferases (Aravind and Ponting, 1999; Hulko et al., 2006), while GAF domains are known as intracellular sensory units that bind small ligands (Anantharaman et al., 2001; Martinez et al., 2002). In addition, we found that BC1008 contains two putative signalling (S)-helices, one that slightly overlaps with the C-terminus of the CHASE3 domain and one that is located between the GAF and the phosphotransferase domain. The S-helix, a conserved helical domain, has recently been described by Anantharaman and co-workers (Anantharaman et al., 2006) and was hypothesised to function as a switch that prevents constitutive activation of linked downstream signalling domains. Based on its domain architecture, we conclude that BC1008 functions as a membraneassociated dual sensor kinase that monitors environmental stimuli by means of its CHASE3 domain and intracellular stimuli by means of its GAF domain.

Bc1008 is located in a conserved gene cluster

In our previous study, we showed that orthologues of *bc1008* were conserved in all eight *B. cereus* group members that were analysed, including *B. thuringiensis* and *B. anthracis* (de Been *et al.*, 2006). We also found a *bc1008* orthologue in the recently sequenced genome of *B. weihenstephanensis* KBAB4. In all these genomes, the *bc1008* orthologues were found close to the *sigB* operon (Fig.

5.1C). Analysis of 20 other available *B. cereus* group genomes, including 18 draft genomes, revealed this genetic makeup to be highly conserved within the *B. cereus* group (results not shown). Only one member, *B. cereus* subsp. *cytotoxis* NVH 391-98, lacked the hybrid kinase gene. Interestingly, this species also lacks the *sigB* operon as has been reported previously (Lapidus *et al.*, 2008).



Fig. 5.1. *Bc1008* gene context conservation and BC1008 domain architecture.

(A) In *Bacillus cereus* ATCC 14579 (BCE), the hybrid kinase-encoding gene *bc1008* is located close to the operons encoding the alternative sigma factor σ^{B} and its three regulatory proteins, RsbV, RsbW and the RR RsbY.

(B) Domain architecture of BC1008. HK and RR phosphotransferase domains are indicated by light-grey rectangles, sensory domains are indicated by dark grey ellipses and linker domains are shown as pentangles: S, Signalling (S)-helix; H, HAMP. Transmembrane helices (TMH) are indicated by black rectangles. The proposed topology is indicated above the detected domains.

(C) Within the *B. cereus* group, *bc1008* orthologues were always found close to the genes encoding σ^{B} and its regulatory proteins: BTH, *B. thuringiensis* konkukian 97-27; BAN, *B. anthracis* Sterne; BWE, *B. weihenstephanensis* KBAB4.

(D) *Bc1008* orthologues were found in the gene neighbourhood of *rsbY* and *bc1007* orthologues in four other low-GC Gram-positives: BCO, *B. coagulans* 36D1; LSP, *Lysinibacillus sphaericus* C3-41; B14905, *Bacillus* sp. B14905; B14911, *Bacillus* sp. NRRL B-14911. Conserved orthologues are indicated by similar colouring.

The observed genomic association is a strong indicator of a regulatory connection between BC1008 and σ^{B} . To obtain further indications for this potential connection, we searched for *bc1008* orthologues in all available, non-B. cereus group, bacterial genome sequences. Orthologues were found in Bacillus coagulans 36D1, Lysinibacillus sphaericus C3-41, Bacillus sp. B14905, and Bacillus sp. NRRL B-14911. Sequence analysis of these orthologues suggested that they contain the same domain architecture and topology as BC1008, except that the proteins of B. coagulans and Bacillus NRRL B-14911 probably do not contain the CHASE3 domain and the associated S-helix (results not shown). No sigB-, rsbV- or rsbW-like genes were found within the genomic regions directly flanking the bc1008 orthologues. However, within these regions, we did find orthologues of rsbY, but also of bc1007, a gene that directly flanks bc1008 in B. cereus and that encodes a putative methylase of chemotactic signal receptors (Fig. 5.1D). The conserved genetic context between *bc1008* and *rsbY* further supported a functional link between them and therefore further supported our proposed regulatory link between BC1008 and σ^{B} .

BC1008 is essential for proper σ^{B} induction under stress conditions

To confirm the proposed regulatory link between BC1008 and σ^{B} , a *B. cereus* bc1008 deletion strain, FM1404, was constructed by allelic replacement of bc1008 with an erythromycin resistance cassette. By means of immunoblotting using anti- σ^{B} antiserum, this mutant strain was tested for its capability to induce σ^{B} protein levels upon mild stress conditions. In previous studies, it was shown that especially a temperature upshift of 30°C to 42°C and, to a lesser extent, a 2.5% NaCl salt shock caused a rapid induction of σ^{B} protein levels when applied to exponentially growing cells (van Schaik *et al.*, 2004a; van Schaik *et al.*, 2005a). Therefore, these particular stress conditions were chosen to test the σ^{B} inducing capabilities of strain FM1404. As shown in figure 5.2A, σ^B protein levels were highly induced upon mild heat shock in wild-type cells, whereas this σ^{B} induction was completely abolished in the bc1008 deletion strain. However, low and constant levels of σ^B were still observed in the deletion strain throughout the different time points. At non-stress conditions, σ^B levels were even slightly elevated in the deletion strain as compared to wild-type. Upon mild salt stress conditions, a slight induction of σ^{B} levels was observed in wild-type cells after 30 min. Again, this induction was abolished in the deletion strain. Instead, constant σ^{B} levels were observed in the deletion strain throughout the different time points, similar to what was observed for heat stress. The above data illustrated that the hybrid kinase BC1008 is essential for a proper induction of σ^{B} protein levels during heat and salt stress conditions. This implied that BC1008 is an important regulator of σ^{B} activity in *B. cereus*. The observed phenotype of strain FM1404 was similar to the phenotype previously found for the B. cereus rsbY deletion mutant (van Schaik et al., 2005a). However, as compared to strain FM1404, the σ^{B} protein levels under stress conditions were more reduced in the *rsbY* deletion mutant. Apparently, σ^{B} induction is more strictly regulated by the RR RsbY than by BC1008, under the given conditions.



Fig. 5.2. Effect of *bc1008* deletion on the induction of σ^{B} upon stress.

(A) Cultures of *B. cereus* ATCC 14579 and it derivative strain FM1404 ($\Delta bc1008$) were grown to midexponential growth phase at 30°C, after which they were exposed to a mild heat (42°C) or salt (2.5% NaCl) shock. Total cellular protein was extracted at the indicated time points. Proteins were separated by SDS-PAGE and $\sigma^{\rm B}$ protein levels were detected with immunoblotting using anti- $\sigma^{\rm B}$ antiserum.

(B) Similarly, strain FM1404 containing either the empty vector pHP13 or the complementation vector pHP13-bc1008full was exposed to a mild heat shock (42°C). Total cellular protein was extracted at the indicated time points and σ^{B} protein levels were detected as described above.

To further confirm the crucial role of BC1008 in σ^{B} induction, we cloned the fulllength *bc1008* gene, including 347 bp of its upstream region, into the low-copy vector pHP13. The resulting vector (pHP13-bc1008full) was introduced into strain FM1404 to complement the *bc1008* deletion *in trans*. Immunoblotting with anti- σ^{B} antiserum showed that, upon heat shock, σ^{B} induction was restored to wild-type levels in the complemented FM1404 strain (Fig. 5.2B).

BC1008 is necessary for proper heat pre-adaptation

In a previous study, it was shown that a *B. cereus sigB* null mutant was less protected against a lethal temperature (50°C) by a pre-adaptation to 42°C than its parental strain, resulting in a more than 100-fold reduced survival of the mutant after 40 minutes at 50°C (van Schaik *et al.*, 2004a). This showed that σ^{B} plays a significant role in the protective heat shock response of *B. cereus*. Because σ^{B} protein levels were highly reduced, but not completely absent, during mild heat shock at 42°C in strain FM1404 (Fig. 5.2A), we tested whether this strain was also impaired in its protective heat shock response. Wild-type cells and cells of strain FM1404 were grown to mid-exponential growth phase and exposed to a heat pre-adaptation of 42°C (30 min), after which they were exposed to the lethal temperature of 50°C. Indeed, viable count measurements showed that strain FM1404 was killed more rapidly than its parental strain, displaying an almost 100-fold reduced survival after 40 minutes at 50°C (Fig. 5.3). These results demonstrated that, just like the B. cereus sigB deletion strain, the bc1008 deletion strain was affected in its protective heat shock response, even though $\sigma^{\scriptscriptstyle B}$ levels were not completely abolished in this strain during heat pre-adaptation (Fig. 5.2A).



Fig. 5.3. The protective heat shock response of *B. cereus* ATCC 14579 and its *bc1008* deletion strain.

Cultures of *B. cereus* ATCC 14579 (circles) and its derivative strain FM1404 (squares), in which *bc1008* had been deleted, were grown to mid-exponential growth phase at 30°C, were then pre-adapted at 42°C for 30 minutes and were finally killed at 50°C. The averages of three independent experiments are shown. The error bars indicate standard deviations.

Transcriptome analyses of the *bc1008* deletion strain upon mild heat shock

The above results demonstrated the vital role of BC1008 in regulating σ^{B} activity under heat stress. Because σ^{B} protein levels were still detectable under the conditions tested, it remained the question to what extent BC1008 contributed to governing the overall σ^{B} regulon. To answer this question, *B. cereus* ATCC 14579 and strain FM1404 were exposed to a mild heat shock ($42^{\circ}C$, 10 min) and their transcriptomes were compared using microarrays. We chose mild heat shock because this condition is known to give the highest induction of σ^{B} (van Schaik et al., 2004a). In addition, the transcriptomes of B. cereus ATCC 14579, FM1400 ($\Delta sigB$) and FM1401 ($\Delta rsbY$), detected after a similar 10-minute heat shock have recently been compared to establish the *B. cereus* σ^{B} regulon (van Schaik et al., 2007). Use of the same conditions allowed us to compare our transcriptome data with those obtained previously. In the previous study, σ^{B} regulon members were defined as those genes that showed a significant and \geq 1.5-fold up-regulation in the wild-type strain as compared to both the sigB and the rsbY deletion strains (WT/ $\Delta sigB$ and WT/ $\Delta rsbY$, resp.). Another criterion was that the gene displayed a significant and ≥ 1.5 -fold up-regulation when *B. cereus* was exposed to a mild heat shock (42°C, 10 min). For comparative purposes, we employed similar criteria to analyse our transcriptome data.

We extracted all genes that exhibited a significant up-regulation of 1.5-fold or more in WT/ $\Delta bc1008$. This resulted in a set of 68 genes, of which 66 were also represented on the previous microarray. Comparison of this gene set to the previously identified set of 45 up-regulated genes in WT/ $\Delta sigB$ and 38 upregulated genes in WT/ $\Delta rsbY$, revealed an overlap of 26 and 29 genes, respectively. These overlapping numbers were comparable to the 28 upregulated genes found in both WT/ $\Delta sigB$ and WT/ $\Delta rsbY$ (van Schaik *et al.*, 2007). The sets of 26 and 29 overlapping genes represented a total of 30 genes, of which 22 belonged to the established σ^{B} regulon. This meant that most of these genes were also heat-inducible and that no large sub-sets of genes shared only between WT/ $\Delta bc1008$ and WT/ $\Delta sigB$ or between WT/ $\Delta bc1008$ and WT/ $\Delta rsbY$ were found. Of the remaining 8 genes that did not belong to the

previously defined σ^{B} regulon, 5 genes (*bc0861*, *bc1154*, *bc1155*, *bc3129*, *bc5390*) were assigned as novel σ^{B} regulon members. In the previous study, *bc0861* and *bc3129* were not significantly induced upon mild heat shock. Since we found these genes to be heat-inducible in our microarray experiments and because they are probably part of known σ^{B} -dependent operons (*bc0863/62* and *bc3132/31/30*, resp.), we assigned them as novel σ^{B} regulon members. Similarly, the putative operon *bc1154/55* was previously not assigned to the σ^{B} regulon because it displayed down-regulation in WT/*ΔsigB* (see discussion). However, the operon was significantly up-regulated in both WT/*Δbc1008* and WT/*ΔrsbY* and was shown to be part of the heat stress stimulon. Moreover, a putative σ^{B} binding site was found 201 bp upstream of *bc1154*. This site (ACGTTTGG-N13-GGGTAAA) matched well with the σ^{B} binding site consensus of *B. subtilis* (Makita *et al.*, 2004) and contained the GTTT (-35) and GGGTA (-10) sites typical for *B. cereus* σ^{B} promoters (van Schaik *et al.*, 2007). Finally, *bc5390* was assigned to the σ^{B} regulon as it was heat-inducible and up-regulated in both WT/*Δbc1008* and WT/*Δbc1008* and WT/*Δbc1008* and WT/*Δbc1008* and WT/*Δbc1008* and were the of the beat stress stimulon. Moreover, a putative σ^{B} binding site was found 201 bp upstream of *bc1154*. This site (ACGTTTGG-N13-GGGTAAA) matched well with the σ^{B} binding site consensus of *B. subtilis* (Makita *et al.*, 2004) and contained the GTTT (-35) and GGGTA (-10) sites typical for *B. cereus* σ^{B} promoters (van Schaik *et al.*, 2007). Finally, *bc5390* was assigned to the σ^{B} regulon as it was heat-inducible and up-regulated in both WT/*Δbc1008* and WT/*ΔrsbY* and because it putatively constitutes an operon with the σ^{B} -dependent gene *bc5391*.

Overall, of the 24 genes previously assigned to belong to the *B. cereus* σ^{B} regulon (and represented on the array), 22 were also significantly down-regulated in the *bc1008* deletion strain (Table 5.3). These results indicated that, at least under mild heat shock conditions, BC1008 is important for full induction of the complete σ^{B} regulon.

Analysis of the 68 significantly up-regulated genes in WT/ $\Delta bc1008$ further revealed that the majority of the most severely affected genes were σ^{B} regulon members, including the *sigB* gene itself. These data suggested that, under heat stress conditions, BC1008 specifically activates σ^{B} and its associated regulon and does not seem to activate any other major pathways.

Table 5.3. Transcriptome analyses of *B. cereus* ATCC 14579 and its *bc1008* deletion strain upon heat shock.

Genes were identified as being σ^{B} -dependent when they displayed a significant upregulation (≥ 1.5 -fold) in at least two of the three data sets WT/ $\Delta sigB$, WT/ $\Delta rsbY$ and WT/ $\Delta bc1008$. In addition, the gene had to display a significant up-regulation (≥ 1.5 -fold) upon heat shock (42°C/30°C). Newly identified σ^{B} regulon members and σ^{B} -dependent promoters are in boldface type. Other σ^{B} -dependent promoters have been experimentally confirmed and/or predicted previously (van Schaik *et al.*, 2004a; van Schaik *et al.*, 2004b; van Schaik *et al.*, 2007).

^a, WT 42°C/30°C expression ratios taken from array experiments described in this study. ^b, *Bc0999* was not represented on the previous arrays (van Schaik *et al.*, 2007) and is included in this table because (i) its expression was highly induced upon heat shock, (ii) this induction was *bc1008*-dependent, and (iii) its expression has been shown to be o^B-dependent (van Schaik *et al.*, 2004b). ^c, Because these genes were replaced by an erythromycin cassette, their expression ratios are not shown. ^d, The expression of this gene is probably influenced by read-through effects of the erythromycin resistance cassette (see discussion). -, Expression ratio was not significant (i.e. the false discovery rate of at least one of the probes was above threshold) and is therefore not shown. *, Gene not represented on array.

gene		expressi	on ratios		annotation	alias	$\sigma^{ m B}$ -dependent promoter sites
	42°C/30°C	WT/∆ <i>sigB</i>	WT/∆rsbY	WT/Abc1008			
bc0861	2.1 ^a	2.1	ı	1.9	hypothetical		in operon with bc0863
bc0862	72.9	63.1	84.3	38.5	protease I	yfkM	in operon with bc0863
bc0863	34.0	72.1	31.5	32.9	catalase	katE	ATGTTTAC-N13-GGGTATC-N47-ATG
bc0995	2.2	1.9	2.0	2.6	hypothetical		none identified
bc0996	2.0	2.1	1.8	2.9	hypothetical		none identified
bc0998	70.9	123.6	85.7	51.7	general stress protein	yfIT	ATGTTTAA-N14-GTGTACT-N45-ATG
bc0999	188.0 ^{a,b}	*	*	53.1	hypothetical	csbD	in operon with bc0998
<i>bc1000</i>	36.3	98.1	42.4	60.7	hypothetical		in operon with bc0998
bc1001	35.8	106.4	29.8	23.0	hypothetical		ATGTTTAA-N13-GTGTATG-N37-ATG
bc1002	38.3	42.6	31.7	27.9	anti-o ^B factor antagonist	rsbV	ATGTTTAA-N13-GGGTAAT-N33-ATG
bc1003	52.4	51.1	38.4	35.8	anti-o ^B factor	rsbW	in operon with bc1002
bc1004	28.1	c	24.4	31.3	RNA polymerase sigma factor o ^в	sigB	in operon with bc1002
bc1005	56.0	10.6^d	47.4	37.6	putative bacterioferritin	orf4	ATGTTTAA-N13-GGGTACT-N20-ATG
bc1006	5.1	4.0	°	6.3	PP2C-type RR; regulates σ^{B} activity	rsbY	none identified
bc1007	3.6	2.4	2.3	0.4 ^d	chemotaxis protein methyltransferase	cheR	in operon with bc1008
bc1008	4.4	3.3	3.1	- c	multi-sensor hybrid kinase	rsbK	none identified
bc1009	87.5	188.5	110.1	33.1	hypothetical		ATGTTTAA-N13-GGGTATG-N40-ATG
<i>bc1010</i>	5.5	8.2	7.0	14.8	hypothetical		ACGTTTAG-N13-GGGTATA-N285-ATG
bc1011	23.7	33.9	25.9	39.6	hypothetical		in operon with bc1010
bc1012	*	*	*	*	hypothetical	ybjQ	in operon with <i>bc1010</i>
bc1154	3.4	0.5	4.0	2.7	ferrochelatase	hemH-2	ACGTTTGG-N13-GGGTAAA-N201-ATG
bc1155	2.2	0.4	2.1	1.7	catalase	katA	in operon with bc1154
bc2108	3.8	2.9	2.8	2.6	ECF sigma factor	sigZ	none identified
bc2638	4.3	4.7	4.4	ı	spore germination protein LC		none identified
bc3129	5.2 ^a	ı	3.1	8.0	Mg ²⁺ and Co ²⁺ transporter	corA	in operon with bc3132
bc3130	10.3	20.7	29.3	25.8	hypothetical		in operon with bc3132
bc3131	42.6	29.5	37.3	41.1	hypothetical		in operon with bc3132
bc3132	2.0	7.6	2.0	5.8	general stress protein	yfIT	AGGTTTAA-N14-GTGTATT-N48-ATG
bc4640	ı	ı	ı	ı	hypothetical	ytfJ	in operon with bc4641
bc4641	3.8	3.3	3.9	2.5	hypothetical	ytfI	ATGAATAA-N13-GGGTACG-N10-ATG
bc5390	3.5	ı	5.7	3.8	cell wall hydrolase	CIND	none identified
bc5391	6.1	11.2	11.0	12.6	spore coat protein	gerQ	in operon with bc5390

Bc1008 is mainly transcribed from a putative σ^{A} -dependent promoter

Analysis of different transcriptome datasets showed that bc1008 and sigB displayed similar expression profiles under varying experimental conditions (Fig. 5.4). A similar co-expression pattern was found between *rsbY* and *sigB*. Although *rsbY* is a σ^{B} regulon member (van Schaik *et al.*, 2007), it is mainly transcribed from a σ^A -dependent promoter (van Schaik *et al.*, 2005a). This is in line with the observation that the fully autoregulatory sigB gene (van Schaik et al., 2004a) generally exhibited higher expression ratios than rsbY (Fig. 5.4). Because the expression ratios of *bc1008* and *rsbY* were comparable and a putative σ^A -dependent promoter was found upstream of *bc1008*, we hypothesised that *bc1008* transcription is also mainly governed by σ^A . To assess from which position *bc1008* is transcribed, we performed RACE experiments on RNA isolated from heat-shocked B. cereus cells. This revealed that bc1008 was transcribed from a position 10 bp downstream of the predicted σ^{A} binding site (Fig. 5.5A). Typically, we also detected several transcripts that started from a position 20 bp downstream of the ATG start codon. Analysis of the region upstream of this transcription start site in DBTBS (Makita et al., 2004) did not reveal promoter sites similar to previously described sigma factor binding sites in Bacillus (Fig. 5.5A). No transcripts were found that suggested any direct dependency of *bc1008* on σ^{B} , nor did we find any transcripts that surpassed the predicted σ^{A} binding site.



Fig. 5.4. Log2 expression ratios of *sigB*, *rsbY* and *bc1008* under different experimental conditions.

For all experiments, the log2 values of the Lowess normalised expression ratios (Cy5/Cy3) are given. Every second experiment includes a Cy5/Cy3 dye swap. Experiments: 1-2, WT (control) vs WT (42°C); 5-6, WT (42°C) vs $\Delta sigB$ (42°C); 9-10, WT (42°C) vs $\Delta rsbY$ (42°C) (GEO accession GSE6005); 3-4, WT (30°C) vs $\Delta sigB$ (30°C); 7-8, WT (30°C) vs $\Delta rsbY$ (30°C) (Van Schaik et al., unpublished results). Experiments 11-12, 13-14 and

15-16: WT (control) vs WT (HCl, pH 5.5, 10 min), WT (HCl, 30 min) and WT (HCl, 60 min), respectively. Experiments 17-18, 19-20 and 21-22: WT (control) vs WT (lactic acid, 2 mM undissociated acid, pH 5.5, 10 min), WT (lactic acid, 30 min) and WT (lactic acid, 60 min), respectively (Van Schaik et al., unpublished results). Experiments: 23-24, WT (control) vs WT (anaerobic stress) (GSE9846); 25-26, WT vs $\Delta ccpA$ (GSE7843).

To further study *bc1008* transcription, we cloned different *bc1008* promoter fragments into plasmid pHT304-18Z (Agaisse and Lereclus, 1994) to create *lacZ* transcriptional fusions. One fragment contained the putative σ^A binding site (fragment 3), while another contained the same region with an additional 79 bp, including the putative unknown sigma factor binding site (fragment 4). Other fragments (5 to 7) were truncations of fragment 4 in which first the predicted σ^A

(fragment 5) and finally the unknown sigma factor binding site (fragment 7) were deleted stepwise (Fig. 5.5B). For control purposes, we analysed the *bc1009* upstream region (fragment 1), which contains a confirmed σ^{B} binding site (van Schaik *et al.*, 2004b). To check if *bc1008* transcription could be initiated from a position beyond its predicted σ^{A} -dependent promoter (i.e. read-through from *bc1009*), we also analysed the activity of fragment 2 (Fig. 5.5B). The recombinant plasmids were introduced into wild-type *B. cereus* cells, which were exposed to a heat shock. Activity measurements showed that only fragments 1, 3 and 4 were active (Fig. 5.5C). As expected, the activities of fragment 3 and 4 were low in comparison to the activity of fragment 1. Fragment 4 displayed a slightly higher activity than fragment 3, indicating that a sigma factor binding site may indeed reside within the region downstream of the *bc1008* translation start. Because fragments 5, 6 and 7 did not display any activities, the predicted σ^{A} -dependent promoter seemed to be most important for *bc1008* transcription under heat stress.

Because σ^{B} activity increases during stationary growth phase (van Schaik *et al.*, 2004a), we also measured β -galactosidase activities after overnight growth. Compared to the previous experiment, this led to increased activities of all fragments, except fragment 7 (Fig. 5.5D). Of the bc1008 upstream fragments, fragment 4 again displayed the highest activity, followed by fragment 3. Again, the truncation of the predicted σ^{A} binding site led to a dramatic decrease in promoter activity. This further demonstrated the importance of this site for proper *bc1008* transcription. Interestingly, low activity was observed for fragment 2, indicating that some bc1008 promoter activity resides within the region upstream of the σ^A promoter. Low activity was also found for fragments 5 and 6. Truncation of this latter fragment, which harboured the putative unknown sigma factor binding site, finally led to a complete reduction of all activity. From the results obtained, we can conclude that *bc1008* transcription is mainly governed by a σ^{A} -dependent promoter. The observed σ^{B} -dependency of *bc1008* upon heat shock (Table 5.3) is not caused by a σ^{B} -dependent promoter directly upstream of *bc1008*, but may be explained by an as-yet unidentified σ^{B} dependent transcriptional regulator that activates the *bc1008* promoter.



Fig. 5.5. Analysis of *bc1008* transcription.

(A) Nucleotide sequences around the *bc1008* translation start. Coding sequences are indicated by grey shading. Convergent arrows indicate a predicted stem-loop. SD indicates the Shine-Dalgarno site. The two transcription start sites determined by 5' RACE are indicated in bold and by hooked arrows (+1). One of the sites was found 10 bp downstream of a predicted σ^A binding site, of which the -35 and -10 regions are underlined. The second transcription start was found 20 bp downstream of the *bc1008* translation start codon. The associated putative -10 region is underlined by a double grey dashed line. Numbers indicate the position on the *B. cereus* genome and correspond with the boundaries of the fragments used in β -galactosidase activity assays.

(B) Overview of the *bc1009* and *bc1008* gene arrangements. Fragments 1-7 were used to construct *lacZ* transcriptional fusions. Boxes indicate known and predicted sigma factor binding sites. The grey box indicates a putative binding site of an unknown sigma factor. (C) and (D) β -galactosidase activity assays of *B. cereus* strains containing *lacZ* transcriptional fusions with one of the seven promoter fragments. EV corresponds to cells that harbour the empty vector pHT304-18Z. Prior to the measurements, cells were heat-shocked (42°C, 30 min) or grown at 30°C for 24 hours.

The BC1008 RR receiver domain is essential for proper induction of σ^{B}

Hybrid HKs often act in a multistep His-Asp-His-Asp signalling pathway (Appleby et al., 1996). BC1008 and RsbY could be involved in such a signalling route to activate σ^{B} . In this case, BC1008 would transfer phosphoryl from its H-box to its fused RR receiver domain. A putative phosphotransferase would subsequently receive phosphoryl from BC1008 and transfer it to the RsbY RR receiver domain. Another scenario for σ^{B} activation could involve direct phosphoryl transfer from BC1008 to RsbY. In this scenario, the BC1008 RR receiver could play a role in fine-tuning BC1008 kinase-activity, similarly to what has been found for the RR receiver domain of the hybrid HK VirA of Agrobacterium tumefaciens (Brencic et al., 2004; Chang and Winans, 1992; Chang et al., 1996). Both scenarios would imply a crucial role for the BC1008 RR receiver domain in the induction of σ^{B} . To test the essential role of this domain in σ^{B} activation, we cloned a truncated bc1008 gene, lacking the region encoding the RR receiver domain, into the low copy plasmid pHP13. The resulting plasmid, pHP13-bc1008trunc, was introduced into strain FM1404 to complement the *bc1008* deletion *in trans*. The resulting strain FM1404trunc was tested for its ability to induce σ^{B} protein levels upon heat shock. As compared to what was observed for strain FM1404 carrying the empty vector pHP13, σ^B protein levels were significantly elevated in strain FM1404trunc at all time points tested (Fig. 5.6A). However, these elevated levels did not increase in time as was observed for the fully complemented strain FM1404compl (Fig. 5.2B). Moreover, the σ^{B} protein levels in strain FM1404trunc never reached the maximum σ^{B} levels observed in strain FM1404compl at 30 minutes heat shock (Fig. 5.6A). These results clearly demonstrated the essential role of the BC1008 RR receiver domain in the activation of σ^{B} .

To show that the truncated *bc1008* gene was correctly translated, BC1008specific antibodies were raised in mice. These were used in immunoblotting experiments to detect the presence of BC1008 and its truncated version in extracts of strains ATCC 14579, FM1404, FM1404compl, FM1404trunc and FM1404ev that had been exposed to a heat shock (42° C, 30 min). This revealed highly specific immunocomplexes corresponding to a mass of ~95 kD in strains ATCC 14579 and FM1404compl and to a mass of ~80 kD in strain FM1404trunc (Fig. 5.6B). These masses correspond well with the theoretical masses of BC1008 (101.5 kDa) and its truncated version (85.9 kDa), respectively. As expected, no specific complexes were observed in strains FM1404 and FM1404ev, which lacked the *bc1008* gene. The above data showed that the truncated *bc1008* gene was properly translated in strain FM1404trunc and therefore support the essential role of the BC1008 RR receiver domain in σ^{B} activation. Based on the observation that σ^{B} protein levels were significantly higher in strain FM1404trunc as compared to σ^{B} levels in strain FM1404ev, we

hypothesise that BC1008 can directly transfer phosphoryl to RsbY without using its RR receiver domain. Because σ^{B} protein levels remained constant throughout the heat shock in strain FM1404trunc and because they never reached the same maximum levels observed in strain FM1404compl, it is likely that the BC1008 RR receiver domain plays a role in fine-tuning the kinase-activity of BC1008 (Fig. 5.6D and discussion). Finally, because of the essential role of the hybrid kinase BC1008 in σ^{B} activation, we propose to rename it to RsbK, which stands for <u>Regulator of sigma b</u> with the "K" indicating that a hybrid <u>K</u>inase is involved.



Fig. 5.6. The BC1008 RR receiver domain is essential for proper induction of σ^{B} .

Cultures (A) of strain FM1404 containing either the empty vector pHP13 or pHP13-bc1008trunc, which encodes a BC1008 version that lacks its RR receiver domain, were grown to mid-exponential growth phase at 30°C, after which they were exposed to a mild heat shock (42°C). Total cellular protein was extracted at the indicated time points. Proteins were separated by SDS-PAGE and $\sigma^{\scriptscriptstyle B}$ protein levels were detected with immunoblotting using anti- σ^{B} antiserum. For control purposes, samples extracted from the complemented strain FM1404compl that had been heat-shocked for 30 min (compl 30) were also included.

(B) To show that the truncated version of *bc1008* was properly translated in strain FM1404trunc, immunoblotting experiments were performed using protein extracts of this strain and of strains ATCC 14579 (WT), FM1404, FM1404compl and FM1404ev. All strains had been heat-shocked (42°C, 30 min). Proteins were separated by SDS-PAGE and BC1008 was detected with anti-BC1008 antiserum.

(C) Model for σ^{B} activation in *B. cereus.* The upstream σ^{B} activation pathway most likely involves direct phosphoryl transfer from RsbK (BC1008) to RsbY. The RsbK RR receiver domain probably plays a role in fine-tuning the RsbK-kinase activity (dashed arrows). Upon specific stress signals, the TCS RsbK-RsbY triggers the downstream σ^{B} activation pathway, which has been described previously (van Schaik *et al.*, 2005a).

Discussion

In this study, we demonstrate that the newly identified multi-sensor hybrid kinase BC1008 plays a crucial role in triggering the σ^{B} -mediated stress response of *B. cereus* ATCC 14579. A *B. cereus* bc1008 deletion strain was shown to be incapable of properly inducing σ^{B} protein levels upon mild heat and salt stress conditions and was affected in its protective heat shock response. Transcriptome analyses revealed that BC1008 is highly dedicated to activating σ^{B} under mild heat stress conditions and that it probably does not activate any other major pathways. Transcriptional analyses of *bc1008* showed that, although *bc1008* is part of the σ^{B} regulon, it is mainly transcribed from a σ^{A} -dependent promoter. Finally, we showed that the fused RR receiver domain of BC1008 is essential for a full induction of σ^{B} and most likely plays a role in fine-tuning BC1008 kinase activity.

The transcriptome data presented in this study showed that, under heat shock conditions, BC1008 is primarily involved in activating σ^{B} and its associated regulon. Indeed, the majority of the highly affected genes in strain FM1404 were established σ^{B} regulon members. For instance, of the 21 genes that were upregulated 5-fold or more in WT/ $\Delta bc1008$, only one gene (bc2646) was not included in the σ^{B} regulon. This gene was not up-regulated in either WT/ $\Delta sigB$ or WT/ $\Delta rsbY$. Therefore, it may be specifically activated by BC1008, independent from σ^{B} and RsbY.

The combined analysis of our transcriptome data with those obtained previously (van Schaik *et al.*, 2007) also allowed us to identify novel σ^{B} regulon members. The most noticeable new members were bc1154 (hemH-2) and bc1155 (katA), which encode a putative ferrochelatase and catalase, respectively and which probably comprise one operon. Both genes were induced upon heat shock and were up-regulated in both WT/ $\Delta bc1008$ and WT/ $\Delta rsbY$, but were down-regulated in WT/ $\Delta sigB$ (Table 5.3). In a previous study, it was demonstrated that the elevated levels of KatA in the sigB deletion mutant may cause this strain to be hydrogen peroxide hyperresistant (van Schaik et al., 2005b). However, this phenotype may be caused by pleiotropic effects related to the antibiotic resistance cassette introduced into the sigB deletion strain, which may increase bc1005 expression levels (Table 5.3). Bc1005 encodes a putative ferritin-like protein, involved in ferric iron binding. These proteins have an important role in the oxidative stress response of bacteria as they sequester free iron and thereby prevent oxidative damage to cellular biomolecules (Smith, 2004). Up-regulation of bc1005 may therefore disturb iron homeostasis, which may lead to the observed up-regulation of *bc1154/55* in the *sigB* deletion mutant.

The results obtained from the promoter studies showed that bc1008 transcription is mainly governed by the household sigma factor, σ^A . Other factors seem to only slightly modulate bc1008 expression. One of the factors that may regulate bc1008 expression is an as-yet unidentified sigma factor that initiates bc1008 transcription from within the bc1008 coding region. This would imply the existence of alternative, N-terminally truncated protein products. Because the associated transcripts were detected at different heat shock time points and the corresponding bands on the agarose gels were clear (i.e. no smears, results not shown), the transcripts probably did not result from

breakdown products of larger stable mRNAs. Moreover, the β -galactosidase assays showed that significant promoter activity indeed resided within the region just upstream of the mapped transcription start site. Sequence analysis of *bc1008* and its orthologues in other bacilli did not reveal any in-frame translation start codons preceded by a clear ribosomal binding site. Moreover, immunoblotting with the anti-BC1008 antiserum revealed only one specific BC1008 product, even in cells that had entered stationary growth phase (results not shown), a condition under which the putative alternative transcription was of significant importance (Fig. 5.5D). Therefore, it remains to be determined whether alternative *bc1008* translation indeed occurs and, if so, what the functional role of this process may be.

Two possible signalling routes for σ^{B} activation in *B. cereus* would be (i) a complex multistep phosphorelay in which a putative third protein phosphotransferase links phosphoryl transfer between the BC1008 and the RsbY RR receiver domains or (ii) a single-step phosphorelay between BC1008 and RsbY, in which the BC1008 RR receiver domain plays a subtle regulatory role. In both scenarios, the BC1008 RR receiver domain would play an important role. The results obtained from the BC1008 truncation experiments indeed revealed that this RR receiver domain is essential for proper induction of σ^B protein levels under heat stress conditions. However, in strain FM1404trunc, σ^{B} protein levels were significantly elevated as compared to the levels in strain FM1404ev. These results suggested that, if BC1008 indeed acts as the cognate HK for RsbY to activate σ^{B} , the truncated version of BC1008 was still capable of transferring phosphoryl to RsbY, albeit in a sub-optimal fashion. Therefore, the most likely signalling route would involve His-Asp phosphotransfer from the BC1008 conserved H-box histidine (H505) directly to the conserved aspartate (D59) in the RsbY RR receiver domain. Apparently, His-Asp phosphoryl transfer within BC1008, from H505 to the conserved aspartate (D827) in its own RR receiver domain, does not constitute the essential first step of an extended His-Asp-His-Asp phosphorelay. However, because σ^{B} protein levels were not fully induced in strain FM1404trunc, we propose that the BC1008 RR receiver domain plays a subtle regulatory role in the BC1008-RsbY signalling cascade. Under non-stress conditions, the BC1008 RR receiver domain may prevent undesired His-Asp phosphorelay between BC1008 and RsbY, while it may stimulate this process under the appropriate stress conditions. This would imply that the BC1008 RR receiver domain does not act constitutively, but is in some way modulated, for example by phosphorylation of its conserved aspartate (D827). A similar regulatory mechanism has been found for the hybrid kinase VirA of Agrobacterium tumefaciens, which phosphorylates its cognate RR VirG upon host-released chemical signals. This signalling route also includes a direct His-Asp phosphorelay between a conserved histidine in the VirA H-box and a conserved aspartate in the VirG RR receiver domain (Huang et al., 1990; Jin et al., 1990a; Jin et al., 1990b). Under non-stimulating conditions, the fused RR receiver domain of VirA specifically inhibits VirA kinase activity and thus prevents improper stimulation of the VirA-VirG signalling pathway (Brencic et al., 2004; Chang and Winans, 1992; Chang et al., 1996).

This study further demonstrates that σ^{B} activation in *B. cereus* significantly differs from known σ^{B} activation pathways in other low-GC Gram-positives. In *B.*
subtilis, but most likely also in many of its closest relatives, a massive protein complex, the stressosome, integrates multiple environmental stress signals to orchestrate dephosphorylation of RsbV and the eventual release and activation of σ^{B} (Chen et al., 2003a; Marles-Wright et al., 2008; Pané-Farré et al., 2005). In addition, B. subtilis uses a second pathway, which monitors energy stress, to regulate σ^{B} activity (Vijay *et al.*, 2000). Both the above pathways are absent in members of the B. cereus group, which contain only one PP2C-type phosphatase, RsbY, for the control of σ^{B} activity (van Schaik *et al.*, 2005a). The results presented in this study strongly suggest that RsbY receives its input signals from the complex, multi-sensor hybrid kinase BC1008. The domain architecture of this unique sensor protein suggests that in members of the B. cereus group, environmental and intracellular stress signalling routes are combined into one single protein. This activation mechanism may reflect unique niches in which B. cereus group members can reside. The BC1008-RsbY module was found to be conserved in only a few other bacilli, where it was not genomically associated to sigB. Therefore, it would be interesting to find out whether these bacilli, including B. coagulans and Lysinibacillus sphaericus also use their BC1008 and RsbY orthologues to trigger sigma factor-mediated stress responses or whether this is a unique feature of the *B. cereus* group.

Acknowledgements

The authors wish to thank Eric-Jan van der Mark, Heidy den Besten, Menno van der Voort and Maarten Mols for their valuable assistance in the lab.

Chapter 6

Novel σ^B activation modules of Gram-positive bacteria involve the use of complex hybrid histidine kinases

Mark de Been, Christof Francke, Roland J. Siezen, Tjakko Abee

Manuscript in preparation

A common bacterial strategy to cope with stressful conditions is to use alternative sigma factors that control a specialised set of target genes. In the food-borne pathogen Bacillus cereus, activation of the major stress-responsive sigma factor σ^{B} is mediated by a unique signalling route that involves the complex multi-sensor hybrid histidine kinase RsbK. In this study, we show that RsbK-type kinases are conserved across a wide variety of bacterial species. Genome context and protein sequence analyses of 118 close RsbK homologues revealed an extreme variability in N-terminal sensory as well as C-terminal regulatory domains and suggested that RsbK-type kinases are subject to complex fine-tuning systems. In addition, we provide strong indications that the use of RsbK-type histidine kinases for the control of σ^{B} -like sigma factors is not restricted to B. cereus group members, but also occurs in at least several lowas well as high-GC Gram-positives, including Geobacillus, Paenibacillus and actinobacteria. These findings illustrate that these Gram-positive bacteria have evolved or obtained a common solution for monitoring the same process: the activation of a stress-responsive alternative sigma factor. This solution markedly differs from the extensively studied and highly conserved RsbRST-mediated σ^B activation route found in *B. subtilis* and related low-GC Gram-positives.

Introduction

Bacteria use dedicated sets of sensory modules to tightly coordinate gene expression in response to specific environmental fluctuations. A commonly used sensory module is the two-component signal transduction system (TCS), which includes a transmembrane sensor histidine kinase (HK) and its cognate response regulator (RR). The mode of signal transduction by TCSs involves a phosphotransfer reaction between a conserved histidine and aspartate residue located in the HK phosphotransferase and RR receiver (REC) domain, respectively. RRs generally function as transcription factors that, upon phosphorylation bind to specific sites on the DNA to alter the expression of genes involved in adaptive responses (Hoch, 2000).

Another bacterial strategy for tight control of gene expression is the use of alternative sigma factors. In exponentially growing cells, most of the transcription is mediated by a "housekeeping" sigma factor that is equivalent to σ^{70} in *Escherichia coli* and σ^{A} in *Bacillus subtilis*. However, under specific conditions, like severe environmental stress, the housekeeping sigma factor gets replaced from the RNA polymerase by alternative sigma factors that bind specific DNA target sites and control specialised regulons (Gruber and Gross, 2003; Paget and Helmann, 2003). One of the best studied alternative sigma factors is the key stress-responsive sigma factor σ^{B} of low-GC Gram-positives of the genera Bacillus, Listeria and Staphylococcus (Price, 2002). Besides mediating the general stress response, σ^{B} plays an important role in virulence in the human pathogens L. monocytogenes and S. aureus (Chaturongakul et al., 2008; Kazmierczak et al., 2005; Novick, 2003), and to a lesser extent in B. anthracis (Fouet et al., 2000). Sigma factors equivalent to σ^{B} have also been found in high-GC Gram-positives, including *Mycobacterium tuberculosis* and *Streptomyces* species (Mittenhuber, 2002). In *S. coelicolor*, σ^{B} acts in a complex network of several paralogous sigma factors, where it plays a prominent role in osmotic and oxidative stress responses, as well as cellular differentiation and the production of antibiotics (Cho et al., 2001; Lee et al., 2005; Viollier et al., 2003).

The σ^{B} network has been studied best in the model low-GC Gram-positive *B.* subtilis. Under non-stress conditions, σ^{B} is held in an inactive state complex by the anti-sigma factor RsbW (Benson and Haldenwang, 1993). Release of σ^{B} from RsbW is accomplished by the anti-anti sigma factor RsbV, which sequesters RsbW under the appropriate conditions. The phosphorylation status of RsbV forms a critical element in the above mechanism as only dephosphorylated RsbV is able to sequester RsbW. As RsbW also acts as a kinase on RsbV, this provides a level of negative feedback on σ^{B} activation (Alper *et al.*, 1996; Dufour and Haldenwang, 1994). Under stress conditions, the main route for σ^{B} activation involves the dephosphorylation of RsbV by one or more specific PP2C-type phosphatases. Whereas the RsbVW partner-switching mechanism is highly conserved in species that contain σ^{B} , the N-terminal domains of the PP2C-type phosphatases vary considerably across species (van Schaik and Abee, 2005). For example, *B. subtilis* contains two σ^{B} -activating PP2C-type phosphatases, RsbP and RsbU. Energy stress is signalled through RsbP, which contains an N-

terminal PAS sensory domain (Brody *et al.*, 2001; Vijay *et al.*, 2000), while environmental stress (i.e heat, osmolytes, ethanol, low pH) is signalled through RsbU, which contains an N-terminus that interacts with the RsbR- and RsbScontaining supramolecular "stressosome", via the regulator RsbT (Akbar *et al.*, 2001; Chen *et al.*, 2003a; Delumeau *et al.*, 2006; Dufour *et al.*, 1996; Kim *et al.*, 2004; Marles-Wright *et al.*, 2008).

In the food-borne human pathogen *B. cereus*, the mechanism of σ^{B} activation has remained mostly unresolved. It has been shown that σ^{B} activation is governed by a single PP2C-type phosphatase, RsbY, which carries an N-terminal REC domain (van Schaik et al., 2005a). This suggested the existence of a partner HK acting on RsbY. Indeed, we have recently identified the hybrid HK, RsbK (BC1008), as a potential partner in the σ^{B} -mediated stress response of B. cereus (de Been et al., unpublished results). A following genome survey indicated that RsbK and RsbY should constitute one functional module for the control of σ^{B} activation in members of the *B. cereus* group, including the pathogens *B. anthracis* and *B. thuringiensis* and the psychrotolerant *B.* weihenstephanensis. Typically, orthologous RsbKY signalling modules were found in only four other bacilli outside the B. cereus group. In contrast to what was found for the B. cereus group members, the RsbKY modules of the other bacilli were not genomically associated to sigB (de Been et al., unpublished results). Therefore, it is conceivable that RsbKY-dependent activation of σ^{B} is a unique feature of the Bacillus cereus group. To further explore this putative unique trait of the B. cereus group, we searched all available microbial and eukaryotic genome sequences for the presence of RsbK-type signalling domains. Moreover, subsequent phylogenetic and gene context analyses revealed that signalling modules involving RsbK-like proteins are present in several other low-GC as well as high-GC Gram-positives and could potentially trigger σ^{B} -mediated stress responses. Based on these results, we propose that, besides the wellcharacterised and conserved σ^{B} activation pathway of *B. subtilis*, the use of RsbK-type hybrid kinases is another common bacterial strategy to activate the stress-responsive sigma factor, σ^{B} .

Materials and Methods

Identification of RsbK homologues

Potential RsbK homologues were retrieved from the National Center for Biotechnology Information (NCBI; ftp.ncbi.nih.gov/genomes/Bacteria/; including 775 bacterial and archaeal genomes on December 4, 2008). Separate BLAST (Altschul *et al.*, 1997) runs were performed with the RsbK (BC1008) HK phosphotransferase domain (residues 495-733) and REC domain (776-891) (Ethreshold $\leq 1e^{-30}$ and $1e^{-15}$ for HK and RR domain hits, resp.). As both domains are characteristic for all TCSs and are easily recognised, such a search should yield all potential candidates. In fact, it has been shown that these domains contain enough information-density to be able to use them in classification and evolutionary studies (de Been *et al.*, 2006; Fabret *et al.*, 1999; Grebe and Stock, 1999). The retrieved BLAST hits were aligned with MUSCLE (Edgar,

2004), after which CLUSTAL W (Thompson *et al.*, 1994) was used to generate two bootstrapped neighbour-joining (NJ) trees, one for each domain. Additional similar sequences from Eukaryotes and low-GC Gram-positive draft genomes were retrieved via the NCBI BLAST server (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and added to the trees in a similar way as has been described previously (de Been *et al.*, unpublished results). The trees were visualised for inspection with LOFT (van der Heijden *et al.*, 2007). All proteins that clustered with at least one of the two RsbK signalling domains (bootstrap \geq 30%) were assigned as RsbK homologues. For the identification of other groups of homologues, a similar approach was used.

Protein sequence and gene context analyses

Protein domain architectures and functionalities were predicted using Pfam (Bateman et al., 2004), SMART (Letunic et al., 2009) and the COG database (Tatusov et al., 2001), while transmembrane helices were predicted using TMHMM (Krogh et al., 2001). Protein sequence sets were analysed for the presence of over-represented sequence strecthes using the MEME expectation maximisation algorithm (Bailey and Elkan, 1994). As conserved gene context is a strong indicator of the biological role of a gene (Dandekar et al., 1998; Overbeek et al., 1999), and even of conserved regulation (de Been et al., 2008; Francke et al., 2008), we analysed all protein sequences encoded within the operons that contained one or more rsbK homologues. Operons were predicted as described previously (Wels et al., 2006), using a maximum distance of 200 bp between genes (see results for exceptions). The genomic regions within a range of ± 5 kb of all *rsbK* homologues were analysed for the presence of genes encoding alternative sigma factors and putative regulators of these sigma factors, including putative PP2C-type phosphatases, anti-sigma factor antagonists and anti-sigma factors. The latter two protein types were identified by the presence of a STAS and HATPase c domain, respectively (Bateman et al., 2004; Letunic et al., 2009). Because the HATPase domains also occur in other protein types, such as HKs, DNA gyrases and topoisomerases, the latter proteins were only assigned as putative anti-sigma factors when they were genomically associated with sigma factors, PP2C-type phosphatases or anti-sigma factor antagonists. In addition, they had to be single HATPase domain proteins (see results for exceptions).

Results and Discussion

RsbK-type hybrid kinases occur in a wide variety of bacterial species

Recently, we found that a functional RbsK is essential for proper σ^{B} activation upon mild stress conditions in *B. cereus* (de Been *et al.*, unpublished results). Using a straightforward BLAST search (see methods) we have mapped the occurrence of RsbK-type HKs in other species. Based on the presence of the TCS-characteristic HATPase and REC domains, we identified a group of 118 potential RsbK homologues across 101 bacterial species, including species from the phyla Proteobacteria (71 RsbK homologues), Firmicutes (19) Actinobacteria

(17), Cyanobacteria (7), Bacteroidetes (2) and undefined (2) (supplementary Fig. S6.1 and S6.2: http://www.cmbi.ru.nl/~mdebeen/Thesis). No RsbK homologues were found in Archaea and Eukaryotes. As the NJ trees for both domains were almost identical we built a concatenated tree. A reduced and collapsed version of a part of this tree, representing the 89 closest homologues to *B. cereus* RsbK, is shown in Fig. 6.1 (left) (complete tree in Fig. S6.3: http://www.cmbi.ru.nl/~mdebeen/Thesis). We found the hybrid HKs of *Bacillus coagulans* 36D1, *Lysinibacillus sphaericus* C3-41, *Bacillus* B14905 and *Bacillus* NRRL B-14911 to be most similar (bootstrap 95.8%) to RsbK of *B. cereus*. The other RsbK homologues found in the low-GC Gram-positives were in *Paenibacillus* JDR-2 (2x), *Geobacillus* Y412MC10 (2x), *Desulfotomaculum reducens* MI-1 and *Clostridium thermocellum* ATCC 27405, although the latter one appeared more distantly related.

RsbK-type HKs display extremely variable sensory and C-terminal regions

Sequence analyses of the RsbK homologues revealed the presence of several Nterminal HK sensory domains (Fig. 6.1 middle), including CHASE3, GAF and PAS/PAC domains, of which the latter two have been implicated in small ligand/cyclic nucleotide binding and redox/light/metabolite sensing, respectively (Galperin, 2004). The GAF sensory domain was highly conserved in the RsbK homologues and was always found next to, and N-terminally from, the HK phosphotransferase domain. In addition, almost all RsbK homologues contained one to several putative HAMP domains, which are thought to link N-terminal sensory domains with intracellular phosphotransferases (Hulko *et al.*, 2006).

Fig. 6.1. Relationship between RsbK homologues in terms of their HK and RR domains, overall domain architecture and associated genomic context.

Left: reduced and collapsed version of the bootstrapped NJ tree that was built using the HK phosphotransferase and REC domain sequences (concatenated) of 118 RsbK-type HKs. The tree shown represents the 89 closest RsbK homologues, including OsaA of S. coelicolor, which has been experimentally studied (Bishop et al., 2004). This tree was rooted on Bphy_5629 of Burkholderia phymatum, which clustered with the remaining 28 more distantly related RsbK homologues, including the experimentally studied HK MXAN_0712 of M. xanthus (Shi et al., 2008). For a complete and uncollapsed version of the tree, including bootstrap support, we refer to Fig. S6.3 (http://www.cmbi.ru.nl/ ~mdebeen/Thesis). Middle: domain architecture of RsbK homologues as defined by Pfam (Bateman et al., 2004), SMART (Letunic et al., 2009) and TMHMM (Krogh et al., 2001). Right: genomic associations of rsbK homologues with genes encoding CheR, CheB, RR (REC), PP2C-type phosphatase, anti-sigma factor (ASF), anti-anti sigma factor (AASF) and $\sigma^{\text{B}}\text{-like}$ proteins. In the case of CheR-, CheB- and RR-encoding genes, only the operons containing the rsbK homologues were considered. Exceptions were made for rsbK and its orthologues in several bacilli and for osaA and its orthologues in several actinobacteria, where the "known" partner RRs (RsbY and OsaB, resp.) are encoded in different operons that flank the rsbK/osaA operons. In the case of sigma factor- and sigma factor regulatorencoding genes, the gene neighbourhoods $(\pm 5 \text{ kb})$ of the rsbK homologues were considered. *, Note that the columns displaying the number of anti-sigma factor antagonists and anti-sigma factors also include RsbR/S- and RsbT-like proteins, respectively.



Based on the N-terminal sensory regions, a subdivision could be made into two RsbK-types: Type I containing transmembrane helices and putative membraneassociated extracellular sensory domains (this type includes *B. cereus* RsbK); and Type II lacking transmembrane helices and containing a multitude of HAMP linker regions (Fig. 6.1). Even within the two types, a high variability was observed between the different N-terminal regions. For example, in *Shewanella woodyi* two type I RsbK-like HKs were identified (Swoo1960 and Swoo1961), which were highly related in terms of their phosphotransferase and GAF domains. However, both proteins displayed marked differences with respect to their N-terminal sensory domains. In Type II RsbK homologues, the N-terminal regions showed extreme variability with respect to the number of detected HAMP domains per HK, which ranged from 4 to as many as 14. These findings are in agreement with a previous study in which it was shown that evolutionary related HKs can contain highly different N-terminal regions due to shufflings and duplications of sensory domains (Alm *et al.*, 2006).

Variability was also found in the number of predicted C-terminal REC domains, which ranged from 1 in most of the RbsK homologues found in low- and high-GC Gram-positives to as many as 3 in the other species. Typically, when considering the homologues that contained 2 or 3 REC domains, the C-terminal one always appeared to be most similar to RsbK REC (~50% identical). Strikingly, the additional REC domains were relatively dissimilar from RsbK REC (~25 id%), the only exception being Mmc1_1215 of *Magnetococcus* MC-1. The distinction between the RsbK-like and the other REC domains was demonstrated by a bootstrapped NJ tree of all the hybrid HK REC domains. In such a tree, the N-terminal, middle, and C-terminal REC domains (considering HKs with three RECs) clustered in three distinct branches (results not shown). These data suggest that in the RsbK-type HKs that harbour multiple REC domains, these domains may fulfil distinct biological roles. This may further indicate the high regulatory complexity of these hybrid HKs.

Connecting the RsbK homologues to cognate RRs

In addition to RsbK itself, two of its homologues have been experimentally characterised. These include the Type II homologues SCO5748 (OsaA) of Streptomyces coelicolor and MXAN_0712 of Myxococcus xanthus. OsaA of S. coelicolor and its putative cognate RR OsaB have been implicated to function in osmoadaptation, aerial mycelium formation and the coordination of antibiotic production (Bishop et al., 2004), while MXAN_0712 of M. xanthus was shown to be essential in fruiting body formation and sporulation (Shi et al., 2008). To obtain additional information about the potential biological role of the other RsbK homologues, especially with respect to the possible regulation of alternative sigma factors, we analysed the genomic regions surrounding RsbK-encoding genes. Almost all (~88%) homologues could be connected to one or more genes encoding an RR (Fig. 6.1 right). The domain composition of these RRs appeared highly variable, ranging from the "classical" composition, containing an Nterminal REC and a C-terminal DNA-binding domain, to "a-typical" compositions, containing a singular REC domain or putative C-terminal guanylate cyclase-, cyclic di-GMP phosphodiesterase- and kinase-type output domains in addition.

Connections of RsbK homologues with CheR and CheB

Interestingly, we found that many RsbK homologues were genomically associated with putative CheR- (~60% of the HKs) and CheB-encoding genes (~42%) (Fig. 6.1). CheR and CheB have been extensively studied in Escherichia coli and B. subtilis, where they play a role in the adaptation (i.e. sensitation and desensitation) of the chemotaxis machinery to persisting stimuli. CheR is a methyltransferase that methylates specific glutamate residues within methylaccepting chemotaxis proteins (MCPs), while the methylesterase/deamidase CheB removes methyl groups from these proteins. MCPs function as stimulus perceptors that transduce their signals to the chemotaxis regulator CheA. The methylation state of the MCPs influences this signalling activity and consequently influences flaggellar rotation (reviewed in: Szurmant and Ordal, 2004; Wadhams and Armitage, 2004). The observed genomic association between RsbK-type hybrid HKs and CheR/B could imply a role for these HKs in bacterial chemotaxis. However, this role has not been described before. Therefore it could well be that the hybrid kinases themselves are the main target of these CheR/B proteins. In MCPs, methylation sites generally appear as glutamate-glutamate (EE) or glutamine-glutamate (QE) pairs that are located in tandemly repeated heptads within coiled-coil regions. In the case of glutamine, the side chain is deamidated by CheB prior to its participation in the methylation cycle (Kehry et al., 1983; Le Moual and Koshland, 1996; Nowlin et al., 1987; Perez et al., 2006; Terwilliger and Koshland, 1984; Zimmer et al., 2000). Interestingly, sequence analysis of the RsbK homologues indeed revealed the presence of conserved EE and QE pairs. These conserved pairs were always found between the cytoplasmic GAF domain and the H-box and occurred in tandemly repeated heptads (Fig. 6.2A), of which some have recently been predicted to constitute a conserved helical domain (Anantharaman et al., 2006). In a previous comparative study of 29 bacterial MCPs (Le Moual and Koshland, 1996), predicted and confirmed methylation sites were assigned to positions "b" and "c" of the "a-b-c-d-e-f-g" heptad repeat, according to the scheme of McLachlan and Stewart (McLachlan and Stewart, 1975). Similarly, we could assign the detected EE and QE pairs of the RsbK homologues to these positions, which were followed almost invariably by a leucine (L) at position "d" (Fig. 6.2A). As compared to established methylation sites in MCPs from E. coli, Salmonella enterica and B. subtilis, the heptads found in the RsbK homologues appeared to be different at positions "ad-e-f-g". However, a recent study on Thermotoga maritima MCPs has revealed that CheR-mediated methylation sites can indeed be distinct from the E. coli, S. enterica and B. subtilis consensus (Perez et al., 2006) (Fig. 6.2B). The above data suggest that the sensitivity of RsbK-type HKs to environmental

signals can be modulated via CheRB-mediated methylation/demethylation. This hypothesis is supported by the fact that the RsbK homologues that are genomically associated with CheRB generally contain more putative methylatable pairs at the "b-c" positions than those that are not associated with CheRB (average of 4.7 pairs/protein versus 2.2 pairs/protein, resp., Fig. S6.4: http://www.cmbi.ru.nl/~mdebeen/Thesis).

Fig. 6.2. Repeated heptads and potential methylation sites in RsbK homologues.

(A) RsbK-type HK protein sequences of the regions located between the cytoplasmic GAF domain and H-box. Protein identifiers (NCBI) are indicated on the left followed by the corresponding location within the protein sequence. Note that the figure only displays sequences of Type I RsbK homologues, which generally contain a higher number of putative methylatable pairs and which display a strong genomic association with CheR/B. Putative methylation pairs (highlighted in black) were assigned to positions "b-c" of the "a-b-c-d-e-f-g" heptad repeats, according to the scheme of McLachlan and Stewart (McLachlan and Stewart, 1975). These assigned positions are in agreement with the positions of known and predicted CheR-mediated methylation sites (Le Moual and Koshland, 1996). The repeated heptads strongly indicate the presence of coiled-coil compositions with an almost invariable leucine (L) at position "d" (boxed residues). In RsbK, the region comprising residues "MQAEEL" to "ASELLR" has previously been predicted to constitute a conserved helical domain, functioning as a molecular switch between signalling domains (Anantharaman *et al.*, 2006).

(B) WebLogo (Crooks *et al.*, 2004) representations of the predicted RsbK-type HK methylation sites (left) and of confirmed methylation sites in MCPs of *E. coli*, *S. enterica*, *B. subtilis* (middle) and *T. maritima* (right). The left image was constructed with the highlighted heptads starting from position 10 in each sequence shown in Fig. 6.2A. References for known methylation sites: *E. coli* Tsr (Kehry *et al.*, 1983), Trg (Nowlin *et al.*, 1987); *S. enterica* Tar (Terwilliger and Koshland, 1984); *B. subtilis* McpB (Zimmer *et al.*, 2000); *T. maritima* MCPs (Perez *et al.*, 2006). Note that the *T. maritima* MCP methylation sites are distinct from those in *E. coli*, *S. enterica* and *B. subtilis* in that they are (e.g.) followed more often by hydrophobic residues (L, V, I, M).

RsbK homologues found in Gram positive bacteria are connected to σ^{B}

The gene context analyses revealed that besides *rsbK* in the *B. cereus* group, several *rsbK* homologues found in low- and high-GC Gram-positives are located in gene clusters encoding one to several proteins related to σ^{B} -mediated responses. The clusters included genes encoding PP2C-type phosphatases, RsbV, RsbW, RsbR, RsbS and RsbT. Moreover, four of these gene clusters also encode a putative σ^{B} -like sigma factor (Fig. 6.1). Some of the *rsbK* homologues, like *fraal6455* in *Frankia alni* ACN14a, were found in gene clusters that encode a single σ^{B} -related regulator, while other *rsbK* homologues are located in gene clusters that encode multiple partner switching proteins for the control of σ^{B} activity. For example, in *S. coelicolor* the *rsbK* homologue *sco7327* is located in a gene cluster that encodes as many as three potential PP2C-type phosphatases, an anti-sigma factor antagonist (RsbV), an RsbRST module and the alternative sigma factor σ^{M} , which is related to σ^{B} (Lee *et al.*, 2005). In addition, one of the PP2C phosphatases was found to be fused to an N-terminal ATPase and thus could function as an anti-sigma factor.



ONH 4

QUI

9 9 9

Øε Øε

RsbK-type hybrid HKs

Escherichia coli (Trg, Tsr) Salmonella enterica (Tar) Bacillus subtilis (McpB)

Thermotoga maritima (TM0429, 1143, 1146 & 1428)

ε

12

<u>₩</u>

Another interesting gene cluster was found in *Geobacillus* Y412MC10, where the *rsbK* homologue *gymc10draft_3662* directly flanks an *rsbY* orthologue, similar to what is found in the *B. cereus* group. Other genes in its direct neighbourhood are *rsbV*, *rsbW* and *sigB*, but also *yflT* and *corA*. These latter two genes putatively encode a general stress protein and an Mg^{2+}/Co^{2+} transporter, respectively. In fact, the orthologues of these genes in *B. cereus* (*bc0998* and *bc3129*, resp.) have been implicated in the σ^{B} -mediated stress reponse (de Been *et al.*, unpublished results; van Schaik *et al.*, 2007). Finally, the RsbK homologue Pjdr2DRAFT_1033 of *Paenibacillus* JDR-2 is likely to be involved in σ^{B} regulation because it is associated with *rsbW* and *sigB* and because the *Paenibacillus* genome harbours an operon encoding RsbY (Pjdr2DRAFT_0995), RsbV, RsbW and YfIT. This operon has probably "jumped" to another location in the genome, as it is flanked by putative transposase- and integrase-encoding elements.

Novel activation routes for σ^{B} in Gram-positive bacteria

As described above, the RsbK-type HK OsaA and its putative cognate RR OsaB of *S. coelicolor* have been implicated in osmoadaptation, differentiation and the production of antibiotics (Bishop *et al.*, 2004). In *S. coelicolor*, these processes are also controlled by σ^{B} (Cho *et al.*, 2001; Lee *et al.*, 2005; Viollier *et al.*, 2003) and indeed, a functional link between OsaAB and σ^{B} has recently been confirmed (Martínez *et al.*, 2009). In this recent study it was shown that *osaC*, which is divergently transcribed from *osaA*, encodes a regulatory protein that consists of an RsbW-like N-terminal domain that functions as a σ^{B} anti-sigma factor. This domain is followed by PAS and GAF sensory domains and a PP2C-type phosphatase. OsaC was linked to *osaB* because the osmotic stress-induced expression of this latter gene was shown to be σ^{B} -dependent and because OsaC was shown to be essential for returning *osaB* and *sigB* expression levels back to "normal" after osmotic stress. In addition, an *osaC* deletion mutant was affected in its adaptive response to osmotic stress, similar to a *sigB* deletion mutant.

Besides OsaC, another "more classical" RsbW protein has been characterised in *S. coelicolor*. This protein (RsbA) acts in an RsbVW-like partner-switching module for the control of σ^{B} (Lee *et al.*, 2004). Unlike *rsbA*, the *rsbV* gene is not located in the *sigB* operon, but in a gene cluster encoding multiple putative sigma factor regulators as well as σ^{M} . Interestingly, this gene cluster also includes *sco7327*, the second *rsbK* homolog of *S. coelicolor*. These findings indicate a complex regulatory connection between σ^{B} and σ^{M} , a phenomenon that has been (partly) confirmed (Lee *et al.*, 2005). It is possible that SCO7327 monitors σ^{M} activation, in accordance with the genomically associated RsbRST module. However, a role for this RsbK homolog in σ^{B} activation cannot be excluded.

Despite the apparent complexity of the above sigma factor activation pathways, it is clear that *osaABC* encodes a primary module for σ^{B} regulation in *S. coelicolor*. One question that needs answering is how signals are transferred to OsaC. It has been suggested by Martínez *et al.* that OsaB may interact with other proteins via its C-terminal coiled-coil region (Martínez *et al.*, 2009). Considering this, we propose that upon (osmotic) stress, the RsbK homolog

OsaA activates its partner RR OsaB, which in turn transduces its signals to OsaC via direct or indirect protein-protein interactions. Because PP2C-type phosphatases are known as key actors for the dephosphorylation of RsbV (van Schaik and Abee, 2005), the PP2C-type phosphatase domain of OsaC most likely acts in a similar fashion to activate σ^{B} . This latter hypothesis agrees with the fact that an *osaC* deletion mutant could be complemented *in trans* by the full *osaC* gene, but not by a truncated *osaC* gene that lacked the PP2C-type phosphatase-encoding region (Martínez *et al.*, 2009).

The proposed model for σ^{B} activation in *S. coelicolor* A3(2) also holds for the high-GC Gram-positives *S. avermitilis, S. griseus, Thermobifida fusca, Salinispora tropica, Salinispora arenicola*, and several *Frankia* species, which all contain the OsaABC module. As shown in this study, the above actinobacterial modules display strong similarities to the σ^{B} -activating RsbKY module of the *B. cereus* group and possible other σ^{B} -activating modules of Gram-positive bacteria. This strongly suggests that the use of RsbK/OsaA-type hybrid HKs is a common strategy for Gram-positive bacteria to control the activity of (σ^{B} -like) alternative sigma factors. As summarised in Fig. 6.3, this strategy is altogether different from the well-characterised σ^{B} activation pathways in *B. subtilis*.

Conclusions

In this study, we provide strong indications that RsbK(Y)-mediated activation of alternative sigma factors is not restricted to members of the B. cereus group, but is used by several other low- as well as high-GC Gram-positives. In addition, we found that RsbK-like hybrid kinases are not restricted to Gram-positive bacteria, but also occur in Proteobacteria, Cyanobacteria and Bacteroidetes. Based on the gene context analyses, it seems that in these bacteria the RsbKtype HKs are used for purposes other than the regulation of alternative sigma factors. This is exemplified by the rsbK homolog abo_1307 of Alcanivorax borkumensis, which is located in an operon that encodes several type II secretory pathway components. In addition, its cognate RR contains a PilZ domain: a domain that has been associated with protein secretion, fimbrial biogenesis and twitching motility (Alm et al., 1996). The finding that similar signalling modules are used for different purposes is analogous to what was found in a recent genome survey, in which it was demonstrated that RsbRST-like signalling modules occur in a wide variety of bacteria where they may interact with varying output modules (Pané-Farré et al., 2005). Apparently, RsbRST- and RsbK-type modules provide common bacterial solutions to the problem of signal integration. However, despite the apparent universal use of these signalling modules across bacteria, considerable variability may occur between these modules. In the case of the RsbK-type HKs, this variability even occurs between homologues involved in the same process. This is clearly demonstrated by the difference in N-terminal sensory domains between RsbK and OsaA. These differences may reflect the different niches in which the associated organisms reside and may illustrate the different solutions these organisms have evolved to monitor the same process: the activation of a stress-responsive alternative sigma factor. These findings further demonstrate the highly modular nature of sigma factor activation routes and signal transduction routes in general.

"B. subtilis module"	B. clausii B. halodurans O. jheyenais L. monocytogenes L. innocua L. welshimeri	RsbQP triggers*	decrease in intracellular (ATP)		σ ^в -dependent transcription
	B. subtilis B. amyloliquefaciens B. pumius B. licheniformis B. coagulans Bacillus NRRL B-14911	RsbRSTU triggers	heat osmolytes (NaCl) ethanol low pH blue light		
<i>lor</i> module"	F. alni Frankia Ccl3 Frankia EAN1pec	OsaABC triggers	osmolytes (KCI, sucrose)		σ ^в -dependent transcription
"S. coelico	S. coelicolor S. avermitilis S. griseus T. fusca S. tropica S. arenicola				
" <i>B. cereus</i> module"	B. cereus B. thuringiensis B. weihenstephanensis B. anthractephanensis Geobacillus Y412MC10 Paenibacillus JDR-2	RsbKY triggers	heat osmolytes (NaCl) ethanol outside		σ ^B -dependent transcription

Fig. 6.3. Established and predicted signalling routes for the control of σ^{B} -like stress-responsive sigma factors in different Gram-positive bacteria.

The partially established σ^{B} activation routes of *B. cereus* and *S. coelicolor* involve homologous hybrid HKs, whereas *B. subtilis* uses two markedly different pathways for the same purpose (see below). Protein domain colourings are the same as in Fig. 6.1. PP2C-type phosphatase domains are indicated in green. Locations of stimulus perception are indicated by "explosion" symbols.

Left: The "B. cereus module" for σ^{B} control involves the hybrid HK RsbK and its putative cognate RR RsbY (de Been *et al.*, unpublished results; van Schaik *et al.*, 2005a). Upon different stresses, RsbK most likely "auto"-phosphorylates on a conserved histidine residue within its H-box, from which phosphoryl is transferred to a conserved asparate within the REC domain of RsbY. This activates the RsbY PP2C domain, which dephosphorylates RsbV, ultimately resulting in the activation of σ^{B} . The RsbK C-terminal REC domain is crucial for proper σ^{B} induction and probably plays a role in fine-tuning RsbK kinase activity (dashed lines) (de Been *et al.*, unpublished results). Another fine-tuning system may include CheRmediated methylation processes of RsbK (this study). Other species, like *Geobacillus* and *Paenibacillus*, may also use the RsbKY module for the control of σ^{B} .

Middle: The "S. coelicolor" module for σ^{B} control involves the RsbK homolog OsaA that analogous to RsbK - most likely directly activates its cognate RR OsaB. In turn, OsaB activates OsaC either directly or indirectly. The PP2C-type phosphatase domain of OsaC is essential for σ^{B} activation. OsaC also contains an N-terminal RsbW-like anti-sigma factor that is involved in the post-osmotic stress control of σ^{B} (Martínez *et al.*, 2009; Perez *et al.*, 2006). OsaABC modules occur in several actinobacteria, where they may also control σ^{B} activity (this study).

Right: The extensively studied "*B. subtilis* modules" for the control of σ^{B} do not involve the use of RsbK-type HKs. The *B. subtilis* stressosome is a ~1.8 MD supramolecular complex that consists of multiple copies of RsbR and RsbS (Chen *et al.*, 2003a; Delumeau *et al.*, 2006; Dufour *et al.*, 1996; Marles-Wright *et al.*, 2008). Different environmental stresses and signals stimulate RsbT kinase activity towards its stressosome substrates (Akbar *et al.*, 2001; Ávila-Pérez *et al.*, 2006; Gaidenko *et al.*, 2006; Kim *et al.*, 2004; Voelker *et al.*, 1995). This leads to dissociation of RsbT from the stressosome to activate RsbU and subsequently σ^{B} (Delumeau *et al.*, 2004). Return of the stressosome into the unphosphorylated state is achieved by RsbX (Chen *et al.*, 2004a; Yang *et al.*, 1996). A second *B. subtilis* σ^{B} activation pathway involves the hydrolase or acyltransferase RsbQ and the PP2C-type phosphatase RsbP. The RsbQP route is triggered upon energy stress, which is either directly sensed by RsbP or is signalled through RsbQ (Brody *et al.*, 2001; Vijay *et al.*, 2000). *, RsbQP-dependent σ^{B} activation has so far only been found in *B. subtilis*.

Chapter 7

Summarising discussion

This thesis describes the functional analysis of two-component signal transduction systems (TCS) in the food-borne pathogen *Bacillus cereus* and related low-GC Gram-positives. It shows the potential of bioinformatics and comparative genomics approaches combined with molecular and transcriptome techniques for elucidating the functional roles of uncharacterised regulatory proteins. The most important results are summarised in Fig. 7.1. This figure is used as a guide in the summary and discussion below. In addition, implications of the results for future research on TCSs of *B. cereus* and other bacteria as well as implications concerning eventual applications in the control and monitoring of pathogenic bacteria, including food-borne pathogens, are briefly discussed.

Initial identification of *B. cereus* group TCSs

B. cereus is recognised as an important cause of food-poisoning and food spoilage worldwide. Within the genus *Bacillus*, *B. cereus* and its closest relatives form a highly homogeneous subdivision, which has been termed the "*B. cereus* group". This group contains other economically and medically important species, including *B. anthracis*, a notorious pathogen that can cause the often fatal disease anthrax in mammals, *B. thuringiensis*, an insect pathogen that is widely used as an insecticide, and *B. weihenstephanensis* that can grow and spoil foods at refrigerator temperatures. Members of the *B. cereus* group are ubiquitously present in the environment and can adapt to a considerable range of environmental challenges (Abee and Wouters, 1999; Jensen *et al.*, 2003; Kotiranta *et al.*, 2000). In bacteria, monitoring of environmental challenges is generally mediated by TCSs, which consist of a sensor histidine kinase (HK) and a partner response regulator (RR).

Because almost nothing was known about TCSs in *B. cereus* and its closest relatives at the onset of this research project, we decided to first identify, and make functional predictions for, the complete arsenal of TCS proteins in members of the B. cereus group. At the time, the complete genome sequence information of eight B. cereus group members was publicly available, including three B. cereus strains (ATCC 14579, ATCC 10987 and E33L or ZK), four B. anthracis strains (Ames 0581, Ames, Sterne and A2012) and one B. thuringiensis strain (konkukian). These genome sequences formed the basis for our analysis of which the results are described in chapter 2. By using Pfam Hidden Markov Models (Bateman et al., 2004), a large set of TCS proteins was identified in the different B. cereus group genomes: 50-58 HKs and 48-52 RRs, divided over 68 and 60 orthologous groups, respectively. Based on the classification scheme of Grebe and Stock (1999), these HKs and RRs were classified into ten major evolutionary subfamilies or classes, of which some could be readily linked to specific biological roles (Fig. 7.1). The majority of the HKand RR-encoding genes (around 40 of each) were found in pairs of adjacent genes, each pair probably encoding one functional TCS. However, we also found a substantial amount of HKs (10-14 per genome) and RRs (7-11) of which the associated genes did not pair with a specific partner gene. To match these socalled "orphans" to potential partner proteins, we performed a congruence-oftrees analysis in combination with literature data-mining. This revealed putative

partners for most of the "orphan" HKs and RRs (Fig. 7.1, connections in grey). Notably, we predicted that members of the *B. cereus* group contain a large number of class four "orphan" HKs (9-14) that can all feed phosphoryl into the Spo0F-Spo0B-Spo0A sporulation initiation cascade (Fig. 7.1, top right). This prediction was further supported by the fact that the HKs most closely related to the B. cereus group class four HKs were only found in other endospore-forming species. In addition, five of the class four HKs have been experimentally confirmed to play a role in sporulation processes in *B. anthracis* (HK39, 40, 48, 49, 50) (Brunsing et al., 2005) and B. thuringiensis (HK39) (Malvar et al., 1994). In addition to the sporulation HKs, predicted biological roles for other B. cereus group TCS proteins included roles in biofilm formation, host-microbe interactions, chemotaxis, nutrient uptake, (I)antibiotic resistance, and many more (Fig. 7.1). As compared to other low-GC Gram-positives, the number of TCS proteins in members of the *B. cereus* group appears to be relatively large. This large number probably directly relates to the relatively large genome sizes of B. cereus group members, as larger genomes tend to encode relatively more signal transduction proteins. Similarly, free-living bacteria tend to encode more signal transduction proteins than highly specialised pathogenic bacteria, which generally have smaller genomes (Galperin, 2005). Considering this, it was of special interest to find that B. anthracis contains a large number of truncated, putatively non-functional, HK and RR genes and completely lacks specific HK and RR genes. A possible scenario is that specialisation of *B. anthracis* as a pathogen has reduced the range of environmental stimuli it encounters, ultimately resulting in the evolutionary disposal of specific TCS genes.

DNA-binding RRs and TCS regulons

To gain more insights into the functionalities of the large number of TCSs detected, we decided to further focus on the *B. cereus* group DNA-binding RRs, their target sites on the DNA (operators) and the associated target genes. To predict RR-specific operators, we designed and implemented a phylogenetic footprinting/shadowing approach. In short: the C-terminal output domain of each RR was scanned against a non-redundant database of low-GC Grampositive genomes and protein hits were used to construct bootstrapped neighbour-joining trees, one for each of the seven RR subfamilies: OmpR, NarL, CitB, LytTR, GInL, Spo0A and AraC. The trees were used to identify groups of orthologous functional equivalent (GOOFE) RRs on the basis of gene context conservation. The upstream regions of putative RR target genes within these GOOFEs were used to predict RR-specific operators. The results of this analysis are described in chapter 3. A large number of novel RR-specific operators was predicted for more than 50% of the DNA-binding RRs of the B. cereus group (Fig. 7.1). In all cases where we found appropriate experimental data, we had identified the operator(s) and corresponding target gene(s) correctly. Analysis of the operator sequences revealed characteristic traits for each RR subfamily. For instance, operators related to the largest subfamily (OmpR) typically consisted of short direct repeats (e.g. TTAAGA-N5-TTAAGA); whereas operators related to the second largest family (NarL) consisted of short inverted repeats (e.g. ATGACA-N2-TGTCAT). These differences indicated a fundamentally different

organisation of the bound RR dimers between the two subfamilies. In addition, we used the RR-specific operator motifs for the genome-wide prediction of additional RR target genes. This allowed relating several RRs to a specific minimal regulon and thereby to a characteristic transcriptional response. Interestingly, many of the predicted regulons comprised genes encoding transport systems. Apparently, stimulus perception by the HK seems to generally lead to activation of a dedicated transport system that either imports (nutrient) or exports (waste product or toxic substance) the compound that constituted or caused the presence of the initial stimulus. Based on the minimal regulon predictions, new biological roles were attributed to various TCSs, including roles in cytochrome c biogenesis (HssRS), transport of carbohydrates, peptides and/or amino acids (YkoGH), and resistance to toxic ions (LiaSR), antimicrobial peptides (BceRS) and beta-lactam antibiotics (BacRS, YcbLM). The success of our phylogenetic footprinting approach is largely dependent on the occurrence of gene context conservation between orthologous regulator genes. Indeed, in most cases (> 75%) where genomic associations between TCS and other genes were found conserved across species, we could retrieve RR-specific operators. For this reason, the availability of more and more bacterial genome sequences is bound to even further increase the success of our phylogenetic footprinting approach.

In chapter 4, we have focused on the TCSs YvrHG and YufLM (Fig. 7.1). The study described in this chapter was partly initiated to confirm and elaborate on the RR-specific operator and TCS regulon predictions described in chapter 3. To do this, deletion mutants of the two respective TCSs were constructed and the transcriptome of each B. cereus mutant strain was compared to that of the wildtype B. cereus strain. This revealed that the respective TCSs were significantly active during exponential growth in rich medium (i.e. BHI). Moreover, analysis of the transcriptome data revealed that our footprinting-based RR operator predictions were correct for both TCSs. Because of this accuracy, we used the predicted RR-specific operators in combination with the generated transcriptome data to guide the identification of additional, more degenerate, operators and associated target genes. This resulted in the identification of more extended TCS regulons. Based on the newly established regulons, biological roles were connected to the two TCSs. These roles included the participation in an intricate genetic network involved in antibiotic resistance, including the confirmed resistance to oxolinic acid (YvrHG) and the confirmed uptake and metabolism of fumarate and the simultaneous repression of fermentation processes (YufLM). The results presented illustrate the power of integrating in silico operator predictions with genome-wide transcriptome analyses. Furthermore, because the two TCSs studied were active in exponentially growing cells in "standard" nutrient-rich medium, the results show that specialised knowledge regarding TCS-specific activating stimuli is not always required to study the functional role of a TCS.

RsbK and the major stress-responsive alternative sigma factor, σ^{B}

Previous work in our laboratory has revealed that one of the key stressresponsive pathways in B. cereus is mediated by the a-typical RR RsbY, which harbours a C-terminal PP2C-type phosphatase domain. Under stressful conditions, this domain is assumed to dephosphorylate the anti-sigma factor antagonist RsbV, resulting in the activation of the major stress-responsive alternative sigma factor, σ^B , which in turn controls a specialised regulon (van Schaik et al., 2004; van Schaik et al., 2005a). Until recently, a partner HK for RsbY had remained unidentified. However, the results described in chapter 5 provide strong indications that RsbY receives its input from the multi-sensor hybrid kinase BC1008 (RsbK), which contains a HK as well as a RR phosphotransferase domain. Genome analyses revealed that *rsbY* is located in a conserved gene cluster which includes *rsbK*. This gene cluster was highly conserved across members of the *B*, cereus group, but was also found in four Bacillus species outside this group. Further support for a link between RsbK, RsbY and σ^{B} came from analyses of available transcriptome datasets, which revealed that *rsbK*, *rsbY* and *sigB* displayed similar expression patterns under different experimental conditions. This led us to further study the predicted role of RsbK in regulating the σ^{B} -mediated stress response. Indeed, an *rsbK* deletion mutant appeared incapable of inducing σ^{B} levels upon mild heat and salt stress conditions. In addition, the *rsbK* deletion strain was impaired in its heat adaptive response. Comparison of the wild-type and rsbK mutant transcriptomes upon heat shock revealed that RsbK was primarily involved in the activation of the σ^{B} mediated stress response. By integrating the transcriptome data with available transcriptome data from the rsbY and sigB deletion mutants (van Schaik et al., 2007), we were able to extend the σ^{B} regulon with new members, including a ferrochelatase- and a catalase-encoding gene. Finally, truncation of the RsbK RR phosphotransferase (receiver) domain demonstrated the importance of this domain for proper σ^{B} induction upon stress. Although σ^{B} induction was clearly disturbed and reduced in this "truncated" strain, σ^{B} protein levels were significantly higher than those observed in the *rsbK* deletion strain, suggesting that the truncated version of RsbK was still capable of activating (i.e. phosphorylating) its partner RR RsbY. These results indicated that the fused RR receiver domain may function in fine-tuning the RsbK kinase activity towards RsbY, comparable to what has been found for the TCS VirAG of Agrobacterium tumefaciens. Under non-stimulating conditions, the fused RR receiver domain of the hybrid HK VirA specifically inhibits VirA kinase activity towards the cognate RR VirG, thus preventing improper stimulation of the signalling pathway (Brencic et al., 2004; Chang and Winans, 1992; Chang et al., 1996). Although the VirAGtype of phosphoryl transfer seems likely, future mutation, interaction and phosphorylation essays must prove the validity of this scenario for RsbKY. Such studies may also point to an alternative signalling route involving a multistep His-Asp-His-Asp phosphorelay, similar to the SIn-Ypd-Ssk phosphorelay in yeast (Posas et al., 1996). This scenario would require a third (Hpt-like) phosphotransferase that mediates phosphoryl transfer between RsbK and RsbY. However, screening the B. cereus group genomes for such phosphotransferases did not reveal any obvious candidates, except for SpoOB and (the N-terminus) of

CheA, which are known to function in sporulation initiation and chemotaxis, respectively (Burbulys *et al.*, 1991; Szurmant and Ordal, 2004) (Fig. 7.1).

Although conserved rsbK/Y gene clusters were found in bacilli outside the B. cereus group, the gene clusters in these bacilli were not genomically associated to sigB as was the case in the B. cereus group. Therefore, it remained unclear whether RsbKY-mediated activation of σ^{B} is a unique trait for members of the *B*. cereus group. To answer this question, we performed a comparative genome analysis focusing on the phosphotransferase domains of RsbK. The results of these analyses are described in **chapter 6**. We found that RsbK-type hybrid HKs are conserved across a wide variety of bacterial species. Genome context and protein sequence analyses of 118 close RsbK homologues revealed an extreme variability in N-terminal sensory as well as C-terminal regulatory domains and suggested that RsbK-type HKs are subject to complex fine-tuning systems, including putative methylation and demethylation processes mediated by CheRand CheB-type proteins, respectively. Finally, we provide strong indications that the use of RsbK-type HKs for the control of σ^{B} -like sigma factors is not restricted to members of the B. cereus group, but also occurs in at least several low- as well as high-GC Gram-positives, including Geobacillus, Paenibacillus and actinobacteria. These findings indicate that these Gram-positive bacteria have evolved or obtained a common solution for monitoring the same process: the activation of a stress-responsive alternative sigma factor. This solution is markedly different from the extensively studied and highly conserved RsbRSTmediated o^B activation route found in *B. subtilis* and related low-GC Grampositives.

Fig. 7.1. Overview of TCSs in eight B. cereus group members.

HKs are indicated in red (sensory) and ecru (phosphotransferase) and RRs are indicated in grey (receiver) and green (DNA-binding domain). Blue/purple domains indicate proteinprotein interaction domains. White and light (red, green, blue/purple) domains indicate the absence/truncation of the respective domain in at least two of the B. cereus group members analysed. Coloured bars surrounding the cell represent the ten different HK subfamilies. Incoming arrows indicate predicted or established HK-specific stimuli. Inside the cell, black and grey arrows represent predicted phosphotransfers between paired and "orphan" HKs and RRs, respectively. RRs for which specific operators were predicted are "connected" to the DNA. Other connection lines illustrate protein-protein interactions. TCSs/proteins that have been experimentally studied by us and that are described in this thesis are highlighted with red asterisks. Those studied elsewhere are highlighted with black asterisks. These include: RsbY (van Schaik et al., 2005a), SctRS (Marraffini and Schneewind, 2006), YvrHG (Fagerlund et al., 2007), HssRS (Stauff and Skaar, 2009), ResDE (Duport et al., 2006; Esbelin et al., 2009; Vetter and Schlievert, 2007; Wilson et al., 2008), several sporulation HKs (Brunsing et al., 2005; Malvar et al., 1994; Scaramozzino et al., 2009; White et al., 2006), Spo0B (Mattoo et al., 2008) and YvfTU (Brillard et al., 2008).

Abbreviations: AIP, auto-inducing peptide; MCP, methyl-accepting chemotaxis protein receptor; CAMP, cationic antimicrobial peptide; TCA, tricarboxylic acid; PMF, proton-motive force.







Fig. 7.2. Domain architecture and topology of the *B. cereus* group HK sensory regions.

The images represent the regions on the N-terminal side of the HK H-box. Black circle: PAS/PAC domain-containing HKs. Grey circle: GAF domain-containing HKs. Domains drawn farther away from each other are separated by ~150 residues or more. Domains drawn closer to each other are separated by fewer than ~50 residues, or by fewer than ~20 residues in the case of transmembrane helices (TMH). Texts below the images indicate in which HK subfamilies the respective domain architecture was found. Numbers show how many times this architecture was found for the subfamily indicated. For example, we found that two of the fourteen HKs of subfamily 4 contained two PAS/PAC domains in their N-terminus (top, right). Abbreviations: SH, signalling-helix; PMH, putative methylation helix (chapter 6). *, the 5TMR_LYT domain harbours three TMHs. Picture adapted from: Schürch, A., de Been, M., Francke, C., unpublished results.

Challenges in TCS research

The results described in this thesis provide extensive insight into the role of TCS-mediated signalling in *B. cereus*. Still, many aspects of these signalling modules remain to be further elucidated, especially with respect to the mechanisms underlying HK stimulus perception and the subsequent information transfer towards downstream HK domains. A valuable first step in understanding these mechanistic principles is the identification of conserved domains and residues located within the HK N-terminal sensory regions. As illustrated in



figure 7.2, many putative sensory domains lie still "hidden" in the N-terminal regions of the B. cereus group HKs. Especially for the HKs of subfamily 1a, we frequently found large regions (>150 aa) that contained no detectable Pfam (Bateman et al., 2004) or SMART (Letunic et al., 2009) domains and that were located between the only two transmembrane helices, the apparent location of environmental stimulus perception (Fig. 7.2). A useful approach to identify sensory domains is to build phylogenetic trees of HK conserved phosphotransferase domains and compare the associated N-terminal regions of related HKs. This may be especially useful for detecting conserved sensory domains with low overall sequence similarity. In a similar way, we found the conserved putative methylation helices in the RsbK-type HKs (chapter 6 and Fig. 7.2). Still, even with the identification of a conserved domain, the nature of the perceived stimulus often remains obscure, as many sensory domains have been implicated in the perception of a wide variety of stimuli. For example, the most frequently detected sensory domains in the B. cereus group HKs were the PAS and GAF domains (Fig. 7.2), each of which have been implicated in the sensing of widely different factors, including redox/light/metabolites and small ligands/cyclic nucleotides/photopigments, respectively (Galperin, 2004). Evidently, further experimental and structural studies on the sensory domains are necessary to gain a proper view on the mode of HK stimulus perception. Still, other bioinformatics and genomics approaches that can provide valuable leads about the perceived stimuli include genome context analyses and TCS regulon predictions (chapters 3 and 4). In addition, many TCSs are involved in quorum-sensing and their HKs directly sense small auto-inducing peptides that are produced by the bacterial cell (Lyon and Novick, 2004; Sturme et al., 2007). In general, identification of the genes encoding these signalling peptides is still a major challenge that may be partly solved by novel in silico techniques.

Potential applications and concluding remarks

Because TCSs do not occur in animals, they form attractive targets for novel bioactives against pathogenic bacteria (Watanabe et al., 2008). Considering that the HK sensory domains form the most easily accessible targets for these bioactives, knowledge about the mode of HK stimulus perception is of crucial importance. This may lead to the rational selection and design of bioactives that can block HK stimulus peception and inhibit downstream signalling cascades that lead to unwanted bacterial behaviour. The potential of such a strategy has been shown in studies on rationally designed synthetic peptides that were able to block quorum-sensing signalling processes in Staphylococcus aureus (Lyon et al., 2000; Otto, 2004; Qiu et al., 2003; Scott et al., 2003). Although current advances in these strategies are still mainly made in the medical field, it is conceivable that they will eventually also lead to applications in other fields, such as those related to food-processing. Finally, specific TCSs and their associated target genes may serve as suitable biomarkers to rapidly determine the stress-responsive state of the bacterial cell. In the case of B. cereus, this may reveal the early onset of sporulation processes or an early increase in stress-responsiveness. Eventually, this may lead to novel intervention strategies for an improved control of *B. cereus*.

References

- Abee, T., and Wouters, J.A. (1999) Microbial stress response in minimal processing. *Int J* Food Microbiol **50**: 65-91.
- Agaisse, H., and Lereclus, D. (1994) Structural and functional analysis of the promoter region involved in full expression of the *cryIIIA* toxin gene of *Bacillus thuringiensis*. *Mol Microbiol* **13**: 97-107.
- Agata, N., Ohta, M., Mori, M., and Isobe, M. (1995) A novel dodecadepsipeptide, cereulide, is an emetic toxin of *Bacillus cereus. FEMS Microbiol Lett* **129**: 17-20.
- Aguilar, P.S., Hernandez-Arriaga, A.M., Cybulski, L.E., Erazo, A.C., and de Mendoza, D. (2001) Molecular basis of thermosensing: a two-component signal transduction thermometer in *Bacillus subtilis*. *Embo J* **20**: 1681-1691.
- Akbar, S., Gaidenko, T.A., Kang, C.M., O'Reilly, M., Devine, K.M., and Price, C.W. (2001) New family of regulators in the environmental signalling pathway which activates the general stress transcription factor o^B of *Bacillus subtilis*. J Bacteriol **183**: 1329-1338.
- Alm, E., Huang, K., and Arkin, A. (2006) The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLoS Comput Biol* 2: e143.
- Alm, R.A., Bodero, A.J., Free, P.D., and Mattick, J.S. (1996) Identification of a novel gene, *pilZ*, essential for type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*. J Bacteriol **178**: 46-53.
- Alper, S., Dufour, A., Garsin, D.A., Duncan, L., and Losick, R. (1996) Role of adenosine nucleotides in the regulation of a stress-response transcription factor in *Bacillus subtilis*. J Mol Biol 260: 165-177.
- Altschul, S.F., Madden, T.L., Schäffer, 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.
- Anantharaman, V., and Aravind, L. (2000) Cache a signalling domain common to animal Ca(2+)-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends Biochem Sci* 25: 535-537.
- Anantharaman, V., Balaji, S., and Aravind, L. (2006) The signalling helix: a common functional theme in diverse signalling proteins. *Biol Direct* **1**: 25.
- Anantharaman, V., Koonin, E.V., and Aravind, L. (2001) Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol **307**: 1271-1292.
- Anderson, I., Sorokin, A., Kapatral, V., Reznik, G., Bhattacharya, A., Mikhailova, N., Burd, H., Joukov, V., Kaznadzey, D., Walunas, T., Markd'Souza, Larsen, N., Pusch, G., Liolios, K., Grechkin, Y., Lapidus, A., Goltsman, E., Chu, L., Fonstein, M., Ehrlich, S.D., Overbeek, R., Kyrpides, N., and Ivanova, N. (2005) Comparative genome analysis of *Bacillus cereus* group genomes with *Bacillus subtilis. FEMS Microbiol Lett* **250**: 175-184.
- Appleby, J.L., Parkinson, J.S., and Bourret, R.B. (1996) Signal transduction via the multistep phosphorelay: not necessarily a road less traveled. *Cell* **86**: 845-848.
- Aravind, L., and Ponting, C.P. (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem Sci* **22**: 458-459.
- Aravind, L., and Ponting, C.P. (1999) The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol Lett* **176**: 111-116.
- Arnaud, M., Chastanet, A., and Débarbouillé, M. (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl Environ Microbiol* **70**: 6887-6891.
- Aronson, A.I., and Shai, Y. (2001) Why *Bacillus thuringiensis* insecticidal toxins are so effective: unique features of their mode of action. *FEMS Microbiol Lett* **195**: 1-8.

- Arthur, M., and Quintiliani, R., Jr. (2001) Regulation of VanA- and VanB-type glycopeptide resistance in enterococci. *Antimicrob Agents Chemother* **45**: 375-381.
- Asai, K., Baik, S.H., Kasahara, Y., Moriya, S., and Ogasawara, N. (2000) Regulation of the transport system for C4-dicarboxylic acids in *Bacillus subtilis*. *Microbiology* **146 (Pt 2)**: 263-271.
- Asano, S.I., Nukumizu, Y., Bando, H., Iizuka, T., and Yamamoto, T. (1997) Cloning of novel enterotoxin genes from *Bacillus cereus* and *Bacillus thuringiensis*. *Appl Environ Microbiol* **63**: 1054-1057.
- Ash, C., Farrow, J.A., Dorsch, M., Stackebrandt, E., and Collins, M.D. (1991) Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. *Int J Syst Bacteriol* **41**: 343-346.
- Ávila-Pérez, M., Hellingwerf, K.J., and Kort, R. (2006) Blue light activates the σ^Bdependent stress response of *Bacillus subtilis* via YtvA. *J Bacteriol* **188**: 6411-6414.
- Baida, G., Budarina, Z.I., Kuzmin, N.P., and Solonin, A.S. (1999) Complete nucleotide sequence and molecular characterization of hemolysin II gene from *Bacillus cereus*. *FEMS Microbiol Lett* **180**: 7-14.
- Baida, G.E., and Kuzmin, N.P. (1995) Cloning and primary structure of a new hemolysin gene from *Bacillus cereus*. *Biochim Biophys Acta* **1264**: 151-154.
- Baikalov, I., Schröder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R.P., and Dickerson, R.E. (1996) Structure of the *Escherichia coli* response regulator NarL. *Biochemistry* 35: 11053-11061.
- 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.
- Bailey, T.L., and Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**: 48-54.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., and Eddy, S.R. (2004) The Pfam protein families database. *Nucleic Acids Res* 32: D138-141.
- Bavykin, S.G., Lysov, Y.P., Zakhariev, V., Kelly, J.J., Jackman, J., Stahl, D.A., and Cherni, A. (2004) Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. *J Clin Microbiol* 42: 3711-3730.
- Bearson, B.L., Wilson, L., and Foster, J.W. (1998) A low pH-inducible, PhoPQ-dependent acid tolerance response protects *Salmonella typhimurium* against inorganic acid stress. *J Bacteriol* **180**: 2409-2417.
- Beecher, D.J., and Wong, A.C. (1994) Identification of hemolysin BL-producing *Bacillus cereus* isolates by a discontinuous hemolytic pattern in blood agar. *Appl Environ Microbiol* **60**: 1646-1651.
- Beecher, D.J., Schoeni, J.L., and Wong, A.C. (1995) Enterotoxic activity of hemolysin BL from *Bacillus cereus*. *Infect Immun* **63**: 4423-4428.
- Beecher, D.J., and Wong, A.C. (1997) Tripartite hemolysin BL from *Bacillus cereus*. Hemolytic analysis of component interactions and a model for its characteristic paradoxical zone phenomenon. *J Biol Chem* **272**: 233-239.
- Beecher, D.J., Olsen, T.W., Somers, E.B., and Wong, A.C. (2000) Evidence for contribution of tripartite hemolysin BL, phosphatidylcholine-preferring phospholipase C, and collagenase to virulence of *Bacillus cereus* endophthalmitis. *Infect Immun* 68: 5269-5276.
- Benson, A.K., and Haldenwang, W.G. (1993) *Bacillus subtilis* sigma B is regulated by a binding protein (RsbW) that blocks its association with core RNA polymerase. *Proc Natl Acad Sci U S A* **90**: 2330-2334.
- Berry, C., O'Neil, S., Ben-Dov, E., Jones, A.F., Murphy, L., Quail, M.A., Holden, M.T., Harris, D., Zaritsky, A., and Parkhill, J. (2002) Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. israelensis. *Appl Environ Microbiol* **68**: 5082-5095.

Bijlsma, J.J., and Groisman, E.A. (2003) Making informed decisions: regulatory interactions between two-component systems. *Trends Microbiol* **11**: 359-366.

- Bishop, A., Fielding, S., Dyson, P., and Herron, P. (2004) Systematic insertional mutagenesis of a streptomycete genome: a link between osmoadaptation and antibiotic production. *Genome Res* 14: 893-900.
- Blanco, A.G., Sola, M., Gomis-Rüth, F.X., and Coll, M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* **10**: 701-713.

Bone, E.J., and Ellar, D.J. (1989) Transformation of *Bacillus thuringiensis* by electroporation. *FEMS Microbiol Lett* **49**: 171-177.

- Bouillaut, L., Ramarao, N., Buisson, C., Gilois, N., Gohar, M., Lereclus, D., and Nielsen-Leroux, C. (2005) FlhA influences *Bacillus thuringiensis* PlcR-regulated gene transcription, protein production, and virulence. *Appl Environ Microbiol* **71**: 8903-8910.
- Brencic, A., Xia, Q., and Winans, S.C. (2004) VirA of *Agrobacterium tumefaciens* is an intradimer transphosphorylase and can actively block *vir* gene expression in the absence of phenolic signals. *Mol Microbiol* **52**: 1349-1362.
- Brillard, J., Susanna, K., Michaud, C., Dargaignaratz, C., Gohar, M., Nielsen-Leroux, C., Ramarao, N., Kolstø, A.B., Nguyen-the, C., Lereclus, D., and Broussolle, V. (2008) The YvfTU two-component system is involved in *plcR* expression in *Bacillus cereus*. *BMC Microbiol* 8: 183.
- Brody, M.S., Vijay, K., and Price, C.W. (2001) Catalytic function of an α/β hydrolase is required for energy stress activation of the σ^{B} transcription factor in *Bacillus subtilis*. *J Bacteriol* **183**: 6422-6428.
- Bron, S. (1990) Plasmids. In *Molecular biological methods for Bacillus*. Harwood, C.R. and Cutting, S.M. (eds). Chichester: John Wiley & Sons Ltd., pp. 75-174.
- Brunsing, R.L., La Clair, C., Tang, S., Chiang, C., Hancock, L.E., Perego, M., and Hoch, J.A. (2005) Characterization of sporulation histidine kinases of *Bacillus anthracis*. J *Bacteriol* **187**: 6972-6981.
- Brunskill, E.W., and Bayles, K.W. (1996) Identification and molecular characterization of a putative regulatory locus that affects autolysis in *Staphylococcus aureus*. J Bacteriol **178**: 611-618.
- Burbulys, D., Trach, K.A., and Hoch, J.A. (1991) Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay. *Cell* **64**: 545-552.
- Carlson, C.R., Caugant, D.A., and Kolstø, A.B. (1994) Genotypic Diversity among *Bacillus* cereus and *Bacillus thuringiensis* Strains. *Appl Environ Microbiol* **60**: 1719-1725.
- Chang, C.H., and Winans, S.C. (1992) Functional roles assigned to the periplasmic, linker, and receiver domains of the *Agrobacterium tumefaciens* VirA protein. *J Bacteriol* **174**: 7033-7039.
- Chang, C.H., Zhu, J., and Winans, S.C. (1996) Pleiotropic phenotypes caused by genetic ablation of the receiver module of the *Agrobacterium tumefaciens* VirA protein. *J Bacteriol* **178**: 4710-4716.
- Chaturongakul, S., Raengpradub, S., Wiedmann, M., and Boor, K.J. (2008) Modulation of stress and virulence in *Listeria monocytogenes*. *Trends Microbiol* **16**: 388-396.
- Chen, C.C., Lewis, R.J., Harris, R., Yudkin, M.D., and Delumeau, O. (2003a) A supramolecular complex in the environmental stress signalling pathway of *Bacillus subtilis*. *Mol Microbiol* **49**: 1657-1669.
- Chen, C.C., Yudkin, M.D., and Delumeau, O. (2004a) Phosphorylation and RsbX-dependent dephosphorylation of RsbR in the RsbR-RsbS complex of *Bacillus subtilis*. *J Bacteriol* **186**: 6830-6836.
- Chen, Y., Succi, J., Tenover, F.C., and Koehler, T.M. (2003b) Beta-lactamase genes of the penicillin-susceptible *Bacillus anthracis* Sterne strain. *J Bacteriol* **185**: 823-830.

- Chen, Y., Tenover, F.C., and Koehler, T.M. (2004b) Beta-lactamase gene expression in a penicillin-resistant *Bacillus anthracis* strain. *Antimicrob Agents Chemother* **48**: 4873-4877.
- Cho, Y.H., Lee, E.J., Ahn, B.E., and Roe, J.H. (2001) SigB, an RNA polymerase sigma factor required for osmoprotection and proper differentiation of *Streptomyces coelicolor*. *Mol Microbiol* **42**: 205-214.
- Comenge, Y., Quintiliani, R., Jr., Li, L., Dubost, L., Brouard, J.P., Hugonnet, J.E., and Arthur, M. (2003) The CroRS two-component regulatory system is required for intrinsic beta-lactam resistance in *Enterococcus faecalis*. J Bacteriol **185**: 7184-7192.
- Cotter, P.D., Emerson, N., Gahan, C.G., and Hill, C. (1999) Identification and disruption of *lisRK*, a genetic locus encoding a two-component signal transduction system involved in stress tolerance and virulence in *Listeria monocytogenes*. *J Bacteriol* **181**: 6840-6843.
- Cotter, P.D., Guinane, C.M., and Hill, C. (2002) The LisRK signal transduction system determines the sensitivity of *Listeria monocytogenes* to nisin and cephalosporins. *Antimicrob Agents Chemother* **46**: 2784-2790.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-1190.
- Cruz Ramos, H., Hoffmann, T., Marino, M., Nedjari, H., Presecan-Siedel, E., Dreesen, O., Glaser, P., and Jahn, D. (2000) Fermentative metabolism of *Bacillus subtilis*: physiology and regulation of gene expression. *J Bacteriol* **182**: 3072-3080
- Cybulski, L.E., del Solar, G., Craig, P.O., Espinosa, M., and de Mendoza, D. (2004) *Bacillus subtilis* DesR functions as a phosphorylation-activated switch to control membrane lipid fluidity. *J Biol Chem* **279**: 39340-39347.
- Daffonchio, D., Cherif, A., and Borin, S. (2000) Homoduplex and heteroduplex polymorphisms of the amplified ribosomal 16S-23S internal transcribed spacers describe genetic relationships in the "*Bacillus cereus* group". *Appl Environ Microbiol* **66**: 5460-5468.
- Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* **23**: 324-328.
- 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.
- de Been, M., Francke, C., Moezelaar, R., Abee, T., and Siezen, R.J. (2006) Comparative analysis of two-component signal transduction systems of *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus anthracis*. *Microbiology* **152**: 3035-3048.
- de Been, M., Tempelaars, M.H., van Schaik, W., Moezelaar, R., Siezen, R.J., and Abee, T. (unpublished results) A novel hybrid kinase is essential for activating the σ^B-mediated stress response of *Bacillus cereus*.
- Delumeau, O., Dutta, S., Brigulla, M., Kuhnke, G., Hardwick, S.W., Voelker, U., Yudkin, M.D., and Lewis, R.J. (2004) Functional and structural characterization of RsbU, a stress signaling protein phosphatase 2C. J Biol Chem 279: 40927-40937.
- Delumeau, O., Chen, C.C., Murray, J.W., Yudkin, M.D., and Lewis, R.J. (2006) Highmolecular-weight complexes of RsbR and paralogues in the environmental signaling pathway of *Bacillus subtilis*. J Bacteriol **188**: 7885-7892.
- de Vries, Y.P., Hornstra, L.M., Atmadja, R.D., van Schaik, W., de Vos, W.M., and Abee, T. (2005) Deletion of *sigB* in *Bacillus cereus* affects spore properties. *FEMS Microbiol Lett* **252**: 169-173.
- Di Franco, C., Beccari, E., Santini, T., Pisaneschi, G., and Tecce, G. (2002) Colony shape as a genetic trait in the pattern-forming *Bacillus mycoides*. *BMC Microbiol* **2**: 33.
- Dierick, K., Van Coillie, E., Swiecicka, I., Meyfroidt, G., Devlieger, H., Meulemans, A., Hoedemaekers, G., Fourie, L., Heyndrickx, M., and Mahillon, J. (2005) Fatal family outbreak of *Bacillus cereus*-associated food poisoning. *J Clin Microbiol* **43**: 4277-4279.

- Doan, T., Servant, P., Tojo, S., Yamaguchi, H., Lerondel, G., Yoshida, K., Fujita, Y., and Aymerich, S. (2003) The *Bacillus subtilis ywkA* gene encodes a malic enzyme and its transcription is activated by the YufL/YufM two-component system in response to malate. *Microbiology* **149**: 2331-2343.
- Dsouza, M., Larsen, N., and Overbeek, R. (1997) Searching for patterns in genomic data. *Trends Genet* **13**: 497-498.
- Dubrac, S., and Msadek, T. (2004) Identification of genes controlled by the essential YycG/YycF two-component system of *Staphylococcus aureus*. J Bacteriol **186**: 1175-1181.
- Dufour, A., and Haldenwang, W.G. (1994) Interactions between a *Bacillus subtilis* antisigma factor (RsbW) and its antagonist (RsbV). *J Bacteriol* **176**: 1813-1820.
- Dufour, A., Voelker, U., Voelker, A., and Haldenwang, W.G. (1996) Relative levels and fractionation properties of *Bacillus subtilis* σ^{B} and its regulators during balanced growth and stress. *J Bacteriol* **178**: 3701-3709.
- Duplessis, M.R., Karol, K.G., Adman, E.T., Choi, L.Y., Jacobs, M.A., and Cattolico, R.A. (2007) Chloroplast His-to-Asp signal transduction: a potential mechanism for plastid gene regulation in *Heterosigma akashiwo* (Raphidophyceae). *BMC Evol Biol* **7**: 70.
- Duport, C., Zigha, A., Rosenfeld, E., and Schmitt, P. (2006) Control of enterotoxin gene expression in *Bacillus cereus* F4430/73 involves the redox-sensitive ResDE signal transduction system. *J Bacteriol* **188**: 6640-6651.
- Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998) *Biological sequence analysis:* probabilistic models of proteins and nucleic acids. Cambridge, United Kingdom: Cambridge University Press.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.
- Ehling-Schulz, M., Fricker, M., and Scherer, S. (2004) *Bacillus cereus*, the causative agent of an emetic type of food-borne illness. *Mol Nutr Food Res* **48**: 479-487.
- Ehling-Schulz, M., Vukov, N., Schulz, A., Shaheen, R., Andersson, M., Märtlbauer, E., and Scherer, S. (2005) Identification and partial characterization of the nonribosomal peptide synthetase gene responsible for cereulide production in emetic *Bacillus cereus*. *Appl Environ Microbiol* **71**: 105-113.
- Ehling-Schulz, M., Fricker, M., Grallert, H., Rieck, P., Wagner, M., and Scherer, S. (2006) Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. *BMC Microbiol* **6**: 20.
- Engelke, G., Gutowski-Eckel, Z., Kiesau, P., Siegers, K., Hammelmann, M., and Entian, K.D. (1994) Regulation of nisin biosynthesis and immunity in *Lactococcus lactis* 6F3. *Appl Environ Microbiol* **60**: 814-825.
- Esbelin, J., Armengaud, J., Zigha, A., and Duport, C. (2009) ResDE-dependent regulation of enterotoxin gene expression in *Bacillus cereus*: evidences for multiple modes of binding for ResD and interaction with Fnr. *J Bacteriol*.
- Evers, S., and Courvalin, P. (1996) Regulation of VanB-type vancomycin resistance gene expression by the VanS(B)-VanR (B) two-component regulatory system in *Enterococcus faecalis* V583. *J Bacteriol* **178**: 1302-1309.
- Fabret, C., and Hoch, J.A. (1998) A two-component signal transduction system essential for growth of *Bacillus subtilis*: implications for anti-infective therapy. *J Bacteriol* **180**: 6375-6383.
- Fabret, C., Feher, V.A., and Hoch, J.A. (1999) Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. *J Bacteriol* **181**: 1975-1983.
- Fagerlund, A., Brillard, J., Fürst, R., Guinebretière, M.H., and Granum, P.E. (2007) Toxin production in a rare and genetically remote cluster of strains of the *Bacillus cereus* group. *BMC Microbiol* **7**: 43.

- Fagerlund, A., Ween, O., Lund, T., Hardy, S.P., and Granum, P.E. (2004) Genetic and functional analysis of the *cytK* family of genes in *Bacillus cereus*. *Microbiology* **150**: 2689-2697.
- Fagerlund, A., Lindbäck, T., Storset, A.K., Granum, P.E., and Hardy, S.P. (2008) *Bacillus cereus* Nhe is a pore-forming toxin with structural and functional properties similar to the ClyA (HIyE, SheA) family of haemolysins, able to induce osmotic lysis in epithelia. *Microbiology* **154**: 693-704.
- Fedhila, S., Gohar, M., Slamti, L., Nel, P., and Lereclus, D. (2003) The *Bacillus thuringiensis* PlcR-regulated gene *inhA2* is necessary, but not sufficient, for virulence. *J Bacteriol* **185**: 2820-2825.
- Fouet, A., Namy, O., and Lambert, G. (2000) Characterization of the operon encoding the alternative σ^{B} factor from *Bacillus anthracis* and its role in virulence. *J Bacteriol* **182**: 5036-5045.
- 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.
- Frankland, G.C., and Frankland, P.R. (1887) Studies on some new micro-organisms obtained from air. *Philos Trans R Soc Lond B Biol Sci* **173**: 257-287.
- Fricker, M., Messelhäusser, U., Busch, U., Scherer, S., and Ehling-Schulz, M. (2007) Diagnostic real-time PCR assays for the detection of emetic *Bacillus cereus* strains in foods and recent food-borne outbreaks. *Appl Environ Microbiol* **73**: 1892-1898.
- From, C., Pukall, R., Schumann, P., Hormazábal, V., and Granum, P.E. (2005) Toxinproducing ability among *Bacillus* spp. outside the *Bacillus cereus* group. *Appl Environ Microbiol* **71**: 1178-1183.
- Gaidenko, T.A., Kim, T.J., Weigel, A.L., Brody, M.S., and Price, C.W. (2006) The blue-light receptor YtvA acts in the environmental stress signaling pathway of *Bacillus subtilis*. *J Bacteriol* **188**: 6387-6395.
- Galperin, M.Y. (2004) Bacterial signal transduction network in a genomic perspective. *Environ Microbiol* **6**: 552-567.
- 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.
- Geng, H., Zhu, Y., Mullen, K., Zuber, C.S., and Nakano, M.M. (2007) Characterization of ResDE-dependent fnr transcription in *Bacillus subtilis*. J Bacteriol **189**: 1745-1755.
- Gohar, M., Faegri, K., Perchat, S., Ravnum, S., Økstad, O.A., Gominet, M., Kolstø, A.B., and Lereclus, D. (2008) The PIcR virulence regulon of *Bacillus cereus*. *PLoS ONE* **3**: e2793.
- Granum, P.E., O'Sullivan, K., and Lund, T. (1999) The sequence of the non-haemolytic enterotoxin operon from *Bacillus cereus*. *FEMS Microbiol Lett* **177**: 225-229.
- Granum, P.E. (2007) *Bacillus cereus*. In *Food Microbiology: Fundamentals and Frontiers*. Doyle, M.P. and Beuchat, L.R. (eds): ASM press, Washington, D.C., pp. 445-455.
- Grebe, T.W., and Stock, J.B. (1999) The histidine protein kinase superfamily. *Adv Microb Physiol* **41**: 139-227.
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33**: D121-124.
- Gruber, T.M., and Gross, C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**: 441-466.
- Guenzi, E., Gasc, A.M., Sicard, M.A., and Hakenbeck, R. (1994) A two-component signaltransducing system is involved in competence and penicillin susceptibility in laboratory mutants of *Streptococcus pneumoniae*. *Mol Microbiol* **12**: 505-515.
- Guinebretière, M.H., Broussolle, V., and Nguyen-The, C. (2002) Enterotoxigenic profiles of food-poisoning and food-borne *Bacillus cereus* strains. *J Clin Microbiol* **40**: 3053-3056.
- Haima, P., Bron, S., and Venema, G. (1987) The effect of restriction on shotgun cloning and plasmid stability in *Bacillus subtilis* Marburg. *Mol Gen Genet* **209**: 335-342.
- Han, C.S., Xie, G., Challacombe, J.F., Altherr, M.R., Bhotika, S.S., Bruce, D., Campbell, C.S., Campbell, M.L., Chen, J., Chertkov, O., Cleland, C., Dimitrijevic, M., Doggett, N.A., Fawcett, J.J., Glavina, T., Goodwin, L.A., Hill, K.K., Hitchcock, P., Jackson, P.J., Keim, P., Kewalramani, A.R., Longmire, J., Lucas, S., Malfatti, S., McMurry, K., Meincke, L.J., Misra, M., Moseman, B.L., Mundt, M., Munk, A.C., Okinaka, R.T., Parson-Quintana, B., Reilly, L.P., Richardson, P., Robinson, D.L., Rubin, E., Saunders, E., Tapia, R., Tesmer, J.G., Thayer, N., Thompson, L.S., Tice, H., Ticknor, L.O., Wills, P.L., Brettin, T.S., and Gilna, P. (2006) Pathogenomic Sequence Analysis of *Bacillus cereus* and *Bacillus thuringiensis* Isolates Closely Related to *Bacillus anthracis. J. Bacteriol.* 188: 3382-3390.
- Hardy, S.P., Lund, T., and Granum, P.E. (2001) CytK toxin of *Bacillus cereus* forms pores in planar lipid bilayers and is cytotoxic to intestinal epithelia. *FEMS Microbiol Lett* **197**: 47-51.
- Hauge, S. (1950) Food poisoning caused by *Bacillus cereus*. *Nord Hyg Tidsskr* **31**: 189-206.
- Hauge, S. (1955) Food poisoning caused by aerobic spore forming bacilli. J Appl Bacteriol 18: 591-595.
- Heinrichs, J.H., Beecher, D.J., MacMillan, J.D., and Zilinskas, B.A. (1993) Molecular cloning and characterization of the *hblA* gene encoding the B component of hemolysin BL from *Bacillus cereus*. J Bacteriol **175**: 6760-6766.
- Helgason, E., Økstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I., and Kolstø (2000) Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis--one species on the basis of genetic evidence. Appl Environ Microbiol 66: 2627-2630.
- Helgason, E., Tourasse, N.J., Meisal, R., Caugant, D.A., and Kolstø, A.B. (2004) Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. *Appl Environ Microbiol* **70**: 191-201.
- 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.
- Heyl, A., Wulfetange, K., Pils, B., Nielsen, N., Romanov, G.A., and Schmülling, T. (2007) Evolutionary proteomics identifies amino acids essential for ligand-binding of the cytokinin receptor CHASE domain. *BMC Evol Biol* **7**: 62.
- Hill, K.K., Ticknor, L.O., Okinaka, R.T., Asay, M., Blair, H., Bliss, K.A., Laker, M., Pardington, P.E., Richardson, A.P., Tonks, M., Beecher, D.J., Kemp, J.D., Kolstø, A.B., Wong, A.C., Keim, P., and Jackson, P.J. (2004) Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. *Appl Environ Microbiol* **70**: 1068-1080.
- Hoch, J.A. (2000) Two-component and phosphorelay signal transduction. *Curr Opin Microbiol* **3**: 165-170.
- Ho, Y.S., Burden, L.M., and Hurley, J.H. (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *Embo J* **19**: 5288-5299.
- Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., Maiden, M.C., Priest, F.G., Barker, M., Jiang, L., Cer, R.Z., Rilstone, J., Peterson, S.N., Weyant, R.S., Galloway, D.R., Read, T.D., Popovic, T., and Fraser, C.M. (2004) Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc Natl Acad Sci U S A* **101**: 8449-8454.
- Hoton, F.M., Andrup, L., Swiecicka, I., and Mahillon, J. (2005) The cereulide genetic determinants of emetic *Bacillus cereus* are plasmid-borne. *Microbiology* **151**: 2121-2124.
- Howell, A., Dubrac, S., Andersen, K.K., Noone, D., Fert, J., Msadek, T., and Devine, K. (2003) Genes controlled by the essential YycG/YycF two-component system of *Bacillus*

subtilis revealed through a novel hybrid regulator approach. *Mol Microbiol* **49**: 1639-1655.

- Hsiao, A., Ideker, T., Olefsky, J.M., and Subramaniam, S. (2005) VAMPIRE microarray suite: a web-based platform for the interpretation of gene expression data. *Nucleic Acids Res* **33**: W627-632.
- Huang, Y., Morel, P., Powell, B., and Kado, C.I. (1990) VirA, a coregulator of Ti-specified virulence genes, is phosphorylated *in vitro*. *J Bacteriol* **172**: 1142-1144.
- Hulko, M., Berndt, F., Gruber, M., Linder, J.U., Truffault, V., Schultz, A., Martin, J., Schultz, J.E., Lupas, A.N., and Coles, M. (2006) The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell* **126**: 929-940.
- Igoshin, O.A., Alves, R., and Savageau, M.A. (2008) Hysteretic and graded responses in bacterial two-component signal transduction. *Mol Microbiol* **68**: 1196-1215.
- Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., Chu, L., Mazur, M., Goltsman, E., Larsen, N., D'Souza, M., Walunas, T., Grechkin, Y., Pusch, G., Haselkorn, R., Fonstein, M., Ehrlich, S.D., Overbeek, R., and Kyrpides, N. (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **423**: 87-91.
- Janausch, I.G., Garcia-Moreno, I., and Unden, G. (2002a) Function of DcuS from *Escherichia coli* as a fumarate-stimulated histidine protein kinase *in vitro*. *J Biol Chem* **277**: 39809-39814.
- Janausch, I.G., Zientz, E., Tran, Q.H., Kröger, A., and Unden, G. (2002b) C4-dicarboxylate carriers and sensors in bacteria. *Biochim Biophys Acta* **1553**: 39-56.
- Jensen, G.B., Hansen, B.M., Eilenberg, J., and Mahillon, J. (2003) The hidden lifestyles of *Bacillus cereus* and relatives. *Environ Microbiol* **5**: 631-640.
- Jernigan, D.B., Raghunathan, P.L., Bell, B.P., Brechner, R., Bresnitz, E.A., Butler, J.C., Cetron, M., Cohen, M., Doyle, T., Fischer, M., Greene, C., Griffith, K.S., Guarner, J., Hadler, J.L., Hayslett, J.A., Meyer, R., Petersen, L.R., Phillips, M., Pinner, R., Popovic, T., Quinn, C.P., Reefhuis, J., Reissman, D., Rosenstein, N., Schuchat, A., Shieh, W.J., Siegal, L., Swerdlow, D.L., Tenover, F.C., Traeger, M., Ward, J.W., Weisfuse, I., Wiersma, S., Yeskey, K., Zaki, S., Ashford, D.A., Perkins, B.A., Ostroff, S., Hughes, J., Fleming, D., Koplan, J.P., and Gerberding, J.L. (2002) Investigation of bioterrorismrelated anthrax, United States, 2001: epidemiologic findings. *Emerg Infect Dis* 8: 1019-1028.
- Jiang, M., Shao, W., Perego, M., and Hoch, J.A. (2000) Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Mol Microbiol* **38**: 535-542.
- Jin, S., Roitsch, T., Ankenbauer, R.G., Gordon, M.P., and Nester, E.W. (1990a) The VirA protein of *Agrobacterium tumefaciens* is autophosphorylated and is essential for *vir* gene regulation. *J Bacteriol* **172**: 525-530.
- Jin, S.G., Prusti, R.K., Roitsch, T., Ankenbauer, R.G., and Nester, E.W. (1990b) Phosphorylation of the VirG protein of *Agrobacterium tumefaciens* by the autophosphorylated VirA protein: essential role in biological activity of VirG. *J Bacteriol* **172**: 4945-4950.
- Jin, T., and Inouye, M. (1994) Transmembrane signaling. Mutational analysis of the cytoplasmic linker region of Taz1-1, a Tar-EnvZ chimeric receptor in *Escherichia coli*. *J Mol Biol* **244**: 477-481.
- Johnson, K.M. (1984) *Bacillus cereus* food-borne illness. An update. *J Food Prot* **47**: 145-153.
- Jordan, S., Hutchings, M.I., and Mascher, T. (2008) Cell envelope stress response in Gram-positive bacteria. *FEMS Microbiol Rev* **32**: 107-146.
- Jordan, S., Junker, A., Helmann, J.D., and Mascher, T. (2006) Regulation of LiaRSdependent 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.

- Joseph, P., Fichant, G., Quentin, Y., and Denizot, F. (2002) Regulatory relationship of twocomponent 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.
- Jung, K., and Altendorf, K. (2002) Towards an understanding of the molecular mechanisms of stimulus perception and signal transduction by the KdpD/KdpE system of *Escherichia coli*. J Mol Microbiol Biotechnol **4**: 223-228.
- Kallipolitis, B.H., Ingmer, H., Gahan, C.G., Hill, C., and Søgaard-Andersen, L. (2003) CesRK, a two-component signal transduction system in *Listeria monocytogenes*, responds to the presence of cell wall-acting antibiotics and affects beta-lactam resistance. *Antimicrob Agents Chemother* **47**: 3421-3429.
- Käll, L., Krogh, A., and Sonnhammer, E.L. (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338**: 1027-1036.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**: D277-D280.
- Kaspar, S., and Bott, M. (2002) The sensor kinase CitA (DpiB) of *Escherichia coli* functions as a high-affinity citrate receptor. *Arch Microbiol* **177**: 313-321.
- Kazmierczak, M.J., Wiedmann, M., and Boor, K.J. (2005) Alternative sigma factors and their roles in bacterial virulence. *Microbiol Mol Biol Rev* **69**: 527-543.
- Kehry, M.R., Bond, M.W., Hunkapiller, M.W., and Dahlquist, F.W. (1983) Enzymatic deamidation of methyl-accepting chemotaxis proteins in *Escherichia coli* catalyzed by the *cheB* gene product. *Proc Natl Acad Sci U S A* **80**: 3599-3603.
- Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Jackson, P.J., and Hugh-Jones, M.E. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J Bacteriol **182**: 2928-2936.
- Kerkhoven, R., van Enckevort, F.H., Boekhorst, J., Molenaar, D., and Siezen, R.J. (2004) Visualization for genomics: the Microbial Genome Viewer. *Bioinformatics* 20: 1812-1814.
- Kern, D., Volkman, B.F., Luginbühl, P., Nohaile, M.J., Kustu, S., and Wemmer, D.E. (1999) Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature* **402**: 894-898.
- Kim, T.J., Gaidenko, T.A., and Price, C.W. (2004) A multicomponent protein complex mediates environmental stress signaling in *Bacillus subtilis*. J Mol Biol **341**: 135-150.
- Klein, C., Kaletta, C., and Entian, K.D. (1993) Biosynthesis of the lantibiotic subtilin is regulated by a histidine kinase/response regulator system. *Appl Environ Microbiol* 59: 296-303.
- Kobayashi, K., Ogura, M., Yamaguchi, H., Yoshida, K., Ogasawara, N., Tanaka, T., and Fujita, Y. (2001) Comprehensive DNA microarray analysis of *Bacillus subtilis* twocomponent regulatory systems. *J Bacteriol* **183**: 7365-7370.
- Koenig, R.L., Ray, J.L., Maleki, S.J., Smeltzer, M.S., and Hurlburt, B.K. (2004) Staphylococcus aureus AgrA binding to the RNAIII-agr regulatory region. J Bacteriol 186: 7549-7555.
- Ko, K.S., Kim, J.W., Kim, J.M., Kim, W., Chung, S.I., Kim, I.J., and Kook, Y.H. (2004) Population structure of the *Bacillus cereus* group as determined by sequence analysis of six housekeeping genes and the *plcR* Gene. *Infect Immun* **72**: 5253-5261.
- Koretke, K.K., Lupas, A.N., Warren, P.V., Rosenberg, M., and Brown, J.R. (2000) Evolution of two-component signal transduction. *Mol Biol Evol* **17**: 1956-1970.

- Körner, H., Sofia, H.J., and Zumft, W.G. (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev* 27: 559-592.
- Kotiranta, A., Lounatmaa, K., and Haapasalo, M. (2000) Epidemiology and pathogenesis of *Bacillus cereus* infections. *Microbes Infect* **2**: 189-198.
- Kreft, J., Berger, H., Härtlein, M., Müller, B., Weidinger, G., and Goebel, W. (1983) Cloning and expression in *Escherichia coli* and *Bacillus subtilis* of the hemolysin (cereolysin) determinant from *Bacillus cereus*. J Bacteriol **155**: 681-689.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305**: 567-580.
- Kuppe, A., Evans, L.M., McMillen, D.A., and Griffith, O.H. (1989) Phosphatidylinositolspecific phospholipase C of *Bacillus cereus*: cloning, sequencing, and relationship to other phospholipases. *J Bacteriol* **171**: 6077-6083.
- Kuroda, M., Kuwahara-Arai, K., and Hiramatsu, K. (2000) Identification of the up- and down-regulated genes in vancomycin-resistant *Staphylococcus aureus* strains Mu3 and Mu50 by cDNA differential hybridization method. *Biochem Biophys Res Commun* 269: 485-490.
- Lapidus, A., Goltsman, E., Auger, S., Galleron, N., Ségurens, B., Dossat, C., Land, M.L., Broussolle, V., Brillard, J., Guinebretière, M.H., Sanchis, V., Nguen-The, C., Lereclus, D., Richardson, P., Wincker, P., Weissenbach, J., Ehrlich, S.D., and Sorokin, A. (2008) Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity. *Chem Biol Interact* **171**: 236-249.
- Lechner, S., Mayr, R., Francis, K.P., Prüss, B.M., Kaplan, T., Wiessner-Gunkel, E., Stewart, G.S., and Scherer, S. (1998) *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int J Syst Bacteriol* **48 Pt 4**: 1373-1382.
- Lee, E.J., Cho, Y.H., Kim, H.S., Ahn, B.E., and Roe, J.H. (2004) Regulation of o^B by an anti- and an anti-anti-sigma factor in *Streptomyces coelicolor* in response to osmotic stress. *J Bacteriol* **186**: 8490-8498.
- Lee, E.J., Karoonuthaisiri, N., Kim, H.S., Park, J.H., Cha, C.J., Kao, C.M., and Roe, J.H. (2005) A master regulator σ^B governs osmotic and oxidative response as well as differentiation via a network of sigma factors in *Streptomyces coelicolor*. *Mol Microbiol* **57**: 1252-1264.
- Le Moual, H., and Koshland, D.E., Jr. (1996) Molecular evolution of the C-terminal cytoplasmic domain of a superfamily of bacterial receptors involved in taxis. *J Mol Biol* **261**: 568-585.
- Lerondel, G., Doan, T., Zamboni, N., Sauer, U., and Aymerich, S. (2006) YtsJ has the major physiological role of the four paralogous malic enzyme isoforms in *Bacillus subtilis*. J Bacteriol **188**: 4727-4736.
- Letunic, I., Doerks, T., and Bork, P. (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res* **37**: D229-232.
- Lewis, R.J., Krzywda, S., Brannigan, J.A., Turkenburg, J.P., Muchová, K., Dodson, E.J., Barák, I., and Wilkinson, A.J. (2000) The trans-activation domain of the sporulation response regulator Spo0A revealed by X-ray crystallography. *Mol Microbiol* **38**: 198-212.
- Liang, X., Zheng, L., Landwehr, C., Lunsford, D., Holmes, D., and Ji, Y. (2005) Global regulation of gene expression by ArIRS, a two-component signal transduction regulatory system of *Staphylococcus aureus*. *J Bacteriol* **187**: 5486-5492.
- Li, M., Lai, Y., Villaruz, A.E., Cha, D.J., Sturdevant, D.E., and Otto, M. (2007) Grampositive three-component antimicrobial peptide-sensing system. *Proc Natl Acad Sci U S A* **104**: 9469-9474.
- Lindbäck, T., Økstad, O.A., Rishovd, A.L., and Kolstø, A.B. (1999) Insertional inactivation of *hblC* encoding the L2 component of *Bacillus cereus* ATCC 14579 haemolysin BL

strongly reduces enterotoxigenic activity, but not the haemolytic activity against human erythrocytes. *Microbiology* **145 (Pt 11)**: 3139-3146.

- Lindbäck, T., Fagerlund, A., Rødland, M.S., and Granum, P.E. (2004) Characterization of the *Bacillus cereus* Nhe enterotoxin. *Microbiology* **150**: 3959-3967.
- Liu, C.Q., Charoechai, P., Khunajakr, N., Deng, Y.M., Widodo, and Dunn, N.W. (2002) Genetic and transcriptional analysis of a novel plasmid-encoded copper resistance operon from *Lactococcus lactis. Gene* 297: 241-247.
- Liu, J., Tan, K., and Stormo, G.D. (2003) Computational identification of the Spo0Aphosphate regulon that is essential for the cellular differentiation and development in Gram-positive spore-forming bacteria. *Nucleic Acids Res* **31**: 6891-6903.
- Liu, W., Eder, S., and Hulett, F.M. (1998a) Analysis of *Bacillus subtilis tagAB* and *tagDEF* expression during phosphate starvation identifies a repressor role for PhoP-P. J *Bacteriol* **180**: 753-758.
- Liu, W., Qi, Y., and Hulett, F.M. (1998b) Sites internal to the coding regions of *phoA* and *pstS* bind PhoP and are required for full promoter activity. *Mol Microbiol* **28**: 119-130.
- López, D., Vlamakis, H., and Kolter, R. (2009) Generation of multiple cell types in *Bacillus* subtilis. FEMS Microbiol Rev 33: 152-163.
- Lund, T., and Granum, P.E. (1997) Comparison of biological effect of the two different enterotoxin complexes isolated from three different strains of *Bacillus cereus*. *Microbiology* **143 (Pt 10)**: 3329-3336.
- Lund, T., De Buyser, M.L., and Granum, P.E. (2000) A new cytotoxin from *Bacillus cereus* that may cause necrotic enteritis. *Mol Microbiol* **38**: 254-261.
- Luxananil, P., Butrapet, S., Atomi, H., Imanaka, T., and Panyim, S. (2003) A decrease in cytotoxic and haemolytic activities by inactivation of a single enterotoxin gene in *Bacillus cereus* CX5. *World J. Microbiol. Biotechnol.* **19**: 831-837.
- Lyon, G.J., and Novick, R.P. (2004) Peptide signaling in *Staphylococcus aureus* and other Gram-positive bacteria. *Peptides* **25**: 1389-1403.
- Lyon, G.J., Mayville, P., Muir, T.W., and Novick, R.P. (2000) Rational design of a global inhibitor of the virulence response in *Staphylococcus aureus*, based in part on localization of the site of inhibition to the receptor-histidine kinase, AgrC. *Proc Natl Acad Sci U S A* **97**: 13330-13335.
- Mahler, H., Pasi, A., Kramer, J.M., Schulte, P., Scoging, A.C., Bär, W., and Krähenbühl, S. (1997) Fulminant liver failure in association with the emetic toxin of *Bacillus cereus*. N Engl J Med **336**: 1142-1148.
- Makita, Y., Nakao, M., Ogasawara, N., and Nakai, K. (2004) DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. *Nucleic Acids Res* **32**: D75-77.
- Malvar, T., Gawron-Burke, C., and Baum, J.A. (1994) Overexpression of *Bacillus thuringiensis* HknA, a histidine protein kinase homology, bypasses early Spo mutations that result in CryIIIA overproduction. *J Bacteriol* **176**: 4742-4749.
- 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.
- Margulis, L., Jorgensen, J.Z., Dolan, S., Kolchinsky, R., Rainey, F.A., and Lo, S.C. (1998) The Arthromitus stage of *Bacillus cereus*: intestinal symbionts of animals. *Proc Natl Acad Sci U S A* **95**: 1236-1241.
- Maris, A.E., Sawaya, M.R., Kaczor-Grzeskowiak, M., Jarvis, M.R., Bearson, S.M., Kopka, M.L., Schröder, I., Gunsalus, R.P., and Dickerson, R.E. (2002) Dimerization allows DNA target site recognition by the NarL response regulator. *Nat Struct Biol* **9**: 771-778.
- Marles-Wright, J., Grant, T., Delumeau, O., van Duinen, G., Firbank, S.J., Lewis, P.J., Murray, J.W., Newman, J.A., Quin, M.B., Race, P.R., Rohou, A., Tichelaar, W., van Heel, M., and Lewis, R.J. (2008) Molecular architecture of the "stressosome," a signal integration and transduction hub. *Science* **322**: 92-96.

- Marraffini, L.A., and Schneewind, O. (2006) Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. *Mol Microbiol* **62**: 1402-1417.
- Martínez-Hackert, E., and Stock, A.M. (1997) Structural relationships in the OmpR family of winged-helix transcription factors. *J Mol Biol* **269**: 301-312.
- Martínez, L.F., Bishop, A., Parkes, L., Del Sol, R., Salerno, P., Sevcikova, B., Mazurakova, V., Kormanec, J., and Dyson, P. (2009) Osmoregulation in *Streptomyces coelicolor*: modulation of SigB activity by OsaC. *Mol Microbiol*.
- Martinez, S.E., Beavo, J.A., and Hol, W.G. (2002) GAF domains: two-billion-year-old molecular switches that bind cyclic nucleotides. *Mol Interv* **2**: 317-323.
- Martin, P.K., Li, T., Sun, D., Biek, D.P., and Schmid, M.B. (1999) Role in cell permeability of an essential two-component system in *Staphylococcus aureus*. J Bacteriol 181: 3666-3673.
- Mascher, T., Margulis, N.G., Wang, T., Ye, R.W., and Helmann, J.D. (2003a) Cell wall stress responses in *Bacillus subtilis*: the regulatory network of the bacitracin stimulon. *Mol Microbiol* **50**: 1591-1604.
- Mascher, T., Zähner, D., Merai, M., Balmelle, N., de Saizieu, A.B., and Hakenbeck, R. (2003b) The Streptococcus pneumoniae cia regulon: CiaR target sites and transcription profile analysis. *J Bacteriol* **185**: 60-70.
- 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.
- Mattoo, A.R., Saif Zaman, M., Dubey, G.P., Arora, A., Narayan, A., Jailkhani, N., Rathore, K., Maiti, S., and Singh, Y. (2008) Spo0B of *Bacillus anthracis* - a protein with pleiotropic functions. *Febs J* **275**: 739-752.
- McLachlan, A.D., and Stewart, M. (1975) Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. J Mol Biol **98**: 293-304.
- Mead, P.S., Slutsker, L., Dietz, V., McCaig, L.F., Bresee, J.S., Shapiro, C., Griffin, P.M., and Tauxe, R.V. (1999) Food-related illness and death in the United States. *Emerg Infect Dis* 5: 607-625.
- Meehl, M., Herbert, S., Götz, F., and Cheung, A. (2007) Interaction of the GraRS twocomponent system with the VraFG ABC transporter to support vancomycinintermediate resistance in *Staphylococcus aureus*. *Antimicrob Agents Chemother* **51**: 2679-2689.
- Mikkola, R., Saris, N.E., Grigoriev, P.A., Andersson, M.A., and Salkinoja-Salonen, M.S. (1999) Ionophoretic properties and mitochondrial effects of cereulide: the emetic toxin of *B. cereus. Eur J Biochem* **263**: 112-117.
- Mittenhuber, G. (2002) A phylogenomic study of the general stress response sigma factor o^B of *Bacillus subtilis* and its regulatory proteins. *J Mol Microbiol Biotechnol* **4**: 427-452.
- Mizuno, T. (2005) Two-component phosphorelay signal transduction systems in plants: from hormone responses to circadian rhythms. *Biosci Biotechnol Biochem* **69**: 2263-2276.

Mock, M., and Fouet, A. (2001) Anthrax. Annu Rev Microbiol 55: 647-671.

- Mohedano, M.L., Overweg, K., de la Fuente, A., Reuter, M., Altabe, S., Mulholland, F., de Mendoza, D., López, P., and Wells, J.M. (2005) Evidence that the essential response regulator YycF in *Streptococcus pneumoniae* modulates expression of fatty acid biosynthesis genes and alters membrane composition. *J Bacteriol* **187**: 2357-2367.
- Molle, V., Fujita, M., Jensen, S.T., Eichenberger, P., González-Pastor, J.E., Liu, J.S., and Losick, R. (2003) The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol* **50**: 1683-1701.
- Mols, M., de Been, M., Zwietering, M.H., Moezelaar, R., and Abee, T. (2007) Metabolic capacity of *Bacillus cereus* strains ATCC 14579 and ATCC 10987 interlinked with comparative genomics. *Environ Microbiol* **9**: 2933-2944.
- Mortimer, P.R., and McCann, G. (1974) Food-poisoning episodes associated with *Bacillus cereus* in fried rice. *Lancet* **1**: 1043-1045.

- Mougel, C., and Zhulin, I.B. (2001) CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. *Trends Biochem Sci* 26: 582-584.
- Mwangi, M.M., and Siggia, E.D. (2003) Genome wide identification of regulatory motifs in *Bacillus subtilis. BMC Bioinformatics* **4**: 18.
- Nakano, M.M., Zuber, P., Glaser, P., Danchin, A., and Hulett, F.M. (1996) Two-component regulatory proteins ResD-ResE are required for transcriptional activation of fnr upon oxygen limitation in *Bacillus subtilis*. *J Bacteriol* **178**: 3796-3802.
- Neumüller, A.M., Konz, D., and Marahiel, M.A. (2001) The two-component regulatory system BacRS is associated with bacitracin 'self-resistance' of *Bacillus licheniformis* ATCC 10716. *Eur J Biochem* **268**: 3180-3189.
- Novick, R.P. (2003) Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol Microbiol* **48**: 1429-1449.
- Nowlin, D.M., Bollinger, J., and Hazelbauer, G.L. (1987) Sites of covalent modification in Trg, a sensory transducer of *Escherichia coli*. *J Biol Chem* **262**: 6039-6045.
- Ohki, R., Giyanto, Tateno, K., Masuyama, W., Moriya, S., Kobayashi, K., and Ogasawara, N. (2003) The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in *Bacillus subtilis. Mol Microbiol* **49**: 1135-1144.
- Okinaka, R., Cloud, K., Hampton, O., Hoffmaster, A., Hill, K., Keim, P., Koehler, T., Lamke, G., Kumano, S., Manter, D., Martinez, Y., Ricke, D., Svensson, R., and Jackson, P. (1999a) Sequence, assembly and analysis of pX01 and pX02. *J Appl Microbiol* 87: 261-262.
- Okinaka, R.T., Cloud, K., Hampton, O., Hoffmaster, A.R., Hill, K.K., Keim, P., Koehler, T.M., Lamke, G., Kumano, S., Mahillon, J., Manter, D., Martinez, Y., Ricke, D., Svensson, R., and Jackson, P.J. (1999b) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J Bacteriol* **181**: 6509-6515.
- Østensvik, Ø., From, C., Heidenreich, B., O'Sullivan, K., and Granum, P.E. (2004) Cytotoxic *Bacillus* spp. belonging to the *B. cereus* and *B. subtilis* groups in Norwegian surface waters. *J Appl Microbiol* **96**: 987-993.
- Otto, M. (2004) Quorum-sensing control in Staphylococci -- a target for antimicrobial drug therapy? *FEMS Microbiol Lett* **241**: 135-141.
- Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D., and Maltsev, N. (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A* **96**: 2896-2901.
- Pager, M.S., and Helmann, J.D. (2003) The σ^{70} family of sigma factors. Genome Biol 4: 203.
- Pané-Farré, J., Lewis, R.J., and Stülke, J. (2005) The RsbRST stress module in bacteria: a signalling system that may interact with different output modules. J Mol Microbiol Biotechnol 9: 65-76.
- Pan, T.M., Wang, T.K., Lee, C.L., Chien, S.W., and Horng, C.B. (1997) Food-borne disease outbreaks due to bacteria in Taiwan, 1986 to 1995. *J Clin Microbiol* **35**: 1260-1262.
- Parkinson, J.S., and Kofoid, E.C. (1992) Communication modules in bacterial signaling proteins. Annu Rev Genet 26: 71-112.
- Pas, J., von Grotthuss, M., Wyrwicz, L.S., Rychlewski, L., and Barciszewski, J. (2004) Structure prediction, evolution and ligand interaction of CHASE domain. *FEBS Lett* 576: 287-290.
- Pelton, J.G., Kustu, S., and Wemmer, D.E. (1999) Solution structure of the DNA-binding domain of NtrC with three alanine substitutions. *J Mol Biol* **292**: 1095-1110.
- Perez, E., Zheng, H., and Stock, A.M. (2006) Identification of methylation sites in *Thermotoga maritima* chemotaxis receptors. *J Bacteriol* **188**: 4093-4100.
- Periago, P.M., van Schaik, W., Abee, T., and Wouters, J.A. (2002) Identification of proteins involved in the heat stress response of *Bacillus cereus* ATCC 14579. *Appl Environ Microbiol* 68: 3486-3495.

- Pietiäinen, M., Gardemeister, M., Mecklin, M., Leskelä, 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.
- Podlesek, Z., Comino, A., Herzog-Velikonja, B., Zgur-Bertok, D., Komel, R., and Grabnar, M. (1995) *Bacillus licheniformis* bacitracin-resistance ABC transporter: relationship to mammalian multidrug resistance. *Mol Microbiol* **16**: 969-976.
- Ponting, C.P., and Aravind, L. (1997) PAS: a multifunctional domain family comes to light. *Curr Biol* **7**: R674-677.
- Posas, F., Wurgler-Murphy, S.M., Maeda, T., Witten, E.A., Thai, T.C., and Saito, H. (1996) Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor. *Cell* **86**: 865-875.
- Price, C.W. (2002) General stress response. In *Bacillus subtilis and its closest relatives: from genes to cells*. Sonenshein, A.L., Hoch, J.A. and Losick, R. (eds): ASM Press, Washington, DC, pp. 369-384.
- Priest, F.G., Barker, M., Baillie, L.W., Holmes, E.C., and Maiden, M.C. (2004) Population structure and evolution of the *Bacillus cereus* group. *J Bacteriol* **186**: 7959-7970.
- Puthiyaveetil, S., Kavanagh, T.A., Cain, P., Sullivan, J.A., Newell, C.A., Gray, J.C., Robinson, C., van der Giezen, M., Rogers, M.B., and Allen, J.F. (2008) The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts. *Proc Natl Acad Sci U S A* **105**: 10061-10066.
- Puthiyaveetil, S., and Allen, J.F. (2009) Chloroplast two-component systems: evolution of the link between photosynthesis and gene expression. *Proc Biol Sci.*
- Qiu, X.Q., Wang, H., Lu, X.F., Zhang, J., Li, S.F., Cheng, G., Wan, L., Yang, L., Zuo, J.Y., Zhou, Y.Q., Wang, H.Y., Cheng, X., Zhang, S.H., Ou, Z.R., Zhong, Z.C., Cheng, J.Q., Li, Y.P., and Wu, G.Y. (2003) An engineered multidomain bactericidal peptide as a model for targeted antibiotics against specific bacteria. *Nat Biotechnol* **21**: 1480-1485.
- Rasko, D.A., Ravel, J., Økstad, O.A., Helgason, E., Cer, R.Z., Jiang, L., Shores, K.A., Fouts, D.E., Tourasse, N.J., Angiuoli, S.V., Kolonay, J., Nelson, W.C., Kolstø, A.B., Fraser, C.M., and Read, T.D. (2004) The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. *Nucleic Acids Res* **32**: 977-988.
- Rasko, D.A., Altherr, M.R., Han, C.S., and Ravel, J. (2005) Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol Rev* **29**: 303-329.
- Read, T.D., Salzberg, S.L., Pop, M., Shumway, M., Umayam, L., Jiang, L., Holtzapple, E., Busch, J.D., Smith, K.L., Schupp, J.M., Solomon, D., Keim, P., and Fraser, C.M. (2002) Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis. Science* **296**: 2028-2033.
- Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R., Holtzapple, E.K., Økstad, O.A., Helgason, E., Rilstone, J., Wu, M., Kolonay, J.F., Beanan, M.J., Dodson, R.J., Brinkac, L.M., Gwinn, M., DeBoy, R.T., Madpu, R., Daugherty, S.C., Durkin, A.S., Haft, D.H., Nelson, W.C., Peterson, J.D., Pop, M., Khouri, H.M., Radune, D., Benton, J.L., Mahamoud, Y., Jiang, L., Hance, I.R., Weidman, J.F., Berry, K.J., Plaut, R.D., Wolf, A.M., Watkins, K.L., Nierman, W.C., Hazen, A., Cline, R., Redmond, C., Thwaite, J.E., White, O., Salzberg, S.L., Thomason, B., Friedlander, A.M., Koehler, T.M., Hanna, P.C., Kolstø, A.B., and Fraser, C.M. (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423**: 81-86.
- Remm, M., Storm, C.E., and Sonnhammer, E.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* **314**: 1041-1052.
- Rhee, S., Martin, R.G., Rosner, J.L., and Davies, D.R. (1998) A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc Natl Acad Sci U S A* **95**: 10413-10418.

Romero, S., Merino, E., Bolívar, F., Gosset, G., and Martinez, A. (2007) Metabolic engineering of *Bacillus subtilis* for ethanol production: lactate dehydrogenase plays a key role in fermentative metabolism. *Appl Environ Microbiol* **73**: 5190-5198.

Ryan, P.A., Macmillan, J.D., and Zilinskas, B.A. (1997) Molecular cloning and characterization of the genes encoding the L1 and L2 components of hemolysin BL from *Bacillus cereus*. *J Bacteriol* **179**: 2551-2556.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Satomura, T., Shimura, D., Asai, K., Sadaie, Y., Hirooka, K., and Fujita, Y. (2005) Enhancement of glutamine utilization in *Bacillus subtilis* through the GlnK-GlnL twocomponent regulatory system. *J Bacteriol* **187**: 4813-4821.

- Scaramozzino, F., White, A., Perego, M., and Hoch, J.A. (2009) A unique GTP-dependent sporulation sensor histidine kinase in *Bacillus anthracis*. *J Bacteriol* **191**: 687-692.
- Schmidt, K., and Gervelmeyer, A. (2003) WHO Surveillance Programmme for Control of Foodborne Infections and Intoxications in Europe: Eighth Report 1999-2000. FAO/WHO Collaborating Centre for Research and Training in Food Hygiene and Zoonoses, Berlin, Germany.
- Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R., and Dean, D.H. (1998) *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol Mol Biol Rev* **62**: 775-806.
- 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.
- Scott, R.J., Lian, L.Y., Muharram, S.H., Cockayne, A., Wood, S.J., Bycroft, B.W., Williams, P., and Chan, W.C. (2003) Side-chain-to-tail thiolactone peptide inhibitors of the staphylococcal quorum-sensing system. *Bioorg Med Chem Lett* **13**: 2449-2453.
- Serizawa, M., Kodama, K., Yamamoto, H., Kobayashi, K., Ogasawara, N., and Sekiguchi, J. (2005) Functional Analysis of the YvrGHb Two-Component System of *Bacillus subtilis*: Identification of the Regulated Genes by DNA Microarray and Northern Blot Analyses. *Biosci Biotechnol Biochem* 69: 2155-2169.
- Serizawa, M., and Sekiguchi, J. (2005) The *Bacillus subtilis* YdfHI two-component system regulates the transcription of *ydfJ*, a member of the RND superfamily. *Microbiology* **151**: 1769-1778.
- Setlow, P. (2000) Resistance of bacterial spores. In *Bacterial stress responses*. Storz, G. and Hengge-Aronis, R. (eds): ASM press, Washington, D.C., pp. 217-230.
- Shi, X., Wegener-Feldbrügge, S., Huntley, S., Hamann, N., Hedderich, R., and Søgaard-Andersen, L. (2008) Bioinformatics and experimental analysis of proteins of twocomponent systems in *Myxococcus xanthus*. J Bacteriol **190**: 613-624.
- Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007) Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res* **35**: 771-788.
- Sidote, D.J., Barbieri, C.M., Wu, T., and Stock, A.M. (2008) Structure of the Staphylococcus aureus AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. Structure 16: 727-735.
- Smith, J.L. (2004) The physiological role of ferritin-like compounds in bacteria. Crit Rev Microbiol 30: 173-185.
- Stauff, D.L., and Skaar, E.P. (2009) *Bacillus anthracis* HssRS signaling to HrtAB regulates heme resistance during infection. *Mol Microbiol*.
- Stauff, D.L., Torres, V.J., and Skaar, E.P. (2007) Signaling and DNA-binding activities of the Staphylococcus aureus HssR-HssS two-component system required for heme sensing. J Biol Chem 282: 26111-26121.
- Stein, T., Heinzmann, S., Kiesau, P., Himmel, B., and Entian, K.D. (2003) The spa-box for transcriptional activation of subtilin biosynthesis and immunity in *Bacillus subtilis*. *Mol Microbiol* **47**: 1627-1636.

- Stenfors Arnesen, L.P., Fagerlund, A., and Granum, P.E. (2008) From soil to gut: *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol Rev* **32**: 579-606.
- Stephenson, K., and Hoch, J.A. (2002) Evolution of signalling in the sporulation phosphorelay. *Mol Microbiol* **46**: 297-304.
- Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochem* **69**: 183-215.
- Stock, J.B., Ninfa, A.J., and Stock, A.M. (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol Rev* **53**: 450-490.
- Strauch, M.A., Trach, K.A., Day, J., and Hoch, J.A. (1992) Spo0A activates and represses its own synthesis by binding at its dual promoters. *Biochimie* **74**: 619-626.
- Sturme, M.H., Francke, C., Siezen, R.J., de Vos, W.M., and Kleerebezem, M. (2007) Making sense of quorum sensing in lactobacilli: a special focus on *Lactobacillus plantarum* WCFS1. *Microbiology* **153**: 3939-3947.
- Sturme, M.H., Nakayama, J., Molenaar, D., Murakami, Y., Kunugi, R., Fujii, T., Vaughan, E.E., Kleerebezem, M., and de Vos, W.M. (2005) An *agr*-like two-component regulatory system in *Lactobacillus plantarum* is involved in production of a novel cyclic peptide and regulation of adherence. *J Bacteriol* **187**: 5224-5235.
- Sun, G., Birkey, S.M., and Hulett, F.M. (1996) Three two-component signal-transduction systems interact for Pho regulation in *Bacillus subtilis*. *Mol Microbiol* **19**: 941-948.
- Szurmant, H., and Ordal, G.W. (2004) Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol Mol Biol Rev* **68**: 301-319.
- Szurmant, H., Nelson, K., Kim, E.J., Perego, M., and Hoch, J.A. (2005) YycH regulates the activity of the essential YycFG two-component system in *Bacillus subtilis*. *J Bacteriol* **187**: 5419-5426.
- Tanaka, K., Kobayashi, K., and Ogasawara, N. (2003) The *Bacillus subtilis* YufLM twocomponent system regulates the expression of the malate transporters MaeN (YufR) and YfIS, and is essential for utilization of malate in minimal medium. *Microbiology* **149**: 2317-2329.
- Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 29: 22-28.
- Tegmark, K., Morfeldt, E., and Arvidson, S. (1998) Regulation of *agr*-dependent virulence genes in *Staphylococcus aureus* by RNAIII from coagulase-negative staphylococci. J *Bacteriol* **180**: 3181-3186.
- Terwilliger, T.C., and Koshland, D.E., Jr. (1984) Sites of methyl esterification and deamination on the aspartate receptor involved in chemotaxis. *J Biol Chem* **259**: 7719-7725.
- 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.
- Throup, J.P., Koretke, K.K., Bryant, A.P., Ingraham, K.A., Chalker, A.F., Ge, Y., Marra, A., Wallis, N.G., Brown, J.R., Holmes, D.J., Rosenberg, M., and Burnham, M.K. (2000) A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. *Mol Microbiol* **35**: 566-576.
- Trach, K.A., and Hoch, J.A. (1993) Multisensory activation of the phosphorelay initiating sporulation in *Bacillus subtilis*: identification and sequence of the protein kinase of the alternate pathway. *Mol Microbiol* **8**: 69-79.
- Trieu-Cuot, P., Carlier, C., Poyart-Salmeron, C., and Courvalin, P. (1990) A pair of mobilisable shuttle vectors conferring resistance to spectinomycin for molecular cloning in *Escherichia coli* and in gram-positive bacteria. *Nucleic Acids Res* **18**: 4296.

- Trieu-Cuot, P., Carlier, C., Poyart-Salmeron, C., and Courvalin, P. (1991) An integrative vector exploiting the transposition properties of Tn1545 for insertional mutagenesis and cloning of genes from gram-positive bacteria. *Gene* **106**: 21-27.
- Turnbull, P.C. (1999) Definitive identification of *Bacillus anthracis*--a review. *J Appl Microbiol* **87**: 237-240.
- Uhl, M.A., and Miller, J.F. (1996a) Central role of the BvgS receiver as a phosphorylated intermediate in a complex two-component phosphorelay. J Biol Chem 271: 33176-33180.
- Uhl, M.A., and Miller, J.F. (1996b) Integration of multiple domains in a two-component sensor protein: the *Bordetella pertussis* BvgAS phosphorelay. *Embo J* **15**: 1028-1036.
- 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.
- van der Heijden, R.T., Snel, B., van Noort, V., and Huynen, M.A. (2007) Orthology prediction at scalable resolution by phylogenetic tree analysis. *BMC Bioinformatics* 8: 83.
- van Schaik, W., and Abee, T. (2005) The role of σ^{B} in the stress response of Gram-positive bacteria targets for food preservation and safety. *Curr Opin Biotechnol* **16**: 218-224.
- van Schaik, W., Tempelaars, M.H., Wouters, J.A., de Vos, W.M., and Abee, T. (2004a) The alternative sigma factor σ^{B} of *Bacillus cereus*: response to stress and role in heat adaptation. *J Bacteriol* **186**: 316-325.
- van Schaik, W., Zwietering, M.H., de Vos, W.M., and Abee, T. (2004b) Identification of o^Bdependent genes in *Bacillus cereus* by proteome and *in vitro* transcription analysis. J *Bacteriol* **186**: 4100-4109.
- van Schaik, W., Tempelaars, M.H., Zwietering, M.H., de Vos, W.M., and Abee, T. (2005a) Analysis of the role of RsbV, RsbW, and RsbY in regulating o^B activity in *Bacillus cereus. J Bacteriol* **187**: 5846-5851.
- van Schaik, W., Zwietering, M.H., de Vos, W.M., and Abee, T. (2005b) Deletion of the *sigB* gene in *Bacillus cereus* ATCC 14579 leads to hydrogen peroxide hyperresistance. *Appl Environ Microbiol* **71**: 6427-6430.
- van Schaik, W., van der Voort, M., Molenaar, D., Moezelaar, R., de Vos, W.M., and Abee, T. (2007) Identification of the σ^{B} regulon of *Bacillus cereus* and conservation of σ^{B} regulated genes in low-GC-content gram-positive bacteria. *J Bacteriol* **189**: 4384-4390.
- Vetter, S.M., and Schlievert, P.M. (2007) The two-component system *Bacillus* respiratory response A and B (BrrA-BrrB) is a virulence factor regulator in *Bacillus anthracis*. *Biochemistry* **46**: 7343-7352.
- Vijay, K., Brody, M.S., Fredlund, E., and Price, C.W. (2000) A PP2C phosphatase containing a PAS domain is required to convey signals of energy stress to the σ^B transcription factor of *Bacillus subtilis*. *Mol Microbiol* **35**: 180-188.
- Vilain, S., Luo, Y., Hildreth, M.B., and Brözel, V.S. (2006) Analysis of the life cycle of the soil saprophyte *Bacillus cereus* in liquid soil extract and in soil. *Appl Environ Microbiol* 72: 4970-4977.
- Viollier, P.H., Kelemen, G.H., Dale, G.E., Nguyen, K.T., Buttner, M.J., and Thompson, C.J. (2003) Specialized osmotic stress response systems involve multiple SigB-like sigma factors in *Streptomyces coelicolor*. *Mol Microbiol* **47**: 699-714.
- Voelker, U., Voelker, A., Maul, B., Hecker, M., Dufour, A., and Haldenwang, W.G. (1995) Separate mechanisms activate σ^B of *Bacillus subtilis* in response to environmental and metabolic stresses. *J Bacteriol* **177**: 3771-3780.
- Wadhams, G.H., and Armitage, J.P. (2004) Making sense of it all: bacterial chemotaxis. Nat Rev Mol Cell Biol 5: 1024-1037.
- Watanabe, T., Okada, A., Gotoh, Y., and Utsumi, R. (2008) Inhibitors targeting twocomponent signal transduction. Adv Exp Med Biol 631: 229-236.

- Wei, Y., Guffanti, A.A., Ito, M., and Krulwich, T.A. (2000) Bacillus subtilis YqkI is a novel malic/Na+-lactate antiporter that enhances growth on malate at low protonmotive force. J Biol Chem 275: 30287-30292.
- Wels, M., Francke, C., Kerkhoven, R., Kleerebezem, M., and Siezen, R.J. (2006) Predicting cis-acting elements of *Lactobacillus plantarum* by comparative genomics with different taxonomic subgroups. *Nucleic Acids Res* **34**: 1947-1958.
- White, A.K., Hoch, J.A., Grynberg, M., Godzik, A., and Perego, M. (2006) Sensor domains encoded in *Bacillus anthracis* virulence plasmids prevent sporulation by hijacking a sporulation sensor histidine kinase. *J Bacteriol* **188**: 6354-6360.
- Whitworth, D.E., and Cock, P.J. (2008) Two-component systems of the myxobacteria: structure, diversity and evolutionary relationships. *Microbiology* **154**: 360-372.
- Whitworth, D.E., and Cock, P.J. (2009) Evolution of prokaryotic two-component systems: insights from comparative genomics. *Amino Acids*.
- Wilson, A.C., Hoch, J.A., and Perego, M. (2008) Virulence gene expression is independent of ResDE-regulated respiration control in *Bacillus anthracis*. J Bacteriol **190**: 5522-5525.
- Yamamoto, H., Murata, M., and Sekiguchi, J. (2000) The CitST two-component system regulates the expression of the Mg-citrate transporter in *Bacillus subtilis*. *Mol Microbiol* **37**: 898-912.
- Yang, X., Kang, C.M., Brody, M.S., and Price, C.W. (1996) Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes Dev* **10**: 2265-2275.
- Yarwood, J.M., McCormick, J.K., and Schlievert, P.M. (2001) Identification of a novel twocomponent regulatory system that acts in global regulation of virulence factors of *Staphylococcus aureus*. J Bacteriol **183**: 1113-1123.
- Yea, C.L., Lee, C.L., Pan, T.M., and Horng, C.B. (1994) Isolation of *Bacillus cereus* in the feces of healthy adults in Taipei City. *Zhonghua Min Guo Wei Sheng Wu Ji Mian Yi Xue Za Zhi* 27: 148-151.
- Yeats, C., Rawlings, N.D., and Bateman, A. (2004) The PepSY domain: a regulator of peptidase activity in the microbial environment? *Trends Biochem Sci* **29**: 169-172.
- Yoshida, K., Ohki, Y.H., Murata, M., Kinehara, M., Matsuoka, H., Satomura, T., Ohki, R., Kumano, M., Yamane, K., and Fujita, Y. (2004) *Bacillus subtilis* LmrA is a repressor of the *ImrAB* and *yxaGH* operons: identification of its binding site and functional analysis of *ImrB* and *yxaGH*. J Bacteriol **186**: 5640-5648.
- Zhou, M., Boekhorst, J., Francke, C., and Siezen, R.J. (2008) LocateP: genome-scale subcellular-location predictor for bacterial proteins. *BMC Bioinformatics* **9**: 173.
- Zhulin, I.B., Taylor, B.L., and Dixon, R. (1997) PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends Biochem Sci* **22**: 331-333.
- Zhulin, I.B., Nikolskaya, A.N., and Galperin, M.Y. (2003) Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. J Bacteriol 185: 285-294.
- Zientz, E., Bongaerts, J., and Unden, G. (1998) Fumarate regulation of gene expression in Escherichia coli by the DcuSR (dcuSR genes) two-component regulatory system. J Bacteriol 180: 5421-5425.
- Zigha, A., Rosenfeld, E., Schmitt, P., and Duport, C. (2007) The redox regulator Fnr is required for fermentative growth and enterotoxin synthesis in *Bacillus cereus* F4430/73. *J Bacteriol* **189**: 2813-2824.
- Zimmer, M.A., Tiu, J., Collins, M.A., and Ordal, G.W. (2000) Selective methylation changes on the *Bacillus subtilis* chemotaxis receptor McpB promote adaptation. *J Biol Chem* **275**: 24264-24272.

Samenvatting voor iedereen

Als wetenschapper spreek je vaak in een voor "normale" mensen onbegrijpelijke taal. Daarom zal ik hier een poging wagen om mijn onderzoek in simpele bewoordingen uit te leggen. Het onderzoek dat in dit proefschrift beschreven staat gaat over *Bacillus cereus*. Dat is een staafvormige bacterie met een lengte van ongeveer 4 μ m (dat is 4 duizendste van een millimeter: voor een indruk, zie Fig. 1.1B). *Bacillus cereus* is een belangrijke veroorzaker van voedselvergiftigingen en –infecties. Produkten die vaak worden verontreinigd door *B. cereus* zijn vlees, groenten, pudding, melk, rijst en pasta. Typische symptomen van een *B. cereus* voedselvergiftiging zijn overgeven en diarree. Ook al zijn deze symptomen vaak mild en van korte duur, er zijn onlangs *B. cereus* uitbraken beschreven met dodelijke afloop. Daarnaast lijkt het aantal gevallen van *B. cereus* uitbraken wereldwijd toe te nemen, vooral in Westerse landen.

B. cereus hoort bij een groep van nauw verwante bacteriën, waaronder de soorten *B. anthracis* en *B. thuringiensis*. Van deze zogenaamde "*B. cereus* groep" is *B. anthracis* - de veroorzaker van miltvuur (anthrax) - het meest beruchte lid, vooral sinds deze bacterie in 2001 werd gebruikt bij poederbriefaanslagen in de Verenigde Staten van Amerika.

In de afgelopen jaren is van een hoop bacteriën, inclusief van een aantal leden van de *B. cereus* groep, de volgorde van de bouwsteentjes van het DNA ontrafeld. Deze volgorde (de erfelijke code) bepaalt welke eiwitten een bacterie kan maken onder bepaalde omstandigheden. Eiwitten zijn een soort machientjes die in de bacterie allerlei chemische reacties uitvoeren, zoals het opnemen en afbreken van voedsel en het uitscheiden van afvalstoffen. Door de erfelijke code te inspecteren kan dus veel, zoniet alle, informatie verkregen worden over de eigenschappen van een bacterie. Maar deze informatie ligt niet zomaar voor het oprapen. Het DNA van *B. cereus* bijvoorbeeld is opgebouwd uit meer dan 5 miljoen bouwsteentjes! En ergens in deze brij van bouwsteentjes liggen de stukjes code (genen) die zorgen voor de aanmaak van eiwitten. Gelukkig is er in de afgelopen jaren veel software ontwikkeld waarmee deze genen op efficiënte wijze geïdentificeerd en hun produkten (eiwitten) geanalyseerd kunnen worden.

Tijdens mijn onderzoek heb ik veelvuldig gebruik gemaakt van bovengenoemde software om de erfelijke code van *B. cereus* en zijn nauwe verwanten te inspecteren. Met deze software heb ik voornamelijk naar twee klassen van eiwitten gezocht en die geanalyseerd: Histidine Kinases (HKs) en Response Regulators (RRs). Er was namelijk al bekend dat HKs en RRs een belangrijke rol spelen bij de aanpassing van bacteriën aan stressvolle condities, zoals een hoge temperatuur of een zuur milieu.

Vergeleken met de meeste bacteriën heeft *B. cereus* een extreem hoog aanpassingvermogen. Dit wordt mede veroorzaakt doordat *B. cereus* sporen kan

vormen: de bacterie gaat dan in een soort stand-by toestand en bouwt een beschermend schild om zich heen waardoor hij deze ongunstige omstandigheden vaak kan overleven (Fig. 1.1C). Wanneer de omstandigheden gunstiger worden kruipt de bacterie weer uit dit schild. Het hoge aanpassingsvermogen van *B. cereus* veroorzaakt veel problemen voor de levensmiddelenindustrie. Tijdens de industriële bereiding van voedsel worden vaak behandelstappen toegepast (zoals verhitting) om alle bacteriën in het product te doden. *B. cereus* kan deze behandelstappen relatief gemakkelijk overleven en blijft zo in het voedsel zitten.

Het hoge aanpassingsvermogen van *B. cereus* zou wel eens te maken kunnen hebben met de bovengenoemde eiwitten Histidine Kinases (HKs) en Response Regulators (RRs) die in het DNA van *B. cereus* gecodeerd liggen. Het DNA van een gemiddelde bacterie codeert voor een handjevol verschillende HKs en RRs. De meeste van deze HKs en RRs vormen specifieke koppels, ook wel tweecomponenten systemen (TCS) genoemd. Grofweg is elk TCS betrokken bij de aanpassing aan één bepaalde stress: zo zijn er TCSs voor de aanpassing aan een hoge temperatuur, aan een zuur milieu, etc.

Een TCS functioneert als volgt: de HK zit als een soort sensortje aan de buitenkant van de bacterie. Wanneer een bepaalde ongunstige omstandigheid zich in de omgeving van de bacterie voordoet, wordt deze stress door de bijpassende HK geregistreerd. Deze HK geeft dan een alarmsignaaltje door aan zijn partner RR. Deze RR, die zich *in* de bacterie bevindt, gaat vervolgens zijn taak uitvoeren: dat is het herkennen en binden aan kleine stukjes erfelijke code (schakelaartjes) waarmee een specifiek groepje van genen wordt geactiveerd en dus een specifiek groepje van eiwitten wordt aangemaakt. Met deze eiwitten kan de bacterie zich vervolgens beter weren tegen de stress die geregistreerd werd door de HK (zie Fig. 1.2 voor een schematische weergave). Bijvoorbeeld door ervoor te zorgen dat de bacterie sporen gaat vormen. Maar er zijn ook TCSs bekend die de zweepstaartjes kunnen aansturen, waarmee de bacterie kan wegzwemmen van vervelende situaties. Simpel gezegd vormen TCSs de zintuigen en het zenuwstelsel van bacteriën, waarmee prikkels worden geregistreerd en de juiste reactie op die prikkels wordt geregeld.

Lijst van termen

term	beschrijving
DNA	Elk levend organisme bevat DNA. De volgorde van de "steentjes" waaruit DNA is opgebouwd
	bepaar de errelijke code.
gen	Een stukje erfelijke code dat zorgt voor de aanmaak van een specifiek eiwit op een specifiek moment.
eiwit	Een soort machientje dat allerlei chemische reacties uitvoert in een organisme.
нк	Histidine Kinase: type eiwit dat als een sensor de omgeving van de bacterie in de gaten houdt. Elke HK reageert op een specifieke prikkel of stress.
RR	Response Regulator: type eiwit dat alarmsignaaltjes ontvangt van een HK. Wanneer dit gebeurt bindt de RR op stukjes erfelijke code (schakelaartjes) om genen te activeren.
TCS	Twee-Componenten Systeem: bestaat uit een HK en een RR die een koppel vormen. TCSs zorgen voor de juiste reactie van een bacterie op zijn omgeving.
spore	Ultieme bacteriële overlevingscapsule. Niet alle bacteriën zijn in staat om sporen te vormen.

Zoals ik hier boven heb beschreven, heb ik me bij mijn onderzoek gericht op het in kaart brengen en analyseren van TCSs in *B. cereus*. Dit heeft interessante kennis opgeleverd over hoe *B. cereus* in staat is zich zo makkelijk aan te passen aan zijn omgeving. Hieronder volgt een kort overzicht van mijn belangrijkste bevindingen per hoofdstuk.

In **hoofdstuk 2** staat beschreven hoe met behulp van software het totale arsenaal aan HKs en RRs in *B. cereus* en een aantal nauwe verwanten in kaart is gebracht. Hieruit bleek dat leden van de *B. cereus* groep beschikken over een relatief groot aantal HKs en RRs (zie Fig. 7.1). Met behulp van software zijn deze HKs en RRs vergeleken met die van bacteriën waaraan veel onderzoek is gedaan. Op deze manier is voorspeld welke TCSs bij welke processen betrokken zijn. Een voorbeeld: *B. cereus* blijkt over een uitgebreid systeem van HKs en RRs te beschikken waarmee het vormen van een spore wordt gereguleerd. Blijkbaar "denkt" *B. cereus* wel drie keer na voordat hij zich omvormt tot een spore.

Hoofdstuk 3 beschrijft de toepassing van een nieuwe methode om de specifieke stukjes erfelijke code (schakelaartjes) die door RRs herkend worden te voorspellen. Deze schakelaartjes liggen in de buurt van de genen die geactiveerd worden zodra de RR aan het betreffende schakelaartje bindt. Door de schakelaartjes en bijbehorende genen van alle RRs in een bacterie in kaart te brengen wordt duidelijk waarvoor een bacterie al zijn verschillende TCSs gebruikt.

Om de voorspellende waarde van bovenstaande methode extra kracht bij te zetten, werden verschillende *B. cereus* mutanten in het lab getest. De resultaten hiervan staan beschreven in **hoofdstuk 4**. In de mutanten werden opzettelijk genen stukgemaakt die coderen voor een bepaald TCS. *B. cereus* beschikt dan dus niet meer over dat bepaalde TCS. Hierdoor kunnen de schakelaartjes voor de betreffende RR ook niet meer aangezet worden wanneer dat nodig is en kan de bacterie niet meer reageren op bepaalde prikkels en stressen. De uitkomsten van de experimenten bleken zeer goed overeen te komen met de eerdere voorspellingen, met andere woorden, de correcte schakelaartjes waren bij de correcte RRs voorspeld.

Bij de identificatie van alle HKs en RRs in *B. cereus* (hoofdstuk 2) werd een zeer complexe HK ontdekt. Met computer-analyses werd voorspeld dat deze HK samenwerkt met een bijzondere RR die betrokken is bij de aanpassing van *B. cereus* aan meerdere ongunstige factoren, waaronder een hoge temperatuur en een hoge zoutconcentratie. Om deze voorspellingen te toetsen werd een *B. cereus* mutant getest die niet meer over deze HK beschikte. De resultaten hiervan staan beschreven in **hoofdstuk 5**. Uit de experimenten bleek dat de voorspellingen inderdaad klopten en dat de *B. cereus* mutant eerder dood ging bij een hoge temperatuur dan de niet-gemuteerde *B. cereus*. De gekoppelde RR is bijzonder omdat die niet aan schakelaartjes op het DNA, maar aan een ander eiwit bindt. Wanneer dit gebeurt, wordt er een hele cascade van reacties in gang

gezet die er uiteindelijk toe leidt dat een belangrijk regulator eiwit actief wordt. Deze regulator is wel in staat om schakelaartjes aan te zetten, waardoor *B. cereus* zich kan wapenen tegen bovengenoemde stressen.

Vergeleken met andere bacteriën, die over hun eigen versie van laatstgenoemde regulator eiwit beschikken, leek de activatie van dit eiwit in *B. cereus* met behulp van een TCS aanvankelijk heel "vreemd". Waarom doet *B. cereus* het op een andere manier? En zijn er ook andere bacteriën die over eenzelfde soort mechanisme beschikken als *B. cereus*? In **hoofdstuk 6** is gevonden dat dit laatste inderdaad het geval is. Het *B. cereus*-achtige mechanisme blijkt zelfs vrij algemeen te zijn voor een bepaalde groep van bacteriën. Daarnaast blijkt uit computer-analyses dat het betreffende TCS mogelijk samenwerkt met nog een ander eiwit, wat het geheel nog complexer maakt. Toekomstige experimenten zullen moeten uitwijzen hoe dit precies werkt.

Het onderzoek dat beschreven staat in dit proefschrift heeft nieuwe inzichten opgeleverd over hoe *B. cereus* zich aanpast aan zijn omgeving met behulp van TCSs. Deze inzichten zijn belangrijk om het gedrag van *B. cereus* te voorspellen tijdens bijvoorbeeld de industrieële bereiding en behandeling van voedsel. Daarnaast kunnen deze inzichten in de toekomst mogelijk toegepast worden om op een heel specifieke manier TCSs te blokkeren. In de medische wereld worden al enige successen geboekt met nieuwe antibiotica die in staat zijn bacteriën "blind" te maken voor hun omgeving door de werking van HKs te verstoren.

Dankwoord

Voordat ik mensen ga bedanken, volgt eerst een korte anecdote. Mijn promotie onderzoek begon destijds met een e-mail van ene Tjakko Abee met als titel "sollicitatie". Even dacht ik in dat hier een mailtje door de spam-filter was geglipt, want ik kende geen Tjakko Abee en de meeste mailtjes met "sollicitatie" in de titel zijn van mensen die me aanspreken met Dear Prof. Dr. de Been en die bij mij willen solliciteren. Even dacht ik er ook aan om het mailtje maar te verwijderen, toen één van mijn collega's riep: "Tjakko Abee, ja dat is een WCFSer." Toch maar even lezen dan. Of ik zou willen solliciteren op een aio-project met als onderwerp: het in kaart brengen van signaal transductie netwerken van *Bacillus cereus*. Dat leek me erg interessant! Daarnaast zou het onderzoek een combinatie van bioinformatica en labwerk worden. Dat was zelfs precies wat ik zocht! En van het één komt het ander en twee maanden later was ik aio met alles erop en eraan, inclusief mijn eigen lading visite-kaartjes waar ik de rest van mijn leven op zou kunnen teren. Ik ben erg blij dat ik dat mailtje toen niet heb verwijderd, want ik had de afgelopen periode niet willen missen.

Allereerst wil ik mijn (co-)promotoren Tjakko, Roland en Christof bedanken voor de fijne samenwerking. Ik heb veel van jullie geleerd en ik waardeer het dat jullie me bijna volledig hebben vrijgelaten om mijn eigen richting aan het onderzoek te geven. Tjakko: jouw enorme enthousiasme heeft me telkens weer gemotiveerd, ook wanneer het qua experimenten even niet wilde vlotten. Roland: door jou ben ik in eerste instantie geïnteresseerd geraakt in bioinformatica en genomics. Christof: bedankt voor je kritische blik. Elke keer wanneer ik dacht: "Zo dit artikel is nu eindelijk echt af", wist jij dat artikel toch weer van een flinke dosis commentaar te voorzien, maar wel op zo'n manier dat ik telkens dacht: "Ja, daar heeft 'ie een goed punt!"

Mijn onderzoek maakte deel uit van het WCFS/TIFN C009 project. Roy: bedankt voor de flair waarmee je dit project geleid hebt. Jouw enthousiasme en betrokkenheid maken je een prettig persoon om mee samen te werken. Daarnaast wil ik natuurlijk ook alle andere C009 teamleden bedanken voor de prettige samenwerking en de gezellige C009 en WCFS/TIFN uitjes. Ik bedank hier alvast Ineke, Stijn en Masja (andere teamleden: zie beneden).

De eerste twee jaar van mijn onderzoek heb ik vooral doorgebracht op het CMBI in Nijmegen. Graag wil ik dan ook alle CMBI-ers bedanken voor de leerzame en gezellige tijd. Mijn dank gaat in het bijzonder uit naar de mensen van de Bamics-groep: mijn kamergenoot-bamixers Jos, Michiel, Miaomiao, Mengjin; de bamixers van de buurkamer Richard, Robert, Tom, Juma, Barzan; de overige bamixers Bas, Bernadet, Greer, Peter, Frank; en de vele bamics studenten die ik in de loop der tijd voorbij heb zien komen, waaronder in het bijzonder "mijn" studenten Marieke en Anita. Iedereen bedankt voor alle leerzame discussies en programmeer-trucjes. En ik heb genoten van de vele *in silico* (Blobby-Volley, Liero, Midtown-Madness, Nintendo-Wii) alsmede *ex silico* (Bamics speurtochten)

competities. Dank tenslotte ook aan Wilmar, Victor en Jos voor het draaiende houden van de bamics-servers.

Naast mijn tijd op het CMBI heb ik een groot deel van mijn tijd pipetterend doorgebracht in Wageningen. Graag wil ik iedereen van de afdeling Food Micro bedanken voor de leerzame en gezellige tijd. In het bijzonder Marcel T, de duizendpoot van ons lab. Marcel: bedankt voor de prettige samenwerking en dank dat je als paranimf aan mijn zijde wilt staan. Willem, Maarten, Heidy, Menno en Janneke: bedankt voor jullie lab-support: het valt niet altijd mee om een BioIT-er te zien huishouden in het lab. Clint: je was een gezellige kamergenoot, ik kijk uit naar onze volgende pot Arkham Horror. Sachin: I enjoyed the car trips with you as my passenger. Zeus and Lídia: thanks for the nice squash-competitions and (Zeus) for the nice time in room 416 (I think we scared away a lot of people with our long discussions about zombie-movies). Julien: merci beaucoup for the nice collaboration and for showing me Avignon.

Naast collega's is er ook een hele lading vrienden en familie die (in)direct betrokken is geweest bij mijn onderzoek. Bas, Saskia, Ward, Hanneke, Jo-Anne, Frank, Jirina, Jan, Maria, Caroline, *****, Marianne, Mark, Sandra, Tamara, Josef, Frank, Paulien, Marian, Kim, Hanneke, Ron, Martijn, Dorri, Rick, Fiepke, Eva, Horst en Hienke: allen bedankt voor o.a. de spelletjesavonden (ookal krijg ik daar soms de neiging van om dobbelstenen bij mensen in de oogjes te gooien), de tripjes naar Lurcy, de squashwedstrijden, LOTR marathons, etc etc. Tamara: ik vind het erg fijn om je als paranimf naast me te hebben. Ton, Trix, Sjoerd, Willem en Ingrid: bedankt voor alle gezellige borreltjes en uitstapjes. We moeten snel maar weer eens een avondje gaan "Jossen". Pa, Ma, Kim en Niels: bedankt dat jullie altijd voor me klaar staan en dat jullie me altijd mijn eigen interesses hebben laten volgen.

En tot slot, Anneke: heel veel dank voor al je begrip, geduld, steun en liefde. Ik ben erg blij dat ik die avond op dat biologenfeest was.

List of publications

de Groot P, Weterings K, **de Been M**, Wittink F, Hulzink R, Custers J, van Herpen M, Wullems G (2004) Silencing of the pollen-specific gene NTP303 and its family members in tobacco affects *in vivo* pollen tube growth and results in male sterile plants. *Plant Mol. Biol.* 55(5): 715 - 726

Cirpus IE, **de Been M**, den Camp HJ, Strous M, Le Paslier D, Kuenen GJ, Jetten MS (2005) A new soluble 10kDa monoheme cytochrome *c*-552 from the anammox bacterium *Candidatus* "*Kuenenia stuttgartiensis*". *FEMS Microbiol. Lett.* 252(2): 273 - 278

Boekhorst J, **de Been M**, Kleerebezem M, Siezen RJ (2005) Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J. Bacteriol.* 187(14): 4928 - 4934

de Been M, Francke C, Moezelaar R, Abee T, Siezen RJ (2006) Comparative analysis of two-component signal transduction systems of *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus anthracis*. *Microbiology*. 152(Pt 10): 3035 - 3048

Mols M, **de Been M**, Zwietering MH, Moezelaar R, Abee T (2007) Metabolic capacity of *Bacillus cereus* strains ATCC 14579 and ATCC 10987 interlinked with comparative genomics. *Environ. Microbiol.* 9(12): 2933 - 2944

Vandenbussche M, Janssen A, Zethof J, van Orsouw N, Peters J, van Eijk M, Rijpkema AS, Schneiders H, **de Been M**, van Tunen A, Gerats T (2008) Generation of a 3D indexed *Petunia* insertion database for reverse genetics. *Plant J.* 54 (6): 1105 - 1114

de Been M, Bart MJ, Abee T, Siezen RJ, 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(10): 2796 - 2809

de Been M, Brillard J, Brousolle V, Abee T. Integrating phylogenetic footprinting approaches with transcriptome profiling reveals new roles for two *Bacillus cereus* two-component signal transduction systems. *Submitted for publication*

de Been M, Tempelaars MH, van Schaik W, Moezelaar R, Siezen RJ, Abee T. A novel hybrid kinase is essential for activating the σ^{B} -mediated stress response of *Bacillus cereus*. Submitted for publication

de Been M, Francke C, Siezen RJ, Abee A. Novel σ^{B} activation modules of Gram-positive bacteria involve the use of complex hybrid histidine kinases. *Manuscript in preparation*

Tempelaars MH, **de Been M**, Abee T. Functional analysis of σ^{H} and identification of its regulon in *Bacillus cereus* ATCC 14579. *Manuscript in preparation*

Curriculum Vitae

Mark de Been was born on the 23rd of April 1980 in Raamsdonksveer, The Netherlands. He finished his secondary education at the Dongemond-college in Raamsdonksveer in 1998. That same year he started his studies in Biology at the Radboud University Nijmegen (then Katholieke Universiteit Nijmegen). As part of his MSc he carried out three research projects at the Radboud University Nijmegen. In the first project, at the department of Molecular Plant Biology, he analysed the phenotypic effects of a pollen-specific gene knockout in Nicotiana tabacum, under supervision of Prof. George Wullems. The second research project was performed at the department of Microbiology, under the supervision of Dr. Marc Strous and Prof. Mike Jetten. The goal of this project was to purify and characterise *c*-type heme proteins from the anammox bacterium *Kuenenia* stuttgartiensis. In the final phase of his studies, he performed a research project at the Centre for Molecular and Biomolecular Informatics (CMBI), where he specialised in bioinformatics and bacterial genomics, under supervision of Prof. Roland Siezen. In this project, he focused on cell wall-anchored surface proteins of Gram-positive bacteria. After finishing his studies cum laude in 2004, he continued his work at the CMBI as a junior researcher to extend his MSc research project.

After concluding the research project at the CMBI by the end of 2004, Mark started his PhD project entitled: "Signal transduction and stress responses in the food-borne pathogen *Bacillus cereus.*" This project was part of a larger project on food safety and food-borne pathogens, which was funded by the Top Institute Food and Nutrition (TIFN, Wageningen). For his PhD research, he conducted his work at both the CMBI and the Food Microbiology laboratory (Wageningen University), under the supervision of Prof. Roland Siezen, Prof. Tjakko Abee and Dr. Christof Francke. The employment at the two departments allowed for a unique combination of genomics and wet-lab research. The results obtained are described in this thesis, and have been published in international peer-reviewed journals and presented at several international scientific conferences.

Mark is currently working on different TIFN projects and preparing to go abroad for a postdoc.

VLAG graduate school activities

Discipline specific activities

Genetics and physiology of food-associated microorganisms, VLAG, 2004 GeneMaths workshop, Applied Maths, 2006

CMBI Conferences, Ravenstein, 2004, 2005, 2007 Netherlands Conference on Bioinformatics, "Images of Life", Groningen, 2004 WCFS / TIFN Food Summit, 2006 Benelux Bioinformatics Conference, Wageningen, 2006 (poster) NBIC Netherlands Bioinformatics Conference, Ede, 2006 NBIC Netherlands Bioinformatics Conference, Amsterdam, 2007 (poster and oral presentation)

Conference on Gram-Positive Genomics, San Diego, USA, 2005 (poster) Conference on Gram-Positive Genomics, Tirrenia, Italy, 2007 (poster) Conference on Gram-Positive Genomics, San Diego, USA, 2009 (oral presentation)

General activities

VLAG PhD week, 2004 Presentation skills, TIFN / Bob de Groof, 2005

Optional courses and activities

Preparation of PhD research proposal

VLAG PhD trip, South Africa, 2005 VLAG PhD trip, Canada, 2008 Working visits and mini-symposiums, TIFN / INRA, 2006, 2007, 2009 WCFS / TIFN days, 2004-2008 WCFS / TIFN C009 project meetings, 2004-2008 WCFS / TIFN BioIT meetings, 2004-2008

Cover design and lay-out: Mark de Been & Anneke Rijpkema