Classification and Functional Characterization of Extracytoplasmic Function (ECF) σ Factors from *Planctomycetes* and *Actinobacteria*

Xiaoluo Huang

Classification and Functional Characterization of Extracytoplasmic Function (ECF) σ Factors from *Planctomycetes* and *Actinobacteria*

Dissertation

zur Erlangung des Doktorgrades

der Fakultät für Biologie

der Ludwig-Maximilian-Universität München

vorgelegt von

Xiaoluo Huang

aus Hubei, China

München, 2015

Erstgutachter: Prof. Dr. Thorsten Mascher

Zweitgutachter: PD Dr. Ralf Heermann

Tag der Abgabe: 15.07.2015

Tag der mündlichen Prüfung: 08.09.2015

Eidesstattliche Versicherung und Erklärung

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

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

München

Xiaoluo Huang

Contents

Abbreviations ... III

List of publications presented in the thesis ... IV

Contributions to the publications ... V

Summary ... VII

Zusammenfassung ... VIII

Chapter 1 ... 1

1 Introduction ... 2

1.1 Overview of bacterial signal transduction systems ... 2

1.1.1 One-component systems ... 2

1.1.2 Two-component systems ... 3

1.1.3 ECF σ factors ... 6

1.2 ECF classification ... 7

1.3 Signaling mechanisms of ECF σ factors ... 10

1.3.1 Regulated proteolysis of the anti-σ factor ... 10

1.3.2 Conformational change of the anti-σ factor ... 11

1.3.3 Activation of ECF σ factors by protein-protein interaction ... 11

1.3.4 Activation of ECF σ factors through a partner switch mechanism ... 14

1.3.5 Regulation of the activity of ECF σ factors by their C-terminal extension ... 14

1.3.6 Transcriptional activation of ECF σ factors ... 15

1.3.7 Activation of ECF σ factors by Ser/Thr Kinases ... 16

1.4 General features of *Planctomycetes* and *Actinobacteria* phyla ... 17

1.4.1 *Planctomycetes* ... 17

1.4.2 *Actinobacteria* ... 18

1.4.2.1 The genus *Streptomyces* ... 19

1.5 Aims of this study ... 22

Chapter 2 ... 24

Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in *Planctomycetes* by comparative genomics ... 24

Supplementary files for Chapter 2 ... 37

Chapter 3 ... 38

Environmental sensing in *Actinobateria*: A comprehensive survey on the signaling capacity of this phylum. ... 38

Supplementary files for Chapter 3 ... 58

Chapter 4 ... 59

Contents

Defining the regulon of genes controlled by σ^E, a key regulator of cell envelope stress in *Streptomyces coelicolor* 59

Supplementary files for Chapter 4 103

Chapter 5 105

5 Discussion 106

5.1 Comparative genomics of ECF σ factors in *Planctomycetes* and *Actinobacteria* 106

5.1.1 Comparative genomics, a powerful tool to identify novel ECF groups 106

5.1.2 The ECF-dependent regulation in the distinct bacterial phylum, *Planctomycetes* 107

5.1.3 The ECF-dependent regulation in *Actinobacteria* 109

5.2 Defining the regulon of genes controlled by σ^E in *Streptomyces* 111

5.2.1 How many genes are controlled by σ^E in *S. coelicolor*? 111

5.2.2 Genes that are differentially transcribed in the *S. coelicolor ΔsigE* mutant 112

5.3 Outlook 115

5.3.1 Assigning the remaining unclassified ECF σ factors into novel ECF groups 115

5.3.2 Validation of the hypothesis obtained from comparative genomics of ECF σ factors 116

5.3.3 Further elucidation of the regulatory network governed by σ^E in *Streptomyces* 116

Supplementary material for Chapter 5 118

References for Chapter 1 and Chapter 5 120

Acknowledgements 135

Curriculum Vitae -Xiaoluo Huang 137

Abbreviations

1CS	one-component system
2CS	two-component system
ECF	extracytoplasmic function
DHp	dimerization and histidine phosphotransferase
CA	catalytic and ATPase
RD	receiver domain
PAS	Per, Arnt, Sim
HAMP	histidine kinases, adenyl cyclases, methyl-accepting proteins and phosphatases
GAF	cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA
Hpt	histidine phosphotransfer
OMP	outer membrane protein
ASD	anti-σ domain
CLD	C-terminal cupin-like domain
SL	σ factor like
ICM	intra-cytoplasmic membrane
ChIP-seq	chromatin immunoprecipitation-sequencing

List of publications presented in the thesis

1. Jogler C, Waldmann J, **Huang X**, Jogler M, Glöckner FO, Mascher T, Kolter R (2012) *J Bacteriol* **194:** 6419-6430.

2. Huang X[#], Pinto D[#], Fritz G, Mascher T (2015) *J Bacteriol* **197:** 2517-2535. ([#]**Co-first author**.)

3. Tran NT[#], **Huang XL**[#], Bush MJ, Chandra G, Pinto D, Hong HJ, Hutchings MI, Mascher T and Buttner MJ. ([#]**co-first author,** in preparation**)**

Contributions to the publications

Chapter 2

Jogler C, Waldmann J, **Huang X**, Jogler M, Glöckner FO, Mascher T, Kolter R (2012) *J Bacteriol* **194:** 6419-6430

Christian Jogler designed the study and supervised Jost Waldmann and Mareike Jogler. He performed protein functional domain analysis and manually cured the final candidate gene set. Christian Jogler wrote major parts of the manuscript and serves as corresponding author. Jost Waldmann developed the software pipeline for comparative genomics and performed the experiment and wrote the corresponding parts of the manuscript. **Xiaoluo Huang** retrieved the sequences of ECF σ factors, performed the classification of ECF σ factors and generated Table and Figures for this part. Mareike Jogler analyzed the dcw gene contend and syntheny. Frank Oliver Glöckner contributed in designing the softwaretool with Jost Waldmann, provided the computer system and helped writing the manuscript. Thorsten Mascher performed statistical analysis of signal transduction proteins and wrote a draft for the signal transduction part in the paper. Roberto Kolter provided the work environment for Jost Waldmann and Christian Jogler, supported the design of the study and wrote major parts of the manuscript.

Chapter 3

Huang X[#], Pinto D[#], Fritz G, Mascher T (2015) *J Bacteriol* **197:** 2517-2535 ([#]**Co-first author**.)

Xiaoluo Huang, Daniela Pinto, Georg Fritz and Thorsten Mascher designed the story. **Xiaoluo Huang** and Daniela Pinto retrieved the data, performed the classification of ECF σ factors and generated tables and figures for this part. **Xiaoluo Huang** performed the classification of the response regulators. **Xiaoluo Huang** retrieved the data for the analysis of the distribution of signal transduction proteins within 119 genomes and generated tables for this part; Georg Fritz did the statistical analysis of the distribution of signal transduction proteins within 119 genomes and generated figures for this part. Daniela Pinto and Georg Fritz classified histidine kinases and one component system. Daniela Pinto and Thorsten Mascher wrote the paper.

Chapter 4

Tran NT[#], **Huang XL**[#], Bush MJ, Chandra G, Pinto D, Hong HJ, Hutchings MI, Mascher T and Buttner MJ. ([#]**Co-first author,** in preparation**)**

Ngat Tran, **Xiaoluo Huang**, Hee-Jeon Hong, Matt Bush, Thorsten Mascher and Mark Buttner designed the experiments. Ngat Tran contributed to S1 mapping and *in vitro* transcription experiments and made the target mutants, and **Xiaoluo Huang** constructed the *ΔsigE attB*$_{\Phi BT1}$::3×FLAG-*sigE* strain, characterized its phenotype and performed the ChIP-seq experiments. Hee-Jong Hong and Ngat Tran performed the microarray analysis**.** **Xiaoluo Huang**, Ngat Tran, Govind Chandra, Daniela Pinto, Matt Bush and Thorsten Mascher performed the bioinformatic analysis. **Xiaoluo Huang**, Matt Bush, Ngat Tran, Matt Hutchings and Mark Buttner wrote the paper.

We hereby confirm the above declaration.

Xiaoluo Huang Prof. Dr. Thorsten Mascher

Summary

Bacteria necessitate multiple signal transduction systems to sense the ever-changing environments and mediate the cellular response accordingly. The major bacterial signal transduction systems are one-component system (1CS), two-component system (2CS) and extracytoplasmic function (ECF) σ factor. Compared to 1CSs and 2CSs, ECF σ factors have only been identified much later and therefore the knowledge about their molecular mechanisms and physiological roles is less profound. This thesis mainly focusses on the study of ECF σ factors from the bacterial phyla, *Planctomycetes* and *Actinobacteria*.

In the first two parts of this study, ECF σ factors from eight planctomycetal genomes and 119 actinobacterial genomes were classified and analyzed in depth. This led to the identification of eight novel ECF groups consisting of 202 protein members from *Planctomycetes* and 18 novel ECF groups consisting of 427 protein members from *Actinobacteria*, respectively. Many of these novel ECF groups were found to show unusual properties. For example, five ECF groups (ECF01-Gob, ECF48, ECF52, ECF53 and ECF56) contained extended C-terminal domains. Four ECF groups (STK1-STK4) were genomically adjacent to serine/threonine kinases. Further analyses of these properties suggested novel ECF-dependent signal transduction mechanisms.

In the third part of this study, we identified the genes controlled by σ^E, a conserved ECF σ factor in the actinobacterial genus, *Streptomyces*. Since σ^E has been characterized to be important for cell envelope stress response in *S. coelicolor*, we firstly defined the σ^E regulon in this organism using a combination of chromatin immunoprecipitation-sequencing (ChIP-seq), DNA microarray and bioinformatic analyses. Thus, 91 target genes were assigned into the σ^E regulon in *S. coelicolor*. By in depth analysis of these genes, it was found that approximately half of them encode proteins showing cell envelope related functions. Amongst the remaining target genes, proteins involved in cell regulation and cell metabolism could be identified. Subsequently, 17 of these 91 targets were validated by S1 mapping or *in vitro* transcription. Using this S1 mapping data set, we identified promoters for all of these 17 targets and established a σ^E binding consensus, consisting of a conserved "AAC" at -35 region and a "TC" at -10 region. Next, we predicted all the σ^E binding sites across 19 *Streptomyces* genomes and established a putative σ^E regulon for every *Streptomyces* genome. Finally, we selected those targets that were conserved in at least 9 *Streptomyces* genomes and built a core σ^E regulon.

Zusammenfassung

Bakterien benötigen eine Vielzahl von Signaltransduktionssystemen, um die sich ständig ändernden Umweltbedingungen wahrzunehmen und eine entsprechende zelluläre Antwort zu vermitteln. Die drei wichtigsten Signaltransduktionssysteme in Bakterien sind Einkomponentensysteme (1CS), Zweikomponentensysteme (2CS) und extracytoplasmatisch-aktive (ECF) σ Faktoren. ECF σ Faktoren wurden im Vergleich zu 1CS und 2CS erst spät entdeckt und sowohl ihre molekularen Wirkmechanismen als auch ihre physiologische Bedeutung sind weniger gut erforscht. Diese Arbeit fokussiert sich hauptsächlich auf die Analyse von ECF σ Faktoren aus den Phyla der Planctomyceten und der Actinobakterien.

Die ersten beiden Teile dieser Arbeit beschäftigen sich mit der Klassifizierung und der detaillierten Analyse von ECF σ Faktoren aus 8 planctomycetischen sowie 119 actinobakteriellen Genomen. Etwa 202 zuvor noch nicht klassifizierte σ Faktoren aus Planctomyceten, sowie 427 aus Actinobakterien, konnten in 26 neue ECF Gruppen eingeordnet werden. Einige von ihnen wiesen ungewöhnliche Eigenschaften auf. So besitzen zum Beispiel 5 ECF Gruppen (ECF01-Gob, ECF48, ECF52, ECF53 und ECF56), zusätzlich zu ihren konservierten σ_2 und σ_4 Domänen, erweiterte C-terminale Domänen, und vier Weitere (STK1-STK4) liegen im Genom in unmittelbarer Nähe zu Serin-/Threoninkinasen. Tiefergehende Analysen deuten auf neuartige ECF-abhängige Signaltransduktionsmechanismen hin.

Im dritten Teil dieser Arbeit wurden die Gene identifiziert, die unter der Kontrolle von σ^E stehen. Hierbei handelt es sich um einen konservierten ECF σ Faktor aus Streptomyceten, welcher dem Phylum der Actinobakterien angehören. Da σ^E eine wichtige Rolle bei der Zellhüll-Stressantwort in *S. coelicolor* spielt, wurde zunächst das σ^E Regulon in diesem Organismus durch eine Kombination von Chromatin-Immunopräzipitation mit anschließender Sequenzierung (ChIP-seq), DNA-Microarray und bioinformatischen Analysen bestimmt. Es konnten dem σ^E Regulon in *S. coelicolor* 91 Zielgene zugeordnet werden. Mit Hilfe einer detaillierten Analyse dieser Gene konnte herausgefunden werden, dass etwa die Hälfte von ihnen für Proteine kodieren, deren Funktion im Zusammenhang mit der Zellhülle steht. Viele weitere scheinen in die Zellregulation und den

Zellmetabolismus involviert zu sein. 17 der 91 Zielgene wurden *in vitro* durch S1-Kartierung oder *in vitro* Transkription bestätigt. Durch die daraus gewonnenen Daten konnten die Promotoren für alle 17 Gene und die Konsensussequenz der σ^E-Bindestelle identifiziert werden. Letztere besteht aus den konservierten Nukleotiden "AAC" in der -35 Region und "TC" in der -10 Region. Im Folgenden wurden alle σ^E Bindestellen in 19 *Streptomyces*-Genomen mittels bioinformatischer Analysen vorausgesagt und die potentiellen σ^E-Regulons für jedes der Genome bestimmt. Schließlich wurden alle Zielgene ausgewählt, die in mindestens 9 *Streptomyces*-Genomen konserviert waren, und daraus ein σ^E-Kernregulon abgeleitet.

Chapter 1

Introduction

Aims

1 Introduction

1.1 Overview of bacterial signal transduction systems

In nature, bacteria constantly encounter various environmental challenges such as fluctuating temperatures and changing nutrition conditions. In order to respond to environmental signals, bacteria employ multiple signal transduction systems to achieve the flow of information from the outside environment to the gene expression module inside the cell. The major signal transduction systems found so far are one-component system (1CS), two-component system (2CS) and extracytoplasmic function (ECF) σ factors (Staroń & Mascher, 2010). Each system is generally composed of two functional modules: a signal input module and a response output module. In the case of 1CSs, the signal input and output modules are fused to one polypeptide (Ulrich et al, 2005), whereas in the case of 2CSs, a histidine kinase and a cognate response regulator generally function as signal input and output, respectively (Casino et al, 2010; Stock et al, 2000). The system governed by ECF σ factors generally employs an anti-σ factor to work as a signal input module and an ECF σ factor to work as a response output module (Helmann, 2002; Mascher, 2013; Staroń & Mascher, 2010). All three systems mediate cellular responses to environmental signals mainly through directing the transcription of downstream genes.

1.1.1 One-component systems

As the simplest form of signal transduction system, 1CSs fuse a signal input domain and a response output domain in a single protein. The input domain senses the signal and further regulates the activity of the output domain, thereby mediating the cellular response. It has been shown that the input domains of most of 1CSs are small molecule binding domains (93%) and the output domains of 1CSs are mainly DNA-binding helix-turn-helix (HTH) domains (84%) (Ulrich et al, 2005).

1CSs are extensively spread in bacteria and involved in various physiological processes. One classical example of 1CSs is the *E. coli* LacI repressor which modulates lactose metabolism (Wilson et al, 2007). The LacI repressor modulates the transcription of the *lacZYA* operon, which contains genes *lacZ*, *lacY* and *lacA*, encoding proteins necessary for lactose utilization (Wilson et al, 2007). Other 1CSs could also act as global regulators and

modulate the transcription of over one hundred genes, such as the cAMP receptor protein, CAP (Kolb et al, 1993; Tutar, 2008).

In individual bacteria, 1CSs are generally far more abundant than 2CSs and ECF σ factors (Ulrich et al, 2005). This reflects an important role of 1CSs for bacteria to sense environmental cues. However, 1CSs have limitations in signal transduction. Although there have been 1CSs reported to locate at the membrane and to be able to sense an extracellular signal (e.g., the bacitracin regulator BcrR (Gebhard et al, 2009) and cholera toxin transcriptional activator ToxR (Miller et al, 1987)), the vast majority of 1CSs are soluble proteins and thus sense intracellular cues (Ulrich et al, 2005).

1.1.2 Two-component systems

2CSs are widely distributed in bacteria and play important roles in sensing extracellular environmental signals (Ulrich et al, 2005). The classical 2CS consists of a histidine kinase and a response regulator. The histidine kinase functions to perceive the signal and the response regulator functions to orchestrate the cellular response (Fig. 1A). The histidine kinase typically contains a sensor domain at the N-terminus, a dimerization and histidine phosphotransferase (DHp) domain in the middle and a catalytic and ATPase (CA) domain at the C-terminus. Upon detecting the signal by the sensor domain, the histidine kinase auto-phosphorylates at a conserved histidine residue locating at the DHp domain, which is catalyzed by the CA domain. The response regulator contains a conserved receiver domain (RD) that is able to catalyze the transfer of the phosphoryl group from the phosphohistidine residue of its cognate histidine kinase to a conserved aspartate residue in the RD. Phosphorylation of the response regulator generally induces its conformational change which further activates the function of its effector domain to elicit a signal output (Capra & Laub, 2012; Casino et al, 2010; Stock et al, 2000).

It should be pointed out that in some histidine kinases, there are also additional domains, which, in most common cases, are PAS (Per, Arnt, Sim) (Zhulin et al, 1997), HAMP (histidine kinases, adenyl cyclases, methyl-accepting proteins and phosphatases) (Aravind & Ponting, 1999) and GAF (cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA) (Arvind & Ponting, 1997), locating between the sensor domain and the DHp domain (Mascher et al, 2006). Such connector domains are either involved in the signal relay or shown to be the sensor for the cytoplasmic stimuli (Möglich et al, 2009; Parkinson, 2010).

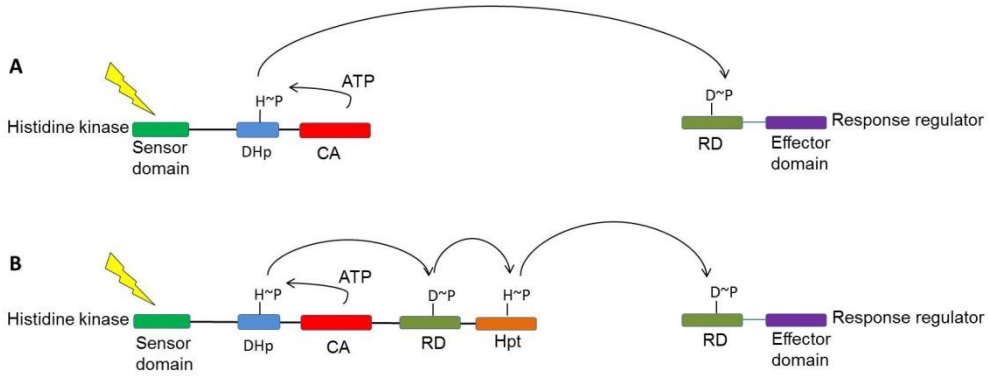


Fig. 1 Schematic overview of signal transduction mechanisms of 2CSs (see text for details). A) typical two component system; B) one example of the extended 2CSs, which involve a phosphorelay process to pass the signal. Abbreviations: DHp, dimerization and histidine phosphotransferase; CA, catalytic and ATPase, RD, receiver domain; Hpt, histidine phosphotransfer domain.

Apart from the typical 2CS paradigm described above, there also exists extended 2CSs, which require a phosphorelay process to pass the signal (Perraud et al, 1999; Zhang & Shi, 2005). Such phosphorelay processes are achieved by multiple phospho-transferring modules. For example, an additional RD and a histidine phosphotransfer (Hpt) domain could be involved in the phosphor-relay process (Fig. 1B). The additional RD and the Hpt domain are located at the C-terminal of the histidine kinase that directly follows the CA domain, which thus lead to the designation of this kinase as a "hybrid histidine kinase". Upon receiving the stimuli, this hybrid histidine kinase firstly auto-phosphorylates at the conserved histidine residue in the DHp domain. Then the phosphoryl group is sequentially transferred to the RD of the hybrid histidine kinase, followed by the histidine phosphotransferase and then is transferred to the terminal response regulator (Perraud et al, 1999).

Based on their domain architecture, histidine kinases can be classified into three groups (Fig. 2) (Mascher et al, 2006). The largest group is comprised of histidine kinases that contain at least two transmembrane helixes and a large extracellular sensor domain. This group is called extracellular- or periplasmic-sensing histidine kinases that detect signals like nutrients or solutes outside the membrane enclosed cytoplasmic parts. Many well understood kinases belong to this group (Mascher et al, 2006). For example, EnvZ from *E. coli* is involved in the response to the extracellular osmolality (Leonardo & Forst, 1996). The two well-characterized sensor histidine kinases NarX and NarQ sense environmental

nitrate and nitrite and are involved in the regulation of transcription of genes involved in anaerobic respiration (Cavicchioli et al, 1996; Stewart, 2003). The second group includes histidine kinases that sense the membrane-associated or internal membrane signal. Histidine kinases within this group contain multiple transmembrane helixes (2 to 20) and very short linkers in between (Mascher et al, 2006). Many well investigated histidine kinases such as the cell envelope stimulus sensors BceS (Dintner et al, 2011; Ohki et al, 2003), LiaS (Jordan et al, 2006; Mascher et al, 2004) and VanS (Hong et al, 2004; Hutchings et al, 2006b) belong to this group. The third group contains histidine kinases that are either cytoplasmically located or membrane anchored and sense intracellular signals. In both cases, the sensor domain of these histidine kinases is located inside the cytoplasm (Mascher et al, 2006). This group of histidine kinases is exemplified by the well understood histidine kinases, CheA from *Proteobacteria* (involved in chemotaxis) (Bilwes et al, 1999; Parkinson, 1976) and KinA from *B. subtilis* (involved in sporulation) (LeDeaux et al, 1995; Stephenson & Hoch, 2001).

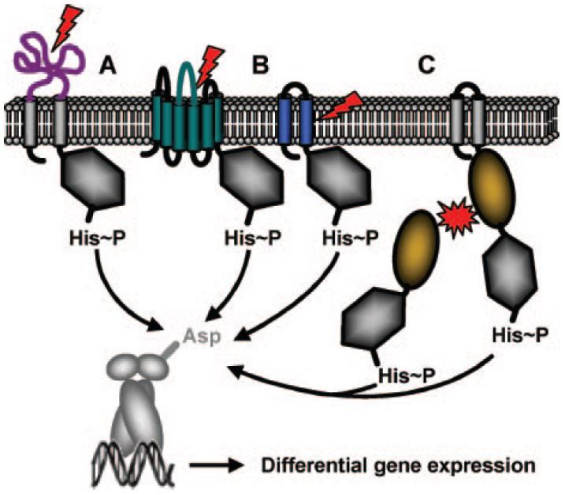


Fig.2 Model of three histidine kinases groups. A) The extracellular- or periplasmic- sensing group; B) the membrane-sensing group (sense membrane-associated or internal membrane signals) and C) the cytoplasmic-sensing group including membrane-anchored or soluble histidine kinases (see text for details). The structural parts of histidine kinase that perceive the stimulus are highlighted by different colors. The stimulus is shown in red color. The figure is taken from (Mascher et al, 2006).

Response regulators can also be classified based on domain architecture (Galperin, 2006). It is found that most response regulators (about 66%) contain a DNA binding effector domain, with NarL family and OmpR family being the major types. These response regulators should thus function as transcriptional regulators. Interestingly, 14% of response regulators only have a receiver domain and might mediate a signal output by themselves. A

considerable percentage of response regulators also contain an enzymatic, RNA-binding or protein-binding effector domain (Galperin, 2006).

Separating the input and output modules into two proteins also makes the 2CS a more flexible design in comparison with the 1CS. The signal does not necessarily pass from one single histidine kinase to one single response regulator. Instead, cross-talk or cross-regulation between histidine kinases and response regulators from different 2CSs can occur. Moreover, in some cases, multiple histidine kinases can phosphorylate one single response regulator, resulting in the signal integration. Conversely, in some cases, one single histidine kinase can phosphorylate multiple response regulators, leading to the signal amplification (Laub & Goulian, 2007).

1.1.3 ECF σ factors

Since the term ECF σ factor first appeared in 1994, these systems have been well investigated in many microorganisms (Lonetto et al, 1994). Recently, owing to the large available genome information in the database, a great number of proteins were assigned into ECF family (Staroń et al, 2009). The ECF σ factors are a group of σ_{70} family proteins, which are involved in the transcription of some specific genes, generally related to the stress response. They recognize the conserved promoter motif, typically with an "AAC" at -35 region. In many cases, they auto-regulate themselves. In addition, ECF σ factors contain two conserved σ_2 and σ_4 domains. The σ_2 and σ_4 domains fulfill the function of RNA polymerase binding and promoter recognition (Fig. 3) (Helmann, 2002; Mascher, 2013; Paget & Helmann, 2003).

Similar to the 2CS, the signal transduction system governed by ECF σ factors also employs different proteins as the signal input and output. Typically, an anti-σ factor works as a signal input module and an ECF σ factor works as a signal output module. When there is no stimulus, the anti-σ factor binds to the ECF σ factor and affects the ability of the ECF σ factor to bind the promoter motif and RNA polymerase. Upon perceiving the stimulus, the anti-σ typically undergoes a regulated proteolysis or conformational change, which in either case results in the release of the ECF σ factor (Brooks & Buchanan, 2008; Helmann, 2002; Mascher, 2013) (Fig. 3) (for detailed signaling mechanisms, see section 1.3).

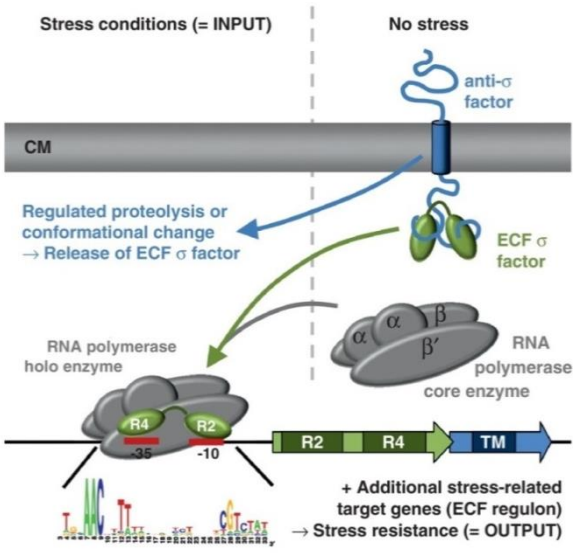


Fig. 3 Overview of typical characteristics of ECF σ factors. Green color indicates the ECF σ factor and the blue color indicates the anti-σ factor and the general signaling mechanism of ECF σ factor. The σ_2 and σ_4 domains of the ECF σ factor are indicated with R2 and R4, respectively. The -35 region and -10 region of a promoter motif are highlighted in red and a typical promoter consensus recognized by ECF σ factors is shown below. The four subunits of RNA polymerase core enzyme are represented in grey color. TM represents the transmembrane helix and CM represents the cytoplasmic membrane. This figure is taken from (Mascher, 2013).

1.2 ECF classification

In 2009, an ECF classification work was published from Thorsten Mascher's lab (Staroń et al, 2009). Approximately 2700 ECF sequences from 369 microbial genomes were firstly retrieved to build an initial ECF dataset. Then, a thorough phylogenetic analysis was carried out based on the conservation of amino acid sequence in the σ_2 and σ_4 domains. 43 major ECF groups (individually containing more than 10 ECF sequences; group numbers 01-43) (Fig. 4) and 24 minor groups (individually containing less than 10 ECF sequences; group numbers 101-124) were thus obtained. This initial classification was further supported by screening the presence or absence of a conserved genomic context, a conserved putative promoter, an anti-σ factor, an additional conserved domain or a reported common physiological function for the ECF members in each group (Staroń et al, 2009).

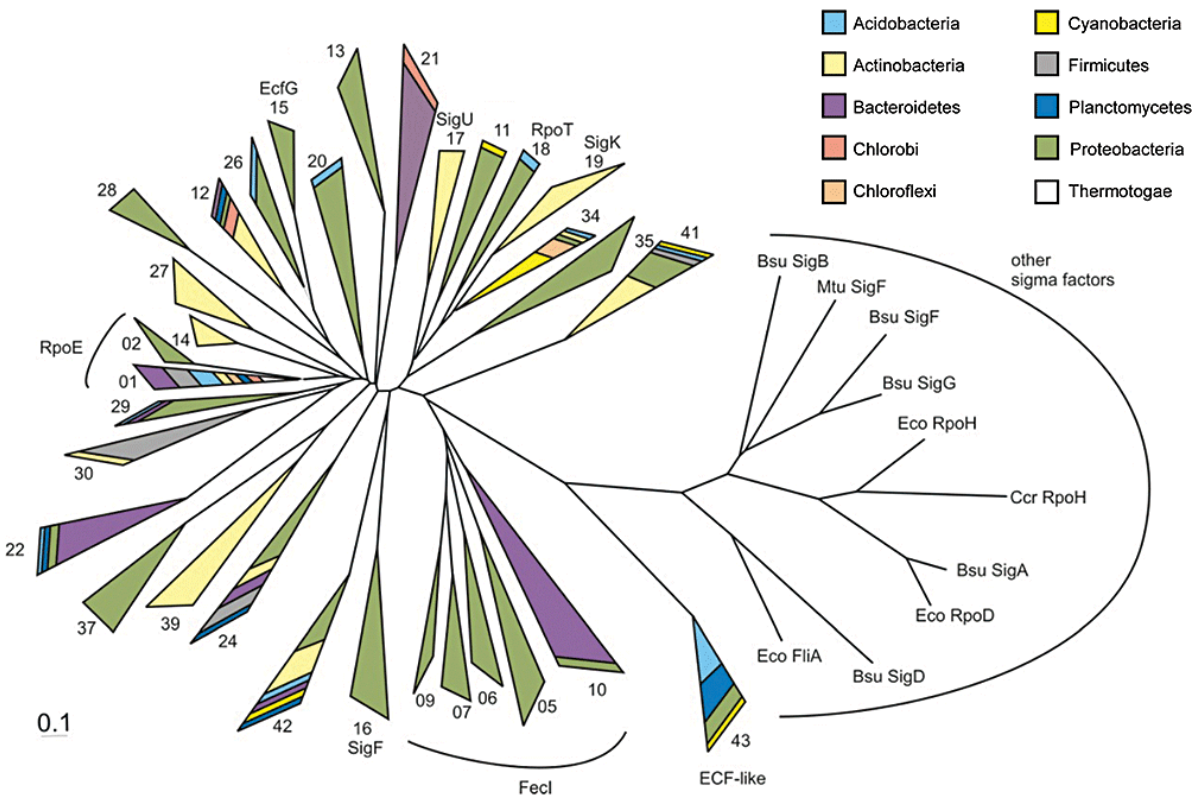


Fig. 4 Phylogenetic tree of the major ECF groups found by Staroń et al (Staroń et al, 2009). For clarity, only the 32 most important ECF groups that have at least 20 protein members in their dataset and/or conserved genomic contexts were shown. The two phylogenetically most distant representatives of each group were selected to build a tree here. A gapless multiple sequence alignment of the conserved σ_2 and σ_4 domain of the ECF σ factors was first carried out, and then the tree was generated using the Least Squares method of the Phylip (Felsenstein, 1989) programs PROTDIST and FITCH, which are run in BioEdit Sequence Alignment Editor (Hall, 1999). The phylum distribution of each group is shown by different colors. Each group is shown by a triangle and the length of the edges in each triangle reflects the phylogenetic diversity within the group. Bsu, *B. subtilis*; Mtu, *M. tuberculosis*; Eco, *E. coli*; Ccr, *C. crescentus*. Some other σ factors (they are not ECF σ factors) are also included in the tree. Typical experimentally validated ECF σ factors (FecI, SigF, RpoE, EcfG, SigU, RpoT and SigK) belonging to different ECF groups are also shown in the figure. Due to minor sequence dissimilarity in the σ_2 and σ_4 domains, group 43 is termed as an ECF-like group. This figure is taken from (Staroń et al, 2009).

Within this analysis, the well characterized RpoE-like and FecI-like σ factors locate in the groups 01-04 and groups 05-10, respectively. The ECF σ factor, RpoE from *E. coli*, which is involved in cell envelope stress response (Raivio & Silhavy, 1999; Ruiz & Silhavy, 2005), belongs to the group 02, whereas the ECF σ factor FecI from *E. coli*, which is involved in citrate-iron uptake (Angerer et al, 1995; Van Hove et al, 1990), is assigned to the group 05. Five groups (ECF11-15) are linked to soluble anti-σ factors and were thus suggested to sense the cytoplasmic located stimuli. Some groups (e.g., ECF01, ECF41 and ECF42) are distributed into different bacterial phyla, however other groups (e.g., ECF05-09 and ECF15) are phylum-specific (Fig. 4).

The classification also identified several ECF groups (e.g., ECF41 and ECF42) with unusual properties (Staroń et al, 2009). ECF41 is a group that is widely distributed in different bacterial phyla. This group of ECF σ factors is highly conserved in amino acid sequence and has an additional C-terminal domain with approximately 100 amino acids (Staroń et al, 2009). In 2012, a work published by Wecke et al. shows that truncation of the C-terminal extension of ECF41 σ factors, $Ecf41_{Bli}$ (from *Bacillus licheniformis*) and $Ecf41_{Rsp}$ (from *Rhodobacter sphaeroides*) significantly affects their activity in the transcription of the downstream target genes. While a partial truncation of their C-terminal extension results in hyperactive alleles, a complete loss of their C-terminal extension results in alleles almost without any activity. Nevertheless, failure in finding out the activating stimulus and phenotype linked to $Ecf41_{Bli}$ and $Ecf41_{Rsp}$ makes the *in vivo* role of this C-terminal extension a mystery even still (Wecke et al, 2012). Further studies such as elucidating the crystal structures of ECF41 σ factors might help to explain their detailed regulatory mechanisms.

ECF42 is another group of ECF σ factors with an additional C-terminal domain and is distributed in different bacterial phyla (Staroń et al, 2009). Compared to ECF41, it has a longer C-terminal extension with approximately 200 amino acids (Staroń et al, 2009). In a recent study, an ECF42 member, ECF10 from *Pseudomonas putida* KT2440 has been shown to be involved in antibiotic resistance and biofilm formation (Tettmann et al, 2014). However, the regulatory mechanism of ECF42 still awaits further experimental elucidation.

In 2011, a novel ECF group was proposed based on the analysis of ECF σ factors resembling a copper-dependent ECF σ factor CorE (Gómez-Santos et al, 2011) and was later termed as ECF44 by Thorsten Mascher (Mascher, 2013) . This group of ECF σ factors contains a cysteine-rich C-terminal domain with approximately 30 amino acids. They also possess the conserved genomic contexts with predicted functions involved in copper trafficking and handling (Gómez-Santos et al, 2011).

It should be pointed out although 68 ECF groups have been defined, they fail to cover the whole ECF family present in the bacterial kingdom. While the identification of the group ECF44 is based on the analysis of a number of CorE-like ECF σ factors (Gómez-Santos et al, 2011), the classification conducted by Staroń et al. is highly biased (Staroń et al, 2009). Of the 369 genomes used for the classification conducted by Staroń et al., 200 are derived from *Proteobacteria* and 124 are derived from the three phyla, *Bacteroidetes, Firmicutes*,

and *Actinobacteria*. Of the 1873 classified ECF sequences, almost 90% are derived from these four phyla, with approximately half of the ECF sequences originating from *Proteobacteria* alone (Staroń et al, 2009).

Indeed, many ECF sequences identified by the rapidly increasing number of genomes sequenced in recent years could not be assigned into any of these 68 ECF groups. In particular, in some bacteria phyla (e.g. *Planctomycetes*), whose genomes are under-represented in the 369 genomes used for the classification by Staroń et al. (Staroń et al, 2009), the vast majority of ECF σ factors could not be assigned into these defined ECF groups (see Chapter 2). Therefore, re-classification of these ECF sequences is necessary. In this thesis, the ECF σ factors from an under-represented bacterial phylum, *Planctomycetes*, and a well characterized bacterial phylum, *Actinobacteria* were chosen as two examples to be classified and analyzed in depth (see Chapter 2 and Chapter 3).

1.3 Signaling mechanisms of ECF σ factors

The seven major types of signaling mechanisms of ECF σ factors have been well summarized by a recent review (Mascher, 2013). Based on the idea in this review, here, each type of signaling mechanism and its typical examples are briefly introduced.

1.3.1 Regulated proteolysis of the anti-σ factor

The ECF group 01-04 has been suggested to be activated through regulated proteolysis of its cognate anti-σ factor (Fig. 5a) (Mascher, 2013). Regulation of σ^E (belonging to ECF02 as identified by Staroń et al. (Staroń et al, 2009)) from *E. coli* is one of the best investigated examples of this mechanism. σ^E is involved in responding to cell envelope stress such as the unfolded outer membrane proteins (OMP) (Raivio & Silhavy, 1999; Ruiz & Silhavy, 2005). The activation of σ^E necessitates the stepwise degradation of its cognate anti-σ factor RseA. RseA is a single-pass transmembrane protein. Its N-terminal domain is located in the cytoplasm where it binds to σ^E (Campbell et al, 2003; De Las Penas et al, 1997; Missiakas et al, 1997). Degradation of RseA begins with a cleavage of its C-terminal periplasmic part between residues Val^{148} and Ser^{149} by a membrane located protease DegS (Ades et al, 1999; Walsh et al, 2003). The C-terminal tails of unfolded OMP activate the activity of DegS through binding to the PDZ domain of DegS (Walsh et al, 2003; Wilken et al, 2004). After cleavage by DegS, the remaining portion of RseA further undergoes a proteolysis by a

membrane embedded protease RseP. This occurs through cleavage of the part of σ^E located in the membrane or near the membrane. This results in the release of the $RseA^{N\text{-terminal}}$-σ^E complex into the cytoplasm (Akiyama et al, 2004; Alba et al, 2002; Kanehara et al, 2002). The cytosolic proteases such as ClpXP and ClpAP further degrade the cytoplasmic part of RseA and thus release the σ^E protein (Chaba et al, 2007). Similar sequential proteolysis mechanisms have also been found in the destruction of other anti-σ factors (Hastie et al, 2013; Hastie et al, 2014; Heinrich & Wiegert, 2009). However, in addition to the proteases, other proteins could also be implicated in modulating the activity of these anti-factors. For example, for σ^E in *E. coli*, RseB protects RseA from degradation by DegS and therefore exerts negative regulatory roles on the activation of σ^E (Cezairliyan & Sauer, 2007).

1.3.2 Conformational change of the anti-σ factor

ECF σ factors can also be disassociated from their bound anti-σ factors through changing the conformation of the anti-σ factors (Fig. 5b). One of the best elucidated examples for this mechanism is the regulation of σ^E (ECF11) from *Rhodobacter sphaeroides.* σ^E is involved in singlet oxygen response in *Rhodobacter sphaeroides* (Campbell et al, 2007; Greenwell et al, 2011).

The activity of σ^E from *R. sphaeroides* is modulated by a cytoplasmic located anti-σ factor ChrR (Campbell et al, 2007; Newman et al, 1999). ChrR contains an N-terminal anti-σ domain (ASD) and a C-terminal cupin-like domain (CLD). The ChrR-ASD domain binds σ^E and sterically blocks the RNA polymerase binding determinants of the σ_2 and σ_4 domain of σ^E. It is found that the inhibitory activity of the ASD to σ^E requires the coordination of a Zn^{2+} ion by at least two conserved cysteine residues (Cys_{35}, Cys_{38}) and a histidine residue (His_6) in the ASD (Campbell et al, 2007). The CLD domain also coordinates a Zn^{2+} ion. However, it appears that the CLD domain is not involved in the inhibitory activity exerted by ChrR, but instead plays a role in the response to singlet oxygen (Campbell et al, 2007; Greenwell et al, 2011). Upon sensing singlet oxygen by ChrR, the ChrR-σ^E complex disassociates, which results in the release of σ^E allowing for the transcription of its target genes (Greenwell et al, 2011).

1.3.3 Activation of ECF σ factors by protein-protein interaction

Activation of ECF σ factors can also be accomplished through protein-protein interaction (Fig. 5c). Such a mechanism is suggested to be characteristic of ECF05-08 and ECF10 groups (Mascher, 2013) and has been elucidated in detail for the FecIR-FecA system from

E.coli. FecI functions as an ECF σ factor to participate in iron uptake in *E.coli* (Angerer et al, 1995; Lonetto et al, 1994; Van Hove et al, 1990). FecR is a single pass cytoplasmic membrane protein that binds to the σ_4 domain of FecI through its N-terminal cytoplasmic domain (Mahren et al, 2002; Welz & Braun, 1998). FecA is a TonB dependent iron transporter that is located in the outer membrane. Apart from iron transportation, it is also critical in sensing extracellular iron-citrate. After binding to iron-citrate, FecA changes its conformation (Ferguson et al, 2002; Härle et al, 1995; Kim et al, 1997; Yue et al, 2003). This signal is further transmitted to FecR through direct protein-protein interaction, which finally results in the activation of FecI (Enz et al, 2003; Enz et al, 2000; Mahren & Braun, 2003; Ochs et al, 1995) to direct the transcription of *fecABCDE* iron uptake operon (Angerer et al, 1995; Van Hove et al, 1990).

It should be pointed out that although FecR binds to FecI, FecR might not function as a typical anti-σ factor. Anti-σ factors generally inhibit the activity of their cognate ECF σ factor (Helmann, 1999; Hughes & Mathee, 1998). This is not the case for FecR. Before the signaling activation, FecR slightly inhibits FecI to bind with RNA polymerase (Mahren & Braun, 2003). However, after induction by ferric citrate, the presence of FecR is critical for the activity of FecI in the transcription of its downstream genes (Ochs et al, 1995). Nevertheless, the detailed mechanism of the positive role of FecR remains obscure. Indeed, in some homologous systems in other microorganisms, the mechanism varies. For example, in *Pseudomonas putida* WCS358, the FecR homolouge PupR does not show an activating effects towards its cognate ECF σ factor PupI (FecI homologue), yet, it displays an obvious inhibitory effect towards PupI (Koster et al, 1994).

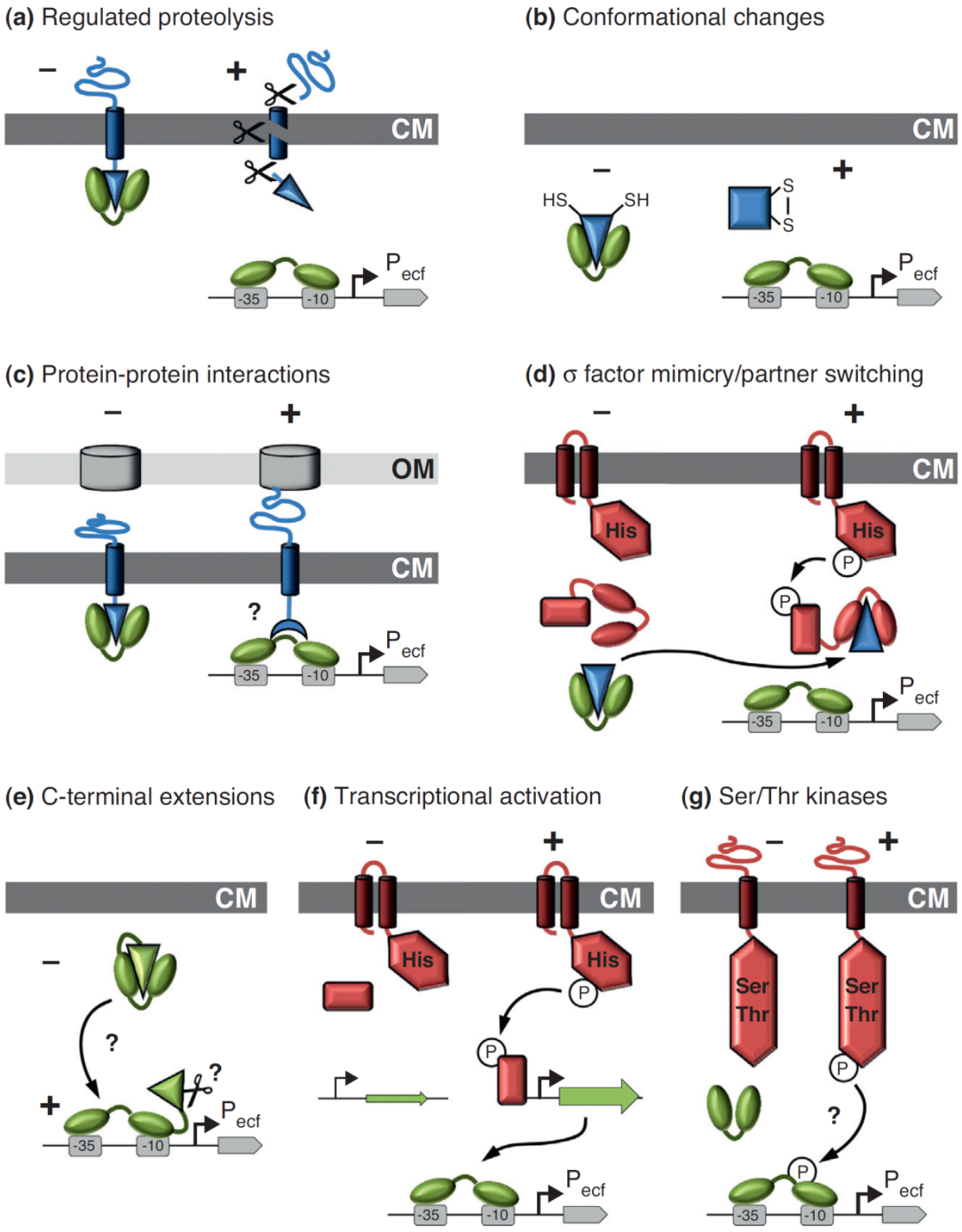


Fig. 5 Models of ECF dependent signaling mechanisms. ECF σ factors are represented in green color and anti-σ factors are represented in blue color. The histidine kinases, serine/threonine kinases and response regulators are represented in red color. "+" indicates the presence of stimulus and "-" indicates the absence of stimulus (see the text for additional details). The figure is taken from (Mascher, 2013).

1.3.4 Activation of ECF σ factors through a partner switch mechanism

The partner switch mechanism employs a third party protein to competitively bind to the anti-σ factor and thus release the ECF factor from its bound anti-σ factor (Fig. 5d), which has been described through the study of σ^{EcfG} (ECF15) from α-*Proteobacteria*. σ^{EcfG} is involved in the general stress response in α-*Proteobacteria* (Francez-Charlot et al, 2009; Francez-Charlot et al, 2015). Release of σ^{EcfG} from its bound anti-σ factor NepR requires an anti-anti σ factor PhyR (Campagne et al, 2012; Francez-Charlot et al, 2009; Herrou et al, 2012). PhyR resembles a response regulator from a 2CS, yet with considerable differences. The receiver domain of PhyR does not locate at the N-terminus of the protein, which is employed by most classical 2CS response regulators, but instead, is located at the C-terminus (Galperin, 2006; Gourion et al, 2006). Moreover, the N-terminal domain of PhyR mimics an ECF σ factor (Gourion et al, 2006; Staroń et al, 2009) and is thus called a σ factor like (SL) domain (Herrou et al, 2012). The difference between a PhyR-SL domain and an ECF σ factor is the lack of a "-10" promoter binding region in the σ_2 domain and many other conserved residues in the conserved σ_2 and σ_4 domains (Staroń et al, 2009). Upon receiving the environmental stimulus, PhyR is phosphorylated by the upstream histidine kinases and changes its conformation to liberate the SL domain from the inhibition by its C-terminal receiver domain. Furthermore, the free SL domain competes for binding to NepR, releasing the σ^{EcfG} (Campagne et al, 2012; Francez-Charlot et al, 2015; Herrou et al, 2012). It has been shown that the binding affinity between PhyR and NepR is much higher than that between σ^{EcfG} and NepR, which thus elucidates the mechanism underlying the release of σ^{EcfG} from the NepR (Campagne et al, 2012).

1.3.5 Regulation of the activity of ECF σ factors by their C-terminal extension

As already discussed in section 1.2, the activity of ECF41 σ factors could be modulated by the additional C-terminal domain (Wecke et al, 2012). Indeed, the activity of ECF44 member, CorE is also controlled by its C-terminal domain (Gómez-Santos et al, 2011). Although it is currently unclear whether other proteins are also involved in regulating the activity of these ECF σ factors, the already known findings suggest the C-terminal extension is an important regulatory module for the activities of these ECF σ factors. Nevertheless, more experimental analysis is necessary to investigate how ECF σ factors with extended domains bind to the RNA polymerase and recognize the promoter motif.

1.3.6 Transcriptional activation of ECF σ factors

ECF σ factors do not necessarily use an anti-σ factor as the signal input. Alternatively, they can cooperate with a 2CS to mediate signal transduction (Fig. 5f), as is well exemplified by the σ^E (ECF39) from *S. coelicolor* (Fig. 6).

σ^E is induced by a wide range of cell envelope disruptors and is important for maintaining the cell envelope integrity (Hong et al, 2002; Paget et al, 1999a). The *ΔsigE* mutant shows approximately 50 times increased sensitivity to lysozyme and altered muropeptide profile compared to the wild type. Moreover, the *ΔsigE* mutant sporulates poorly and over-produces the pigment, actinorhodin on the medium deficient in Mg^{2+} (Paget et al, 1999a).

Contrary to many ECF factors that auto-regulate their own transcription, the transcription of the *sigE* gene necessitates an induced expression of a 2CS CseBC upon envelope stimulus perception (Hong et al, 2002; Paget et al, 1999b). The *sigE*, *cseB* and *cseC* genes together with another gene, *cseA* are located in the same operon (Paget et al, 1999b). It is suggested that upon receiving the environmental stimulus, the histidine kinase CseC auto-phosphorylates itself and transfers the phosphoryl group to its cognate response regulator CseB. The phosphorylated CseB is then activated and further directs the transcription of the *sigE* operon (Fig. 6) (Hong et al, 2002; Paget et al, 1999b). It appears that the transcription of the *sigE* operon is solely dependent on CseB and that about 90% of the transcripts stop at the terminal of the *sigE* gene (Hong et al, 2002; Paget et al, 1999b). The activation of the expression of σ^E elicits a signal output and is suggested to direct the transcription of the envelope-stress-responsive genes (Fig. 6) (Hong et al, 2002; Hutchings et al, 2006a).

It should be pointed out here that CseA is a lipoprotein (Hutchings et al, 2006a), but the detailed function is unknown. The *ΔcseA* mutant shows an increased transcription of *sigE* (Hutchings et al, 2006a). It is suggested that CseA might directly interact with CseC and thus modulates the activity of this system. Alternatively, loss of CseA might destabilize the cell envelope, which thus leads to the activation of CseBC (Hutchings et al, 2006a). But, these hypotheses still await further experimental validation.

Two *in vivo* targets have been known to be controlled by σ^E before. One is *hrdD*, encoding an alternative house-keeping σ factor (Paget et al, 1999a). The other one is the *cwg* operon, which encodes proteins predicted to be involved in the synthesis of cell wall glycan (Hong

et al, 2002). But no clear phenotype has been found for either *hrdD* mutant (Buttner et al, 1990) or *cwg* mutant (Hong et al, 2002). Therefore, the complete functional roles of σ^E in cell envelope response were unknown. In this study, a large number of genes were identified to be controlled by σ^E, which provided a detailed biological understanding of the regulation governed by σ^E (see Chapter 4).

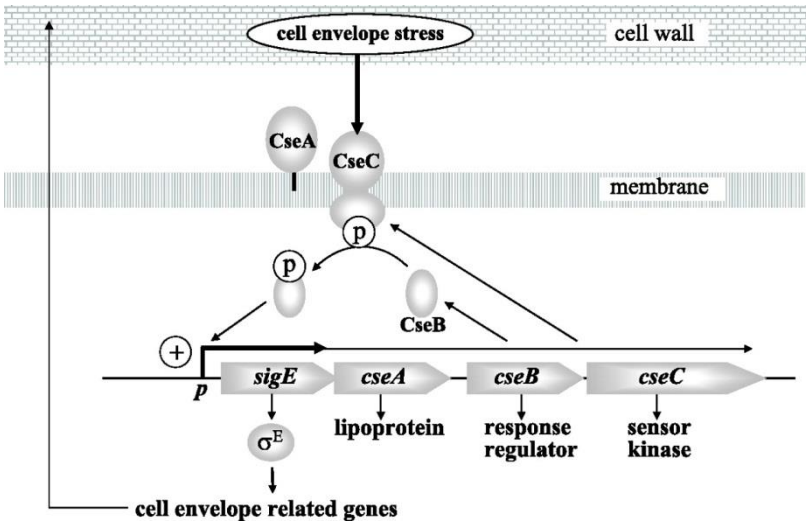


Fig. 6 Model of the σ^E dependent cell envelope stress response in *S. coelicolor* (see the text for details). This figure is taken from (Hutchings et al, 2006a).

1.3.7 Activation of ECF σ factors by Ser/Thr Kinases

Although there was no experimental evidence supporting that Ser/Thr kinases can modulate the activity of ECF σ factors, the conserved co-existence of Ser/Thr kinases and ECF σ factors belonging to group 43 on the same genomic context (Staroń et al, 2009) strongly suggests a regulatory relationship between them. Bacterial Ser/Thr kinases have been shown to be involved in many physiological processes including cell development (Nádvorník et al, 1999; Nariya & Inouye, 2005), cell division (Beilharz et al, 2012; Ruggiero et al, 2012), cell virulence (Papavinasasundaram et al, 2005; Wiley et al, 2006) and cell metabolism (Atsushi et al, 1994; Cowley et al, 2004). Typically, they exert function through phosphorylation of their downstream regulators or functional proteins (Dworkin, 2015; Pereira et al, 2011). It is hypothesized that Ser/Thr kinases might contribute to ECF σ factor-dependent signal transduction through integrating or sensing the upstream signal and activating the ECF σ factor by phosphorylation (Fig. 5g) (Mascher, 2013). Nevertheless, further experimental evidence is necessary to demonstrate this hypothesis.

1.4 General features of *Planctomycetes* and *Actinobacteria* phyla

Since our thesis focuses on the classification and functional characterization of ECF σ factors from *Planctomycetes* and *Actinobacteria*, the general features of these two phyla of bacteria will be briefly introduced in the following paragraphs.

1.4.1 *Planctomycetes*

Planctomycetes are widely distributed in different environmental niches, e.g., soil (Buckley et al, 2006), wetland (Kulichevskaya et al, 2015; Kulichevskaya et al, 2007), fresh water (Bondoso et al, 2011) and ocean (Shu & Jiao, 2008). They are quite different from many common bacterial phyla such as *Firmicutes*, *Proteobacteria* and *Actinobacteria*, and show many unusual features. For example, rather than using binary fission to divide, almost all the *Planctomycetes* bacteria use a budding process to reproduce themselves (Franzmann & Skerman, 1984; Fuerst & Sagulenko, 2011; Lee et al, 2009; Tekniepe et al, 1981). *Planctomycetes* also contain complex endomembrane systems (Acehan et al, 2014; Lindsay et al, 1997; Lindsay et al, 2001; Sagulenko et al, 2014; Santarella-Mellwig et al, 2013). Historically, this endomembrane system was suggested to form compartmentalized cell regions: paryphoplasm (a region between a cytoplasmic membrane and the intra-cytoplasmic membrane (ICM)) and pirellulosome (or riboplasm; a ribosome-rich part, which is surrounded by ICM and contains the condensed DNA) (Fuerst & Sagulenko, 2011; Lindsay et al, 1997; Lindsay et al, 2001). For the planctomycete *Gemmata obscuriglobus*, it was also suggested that a nuclear-like structure was additionally formed inside the pirellulosome, in which double membranes surround the chromosome (Lindsay et al, 2001). However, more recent studies challenge (Acehan et al, 2014; Santarella-Mellwig et al, 2013) or support (Sagulenko et al, 2014) the idea of intracellular compartmentalization in *Gemmata obscuriglobus*. As a challenging view, the complex endomembrane network is suggested to be formed by the invaginations of the cytoplasmic membrane and the cytoplasm is all interconnected. Therefore, the membrane organization of *G. obscuriglobus* is likely to be an extension of the typical Gram-negative bacterial membrane system (Acehan et al, 2014; Santarella-Mellwig et al, 2013). Another interesting feature of the *Planctomycetes* is that they can produce sterols that are typically synthesized by eukaryotes (Pearson et al, 2003). Additionally, the planctomycete *G. obscuriglobus* has been shown to

be capable of taking up proteins from the environment in a way that is similar to "endocytosis" in eukaryotes (Lonhienne et al, 2010).

1.4.2 *Actinobacteria*

Actinobacteria is comprised of gram-positive bacteria with high G+C content in their DNA. In particular, some members' DNA has a G+C content that exceeds 70% (Bentley et al, 2002; Ikeda et al, 2003). The bacteria belonging to this phylum can grow in multiple environmental niches. For example, *Streptomyces* has been found in the soil (Bontemps et al, 2013), sea water (Zhu et al, 2011), marine sediments (Veyisoglu & Sahin, 2014), bamboo litter (Lee & Whang, 2014), and intestinal tract of a small terrestrial crustacean, *Armadillidium vulgare* (Shibazaki et al, 2011). *Mycobacteria* are largely isolated from clinical samples (Abbadi et al, 2009; Ramos et al, 2013; Shinnick & Good, 1994). *Bifidobacteria* mainly inhabit the gastrointestinal tract of animals, (e.g., mammals and birds) (Turroni et al, 2014; Turroni et al, 2011).

Actinobacteria has been well characterized and many bacteria belonging to this phylum show remarkable properties. Some *Actinobacteria* are pathogenic. For example, *Mycobacterium tuberculosis* and *Mycobacterium leprae* are pathogens of two well-known diseases, tuberculosis and leprosy, respectively (Cook et al, 2009; Sasaki et al, 2001). *Corynebacterium diphtheriae* causes the disease diphtheria (Hadfield et al, 2000). *Streptomyces scabies* is the causative agent of the potato common scab disease (Lambert & Loria, 1989; St-Onge et al, 2008; Wanner, 2006). However, some *Actinobacteria* are of great biotechnological importance (Ventura et al, 2007). For example, approximately two thirds of the commercial antibiotics are derived from *Streptomyces spp.* (de Lima Procópio et al, 2012). *Bifidobacteria* are widely used as probiotics. *Bifidobacteria* benefit the health of the host in many aspects such as showing an antagonist effect towards intestinal pathogens, modulating the host immune system and degrading complex carbohydrates (Picard et al, 2005; Turroni et al, 2014; Turroni et al, 2011). *Corynebacterium glutamicum* has been widely used as an industrial workhorse for the production of L-amino acids. In particular, it is used to produce the flavor enhancer L-glutamate and the food additive L-lysine in a very large scale (over 1 million tons per year). In recent years, this bacterium has also been used as a genetic platform to produce more compounds using technologies of metabolic engineering and synthetic biology (Becker & Wittmann, 2012; Woo & Park, 2014).

1.4.2.1 The genus *Streptomyces*

The *Streptomyces spp.* is one of the best investigated bacterial genus in *Actinobacteria*. *Streptomyces* are of great interest for research because of their two important features: 1) they produce a variety of secondary metabolites including antibiotics and many other bioactive molecules (Běhal, 2000; de Lima Procópio et al, 2012; Ōmura, 1992); 2) they show a distinct morphological differentiation (Chandra & Chater, 2014; Elliot et al, 2007; Flärdh & Buttner, 2009). *Streptomyces* morphologically resemble the filamentous-fungi and show complex multi-cellular lifestyle (Elliot et al, 2007; Flärdh, 2003).

The life cycle of *Streptomyces* (Fig. 7) starts from the germination of a free spore. One or two germ tubes emerge from the spore and grow into the substrate, they then further elongate by tip extension and branching to form a substrate mycelium. As the growth proceeds and the nutrients become depleted in the substrate mycelium, a complex signal cascade triggers the coating of parts of mycelial filaments with a hydrophobic sheath, which enable them to break the surface tension and grow into the air to form aerial hyphae. The individual aerial hyphae further grow to become a long unbranched cell generally containing over 50 copies of genomes. Subsequently, prespore compartments are formed by division along the length of the aerial hyphae. Dividing aerial prespore compartments further undergo morphological and metabolic changes such as remodeling and thickening the cell wall, condensation of the chromosomes and production of a spore associated pigment to develop into mature spores (Angert, 2005; Chandra & Chater, 2014; Flärdh & Buttner, 2009).

Three species in this genus, *Streptomyces coelicolor* A3(2), *Streptomyces griseus* and *Streptomyces venezuelae* have been selected as model organisms to investigate the genetics and cell biology of *Streptomyces* (Chandra & Chater, 2014). *S. coelicolor* A3(2) is the most widely used *Streptomyces* strain in the laboratory (Kieser et al, 2000). *S. griseus* is the producer of the aminoglycoside antibiotic, streptomycin (Mansouri & Piepersberg, 1991; Schatz et al, 1944). *S. venezuelae* is the first bacterium found to produce the antibiotic, chloramphenicol (Ehrlich et al, 1947; Ehrlich et al, 1948). Contrary to *S. coelicolor* A3 (2) that only sporulates on solid medium, this organism is able to sporulate thoroughly and rapidly in the liquid culture. This property makes it being a good workhorse for the investigation of gene expression synchronous to the cell growth in the submerged culture (Bibb et al, 2012; Flärdh & Buttner, 2009).

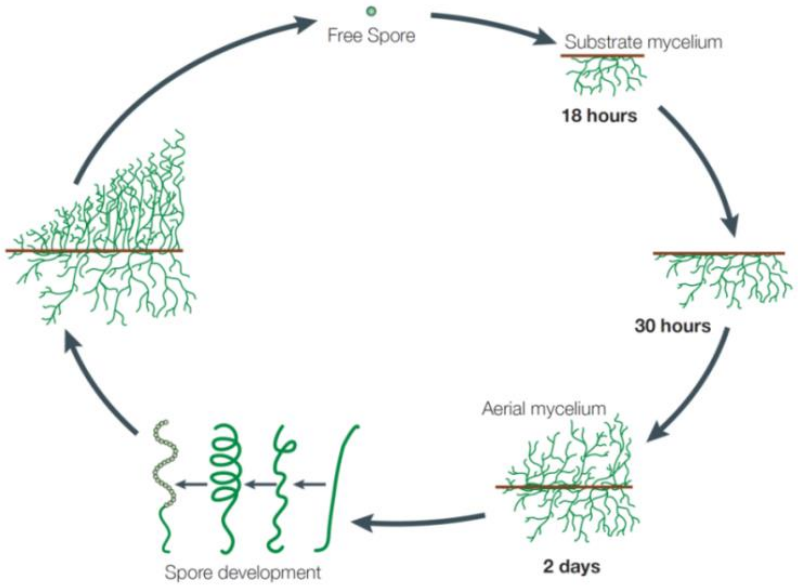


Fig. 7 *Streptomyces coelicolor* life cycle (See the text for details). This figure is taken from (Angert, 2005).

So far, a number of *Streptomyces* genomes have been sequenced, which provide much information about the molecular mechanisms underlying the morphological development, signal transduction, secondary metabolism in this genus of bacteria (Bentley et al, 2002; Cruz-Morales et al, 2013; Ikeda et al, 2003; Medema et al, 2010; Ohnishi et al, 2008). The *S. coelicolor* genome has a size of approximately 8.7 Mb and contains 7,825 predicted genes. These include over 20 gene clusters which may be involved in the synthesis of secondary metabolites. The genome also encodes a large number of regulatory proteins (12.3% of all the encoded proteins) including over 60 σ factors, over 80 histidine kinases and 79 response regulators. The origin of replication (oriC) and almost all the predicted essential genes (encoding proteins related to DNA replication, cell division, and transcription, translation and amino-acid biosynthesis) lie in the center of the linear chromosome of *S. coelicolor*. In contrast, genes encoding proteins with a putative non-essential function (e.g. hydrolytic exoenzymes and those involved in the secondary metabolites) locate at the arms of the chromosome (Bentley et al, 2002). The genome of *Streptomyces griseus* IFO 13350 has a size of about 8.6 Mb and contains 7138 protein open reading frames (Ohnishi et al, 2008), whereas the genome of *Streptomyces venezuelae* ATCC 10712 (GenBank Accession No. NC_018750) has a genome size of 8.2 Mb and encodes 7238 proteins.

Through multiple genetic studies of the model streptomycetes, several tens of key genes involved in the morphological development have been identified (Elliot et al, 2007; Flärdh & Buttner, 2009). Many of these genes are designated with "*bld"* (including *bldA*, *bldB*, *bldC*, *bldD*, *bldG*, *bldH*, *bldJ*, *bldK*, *bldL*, *bldM* and *bldN)* or "*whi"* (including *whiA*, *whiB*, *whiD*, *whiE*, *whiG*, *whiH*, *whiI*, *whiJ*, *whiL*, *whiM* and *whiO*) (Elliot et al, 2007). The *bld* mutants fail to form aerial mycelium under many culture conditions, resulting in a bald appearance of the colony (Champness, 1988; Merrick, 1976). The *whi* mutants could form aerial mycelium, but are not able to sporulate efficiently. These mutants could not produce spore pigment in their aerial mycelium and keep white appearance on the medium (Chater, 1972; Hopwood et al, 1970). Many of these *bld* and *whi* genes have been well characterized. For example, *whiA* encodes a transcriptional regulator, which modulates the transcription of over 200 targets, including those encode proteins involved in cell division, chromosome segregation and aerial growth (Bush et al, 2013). *bldM* encodes a 2CS-typed response regulator, which controlls the transcription of over 100 target genes, including a number of key development-related genes (Al-Bassam et al, 2014). Mutants of some other genes that are not designated with a *"bld"* or *"whi"* also show a similar phenotype with *bld* and *whi* mutants (Elliot et al, 2007; Flärdh, 2003). One example of these genes is *ramS*, which encodes a SapB precursor. SapB is a lantibiotic-like peptide, which functions as an important biosurfactant to facilitate the growth of the aerial hyphae from the aqueous phase to the air (Flärdh & Buttner, 2009; Kodani et al, 2004).

A number of signal transduction systems have been well characterized in *Streptomyces* (Hutchings et al, 2004; Liu et al, 2013; Paget et al, 2002). These include three ECF σ factors, σ^E, σ^R and σ^{BldN}. As already introduced in the section 1.3.6, σ^E is involved in the cell envelope stress response (Hong et al, 2002; Paget et al, 1999a). The transcription of σ^E is modulated by a two component system, CseBC and presumably a lipoprotein, CseA as well (Hong et al, 2002; Hutchings et al, 2006a; Paget et al, 1999b). σ^R has been shown to be involved in the thiol-oxidative stress response in *S. coelicolor* (Paget et al, 1998). Similar to σ^E from *R. sphaeroides* (Campbell et al, 2007; Greenwell et al, 2011) (see section 1.3.2), the activity of σ^R from *S.coelicolor* is modulated by its cognate anti-σ factor RsrA through a redox dependent conformational change (Bae et al, 2004; Kang et al, 1999; Paget et al, 2001). The σ^R regulon consists of over 100 genes, with a great proportion of them encoding proteins involved in thiol redox homeostasis and protein quality control (Kim et al, 2012). σ^{BldN} is encoded by the gene *bldN* and plays an important role in the formation of aerial

mycelium (Bibb et al, 2000). The activity of σ^{BldN} is suggested to be modulated by its cognate anti-σ factor, RsbN (Bibb et al, 2012). The regulon controlled by σ^{BldN} in *S.venezuale* includes *rsbN, bldM*, and the genes encode chaplin (ChpB, C, E, F, G and H) and rodlin proteins (RdlA, B and C) (Bibb et al, 2012). In *Streptomyces*, the chaplin proteins and rodlin proteins together constitute the main components of the hydrophobic sheath coating aerial mycelium and spores (Flärdh & Buttner, 2009).

1.5 Aims of this study

The core aim of this study is to classify and functionally characterize ECF σ factors from *Planctomycetes* and *Actinobacteria*, although other signal transduction systems (1CSs and 2CSs) are also included in the analysis. Both bioinformatic methods and genome-wide experimental analysis would then be used.

In Chapter 2 and Chapter 3, we sought to classify and analyze the ECF σ factors from eight *Planctomycetes* genomes and 119 actinobacterial genomes. Since many ECF σ factors found from these genomes could not be assigned into any of the previously identified 68 ECF groups, we postulated that it is likely to identify novel ECF groups from these ECF σ factors. Therefore, the sequences properties of these ECF σ factors related to building a novel ECF group should be analyzed. These include the conservation of their amino acid sequences, genomic contexts, domain architectures and possible recognized target promoter motifs. Since the activity of ECF σ factors were generally regulated by their cognate anti-σ factors (Brooks & Buchanan, 2008; Hughes & Mathee, 1998; Mascher, 2013), the genomic context analysis should give a special focus on the identification of the cognate anti-σ factors of these previously unclassified ECF σ factors. In order to unravel the ECF-dependent regulation in *Planctomycetes* and *Actinobacteria*, it would be ncessary to find cues related to signal transduction mechansims and physiological funciton of these previously unclassified ECF σ factors from the idenitifed conserved sequence properties. The distribution and abundance of different ECF groups (including those groups identified from both the previous study and this study) in these two phyla of bacteria should also be analyzed.

In Chapter 4, we aimed at defining the regulon controlled by σ^{E}, an ECF σ factor that is a key regulator of the cell envelope stress response in *S. coelicolor*. The first aim of this part

of study was to define the σ^E regulon using chromatin immunoprecipitation-sequencing (ChIP-seq), DNA microarray and thorough bioinformatic analysis in *S.coelicolor*. In order to validate the targets, biochemical (e.g. S1 mapping and *in vitro* transcription) and genetic experiments (generating mutants of the targets) should be done. The second aim of this part of study was to predict the σ^E regulons from 19 *Streptomyces* genomes and establish a σ^E-core regulon. The σ^E-core regulon should consist of the conserved targets controlled by σ^E in *Streptomyces*. The final aim of this part of study was to unravel the key mechanisms underlying the cell envelope stress response governed by σ^E through the analysis of the function of the σ^E targets in depth.

Chapter 2

Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in *Planctomycetes* by comparative genomics

Jogler C, Waldmann J, **Huang X**, Jogler M, Glöckner FO, Mascher T, Kolter R (2012) *J Bacteriol* **194:** 6419-6430

JB
Journals.ASM.org

Identification of Proteins Likely To Be Involved in Morphogenesis, Cell Division, and Signal Transduction in *Planctomycetes* by Comparative Genomics

Christian Jogler,[a,b] Jost Waldmann,[c] Xiaoluo Huang,[d] Mareike Jogler,[b] Frank Oliver Glöckner,[c] Thorsten Mascher,[d] and Roberto Kolter[a]

Harvard Medical School, Boston, Massachusetts, USA[a]; DSMZ, Braunschweig, Germany[b]; MPI for Marine Microbiology, and Jacobs University , Bremen, Germany[c]; and Ludwig Maximilians University, Munich, Germany[d]

Members of the *Planctomycetes* clade share many unusual features for bacteria. Their cytoplasm contains membrane-bound compartments, they lack peptidoglycan and FtsZ, they divide by polar budding, and they are capable of endocytosis. Planctomycete genomes have remained enigmatic, generally being quite large (up to 9 Mb), and on average, 55% of their predicted proteins are of unknown function. Importantly, proteins related to the unusual traits of *Planctomycetes* remain largely unknown. Thus, we embarked on bioinformatic analyses of these genomes in an effort to predict proteins that are likely to be involved in compartmentalization, cell division, and signal transduction. We used three complementary strategies. First, we defined the *Planctomycetes* core genome and subtracted genes of well-studied model organisms. Second, we analyzed the gene content and synteny of morphogenesis and cell division genes and combined both methods using a "guilt-by-association" approach. Third, we identified signal transduction systems as well as sigma factors. These analyses provide a manageable list of candidate genes for future genetic studies and provide evidence for complex signaling in the *Planctomycetes* akin to that observed for bacteria with complex life-styles, such as *Myxococcus xanthus*.

In many respects, the *Planctomycetes* are extremely unusual bacteria (12). There is still controversy among evolutionary biologists regarding the phylogeny of *Planctomycetes*. They have been proposed to be the deepest-branching bacterial phylum (4, 23) or the most rapidly evolving bacteria (62). Their remote relatedness to *Chlamydia* (55, 61) has led to a broadly accepted hypothesis that there is a superphylum consisting of the *Planctomycetes*, *Verrucomicrobia*, and *Chlamydia* (PVC superphylum) (46, 59). Regardless of their exact phylogeny, all *Planctomycetes* share a very unusual cell plan that differs dramatically from those of Gram-positive and Gram-negative bacteria (13, 28). They lack peptidoglycan (PG); instead, they have a proteinaceous cell wall, a fact which renders them resistant to β-lactam antibiotics (6, 26, 27). Underneath the proteinaceous cell wall lies the cytoplasmic membrane (CM). Strikingly, the cytoplasm of planctomycetes is divided by an intracytoplasmic membrane (ICM) into the paryphoplasm and the pirellulosome or riboplasm (28). The DNA is highly condensed and forms a nucleoid that occupies only a fraction of the pirellulosome (12). Interestingly, in *Gemmata obscuriglobus*, the nucleoid is surrounded by two closely opposed double membranes, thus resembling the enveloped nucleus of a eukaryotic cell (12, 13). However, in contrast to eukaryotes, this nucleus-like structure contains ribosomes. However, since most ribosomes are situated within the pirellulosome, many mRNAs generated in the nucleus-like compartment of *G. obscuriglobus* must somehow be transported through both double membranes to be translated. However, nothing is known about such transfer mechanisms and whether they are related to eukaryotic mRNA transport (12). Planctomycetes are also unusual in that their membranes contain sterols (42). Furthermore, *Planctomycetes* and some other members of the PVC superphylum are the only organisms among bacteria and archaea that encode proteins with a high level of structural similarity to eukaryotic membrane coat (MC) proteins, which play a major role in the formation of coated vesicles. The planctomycetal MC-like proteins have β-propeller domains followed by stacked pairs of α-helices (SPAHs), as do their eukaryotic counterparts. In addition, a significant amount of the MC-like proteins are localized in close proximity to the ICM or to vesicles within the paryphoplasm (48). Such vesicles have been shown to allow the endocytosis-like uptake of proteins into the paryphoplasm of *G. obscuriglobus* cells (29), a trait long believed to be a hallmark of eukaryotic cells (21). In contrast to all other bacteria, the polar planctomycetes divide via budding, in an FtsZ-independent manner. As a consequence, their cell division more resembles that of budding yeast than that of the typical bacterium, which relies on FtsZ. Thus, planctomycetes challenge the prokaryote-eukaryote dichotomy in several ways (10).

In the face of these unusual features, one might expect that planctomycetal genomes might contain many *Archaea*-like or *Eukarya*-like genes. However, this seems not to be the case (1, 11, 55). Thus, we embarked on a comparative genomics study of planctomycetes with the aim of identifying proteins likely to be related to their unusual biology. The results of these analyses now provide candidate proteins for performing subsequent genetic analyses. Given that genetic tools are available only for *Planctomyces limnophilus*, we focused mainly on this model organism.

Received 24 July 2012 **Accepted** 14 September 2012
Published ahead of print 21 September 2012
Address correspondence to Christian Jogler, christian@jogler.de.
Supplemental material for this article may be found at http://jb.asm.org/.

doi:10.1128/JB.01325-12

TABLE 1 Selected features of sequenced planctomycetal genomes except for anammox bacteria[a]

Organism	Chromosome size (Mb)	Sequence type	GenBank accession no.	Plasmid size (kb)	CDS	No. of genes	% G+C content	% of PCGs with function	No. of tRNAs	No. of rRNA genes	Reference or sources
Rhodopirellula baltica SH1	7.1	Complete	NC_005027.1		7,403	7,325	55.40	34.22	76	3	17
Planctomyces limnophilus	5.4	Complete	NC_014148.1	37	4,313 (+60)	4,198 (+60)	53.7 (57.0)	ND	61	4	26a
Pirellula staleyi	6.2	Complete	NC_013720.1		4,825	4,717	57.50	ND	46	3	6a
Isosphaera pallida	5.5	Complete	NC_014962.1	56	3,791 (+32)	3,690 (+32)	62.4 (67.0)	59.33	48	9	17a
Planctomyces brasiliensis	6.0	Draft	NZ_AEIC00000000[b]		4,865	4,769	56	57.55	45	6	JGI, NCBI
Gemmata obscuriglobus	9.1	Draft	NZ_AEIC00000000	ND	8,080	7,989	67.20	39.48	85	6	JGI, NCBI
Planctomyces maris	7.8	Draft	NZ_ABCE00000000	ND	6,530	6,480	50.50	40.66	48	1	JGI, NCBI
Blastopirellula marina	6.7	Draft	NZ_AANZ00000000	ND	6,079	6,025	57.00	39.64	48	6	JGI, NCBI

[a] PCG, protein-coding genes; ND, not determined; CDS, coding sequences. Numbers in parentheses provide data for inclusion of plasmids.
[b] Downloaded from TrEMBL.

MATERIALS AND METHODS

Phylogenetic analyses of *Planctomycetes*. Aligned 16S rRNA gene sequences were derived from the SILVA database (release 108 [http://www.arb-silva.de/download/arb-files]) (47), and the alignment was corrected manually. Phylogenetic trees were calculated by employing the ARB software package (31). The RAxML module (rate distribution model GTRGAMMA; rapid bootstrap analysis algorithm) performed the maximum likelihood analysis, while neighbor-joining trees were calculated with the ARB Neighbor Joining tool using Felsenstein correction. Bootstrap values were determined by using 1,000-fold resampling (31). GenBank accession numbers of the sequences used are listed in Table S1 in the supplemental material.

Distilling the planctomycetal core genome. First, an NCBI BLAST database ("core contributors") containing all available protein sequences of *Planctomycetes* was generated (Table 1), preserving their taxonomic information. Second, an all-against-all BLASTP search was used to extract all homologues within the data set. To exclude matches to the originating genome, the mask option (negative_gilist) of BLAST was used. If a protein matched the same database entry multiple times, only the best BLAST match was considered. Hits with a query coverage higher than 60% and below the E value cutoff of 1e−5 were taken into account. Additionally, the number of species in the associated list of BLAST matches was counted for each query protein. Queries not matching all taxa in the database, except the originating one, were discarded. Subsequently, sequences of queries fulfilling the selection criteria were compiled into a list, and all genes among the list linked by reciprocal BLAST hits were clustered. If a BLAST match did not link back to the query, the corresponding sequence was rejected. Consequently, each cluster represents a group of homologous genes, with a high probability of sharing the same function.

Trimming the planctomycetal core genome using *in silico* subtraction. We compiled a set of genes from well-studied model organisms (*Escherichia coli* strain K-12 substrain MG1655 [GenBank accession number U00096] and *Bacillus subtilis* subsp. *subtilis* strain 168 [accession number AL009126]), and a second BLAST database ("blacklist") was generated out of those. All genes of the planctomycete core genome were compared against this blacklist database by using BLAST. Genes reaching the E value cutoff of 1e−5 and a minimum coverage of 65% led to the deletion of the entire cluster from the planctomycetal core genome set. Corresponding Perl scripts are available upon request. A list of all protein sequences and the resulting clusters is available in Table S2 in the supplemental material.

Analysis of the *Planctomyces limnophilus* core genome. Since *P. limnophilus* is the only planctomycete that can be genetically manipulated, we focused on its core genome in order to identify candidate proteins for subsequent genetic experiments. Consequently, all sequences except those with a *P. limnophilus* origin were deleted from the 114 core genome clusters obtained after the *in silico* subtraction of *E. coli* and *B. subtilis* genomes. The resulting clusters were manually inspected, and proteins likely to be involved in cell division, shape determination, or compartmentalization were selected as starting points for a "guilt-by-association" approach (see Table S3 in the supplemental material). The genomic context of each selected gene was manually inspected to identify putative operons (POs) harboring at least two members of the core genome. Five putative operons fulfilled these criteria and were manually analyzed by using BLAST, the NCBI database, and the Conserved Domain Database (CDD) for functional annotation (36).

Cell shape and cell division proteins. By screening the literature, we composed a list of 64 proteins involved in cell division, filament formation, and cell shape determination from various bacteria, archaea, and yeasts (see Table S4 in the supplemental material). The query protein was compared against the NCBI database by using BLAST with default settings and "*Planctomycetes*" as the taxon filter. Putative homologues were reverse analyzed against the database. If the reverse BLAST experiment revealed proteins from the same family of the original query, they are highlighted in green in Table S4 in the supplemental material; if not, they are highlighted in yellow and are referred to as "-like" proteins. Protein identifications, E values, and percent identity values are given in Table S4 in the supplemental material.

Gene synteny and content of *dcw* operon genes among *Planctomycetes*. Based on previous studies, we used planctomycete proteins homologous to those of *E. coli* encoded within the *dcw* operon as a starting point (see Table S4 in the supplemental material) (39, 46). Subsequently, we manually determined the synteny of the corresponding genes and visualized the gene order (see Fig. 4), while domains were determined as described above.

Analysis of planctomycetal signal transduction. The Microbial Signal Transduction (MiST2) database (http://mistdb.com/) was employed to gain an overview of regulatory proteins encoded in the eight planctomycetal genomes (57). Extracytoplasmic function sigma factor (ECF) sequences were retrieved from the MiST database and cross-checked by an ECF finder analysis (http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder/), and all positive sequences were added to a comprehensive ECF data set that contained a total of 362 sequences (see Table S5 in the supplemental material). Subsequently, the protein sequences were aligned by using the ClustalW algorithm at the European Bioinformatics Institute website (http://www.ebi.ac.uk/). After the manual removal of gaps and N- or C-terminal extensions, phylogenetic distances were calculated for the resulting multiple-sequence-alignment files of the ECF core proteins by employing the Fitch-Margoliash and least-squares distance algorithms (9), as implemented in the BioEdit sequence alignment editor (18). Individual branches of the tree were subjected to detailed follow-up analyses, as described previously (53), to identify potential anti-σ factors and/or additional associated proteins. These analyses are based on genomic context conservation as a measure of a potential functional link between their respective gene products.

RESULTS AND DISCUSSION

The planctomycetal core genome. The *Planctomycetes* clade displays enormous phylogenetic depth, and its phylogeny is still con-

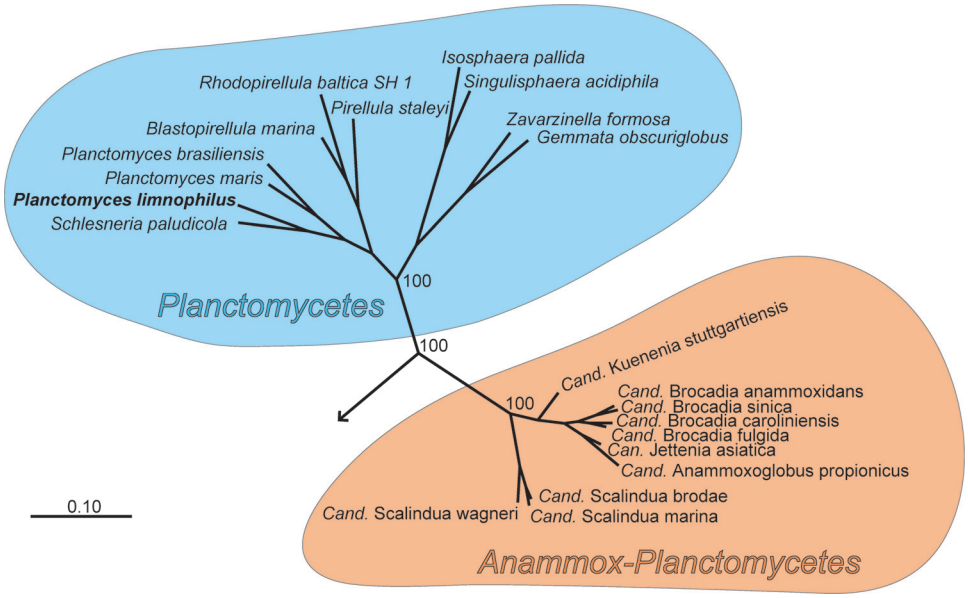


FIG 1 The class *Planctomycetia* is split into two distinct orders. Shown is a maximum likelihood phylogenetic tree of planctomycetal 16S rRNA gene sequences. Bootstrap values are shown only for deep branches, and different *Escherichia coli* sequences were used as an outgroup, represented by an arrow (GenBank accession numbers of the *E. coli* sequences used are GU594315, GU594305, GU594306, GU594304, GU594302, and GU594316). "*Ca*. Kuenenia stuttgartiensis," "*Candidatus* Kuenenia stuttgartiensis."

troversial (23, 46, 59). Thus, prior to determining the planctomycete core genome, we used 16S rRNA gene analysis to decide which species to include in our study for the practical purpose of determining a group of organisms sufficiently related to each other to produce a meaningful core genome. The maximum likelihood 16S rRNA gene tree shows, with excellent bootstrap support (bootstrap value of 100), two distinct branches. One branch was comprised of anammox planctomycetes, while the other branch contained all other planctomycetal species cultivated thus far (Fig. 1). This result provided us with a phylogenetic criterion to leave anammox bacteria out of our analysis. Within this work, we thus differentiate between "*Planctomycetes*" and "anammox bacteria." However, whether or not this has taxonomic implications is beyond the scope of this study.

Despite phylogenetic controversies, *Planctomycetes* are members of the domain *Bacteria* and, as such, would be expected to divide by using FtsZ. However, FtsZ is absent in planctomycetal genomes (2). In order to identify proteins that might be involved in cell division, shape determination, and compartmentalization, we performed two independent lines of analysis. Our planctomycetal core genome approach led to 114 predicted protein clusters containing 2,908 proteins from all eight analyzed planctomycetes after the *in silico* subtraction of *E. coli* and *B. subtilis* genomes (Fig. 2). Each cluster contained between 8 and 764 proteins, with an average of 25 proteins. This subtraction yielded important insights, since planctomycetes are difficult to cultivate, and genetic tools are available only for *P. limnophilus*. Since *P. limnophilus* is the only member of this phylum that has been genetically modified, we focused on its core genome for subsequent analyses. *P. limnophilus* contributed 274 proteins to the 114 core clusters, resulting in an average of about 2 proteins per cluster. However, even in this model, planctomycete mutant construction is rather slow. Thus, we wanted to further reduce the number of candidate genes for subsequent mutational studies. Consequently, all *P. lim-*

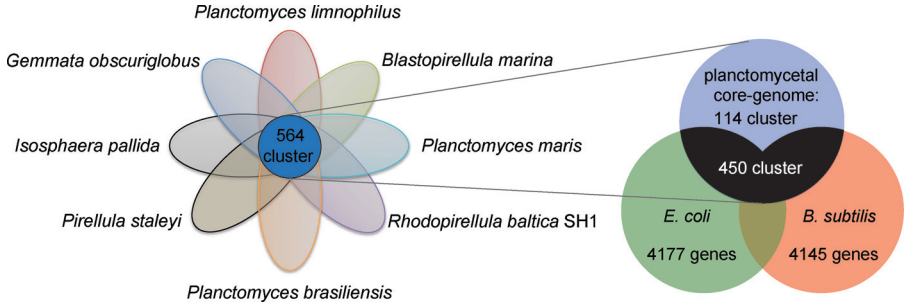


FIG 2 The planctomycetal core genome. The planctomycetal core genome was determined by comparing all eight sequenced *Planctomycetes* species against each other by using the BLAST algorithm with a coverage of >60% and an E value cutoff of 1e−5 as parameters. The Venn diagram visualizes the 564 clusters fulfilling these criteria. After *in silico* subtraction, using proteins encoded by *Escherichia coli* (GenBank accession number U00096) and *Bacillus subtilis* (accession number AL009126), 450 planctomycetal core genome clusters were eliminated, and 114 planctomycete-specific clusters remained. Genes within these 114 clusters fulfill two criteria: they are conserved among planctomycetes and absent in the genomes of the two most intensively studied model organisms, *E. coli* and *B. subtilis*.

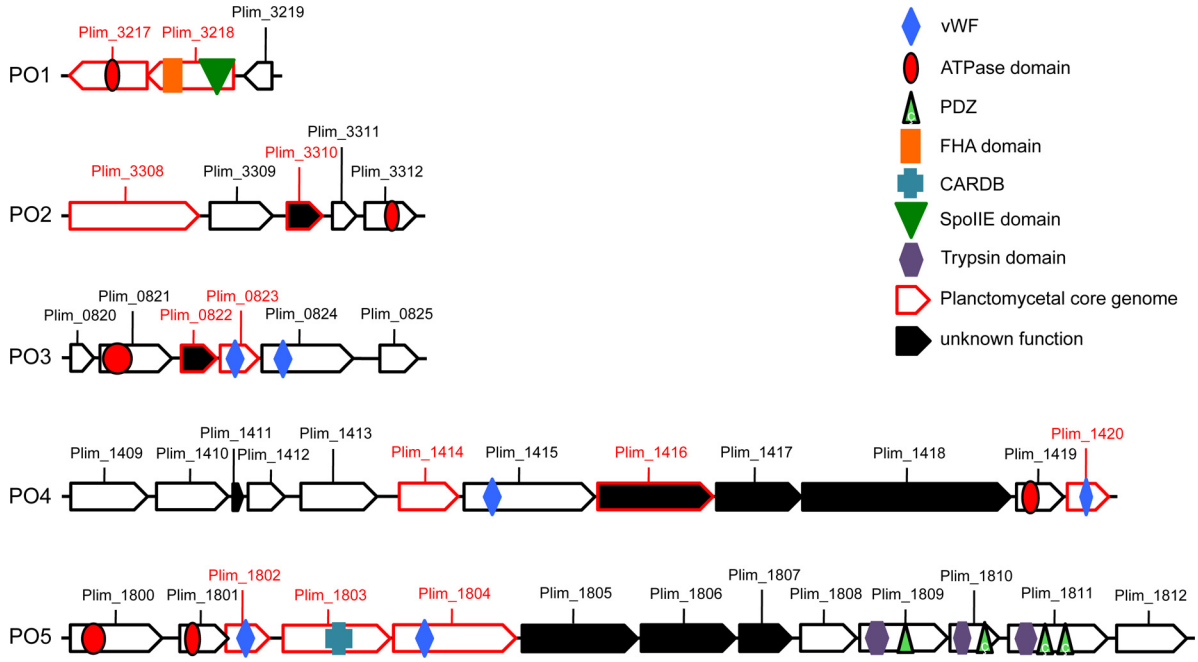


FIG 3 Selected putative *P. limnophilus* operons. Putative *P. limnophilus* operon 1 (PO1) to PO5 are shown, which fulfill two criteria: (i) they contain one gene that was evaluated as being putatively involved in cell division or compartmentalization based on a manual inspection of the *P. limnophilus* core genome, and (ii) they harbor at least one more core genome member. Core genome-related genes are shown as red outlined arrows. Identifiers for all genes are given, and selected putative domains are highlighted (vWF, von Willebrand factor; FHS, Forkhead associated).

nophilus core proteins were manually evaluated, and six clusters were selected for in-depth analyses based on their putative cell biological relevance (see Table S3 in the supplemental material). One of these clusters, cluster 23, contained 39 MC-like proteins encoded by all eight planctomycetal genomes under study, and 5 proteins belong to *P. limnophilus*. Despite the presence of MC-like proteins, no "smoking gun" was found, and given that most proteins are of unknown function, we followed a "guilt-by-association" approach to identify putative targets. First, we manually inspected the neighborhood within the six selected *P. limnophilus* core genome clusters. Second, we focused on genomic regions in which at least two core genes were situated within a putative operon. Five such POs fulfilled these criteria (PO1 to PO5) (Fig. 3), while MC-like proteins were not among them. However, the five putative operons encode proteins containing putative cell division-related domains (e.g., von Willebrand factor [vWF], PDZ, and ATPase domains). Such domains have been found to be associated with putative novel cell division mechanisms in *Archaea* (34, 35). Individually, each PO has features that render it attractive for further study.

In PO1, we identified forkhead-associated (FHA), SpoIIE, and LpxE domains. Such domains might be involved in cell division and membrane organization, as FHA domains play a critical role in the hyphal fusion of *Neurospora crassa*, while SpoIIE is required for asymmetric cell division during sporulation in *Bacillus* (37). LpxE domains are known to selectively dephosphorylate lipid A molecules, which are present in the outer membranes of Gram-negative bacteria (20). This finding is unexpected, since planctomycetes lack such an outer membrane. However, further genes required for the biosynthesis of lipid A were found in *Rhodopirellula baltica*, indicating that planctomycetes might produce lipid A (17, 24).

PO2 encodes membrane-bound dehydrogenase and heme-binding domains. This configuration is frequently found in *Planctomycetes* and *Verrucomicrobia* but is not found in any other bacteria sequenced thus far. Furthermore, dihydrolipoamide dehydrogenase and glycine cleavage H protein domains were identified. Most strikingly, some similarity to the MinD superfamily "P-loop ATPase ferredoxin" domains were detected. MinD participates in the spatial regulation of the formation of the cytokinetic FtsZ ring and thus has a major function during cell division (41).

The third identified operon, PO3, encodes proteins which appear to be related to signal transduction. Two of them contain von Willebrand factor type A domains. Thus, they might be involved in processes such as membrane modeling, cellular differentiation, and adhesion. Genes encoding proteins of unknown function, such as Plim_0822, might be responsible for planctomycete-specific traits, and their being near the "usual suspects," such as von Willebrand factor domains, makes them interesting targets for subsequent genetic experiments.

PO4 encodes 12 proteins. While two might be related to sugar metabolism, four are hypothetical or contain domains of unknown function. Thus, mainly the vWF and ATPase domains point toward the involvement of this putative operon in cell division or membrane remodeling. Interestingly, Plim_1419 might resemble a gas vesicle-related protein, GvpN. However, *gvpN* deletion mutants of *Halobacterium salinarium* revealed that GvpN is not essential for gas vesicle assembly but may enhance their synthesis in some way (40).

PO5 is the largest putative operon, and its component proteins contain several ATPase domains. In addition, PDZ domains that may mediate protein-protein interactions are found together with trypsin domains, indicating putative proteases. Furthermore, a

TABLE 2 Cell division- and cytoskeleton-related genes of planctomycetes[a]

Protein	Presence of encoding gene in organism							
	Rhodopirellula baltica SH1	*Planctomyces limnophilus*	*Pirellula staleyi*	*Isosphaera pallida*	*Planctomyces brasiliensis*	*Gemmata obscuriglobus*	*Planctomyces maris*	*Blastopirellula marina*
ClpX	+	+	+	+	+	+	+	+
ClpP	+	+	+	+	+	+	+	+
CpaE	+	+	+	+	+	+	+	+
Ddl	+	+	+	+	+	+	+	−
EnvA	+	+	+	+	+	+	+	+
FtsE	+	+	+	+	+	+	+	+
FtsI	−	+	−	−	+	−	+	−
FtsK	+	+	+	+	+	+	+	+
FtsW	−	+	−	−	+	−	+	−
RodA	−	+	−	−	+	−	+	−
MraW	+	+	+	+	+	+	+	+
MraY	+	+	+	+	+	+	+	+
MraZ	−	+	−	−	+	−	+	+
MreB	−	+	−	−	+	−	+	+
MurC	−	+	−	−	+	−	+	−
MurD	−	+	−	−	+	−	+	−
MurE	+	+	+	−	+	−	+	+
MurF	−	+	−	−	+	−	+	−
MurG	−	+	−	−	+	−	+	−
ParA	+	+	+	+	+	+	+	+

[a] This table summarizes data shown in Table S4 in the supplemental material; only those proteins that are encoded by at least one planctomycetal genome are shown. Proteins are grouped by function, and boldface type indicates *dcw* operon-related proteins.

cell adhesion domain found in bacteria (CARDB) was identified. In addition, the fact that three core genome members (Plim_1802 to Plim_1804) are next to each other makes PO5 a promising candidate for future investigations. For a more detailed description of the five putative operons, see the supplemental material.

As expected, the comparative genomic approach revealed several target proteins from *P. limnophilus* for future genetic analyses. Interestingly, while they are part of the core genome, MC-like-protein-encoding genes show a patchy distribution in the *P. limnophilus* genome and do not cluster with other core genome members. However, such proteins were recently demonstrated to be involved in the endocytosis of proteins in *G. obscuriglobus* and are thus, without a doubt, worth studying in detail (29, 48).

Cell division and cell shape determination. All eight planctomycete genomes used in this study were screened for genes encoding one of 64 cell morphogenesis- or division-related proteins that were described for different bacteria, archaea, or yeasts, and the results are summarized in Table S4 in the supplemental material. Table 2 provides a selection of proteins that were found in at least one planctomycete. As shape-determining proteins, we included representatives from all bacterial actin-like families, ParM, FtsA, AlfA, Alp7, and MreB (49). The gene encoding the bacterial actin homologue MreB was found in *Planctomyces* species and in *Blastopirellula marina*. However, we failed to identify any bacterial tubulin (Btub)-like protein using BtubA and BtubB as well as FtsZ as queries (32, 45). Recently, a new family of FtsZ-like proteins (FtsZl-1) was described, and *B. marina*, *G. obscuriglobus*, *Pirellula staleyi*, and *R. baltica* were shown to encode such a protein (34). We failed to detect homologues of such FtsZl-1 proteins in all other sequenced planctomycetes using almost the same approach as that which led to the discovery of FtsZl-1 (34). In addition, FtsZl-1 proteins lack the T6-H7 structure and thus might not be able to form filaments at all (34). Thus, FtsZl-1 might very well fulfill a function other than Z-ring formation during cell division in planctomycetes. In addition, we screened for proteins involved in archaeal cell division. Two FtsZ-independent division mechanisms are known for the domain *Archaea* (35). One is homologous to the eukaryotic ESCRT-III system and requires Snf and VPS4 proteins. While we could not detect the presence of Snf-encoding genes in planctomycetal genomes using BLAST (see Table S4 in the supplemental material), VPS4 revealed several putative homologues. However, a critical diagnostic feature, the amino-terminal microtubule interaction and transport (MIT) domain, present in all archaeal VPS4 proteins, is missing in their putative planctomycetal counterparts. Thus, we conclude that planctomycetes lack VPS4. The third archaeal division mechanism is still puzzling but seems to involve actin-like proteins. As mentioned above, by our search approach, MreB homologues could be found in only four planctomycetes and can thus not explain a universal mechanism of planctomycetal cell division. Finally, we did not detect homologues of the protein putatively involved in Z-ring formation in anammox bacteria, kustd_1438 (58). The absence of this protein among planctomycetes other than anammox bacteria further supports our initial assumption to focus on one part of the planctomycetal order, as two independent cell division mechanisms might further support the differences between both lineages. Since the cytokinesis of planctomycetes parallels the division of yeast, we screened for septin proteins known to be crucial for septum formation in *Saccharomyces cerevisiae* (3, 43). However, we did not identify any homologues among planctomycetes using the basic BLAST algorithm.

In contrast to putative Z-ring-forming proteins, FtsW was found in *Planctomyces* species, while genes for other cell division-related proteins, such as FtsE, FtsK, ParA, CpaE, EnvA, MraW, and MraY, were detected in all planctomycetal genomes. However, MraZ was found only in *Planctomyces* and *Blastopirellula*

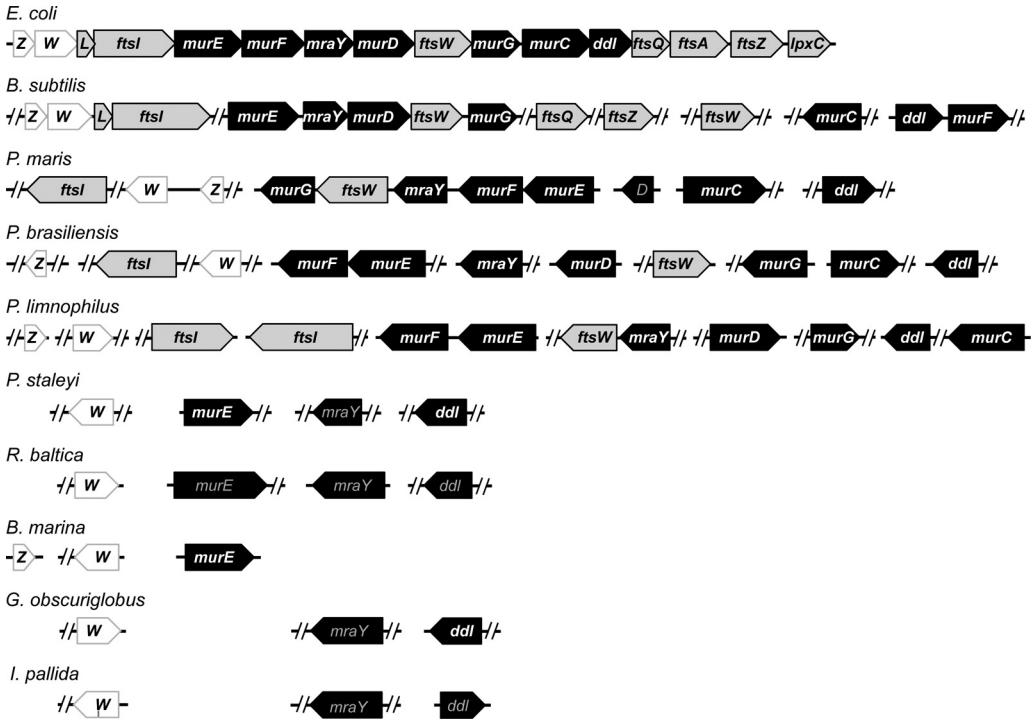


FIG 4 Content and synteny of *dcw* operon genes among planctomycetes. The cell division-related (gray) and peptidoglycan synthesis-related (black) genes arranged in the *dcw* operons of *E. coli* and *B. subtilis* are shown in comparison to homologous genes from planctomycetal genomes. Genes of unknown function are shown in white. Genes with weak similarity to the *E. coli* query are labeled in gray (W, *mraW*; Z, *mraZ*; L, *ftsL*).

species. Besides cell division, "peptidoglycan-related" proteins such as MurC, MurD, MurE, MurF, and MurG as well as Ddl and FtsI showed a more complex distribution pattern. While all these proteins are encoded by *Planctomyces* species, all except MurE and Ddl are absent in all other planctomycetes. Besides *Planctomyces*, MurE was also found in *B. marina*, *R. baltica*, and *P. staleyi*, while Ddl was absent in *B. marina* (Table 2).

Beyond gene content, we investigated the gene synteny of cell morphogenesis and division genes. Here our starting point was the *dcw* operon structure of *E. coli* as a comparison. The *dcw* operon is highly conserved among "rod-shaped" bacteria, and "genomic channeling" has been proposed to force the colocalization of genes involved in cell division and peptidoglycan synthesis to optimize septum formation (39). While genomic channeling nicely explained the persistence of the *dcw* operon over a broad taxonomic range (39) (Fig. 4), this mechanism is supposed to be restricted to rod-shaped cells, and transitions to other morphologies seem to involve genomic rearrangements and a loss of gene order (54). This was supported by the first investigation of planctomycetal *dcw* gene clusters (46). Despite the few planctomycetal genomes available at that time, the conclusion was drawn that the last planctomycetal ancestor (LPA) contained a complete *dcw* operon and that gene content and order were gradually lost (46). This hypothesis is supported by our data shown in Table 2 and Fig. 4: all three *Planctomyces* species contain most *dcw* operon genes found in all planctomycetes. However, gene synteny is less conserved among *Planctomyces* than gene content. While *Planctomyces maris* has four putative operons and one single *murD*-like gene, the members of the *dcw* operon of *E. coli* are scattered among eight different putative operons and one solitary *murD* gene in *Planctomyces brasiliensis*. In *P. limnophilus*, the situation is even more complex, as *dcw* genes are localized across nine different putative operons. Among all other planctomycetes, *P. staleyi* and *R. baltica* contain four *dcw* operon genes, while *B. marina*, *G. obscuriglobus*, and *Isosphaera pallida* harbor three, none of which colocalize. Thus, gene synteny is less conserved than gene content in regard to planctomycetal *dcw* genes, while the gradual loss becomes more obvious, since our study used more than double the number of planctomycetal genomes used in previous analyses (46). Interestingly, gene content and synteny reflect the 16S rRNA gene tree phylogeny (Fig. 1). However, the presence of peptidoglycan (PG)-related genes is difficult to explain, as planctomycetes are known to lack a PG cell wall (12), and genes without function tend to erode quickly. For PG-wall-less *Chlamydia*, cell division has been shown to be altered by beta-lactam antibiotics. Consequently, speculations of an involvement of PG in Z-ring replacement arose (2). However, consistent with previous studies, we failed to detect any growth inhibition caused by beta-lactam antibiotics such as ampicillin among planctomycetes thus far (6, 22, 26).

Taken together, our finding that several cell division- and peptidoglycan-associated proteins known from bacteria are encoded by some, if not all, *Planctomycetes* (Table 2) might suggest that their last ancestor might have divided in an FtsZ-dependent manner and was confined by a PG layer.

Putative cell division-related operons. Following a guilt-by-association approach, we screened putative planctomycetal *dcw* gene-containing operons for the presence of core genome members. The rationale driving this analysis is the assumption that cell division-related *dcw* operon genes might be found near genes involved in the FtsZ-independent division of planctomycetes. Genes

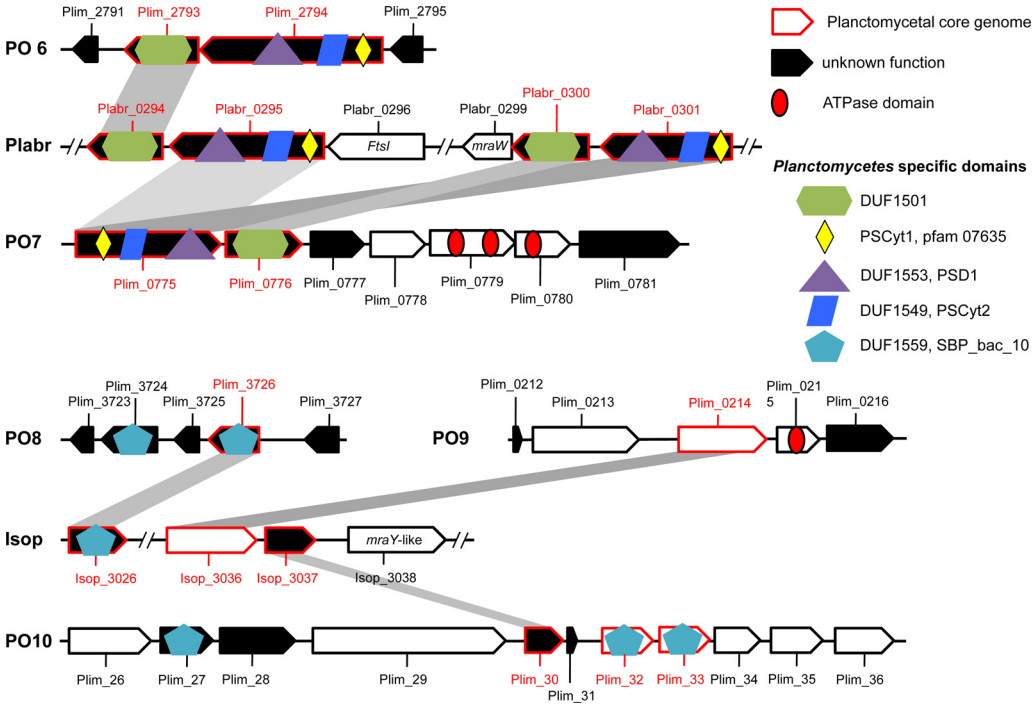


FIG 5 Putative cell division-related planctomycetal operons. The putative Plabr and Isop operons fulfilled two criteria: (i) they contain a cell division- or peptidoglycan synthesis-related *dcw* gene, and (ii) at least one additional core genome member is present within the same putative operon. In a second round, corresponding *P. limnophilus* genes were identified, which turned out to be localized on 5 different putative operons (PO6 to PO10). Selected protein domains are highlighted, while core genome members are boxed in red.

encoding FtsZ-substituting proteins most likely arose in the LPA and would consequently be absent from genomes of other bacteria such as *E. coli* or *B. subtilis*. Such planctomycetal cell division genes are likely to be conserved among all planctomycetes and to be part of their core genome. However, only two putative operons fulfilled our screening criteria. One was found within the genome of *P. brasiliensis*, spanning 25 kb and containing 14 genes, 2 of which are *dcw* operon related (*ftsI* and *mraW*) and 4 of which are members of the planctomycetal core genome (Plabr_0294, Plabr_0295, Plabr_0300, and Plabr_0301) (Fig. 5). The second putative operon found in the genome of *I. pallida* had a size of more than 38 kb and contained 26 genes. Besides the *marY*-like *dcw* gene, three planctomycetal core genome members, Isop_3026, Isop_3036, and Isop_3037, were identified (Fig. 5). However, since both organisms are not genetically manipulable, we performed a second screening to compare the identified core genome members putatively related to cell division from *I. pallida* and *P. brasiliensis* operons against the genome of *P. limnophilus*. This screening revealed two and three putative *P. limnophilus* operons, respectively (Fig. 5; for a detailed description of individual genes and proteins, see the supplemental material).

Many proteins encoded by the genes of PO6 to PO10 contained domains putatively related to cell division, such as ATPase domains, or to chromosome segregation domains. In addition, many genes contain domains of unknown function that are found exclusively in *Planctomycetes* or *Verrucomicrobia* (Fig. 5). Most strikingly, three of four putative operons identified by the guilt-by-association approach turned out to contain two or three genes that are part of the planctomycetal core genome. In PO6 and PO7, two core genome members with identical domain architectures colocalize (Plim_2793 and Plim_2794, and Plim_0775 and Plim_0776, respectively). Interestingly, the same domain architecture and localization pattern were observed in two areas of a *P. brasiliensis* putative operon, where the first two genes, Plabr_0294 and Plabr_0295, colocalize with *ftsI* and the other two, Plabr_0300 and Plabr_0301, colocalize with *mraW*. This finding makes *P. limnophilus* PO6 and PO7 interesting candidates for future genetic studies.

All 10 identified putative operons (Fig. 3 and 5) provide some indications that their gene products are involved in cell shape determination, membrane organization, chromosome segregation, or cell division. The planctomycetal proteins of unknown function, containing known domains or not, might be the most interesting ones, as new biological functions might await elucidation. The rationale for our study was the *in silico* identification of putative targets for subsequent genetic experiments, and all the putative operons identified need future experimental verification to reveal their putative relevance in planctomycetal key traits. Wet laboratory experiments must now be carried out to determine whether colocalized genes are cotranscribed or not. Experimentally verified operons from the gene clusters identified in this study should then be subjected to subsequent mutagenesis in *P. limnophilus*. Such experimental approaches should help to ultimately identify the function of genes involved in planctomycete-specific traits.

Planctomycetal signal transduction and gene regulation. Complex life-styles and cellular differentiation cycles necessitate the presence of equally complex signaling cascades and regulatory networks to orchestrate and coordinate the corresponding massive spatiotemporal modulations of gene expression patterns, as

TABLE 3 Overview of and statistics on signal-transducing and regulatory proteins encoded by *Planctomycetes* genomes[a]

Organism	Genome size (Mb)	Total no. of signal-transducing and regulatory proteins	No. of proteins										
			1CSs		2CSs				Chemotaxis			ECFs	
			Total	STPKs	HKs	HHKs	RRs[b]	HRRs	Total[c]	MCPs	CheA	Total	TM-ECFs
Blastopirellula marina	6.65	374	192	40	25	16	52 (1)	7	24	9	4	49	0
Gemmata obscuriglobus	9.16	635	288	107	41	26	83 (4)	15	60	29	5	115	63
Isosphaera pallida	5.53	220	132	61	15	14	31 (1)	3	7	2	2	14	0
Pirellula staleyi	6.2	285	153	32	22	6	45 (1)	2	20	4	4	33	0
Planctomyces brasiliensis	6.01	279	155	30	24	14	42 (1)	5	5	1	1	28	0
Planctomyces limnophilus	5.46	232	132	28	20	8	34 (1)	4	6	2	1	23	0
Planctomyces maris	7.78	376	220	42	22	18	47 (1)	3	8	1	1	52	0
Rhodopirellula baltica	7.15	344	190	55	29	11	56 (2)	6	2	0	0	48	2

[a] Data were extracted from the MiST2 database. Abbreviations: HKs, histidine kinases; HHKs, hybrid histidine kinases; RRs, response regulators; HRRs, hybrid response regulators; STPKs, Ser/Thr protein kinases; TM, transmembrane.

[b] Numbers in parentheses indicate the distribution of planctomycete-specific RRs illustrated in Fig. 6A.

[c] The total number of chemotaxis proteins does not include CheY-like RRs, which cannot be distinguished from standard RRs based on their domain architectures. They are therefore included in the numbers of RRs of 2CSs.

has been observed for model organisms such as *B. subtilis*, *Caulobacter crescentus*, and *Streptomyces coelicolor* (25, 30, 38). An understanding of the regulatory and signaling potential embedded in the genome sequences of *Planctomycetes* species is therefore a prerequisite to the unraveling of differentiation processes in this phylum. We therefore decided to closely examine the repertoire of signal-transducing proteins of the planctomycete species.

Based on mechanistic principles and phylogenetic relationships, three fundamental mechanisms of bacterial signal transduction can be distinguished: one-component systems (1CSs), two-component systems (2CSs), and extracytoplasmic function sigma factors (ECFs) (52, 53, 56). Based on data from well-represented bacterial phyla such as the *Proteobacteria* or the *Firmicutes*, an "average" genome harbors about four times more 1CSs than 2CSs and four times more of the latter than ECFs. However, while the proportions of 1CSs and 2CSs are relatively constant for most phyla (56), the numbers of ECFs seem to be much more variable among different species (52). With very few exceptions, the absolute numbers of signaling proteins correlate directly with the size of the genome (56). The distribution of planctomycetal signaling proteins supports both the correlation between genome size and the number of signaling proteins as well as the ratio between 1CSs and 2CSs (Table 3). In contrast, all planctomycetes are particularly rich in ECFs, which are equal in number to or sometimes even outnumber the 2CSs, indicating an important role of these proteins in the regulatory processes of this phylum.

1CSs. 1CSs are proteins that connect an input with an output on a single polypeptide chain. The majority (about 85%) of 1CSs from most bacteria harbor a DNA-binding output domain, as is the case for paradigms such as LacI or TetR. The output of the remaining 15% is mostly realized using protein kinases/phosphatases or regulators involved in the making and breaking of secondary messengers, such as adenylate/diguanylate cyclases (56). In planctomycetes, the percentage of DNA-binding 1CSs is much lower, while 20 to 45% of all 1CSs are Ser/Thr protein kinases (STPKs) (Table 3). Such proteins, which are common among eukaryotes, normally play a minor role in bacterial signal transduction, and their physiological role has only very recently been elucidated for a few examples, most of which are involved in differentiation or developmental processes (44). The reason for the very different abundances of STPKs between prokaryotes and eukaryotes might reside in the fact that in bacteria, with their chromosome located in the cytoplasm, most adaptation and differentiation processes are directly regulated via differential transcription. For this, bacteria use predominantly sensory histidine kinases (HKs) as part of 2CSs to activate their cognate response regulators (RRs), which, once activated, function mostly as DNA-binding proteins. In contrast, the eukaryotic chromosomes are spatially separated from the cytoplasm in the nucleus. Hence, direct protein modifications (including phosphorylations) play a much more important role in cellular physiology and differentiation. Thus, eukaryotes require larger numbers of protein kinases to modulate the protein phosphorylation pattern in response to diverse stimuli. Considering that the chromosome of, e.g., *Gemmata obscuriglobus* is surrounded by two nuclear membranes, the high abundance of STPKs might point toward a similar role of protein phosphorylation in the differentiation processes of planctomycetes (12). However, this does not imply any necessary evolutionary homology but implies merely a convergence of signal transduction system characteristics due to analogies in cell organization.

2CSs. 2CSs consist of a sensor protein, the HK, which, in the presence of a suitable stimulus, autophosphorylates. Subsequently, a cognate partner protein, the RR, becomes active through a phosphoryl group transfer from the histidine kinase (HisKA) domain of the HK to the receiver domain of the RR (15). This phosphorylation then activates the output domain of the RR, which, in most cases (about 85%), binds DNA to orchestrate the differential expression of its target genes (14, 56). Usually, the two partner proteins are encoded by neighboring genes that are organized in an operon. In planctomycetal genomes, the absolute number of 2CS proteins correlates well with genome sizes, and the ratio of 1CSs to 2CSs is comparable to those in other species (Table 3). However, RRs outnumber HKs, and only 50 to 70% of the RRs are DNA-binding proteins (bacterial average, 85%). Most planctomycetal RRs consist of a receiver domain only and lack obvious output domains. Many planctomycetal 2CS sensor proteins represent complex (hybrid) HKs, containing multiple putative input domains and often more than one HisKA domain. Hybrid HKs are part of complex phosphorelays, involving the integration/modulation of the HK activity itself or the activity of downstream RRs (7). The majority (about 70 to 80%) of HKs and

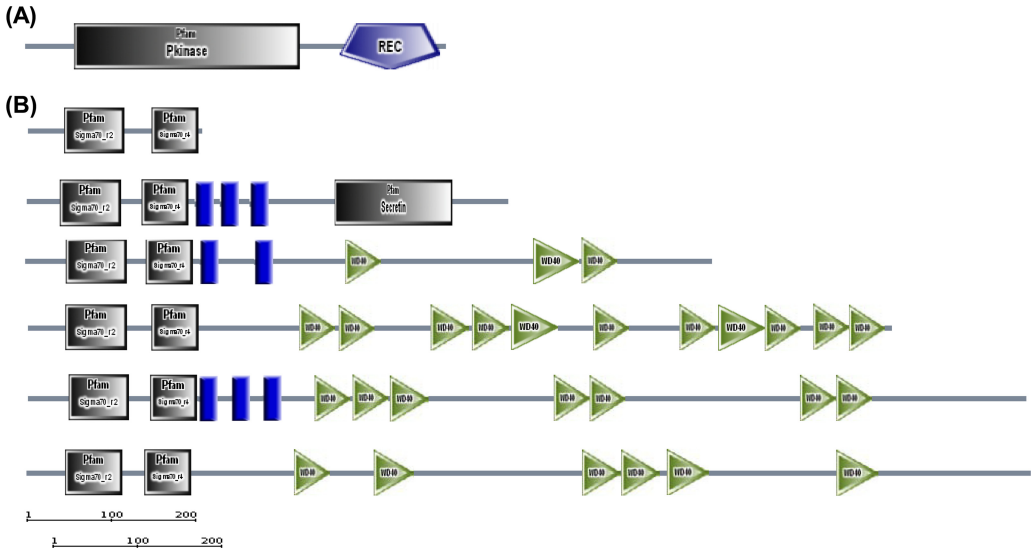


FIG 6 Domain architectures of special 2CSs and ECFs in planctomycetes. The typical domain architecture is shown based on results from the SMART database (http://smart.embl-heidelberg.de/). (A) Novel type of RRs in *Planctomycetes* species consisting of an N-terminal Ser/Thr/Tyr protein kinase domain (Pfam designation, Pkinase) and a C-terminal receiver domain. (B) Typical C-terminal extensions of the ECF01-Gob proteins. Standard ECFs (shown on the top) are comprised of only two regions, the σ2 (Pfam designation, Sigma70_r2) and σ4 (Pfam designation, Sigma70_r4) domains, with a length of about 200 amino acids. The C-terminal extensions of ECF01-Gob ECFs in planctomycetes contain up to 1,000 amino acids, which, among other domains (see the text for details), contain putative transmembrane regions (shown in blue rectangles) as well as secretin and WD40 domains with different arrangement styles. The scale bar represents the protein length in amino acids.

RRs are encoded by orphan genes unrelated to their respective partners, pointing toward a more complex mode of 2CS-dependent signal transduction in planctomycetes. Among such genes, one was found to encode a novel type of RR conserved in all eight planctomycetal genomes but restricted to this phylum. Its domain architecture consists of an N-terminal Ser/Thr/Tyr protein kinase (Pfam designation, Pkinase) and a C-terminal receiver (REC) domain (Fig. 6A), which is normally found in the N terminus.

The widely distributed domain architecture of RRs consisting solely of a receiver domain might point toward a role of these proteins in the shuttling of phosphoryl groups between histidine kinases and histidine-containing phosphotransfer proteins (containing a Pfam HPt domain) within complex 2CS-based phosphorelays involved in integrating different stimuli, as described previously for the decision-making process in *B. subtilis* sporulation or the Rcs phosphorelay of enterobacteria (8, 33, 60). This hypothesis seems attractive in light of the greater abundance of RRs than HKs. A second explanation could be that such "receiver-only" RRs mediate their output through direct protein-protein interactions, as most classically exemplified by the CheY-like RRs involved in the chemotaxis systems discussed below. Taken together, only a small fraction of planctomycetal 2CSs provides a simple one-to-one connection between a stimulus input and a cellular response in the form of differential gene expression. Instead, the majority of 2CS proteins seem to be involved in complex, highly interconnected, multistep phosphorelays required to orchestrate some aspects of the complex life cycle described for this phylum. Interestingly, the overall situation is very reminiscent of that of the deltaproteobacterium *Myxococcus xanthus* (50). However, the unusual domain configuration of some planctomycetal RRs (Fig. 6A) remains enigmatic. Probably, upon the phosphorylation of the receiver domain by a cognate HK, the N-terminal protein kinase domain is activated to phosphorylate its target protein(s) as an output of 2CS-mediated signal transduction. While both the physiological role and the signaling mechanism remain elusive, the conservation of these novel RRs within all planctomycetal species and their absence outside this phylum make these proteins very interesting candidates for future functional studies.

Chemotaxis. While bacterial chemotaxis is derived from 2CSs, we treat such proteins as a separate entity, because of the large number of unique domains and proteins dedicated to orchestrating bacterial motility. All planctomycetes included in this study are motile in one stage of their life cycle, while *I. pallida* lacks a flagellum, and its motility is achieved through gliding (16). Thus, chemotaxis is to be expected for all the analyzed bacteria. We focused on the distribution and abundance of two central chemotaxis proteins, so-called methyl-accepting chemoreceptor proteins (MCPs) and the CheA-like HK, in order to gain some insight into the motility intelligence between different species. While the first gives an indication of the diversity of different stimuli sensed, the second is a reliable measure of the number of independent chemotaxis systems encoded in a given bacterial genome. Two major conclusions can be drawn from the data shown in Table 3. First, there is a huge variance in the complexity of the chemotaxis systems among the eight species, which is mostly independent of genome size. While *B. marina*, *P. staleyi*, and especially *G. obscuriglobus* have a high motility intelligence, with four to five independent chemotaxis systems and 20 to 60 MCPs, the remaining planctomycetes contain only a very limited number of chemotaxis proteins, based on one or two chemotaxis modules. Second, our analysis revealed the complete lack of any chemotaxis machinery in the motile and flagellated organism *R. baltica*, which is in agreement with data from previous studies (17). This can mean that *R. baltica* can swim but has no means to direct its movement, or it somehow has acquired, developed, or adapted a completely different mechanism for adjusting its flagellar motor. Both alterna-

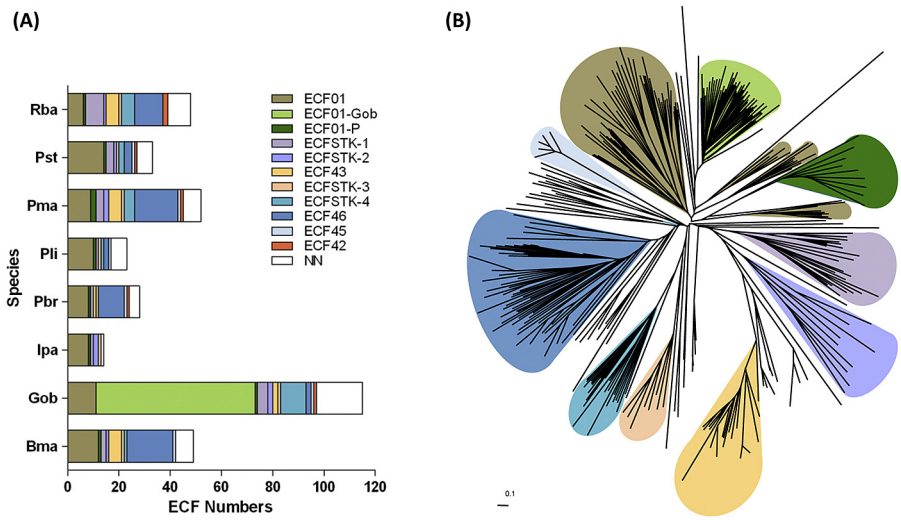


FIG 7 Phylogenetic tree and classification of ECF sigma factors from planctomycetes. Shown is a phylogenetic tree of all ECF sequences extracted from the MiST2 database, generated by the least-squares distances method. A subsequent group analysis of unclassified planctomycetal ECFs was done based on their phylogenetic distances, genomic context conservation, or the presence of promoter motifs, which resulted in the assignment of the novel groups ECF45, ECF46, ECFSTK1, ECFSTK2, ECFSTK3, and ECFSTK4 as well as the ECF01-like groups ECF01-Gob and ECF01-P. The corresponding sequence information is shown in Table S5 in the supplemental material. Different groups are shown in different colors and are highlighted correspondingly in the tree map. (A) Abundance and distribution of ECFs in the different planctomycete species. Abbreviations: Rba, *Rhodopirellula baltica* SH1; Pst, *Pirellula staleyi* DSM 6068; Pma, *Planctomyces maris* DSM 8797; Pli, *Planctomyces limnophilus* DSM 3776; Pbr, *Planctomyces brasiliensis* DSM 5305; Ipa, *Isosphaera pallida* ATCC 43644; Gob, *Gemmata obscuriglobus* UQM 2246; Bma, *Blastopirellula marina* DSM 3645. (B) Phylogenetic tree of ECFs in planctomycete species. Eight ECFs were excluded because of their phylogenetic distance from other ECFs. These proteins are highlighted in Table S5 in the supplemental material.

tives seem equally unlikely but point toward very interesting future experiments in this planctomycetal model organism.

ECFs. ECFs belong to the pool of alternative σ factors that recognize specific promoter sequences to control gene expression. Their activity is usually regulated by cognate anti-σ factors (5, 19). Recently, a classification of ECFs was reported, based on about 400 microbial genomes (53). Given the genome sequencing bias toward model bacteria, or those with medical or biotechnological relevance, many phyla were underrepresented despite their ecological relevance. Not surprisingly, most ECFs from the dominant phyla can be classified, while a majority of ECFs from underrepresented phyla, including the *Planctomycetes*, could not be assigned to any defined ECF group (53). Because of the large abundance of unclassified ECFs in *Planctomycetes*, we performed an in-depth analysis of these proteins mimicking the initial ECF classification (53). The overall abundance of ECFs in the eight sequenced planctomycetal species is shown in Fig. 7A. Branches of the phylogenetic tree corresponding to known or novel conserved ECF groups are color coded in Fig. 7B, and their genomic abundance and distribution are indicated by the same colors in Fig. 7A. ECFs were found to represent a widely distributed signaling principle for the phylum *Planctomycetes*. *G. obscuriglobus* in particular is one of the most ECF-rich bacteria sequenced to date, with its genome encoding 115 ECFs. This is only three proteins short of the current record holder, the deltaproteobacterium *Plesiocystis pacifica*. Based on the original set of 43 group-specific hidden Markov models (HMMs) described previously by Staron et al. (53), only 96 out of the 362 proteins could be classified into one of four conserved ECF groups present in planctomycetes (data not shown). The majority of these belong to the highly diverse ECF01 group of RpoE-like σ factors that are functionally linked to typical membrane-anchored anti-σ factors harboring an antisigma domain. The other classifiable ECFs were assigned either to the two poorly characterized groups ECF22 and ECF42, which both lack a designated anti-σ factor, or to the more distantly related ECF-like proteins of the ECF43 group, which seem to be functionally linked to STPKs (53). In this new planctomycete-specific analysis, we defined eight novel ECF groups based on phylogenetic relationships and genomic context conservation, raising the count of classified planctomycetal ECFs by a factor of 3, from 96 to 304 (Fig. 7; see also Table S5 in the supplemental material for details). Most strikingly, the ECF01-Gob branch (Fig. 7B, light green) contains more than half of all ECFs encoded in the genome of *G. obscuriglobus* (62 of 115) but not a single protein from any of the other seven planctomycetal species. This strongly suggests that more than 50% of all ECFs in this organism are evolutionarily young paralogues. In addition, these *Gemmata*-specific ECFs comprise a diverse and unusual domain architecture (Fig. 6B). They share very large C-terminal extensions, sometimes exceeding the length of the ECF core by a factor of more than 4 (19). The extension of about 20 ECF01-Gob proteins contains multiple WD40 repeats that might be involved in the assembly of multiprotein complexes (51). In addition, most ECF01-Gob proteins are membrane anchored, harboring 1 to 3 putative transmembrane helices between the ECF core and the C-terminal extension. These are the first membrane-anchored ECFs described for bacteria. Given the presence of a nucleus-like cellular organization in this organism, it is a highly attractive, but also purely speculative, hypothesis that this unusual type of transmembrane ECF has specifically evolved in this species to facilitate signal transduction between the cytoplasm and the nucleoid.

Besides ECF01-Gob, we identified four well-separated novel ECF groups (ECFSTK1 to ECFSTK4) that are all functionally linked to STPKs, according to genomic context conservation.

Such a link was previously described only for the ECF-like proteins of the ECF43 group (53). This result demonstrates the wide distribution of signaling modules consisting of a protein kinase and an ECF-type σ factor. While it is tempting to speculate that, in these modules, the STPKs function as sensors that phosphorylate and thereby activate their cognate ECF partner, such a novel mechanism of ECF-dependent signal transduction awaits experimental verification.

Taken together, our ECF analysis of the phylum *Planctomycetes* indicates that the exploration of this third pillar of bacterial signal transduction has only just begun. Especially in phyla underrepresented in the data set of the initial analysis (53), there is a large potential for finding novel conserved mechanisms not present in the more common model organisms. The groups described above not only add to the growing list of conserved ECFs but also allow a first glimpse into how planctomycetes employ such proteins as a central mechanism of signal transduction.

Conclusion. By employing various bioinformatic methods, we identified proteins putatively involved in planctomycetal cell division and in their conspicuous cell plan. In addition, we identified planctomycetal signal transduction systems as well as sigma factors and evidence for a new signaling logic among *Planctomycetes*. As the genetic manipulation of *Planctomycetes* remains tedious, our findings will help to guide future wet laboratory experiments with a manageable list of the most promising target proteins to study hallmark planctomycetal traits.

REFERENCES

1. **Bauer M, et al.** 2004. *Archaea*-like genes for C1-transfer enzymes in *Planctomycetes*: phylogenetic implications of their unexpected presence in this phylum. J. Mol. Evol. **59**:571–586.
2. **Bernander R, Ettema TJ.** 2010. FtsZ-less cell division in Archaea and Bacteria. Curr. Opin. Microbiol. **13**:747–752.
3. **Bertin A, et al.** 2012. Three-dimensional ultrastructure of the septin filament network in *Saccharomyces cerevisiae*. Mol. Biol. Cell **23**:423–432.
4. **Brochier C, Philippe H.** 2002. Phylogeny: a non-hyperthermophilic ancestor for bacteria. Nature **417**:244.
5. **Butcher BG, Mascher T, Helmann JD.** 2008. Environmental sensing and the role of extracytoplasmic function (ECF) sigma factors, p 233–261. *In* El-Sharoud WM (ed), Bacterial physiology—a molecular approach Springer-Verlag GmbH, Berlin, Germany.
6. **Cayrou C, Raoult D, Drancourt M.** 2010. Broad-spectrum antibiotic resistance of *Planctomycetes* organisms determined by Etest. J. Antimicrob. Chemother. **65**:2119–2122.

6a. **Clum A, et al.** 2009. Complete genome sequence of *Pirellula staleyi* type strain (ATCC 27377). Stand. Genomic Sci. **1**:308–316.

7. **de Been M, Francke C, Siezen RJ, Abee T.** 2011. Novel sigmaB regulation modules of Gram-positive bacteria involve the use of complex hybrid histidine kinases. Microbiology **157**:3–12.
8. **de Hoon MJ, Eichenberger P, Vitkup D.** 2010. Hierarchical evolution of the bacterial sporulation network. Curr. Biol. **20**:R735–R745.
9. **Felsenstein J.** 1989. Mathematics vs. evolution: mathematical evolutionary theory. Science **246**:941–942.
10. **Forterre P, Gribaldo S.** 2010. Bacteria with a eukaryotic touch: a glimpse of ancient evolution? Proc. Natl. Acad. Sci. U. S. A. **107**:12739–12740.
11. **Fuchsman CA, Rocap G.** 2006. Whole-genome reciprocal BLAST analysis reveals that *Planctomycetes* do not share an unusually large number of genes with Eukarya and Archaea. Appl. Environ. Microbiol. **72**:6841–6844.
12. **Fuerst JA, Sagulenko E.** 2011. Beyond the bacterium: *Planctomycetes* challenge our concepts of microbial structure and function. Nat. Rev. Microbiol. **9**:403–413.
13. **Fuerst JA, Webb RI.** 1991. Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. Proc. Natl. Acad. Sci. U. S. A. **88**:8184–8188.
14. **Galperin MY.** 2010. Diversity of structure and function of response regulator output domains. Curr. Opin. Microbiol. **13**:150–159.
15. **Gao R, Stock AM.** 2009. Biological insights from structures of two-component proteins. Annu. Rev. Microbiol. **63**:133–154.
16. **Giovannoni SJ, Godchaux W III, Schabtach E, Castenholz RW.** 1987. Cell wall and lipid composition of *Isosphaera pallida*, a budding eubacterium from hot springs. J. Bacteriol. **169**:2702–2707.
17. **Glöckner FO, et al.** 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc. Natl. Acad. Sci. U. S. A. **100**:8298–8303.

17a. **Göker M, et al.** 2011. Complete genome sequence of *Isosphaera pallida* type strain (IS1B). Stand. Genomic Sci. **4**:63–71.

18. **Hall T.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**:95–98.
19. **Helmann JD.** 2002. The extracytoplasmic function (ECF) sigma factors. Adv. Microb. Physiol. **46**:47–110.
20. **Ingram BO, Sohlenkamp C, Geiger O, Raetz CR.** 2010. Altered lipid A structures and polymyxin hypersensitivity of *Rhizobium etli* mutants lacking the LpxE and LpxF phosphatases. Biochim. Biophys. Acta **1801**:593–604.
21. **Jermy A.** 2010. Evolution: bacterial endocytosis uncovered. Nat. Rev. Microbiol. **8**:534.
22. **Jogler C, Glockner FO, Kolter R.** 2011. Characterization of *Planctomyces limnophilus* and development of genetic tools for its manipulation establish it as a model species for the phylum *Planctomycetes*. Appl. Environ. Microbiol. **77**:5826–5829.
23. **Jun SR, Sims GE, Wu GA, Kim SH.** 2010. Whole-proteome phylogeny of prokaryotes by feature frequency profiles: an alignment-free method with optimal feature resolution. Proc. Natl. Acad. Sci. U. S. A. **107**:133–138.
24. **Kerger BD, et al.** 1988. The budding bacteria, Pirellula and Planctomyces, with atypical 16S ribosomal-RNA and absence of peptidoglycan, show eubacterial phospholipids and uniquely high proportions of long-chain beta-hydroxy fatty-acids in the lipopolysaccharide lipid-A. Arch. Microbiol. **149**:255–260.
25. **Kirkpatrick CL, Viollier PH.** 2012. Decoding Caulobacter development. FEMS Microbiol. Rev. **36**:193–205.
26. **König E, Schlesner H, Hirsch P.** 1984. Cell-wall studies on budding bacteria of the *Planctomyces Pasteuria* group and on a *Prosthecomicrobium* sp. Arch. Microbiol. **138**:200–205.

26a. **Labutti K, et al.** 2010. Complete genome sequence of *Planctomyces limnophilus* type strain (Mu 290). Stand. Genomic Sci. **3**:47–56.

27. **Liesack W, Konig H, Schlesner H, Hirsch P.** 1986. Chemical-composition of the peptidoglycan-free cell envelopes of budding bacteria of the *Pirella Planctomyces* group. Arch. Microbiol. **145**:361–366.
28. **Lindsay MR, Webb RI, Fuerst JA.** 1997. Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus *Pirellula*. Microbiology **143**:739–748.
29. **Lonhienne TG, et al.** 2010. Endocytosis-like protein uptake in the bacterium *Gemmata obscuriglobus*. Proc. Natl. Acad. Sci. U. S. A. **107**:12883–12888.
30. **Lopez D, Kolter R.** 2010. Extracellular signals that define distinct and coexisting cell fates in Bacillus subtilis. FEMS Microbiol. Rev. **34**:134–149.
31. **Ludwig W, et al.** 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32**:1363–1371.
32. **Lutkenhaus J.** 2007. Assembly dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. Annu. Rev. Biochem. **76**:539–562.
33. **Majdalani N, Gottesman S.** 2005. The Rcs phosphorelay: a complex signal transduction system. Annu. Rev. Microbiol. **59**:379–405.
34. **Makarova KS, Koonin EV.** 2010. Two new families of the FtsZ-tubulin protein superfamily implicated in membrane remodeling in diverse bacteria and archaea. Biol. Direct **5**:33.
35. **Makarova KS, Yutin N, Bell SD, Koonin EV.** 2010. Evolution of diverse cell division and vesicle formation systems in *Archaea*. Nat. Rev. Microbiol. **8**:731–741.
36. **Marchler-Bauer A, et al.** 2011. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. **39**:D225–D229.
37. **Margolis P, Driks A, Losick R.** 1991. Establishment of cell type by compartmentalized activation of a transcription factor. Science **254**:562–565.
38. **McCormick JR, Flardh K.** 2012. Signals and regulators that govern *Streptomyces* development. FEMS Microbiol. Rev. **36**:206–231.
39. **Mingorance J, Tamames J, Vicente M.** 2004. Genomic channeling in bacterial cell division. J. Mol. Recognit. **17**:481–487.
40. **Offner S, Wanner G, Pfeifer F.** 1996. Functional studies of the gvpACNO

operon of *Halobacterium salinarium* reveal that the GvpC protein shapes gas vesicles. J. Bacteriol. **178**:2071–2078.
41. **Park KT, et al.** 2011. The Min oscillator uses MinD-dependent conformational changes in MinE to spatially regulate cytokinesis. Cell **146**:396–407.
42. **Pearson A, Budin M, Brocks JJ.** 2003. Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus. Proc. Natl. Acad. Sci. U. S. A. **100**:15352–15357.
43. **Pereira G, Yamashita YM.** 2011. Fly meets yeast: checking the correct orientation of cell division. Trends Cell Biol. **21**:526–533.
44. **Pereira SF, Goss L, Dworkin J.** 2011. Eukaryote-like serine/threonine kinases and phosphatases in bacteria. Microbiol. Mol. Biol. Rev. **75**:192–212.
45. **Pilhofer M, Ladinsky MS, McDowall AW, Petroni G, Jensen GJ.** 2011. Microtubules in bacteria: ancient tubulins build a five-protofilament homolog of the eukaryotic cytoskeleton. PLoS Biol. **9**:e1001213. doi:10.1371/journal.pbio.1001213.
46. **Pilhofer M, et al.** 2008. Characterization and evolution of cell division and cell wall synthesis genes in the bacterial phyla *Verrucomicrobia*, *Lentisphaerae*, *Chlamydiae*, and *Planctomycetes* and phylogenetic comparison with rRNA genes. J. Bacteriol. **190**:3192–3202.
47. **Pruesse E, et al.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. **35**:7188–7196.
48. **Santarella-Mellwig R, et al.** 2010. The compartmentalized bacteria of the *Planctomycetes-Verrucomicrobia-Chlamydiae* superphylum have membrane coat-like proteins. PLoS Biol. **8**:e1000281. doi:10.1371/journal.pbio.1000281.
49. **Shaevitz JW, Gitai Z.** 2010. The structure and function of bacterial actin homologs. Cold Spring Harb. Perspect. Biol. **2**:a000364. doi:10.1101/cshperspect.a000364.
50. **Shi X, et al.** 2008. Bioinformatics and experimental analysis of proteins of two-component systems in *Myxococcus xanthus*. J. Bacteriol. **190**:613–624.
51. **Smith TF, Gaitatzes C, Saxena K, Neer EJ.** 1999. The WD repeat: a common architecture for diverse functions. Trends Biochem. Sci. **24**:181–185.
52. **Staron A, Mascher T.** 2010. General stress response in alpha-proteobacteria: PhyR and beyond. Mol. Microbiol. **78**:271–277.
53. **Staron A, et al.** 2009. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. Mol. Microbiol. **74**:557–581.
54. **Tamames J, Gonzalez-Moreno M, Mingorance J, Valencia A, Vicente M.** 2001. Bringing gene order into bacterial shape. Trends Genet. **17**:124–126.
55. **Teeling H, Lombardot T, Bauer M, Ludwig W, Glockner FO.** 2004. Evaluation of the phylogenetic position of the planctomycete 'Rhodopirellula baltica' SH 1 by means of concatenated ribosomal protein sequences, DNA-directed RNA polymerase subunit sequences and whole genome trees. Int. J. Syst. Evol. Microbiol. **54**:791–801.
56. **Ulrich LE, Koonin EV, Zhulin IB.** 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. **13**:52–56.
57. **Ulrich LE, Zhulin IB.** 2010. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. Nucleic Acids Res. **38**:D401–D407.
58. **van Niftrik L, et al.** 2009. Cell division ring, a new cell division protein and vertical inheritance of a bacterial organelle in anammox *Planctomycetes*. Mol. Microbiol. **73**:1009–1019.
59. **Wagner M, Horn M.** 2006. The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr. Opin. Biotechnol. **17**:241–249.
60. **Wang D, Korban SS, Zhao Y.** 2009. The Rcs phosphorelay system is essential for pathogenicity in *Erwinia amylovora*. Mol. Plant Pathol. **10**:277–290.
61. **Weisburg WG, Hatch TP, Woese CR.** 1986. Eubacterial origin of *Chlamydiae*. J. Bacteriol. **167**:570–574.
62. **Woese CR.** 1987. Bacterial evolution. Microbiol. Rev. **51**:221–271.

Supplementary files for Chapter 2

The following supplementary files could be found at http://jb.asm.org/content/194/23/6419/suppl/DCSupplemental .

Supporting text.

- Selected morphology related genes and proteins in *Planctomycetes*
- Detailed description of PO1-5 found using the "guilt by association approach" on the planctomycetal core genome
- Detailed description of PO6-10 found using the "guilt by association approach " on putative cell-division related operons found in the core genome of *P. limnophilus*
- Detailed description of the newly identified ECF groups of *Planctomycetes*

Table S1. Accession numbers of 16S rRNA genes used for phylogenetic analysis.

Table S2. The planctomycetal core-genome.

Table S3. The *Planctomyces limnophilus* core genome.

Table S4. Cell division and cytoskeleton related planctomycetal proteins.

Table S5. Comprehensive planctomycetal ECF-dataset.

Chapter 3

Environmental sensing in *Actinobateria*: A comprehensive survey on the signaling capacity of this phylum.

Huang X[#], Pinto D[#], Fritz G, Mascher T (2015) *J Bacteriol* **197:** 2517-2535. ([#]**Co-first author**.)

JB Journals.ASM.org

Environmental Sensing in *Actinobacteria*: a Comprehensive Survey on the Signaling Capacity of This Phylum

Xiaoluo Huang, Daniela Pinto, Georg Fritz,* Thorsten Mascher*

Department Biology I, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany

ABSTRACT

Signal transduction is an essential process that allows bacteria to sense their complex and ever-changing environment and adapt accordingly. Three distinct major types of signal-transducing proteins (STPs) can be distinguished: one-component systems (1CSs), two-component systems (2CSs), and extracytoplasmic-function σ factors (ECFs). Since *Actinobacteria* are particularly rich in STPs, we comprehensively investigated the abundance and diversity of STPs encoded in 119 actinobacterial genomes, based on the data stored in the Microbial Signal Transduction (MiST) database. Overall, we observed an approximately linear correlation between the genome size and the total number of encoded STPs. About half of all membrane-anchored 1CSs are protein kinases. For both 1CSs and 2CSs, a detailed analysis of the domain architectures identified novel proteins that are found only in actinobacterial genomes. Many actinobacterial genomes are particularly enriched for ECFs. As a result of this study, almost 500 previously unclassified ECFs could be classified into 18 new ECF groups. This comprehensive survey demonstrates that actinobacterial genomes encode previously unknown STPs, which may represent new mechanisms of signal transduction and regulation. This information not only expands our knowledge of the diversity of bacterial signal transduction but also provides clear and testable hypotheses about their mechanisms, which can serve as starting points for experimental studies.

IMPORTANCE

In the wake of the genomic era, with its enormous increase in the amount of available sequence information, the challenge has now shifted toward making sense and use of this treasure chest. Such analyses are a prerequisite to provide meaningful information that can help guide subsequent experimental efforts, such as mechanistic studies on novel signaling strategies. This work provides a comprehensive analysis of signal transduction proteins from 119 actinobacterial genomes. We identify, classify, and describe numerous novel and conserved signaling devices. Hence, our work serves as an important resource for any researcher interested in signal transduction of this important bacterial phylum, which contains organisms of ecological, biotechnological, and medical relevance.

Bacterial survival critically depends on the ability to swiftly respond to environmental changes. To efficiently monitor the surrounding environment, microbial genomes encode numerous and highly diverse proteins that can sense a given extracellular stimulus, transmit the signal to the cytoplasm, and elicit a proper response. These signal-transducing proteins (STPs) can be divided into three major groups: one-component systems (1CSs), two-component systems (2CSs), and extracytoplasmic-function σ factors (ECFs).

The vast majority of STPs in bacteria are 1CSs. These systems are composed of a single protein that contains an input domain, which senses the stimulus, and an output domain, which elicits the response by binding nucleic acids, modifying proteins, or performing an enzymatic reaction (1). 2CSs, which are typically composed of a histidine kinase (HK) and a response regulator (RR), represent the second-most-abundant signaling principle. The HKs are usually membrane-associated proteins with an extracytoplasmic N-terminal input domain and a cytoplasmic C-terminal transmitter domain. Upon stimulus perception, the HKs autophosphorylate at a highly conserved histidine residue. This phosphohistidine then serves as a phosphoryl group donor to activate the cognate RR through phosphorylation of an invariant aspartate residue. RRs are usually soluble proteins that contain an N-terminal receiver domain, as the target site of the phosphotransfer, and a C-terminal output domain. The output domains of 2CSs are often phylogenetically related to those found in 1CSs (1) and hence also bind nucleic acids, modify proteins, perform some enzymatic activity, or, less frequently, bind other proteins (2). The third pillar of bacterial signal transduction is represented by ECFs. Like other σ factors, ECFs are components of the RNA polymerase holoenzyme that determine the promoter specificity (3). In contrast to the more complex and essential housekeeping σ factors, ECFs contain only two of the four conserved domains of σ^{70} proteins, termed σ_2 and σ_4, which are sufficient for interaction with the RNA polymerase core enzyme and for mediating promoter

Received 5 March 2015 **Accepted** 12 May 2015

Accepted manuscript posted online 18 May 2015

Citation Huang X, Pinto D, Fritz G, Mascher T. 2015. Environmental sensing in *Actinobacteria*: a comprehensive survey on the signaling capacity of this phylum. J Bacteriol 197:2517–2535. doi:10.1128/JB.00176-15.

Editor: I. B. Zhulin

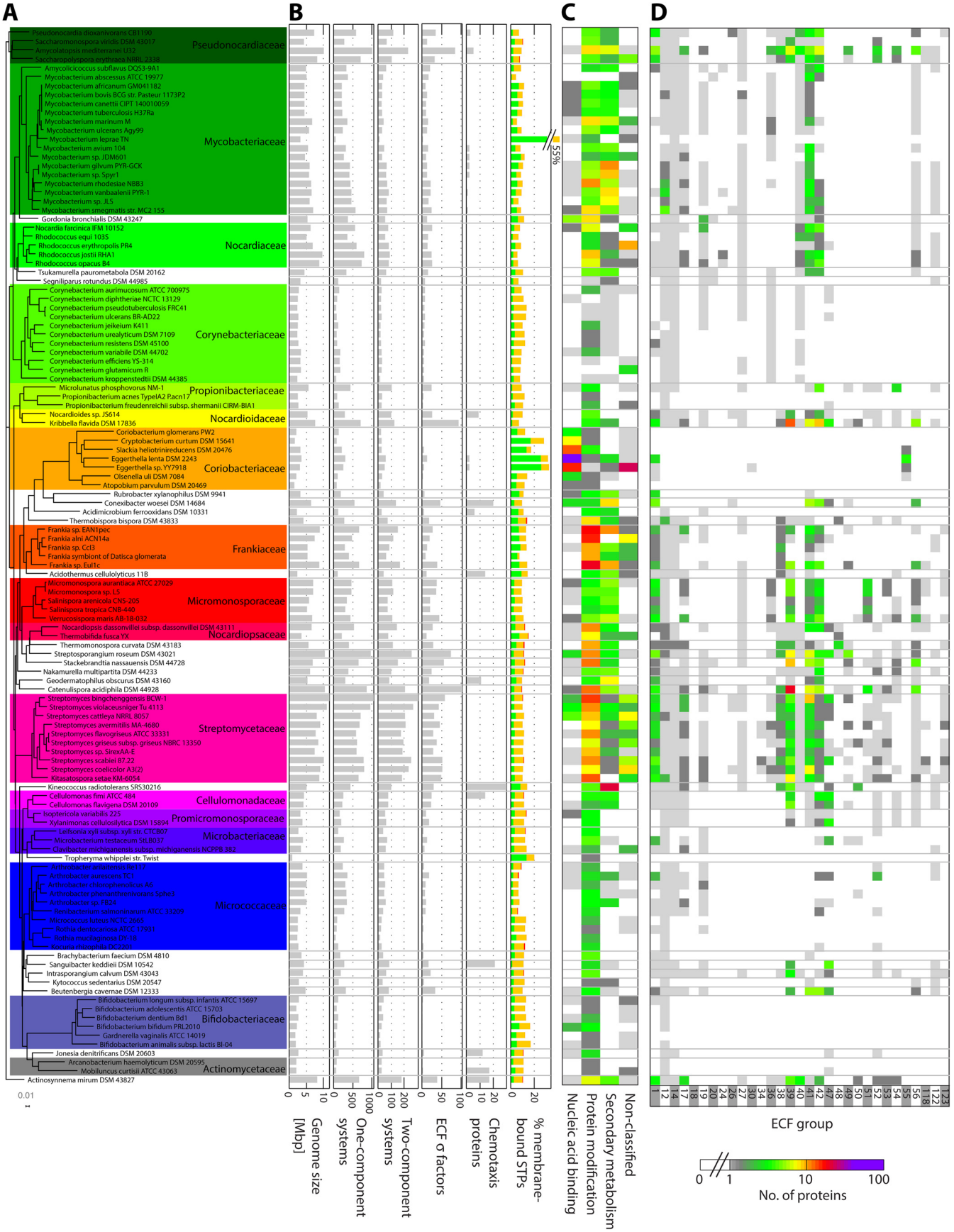
Address correspondence to Thorsten Mascher, thorsten.mascher@tu-dresden.de.

* Present address: Georg Fritz, LOEWE-Center for Synthetic Microbiology, Philipps-Universität Marburg, Marburg, Germany; Thorsten Mascher, Institut für Mikrobiologie, Technische Universität Dresden, Dresden, Germany.

X.H. and D.P. contributed equally to this work.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.00176-15.

doi:10.1128/JB.00176-15



recognition. The activity of the ECFs is usually controlled by membrane-associated anti-σ factors (ASFs) that tightly bind (and thereby inactivate) the cognate ECFs (4). Upon perceiving an appropriate inducing stimulus, the ASFs are inactivated through modifications, conformational changes, or regulated proteolysis, thereby releasing the ECF to recruit RNA polymerase core enzyme and ultimately allowing transcription initiation from alternative and ECF-specific target promoters.

In the wake of the genomic era, with its enormous increase in the amount of available sequence information, the challenge has now shifted toward making sense and use of this treasure chest. Such analyses are a prerequisite to provide meaningful information that can help guide subsequent experimental efforts, such as mechanistic studies on novel signaling strategies. With respect to STPs, this provoked the need to phylogenetically group and classify them in order to identify conserved features that ultimately allow the development of hypotheses about their physiological roles and signaling mechanisms. Over the last decade, classification systems were proposed for 1CSs, HKs, RRs, and ECFs (1, 2, 5, 6). In 2005, Ulrich et al. proposed a classification of 1CSs based on the specific combinations of input and output domains (1). One year later, Galperin suggested classifying RRs based on their output domains (2). A functional grouping of HKs was based on the membrane topology, number of transmembrane (TM) helices, and sequential arrangement of the sensory domains within the N-terminal input domains (5). In the case of ECFs, a combination of sequence similarity, the domain architectures of both the ECFs and their the cognate ASFs, genomic context conservation, and target promoter motifs was used to develop a classification scheme (6).

All of these studies indicated a number of unique features of STPs from actinobacterial genomes: for instance, they do not encode a number of RR types found in other bacteria (e.g., RECSARP or NtrC type) and virtually lack chemotaxis-related proteins (2). Moreover, many actinobacterial genomes are particularly ECF rich and encode over a dozen unique ECF groups (6). For these reasons, and to account for the significantly increased number and diversity of actinobacterial genomes available now, we decided to comprehensively analyze the STP landscape of *Actinobacteria*. Our results significantly expand our knowledge compared to the earlier studies, which were based on the rather limited number of actinobacterial genome sequences available at that time. Our goal was to extract a comprehensive picture of how this phylum perceives the environment.

MATERIALS AND METHODS

Building the actinobacterial genome collection. The phylogenetic tree presented in Fig. 1 was built from the set of 16S rRNA gene sequences of all genomes (see Table S1 in the supplemental material), retrieved via the NCBI Nucleotide Database (http://www.ncbi.nlm.nih.gov/nucleotide). The multiple-sequence alignment of the sequences was generated in Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (7). Its gapless version was used to generate the phylogenetic tree, which was built in the BioEdit Sequence Alignment Editor (http://www.mbio.ncsu.edu/bioedit/bioedit.html) (8) using the neighbor-joining method (9) and visualized in Dendroscope (10) (available at http://ab.inf.uni-tuebingen.de/software/dendroscope/).

Characterization of one- and two-component systems. For the complete set of proteins identified as 1CSs in the Microbial Signal Transduction (MiST) database (11), the protein annotation, organism, number of transmembrane helices (TMHs) predicted by TMHMM (available at http://www.cbs.dtu.dk/services/TMHMM/) (12), and conserved protein domains predicted by Pfam (http://pfam.xfam.org) (13) were extracted and used for a semiautomated classification with custom scripts written in MATLAB (The MathWorks Inc.). From the complete set of 1CSs, 1,999 proteins with at least one transmembrane helix were selected for further analysis. As a selection criterion, we considered TMH predictions by TMHMM Server v. 2.0 (12), which predicts only membrane-spanning TMHs. Note that the TMHs graphically represented in the MiST database (11) are predictions by the DAS-TMfilter server (http://mendel.imp.ac.at/sat/DAS/DAS.html) (14), which include hydrophobic stretches as small as 2 amino acids long, thus leading to a large number of false positives in an automated screen. The selected proteins were then classified, based on their domain architecture (as predicted by Pfam [13]), as protein kinases or phosphatases, guanylate cyclases, and DNA- or RNA-binding proteins (Table 1; see Table S2 in the supplemental material).

The complete set of identified RRs was investigated based on the nature of their output domains as predicted by Pfam (13). The number of proteins with each individual domain was manually determined. Proteins with uncommon domain architectures were further analyzed regarding their genomic context conservation and taxonomical span (see Table S3 in the supplemental material). The first was investigated using the tree-based genome browser tool in the MicrobesOnline database (15) (http://www.microbesonline.org; March 2011 update) and the second using the NCBI Conserved Domain Architecture Retrieval Tool (CDART) (16) (http://www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi).

Similarly, the complete set of proteins in our genome collection identified by the MiST database (11) as HKs were further analyzed. These proteins were preclassified based on their Pfam (13) domain architecture and transmembrane helices predicted by the TMHMM Server (12) using custom scripts written in MATLAB (The MathWorks Inc.). After manual validation of the classification, one representative of each group was used to evaluate the genomic context and taxonomical span (Table 2; see Table S4 in the supplemental material). As for response regulators, the first was investigated using the tree-based genome browser tool in the MicrobesOnline database (15) and the second using NCBI CDART (16).

Classification of ECFs. Of the 2,203 ECFs identified in the 119 actinobacterial genomes, 526 ECFs could not be associated with any of the ECF groups defined previously (6). These protein sequences were then further analyzed (see Tables S5 and S6 in the supplemental material). A multiple-sequence alignment was generated in Clustal Omega (7) from the sequences of all unclassified ECF σ factors, trimmed to contain only the conserved regions σ_2 and σ_4. The unrooted tree was generated from the gapless multiple-sequence alignment using the neighbor-joining method (9) implemented in the BioEdit Sequence Alignment Editor (8). The grouping was then manually performed on the resulting tree. As before, the genomic context analysis was performed using the tree-based genome browser tool in MicrobesOnline (15).

FIG 1 Distribution pattern of STPs in the phylum *Actinobacteria*. (A) Phylogenetic tree (based on 16S rRNA) of all the organisms analyzed here. The names of the families represented in our collection by more than one genome are shown on the right in a larger font. Each family is color coded. (B) Distributions of genome sizes and each type of STP by organism. For the percentage of membrane-bound STPs, the color codes for the type of STP is as follows: 1CSs, green; 2CSs, orange; and ECFs, red. (C) The distribution of 1CSs is illustrated by a two-dimensional heat map, where the colors indicate the number of 1CSs of a given type in a given genome. (D) The distribution of ECFs into groups is also illustrated by a two-dimensional heat map, where the colors indicate ECF numbers. The color code is the same as for panel C and is shown below. For clarity, only ECF groups containing more than 10 proteins are shown.

TABLE 1 Classification of membrane-anchored actinobacterial 1CSs

Group identifier	No. of proteins	Length[a]	Protein domain architecture[b]	Taxonomical span[c]	Conserved genomic context
Kinases					
1CS_1.1	633	560 ± 136	TMH_n-Pkinase Pkinase-TMH_n TMH_n-Pkinase-TMH_n	Widespread	None
1CS_1.2	163	640 ± 61	Pkinase-$TMH_{1–2}$-$PASTA_{1–5}$	At, B, Cf, F	Transpeptidase, FtsW
1CS_1.3	15	735 ± 106	Pkinase-$NHL_{1–4}$	At, Cf, Dt, F, Pr	None
1CS_1.4	12	852 ± 82	Pkinase-TMH-$WD40_{2–7}$	At, Cf, Cy, Dt, Pl, V	None
1CS_1.5	9	609 ± 32	Pkinase-TMH_PknH_C	At	None
1CS_1.6	8	761 ± 37	Pkinase-TMH-PQQ_2_2	At, Cf, Cy, Dt	None
1CS_1.7	5	789 ± 13	TMH_2-PAP2-Pkinase-UPF0104	At	None
1CS_1.8	5	587 ± 83	Pkinase-TMH-DUF4352	At, Cf	None
1CS_1.9	5	619 ± 26	Pkinase-TMH-Lipoprotein_21	At	None
1CS_1.unclassified	47	NA	Various	NA	NA
Phosphatases					
1CS_2.1	117	392 ± 44	SpoIIE-TMH TMH-$SpoIIE_{1–4/16}$ TMH-PP2C_2	Widespread	None
1CS_2.2	35	527 ± 153	$TMH_{2–8}$-HD $TMH_{1–10}$-GGDEF-HD TMH-(7TMR-HDED)-7TM_7MR_HD-HD	Widespread	None
1CS_2.3	7	609 ± 52	(TMH)-CHASE-TMH-HAMP-$SpoIIE_{1–2}$	At, Cy	None
1CS_2.4	5	680 ± 36	MASE1-(PAS/GAF)-SpoIIE	At, Cy, Pt, Sp	None
1CS_2.unclassified	6	NA	Various	NA	NA
Guanylate cyclases					
1CS_3.1	192	419 ± 92	$TMH_{1–10}$-GGDEF	Widespread	None
1CS_3.2	130	783 ± 131	$TMH_{1–10}$-GGDEF-EAL $TMH_{2/5}$-$GGDEF_2$-EAL TMH_2-GGDEF-EAL-TMH TMH_{10}-GGDEF-TMH_9-GGDEF-EAL	Widespread	None
1CS_3.3	118	548 ± 67	$TMH_{2–7}$-HAMP-Guanylate_cyc	Widespread	None
1CS_3.4	43	931 ± 110	$TMH_{1–2/5–6/8–9}$-$PAS_{1–2/4}$-GGDEF-EAL $TMH_{5–7}$-GAF-GGDEF-EAL TMH_5-GGDEF-EAL-$GAF_{1–2}$	Widespread	None
1CS_3.5	9	757 ± 133	MASE1-($PAS_{2–3}$)-(GAF)-GGDEF-(EAL)	Ac, At, Cy, F, Pr	None
1CS_3.6	6	725 ± 27	$TMH_{1–3}$-HAMP-GAF-GGDEF	At, Cf, Cy, Dt, F, Nt, Pr	None
1CS_3.7	6	641 ± 177	$TMH_{6–9}$-GAF/$PAS_{1–3}$-GGDEF	Widespread	None
1CS_3.8	6	543 ± 156	$TMH_{2–7}$/PTS_EIIC-(PAS)-EAL	Widespread	None
1CS_3.9	5	1,362 ± 48	TMH_2-PAS-GGDEF-($TMH_{1–3}$)-(PAS)-$GGDEF_{1–2}$	At	Acyl-CoA dehydrogenase
1CS_3.unclassified	12	NA	Various	NA	NA
DNA-binding proteins					
1CS_4.1	188	469 ± 88	$TMH_{1–13}$-GerE TMH_8-GerE-TMH_{12}	Widespread	None
1CS_4.2	62	372 ± 409	$HTH_{1–2}$-$TMH_{1–8}$ $TMH_{1–4}$-HTH	Widespread	None
1CS_4.3	34	232 ± 49	TetR_N-$TMH_{1–2}$-(TetR_C) TMH_4-TetR_N	Widespread	None
1CS_4.4	19	282 ± 22	HTH_25-TMH-DUF4115	Widespread	FtsK, 2-methylthioadenine synthetase, CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase
1CS_4.5	7	371 ± 75	HTH_31-TMH-DUF2690	At	None
1CS_4.6	6	279 ± 97	DUF2637-HTH	At	None
1CS_4.7	6	368 ± 55	TMH-DUF4066-HTH_18	Ac, At, B, Cf, Cy, Df, F, Gm, Pl, Pr, Sp, V	None

(Continued on following page)

TABLE 1 (Continued)

Group identifier	No. of proteins	Length[a]	Protein domain architecture[b]	Taxonomical span[c]	Conserved genomic context
1CS_4.unclassified	35	NA	Various	NA	NA
RNA-binding proteins					
1CS_5.1	1	200	TMH_2-ANTAR	At, B, F, Fu, Nt, Pr, Sp, Sy, T	None

[a] Amino acids (mean ± standard deviation).
[b] Protein domain designations as in the Pfam database. Note that when TMHs are not explicitly mentioned in the domain architecture, they are part of one of the assigned domains.
[c] Ac, *Acidobacteria*; Aq, *Aquificae*; Ar, *Armatimonadetes*; At, *Actinobacteria*; B, *Bacteroidetes*; Ca, *Caldiserica*; Cf, *Chloroflexi*; Ch, *Chlorobi*; Cl, *Chlamydiae*; Cr, *Chrysiogenetes*; Cy, *Cyanobacteria*; Df, *Deferribacteres*; Dg, *Dictyoglomi*; Dt, *Deinococcus-Thermus*; E, *Elusimicrobia*; Fb, *Fibrobacteres*; Fu, *Fusobacteria*; Gm, *Gemmatimonadetes*; I, *Ignaeribacteriae*; L, *Lentisphaerae*; M, *Marinimicrobia*; Nn, *Nitrospinae*; Nt, *Nitrospirales*; Pl, *Plantomycetes*; Pr, *Proteobacteria*; Sp, *Spirochaetes*; Sy, *Synergistetes*; T, *Tenericutes*; Td, *Thermodesulfobacteria*; Tt, *Thermotogae*; V, *Verrucomicrobia*. NA, not applicable. In this context, "widespread" refers to 19 to 31 bacterial phyla.

Characterization of ECFs containing C-terminal extensions. Four ECF groups were composed of longer ECFs. A multiple-sequence alignment was built in Clustal Omega (7) from the complete protein sequences of these ECFs and representatives of standard ECFs, allowing visualization of the C-terminal extension (not shown). The complete protein sequences of these ECF σ factors were then submitted to TMHMM Server 2.0 (12) and Pfam (13) for prediction of TMHs and identification of protein domains, respectively. Multiple-sequence alignments were then generated in Clustal W2 (17) (http://www.ebi.ac.uk/Tools/msa/clustalw2/) from trimmed protein sequences encompassing individual identified conserved domains and were visualized with CLC Sequence Viewer software (CLC bio). The amino acid frequency distribution in the C-terminal extensions was also calculated in the CLC Sequence Viewer software.

Identification and characterization of ASFs. Protein sequences of conserved genes located next to and presumably cotranscribed with ECFs were retrieved from MiST (11) (see Table S7 in the supplemental material). They were then submitted to TMHMM Sever 2.0 (12) and Pfam (13) for identification of transmembrane helices and protein domains, respectively. Multiple-sequence alignments were generated in Clustal W2 (17) and visualized with CLC Sequence Viewer software. Sequence logos of predicted segments located in the cytoplasm were generated in the WebLogo tool (18) (http://weblogo.berkeley.edu/logo.cgi) and illustrate the degree of amino acid conservation through graphical representation of a position weight matrix. Secondary-structure prediction of segments located in the periplasm were made in the PSIPRED Protein Sequence Analysis Workbench (available at http://bioinf.cs.ucl.ac.uk/psipred/) using the PSIPRED v3.3 prediction method (19) and graphically represented through the Prosite MyDomains tool (http://prosite.expasy.org/cgi-bin/prosite/mydomains/).

Identification of group-specific target promoters. Initially, a library of upstream regulatory sequences was generated for each new ECF group (see Table S8 in the supplemental material). All 250-nucleotide-long sequences located immediately upstream of the start codon of the first gene in the ECF σ factor-encoding operon were retrieved from Microbes Online (15) or MiST (11). MicrobesOnline's operon predictions were used, except in cases in which the analyzed genome was not part of that database. In such cases, ECF-encoding operons were defined, as previously (20), as all consecutive genes adjacent to the ECF σ factor gene in the same orientation and separated by less than 50 nucleotides. Then, BioProspector (21; http://ai.stanford.edu/~xsliu/BioProspector/) was used to identify overrepresented motifs in those sequences, mostly as described previously (20). The parameter settings used to search for two-block motifs that may not occur in all input sequences and only on their forward strands were as follows: lengths of the upstream and downstream blocks (*W* and *w*, respectively), 5 to 7 nucleotides; minimum distance (*g*) and maximum distance (*G*) separating the two blocks, 15 to 19 nucleotides and 16 to 20 nucleotides, respectively. These parameters were iteratively varied to encompass all possible combinations in which the difference between the maximum and minimum distances separating the two blocks was not more than 1 nucleotide. Third, the 10 highest-scoring motifs selected from 40 reinitializations in each run were manually analyzed. The collection of 450 sequence motifs obtained for each ECF group was initially restricted to those in which the number of motif hits was equal to or lower than the number of input sequences. Then, for each remaining motif, the number of sequences with multiple motif hits was manually determined. From those motifs with the lowest number of sequences with multiple motif hits, the one found in the highest number of sequences and with the highest score was selected. Finally, the sequence logos were generated, using the WebLogo tool (18), from all the motif-containing sequences except those that contained additional, lower-scored motif hits (i.e., only one hit per sequence was used).

RESULTS AND DISCUSSION

In order to generate the genome collection for our analysis, all 299 actinobacterial genomes present in the MiST database (11) were selected. This initial set was then reduced to exclude unfinished draft genomes and to eliminate the redundancy by including only one genome per species, which was chosen based on containing the highest number of STPs for the species. Among the genomes of *Mycobacterium* species strains JLS, KMS, and MCS, only the first was maintained due to the similarity between their STPs' profiles. The remaining set of 119 genomes, which were used for further analysis, is listed in Table S1 in the supplemental material. Information regarding the organisms' lifestyles, as well as abundances and distributions of STPs, was retrieved from the MiST database (11).

Distribution of STPs. The analyzed actinobacterial genomes have GC contents ranging from 40 to 75%, are 0.9 to 12 Mbp in size, and encode numbers of proteins ranging from 808 to 10,022, with an average of 4,380 proteins per genome (see Table S1 in the supplemental material). Of these, on average, about 10% are involved in signal transduction (see Table S9 in the supplemental material). The morphological, metabolic, and habitat diversity of these organisms (see Table S1 in the supplemental material) suggests that their genomes may encode a corresponding diversity of signal-transducing systems.

Our definition of the different types of STPs follows that of the MiST database (11). Briefly, 1CSs are single proteins that contain both input and output domains but lack phosphotransfer domains typical of 2CSs. These 2CSs include, first, HKs, defined as proteins that have a transmitter unit (consisting of the catalytic HATPase_c domain and the DHp domain as the site of autophosphorylation) but not a receiver domain (a more detailed descrip-

TABLE 2 Domain-based classification of actinobacterial HKs

Group identifier	No. of proteins	Length[a]	Protein domain architecture[b]	Taxonomical span[c]	Predicted sensing	Notes[d]
HK_01	42	547 ± 36	TMH-CHASE3-TMH-HAMP-HisKA-HATPase_c	At, Pr, Cy, B, Dt, F, Pl, V, Df	Extracytoplasmic	CHASE-CHASE6 sensor-like GenCon: in operon with RRs
HK_02	68	547 ± 32	TMH-(PAS)-TMH-PAS-HATPase_c	Pr, F, At, Sy, Sp, Df	Extracytoplasmic	CitA/DcuS-like GenCon: RR-HK-transporter
HK_03	730	507 ± 106	TMH-[50–300 aa]-TMH-HAMP-HisKA-HATPase_c	ND	Extracytoplasmic	Prototypical sensors
HK_04	30	409 ± 24	TMH-[60 aa]-TMH-HAMP-HisKA-HATPase_c	ND	Extracytoplasmic	NarX/Q-like
HK_05	68	382 ± 26	TMH-[30–40 aa]-TMH-HAMP-HisKA-HATPase_c	ND	Extracytoplasmic	PrmB-like
HK_06	26	374 ± 28	TMH-[25–30 aa]-TMH-HAMP-HisKA-HATPase_c	ND	Extracytoplasmic	VanS-like
HK_07	384	461 ± 108	TMH-HAMP-HisKA-HATPase_c	ND	Membrane?	
HK_08	244	419 ± 60	TMH_3-(HAMP)-HisKA-HATPase_c	ND	Membrane?	
HK_09	5	404 ± 117	$TMH_{0–5}$-HisKA-GerE	F, At, Ap, Pr	Membrane?	Overrepresented in *Firmicutes*
HK_10	35	724 ± 141	$TMH_{2–6}$-(HisKA)-HATPase-$TMH_{1–6}$	At	Membrane	
HK_11	71	435 ± 33	PspC-$TMH_{4–6}$-HATPase_c	At	Membrane	GenCon: downstream of *pspC*
HK_12	26	722 ± 234	TMH_9-HisKA-HATPase_c	ND	Membrane	
HK_13	6	735 ± 70	$TMH_{2–5}$-HisKA_3-(HATPAse_c)-$TMH_{3–5}$-HisKA-HATPase_c	ND	Membrane?	Restricted to *Actinomycetales*
HK_14	11	541 ± 52	TMH_4-HisKA-HATPase_c-TMH	ND	Membrane	Restricted to *Actinomycetales*
HK_15	4	701 ± 138	TMH_2-HAMP-HisKA-HATPase_c-$TMH_{11–14}$	ND	Extracytoplasmic or membrane	
HK_16	139	421 ± 108	TMH-[5–25 aa]-TMH-(HAMP)-HisKA-HATPase_c	ND	Membrane	LiaS-like
HK_17	937	416 ± 43	$TMH_{4/5}$-HisKA-HATPase_c-(TMH)	ND	Membrane	DesK-like
HK_18	162	441 ± 59	$TMH_{6/7}$-HATPase_c	ND	Membrane	ComD/ArgC fit descriptor
HK_19	1	664	TMH-7TMR_DISM_7TM-HisKA-HATPase_c	F, Pr, Sp, B, At	Membrane	
HK_20	5	841 ± 204	$TMH_{11/12}$-HisKA-HATPase_c-(TMH)	ND	Membrane	PutP-like
HK_21	41	606 ± 74	$TMH_{8/10}$-HisKA-HATPase_c	ND	Membrane	ComP-like
HK_22	2	680 ± 3	$TMH_{8/10}$-PAS-HisKA-HATPase_c	ND	Membrane	Restricted to *Propinibacterineae*
HK_23	2	795 ± 177	MASE1-(PAS_3)-HisKA-HATPase	At	Membrane/cytoplasm	
HK_24	11	567 ± 58	$TMH_{3–10}$-GAF-HisKA-HATPase_c	ND	Membrane/cytoplasm	
HK_25	2	721 ± 6	$TMH_{2/3}$-PAS-GAF-HisKA-HATPase_c	ND	Membrane/cytoplasm	Restricted to *Mycobacterium*
HK_26	12	881 ± 67	TMH_2-PAS-GAF-SpoIIE-HATPase_c	ND	Membrane/cytoplasm	Restricted to *Streptomyces*
HK_27	71	858 ± 30	KdpD-Usp-TMH_3-HisKA-HATPase_c	Pr, At, F, B, T	Cytoplasm	KdpD-like GenCon: part of *kdp* operon
HK_28	40	415 ± 73	PAS-(HisKA)-HATPase_c	ND	Cytoplasm	NtrB-like
HK_29	9	554 ± 106	$PAS_{2/3}$-(HisKA)-HATPase_c	ND	Cytoplasm	KinA-like
HK_30	3	798 ± 32	PAS-GAF-PHY-HisKA-HATPase_c	Pr, Cy, B, At, Pl, V	Cytoplasm	
HK_31	100	499 ± 12	H_kinase_N-PAS-HisKA-HATPase_c	At, F	Cytoplasm	
HK_32	5	1598 ± 5	Pkinase-AAA_16-(TRP_2/GAF_2)-HisKA-HATPase_c	F, At, Cy	Cytoplasm	
HK_33	30	485 ± 11	cNMP_binding-HATPase_c	At, Pr, Cy, Ad, B, V, Cf, Dt, F, Nt, Ar	Cytoplasm	GenCon: near thioredoxin reductase
HK_34	104	386 ± 103	HisKA-HATPase_c	At, F, Pr, B, Sp, Cy, Cf, V	Cytoplasm	
HK_35	806	186 ± 103	HATPase	F, At, Pr, Cf, Sp, Cy	Cytoplasm	
HK_36	7	217 ± 95	HisKA	At, F, Pr, Sp, B, Cy, Cf	Cytoplasm	
HK_37	5	249 ± 9	STAS-HATPase_c HATPAse-STAS	At, Pr, F, Sp, B, Tt, Cf	Cytoplasm	Overrepresented in actinobacteria GenCon: downstream of extracellular solute-binding protein and HK

(Continued on following page)

TABLE 2 (Continued)

Group identifier	No. of proteins	Length[a]	Protein domain architecture[b]	Taxonomical span[c]	Predicted sensing	Notes[d]
HK_38	170	514 ± 86	$GAF_{1/2}$-HisKA-HATPase_c	At, F, Pr, Cf, Cy, Dt, Sp	Cytoplasm	
HK_39	10	643 ± 88	PAS-$GAF_{1/2}$-(PAS)-HisKA-HATPase_c	At, Pr, Cf, Cy, V, F, Dt	Cytoplasm	
HK_40	170	790 ± 141	$PAS_{1/2}$-($GAF_{1/2}$)-SpoIIE-HATPase_c	At, F	Cytoplasm	
HK_41	26	815 ± 97	HATPase_c-(PAS)-$GAF_{1/2}$-SpoIIE	At, V, Pr	Cytoplasm	Overrepresented in actinobacteria
HK_42	29	486 ± 175	HATPase_c-SpoIIE	Pr, At, Dt, Cy, F, Ad, Pl, V, B	Cytoplasm	GenCon: part of operon of regulators of sigma B activity
HK_43	24	622 ± 146	SpoIIE-HATPase_c	At, Pr, F, B, Pl, Sp	Cytoplasm	Overrepresented in actinobacteria
HK_44	12	600 ± 74	TMH-PAS-HisKA-HATPase	ND	Cytoplasm	
HK_45	11	389 ± 84	(HAMP)-HisKA-HATPase	ND	Cytoplasm	
HK_46	167	979 ± 191	TMH-NIT-HAMP-HATPase	At	Cytoplasm	GenCon: part of operon encoding proteins of unknown function
HK_47	4	292 ± 53	HATPase_c-HTH	At	Cytoplasm	
HK_48	12	692 ± 134	(PAS/PAS-GAF-PAS)-SpoIIE-HATPase_c-STAS_2	At	Cytoplasm	
HK_49	3	739 ± 3	RsbU_N-PAS-SpoIIE-HATPase_c	At	Cytoplasm	GenCon: downstream of 4 anti-anti-sigma regulatory factors
HK_50	34	325 ± 29	MEDS-HATPase_c	At	Cytoplasm	
HK_00	12		TMH_2-CHASE3-TMH-HAMP-GAF-PAS-HisKA-HATPase_c HK_CA-GAF-PAS-HisKA-HATPase_c GAF-PAS_2-HisKA-HATPase_c HAMP-PAS-HisKA-HATPase_c MASE1-SpoIIE-HATPase_c HATPase_c-RRXRR-HNH HATPase-DUF3883 HATPase_c-PCMT HATPase-DUF4325 DNA_ligase_A_M-DNA_ligase_A_C-His_kinase SSF-HisKA-HATPase_c ABC_trans-HisKA-HATPase_c			Unique architectures

[a] Amino acids (mean ± standard deviation).
[b] Protein domain designations as in the Pfam database. aa, amino acids.
[c] Ad, *Acidobacteria*; Ar, *Armatimonadetes*; At, *Actinobacteria*; B, *Bacteriodetes*; Cf, *Chloroflexi*; Cy, *Cyanobacteria*; Df, *Deferribacteres*; Dt, *Deinococcus-Thermus*; F, *Firmicutes*; Nt, *Nitrospirae*; Pl, *Plantomyces*; Pr, *Proteobacteria*; Sp, *Spirochetes*; Sy, *Synergistales*; T, *Tenericutes*; Tt, *Thermotogae*; V, *Verrucomicrobia*; ND, not determined.
[d] GenCon, genomic context conservation.

tion of the domains can be found in Table S10 in the supplemental material). The second component, RRs, is defined as proteins that contain a receiver but not a transmitter domain. Also included as 2CSs are hybrid histidine kinases (HHKs) and hybrid response regulators (HRRs), which are proteins that have both transmitter and receiver domains. HHKs contain transmitter domains located N terminal to the receiver domain, while HRRs have transmitter domains located C terminal to the receiver domain. Chemotaxis (Che) proteins are specific types of 2CSs that are defined and classified according to the presence of conserved protein domains (e.g., CheW, CheB, or CheD) (11, 22). Finally, ECFs are members of the σ^{70} family of σ factors that contain only the conserved regions σ_2 and σ_4 (23).

Of a total of 51,138 STPs, 77% (39,590) are 1CSs, 5% of which (1,957) are membrane associated. Eighteen percent (9,141) of all the proteins are part of 2CSs, of which 54% (4,928) are HKs, 44% (4,032) are RRs, and 2% (181) are HHKs and HRRs. Only 0.4% (193) are chemotaxis-related proteins, while 4.3% (2,214) are ECF σ factors. While there is no strong relationship between the distribution of STPs and the organisms' morphologies, metabolisms, or habitats (see Table S1 in the supplemental material), a careful analysis of the distribution of STPs among taxonomical families revealed that the distribution is not homogeneous. While the families *Actinomycetaceae*, *Bifidobacteriaceae*, *Coriobacteriaceae*, and *Corynebacteriaceae* possess relatively low numbers of STPs, members of the families *Nocardiaceae*, *Pseudonocardiaceae*, and *Streptomycetaceae* are particularly STP rich (Fig. 1A and B). Given the genome size distribution within the respective families, this observation seems to be in line with previous reports describing a correlation between the genome size and the number of STPs (24,

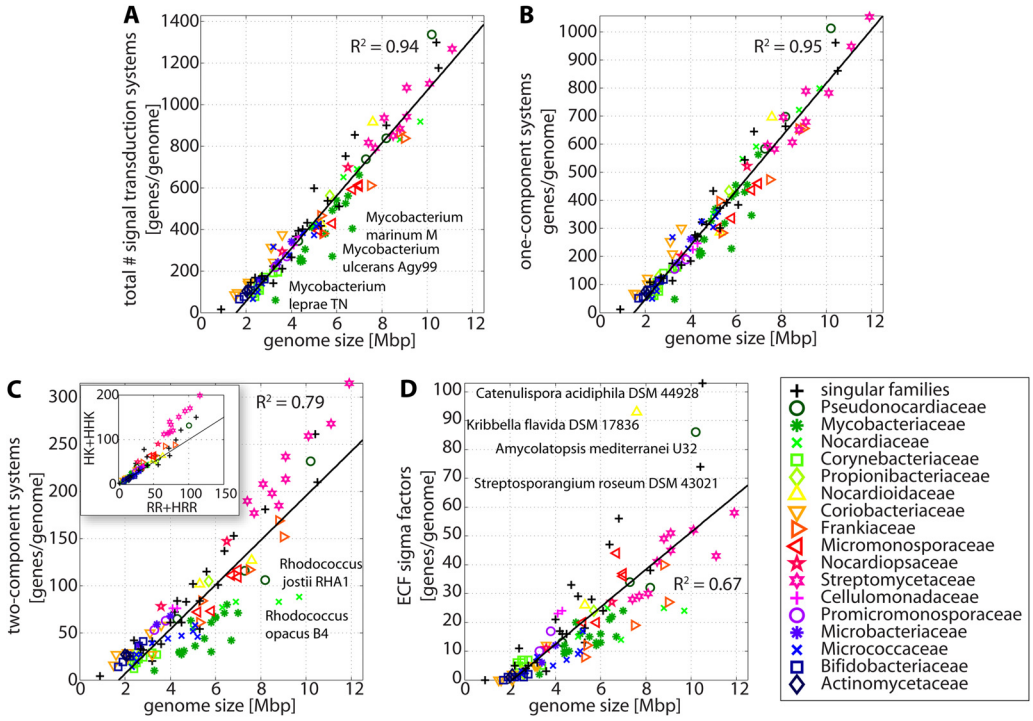


FIG 2 Correlation of STP numbers with genome sizes. Shown are scatter plots of the total numbers of STPs (A), 1CSs (B), 2CSs (C), and ECFs (D) encoded in a given genome as a function of the organism's genome size. The inset in panel C shows the correlation between HKs and RRs. In all cases, the total numbers of proteins are represented, and the genomes are color coded by taxonomical family. The black lines represent the best fit of a linear equation, and the points that deviate the most from that line are identified by the organism's name. In all panels, the size of the symbol is proportional to the organism's genome size.

25). Indeed, a positive and almost linear correlation (coefficient of determination [R^2] = 0.94) between the total number of STPs and the genome size was also observed in the actinobacterial genomes (Fig. 2A).

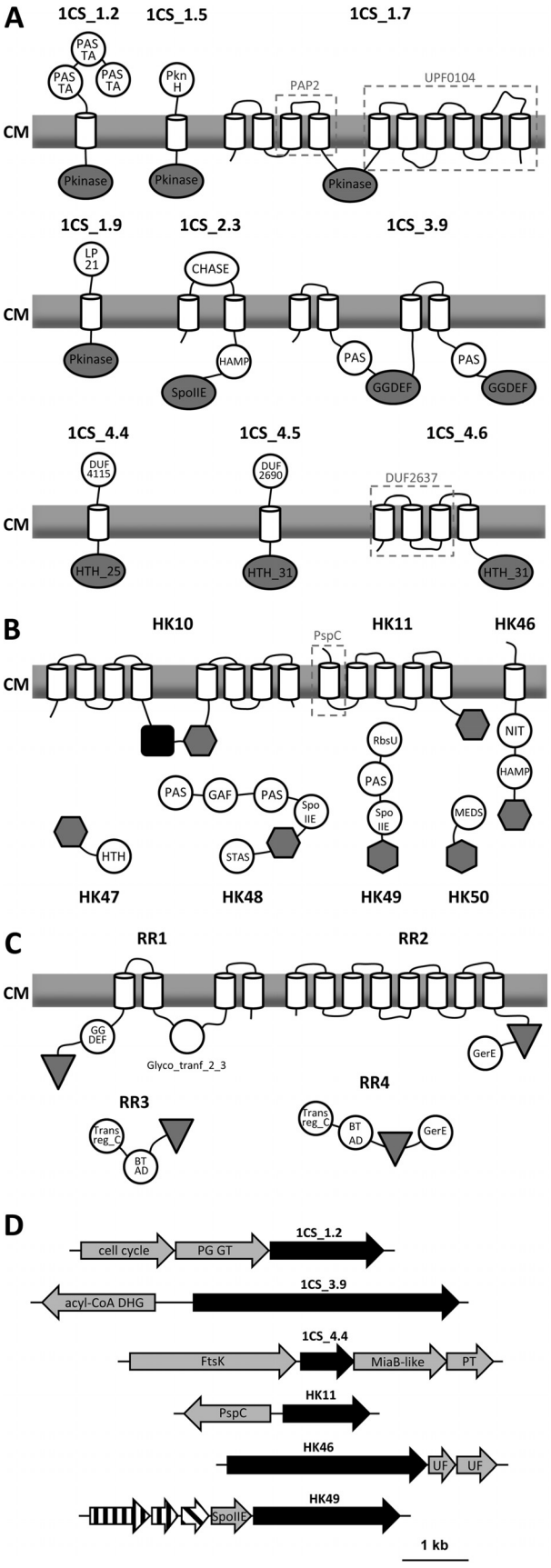
However, some *Mycobacterium* spp. (*Mycobacterium marinum* M, *Mycobacterium ulcerans* Agy99, and *Mycobacterium leprae* TN) deviate from this rule by harboring fewer STPs than were expected based on their genome sizes alone. This could be due to genome reduction during evolution, which might introduce a bias into the relationship between STPs and genome size as determined here and depicted in Fig. 2. For example, *M. leprae* is an obligatory intracellular pathogen that underwent a significant genome reduction during evolution, resulting in less than half of its genome consisting of functional genes (26).

To gain more detailed insight into the distribution of STPs, we subsequently analyzed the distribution of each type of STP. Given that the vast majority of STPs are 1CSs, an identical linear correlation (R^2 = 0.94) was observed between the genome size and the number of 1CSs (Fig. 2B), while this correlation was less well preserved (R^2 = 0.79) for 2CSs (Fig. 2C). This deviation is mainly caused by significantly increased numbers of HKs over RRs in 2CS-rich genomes (Fig. 2C, inset), indicative of an increased need and hence ability to integrate signals in more complex organisms. The *Mycobacteriaceae* and *Nocardiaceae* (e.g., *Rhodococcus jostii* RHA1 and *Rhodococcus opacus* B4) are 2CS-poor bacterial families. The weakest correlation with genome size (R^2 = 0.69) was observed for the ECFs (Fig. 2D), with four outliers that are particularly ECF rich relative to their genome sizes (*Catenulispora acidiphila*, *Kribbella flavida*, *Amycolatopsis mediterranei*, and *Streptosporagium roseum*). As a general trend, ECFs seem to be underrepresented (and often absent) in small genomes, while they tend to be enriched in organisms with large genomes. In fact, investigating the correlation between the different types of STPs revealed that for organisms with small genomes, the STPs are almost exclusively made up of 1CSs and 2CSs, while actinobacteria with larger genomes and hence more complex lifestyles start accumulating other types of STPs, particularly ECFs and chemotaxis-related proteins (see Fig. S1 in the supplemental material).

Below, we analyze each signaling principle separately, with a special focus on extracellular sensing. In doing so, we emphasize and highlight systems that are prominent in or even unique to the phylum *Actinobacteria*.

1CSs. One-component systems represent the most abundant and simplistic signaling principle in bacteria, because they combine the stimulus-perceiving input domain and the cognate output domain, which mediates the cellular response (1). The vast majority of 1CSs are soluble regulatory proteins that respond to intracellular cues. Given our focus on the perception of environmental signals, we restricted our analysis to membrane-associated 1CSs (i.e., those proteins that contain at least one TMH according to TMHMM analyses), which comprise approximately 5% (1,957) of all 1CSs.

These proteins were then further classified according to their domain architectures. Based on their output domains, about 18% of them are nucleic acid (mostly DNA) binding proteins and 27% are involved in second-messenger sensing. Remarkably, the remaining half of the membrane-associated 1CSs are involved in mediating protein modification as their predicted output, mostly



through functioning as Ser/Thr protein kinases (Table 1). Below, we highlight some prominent features of the identified 1CS groups.

Protein kinases. Protein kinases are the predominant type of membrane-anchored 1CSs in the *Actinobacteria*. Among them, Ser/Thr kinases are the most common type (about 80% of all protein kinases). Bacterial Ser/Thr kinases, similarly to their eukaryotic counterparts, can phosphorylate a myriad of substrates and thereby structure complex signaling networks involved in diverse cellular processes, e.g., pathogenesis (27), cell division (28), control of gene expression (29), stress response (30), and quorum sensing (31). Most bacterial genomes encode few (if any) Ser/Thr kinases. Higher numbers are found only in bacteria with more complex lifestyles, like the *Planctomycetes*, 20 to 45% of whose 1CSs are Ser/Thr kinases (32). While *Actinobacteria* are not as rich in protein kinases (with an average of only 2% of their 1CSs being protein kinases), some families are particularly protein kinase rich: *Frankiaceae* (5.5%), *Actinomycetaceae* (4.9%), *Nocardiopsaceae* (4.4%), *Bifidobacteriaceae* (4.2%), and *Microbacteriaceae* (3.1%).

Based on their domain architectures, nine different groups of protein kinases can be distinguished among the 902 such proteins in *Actinobacteria* (Table 1), with about 800 being associated with one of two major groups. Protein kinases of group 1CS_1.1 contain a variable number of TMHs in addition to the kinase domains. Such architectures are widespread in the bacterial world, and no functional predictions can be made with respect to the stimulus sensed or the physiological role these kinases play.

In contrast, the second-most-abundant group, 1CS_1.2, is well defined. These membrane-anchored kinases contain one to five extracellular PASTA domains (Fig. 3A; see Table S10 in the supplemental material) that are implicated in sensing cell wall components and regulate aspects of cell wall homeostasis and remodeling (33). The best-understood example of this group is PknB of *Mycobacterium tuberculosis*, which senses muropeptides and mediates the exit of cells from dormancy (34). Genes encoding 1CS_1.2 proteins are frequently preceded by genes encoding penicillin-binding proteins and the cell cycle protein FtsW (Fig. 3D). Such kinases are also found in the phyla *Bacteriodetes*, *Chloroflexi*, and *Firmicutes*.

In contrast, a number of minor protein kinase groups, containing only a few proteins each, are restricted to the *Actinobacteria*. Group 1CS_1.5 is restricted to the family *Mycobacteriaceae* and contains an extracellular PknH_C domain with unknown function (Fig. 3A; see Table S10 in the supplemental material). Such domains are also found in numerous other proteins, such as a

FIG 3 (A to C) Schematic representations of domain architectures of 1CSs (A), HKs (B), and RRs (C) that are found only in *Actinobacteria*. Cytoplasmic membranes (CM) are represented in gray, and TMHs are shown as cylinders. HK domains responsible for dimerization and phosphoacceptance are represented as squares, while ATPase domains are represented as hexagons. RR domains are represented as triangles. Other domains are represented by circles with their Pfam designations. (D) Relevant genomic context conservation. An ASF-coding gene is represented by the obliquely hatched arrow, while genes encoding anti-anti-sigma factors are represented by vertically hatched arrows. CoA, coenzyme A; DHG, dehydrogenase; PG GT, peptidoglycan glycosyltransferase; PT, CDP-diacylglycerol-3-phosphate 3-phosphatidyltransferase; UF, unknown function. The remaining acronyms represent Pfam domains for which a description can be found in Table S10 in the supplemental material.

number of lipoproteins from *M. tuberculosis*. They contain two conserved cysteine residues that likely form a disulfide bridge (13). Group 1CS_1.7 is restricted to the order *Actinomycetales* and is characterized by a cytoplasmic PAP2 and a membrane-integral UPF0104 domain (Fig. 3A; see Table S10 in the supplemental material). While the first provides a putative phosphatase activity, the second is uncharacterized but contains a highly conserved proline-glycine motif. Group 1CS_1.9 is restricted to the family *Frankiaceae* and contains an extracytoplasmic lipoprotein_21 domain with unknown function (Fig. 3A; see Table S10 in the supplemental material). This domain is also found in some lipoproteins from mycobacteria, including LppP, which is required for optimal growth of *M. tuberculosis* (information retrieved from Pfam [13]).

Protein phosphatases. Most of the 170 phosphatases are derived from two of the four distinct phosphatase groups that have been identified (Table 1). The domain architectures of both groups can be found in many other bacterial phyla, and no group-specific genomic context conservation was observed that could help shed light on the physiological roles of these proteins. The domain architecture of group 1CS_2.3 contains one sensory input domain (CHASE) and one signal transduction domain (HAMP), in addition to the phosphatase domain (SpoIIE) (Fig. 3A; see Table S10 in the supplemental material). Members of the phylum *Cyanobacteria* also encode such proteins. Members of the small group 1CS_2.4 contain a membrane-associated sensor domain (MASE1) (Fig. 3A; see Table S10 in the supplemental material) and are also found in the phyla *Cyanobacteria*, *Proteobacteria*, and *Spirochaetes* (Table 1).

Guanylate cyclases. Guanylate cyclases are the second-most-abundant type of membrane-anchored 1CSs in *Actinobacteria*. The 527 proteins can be subdivided into nine groups, three of which, 1CS_3.1 to 1CS_3.3, contain more than 100 members each (Table 1). The domain architectures of most groups can be found in many other bacterial phyla, and no group-specific genomic context conservation was observed. Only the smallest group, 1CS_3.9, is actinobacterium specific, and its five members are derived from the genus *Frankia* (Fig. 3A and Table 1).

Of the 119 actinobacterial genomes in our data set, 88 encoded membrane-associated guanylate cyclases (MAGCs). Of these, 58% have between one and four such enzymes, and 34% contain between five and nine. Only 17% of the analyzed genomes encode 10 or more MAGCs. The most dramatic example is *Kineococcus radiotolerans* SRS30216, which encodes 42 MAGCs.

The reasons why bacteria harbor multiple guanylate cyclases have been debated for many years. As reviewed previously (35, 36), enzymes involved in cyclic di-GMP (c-di-GMP) metabolism might be expressed at different times, as is the case for *Escherichia coli* YhjH and *Yersinia pestis* HmsT, involved in motility and virulence regulation, respectively. Moreover, such enzymes might have distinct localization patterns contributing to distinct local concentrations, as is the case for the *Salmonella* species curli regulator CsgG and those involved in regulation of the *Caulobacter crescentus* cell cycle. Finally, as is also apparent in our classification (Table 1), guanylate cyclase DGGE domains might cooccur with different signal input domains (e.g., PAS, GAF, or MASE1), which might also reflect a higher potential for signal integration.

DNA-binding 1CSs represent the third-most-abundant type of membrane-anchored 1CSs in the *Actinobacteria*. The 357 proteins fall into seven groups, with more than half of all the proteins containing a GerE output domain (group 1CS_4.1) (Table 1; see Table S10 in the supplemental material). Group 1CS_4.4 is characterized by an N-terminal intracellular HTH_25 output domain and an extracellular C-terminal domain with unknown function (DUF4115). The two domains are separated by a single TMH (Fig. 3A; see Table S10 in the supplemental material). This domain architecture can be found in many bacterial phyla, but only the 19 actinobacterial members additionally share a genomic context: genes encoding 1CS_4.4 proteins are frequently preceded by genes encoding FtsK-like DNA translocases. Two downstream genes, encoding MiaB-like 2-methylthioadenine synthetases and CDP-diacylglycerol-3-phosphate 3-phosphatidyltransferases, are potentially cotranscribed (Fig. 3D). The physiological relevance of this conservation remains to be determined. Two small actinobacterium-specific groups of HTH_31-containing 1CSs, 1CS_4.5 and 1CS_4.6, differ in their putative input domains. While the first contains an extracellular DUF2690 domain with unknown function, the latter contains a membrane-embedded DUF2637 domain for perceiving a stimulus at or within the membrane interface (Fig. 3A; see Table S10 in the supplemental material). Both groups lack any genomic context conservation.

Only a single membrane-anchored RNA-binding 1CS (with a C-terminal ANTAR output domain) can be found in all 119 actinobacterial genomes (Table 1; see Table S10 in the supplemental material).

2CSs. In contrast to 1CSs, which combine input and output domains on a single polypeptide chain, these two domains are separated on two different proteins for 2CSs. The stimulus-perceiving input domain is usually located at the N-terminal end of HKs, while the output domain can be found at the C-terminal end of the cognate RR. Signal transduction requires specific communication between the two partner proteins, which is based on a phosphoryl group transfer; upon stimulus perception, ATP-dependent autophosphorylation of a highly conserved histidine residue, located in the DHp (dimerization and histidine phosphorylation) domain, is mediated by a C-terminally located HATPase_c catalytic domain (see Table S10 in the supplemental material). Together, the DHp and HATPase_c domains form the transmitter unit that characterizes HKs of 2CSs (5). Phosphohistidine then serves as a phosphodonor for activating the cognate RR at an invariant aspartate residue located in the N-terminal REC (receiver) domain (see Table S10 in the supplemental material). This phosphotransfer usually results in dimerization of the RRs, thereby activating the C-terminal output domain. While most RRs are transcriptional regulators, a variety of (often homologous) output domains similar to that observed for 1CSs can also be found in RRs (2). The separation of input and output on two proteins simplifies the response to extracellular cues and also allows signal integration and amplification processes in more complex regulatory cascades, best exemplified by the 2CS-dependent phosphorelay that orchestrates the commitment to sporulation in *Bacillus subtilis* (37). Accordingly, over 50% of 2CSs are predicted to connect environmental stimuli with cellular responses, based on the presence of extracytoplasmic input domains. This estimation is derived from a comprehensive analysis of the input domain architectures of over 4,500 HKs (5). Nevertheless, many 2CSs are also employed in responding to cellular cues (5). Over 9,100 proteins were extracted from the MiST databases as part of 2CSs. Of these, 55% represent HKs, while the remaining 45% are classified as RRs (see Table S9 in the supplemental material). In order to

further understand two-component signaling in actinobacteria, we looked in detail for each component (HKs and RRs) individually. For both types of proteins, an approach similar to that outlined for the membrane-anchored 1CSs was applied: the proteins were grouped based on their domain architectures (Table 2; see Tables S3 and S4 in the supplemental material). For each group, the phylogenetic distribution was analyzed to identify actinobacterium-specific groups. Such groups of HKs and RRs are presented below.

Histidine kinases. Of the 4,928 HKs extracted from the MiST database, 4,916 could be classified into 50 groups based on their domain architectures (Table 2; see Table S4 in the supplemental material). Some of these groups were found only in actinobacterial proteins and contain unusual domain architectures, as illustrated in Fig. 3B and described below.

In the 35 members of the actinobacterium-specific group HK10, the transmitter unit is flanked on both sides by 2 to 6 TMHs (Fig. 3B). The architecture may suggest that these HKs have a sensing mechanism linked to these transmembrane regions. The mechanistic reason for this unusual domain architecture, particularly the function of the C-terminal TMHs, remains to be identified.

The input domain of group HK11 proteins also contains 4 to 6 TMHs, indicative of a membrane-associated sensing mechanism. Remarkably, the N-terminal TMH represents a conserved phage shock protein C (PspC) domain (Fig. 3B; see Table S10 in the supplemental material). In proteobacteria, PspC is indirectly involved in transcriptional (auto)regulation of its encoding *pspABCDE* operon. The resulting PSP response plays a significant role in competition for survival under nutrient- or energy-limited conditions (38). Another PspC-like protein is encoded by an upstream and divergently oriented gene, a genomic context that is conserved in most of the 71 members of HK11 (Fig. 3D). Together, these observations indicate that HK11 proteins may play an important role in orchestrating the phage shock protein-like response of *Actinobacteria*. This might represent a novel mechanism that seems to combine the function of proteobacterial PspC proteins as sensors/membrane anchors with control of the PSP-like Lia response of *Firmicutes* bacteria by unique LiaRS-like 2CSs (39, 40).

Proteins of the large (167-member) and actinobacterium-specific group HK46 are anchored to the membrane by an N-terminal TMH. Their unifying hallmark feature is a cytoplasmic NIT domain (see Table S10 in the supplemental material) located directly C terminal to the TMH, which is normally associated with microbial responses to nitrate and nitrite (41). Additionally, genes encoding HK46 proteins are cotranscribed with genes encoding conserved proteins with unknown functions.

The actinobacterium-specific groups HK47 to HK50 all represent soluble HK-like proteins that contain only a HATPase_c domain but seem to lack the His-containing DHp domain (Fig. 3B). This architecture indicates a cytoplasmic sensing mechanism and potentially the phosphorylation of a DHp-containing partner protein, which remains to be identified. The three HK47 proteins contain a C-terminal DNA-binding output domain. Group HK48 is restricted to the order *Actinomycetales*, and the 12 proteins show a rather complex domain architecture, including PAS and GAF domains (see Table S10 in the supplemental material) that might play a sensory role (42, 43). The presence of a SpoIIE domain (found in phosphatases, adenylate cyclases, and sporulation proteins [44]) and a STAS_2 domain (often present in the C-terminal region of sulfate transporters and ASF antagonists [45]) (see Table S10 in the supplemental material) might indicate a unique sensing and signaling mechanism for group HK48 proteins that remains to be investigated.

The presence of SpoIIE and RsbU domains in group HK49 proteins (Fig. 3B; see Table S10 in the supplemental material) points to a role of these HKs in more complex phosphorelay cascades, e.g., in differentiation and/or general stress responses that might also involve alternative σ factors. This HK group is restricted to the order *Actinomycetales*. The hallmark feature of group HK50 proteins is the presence of a MEDS (methanogen/methylotroph DcmR sensory) domain (see Table S10 in the supplemental material) that likely functions in sensing hydrocarbon derivatives (46).

Response regulators. The 4,042 actinobacterial RRs extracted from the MiST database were also analyzed and grouped according to their output domains (see Table S3 in the supplemental material). Eighty percent of all RRs contain either GerE or Trans_reg_C output domains (see Table S10 in the supplemental material). The first is a LuxR-type DNA-binding helix-turn-helix domain, while the second is a C-terminal transcription-regulatory domain that also plays a role in DNA binding (44). Some actinobacterium-specific types of RRs with unusual domain architectures are illustrated in Fig. 3C. They include two membrane-anchored RR types (RR1 and RR2) and two groups of soluble RRs characterized by the presence of an N-terminal Trans_reg-C domain and an additional bacterial transcriptional activator (BTAD) domain (see Table S10 in the supplemental material), which can be found in the DnrI/RedD/AfsR family of transcriptional regulators (47). The regulatory mechanisms of such unique types of RRs remain to be determined.

Chemotaxis proteins. Chemotaxis is a special form of 2CS-dependent regulation that is characterized by a unique type of CheA-like HK and a number of typical protein domains restricted to chemotaxis regulation, including CheW, CheZ, and CheR. All the proteins containing such domains are classified as chemotaxis proteins and have been extracted from the MiST database (see Table S9 in the supplemental material). An earlier study, based on only 17 actinobacterial genomes, concluded that chemotaxis proteins are absent from this phylum (2). Our own analysis of 119 actinobacterial genomes confirms that chemotaxis proteins are indeed very rare. Nevertheless, a few noteworthy exceptions to this rule could be identified and are briefly discussed below.

The genomes of five motile actinobacterial species (*Conexibacter woesei* DSM 14684, *K. radiotolerans* SRS30216, *Jonesia denitrificans* DSM 20603, *Cellulomonas fimi* ATCC 484, and *Mobiluncus curtisii* ATCC 43063) encode complete sets of chemotaxis proteins, and their genes are located in the immediate vicinity of flagellar operons. Hence, these proteins might be involved in chemotactic motility. In addition, *Nocardioides* sp. strain JS614 also contains a complete set of chemotaxis proteins, even though the organism has been described as being nonmotile (48). Additionally, a number of nonmotile actinobacteria contain relatively high numbers of chemotaxis-related STPs, e.g., 21 in *Sanguibacter keddieii* DSM 10542. This observation raised questions regarding the function and functionality of these chemotaxis proteins in these organisms. Three explanations can be envisaged.

First, an incomplete set of chemotaxis proteins might represent an intermediate of reductive evolution. Potentially, these species derive from a motile ancestor. After they assumed a sessile lifestyle later in evolution, chemotaxis proteins were no longer required and thus were gradually lost. If this assumption is true, this should

result in the complete loss of genes encoding chemotaxis-related functions.

Second, such chemotaxis proteins might have acquired new regulatory functions that are no longer associated with motility. A few such cases have been described in the literature. The *che3* operon of *Myxococcus xanthus* is required for differentiation (49). Moreover, it was suggested that one out of the four *che* operons of *Pseudomonas aeruginosa* regulates some pathogenicity genes, while one *che* operon of *Pseudomonas fluorescens* seems to be involved in cellulose biosynthesis (50). In actinobacteria, we identified several examples that might be in line with this idea. Here, eight nonmotile organisms (*Pseudonocardia dioxanivorans*, *A. mediterranei*, and six mycobacteria) harbor only CheB- and/or CheR-like proteins, and the corresponding genes are located next to genes encoding STAS domain-containing proteins (see Table S10 in the supplemental material). This domain is found in sulfate transporters and bacterial anti-σ factor antagonists (45), suggesting that these CheB/CheR proteins might be involved in methyl-mediated signal transduction processes unrelated to motility.

Third, potentially missing chemotaxis proteins might be only weakly conserved and hence misannotated. In these cases, the organisms would indeed use chemotaxis-related proteins for regulating their motility. Because of the high conservation of chemotaxis pathways in bacteria, this hypothesis is probably the least likely but nevertheless cannot be ruled out.

Extracytoplasmic-function σ factors. ECFs represent the third pillar of bacterial signal transduction, with an average of six ECFs per bacterial genome (6). We previously analyzed more than 2,700 predicted ECFs from 369 microbial genomes belonging to 11 different phyla and could define 67 ECF groups based on the sequence similarity and domain architecture of both the ECFs and their cognate anti-σ factors, genomic context conservation, and putative target promoter motifs (6). Nevertheless, numerous actinobacterial ECFs could not be classified at the time. The significantly increased number of genomes available now inspired a phylum-specific reanalysis of actinobacterial genomes for the present analysis. We could identify 2,203 ECFs in our collection of 119 actinobacterial genomes. Of these, 76% (1, 677) belonged to one of the ECF groups defined in our initial study (6), while the remaining 24% (526) could not be classified. These were then subjected to further in-depth analyses, as described below.

ECF distribution and abundance. In actinobacteria, the most abundant ECF groups are ECF01, ECF39, ECF41, and ECF42. Together, they account for more than half of all actinobacterial ECFs (Fig. 4; see Table S5 in the supplemental material) (6). Eight ECF groups (ECF14, ECF17, ECF19, ECF27, ECF36, ECF38, ECF39, and ECF40) could be found exclusively in the phylum *Actinobacteria* (6). They represent almost 20% of all actinobacterial ECFs.

Members of the *Corynebacteriaceae*, *Coriobacteriaceae*, *Microbacteriaceae*, *Micrococcaceae*, *Bifidobacteriaceae*, and *Actinomycetaceae* have low numbers and diversity of ECFs (Fig. 1). For example, members of the families *Coriobacteriaceae* and *Bifidobacteria* possess ECFs of only 2 ECF groups: ECF01 (both) and ECF30 (*Coriobacteriaceae*) or ECF12 (*Bifidobacteria*). In contrast, *Nocardioidaceae*, *Streptomycetaceae*, and *Catenulisporaceae* are particularly ECF-rich families, with ECFs belonging to 15 to 20 different ECF groups (Fig. 1).

Identification and classification of novel ECF groups. Given that a quarter of all actinobacterial ECFs could not be assigned to any of the initially defined ECF groups (see Table S5 in the supplemental material), we aimed to classify them. A strategy identical to the one used previously (6), relying on sequence similarity of the ECFs and genomic context conservation, was pursued. We defined 18 new ECF groups (10 groups with more than 10 ECF sequences [ECF47 to ECF56] and 8 "minor" groups [ECF125 to ECF132], each containing less than 10 sequences). This allowed us to classify 427 of the 526 ECFs not covered by the previous classification (6). Hence, only about 4% of all actinobacterial ECFs remain unclassified (Fig. 4).

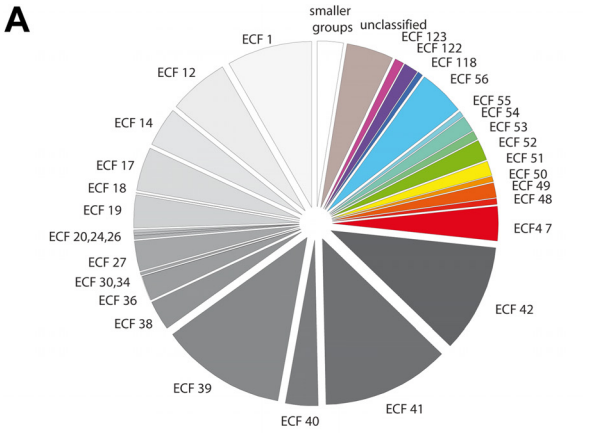
The vast majority of the novel groups identified here are taxonomically restricted to the *Actinobacteria* (see Table S11 in the supplemental material). The exceptions are ECFs of groups ECF55, ECF56, and ECF127, which are also found in other phyla, e.g., the *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, and *Chloroflexi*.

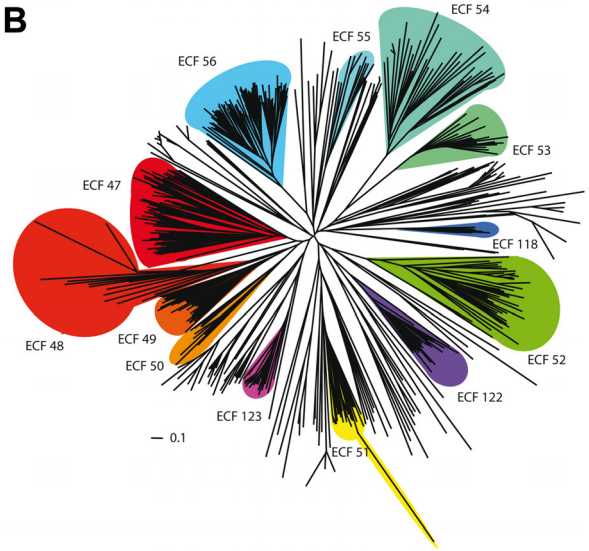
Descriptions of novel ECF groups. (i) ECF47. ECFs of group ECF47 occur only in the *Actinobacteria* but are widely distributed within the phylum, having been identified in 19 out of the 39 actinobacterial families analyzed (Fig. 1; see Table S11 in the supplemental material). Genes encoding these ECFs are putatively cotranscribed with their cognate ASFs, but genomic context conservation does not go beyond the σ/anti-σ pair.

The ASFs are membrane associated via a putative alanine- and valine-rich transmembrane helix that shows low similarity to that of RskA (regulator of ECF19 SigK) proteins, as reflected by a hit of the Pfam domain RskA (see Table S10 in the supplemental material) and by the multiple-sequence alignment shown in Fig. 5. Moreover, it was possible to identify an N-terminal anti-σ domain (ASD), which is a structural motif with reduced sequence conservation (only 21% sequence identity over 63 aligned residues), as defined based on structural studies of *E. coli* RseA and *Rhodobacter sphaeroides* ChrR (51). This domain is commonly found in both membrane-associated and soluble ASFs and is involved in σ/anti-σ interaction. In contrast to the N-terminal regions, the C-terminal periplasmic regions of these ASFs are very diverse both in sequence and in predicted secondary structure (Fig. 5; see Fig. S2 in the supplemental material).

(ii) ECF48. ECFs of group ECF48 are also restricted to the phylum *Actinobacteria* (see Table S11 in the supplemental material). They possess long C-terminal extensions that are unusual, since they contain one putative transmembrane helix (Fig. 6). Thus, ECF48 proteins represent membrane-associated ECFs, a feature that has so far been found only in the planctomycete-specific group ECF01-Gob (32). Moreover, this C-terminal extension contains a conserved HXXXCXXC sequence motif (Fig. 6B) characteristic of the zinc-containing anti-σ (ZAS) domain (see Table S10 in the supplemental material), in which the two conserved cysteine residues usually coordinate a zinc ion. Upon zinc release, a disulfide bond is formed, causing a drastic conformational change in the ASF, ultimately leading to the release of the σ factor (52). The ZAS domain can be either redox sensitive or insensitive, which is mainly determined by the identities of the amino acid residues flanking the two conserved cysteine residues (53). In the case of the ECF48-associated ZAS domains, these flanking amino acid residues do not support redox sensing. Additionally, ECF48 proteins contain long (208- ± 53-amino-acid) putatively periplasmic regions that are proline rich (Fig. 6); their relevance remains to be identified.

The domain architecture of ECF48 proteins can be interpreted in two different ways. Stimulus perception by the extracytoplas-

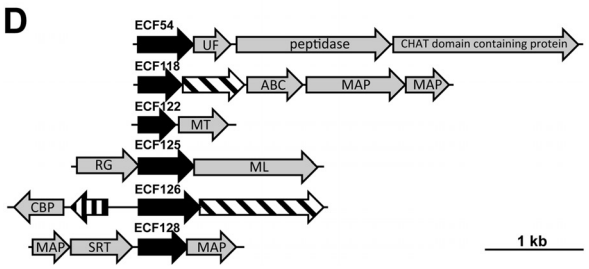




C

ECF group	Regulation of ECF σ factors' activity	
	Type	Conserved domains
ECF47	ASF	1 TMH + RskA
ECF48	CTE	ZAS + 1 TMH
ECF49	ASF	1 TMH + RskA
ECF50	ASF	1 TMH
ECF51	ASF	1 TMH + RskA
ECF52	CTE	ZAS + 0-6 TMH
ECF53	CTE	ZAS + CAP/Glyco_hydro_cc/AbfB/Beta_helix
ECF54	ND	NA
ECF55	ND	NA
ECF56	CTE	SnoaL_2
ECF118	ASF	ZAS + 6 TMH
ECF122	ND	NA
ECF123	ASF	1 TMH

ND - not determined; NA – not applicable; ASF – anti-σ factor; CTE – C-terminal extension; TMH – transmembrane helix



mic C-terminal domain may result in a conformational change that is then transduced through the membrane to activate the cytoplasmic ECF output domain. In this case, the ECF would stay intact and mediate transcription initiation from its site at the membrane. Alternatively, a stimulus could trigger regulated proteolysis to release the cytoplasmic ECF domains for transcription initiation. This hypothesis is supported by a recent report demonstrating regulated proteolysis for an unusual membrane-anchored ECF07 protein from *Pseudomonas putida* PP2192, which is cleaved at the TM domain by RseP, releasing an active soluble ECF into the cytoplasm (54). Future experimental studies will be required to distinguish between these two possibilities.

(iii) ECF49 to ECF51. ECFs of groups ECF49 to ECF51 are exclusively present in *Actinobacteria* (see Table S11 in the supplemental material), and the vast majority of those of ECF51 are found in *Micromonosporaceae* and *Streptomycetaceae* (Fig. 1). These σ factors are putatively cotranscribed with their cognate ASFs, but genomic context conservation does not go beyond the σ/anti-σ pair. The putative cognate ASFs are likely membrane associated, and in the cases of groups ECF49 and ECF51, the alanine- and valine-rich transmembrane helix shows weak similarity to RskA proteins (Fig. 5). The remaining regions of the protein are overall very diverse in terms of sequence and predicted secondary structure (Fig. 5; see Fig. S2 in the supplemental material).

(iv) ECF52. ECFs of group ECF52 occur only in the phylum *Actinobacteria* (see Table S11 in the supplemental material). They possess long C-terminal extensions that, similar to what has been described for ECF48, contain a redox-insensitive ZAS domain with its characteristic HXXXCXXC signature. However, these ECFs contain variable numbers of transmembrane helices (between one and six, as identified by TMHMM Server 2.0) and a long (397- ± 257-amino-acid) proline-rich C-terminal extension (Fig. 6). In this region, we identified carbohydrate-binding domains (e.g., PF08305 and PF00553) (see Table S10 in the supplemental material) that, together with the genomic localization of their encoding genes in the vicinity of genes encoding proteins involved in carbohydrate metabolism (e.g., glycosyl transferases, xylanases, or pyruvate dehydrogenases), might suggest a role of ECF52 σ factors in regulation of a certain aspect of carbohydrate metabolism.

(v) ECF53. ECF53 σ factors have a very narrow taxonomical distribution and are found almost exclusively in organisms that belong to the family *Streptomycetaceae* (Fig. 1). These σ factors constitute an unusual group, since they possess a conserved σ_2 region but not a well-conserved σ_4 region (Fig. 6). Moreover, a

FIG 4 Classification of actinobacterial ECFs. (A) Distribution of actinobacterial ECFs into old and newly defined groups. The proportions of ECFs now classified into groups containing fewer than 10 proteins (white) and those that remain unclassified (brown) are also represented. (B) Phylogenetic tree of previously unclassified ECFs and those of groups ECF118, ECF122, and ECF123 created from a gapless multiple-sequence alignment of regions σ_2 and σ_4. Shading following the same color code as for panel A highlights each branch that represents a new group. (C) Putative ways in which the activities of ECFs of each group are regulated. (D) Genomic conservation in ECF groups. Genes encoding ECFs are represented in black, putative ASFs by diagonal hatching, and a putative anti-anti-σ factor by vertical hatching. ABC, ABC transporter; CBP, calcium-binding protein; MAP, membrane-associated protein; MT, methyltransferase; RG, transcriptional regulator; SRT, sortase; UF, unknown function. Only groups containing more than 10 proteins are represented. See the text for details.

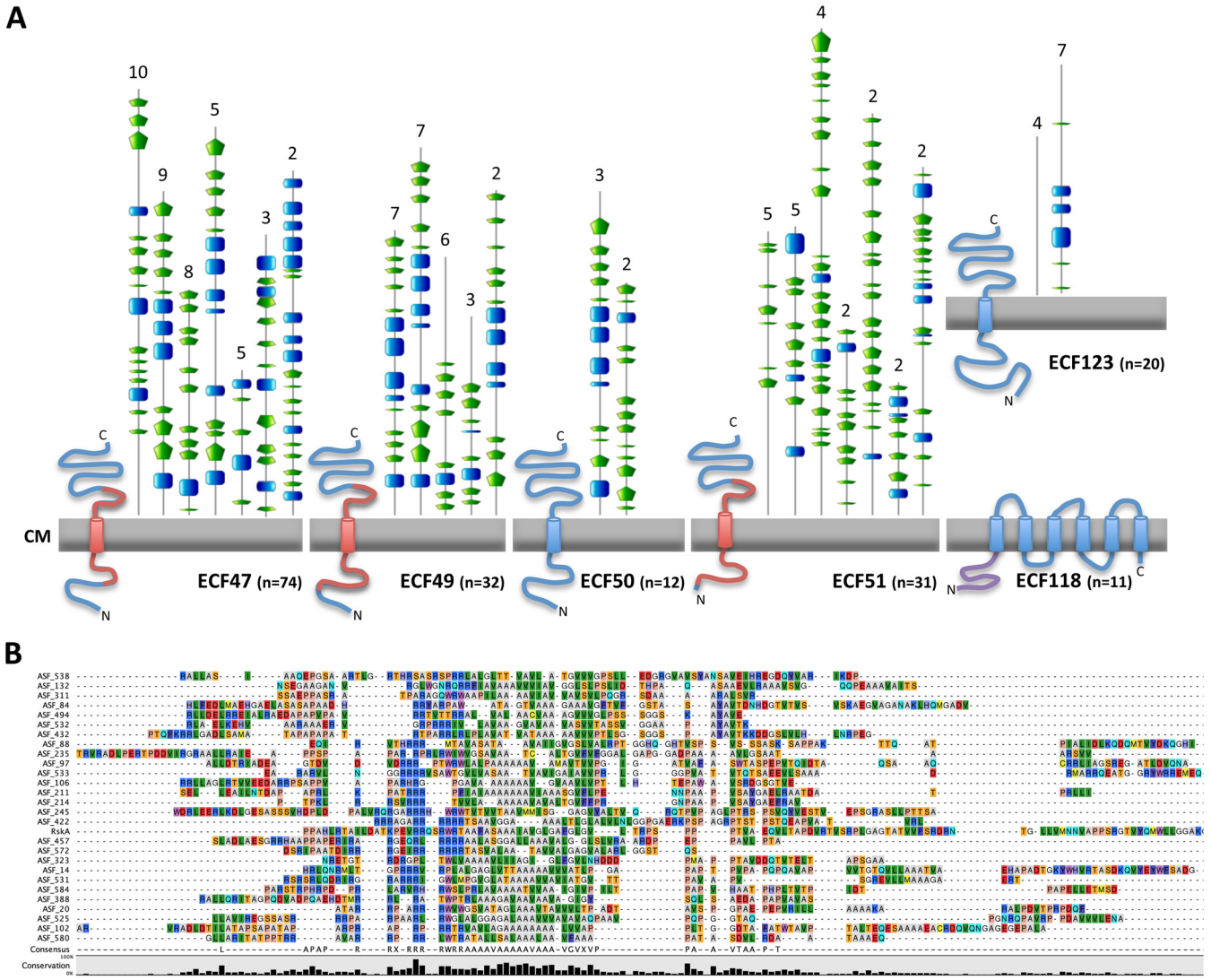


FIG 5 Putative ASFs. (A) Schematic representation of the architecture of the putative ASFs. Cytoplasmic membranes (CM) are represented in gray, and predicted transmembrane helices as cylinders. The secondary structures of the predicted periplasmic regions are shown above. The blue squares represent helices, and the green pentagons represent strands. The numbers at the top are the numbers of proteins showing that secondary structure. RskA domains are colored red, and ZAS domains are colored purple. The total number of ECFs in each group is given in parentheses. For clarity, only groups containing more than 10 proteins are represented. (B) Multiple-sequence alignment of ASFs containing RskA domains. The sequences are identified by the ASF numbers (see Table S7 in the supplemental material), and the degree of sequence conservation is represented in a bar graph at the bottom. Amino acid residues are colored with RasMol colors, and gaps are represented by dashes.

redox-insensitive ZAS domain (Fig. 6; see Table S10 in the supplemental material) was identified in their C-terminal extension. Beyond this, these extensions are highly variable between different ECF53 proteins and may contain one of several different domains with predicted enzymatic activities, e.g., glycosyl hydrolase catalytic core domains or alpha-L-arabinofuranosidase B domains (see Table S10 in the supplemental material). This observation again points to a potential role of these ECFs in regulating carbohydrate metabolism.

(vi) ECF54. ECF54 proteins are restricted to the phylum *Actinobacteria* (see Table S11 in the supplemental material). Their genes are located either upstream or downstream of a small gene encoding a protein containing a carboxypeptidase-regulatory-like domain, a gene encoding a peptidase S8/S53, and a larger gene encoding a protein containing a C-terminal CHAT domain and three weakly conserved N-terminal tetratricopeptide repeats (Fig. 4). The CHAT domain (see Table S10 in the supplemental material) appears to be related to peptidases (information retrieved from InterPro [44]), and the tetratricopeptide repeats are involved in protein-protein interactions (55). Two scenarios can be hypothesized: either (i) the carboxypeptidase-regulatory-like domain protein is involved in the regulation of the ECF by means of the peptidase encoded adjacently or (ii) it is involved only in the regulation of the peptidase and the proximity to the ECF-encoding gene reflects only that these genes are under the transcriptional control of that ECF.

(vii) ECF56. ECFs of group ECF56 can be found in the phyla *Actinobacteria*, *Proteobacteria*, and *Gemmatimonadetes* and are

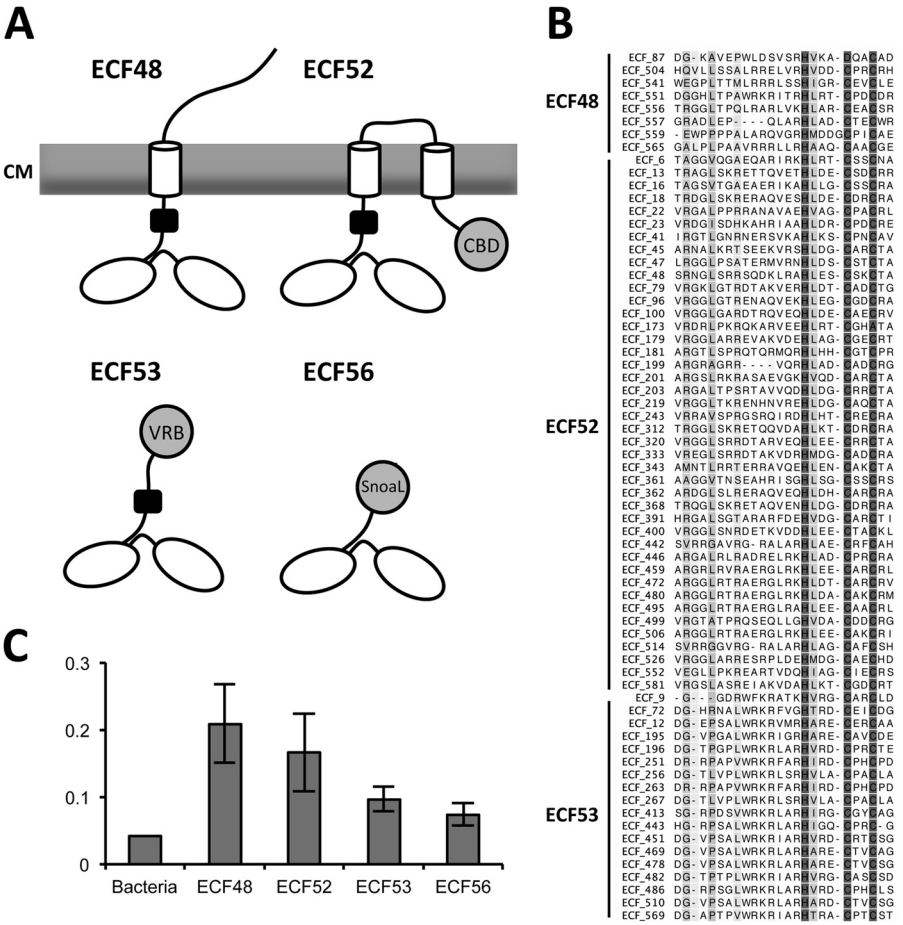


FIG 6 ECFs containing C-terminal extensions. (A) Schematic representation of the architecture of ECFs containing C-terminal extensions. The σ_2 and σ_4 regions are represented as ovals; the ZAS domain is represented as a black box, TMHs are shown as cylinders, and other domains are shown as circles. CBD, carbohydrate-binding domain; CM, cytoplasmic membrane; SnoaL, SnoaL_2 domain; VRB, variable domain. (B) Multiple-sequence alignment of the ZAS domain (PF13490) identified in members of the groups ECF48, ECF52, and ECF53. The degree of sequence conservation is color coded from white (no conservation) to dark gray (full conservation), clearly revealing the HXXXCXXC motif putatively responsible for zinc binding. (C) Proline frequency in bacteria (average) and in the C-terminal extensions of ECFs of groups ECF48, ECF52, ECF53, and ECF56. The error bars denote standard deviations.

consistently present in 22 out of the 39 actinobacterial families analyzed (Fig. 1; see Table S11 in the supplemental material). ECF σ factors of this group are small (336- ± 21-amino-acid) proteins with the characteristic conserved σ_2 and σ_4 regions followed by a SnoaL_2 domain (Fig. 6; see Table S10 in the supplemental material). This domain was originally described in SnoaL-like proteins, which are polyketide cyclases involved in biosynthesis of nogalamycin, an anthracycline antibiotic produced by *Streptomyces nogalater* (56) and in a large number of other bacterial sequences (information retrieved from InterPro [44]). Additional genomic context conservation could not be identified for this ECF group.

(viii) ECF125. ECF125 σ factors are restricted to *Actinobacteria* (see Table S11 in the supplemental material). Their genes are located upstream of a gene encoding a putative metalloprotein from the "glyoxalase/bleomycin resistance protein/dioxigenase" superfamily (Fig. 4). Such metalloproteins can be involved in peptide antibiotic resistance (57) and detoxification of metabolic subproducts (58). A putative cognate ASF was not identified for ECF125 σ factors. Instead, a transcriptional regulator of the TetR family is frequently encoded in the vicinity of ECF125 genes.

(ix) ECF126. ECF126 σ factors can be found only in *Actinobacteria* (see Table S11 in the supplemental material). Their encoding genes are frequently located in the immediate vicinity of those encoding their cognate ASFs, a putative anti-anti-σ factor and a calcium-binding protein (Fig. 4). Their cognate ASFs are long (447- ± 58-amino-acid) and soluble proteins containing an N-terminal redox-insensitive ZAS domain and, C-terminally, a domain similar to the N-terminal domain of the mycothiol maleylpyruvate isomerase.

(x) ECF127 and ECF128. ECF127 σ factors can be found in *Actinobacteria* and *Chloroflexi*, while the ECF128 σ factors are restricted to *Actinobacteria* (see Table S11 in the supplemental material). Genes encoding putative ASFs were not identified. Instead, ECF127 genes are located upstream of genes encoding Rieske proteins (see Table S10 in the supplemental material), which are iron-sulfur proteins of cytochrome complexes (59), and ECF128 genes are surrounded by genes encoding two membrane-associated proteins with unknown functions and a sortase (Fig. 4).

(xi) ECF130 and ECF132. Both groups ECF130 and ECF132 are restricted to *Actinobacteria*, and their genes are linked to those encoding the cognate ASFs (see Table S11 in the supplemental material). While the ASFs of group ECF130 are small soluble pro-

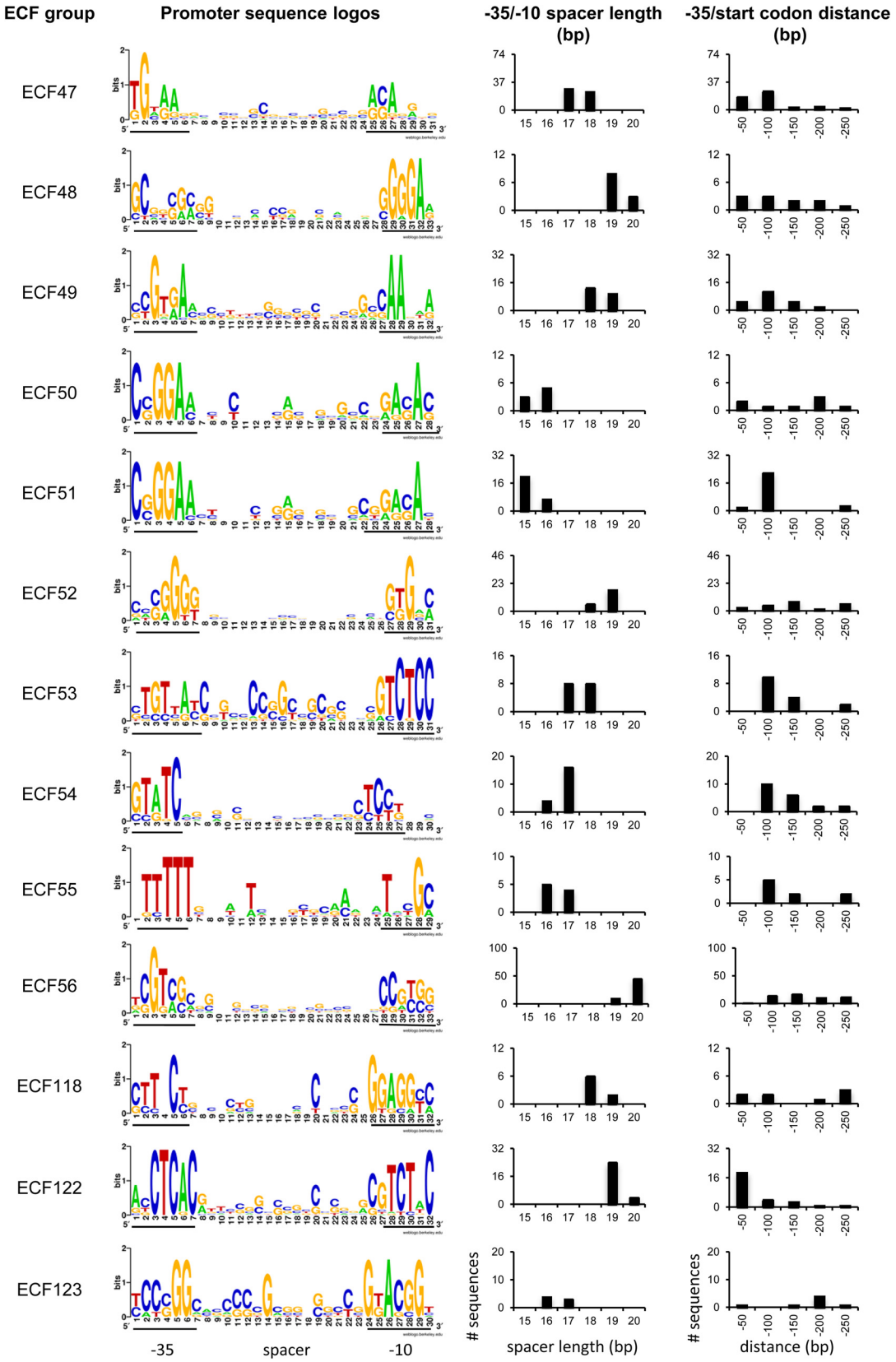


FIG 7 Putative ECF target promoters. Shown are sequence logos illustrating the −35 and −10 motifs, as well as the corresponding spacer sequences. The exact motifs identified by BioProspector are underlined beneath each logo. The bar charts represent the distributions of spacer lengths found in the identified promoters and the distance between the most upstream residue of the −35 motif and the start codon. The categories are as follows: −50, distances between 0 and −50; −100, distances between −51 and −100; −150, distances between −101 and −150; −200, distances between −151 and −200; −250, distances between −201 and −250. Target promoters were predicted only for groups containing more than 10 proteins and whose putative target promoter motifs were not identified previously.

teins with an ASD, those of ECF132 are membrane-associated proteins with an N-terminal ASD and a proline-rich C terminus.

(xii) ECF55, ECF129, and ECF131. ECFs of group ECF55 can be identified in the phyla *Actinobacteria*, *Firmicutes*, and *Bacteroidetes*. However, in *Actinobacteria*, they are restricted to the family *Coriobacteriaceae* (Fig. 1). ECFs of groups ECF129 and ECF131 are restricted to *Actinobacteria* (see Table S11 in the supplemental material). For all the groups, no conserved genomic context was observed and no putative ASF was identified.

Identification of group-specific ECF target promoter motifs. The combined body of evidence derived from both comparative genomic predictions and experimental studies strongly suggests that ECFs belonging to the same group recognize similar target promoters (60, 61). Given that one of the hallmark features of most ECFs is the autoregulation of their own expression (62), it is to be expected that such target promoters can be found upstream of the respective transcriptional units, which facilitates their identification by searching for overrepresented bipartite sequence motifs in the promoter regions from within one ECF group.

We therefore attempted to identify the target promoters of newly defined ECF groups (ECF47 to ECF56) and of previously described ECF groups for which no promoter had yet been identified (ECF118, ECF122, and ECF123). Indeed, putative promoter sequences were identified for all these ECF groups, albeit in only about 70% of the promoter sequences (Fig. 7; see Table S8 in the supplemental material). In 7% of those sequences, additional—albeit degenerated—putative promoter motifs could also be found. One-fifth of all the promoter sequences were located very close to the start codon, so the +1 position of the mRNA would reside within the ribosome binding site or even directly upstream of the start codon. This observation indicates that (i) the putative promoter might not be a real promoter; (ii) a leaderless mRNA (without a ribosome binding site) is generated from such promoters, and distinct strategies of regulation of translation initiation are employed (63); or (iii) the start codon is misannotated in those ECFs. The last possibility is supported by a global study of *M. tuberculosis* H37Rv, which demonstrated that about 7% of all coding sequences were indeed misannotated in the strain (64).

Final considerations. In the last 5 years alone, the number of completed microbial genomes has tripled. Next-generation sequencing efforts have further expanded by almost an order of magnitude the available sequence space of unfinished draft genomes. This massive increase in sequence information significantly boosts the complexity of comparative genomics analyses but also facilitates grouping of previously unclassified proteins with similar characteristics, thereby enabling the generation of new hypotheses regarding their functions or mechanisms of action.

This is reflected in the study by Jogler et al. (32), in which the analysis of eight genome sequences of *Planctomycetes* resulted in the definition of eight new ECF groups, thereby classifying almost 80% of the *Planctomycetes* ECFs that could not be grouped by our original classification (6). The same was observed in the present study: the analysis of 119 actinobacterial genomes resulted in the identification of 18 new ECF groups. This allowed the classification of 81% of the actinobacterial ECFs that were not covered by any of the original ECF groups (5, 6, 32).

Actinobacteria can live in aquatic (65) or terrestrial (66) environments and can also be pathogenic to both animals (67) and plants (68). Moreover, the phylum includes rod-shaped as well as filamentous bacteria, and some of the organisms can also undergo differentiation into spores (69) or other dormant forms (70). These complex lifestyles are mirrored in the complexity of signal transduction and in the number and diversity of STPs shown here but also in signaling networks already elucidated (71). Our initial study on ECF classification (6), in which *Actinobacteria* were highlighted as one of the most ECF-rich phyla, is a good example of this. Similarly, the current study revealed six actinobacterium-specific 1CS groups, eight actinobacterium-specific HK groups, three actinobacterium-specific RR architectures, and seven new actinobacterium-specific ECF groups.

However, the data presented in this study, which is based on a sequence analysis of over 50,000 STPs, not only serve as a key resource for researchers interested in actinobacterial signal transduction, they also provide a comprehensive platform for the generation of hypotheses regarding the biological roles and regulatory mechanisms of newly identified STP groups. (i) A group-specific protein domain architecture may provide crucial hints about the regulatory mechanism (e.g., as described for the role of HK11 in mediating the actinobacterial phage shock response or the regulatory relevance of C-terminal extensions in ECF group ECF48 or ECF52). (ii) Genomic context conservation may point to important accessory regulatory proteins or conserved target genes, as indicated in Fig. 3D and 4D for 1CSs and ECFs, respectively. (iii) Predicted ECF group-specific target promoter motifs (Fig. 6) are key to identifying the corresponding regulon. All of these indications can help to formulate clear hypotheses that can help to focus and streamline subsequent experimental efforts. Ultimately, the information provided in this report will therefore help to mechanistically explore the abundance, diversity, and uniqueness of actinobacterial STPs that contribute to the remarkable adaptability of these organisms to complex environments and distinct lifestyles.

ACKNOWLEDGMENTS

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (MA2837/2-2 to T.M.). D.P. receives funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 628509. X.H. is the recipient of a stipend from the Chinese Scholarship Council (CSC).

REFERENCES

1. **Ulrich LE, Koonin EV, Zhulin IB.** 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol **13:**52–56. http://dx.doi.org/10.1016/j.tim.2004.12.006.
2. **Galperin MY.** 2006. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. J Bacteriol **188:**4169–4182. http://dx.doi.org/10.1128/JB.01887-05.
3. **Helmann JD, Chamberlin MJ.** 1988. Structure and function of bacterial sigma factors. Annu Rev Biochem **57:**839–872. http://dx.doi.org/10.1146/annurev.bi.57.070188.004203.
4. **Mascher T.** 2013. Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. Curr Opin Microbiol **16:**148–155. http://dx.doi.org/10.1016/j.mib.2013.02.001.
5. **Mascher T, Helmann JD, Unden G.** 2006. Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol Biol Rev **70:**910–938. http://dx.doi.org/10.1128/MMBR.00020-06.
6. **Staroń A, Sofia HJ, Dietrich S, Ulrich LE, Liesegang H, Mascher T.** 2009. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. Mol Microbiol **74:**557–581. http://dx.doi.org/10.1111/j.1365-2958.2009.06870.x.
7. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG.** 2011.

Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7:**539. http://dx.doi.org/10.1038/msb.2011.75.

8. **Hall TTA.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41:**95–98.
9. **Saitou N, Nei M.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:**406–425.
10. **Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R.** 2007. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics **8:**460. http://dx.doi.org/10.1186/1471-2105-8-460.
11. **Ulrich LE, Zhulin IB.** 2010. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. Nucleic Acids Res **38:**D401–D407. http://dx.doi.org/10.1093/nar/gkp940.
12. **Sonnhammer EL, von Heijne G, Krogh A.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol **6:**175–182.
13. **Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD.** 2012. The Pfam protein families database. Nucleic Acids Res **40:**D290–D301. http://dx.doi.org/10.1093/nar/gkr1065.
14. **Cserzö M, Eisenhaber F, Eisenhaber B, Simon I.** 2002. On filtering false positive transmembrane protein predictions. Protein Eng **15:**745–752. http://dx.doi.org/10.1093/protein/15.9.745.
15. **Alm EJ, Huang KH, Price MN, Koche RP, Keller K, Dubchak IL, Arkin AP.** 2005. The MicrobesOnline Web site for comparative genomics. Genome Res **15:**1015–1022. http://dx.doi.org/10.1101/gr.3844805.
16. **Geer LY, Domrachev M, Lipman DJ, Bryant SH.** 2002. CDART: protein homology by domain architecture. Genome Res **12:**1619–1623. http://dx.doi.org/10.1101/gr.278202.
17. **Thompson JD, Higgins DG, Gibson TJ.** 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. http://dx.doi.org/10.1093/nar/22.22.4673.
18. **Crooks GE, Hon G, Chandonia J, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome Res **14:**1188–1190. http://dx.doi.org/10.1101/gr.849004.
19. **McGuffin LJ, Bryson K, Jones DT.** 2000. The PSIPRED protein structure prediction server. Bioinformatics **16:**404–405. http://dx.doi.org/10.1093/bioinformatics/16.4.404.
20. **Rhodius VA, Segall-Shapiro TH, Sharon BD, Ghodasara A, Orlova E, Tabakh H, Burkhardt DH, Clancy K, Peterson TC, Gross C, Voigt C.** 2013. Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. Mol Syst Biol **9:**702. http://dx.doi.org/10.1038/msb.2013.58.
21. **Liu X, Brutlag D, Liu J.** 2001. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac Symp Biocomput **6:**127–138.
22. **Wuichet K, Alexander RP, Zhulin IB.** 2007. Comparative genomic and protein sequence analyses of a complex system controlling bacterial chemotaxis. Methods Enzymol **422:**1–31.
23. **Gruber TM, Gross C.** 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol **57:**441–466. http://dx.doi.org/10.1146/annurev.micro.57.030502.090913.
24. **Ulrich LE, Zhulin IB.** 2007. MiST: a microbial signal transduction database. Nucleic Acids Res **35:**D386–D390. http://dx.doi.org/10.1093/nar/gkl932.
25. **Galperin MY.** 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol **5:**35. http://dx.doi.org/10.1186/1471-2180-5-35.
26. **Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honoré N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell, T, Fraser, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG.** 2001. Massive gene decay in the leprosy bacillus. Nature **409:**1007–1011. http://dx.doi.org/10.1038/35059006.
27. **Parandhaman DK, Hanna LE, Narayanan S.** 2014. PknE, a serine/threonine protein kinase of Mycobacterium tuberculosis initiates survival crosstalk that also impacts HIV coinfection. PLoS One **9:**e83541. http://dx.doi.org/10.1371/journal.pone.0083541.
28. **Molle V, Kremer L.** 2010. Division and cell envelope regulation by Ser/Thr phosphorylation: Mycobacterium shows the way. Mol Microbiol **75:**1064–1077. http://dx.doi.org/10.1111/j.1365-2958.2009.07041.x.
29. **Parandhaman DK, Sharma P, Bisht D, Narayanan S.** 2014. Proteome and phosphoproteome analysis of the serine/threonine protein kinase E mutant of Mycobacterium tuberculosis. Life Sci **109:**116–126. http://dx.doi.org/10.1016/j.lfs.2014.06.013.
30. **Kumar D, Palaniyandi K, Challu VK, Kumar P, Narayanan S.** 2013. PknE, a serine/threonine protein kinase from Mycobacterium tuberculosis has a role in adaptive responses. Arch Microbiol **195:**75–80. http://dx.doi.org/10.1007/s00203-012-0848-4.
31. **Cluzel M-E, Zanella-Cléon I, Cozzone AJ, Fütterer K, Duclos B, Molle V.** 2010. The Staphylococcus aureus autoinducer-2 synthase LuxS is regulated by Ser/Thr phosphorylation. J Bacteriol **192:**6295–6301. http://dx.doi.org/10.1128/JB.00853-10.
32. **Jogler C, Waldmann J, Huang X, Jogler M, Glöckner FO, Mascher T, Kolter R.** 2012. Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in Planctomycetes by comparative genomics. J Bacteriol **194:**6419–6430. http://dx.doi.org/10.1128/JB.01325-12.
33. **Yeats C, Finn RD, Bateman A.** 2002. The PASTA domain: a beta-lactam-binding domain. Trends Biochem Sci **27:**438. http://dx.doi.org/10.1016/S0968-0004(02)02164-3.
34. **Mir M, Asong J, Li X, Cardot J, Boons G-J, Husson RN.** 2011. The extracytoplasmic domain of the Mycobacterium tuberculosis Ser/Thr kinase PknB binds specific muropeptides and is required for PknB localization. PLoS Pathog **7:**e1002182. http://dx.doi.org/10.1371/journal.ppat.1002182.
35. **Povolotsky TL, Hengge R.** 2012. "Life-style" control networks in Escherichia coli: signaling by the second messenger c-di-GMP. J Biotechnol **160:**10–16. http://dx.doi.org/10.1016/j.jbiotec.2011.12.024.
36. **Hengge R.** 2009. Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol **7:**263–273. http://dx.doi.org/10.1038/nrmicro2109.
37. **Trach K, Burbulys D, Strauch M, Wu JJ, Dhillon N, Jonas R, Hanstein C, Kallio P, Perego M, Bird T.** 1991. Control of the initiation of sporulation in Bacillus subtilis by a phosphorelay. Res Microbiol **142:**815–823. http://dx.doi.org/10.1016/0923-2508(91)90060-N.
38. **Brissette JL, Weiner L, Ripmaster TL, Model P.** 1991. Characterization and sequence of the Escherichia coli stress-induced psp operon. J Mol Biol **220:**35–48. http://dx.doi.org/10.1016/0022-2836(91)90379-K.
39. **Darwin AJ.** 2005. The phage-shock-protein response. Mol Microbiol **57:**621–628. http://dx.doi.org/10.1111/j.1365-2958.2005.04694.x.
40. **Mascher T.** 2014. Bacterial (intramembrane-sensing) histidine kinases: signal transfer rather than stimulus perception. Trends Microbiol **22:**559–565. http://dx.doi.org/10.1016/j.tim.2014.05.006.
41. **Shu CJ, Ulrich LE, Zhulin IB.** 2003. The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors. Trends Biochem Sci **28:**121–124. http://dx.doi.org/10.1016/S0968-0004(03)00032-X.
42. **Taylor BL, Zhulin IB.** 1999. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev **63:**479–506.
43. **Aravind L, Ponting CP.** 1997. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci **22:**458–459. http://dx.doi.org/10.1016/S0968-0004(97)01148-1.
44. **Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Muellenet P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJA, Scheremetjew M, Tate J, Thimmajanarthanan M, Thomas PD, Wu CH, Yeats C, Yong S-Y.** 2012. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res **40:**D306–D312. http://dx.doi.org/10.1093/nar/gkr948.
45. **Aravind L, Koonin EV.** 2000. The STAS domain—a link between anion transporters and antisigma-factor antagonists. Curr Biol **10:**R53–R55. http://dx.doi.org/10.1016/S0960-9822(00)00335-3.
46. **Anantharaman V, Aravind L.** 2005. MEDS and PocR are novel domains with a predicted role in sensing simple hydrocarbon derivatives in prokaryotic signal transduction systems. Bioinformatics **21:**2805–2811. http://dx.doi.org/10.1093/bioinformatics/bti418.

47. **Yeats C, Bentley S, Bateman A.** 2003. New knowledge from old: in silico discovery of novel protein domains in Streptomyces coelicolor. BMC Microbiol **3:**3. http://dx.doi.org/10.1186/1471-2180-3-3.
48. **Coleman NV, Wilson NL, Barry K, Brettin TS, Bruce DC, Copeland A, Dalin E, Detter JC, Del Rio TG, Goodwin L, Hammon NM, Han S, Hauser LJ, Israni S, Kim E, Kyrpides N, Land ML, Lapidus A, Larimer FW, Lucas S, Pitluck S, Richardson P, Schmutz J, Tapia R, Thompson S, Tice HN, Spain JC, Gossett JG, Mattes TE.** 2011. Genome sequence of the ethene- and vinyl chloride-oxidizing actinomycete *Nocardioides* sp. strain JS614. J Bacteriol **193:**3399–3400. http://dx.doi.org/10.1128/JB.05109-11.
49. **Kirby JR, Zusman DR.** 2003. Chemosensory regulation of developmental gene expression in Myxococcus xanthus. Proc Natl Acad Sci U S A **100:**2008–2013. http://dx.doi.org/10.1073/pnas.0330944100.
50. **Wadhams GH, Armitage JP.** 2004. Making sense of it all: bacterial chemotaxis. Nat Rev Mol Cell Biol **5:**1024–1037. http://dx.doi.org/10.1038/nrm1524.
51. **Campbell EA, Greenwell R, Anthony JR, Wang S, Lim L, Das K, Sofia HJ, Donohue TJ, Darst SA.** 2007. A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria. Mol Cell **27:**793–805. http://dx.doi.org/10.1016/j.molcel.2007.07.009.
52. **Li W, Bottrill AR, Bibb MJ, Buttner MJ, Paget MSB, Kleanthous C.** 2003. The role of zinc in the disulphide stress-regulated anti-sigma factor RsrA from Streptomyces coelicolor. J Mol Biol **333:**461–472. http://dx.doi.org/10.1016/j.jmb.2003.08.038.
53. **Jung Y-G, Cho Y-B, Kim M-S, Yoo J-S, Hong S-H, Roe J-H.** 2011. Determinants of redox sensitivity in RsrA, a zinc-containing anti-sigma factor for regulating thiol oxidative stress response. Nucleic Acids Res **39:**7586–7597. http://dx.doi.org/10.1093/nar/gkr477.
54. **Bastiaansen KC, Ibañez A, Ramos JL, Bitter W, Llamas M.** 2014. The Prc and RseP proteases control bacterial cell-surface signalling activity. Environ Microbiol **16:**2433–2443. http://dx.doi.org/10.1111/1462-2920.12371.
55. **D'Andrea LD, Regan L.** 2003. TPR proteins: the versatile helix. Trends Biochem Sci **28:**655–662. http://dx.doi.org/10.1016/j.tibs.2003.10.007.
56. **Torkkell S, Kunnari T, Palmu K, Mäntsälä P, Hakala J, Ylihonko K.** 2001. The entire nogalamycin biosynthetic gene cluster of Streptomyces nogalater: characterization of a 20-kb DNA region and generation of hybrid structures. Mol Genet Genomics **266:**276–288. http://dx.doi.org/10.1007/s004380100554.
57. **Dumas P, Bergdoll M, Cagnon C, Masson JM.** 1994. Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. EMBO J **13:**2483–2492.
58. **Sousa Silva M, Gomes RA, Ferreira AEN, Ponces Freire A, Cordeiro C.** 2013. The glyoxalase pathway: the first hundred years and beyond. Biochem J **453:**1–15. http://dx.doi.org/10.1042/BJ20121743.
59. **Schneider D, Schmidt CL.** 2005. Multiple Rieske proteins in prokaryotes: where and why? Biochim Biophys Acta **1710:**1–12. http://dx.doi.org/10.1016/j.bbabio.2005.09.003.
60. **Rhodius VA, Suh WC, Nonaka G, West J, Gross CA.** 2006. Conserved and variable functions of the sigmaE stress response in related genomes. PLoS Biol **4:**e2. http://dx.doi.org/10.1371/journal.pbio.0040002.
61. **Dufour YS, Landick R, Donohue TJ.** 2008. Organization and evolution of the biological response to singlet oxygen stress. J Mol Biol **383:**713–730. http://dx.doi.org/10.1016/j.jmb.2008.08.017.
62. **Helmann JD.** 2002. The extracytoplasmatic function (ECF) sigma factors. Adv Microb Physiol **46:**47–110.
63. **Moll I, Grill S, Gualerzi CO, Bläsi U.** 2002. Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. Mol Microbiol **43:**239–246. http://dx.doi.org/10.1046/j.1365-2958.2002.02739.x.
64. **DeJesus M, Sacchettini JC, Ioerger TR.** 2013. Reannotation of translational start sites in the genome of Mycobacterium tuberculosis. Tuberculosis **93:**18–25. http://dx.doi.org/10.1016/j.tube.2012.11.012.
65. **Makovcova J, Slany M, Babak V, Slana I, Kralik P.** 2014. The water environment as a source of potentially pathogenic mycobacteria. J Water Health **12:**254–263. http://dx.doi.org/10.2166/wh.2013.102.
66. **Park Y, Kook M, Ngo HTT, Kim K-Y, Park S-Y, Mavlonov GT, Yi T-H.** 2014. Arthrobacter bambusae sp. nov., isolated from soil of a bamboo grove. Int J Syst Evol Microbiol **64:**3069–3074. http://dx.doi.org/10.1099/ijs.0.064550-0.
67. **Vázquez-Boland JA, Giguère S, Hapeshi A, MacArthur I, Anastasi E, Valero-Rello A.** 2013. Rhodococcus equi: the many facets of a pathogenic actinomycete. Vet Microbiol **167:**9–33. http://dx.doi.org/10.1016/j.vetmic.2013.06.016.
68. **González AJ, Trapiello E.** 2014. Clavibacter michiganensis subsp. phaseoli subsp. nov., pathogenic in bean. Int J Syst Evol Microbiol **64:**1752–1755. http://dx.doi.org/10.1099/ijs.0.058099-0.
69. **McCormick JR, Flärdh K.** 2012. Signals and regulators that govern Streptomyces development. FEMS Microbiol Rev **36:**206–231. http://dx.doi.org/10.1111/j.1574-6976.2011.00317.x.
70. **Shleeva M, Mukamolova GV, Young M, Williams HD, Kaprelyants AS.** 2004. Formation of "non-culturable" cells of Mycobacterium smegmatis in stationary phase in response to growth under suboptimal conditions and their Rpf-mediated resuscitation. Microbiology **150:**1687–1697. http://dx.doi.org/10.1099/mic.0.26893-0.
71. **Flärdh K, Buttner MJ.** 2009. Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. Nat Rev Microbiol **7:**36–49. http://dx.doi.org/10.1038/nrmicro1968.

Supplementary files for Chapter 3

The following supplementary files could be found in the attached CD.

Fig. S1. Correlation between different types of STPs.

Fig. S2. Sequence motifs in ASFs.

Table S1. General information about the analyzed genomes ("Genome properties" sheet) and organism's lifestyles ("General characteristics" sheet).

Table S2. Sequence list of actinobacterial 1CSs. Sequences are grouped accordingly to the classification depicted in Table 1.

Table S3. Distribution of RR output domains by organism ("Distribution of output domains" sheet) and information about the identified uncommon domain architectures ("Other domain architectures" sheet).

Table S4. Sequence list of actinobacterial HKs. Sequences are grouped accordingly to the classification depicted in Table 2.

Table S5. Distribution of ECFs by group and organism.

Table S6. Sequence list of newly classified ("Newly classified ECFs" sheet) and unclassified ("Unclassified ECFs" sheet) ECFs.

Table S7. Sequence list of ASFs grouped by the corresponding ECF group.

Table S8. List of sequences used for promoter analysis.

Table S9. Distribution of signal transducing proteins by type and organism.

Table S10. Description of protein domains.

Table S11. Summary of the main characteristic of newly described ECF groups.

Chapter 4

Defining the regulon of genes controlled by σ^E, a key regulator of cell envelope stress in *Streptomyces coelicolor*

Tran NT[#], **Huang XL**[#], Bush MJ, Chandra G, Pinto D, Hong HJ, Hutchings MI, Mascher T and Buttner MJ. ([#]**co-first author,** in preparation**)**

Defining the regulon of genes controlled by σ^E, a key regulator of cell envelope stress in *Streptomyces coelicolor*

*Ngat T. Tran[1,3,#], Xiaoluo Huang[1, 2,#], Matthew J. Bush[1], Govind Chandra[1], Daniela Pinto[2], Hee-Jeon Hong[1,5], Matthew I. Hutchings[1,6], Thorsten Mascher[2,4], Mark J. Buttner[1]**

[1]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK.
[2]Department Biology I, Ludwig-Maximilians-Universität München, Großhaderner Str. 2-4, 82152, Planegg-Martinsried, Germany.

#These two authors contribute equally to this work.

* Corresponding author.
Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH. Tel: +44 1603 450757. Email:mark.buttner@jic.ac.uk

Present addresses:
[3]Division of Infectious Diseases, Boston Children's Hospital, Harvard Medical School, Boston 02115, Massachusetts, USA
[4]Institut für Mikrobiologie, Technische Universität Dresden, 01062 Dresden, Germany
[5]Department of Biochemistry, University of Cambridge, Cambridge, CB2 1QW, UK.
[6]School of Biological Sciences and School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, Norwich Research Park. NR4 7TJ, UK

ABSTRACT

The extracytoplasmic function (ECF) σ factor, σ^E is a key regulator of the cell envelope stress response in *S. coelicolor*. Although its role in maintaining normal cell wall integrity has been known for over a decade, a comprehensive analysis of the genes under its control has yet to be undertaken. Here, using a combination of chromatin immunoprecipitation-sequencing (ChIP-seq), DNA microarray profiling and bioinformatic analysis, we define the σ^E regulon consisting of 91 key target genes. Approximately half of these genes encode proteins that are implicated in cell envelope function. 17 novel targets were validated by S1 mapping or *in vitro* transcription, establishing a σ^E binding consensus comprised of a conserved "AAC" at -35 region and a "TC" at -10 region. Subsequently, bioinformatic analysis identified σ^E regulated genes from 19 *Streptomyces* species, allowing us to establish a core regulon of σ^E function. Taken together, our study provides biological understanding of the σ^E-mediated response to cell envelope stress throughout the *Streptomyces* genus.

IMPORTANCE

The cell envelope stress response is critical for all bacteria to maintain envelope structure and survive in the various environmental niches in which they encounter antibacterial agents. Although, regulatory systems implicated in this response have been well characterized in a variety of organisms, there is little understanding of the response in antibiotic producer, *Streptomyces*. In the model organism *S. coelicolor*, the extracytoplasmic function (ECF) σ factor, σ^E has been shown to mediate the response towards a range of cell envelope disruptors, including vancomycin. In the clinic, resistance to vancomycin is an increasing problem as it is the front-line therapy against infections caused by methicillin-resistant *Staphylococcus aureus* (MRSA). Here, the genome-wide identification of the genes under σ^E control explains how this genus of bacteria is able to survive cell wall-induced stress.

INTRODUCTION

The bacterial cell envelope, made up of the cell wall and cell membranes, is critical in counteracting the high intracellular osmotic pressure to maintain cell shape (1). It also provides an essential defensive barrier against various environmental stress agents. Importantly, the cell envelope facilitates the ability of the cell to monitor the external environment and subsequently modulate cell behaviour (2, 3). Numerous antibiotics found so far target the bacterial cell envelope. For example, penicillin and other β-lactams mimic the D-alanyl-D-alanine terminus of the pentapeptide side chain of peptidoglycan and thus block the activity of penicillin binding proteins (PBPs) in the elongation and cross-linking of peptidoglycan precursors (4). Furthermore, vancomycin and other glycopeptide antibiotics bind to the D-alanyl-D-alanine terminus and thereby inhibit peptidoglycan cross-linking (5). In order to maintain cell viability, bacteria employ two major forms of signalling system (two-component systems and extracytoplasmic function (ECF) σ factors) to sense and respond to environmental stresses (2, 6). Both systems are functionally analogous in that they generally consist of a membrane protein (a sensor kinase or an anti-σ factor), that acts as a stress sensor and a transcription factor (a response regulator or a σ factor) that modulates gene expression. In the case of two-component systems, a signal leads to the autophosphorylation of a membrane-bound sensor kinase. The kinase subsequently phosphorylates its cognate response regulator, activating it as a transcription factor to affect the expression of genes involved in the cellular response (7, 8). Similarly, ECF σ factors have been shown to control the cellular stress response typically via an interaction involving their cognate anti-σ factors (9, 10), which are generally located in the membrane (11). In the absence of the signal, the anti-σ factor binds to its cognate ECF σ factor and inhibits its activity. However, detection of the signal by the anti-σ leads to the proteolysis, a conformational change or other changes of the anti-σ factor. In any case, the net result is the release of the ECF σ factor which is then able to bind its target promoters and elicit a specific transcriptional response (9, 10).

ECF σ factors belong to the σ^{70} family but differ from canonical housekeeping σ factors in that they have only two conserved σ domains, homologous to the σ_2 and σ_4 domains from housekeeping σ factors (12, 13). Subsequently the DNA-sequence motifs recognised by ECF σ factors, which typically contain an “AAC” at the -35 region, and conserved “G” and “T” residues at -10 region (11), are markedly different from those recognised by the

σ^{70}-family members as exemplified by the−35 (“TTGACA”) and −10 (“TATAAT”) consensus motifs in *B.subtilis and E.coli* (14, 15).

The roles of one two-component system (CpxAR) and one ECF σ factor (σ^E) in the cell envelope stress response of *E. coli* have been well established (16). The CpxAR system is induced by a variety of cell envelope stresses such as alkaline pH (17), indole (18) and increased osmolality (19). Overexpression of the outer membrane protein NplE (20), an altered membrane composition (21) and accumulation of mis-folded MalE aggregates (22) or pilus subunits (23) have all been demonstrated to additionally induce the CpxAR system. Activation of CpxAR results in the elevated expression of a subset of target genes that are involved in envelope protein folding and degradation, such as the periplasmic protease DegP, the periplasmic disulfide oxidoreductase DsbA and the foldase chaperone PpiA (24-26). The ECF σ factor, σ^E, mainly responds to stresses that affect the folding of outer membrane proteins (OMPs) (16, 27). For example, mutations in the OMP folding chaperone induce the σ^E stress response (28). The σ^E regulon includes a variety of genes involved in OMP folding (29, 30) as well as a number of small RNAs that down-regulate OMP expression, thereby reducing the flow of OMPs to the cell envelope (31-33).

In *Bacillus subtilis*, four two-component sytems (LiaRS, BceRS, YvcPQ, YxdJK) and at least three of its seven ECF σ factors, σ^M, σ^X , σ^W, have been shown to have roles in the response to cell envelope stress (2). For example, BceRS is strongly induced by bacitricin and involved in bacitracin detoxification (34). σ^M is activated by a wide variety of sources of envelope stress such as vancomycin, bacitracin, phosphomycin and cationic antimicrobial peptides (34-36). Much effort has also been made to define the regulatory networks linked to these signalling systems. It was demonstrated that σ^M contributes to the transcription of genes whose functions are related to transcriptional control, cell wall biosynthesis, cell shape determination, cell division, DNA monitoring and repair, and detoxification (37). Approximately 57 genes (30 operons) are likely to be direct targets of σ^M under antibiotic stress conditions, including several targets also assigned to either or both the σ^X and σ^W regulons (37).

Streptomyces coelicolor is a soil dwelling, saprophytic actinobacterium with a complex differentiating life cycle, including mycelial growth and sporulation (38). It is a well-

established model organism in which to study both development and signal transduction in the *Streptomyces* genus (38, 39). *S. coelicolor* encodes 67 paired two-component systems (39), and 51 ECF σ factors (collected from MiST2 database, http://mistdb.com/) (40). Despite this, only the two-component system VanRS and the ECF σ factor σ^E, have been so far described to play a role in the cell envelope stress response (41-46). VanRS responds to the glycopeptide antibiotics, e.g.vancomycin, ristocetin, chloroeremomycin and a glycopeptide antibiotic A47934, but not to other cell envelope disruptors such as moenomycin A, bacitracin and ramoplanin (41, 42). In contrast, σ^E expression is induced by a wide range of antibiotics, e.g. penicillins (penicillin G, amoxycillin, ampicillin, ticarcillin), glycopeptides (teicoplanin, ristocetin, vancomycin, chloroeremomycin), cephalosporins (cefaclor, cephalexin), a phosphoglycolipid (moenomycin A), a peptide (bacitracin), and a cyclic depsipeptide (ramoplanin), that targets the cell wall (44). A *sigE* mutant shows a 50-fold increase in sensitivity to the cell wall hydrolytic enzyme lysozyme and an altered cell wall muropeptide profile (45). The ECF σ factor σ^E seems therefore to play a more fundamental role than VanRS in the response of *S.coelicolor* to cell envelope stress.

The *sigE* gene is located in a four gene operon, with *cseA* encoding a lipoprotein, *cseB* encoding a response regulator and *cseC* encoding a sensor kinase, together directly downstream of *sigE*. Approximately 90% of transcription has been shown to terminate directly downstream of the *sigE* gene (43) and it appears that the transcription of *sigE* is completely dependent on the two-component system CseB (43, 44). It is likely that upon detection of a cell envelope stress, the sensor kinase CseC is auto-phosphorylated before phosphorylating its cognate response regulator CseB which in-turn directs the transcription of *sigE* (43, 44). *sigE* seems to be the sole target of CseB since the *sigE* mutant and *cseB* mutant show the same phenotype and overexpression of the σ^E protein almost complements the phenotype of increased lysozyme sensitivity of *S. coelicolor* lacking CseB (43). The function of CseA remains unknown. However, deletion of *cseA* results in an increased activity of σ^E in the cell and it has been suggested that CseA may interact with the membrane anchored sensor kinase CseC to modulate the activity of the CseBC-σ^E system (46). The requirement of a two-component system for transcription of *sigE* sets this system apart from other well characterized ECF σ factor regulatory mechanisms, which instead generally employ an anti-σ factor to modulate ECF σ factor

activity. Despite its name and proposed function, *S. coelicolor* σ^E is not similar to σ^E from either *E. coli* or *M. tuberculosis* where cognate anti-σ factors are present. Other than *S. coelicolor* σ^E, only one ECF σ factor has been shown to be dependent upon a two-component system (9). The HrpL-like ECF σ factor is indirectly regulated by a HrpXY-like two-component system and is involved in virulence-associated functions in *Pseudomonas syringae* (47-50).

Despite the critical role of σ^E in sensing cell envelope stress in *S. coelicolor*, only two *in vivo targets* have so far been described, i. e. a gene encoding the housekeeping σ-factor HrdD (45), the function of which is unknown (51) and a predicted cell wall glycan biosynthesis operon *cwg* (44). Therefore, the complete physiological function of σ^E in the cell envelope stress response remains unknown. In this study, using a combination of ChIP-seq, DNA microarray and bioinformatics analysis, 91 genes were assigned into the σ^E regulon in *S. coelicolor*. Approximately half of them were found to be directly involved in cell envelope-related functions. Other targets are also implicated in signal transduction and cell metabolism. Furthermore, core-regulon analysis here elucidates the conservation of the σ^E regulation network over the whole of the *Streptomyces* genus.

RESULTS AND DISCUSSION

Identification of σ^E targets by chromatin immuno-precipitation sequencing (ChIP-seq) and constitution of the σ^E regulon. To define the genes under control of the ECF σ factor, σ^E, we decided to employ chromatin immuno-precipitation coupled with high-throughput sequencing (ChIP-seq), a technique that has been widely used to study transcription factor regulons in *Streptomyces* (52, 53). Therefore, we constructed a strain of *Streptomyces coelicolor* that lacked *sigE* at its native locus but expressed an N-terminal, triple FLAG-tagged version of σ^E from the ΦBT1 integration site. As shown in Fig. 1, expression of 3xFLAG-σ^E *in trans*, under control of its native promoter, restores the resistance of *S. coelicolor* to lysozyme to wild-type levels. Furthermore, after treatment with vancomycin, 3xFLAG-σ^E was readily detected using anti-σ^E polycolonal antibody (Fig. S1) and anti-FLAG antibody (data not shown) in western blot assays suggesting suitable conditions for ChIP-seq experiments.

ChIP-seq was conducted with M2 anti-FLAG antibody after 30 minutes of treatment with vancomycin to induce cell-envelope stress. The wild type M600 *S. coelicolor* strain was used as a negative control to eliminate any false signals that arise from cross-reaction of the antibody with other DNA binding proteins. In addition, total (non-immunoprecipitated) input DNA was also subjected to sequencing. This very useful additional control enables non-uniform shearing of the chromosome to be taken into account (54). Using $P< 10^{-4}$, as the threshold for significance, a total of over 200 peaks were detected in the FLAG-tagged SigE strain (Table S1). Importantly, only a small number of low-significance peaks were detected in the wild type M600 strain in which the non-tagged version of σ^E is expressed (Fig. S2 and data not shown).

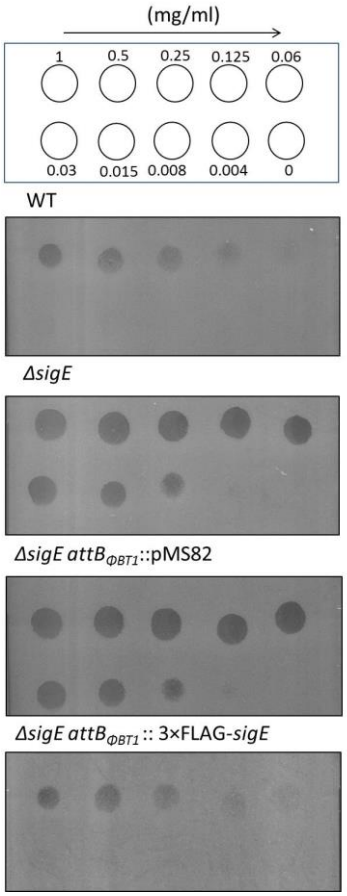


Fig 1. Lysozyme sensitivity of wild type *S. coelicolor* M600 (WT), *sigE* null mutant J2130 (*ΔsigE*), J2130 harbouring empty pMS82 vector (*ΔsigE attB*$_{\Phi BT1}$::pMS82), and J2130 expressing the N-terminal, triple FLAG-tagged σ^E (*ΔsigE attB*$_{\Phi BT1}$:: 3×FLAG-*sigE).* 5 µl lysozyme in a 2-fold dilution series was spotted onto DNA agar after plating 2×10^6 spores of each strain to create a confluent lawn. Images were taken after growth at 30 ˚C for 2 days.

Next, in order to constitute a σ^E regulon, we identified the likely promoter sites from the most enriched DNA regions in every ChIP-seq target (Table S1). The promoters of σ^E

target genes show strong conservation in the -35 (“AAC”) and -10 (“TC”) regions. The “AAC” sequence has been widely reported to be a typical conserved binding region for many ECF σ factors such as σ^M, σ^W and σ^X from *B. subtilis* (55), and σ^E from *E. coli* (29). The similarity between the binding consensus of *S. coelicolor* σ^E and that of other ECF σ factors may reflect a common mechanism of DNA-recognition in this protein family. Subsequently, those targets that contain a promoter site located at greater than 400 bp from the downstream gene were discarded for the subsequent analysis, leaving 91 target genes under σ^E control (Table 1).

Table 1

Locus[1]	TM[2]	Description[3]	G[4]	S[5]	Promoter sequence[6]	P_{dis}[7]	Evidence[8]			
							P_{GS}	C	A	*In vitro*
sco0662-0664*	6; 0; 0	Putative binding protein dependent transport protein; putative 2-hydroxyacid-family dehydrogenase; hypothetical protein (Pfam: DUF2264)	1	+	AGC**AACC**TCGGCTACAACATGGT-**GGTCT**T	360	+	+	-	
*sco0736**©	1	Putative protein, L,D-transpeptidase domain protein	2	-	CGC**AACC**AAAGCCGCCGGACGGC-**GGTCT**A	70	+	++	+	
sco0849-0848*	4; 0	Hypothetical protein (Pfam: DUF2269); putative oxidoreductase	5	-	AGG**AACG**GATAGGTGTTCTTCGC-**CATCC**C	306	-	+	-	
sco0877-0879	0; 0; 0	Putative transcriptional regulator (AAA_16, LuxR-type HTH domains); putative hydrolase; putative protein (AAA domaim)	4	+	GCC**AACG**AGGGCCGGACGCCGGC-**CGTCC**C	18	-	+	-	
sco1023-1024*	2; 0	Hypothetical protein; Hypothetical protein (CASH, Multiple PbH1, AAA domains)	5	+	GGG**AACC**CCGTCCTGGGTCCGCG-**CGTTG**G	97	-	+	N	
sco1168	0	Hypothetical protein	6	+	CTG**AACC**TCACGCGCGCGGAGCA-**CGTCG**T	249	+	+	+	
sco1647	0	Hypothetical protein (Pfam: Pup_ligase)	3	-	ACC**AACC**CCGACGACTGGGCCCG-**CATCT**C	362	+	+	-	
sco1738	0	Hypothetical protein (Pfam: IMS; Pfam: IMS_C)	7	-	TCG**CACG**ACGCCTGCGCGCACAC-**CGTCT**T	384	-	+	-	
sco1755	0	Hypothetical protein	6	-	GAC**AACC**AGAAGGCAGGGTTCCG-**GGTCT**A	38	+	++	+	
*sco1875**©	1	HMW PBP, cell wall biosynthesis	2	+	GGG**AACG**ACCGCCGCCCGCCCGTT**CGTCC**T	18	+	+	++	
*sco2055**	1	Hypothetical protein	5	+	CGG**AACC**AATTCTCTCAGACCCG-**CGTCC**G	171	+	+	-	
sco2168-2167*	0; 0	Hypothetical protein (Pfam: PspA_IM30); Hypothetical protein	7	-	GGG**AACG**ATCCGGCAACGCCGGT-**CGTCT**G	262	+	++	+	
*sco2255**©	3	Hypothetical protein (Pfam: Exo_endo_phos)	4	-	CGG**AACT**CCGCGGGACGGCCCGTA**CGTCC**T	105	-	+	+	
sco2294-2293	0; 10	Putative AraC family transcription regulator; Hypothetical protein (Pfam: EamA, Pfam: EamA)	4	-	GGA**CACC**GCGGGATTCCTCTGAT-**CGTCT**G	23	+	++	+	
*sco2334**	9	Hypothetical protein	5	+	GCC**AACG**TTTCCGTTCGAATTAT-**CGTCT**T	54	+	++	+	S1, IVT
sco2368	0	Hypothetical protein (Pfam:TerD)	7	-	GGC**AACG**TCTCGCGCGCCTACGG-**CGTCT**T	317	+	++	N	
*sco2419**©*-2410*	1 for all	Gene clusters encoding envelope proteins	5	-	TAC**GACC**ACTACTTCAACCTCTT-**CCTCT**C	224	+	+	+	
sco2611©*-2609(mreBCD*[1]*)**	0; 1; 5	Lateral cell wall biosynthesis	2	-	GGG**AACG**GATCCCACCGTTGGCC-**CGTCT**C	157	+	++	+	S1, IVT
*sco2629**	1	Hypothetical protein	5	+	TTC**AACT**ACAAGTTCCCCGACAC-**GGTCT**T	171	-	+	+	
*sco2807**©	5	Hypothetical protein (Pfam: PAP2_3)	5	+	GGC**AACC**CGAGGGGCGATGCCCG-**CGTCT**A	122	+	+	+	
*sco2892**©	1	Hypothetical protein (Pfam: Lipase_GDSL_2)	8	+	CGG**AACG**GAACACAAGTTCCCGG-**CGTCT**G	113	+	++	+	
*sco2897**	1	HMW PBP, cell wall biosynthesis	2	+	GGG**AACG**GAACCCGCGGTGCGAG-**AGTCT**T	260	+	++	++	S1
sco2939	0	Hypothetical protein (Pfam: SUKH-4)	7	-	GGC**AACG**AGTGCGTCCCCCCACG-**CGTCC**T	36	+	++	+	
sco2974(pkaA)©	0	Ser/Thr protein kinase	4	-	GGC**AACC**ACGGGACCGGGTCGAG-**CGTCT**T	108	+	++	-	
sco2975	0	Hypothetical protein (FHA-Forkhead associated, AAA domains)	4	+	CGT**GACC**GATCTCAAGCGGACGG-**CATTC**G	221	-	++	-	
*sco3034(whiB)**	0	Sporulation regulatory protein	4	-	CGG**AACG**GGATCGATCGCCGGGG-**CGTCC**T	238	+	+	+	
*sco3044**	1	Hypothetical protein (Pfam:LytR_cpsA_psr, Pfam: LytR_C)	2	+	AGT**GACC**TGAGGGGCCCCGCACG-**CGTCT**G	335	+	++	++	S1
*sco3098**	0	Putative secreted protein (Pfam: Transglycosylas, LysM domain)	2	-	GTC**AACC**GCCGCGTGGTCCCCGT-**CGTCT**T	15	+	++	-	
*sco3194**	1	Putative Lipoprotein, L,D-transpeptidase domain protein	2	+	GGG**AACC**CCACGGGCCGCCGGGCA**CCTCT**A	46	+	++	+	S1, IVT
sco3202(hrdD)	0	RNA polymerase principal sigma factor	4	-	GGC**AACC**CTCAGGCGGTACGGGC-**CGTCT**T	375	+	++	+	S1, IVT
sco3342-3341*	0; 3	Putative glycine-rich secreted protein (Pfam: DUF4232); Hypothetical protein	5	-	GGG**AACG**GTGTGCCGGGCCGAGCG**GCTCT**T	74	-	++	+	
*sco3396**©	2	Hypothetical protein (Pfam: Esterase)	8	-	CGG**AACC**TCGCCCGACATTTCCT-**CATCT**G	151	+	++	++	S1, IVT
*sco3397**©	5	Possible lysyl-tRNA synthetase MprF (Pfam: DUF2156)	2	+	GTG**AACC**TCTCCCTCCGAGACAC-**CGTCC**T	95	+	++	++	S1
sco3419	0	Hypothetical protein (Pfam: PadR)	4	-	CTC**AACG**GCGACACCATGCTGGA-**CGCCT**T	137	-	++	-	
sco3424	0	Putative regulator, similar to AbaA and BldB	4	-	GGG**AACG**ACTTCTCGGGCCCCGG-**CGTCG**T	164	-	+	-	
sco3481	0	Hypothetical protein (Pfam: Glyco_hydro_43)	3	-	TGG**AACG**ACTACCTGGTCGCCAC-**CGTCT**T	207	+	++	-	
sco3548	0	Putative anti-sigma factor	4	-	TGC**AACC**AGGAGCGCATTCTCAA-**GATCT**T	182	+	++	-	
sco3559	0	Oxidoreductase	3	-	GGG**CACG**GCGCCGGGTTGCGTAG-**GGTCT**T	4	-	++	-	
sco3712	0	Putative hydrolase, similar to the polysaccharide deacetylase (Pfam: Polysacc_deac_1)	3	+	GGG**ATCC**CGCGGCGGGTTTCTCC-**CGTCC**T	5	+	++	++	S1

*sco3728**	1	Hypothetical protein	5	-	GGG**AACG**GATCGGCGGCCGGCAG-**CGTCG**T	46	+	+	N		
sco3761	0	Hypothetical protein	6	-	GGG**AACC**TCGGCATGACCGTGTT-**CGTCT**C	47	+	+	+		
sco3900-3899	0; 0	Hypothetical protein (Pfam: PadR); Hypothetical protein (Pfam: NAD_binding_5)	4	-	CAA**AACC**CCCGCGGCCCGAAGTT-**CACCT**C	142	-	+	+		
sco3972	0	Hypothetical protein (Prim-Pol domain)	9	+	TGG**AACC**CGGCGACGGACCCGGG-**CGTCC**T	317	+	++	+		
*sco4042**	1	Hypothetical protein (Pfam: LytR_C)	2	-	TCG**AACC**TCGGAACGTCGACTGAT**CATCT**A	60	+	+	-		
*sco4069**	2	Hypothetical protein	5	+	CCG**AACC**CGGCAGGCCCCGGCTC-**CGTCT**C	259	+	++	+		
sco4120©	0	Hypothetical protein (Pfam: PAS_4, PP2C_SIG domain, Pfam: HATPase_c)	4	+	AGG**AACT**CCCCCGGCCACCGGGG-**CGTCT**G	145	+	+	N		
*sco4133**	5	Hypothetical protein (acidPPc domain)	5	+	TGG**AACG**TATCAACGGGGACCGTG**CGTTC**C	84	-	+	N		
*sco4134**©	0	Putative lipoprotein (Pfam: FAD_binding_4, Pfam: BBE)	8	-	GGG**AACC**CGCGCCCCCACACCCC-**CGTCT**C	33	+	++	++	S1	
sco4159-4158	0; 0	GlnR, transcriptional regulatory protein; putative LacI-family regulatory protein	4	-	GCG**AACC**GGGCACGACCACAAAC-**CGTCC**C	16	+	+	+		
sco4253	0	Hypothetical protein (Pfam: Phage_sheath_1)	6	-	AGA**AACG**CCGGGCGTCCGCCAGG-**GGTCT**T	158	-	+	N		
sco4263	0	Putative transcriptional regulator (AAA, LuxR type HTH domains), possible BldA target	4	+	CAC**CACC**GTTCACCGCAGTCGTT-**CGTCT**G	38	+	++	++	S1	
*sco4289**	0	Putative secreted protein	5	+	GAC**AACG**TCACGGACGGTTCCCC-**CGCCT**G	110	-	++	N		
*sco4439**©	0	LMW PBP; cell wall biosynthesis	2	-	TGG**AACC**AGTAGGTATGTCGTTCT**CGTCT**T	222	+	++	+		
sco4468-4467	0; 0	Gene clusters encoding hypothetical proteins	6	-	GAC**AACC**GCCCCCAACGCCGTGC-**CGTCT**G	169	+	++	+		
*sco4471**©	1	Hypothetical protein	2	+	CGG**AACC**CGCTCGTTCGTCGCGT-**CCTCT**C	38	+	++	++	S1, IVT	
sco4494	0	Hypothetical protein (Elp3 domain)	3	+	AGT**AACC**GGGGCGTACCGTTGACC**CGTCT**G	19	+	+	+		
*sco4582**	1	Hypothetical protein (Pfam: SLT)	5	-	GGC**AACC**CGACCGGAACCTGTGC-**CCTCC**C	345	+	+	N		
*sco4613**©	3	Hypothetical protein	5	+	CGC**AACC**ACCCGGCGCGGTCGGAA**CGTCC**T	88	+	+	+		
*sco4651(yjr2)**	0	Putative lipoprotein	5	+	AGA**AACC**ACAAGATCGTTCGAAC-**CGTTT**C	105	-	+	-		
*sco4847**	1	LMW PBP; cell wall biosynthesis	2	-	CGC**AACC**CGATGACCCCGACGAC-**CGTCC**C	271	+	++	+	S1	
*sco4849**	4	Hypothetical protein (Pfam: Metallophos)	3	+	GAG**GACG**TCACGGACGCCCTGAG-**CGTCC**C	20	-	++	+		
*sco4904**	4	Hypothetical protein (Pfam: VanZ)	7	-	CGG**AACC**GCACACGGCGGGGGGCG**CGTCT**A	7	+	++	+		
*sco4934**©	0	Hypothetical Lipoprotein, L,D-transpeptidase domain protein	2	+	GGC**AACC**GCCGCCCGGGGTTTCGT**CGTCT**C	172	+	++	+	S1	
*sco4968**	1	Hypothetical protein	5	+	CGG**AACG**GCGTACCAGCCGCTGAA**GGTCT**A	347	+	+	-		
*sco5030**©	7	Hypothetical protein (Pfam: UPF0118)	5	+	CTC**AACC**TCGCGCAGCCCCTCAC-**CGTCT**T	94	+	++	+	S1	
*sco5039**©	0	HMW PBP, cell wall biosynthesis	2	+	CAC**AACC**TTGAACCCCGCTCGTA-**CGTCG**G	335	+	+	+		
sco5049	0	Hypothetical protein (Pfam: Nitroreductase)	3	-	GCG**AACT**GTCCGACTTGAATTTCA**CCTTT**C	212	-	+	+		
*sco5213**	6	Hypothetical protein (Pfam: Peptidase_M50B)	8	-	GCG**AACC**GGCTCCGGGTCCTCGA-**CGTCT**T	198	+	++	+		
*sco5255**	0	Signal peptidase protein	8	+	GCA**AACA**GGCGGGAAAGCATGAAG**CGTTC**C	132	-	+	-		
sco5310	0	Hypothetical protein	6	-	GGG**AACG**GGCCGCCACGCGCGCA-**CGTTC**T	124	-	+	+		
*sco5358**©	1	Hypothetical protein (Pfam: LytR_cpsA_psr)	2	+	TGC**AACC**CTGTCCCGAGTTCCGC-**CGTCT**G	108	+	++	++	S1	
sco5535-5336(accB-accE)	0; 0	Acetyl-CoA carboxylase complex subunits	3	+	TGT**GACC**TCTACAAGCCAGAGGC-**CCTCT**G	117	+	+	+		
sco5705	0	Hypothetical protein (Pfam: DUF448)	6	+	GCG**AACG**CGCTCTCCCCGGCCCG-**CGTCT**C	304	+	+	-		
*sco5742**©	1	Hypothetical protein	5	-	GGG**CACC**TGAAGGGGCGTTCGTT-**CGTCT**G	49	+	++	++		
*sco5856**	2	Hypothetical protein	5	+	CGG**AACT**AATGGTTTCGGCCGCA-**CGTCC**C	52	-	++	+		
sco5981	0	Hypothetical protein (Pfam: Sulfotransfer_3)	6	-	GCG**AACC**CTCAGCCTCCTCAGAC-**CCTCT**T	29	+	++	+		
sco6028	0	Putative ribonuclease	9	-	CGG**AACG**TTCCGTCGGCGGGCTC-**CGTCG**A	122	+	+	-		
sco6130	0	Hypothetical protein (HATPase_c domain)	4	+	CTC**CACC**CCCGTCCCCACGTGAG-**CGCCT**C	21	-	+	-		
sco6178-6177	0;0	Putative deacetylase; Hypothetical protein (Pfam: Polyketide_cyc2)	3	-	ATG**AACC**GCGTATATACACGCAG-**CGTAT**A	50	-	++	-		
*sco6179-6190**	0 for all	*cwg* operon, cell wall glycan synthesis[1]	2	+	CGC**AACC**TGGTCCCCGTTTTCGT-**CGTCT**T	147	+	++	+	S1, IVT	
sco6262-6263	0;0	Putative helicase; hypothetical protein (Pfam: SWIM)	9	+	CGA**GACC**ACCGGTGCCGGTCTCGA**CGTCT**T	389	+	+	+		
sco6357-6353*	4; 4; 6; 0; 1	Hypothetical protein (Pfam: SNARE_assoc); Hypothetical protein (acidPPc domain); Hypothetical protein (acidPPc domain) ; putative two-component regulator;	5	-	GGG**AACG**TTCCTCACTCCGCCAT-**CGTCT**A	88	+	++	+		

sco6357-6353*	4; 4; 6; 0; 1	Hypothetical protein (Pfam: SNARE_assoc); Hypothetical protein (acidPPc domain); Hypothetical protein (acidPPc domain) ; putative two-component regulator; putative two-component histidine kinase	5	-	GGG**AACG**TTCCTCACTCCGCCAT-**CGTCT**A	88	+	++	+	
*sco6379**	4	Hypothetical protein (Pfam: SNARE_assoc)	5	-	TGG**AACG**GTCCTCACCCCGCTGC-**CGTCT**A	88	+	++	N	
sco6750	0	Putative IPP isomerase	3	-	GCG**GACG**GCCCGGGGGCGCGCACG**CACCG**G	234	-	+	N	
*sco6773**	1	Putative peptidase (Lysin motif domain, Pfam: Peptidase_M23)	8	+	GGG**AACC**CTTCGCTTGTCCCTGTG**CGTCT**T	224	+	++	+	
sco6832-6833	0;0	MutA, methylmalonyll-CoA mutase; probable isobutyryl-CoA mutase, small subunit	3	+	GGC**GACC**GTGCTGCGGAGCCCAA-**CATCT**T	242	+	+	-	
sco6979-6982*	0; 10; 0; 0	Putative solute-binding lipoprotein; Putative ABC transpoter membrane component;Putative ABC transporter ATP binding component; hypothetical protein (Pfam: AP_endonuc_2)	1	+	CTC**AACC**TCCGCCAGGGGTACGCC**CGTCT**G	322	+	+	-	
*sco7233**	1	Hypothetical protein, possible BldA target	5	+	GGC**AACC**CGAAGGATCTCCATCC-**CCTCC**T	69	+	++	++	S1
sco7657©-7658*	1; 0	Hypothetical protein (Pfam: Lysyl oxidase) ; hypothetical protein	8	+	GAC**AACC**GGGCATCCGAGCGCTC-**CCTCT**C	75	+	++	++	S1
sco7730	0	Hypothetical protein (Pfam: Acetyltransf_CG)	3	+	CGA**GACC**GACGCCCCGGCGCGGAC**CATCC**T	245	-	+	+	
scot11	0	tRNA-Met	10	-	GGG**AACC**GCGCGGCACGCTGCGG-**AGTCC**T	107	+	+	N	

1. σ^E targets included here meet the following criteria: 1) a clear ChIP-seq enrichment peak is found in the upstream region of the target; 2) a predicted promoter is found within 400 bp of the upstream region of the target. Genes that are transcribed in the same direction and contain an intergenic region that is less than 50 bp were assigned into the same operon (29, 56). The *cwg* operon (44) and *mreBCD* operon (57, 58) were defined based on published references. Any target gene or the first target gene in a predicted operon with predicted cell enveloped related function is marked with a "*". The whole *cwg* operon and *mreBCD* operon were marked with "*". The targets belonging to the defined σ^E core regulon from 19 *Streptomyces* genomes (Fig. 5A) were marked with a "©".
2. "TM" represents the number of predicted transmembrane helices identified by TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (59, 60).
3. Description of each target was based on either the annotation in StrepDB database (http://strepdb.streptomyces.org.uk), the domain architecture identified by Embl Smart database (http://smart.embl-heidelberg.de/) (61, 62) and the Pfam database (http://pfam.xfam.org/) (63) or the published references (see the text in the part of "functional analysis of σ^E regulon" and references cited by the text). Description of different genes in one operon is separated by ";".
4. Based on the proposed function of each target or the first target in an operon, they were assigned into 10 different functional groups (G). "1"= transportation; "2" = cell wall and cell membrane synthesis or modification; "3" = cell metabolism; "4"= cell regulation; "5" = envelope protein with unknown function; "6" = unknown protein; "7" = stress defense; "8" = envelope associated enzyme; "9" = DNA related function and "10" = tRNA.
5. "S" = the strand on which the gene is found (forward =+; reverse = -).
6. The promoter sequence of each target was predicted either by S1 nuclease mapping manually or via identification of the conserved sequence motifs underlying the highly enriched (≥200 bp) DNA regions from our ChIP-seq study, using the BioProspector program (http://ai.stanford.edu/~xsliu/BioProspector/, (64)). The predicted "-35" region and "-10" region of the promoter sequences are underlined.
7. "P_{Dis}" = the distance between the last base of the predicted promoter region to the first base of the starting codon of the target gene.
8. Validation of each target gene or the first target gene in an operon. "P_{GS}" = genome-wide bioinformatics prediction of the σ^E binding sites using the Virtual Footprint version 3.0 tool incorporated into the PRODORIC server (http://www.prodoric.de/vfp/vfp_regulon.php) (65, 66); "+" indicates that the promoter is captured by the prediction; "-" indicates that the promoter is not captured by the prediction. "C" = ChIP-seq profile; "+" indicates a ChIP-seq p-value between 10^{-5} and 10^{-30}; "++" indicates a ChIP-seq p-value below 10^{-30}. "A"= DNA microarray data; "+" indicates the transcription of the gene is induced less than 2 fold more in the wild type compared to the *sigE* mutant after treatment by vancomycin for 30 minutes; "++"indicates the transcription of the gene is induced at least 2 fold in the wild type compared to the *sigE* mutant after treatment by vancomycin for 30 minutes; "N" indicates, there is no detectable difference in expression between the wild type and *sigE* mutant after treatment by vancomycin for 30 minutes; "-"indicates the transcription of the gene is induced more in the *sigE* mutant than that in the wild type. "*in vitro*"= the targets are either confirmed by S1 nuclease mapping (S1) or *in vitro* transcription (IVT).

Further, in order to determine how σ^E influences the expression of its target genes, the *sigE* mutant were subjected to time-resolved, genome-wide transcriptional profiling before and after treatment with vancomycin which is exactly the same as those were done in the wild-type *S. coelicolor* before (67). Many SigE targets show a significant decrease in expression in the Δ*sigE* mutant compared to the wild-type (Table S1, Table 1), suggesting their transcriptional dependence on *sigE*. A number of σ^E targets however, do not show a clear transcriptional dependence upon *sigE*. Such targets commonly exhibit an increase in expression in wild-type *S. coelicolor* in response to vancomycin (re-analysed from the transcriptomic data of wild type *S. coelicolor* published by Hesketh et al. 2011, data not

shown (67)), in line with a role as part of the σ^E-response whilst indicating more complex transcriptional regulation exists for these targets.

Validation and classification of σ^E targets by S1 mapping. In order to validate the σ^E targets identified by the ChIP-seq and microarray experiments, 17 targets (*sco4471, sco4263, sco7657, sco3396, sco3397, sco2334, sco4134, sco5030, sco3044, sco7233, sco3712, sco2897, sco4847, sco5358, mreB, sco3194* and *sco4934*) were selected for promoter mapping via S1 nuclease. Results confirm that the genes identified here by ChIP-seq are indeed dependent upon σ^E for their expression (Fig. 2). This was further reinforced by *in vitro* transcription experiments that included *sco4471, mreB, sco3396, sco2334, sco3194* (Fig.S3 and data not shown). Subsequently, σ^E-dependent genes were divided into three groups based on the number of promoters upstream of each gene and their dependence on σ^E for their expression. Group one (*sco4471, sco4263, sco7657, sco3396, sco3397)* includes genes that have a single promoter and are completely dependent upon σ^E for their expression (Fig. 2A). In line with S1 mapping data, microarray transcriptional profiling data reveals that the transcription of group one targets is induced in the presence of vancomycin and is entirely dependent upon *sigE* (Fig. S4). Group two (*sco2334, sco4134, sco5030, sco3044, sco7233, sco3712, sco2897, sco4847* and *sco5358*) includes genes that have a single promoter and are partially dependent upon σ^E for their expression (Fig. 2B and data not shown). Once again, in agreement with S1 nuclease protection assays, expression profiling data reveals a clear response to the presence of vancomycin and a partial dependence upon *sigE* (Fig. S5). Finally group three (*mreB, sco3194* and *sco4934*) includes genes that have more than one promoter, one of which is dependent upon σ^E for its expression (Fig. 2C and data not shown). The transcription of all gene targets in this class is increased upon addition of vancomycin but the dependence on *sigE* is more subtle (especially *sco3194* and *sco4934*) (Fig. S6). Based on these criteria, the previously published σ^E-targets *cwgA* and *hrdD* could also be assigned into groups 2 and group 3 respectively.

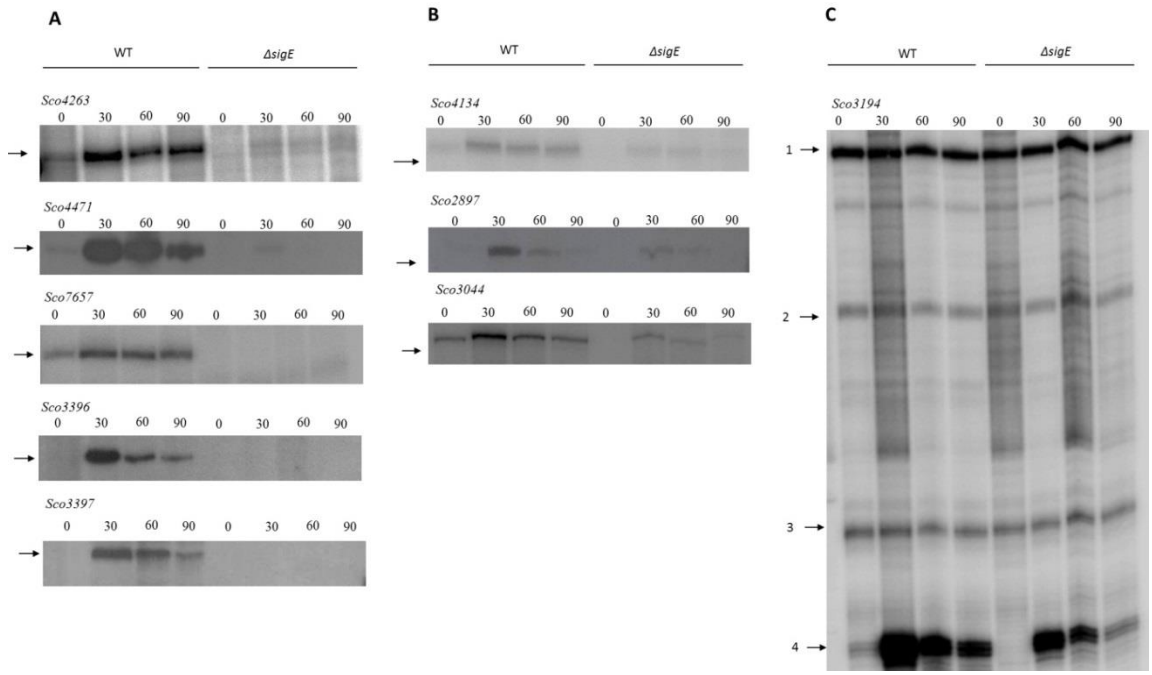


Fig 2. Examples of S1 nuclease protection assays of σ^E-target probes showing complete dependence of transcription upon σ^E (A), a partial dependence of transcription upon σ^E, from a single promoter (B) and a partial dependence of transcription upon σ^E, from multiple promoters (C). RNA was prepared from *S. coelicolor* M600 (WT) and the *sigE* null mutant J2130 (*ΔsigE*) after 0, 30, 60 and 90 mins treatment with 10 µg/ml (w/v) vancomycin. In each case the position of the protected transcript is indicated by an arrow. In (C) the arrow marked "4" indicates the position of the σ^E-dependent transcript.

The S1 mapping data was then used to identify the -10 and -35 recognition sequences for the 17 novel targets tested. These promoter sequences show a high correlation with our predicted promoter sequences (16/17 showed exact correlation, data not shown). Subsequently, based on these validated promoter sequences (Fig. 3), a σ^E consensus was generated using WebLogo (68). It is noteworthy that no clear difference exists in the predicted -35 and -10 binding motifs between those promoters of genes that are completely dependent upon σ^E and those that exhibit partial dependence. This suggests that some more subtle difference in promoter sequence, or else some other factor dictate whether a given promoter is solely transcribed by σ^E or whether it is transcribed via the action of multiple σ factors.

The partial dependence of gene targets upon σ^E suggests that there are likely to be other σ factors involved in their transcription. The extensive overlap in the binding specificity of different σ factors in bacteria has been well described. For example, there exists clear overlapping specificity between the binding of the heat shock σ factor, σ^{32}, and the housekeeping σ factor, σ^{70}, which share a similar -35 consensus motif ("CTTGAA" and

"TTGACA" respectively) (69). Most relevant to this work is the existence of multiple ECF σ factors, involved in the cell envelope stress response, that together contribute to the transcription from a common promoter, e.g. σ^M, σ^W and σ^X in *B. subtilis* (36, 55). The predicted consensus binding motifs for these ECF σ factors are highly similar (55) and it has been shown that just one or two DNA bases changes at the -10 region of the binding consensus are sufficient to alter the relative contributions of σ^W and σ^X towards the transcription of the same promoter (70).

S. coelicolor has more than 60 σ factors (71). Our results suggest that other σ factors are indeed involved in the transcription of σ^E-target genes. Furthermore, these σ factors may themselves be involved in the cell-envelope stress response. S1-mapping data reveals that transcription by such σ factors frequently occurs in response to the addition of vancomycin, as has been demonstrated for σ^E itself. The partial dependence upon σ^E for the majority of ChIP-seq targets tested by S1-mapping (14/19) implies the presence of a large network of σ factors that cooperate with σ^E to maintain cell envelope integrity in *Streptomyces*.

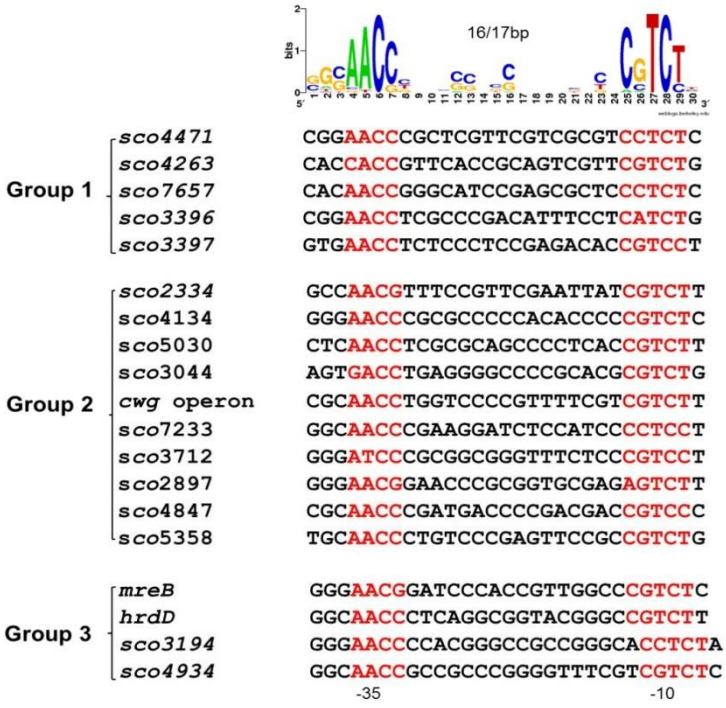


Fig 3. Alignment of the σ^E -10 and -35 recognition sequences, identified from the S1 mapping of 17 targets tested (and additionally including *hrdD* and *cwg*). The sequences are divided into group 1 (one promoter, completely dependent upon σ^E), group 2 (one promoter, partially dependent upon σ^E) and group 3 (multiple promoters, one at least partially dependent upon σ^E). Above the alignment is shown the corresponding σ^E consensus sequence, generated using WebLogo (68).

Functional analysis of σ^E regulon. 15 genes encode proteins involved in signal transduction and gene regulation (including the σ factor, HrdD) (Table 1) suggesting a complex signal transduction network is induced by σ^E. Indeed, HrdD is predicted to regulate the expression of over 80 genes, including 31 genes that themselves encode regulatory proteins (72). 12 genes also encode proteins involved in cell metabolism (Table 1). Over half of the genes under control of σ^E encode proteins relating to the cell-envelope (Table 1). These proteins include those involved in cell wall peptidoglycan elongation and assembly, cell wall-associated polysaccharide synthesis, lateral cell synthesis and sporulation as well as membrane modification and maintenance of integrity. These genes are highly significant σ^E-targets, as defined by ChIP-seq and clearly depend upon σ^E for their expression, as revealed by our microarray experiments, together suggesting that they play a major role in maintaining cell envelope integrity in *Streptomyces*.

I Cell wall peptidoglycan elongation and assembly. The σ^E targets *sco2897,* and *sco5039* encode proteins predicted to be "pencilling binding proteins" (PBPs) (Table 1). PBPs are involved in the final stage of peptidoglycan synthesis, facilitating the polymerisation and cross-linking of the peptidoglycan precursor, the disaccharide unit, and lipid II(73, 74). Penicillin or other β-lactam antibiotics are able to bind to these proteins and thus block their activity in peptidoglycan assembly (4). PBPs are broadly divided into two classes: the high molecular weight mass (HMM) PBPs and the low molecular weight mass (LMM) PBPs. Based on their structure and specific catalytic activity, the HMM PBPs are further sub-divided into two classes: A and B (73, 74). Class A has an N-terminal glycosyltransferase domain, involved in glycan chain elongation and a C-terminal transpeptidase domain, involved in the cross-linking of pentapeptide stems of glycan units (73, 74). This class of PBP is critical for cell growth in some bacteria such as in *E. coli*, where deletion of the two class A PBPs, (PBP1a and PBP1b) is lethal (75). Similarly, in *Streptococcus pneumoniae*, the class A PBPs, PBP1a and PBP2a together appear to be essential for viability (76). *sco2897*, and *sco5039* together with another class A PBP *sco3901* (a *sigE* target, identified by ChIP-seq but with the σ^E binding consensus greater than 400bp upstream, Table S1) encode proteins belonging to this subclass and are the only three class A HMW PBPs among more than 20 PBPs in *S. coelicolor* (74). It has been shown that deletion of any of these three PBPs result in slight decreased vancomycin resistance (67).

In contrast to class A HMW PBPs, class B HMW PBPs do not include an N-terminal glycosyltransferase domain, but rather an N-terminal domain thought to be involved in cell morphogenesis via interaction with protein partners (73, 74). For example, in *M. tuberculosis*, PBPA in this group is required for cell division and maintenance of cell shape. Phosphorylation of PBPA by the serine/threonine kinase PknB is suggested to regulate the positioning of PBPA at the cell septum and thereby modulate peptidoglycan synthesis (77). In *E. coli*, PBP3 (FtsI) is recruited by the cell division protein, FtsW, to the site of cell division (78). Some of the class B HMW PBPs, e.g. PBP2a from methicillin-resistant *S. aureus* (MRSA) (79-81); PBP5fm from *Enterococcus faecium* (82, 83); have been demonstrated to have a low affinity for penicillin and thus give rise to β-lactam resistance. The σ^E target *sco1875* encodes a member of the class B HMW PBPs (74) (Table 1). It shows greater than 50% sequence identity to *Streptomyces griseus* PBPA although a *pbpA* deletion in this organism results in no clear phenotype (84). In *Streptomyces coelicolor*, a Δ*sco1875* mutant exhibits increased sensitivity towards both vancomycin and bacitracin, further in line with a role in cell envelope stress response (67).

The LMM PBPs are involved in the cleavage of the terminal alanine in the pentapeptide stems of the glycan chain (DD-carboxypeptidase activity) or the internal peptide bond between two pentapeptide stems (endopeptidase activity) and therefore modulate peptidoglycan maturation or recycling (73, 74). The σ^E targets *sco4847* and *sco4439* encode putative D-alanyl-D-alanine carboxypeptidases (Table 1), but their function in *Streptomyces* is yet to be clarified.

Among the six σ^E targets in total that encode PBPs, *sco2897* and *sco4847* have been confirmed *in vitro* by S1 nuclease mapping to be transcribed from a single promoter that is partially dependent on σ^E (Fig. 2 and data not shown). Microarray transcriptional profiling also suggests that the *sco5039, sco4439* and *sco1875* genes are transcribed in a partially *sigE*-dependent manner (Table S1 and data not shown). Taken together, these findings implicate the increase in PBP expression, via σ^E-directed transcription, as an important component in facilitating the response to cell envelope damage in *Streptomyces*.

II An alternative pathway of peptidoglycan cross-linking. The σ^E targets *sco3194*, *sco4934* and *sco0736* encode proteins that each contain a L,D-transpeptidase catalytic domain (Pfam: YkuD) (Table 1). Such proteins cross-link the peptidoglycan precursor by forming a L-Lys_3/D-Asx-L-Lys_3 (3-3) bridge between two peptide chains of the peptidoglycan precursor, which bypass the typical D-Ala_4/D-Asx-L-Lys_3(4-3) transpeptidase activity of PBPs inhibited by β-lactams, and therefore enable the cell to display resistance to β-lactams (85). 3-3 cross links have been found to be rich in *M. Tuberculosis* and are suggested to play a role in the adaptive response during stationary phase (86). L,D-transpeptidase activity is also employed by *E. coli* in the attachment of Braun's lipoprotein (BLP) to the peptidoglycan (87). BLP is involved in cell envelope integrity through the connection of the outer membrane to the peptidoglycan layer (88, 89). The transcription of *sco3194*, *sco4934* and *sco0736* is highly induced by vancomycin and shown to be partially dependent on σ^E as shown from microarray and S1 mapping data (Table S1, Fig 2 and data not shown). The σ^E-mediated transcription of genes encoding proteins with L,D-transpeptidase activity, suggests that *S. coelicolor* may employ a similar 3-3 cross-linking mechanism to improve cell envelope integrity in response to cell envelope stress.

III Cell wall-associated polysaccharide synthesis. The σ^E targets *sco3044* and *sco5358* encode proteins in the LytR-CpsA-Psr (LCP) family (Table 1). The LCP proteins have been shown to be involved in the attachment of wall teichoic acid (WTA), capsular polysaccharides or else secondary cell wall polysaccharide to the peptidoglycan of the bacterial cell wall (90-93). WTA constitutes up to 60% of the Gram-positive cell wall and has roles in the regulation of cell division, cell shape determination, antibiotic resistance and pathogenesis (94). The transcription of LCP protein encoding gene, *msrR* from *Staphylococcus aureus* is induced by cell wall disruption agents such as β-lactams, glycopeptides, and lysostaphin (95). Deletion of *msrR* results in increased sensitivity to methicillin (oxacillin) and teicoplanin (95). In addition, deletion of genes encoding all three LCP homologs in *S. aureus* leads to a release of WTA into the extracellular medium (91) as well as a disruption in septum placement and cell separation (96). Recently, a LCP-like protein has also been shown to be involved the glycosylation of the GspA (AcaC) surface protein in *Actinomyces oris*. The glycosylated GspA is suggested to be attached to the cell wall by the sortase SrtA (97). Our finding that the *sco3044* and

sco5358 genes are σ^E targets and their partial dependence on σ^E as revealed by expression profiling (Fig. S5 A, B) data implicates the σ^E-response in the maintenance of cell wall components other than peptidoglycan.

IV Lateral cell wall synthesis and sporulation. ChIP-seq, microarray expression profiling and S1 mapping has here identified *mreB* as a σ^E target (Table 1). MreB is an actin-homolog and is found in a variety of bacteria. In rod-shaped bacteria such as *E. coli*, *B. subtilis* and *C. crescentus*, it acts as a cytoskeletal element underneath the cytoplasmic membrane to direct cell wall biosynthesis with the membrane proteins MreC and MreD, the encoding genes of which are located in the same cluster with *mreB* (98-100). In stark contrast to its function in other Gram-positive bacteria, in actinomycetes such as *S. coelicolor*, MreB does not direct lateral growth which instead occurs via an MreB-independent mechanism of tip-extension (58, 101, 102). Rather, MreB localises to the spore septa prior to cell division before spreading out, leading to the thickening of the spore coat (58, 101). Interestingly, *S.coelicolor* lacking *mreB* sporulates poorly and leads to the overproduction of actinorhodin (data not shown). The *sigE* null mutant exhibits similar characteristics; in line with our finding that *mreB* is a target of σ^E in response to cell envelope stress.

This study also identifies the *whiB* gene, encoding the key developmental regulator, WhiB, as a σ^E target (Table 1). WhiB belongs to the actinomycete-specific WhiB-like (Wbl) protein family (103, 104). Wbl proteins are all small (81-122 amino acids) and habour four conserved cystines that are together able to bind an oxygen-sensitive [4Fe–4S] cluster (105, 106). The detailed biochemical function of this protein family remains currently unknown, however evidence to-date so far suggests that WhiB and Wbl proteins function as transcription factors (103, 104, 107, 108). WhiB is essential for the initiation of sporulation septation in *S.coelicolor* (103), and in *M. tuberculosis*, the expression of WhiB orthologue WhiB2 is additionally induced by cell wall-inhibiting agents (isoniazid, ethambutol, and cycloserine) (109). *whiB* has two promoters, and the upstream promoter was previously shown to be recognised by σ^E in a run-off assay (110, 111). The transcription of *whiB* is highly induced by vancomycin in a σ^E dependent manner (Table 1 and data not shown). Therefore, *whiB* might be important in maintaining cell development as part of the σ^E–mediated response to cell envelope stress.

V Membrane modification. The *sco3397* gene encoding the multiple peptide resistance factor (MprF) protein is a clear σ^E target, as determined by ChIP-seq (Table 1). The transcription of *sco3397* is highly induced by vancomycin and is completely dependent on *sigE* as shown from microarray data and S1 mapping data (Fig. S4 and Fig. 2). MprF proteins are lysylphosphatidylglycerol (LPG) synthases that catalyse the transfer of L-lysine from lysyl-tRNA to the negatively charged lipid phosphatidylglycerol, thus neutralising the membrane surface charge. This leads to a resistance to the action of cationic antimicrobial peptides (CAMPs) via their repulsion (112). MprF has also been shown to affect the susceptibility of *S. aureus* towards various antibiotics such as vancomycin and daptomycin (113-115). In *S. coelicolor*, an *mprF* mutant shows increased sensitivity towards vancomycin and bacitracin (67), further in line with a role for MprF in the cell envelope stress response.

VI Maintenance of membrane integrity. The σ^E target *sco2168* encodes a PspA (phage shock protein A) homologue (116) (Table 1). PspA is the major effector of the phage shock protein (Psp) system that, in many organisms is involved in the adaptive response towards multiple extra-cytoplasmic stresses, most likely blocking stress-induced membrane damage and the resulting dissipation of the proton motive force (PMF) (117, 118). In *S. lividans, pspA* expression has been shown to be upregulated under various stress conditions where PspA is important for its survival (116). Therefore, in *S. coelicolor,* PspA might play an important role in the cell envelope stress response governed by σ^E.

The σ^E target *sco4471* encodes a protein involved in maintaining the integrity of the cell envelope. *sco4471*, encoding a predicted secreted protein, is a clear σ^E target as identified by ChIP-seq and microarray expression profiling reveals a complete dependence on σ^E for its transcription (Table 1, Fig. S4). Furthermore, the gene expression of *sco4471* shows a dramatic increase in response to vancomycin (Fig. S4, Table S1, more than 20-fold higher in the wild type *S.coelicolor* strain than the *sigE* mutant after treatment with vancomycin). Given its apparent importance as a member of the σ^E regulon, we set out to identify the function of this protein in the cell envelope response. Deletion of the gene encoding Sco4471 resulted in a four-fold increase in sensitivity to lysozyme compared to wild type *S. coelicolor* (Fig. 4A). Thus a lack of *sco4471* expression in a *sigE* mutant may

go some way to explaining its increased sensitivity to the cell wall hydrolytic enzyme (45). *S. coelicolor* Δ*sco4471* also showed approximately a two-fold increase in sensitivity to the cell wall specific antibiotics carbenicillin, vancomycin and bacitracin (data not shown). Overall, these results suggest that Sco4471 plays a role in maintaining the integrity of the cell envelope. As shown in Fig. 4B, the *sco4471* mutant also displays an alteration in spore size and shape suggesting a role for Sco4471 in spore formation.

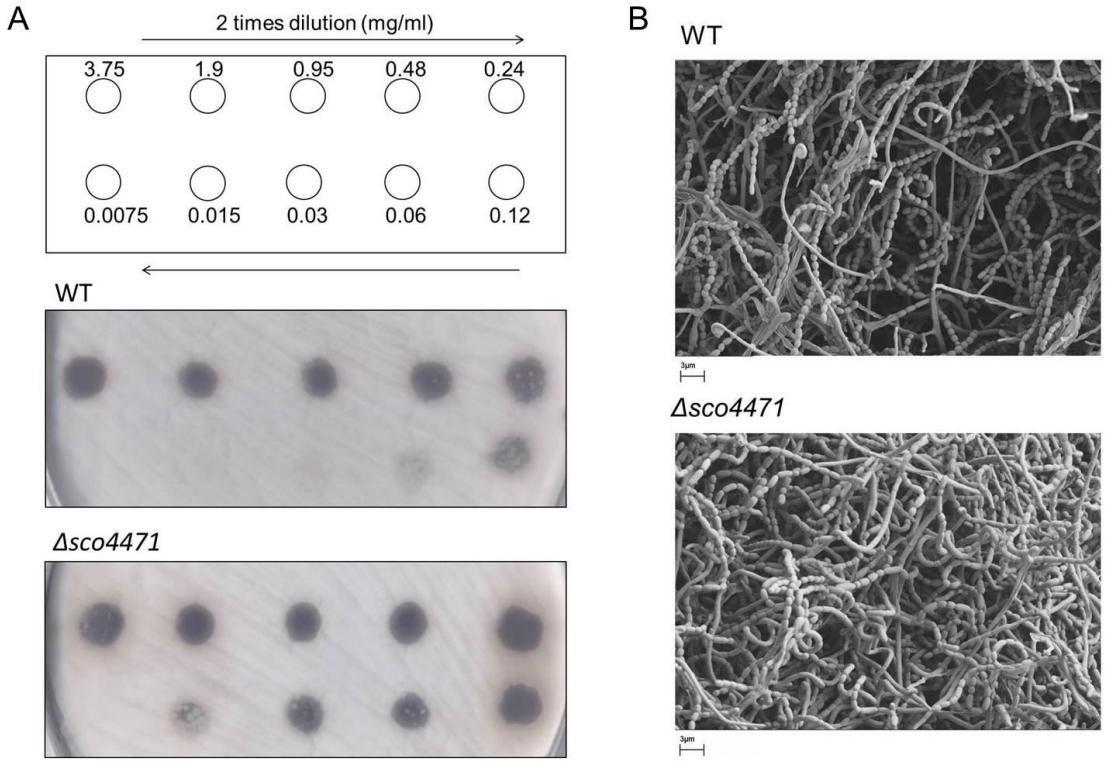


Fig. 4 A) Lysozyme sensitivity of *Δsco4471::apr* mutant and the wild type *S. coelicolor* M600 (WT). Lysozyme at different concentrations was spotted on the confluent spore lawn made by spreading $9.4x10^7$ spores on SFM plates. Images were taken after growth at 30 ˚C for 3 days. B) Scanning electron microscopy images of *Δsco4471::apr* mutant and the wild type *S. coelicolor* M600 (WT) examined after 5 days.

The σ^E core regulon. Following our identification of the genes under σ^E control, we sought to establish the role of σ^E across the entire *Streptomyces* genus. Given the high conservation of the σ_2 and σ_4 domains (that bind the -10 and -35 promoter elements respectively (119-122)) across σ^E orthologues (Fig. S7), we anticipated that the same σ^E orthologues would recognize highly similar or identical promoter motifs. Subsequently, we defined the bioinformatically predicted, σ^E regulon of 19 *Streptomyces* species via genome-wide prediction of the σ^E binding consensus sequences. Two promoter position

weight matrices (PWM)s with a 16 bp and 17 bp spacer between the "-35" region and "-10" region respectively were first generated from 19 refined (*in vitro*) validated σ^E target promoter sequences. These two PWMs were then used to predict all possible σ^E binding sites across all 19 genomes. In *S. coelicolor*, this prediction detects each of the 19 *in vitro* validated σ^E targets and over 70% of the targets obtained from our ChIP-seq experiments (Table 1), suggesting suitable parameters for accurate prediction. On average, each genome contained over 5000 predicted σ^E binding sites (Table S3). Since σ^E is unlikely to control the expression of so many genes, it is likely that an additional promoter motif, or else some other factors help σ^E orthologues to discriminate binding sites *in vivo*.

The promoters recognized by ECF σ factors have been shown to mainly locate in the closed upstream region of their controlled genes, which, in the vast majority of cases, is within 200 bp from the start codon of the downstream gene (37, 123). Therefore, to establish a σ^E "core" regulon, only those predicted promoter sites between 10 bp and 200 bp upstream of the start codon of the downstream gene were selected for further analysis. This gave rise to on average approximately 500 predicted targets from each of the 19 *Streptomyces* genomes. *S. bingchenggensis* contains the largest predicted σ^E regulon with 648 targets, whereas *S.cattleya* has a smallest σ^E regulon with 344 targets (Table S5). As a general trend, the larger genomes seem to have more targets than the smaller ones, suggesting a correlation between the genome size and potential complexity of σ^E regulon.

For most species, approximately 40% of predicted targets are specific to each genome. However, *S. pratensis* ATCC 33331, *S. sp.* PAMC26508, *S. hygroscopicus* subsp. jinggangensis 5008 and *S. hygroscopicus* subsp. jinggangensis TL01 have very few genome-specific, σ^E predicted-targets (Table S5). This variation may reflect the specific evolutionary pressure of each species in response to cell envelope stress. Indeed, *S. hygroscopicus* subsp. jinggangensis 5008 and *S. hygroscopicus* subsp. jinggangensis TL01 belong to the same species, and *S. pratensis* ATCC 33331 is phylogenetically close to *S. sp.* PAMC26508 (Fig. 5). Therefore the relatively low number of specific targets in these species suggests that similar σ^E regulatory patterns are found within genetically similar organisms.

118 predicted σ^E targets are conserved across at least 9 *Streptomyces* genomes (Fig. 5, Table 1 and Table S6). These targets therefore constitute a σ^E core regulon and may be representative of the overarching σ^E-mediated response to cell envelope stress. The majority can be assigned to three functional groups: 1) cell envelope related function; 2) cell metabolism and 3) cell regulation (Table 1 and Table S6). 21 of these targets (*S. coelicolor* gene locus*: sco0736, sco1875, sco2255, sco2419, mreB, sco2807, sco2892, pkaA, sco3396, mprF, sco4120, sco4134, sco4439, sco4471, sco4613, sco4934, sco5030, sco5039, sco5358, sco5742, sco7657*) are part of the 91 genes under σ^E-control (Table 1), identified here by ChIP-seq and include 9 targets (*mreB, sco3396, mprF, sco4134, sco4471, sco4934, sco5030, sco5358* and *sco7657*) validated in our S1 mapping or *in vitro* transcription experiments (Table 1, Fig. 5A). *mreB* is present in all the 19 predicted σ^E regulons indicating its pivotal role in cell envelope stress response in *Streptomyces.* This is in line with previous finding in *S.coelicolor* that MreB is important in the spore cell wall synthesis (58, 101). *sco4471* is present in 18/19 predicted σ^E regulons, which also reflects its importance in the σ^E dependent cell envelope stress response. Indeed, our work has revealed that a *sco4471* mutant is more sensitive to lysozyme as well as cell-wall specific antibiotics, suggesting that Sco4471 plays a key role in maintaining the integrity of the cell envelope. As already discussed, *sco1875, sco4439* and *sco5039* encode PBPs, which are likely to be involved in the synthesis of peptidoglycan. *sco0736* and *sco4934* encode L, D transpeptidases which may be involved in synthesis of peptidoglycan through the formation of a L-Lys_3/D-Asx-L-Lys_3 (3-3) bridge between two peptide chains of the peptidoglycan precursor. *sco5358* encodes a LytR_cpsA_psr family protein which may be involved in cell wall teichoic acid deposition and *mprF* is thought to encode a protein required for the resistance towards cationic antimicrobial peptides and antibiotics *sco2892, sco3396, sco4134* and *sco7657* are predicted to encode envelope associated enzymes, whereas *sco2419, sco2807, sco4613, sco5030* and *sco5742* encode unknown envelope proteins (Table 1). Finally, *sco2974 (pkaA)* encodes a Ser/Thr protein kinase and, *sco2255* and *sco4120* are predicted to encode regulatory proteins (Table 1).

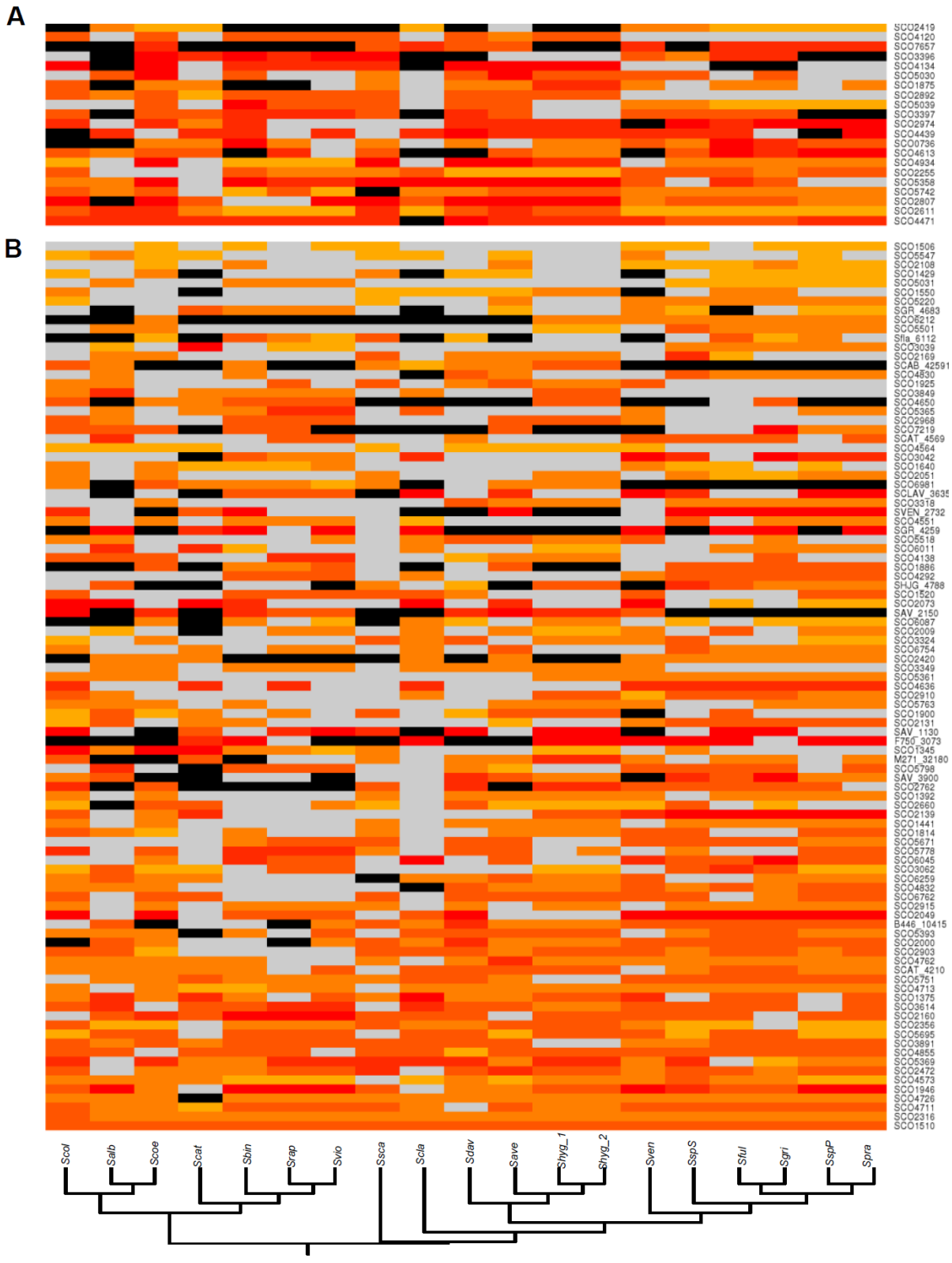


Fig. 5. The core σ^E regulon in 19 *Streptomyces* genomes. The predicted σ^E targets that were conserved in at least 9 *Streptomyces* genomes were assigned into the core σ^E regulon. "Black" indicates no orthologue of the target is found in the designated genome. "Grey" indicates the orthologue of the target is found, but the σ^E binding consensus is not present between 10-200bp upstream of the target orthologue. "Yellow" to "red" linear gradient indicates both the orthologue of the target is found and a σ^E binding consensus is found 10-200 bp upstream of the target orthologue. The σ^E binding consensus of each target was predicted by the Virtual Footprint version 3.0 tool incorporated into the PRODORIC server

(http://www.prodoric.de/vfp/vfp_regulon.php) (65, 66) and a PRODORIC score was given to reflect the quality of the prediction. The "Yellow" to "red" linear gradient here indicates the PRODORIC score of the σ^E binding consensus from the minimum value to the maximum value. **A)** The validated core σ^E regulon. It consisted of targets that are not only identified by the genome-wide bioinformatics prediction, but also present in Table 1. **B)** The extended core σ^E regulon. It consisted of targets that are solely identified by the genome-wide bioinformatics prediction, but not present in Table 1. The *Streptomyces* strains were listed according to the guide tree generated from multiple sequence alignment of their 16s rDNA by T-Coffee (http://www.ebi.ac.uk/Tools/msa/tcoffee/)(124).

97 of the 118 σ^E core regulon targets identified in the bioinformatics analysis were not identified as part of the validated σ^E regulon that results from our ChIP-seq study (shown in Table 1) and thus constitute an "extended core-regulon"(Fig. 5B, Table S6). These include 19 highly conserved targets (*sco1375*, *sco1510*, *sco1946* (pgK), *sco2160*, *sco2316*, *sco2356*, *sco2472*, *sco3891*, *sco4573 (nuoL)*, *sco4711 (rpsQ)*, *sco4713 (rplX)*, *sco4726 (rpmJ)*, *sco4762 (groEL1)*, *sco5369*, *sco5695*, *sco5751*, *sgr_2144*, *sgr_2690* and *sgr_3371*) that are present in 16 or more of the 19 *Streptomyces* genomes selected (Table 2). *sco2316* (encoding a hypothetical protein) and *sco1510* (encoding a peptidyl-prolyl cis-trans isomerase) are present in all the 19 predicted σ^E regulons, implying their central role in the σ^E response. *rpsQ*, *rplX* and *rpmJ* encode three ribosome components and *groEL1* encodes a protein folding chaperon, which might suggest σ^E also has a role in maintaining protein homeostasis under stress conditions.

Overall, the establishment of a core-σ^E reguon among *Streptomyces* species has revealed the key genes that together form the central contribution of σ^E towards the cell envelope stress response. Outside the *Streptomycetes*, the *sigE-cseABC* operon (encoding σ^E and the two component system CseBC and a lipoprotein CseA (43-46)), seems to be absent (identified by tree-based genome browser tool in MicrobesOnline database, data not shown, http://www.microbesonline.org/ (125)). Therefore the conserved mechanisms that underly the σ^E dependent cell envelope stress response are probably specific to the *Streptomyces* genus.

Table 2-Selected highly conserved predicted σ^E targets in 19 *Streptomyces* genomes[1]

Locus[2]	TMH[3]	Description[4]	Orthologues[5]	OP[6]
sco1510	1	Peptidyl-prolyl cis-trans isomerase	18	18
sco2316	0	Hypothetical protein, Pfam: Prenyltrans_2	18	18
sco2611(mreB)#	0	Lateral cell wall biosynthesis	18	18
sco4471#	1	Hypothetical protein	17	17
sco4726(rpmJ)	0	50S ribosomal protein L36	17	17
sco4573(nuoL)	14	NADH dehydrogenase subunit	18	17
sco4711(rpsQ)	0	30S ribosomal protein S17	18	17
sco5742#	1	Hypothetical protein	17	16
sco1946(pgK)	0	Phosphoglycerate kinase	18	16
sco2356	1	Hypothetical protein	18	16
sco2472	1	Hypothetical protein, Pfam: DUF218	18	16
sco3891	1	Hypothetical protein	18	16
sco4713(rplX)	0	50S ribosomal protein L24	18	16
sco5369	1	F0F1 ATP synthase subunit B chain	18	16
sco5695	4	Metalloprotease	18	16
sco2807#	5	Hypothetical protein, Pfam: PAP2_3	17	15
sgr_2690	0	succinate dehydrogenase iron-sulfur subunit gene: sdhB	18	16
sco1375	0	Hypothetical protein	18	15
sco2160	1	Hypothetical protein	18	15
sco4762(groEL1)	0	Chaperonin GroEL	18	15
sco5751	1	Hypothetical protein, Pfam:HTH_25, Pfam:DUF4115	18	15
sgr_3371	0	aspartate-semialdehyde dehydrogenase	18	15
Sgr_2144	0	ABC transporter ATP-binding protein	18	15

1, Targets that were conserved in at least 16 *Streptomyces* genomes were shown here. The genome accessions of 19 *Streptomyces* species were listed in Table S2.

2, The targets from either *S.coelicolor* or *Streptomyces griseus* subsp. griseus NBRC 13350 were chosen as representatives. The targets belonging to the validated σ^E regulon in *S. coelicolor* as shown in Table 1 were marked with a "#".

3, "TM" represents the numbers of transmembrane helices identified by TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (59, 60).

4, Description of each target is based on the information extracted from the genome annotation deposited in the GenBank database (http://www.ncbi.nlm.nih.gov/), the domain architectures identified by Embl Smart database (http://smart.embl-heidelberg.de/) (61, 62) or the published reference (*mreB*, (57, 58)). For those hypothetical proteins which do not have a clear functional annotation, the domain information is included in the description.

5, "Orthologues"= the number of orthologues found in the remaining 18 *Streptomyces* genomes (Table S2), outside of *S. coeilicolor* or *Streptomyces griseus* subsp. griseus NBRC 13350.

6, "OP"= the number of orthologues, which has a predicted σ^E promoter within 10 bp to 200 bp upstream region, found in the remaining 18 *Streptomyces* genomes (Table S2), outside of *S. coeilicolor* or *Streptomyces griseus* subsp. griseus NBRC 13350.

Conclusions. In this study, the identification of the genes under control of σ^E via a combination of ChIP-seq and microarray transcriptional profiling reveals a complex regulatory network governed by σ^E in *S. coelicolor* in response to cell envelope-induced stress. The identification of a plethora of target genes encoding cell envelope proteins is in strong agreement with a role for σ^E in governing the cell envelope stress response (Fig. 6). Recently, mass spectrometry was employed to analyze changes in the *S. coelicolor* proteome upon vancomycin-induced stress (126). In line with our study, several proteins encoded by σ^E targets increase in abundance in response to vancomycin treatment. These include the proteins encoded by the genes *sco1647*, *sco2368*, *sco4494* and the gene encoding the primary sigma factor, σ^{HrdD}.

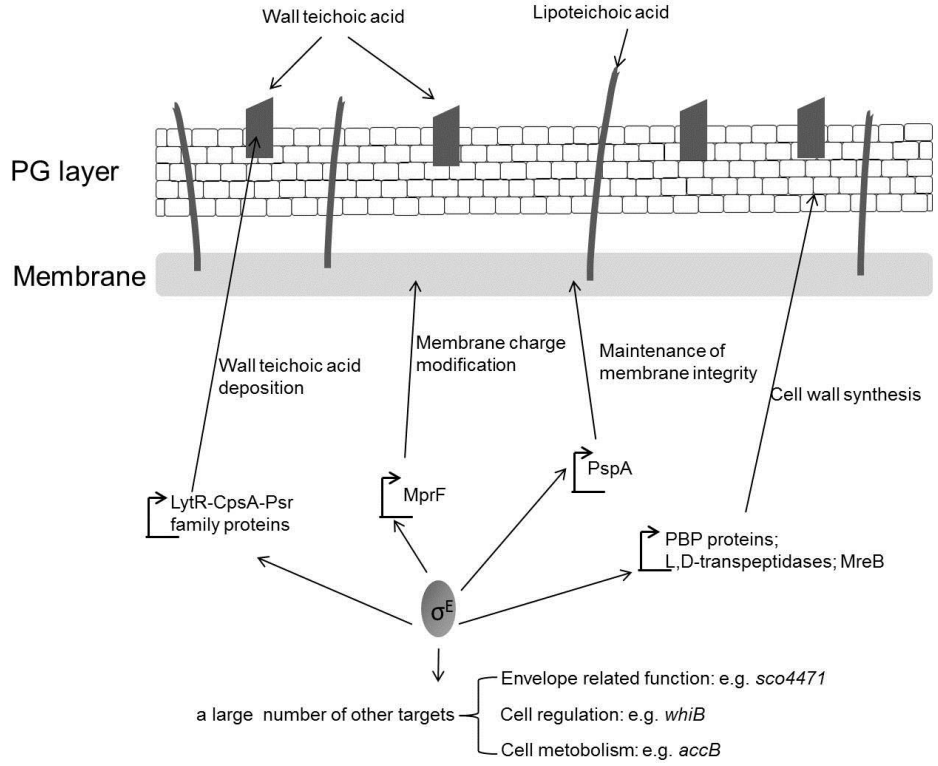


Fig. 6. Key mechanisms underlying σ^E dependent cell envelope stress response. Here, LytR-CpsA-Psr (LCP) family proteins refer to σ^E targets, *sco3044* and *sco5358*. At least one deduced function of LCP proteins is wall teichoic acid deposition. PBP proteins refer to σ^E targets *sco1875*, *sco2897*, *sco4439*, *sco4847* and *sco5039*. L, D transpeptidases refer to σ^E targets *sco0736*, *sco3194* and *sco4934*. Abbreviations: PG, Peptidoglycan; PBP, Penicillin binding proteins.

The observation that the majority of *in vitro* validated targets exhibit only a partial dependence on σ^E in microarray profiling, suggests that the regulation of such genes is

complex and is also subject to control by other σ factors. It would be interesting to elucidate the role of these other σ factors and whether or not they cooperate with σ^E to mediate the response to environmental stresses. The data presented here has allowed us to extend the σ^E regulon across 19 *Streptomyces* species via promoter based genomic searches. This has identified key genes with central roles in the "core-σ^E regulon", present across all *Streptomyces*. Three targets (*sco1510, sco2316* and *mreB*) were predicted to be subject to σ^E control in every *Streptomyces* genome analyzed, indicating a critical role in the σ^E dependent stress response. Taken together, our work provides a detailed picture of how σ^E protects *Streptomyces* from cell-envelope induced stress.

MATERIALS AND METHODS

Bacterial strains, plasmids, oligonucleotides and growth conditions. Bacterial strains, plasmids and primers in this study are listed in Table S7. Unless otherwise stated here, growth media and culture conditions were as described previously (127).

Construction of a 3×FLAG-σ^E-complemented *S.coelicolor* strain. In order to engineer an *S. coeilicolor* strain expressing a form of σ^E with an N-terminal, triple-FLAG tag (DYKDHDGDYKDHDIDYKDDDDK), a pMS82 (128)-derived construct, pMS82-3FLAG-*sigE*, was created via a two-step fusion-PCR approach. In the first step, the cosmid STE94 was used as a template for two separate PCR-reactions. The first reaction amplified the promoter region of the *sigE* gene using the primer pair $P1_{3NFLAGSigE}$ and $P2_{3NFLAGSigE}$. The second reaction amplified the coding region of the *sigE* gene using the primer pair $P3_{3NFLAGSigE}$ and $P4_{3NFLAGSigE}$. Together the $P2_{3NFLAGSigE}$ and $P3_{3NFLAGSigE}$ primers contain the sequence encoding the triple-FLAG tag via a 24bp overlapping section. In the second step, the primers $P1_{3NFLAGSigE}$ and $P4_{3NFLAGSigE}$ were used to amplify the entire *sigE* gene and its promoter, fusing the two products from step 1 together and incorporating the 3xFLAG tag sequence between them. The $P1_{3NFLAGSigE}$ and $P4_{3NFLAGSigE}$ primers additionally contain the HindIII and KpnI sites respectively to enable cloning into HindIII, KpnI-cut pMS82. The resulting vector was named pMS82-3FLAG-*sigE*. The plasmid pMS82-3FLAG-*sigE* was then introduced into the *ΔsigE* mutant J2130 (45) by conjugation using the *dam dcm hsdS E. coli* strain ET12567 containing pUZ8002 (45, 129).

Lysozyme sensitivity test. Lysozyme sensitivity tests for the wild type strain M600, *SigE* mutant J2130 and 3×FLAG-σ^E-complemented Δ*sigE* were performed as the same as described by Paget *et al.* (45). Briefly, 2×10^6 spores of *S.coelicolor* were spread onto a Difco Nutient Agar (DNA) plate to make a confluent lawn. 5 µl of 1 mg/ml lysozyme was then diluted in a two-fold series and the enzyme spotted onto the spore lawn before incubation at 30 ^oC for two days. Separately, lysozyme sensitivity tests were additionally carried out for the wild type strain M600 and Δ*sco4471* mutant. Lysozyme (ranging from 3.75 mg/ml to 0.0075 mg/ml through continuous two-fold dilutions) was spotted onto the confluent spore lawn generated from $9.4x10^7$ spores of these two strains on the SFM medium. The lysozyme sensitivity of M600 and Δ*sco4471* mutant was then visualized after growth at 30 ^oC for three days.

Western blot analysis. To begin, 3×FLAG-σ^E-complemented Δ*sigE* and Δ*sigE* J2130 were treated in 5 ml TES buffer (250mM N-[tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid, pH7.2) for 10 min at 50 ^oC and germination carried out in 5 ml 2×PG (0.5 ml of 10% yeast extract, 0.25 ml of 20% casamino acids, 0.05ml of 1M $CaCl_2$ and 4.2 ml of H_2O) medium for 2-3 hours. Following this, germinated spores were span down at 4500 xg for 10 min and inoculated into 50 ml NMMP in 250 ml canonical flasks with springs to achieve a final OD_{450} of about 0.010, then grown at 30 ^oC, shaking at 250 rpm. At OD_{450}~ 0.6, vancomycin was added to a final concentration of 10 µg/ml and the cells further grown with shaking for 15min, 30min, 45min or 60 min respectively.

For Western blotting, for each time point, post-vancomycin induction (0, 15, 30, 45 and 60 min), 5ml of each culture was taken and span down at 3000 rpm for 1min. Cells were washed in 5ml ice-cold sonication buffer [20 mM Tris pH 8.0, 5 mM EDTA, 1 x EDTA-free protease inhibitors (Roche)] and finally resuspended into 1 ml. The samples were then sonicated on ice for five cycles (15 sec on/15 sec off) at 4.5 micron amplitude plus two cycles (15 sec on/15 sec off) at 9 micron amplitude. Subsequently, cell debris was removed from the lysates by centrifugation at 16,000 xg for 15 min at 4˚C. The supernatant was harvested and its protein concentration was determined by the Bradford assay (Biorad). Supernatant with the protein amount of 10 ug from each sample was then loaded on a 12.5% polyacrylamide SDS-PAGE gel. After electrophoresis, proteins were

then transferred to a Hybond-C Extra nylon membrane (Amerhsam Pharmacia Biotech) by the Invitrogen XCell II Blot system. For detection of 3xFLAG-σ^E, anti-σ^E polyclonal antibody raised from rabbit (diluted 1:300) and anti-rabbit IgG alkaline phosphatase secondary antibody (sigma A8025) (diluted 3:5000) were used. After detection, 3xFLAG-σ^E was visualized on the membrane directly by the SigmaFast system (Sigma) that employs BCIP/NBT (5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium) as a substrate.

RNA isolation and DNA microarray analysis. RNA isolation from *S.coelicolor ΔsigE* mutant J2130 was performed as described by Hong et al (44). Total RNA was isolated from mycelia harvested from 5 ml liquid cultures using an RNeasy Midi Kit (Qiagen) according to the manufacturer's instruction with some modifications. The cell pellet was resuspended in TE buffer containing lysozyme (10 mM Tris, pH 8, 1mM EDTA, 15 mg/ml lysozyme) and incubated at room temperature for 60 minutes. RLT buffer (Qiagen) was added (4 ml) and samples were sonicated 3 cycles ON-OFF on ice at 18 micron amplitude and for 20 seconds. Samples were then extracted twice with Phenol: Chloroform: Isoamyl Alcohol 25:24:1 saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (2 ml) and once with chloroform (4 ml). Extracts were mixed with 100% ethanol and applied to RNeasy Midi columns. Purified RNA was eluted with 300 µl RNase-free water. Affymetrix Gene Chip hybridization and data collection were essentially as described before (67, 130). The data analysis was conducted as described by Bush et al. (52). The expression valulues were then retreived from the *MArrayLM* object obtained in the above analysis and directly used to generate the graphs shown in this paper.

Chromatin immunoprecipitation Sequencing. *S.coelicolor* strain wild type M600 and 3×FLAG-σ^E-complemented Δ*sigE* spores were germinated and grown as described for the Western blot analysis. For the Chromatin immunoprecipitation (ChIP), the cell-envelope stress response was induced by treatment with vancomycin to a final concentration of 10 µg/ml and for 30 min. Following this, formaldehyde was added to cultures at a final concentration of 1% (v/v) and incubation was continued at 30^{o}C with shaking for a further 30 min.Glycine was then added to a final concentration of 125 mM to stop the cross-linking. Cells were then harvested, lysed, sonicated and the immunoprecipitation conducted via M2 (Sigma Aldrich A2220) gel suspension. Subsequent steps were

conducted as described by Bush et al. (52). Notably, for each tested strain, while immunoprecipitated DNA was used as a ChIP (input) sample, the non-immunoprecipitated total DNA was used as a reference sample. Sequence analysis was conducted as described by Bush et al. (52).

Prediction of the promoter motif for all the ChIP-seq targets

Initially, at least 200 bp sequences surrounding the highest enriched "25 bp" genomic region of all the ChIP-seq targets were extracted. Then, over-represented 2-block motifs mimicking a typical promoter with conserved "-35" and "-10" regions were identified in the forward strand of these sequences by the BioProspector program using the parameters: "W=4", "w=5", "G=17", "g=16" and "G-g=1 bp" ("W" and "w" stand for the length of the upstream and downstream motifs, respectively; "G" and "g" stand for the maximum and minimum distances between the 2 blocks, respectively) (64). 2-block motifs were obtained from iterative searches using these parameters. After 40 reinitializations, the highest scoring motifs were then selected to represent the σ^E binding sites since they highly resemble the previous reported σ^E promoter motif.

Bioinformatic prediction of the σ^E regulon and establishment a σ^E core-regulon

In order to predict the σ^E regulon, two promoter PWMs, PWM_19_16 and PWM_19_17 were firstly built from the 19 validated promoter sequences as shown in Fig. 3 by restricting the spacer between "-35" region and "-10" region to be 16 bp and 17 bp, respectively. In the case of PWM_19_16, one base was removed from non-conserved region of *sco3194* and *sco4934* promoters respectively, whereas, in the case of PWM_19_17, one base was added into the non-conserved region of each promoter with 16 bp between "-35" region and "-10" region. Then, these promoter PWMs were used to search for all putative σ^E binding sites from 19 *Streptomyces spp.* chromosome sequence using the Virtual Footprint version 3.0 tool incorporated into the PRODORIC server (http://www.prodoric.de/vfp/vfp_regulon.php) (65, 66) with the parameters: "Non-Occurance Penalty=None", "Sensitivity =1", "Core Sensitivity/Size =1/6". The list of all putative promoters was then restricted to those located in the same strand and between 10 and 200 bp upstream from the start codon of the closest predicted coding sequence. Their putative downstream target genes were then assigned into the σ^E regulons. Subsequently, orthologs of these targets were determined by a reciprocal best-hits analysis using

BLASTP program. Briefly, for every target in a given *Streptomyces* genome, its protein product was firstly used to search for all the similar proteins encoded in other *Streptomyces* genomes by BLASTP program. If two proteins were top-BLASTP hits of each other, they were considered as reciprocal best hits. The proteins, which were reciprocal best hits, were then defined as orthologs if their identity and coverage were at least 25% and 50%, respectively. Alternatively, the proteins, which were not reciprocal best hits, were defined as orthologs if their identity and coverage were at least 70% and 80%, respectively. The σ^E core regulon was defined as those orthologs predicted to have a σ^E dependent promoter in at least 9 of the 19 genomes analysed.

S1 nuclease mapping.

To generate the probes, a reverse primer within 80 bp downstream of the start codon of each gene was first labelled with [ɣ-32P] ATP. Amplification was then conducted from a template using the labelled reverse primer and a forward primer approximately 400 bp upstream of the start codon. For all assays, 30 μg of RNA and 25 pmol of labelled probe were dissolved in 20 μl of NaTCA buffer and hybridized at 45°C overnight after denaturation at 65°C for 15 min. Primer sequences used in S1 nuclease mapping are listed in Table S7. Sequencing ladders were generated by using Sequenase™ Version 2.0 DNA Sequencing Kit (USB Europe GMBH).

Purification of σ^E and *in vitro* transcription assays. σ^E was overexpressed and purified to homogeneity as described previously (45). Run-off transcription assays were performed using [α-^{32}P]-CTP (Perkin Elmer) at 3000 Ci mmol^{-1} as described previously (131). Reaction mixtures contained 1.25 pmol of *E. coli* core RNA polymerase (Epicentre Technologies) and 6 pmol of σ^E. Transcripts were analysed on a 6% (w/v) polyacrylamide-7 M urea gel using a heat-denatured, ^{32}P-labelled *Hpa*II digest of pBR322 as size standards. Cold RNase-free PCR probes generated as for the S1 mapping experiments were used as templates.

Gene disruption of *sco2611* (*mreB*) and *sco4471.* Using the 'Redirect' PCR targeting method of Gust *et al*. (132, 133), *sco2611 (mreB)* and *sco4471* mutants were generated in which the coding regions were replaced with a single apramycin resistance (*apr*) cassette. Cosmid C49 (*mreB*) or cosmid D65 (*sco4471)* was introduced into *E. coli* BW25113

containing pIJ790 and the relevant gene was replaced with the *apr-oriT* cassette amplified from pIJ773 using appropriate primer pairs SCO2611KOFW, SCO2611KORV and SCO4471KOFW, SCO4471KORV respectively (Table S7). The resulting disrupted cosmids were confirmed by restriction digestion and by PCR analysis using appropriate flanking primers (Table S7), and introduced into *S. coelicolor* by conjugation via the methylation-deficient *E. coli* strain ET12567 (*dam dcm hsdS*) carrying the driver plasmid pUZ8002. Null mutant derivatives, generated by double crossing over, were identified by their apramycin-resistant, kanamycin-sensitive phenotypes, and their chromosomal structures were confirmed by PCR analysis using appropriate flanking primers (Table S7) and by Southern hybridization. Finally, for Δ*mreB*, the apramycin marker was flipped out from chromosome according using the flip recombinase (132, 133), leaving a clean deletion with a short scar-sequence.

Scanning electron microscopy of *S. coelicolor* Δ*sco4471*. For scanning electron microscopy, five day old colonies were mounted on the surface of an aluminum stub with optimal cutting temperature compound (BDH Laboratory Supplies, Poole, England). The stub was then immediately plunged into liquid nitrogen slush at approximately -210^{o}C to cryo-preserve the material and transferred to the cryostage of an ALTO 2500 cryotransfer system (Gatan, Oxford, England) attached to a Zeiss Supra 55 VP field emission gun scanning electron microscope (Zeiss SMT, Germany). The surface frost was sublimated at -95°C for 3 min before the sample was sputter coated with platinum for 2 min at 10 mA at below -110°C. After sputter-coating, the sample was moved onto the cryo-stage in the main chamber of the microscope, held at approximately -130^{o}C. The sample was imaged at 3kV and digital TIFF files were stored.

ACKNOWLEDGEMENTS

This work is supported by a Dorothy Hodgkin Postgraduate Award to Ngat T. Tran from the BBSRC, by a postgraduate stipend to Xiaoluo Huang from China Scholarship Council, by the DFG grant MA2837/2-2, and by the MET Institute Strategic Programme Grant to the John Innes Centre from the BBSRC.

References

1. **Silhavy TJ, Kahne D, Walker S.** 2010. The bacterial cell envelope. Cold Spring Harb. Perspect. Biol. **2:**a000414.
2. **Jordan S, Hutchings MI, Mascher T.** 2008. Cell envelope stress response in Gram-positive bacteria. FEMS Microbiol. Rev. **32:**107-146.
3. **Costerton J, Cheng K-J.** 1975. The role of the bacterial cell envelope in antibiotic resistance. J. Antimicrob. Chemother. **1:**363-377.
4. **Strominger JL, Tipper DJ.** 1965. Bacterial cell wall synthesis and structure in relation to the mechanism of action of penicillins and other antibacterial agents. Am. J. Med. **39:**708-721.
5. **Kahne D, Leimkuhler C, Lu W, Walsh C.** 2005. Glycopeptide and lipoglycopeptide antibiotics. Chem. Rev. **105:**425-448.
6. **Raivio TL.** 2005. Envelope stress responses and Gram-negative bacterial pathogenesis. Mol. Microbiol. **56:**1119-1128.
7. **Gao R, Stock AM.** 2009. Biological insights from structures of two-component proteins. Annu. Rev. Microbiol. **63:**133.
8. **Stock AM, Robinson VL, Goudreau PN.** 2000. Two-component signal transduction. Annu. Rev. Biochem. **69:**183-215.
9. **Mascher T.** 2013. Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. Curr. Opin. Biotechnol. **16:**148-155.
10. **Brooks BE, Buchanan SK.** 2008. Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. Biochim. Biophys. Acta. **1778:**1930-1945.
11. **Staroń A, Sofia HJ, Dietrich S, Ulrich LE, Liesegang H, Mascher T.** 2009. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. Mol. Microbiol. **74:**557-581.
12. **Lonetto MA, Brown KL, Rudd KE, Buttner MJ.** 1994. Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. Proc. Natl. Acad. Sci. USA **91:**7573-7577.
13. **Paget M, Helmann JD.** 2003. The sigma70 family of sigma factors. Genome Biol **4:**203.
14. **Helmann JD.** 1995. Compilation and analysus of *Bacillus Subtilis* σ^A-dependent promoter sequences: evidence for extended contact between RNA polymerse and upstream promoter DNA. Nucleic Acids Res. **23:**2351-2360.
15. **Travers AA.** 1987. Structure and function of *E. coli* promoter DNA. Crit. Rev. Biochem. Mol. Biol. **22:**181-219.
16. **Ruiz N, Silhavy TJ.** 2005. Sensing external stress: watchdogs of the the *Escherichia coli* cell envelope. Curr. Opin. Microbiol. **8:**122-126.
17. **Danese PN, Silhavy TJ.** 1998. CpxP, a stress-combative member of the Cpx regulon. J. Bacteriol. **180:**831-839.
18. **Hirakawa H, Inazumi Y, Masaki T, Hirata T, Yamaguchi A.** 2005. Indole induces the expression of multidrug exporter genes in *Escherichia coli*. Mol. Microbiol. **55:**1113-1126.
19. **Jubelin G, Vianney A, Beloin C, Ghigo J-M, Lazzaroni J-C, Lejeune P, Dorel C.** 2005. CpxR/OmpR interplay regulates curli gene expression in response to osmolarity in *Escherichia coli*. J. Bacteriol.**187:**2038-2049.

20. **Snyder WB, Davis L, Danese PN, Cosma CL, Silhavy TJ.** 1995. Overproduction of NlpE, a new outer membrane lipoprotein, suppresses the toxicity of periplasmic LacZ by activation of the Cpx signal transduction pathway. J. Bacteriol. **177:**4216-4223.
21. **Mileykovskaya E, Dowhan W.** 1997. The Cpx two-component signal transduction pathway is activated in *Escherichia coli* mutant strains lacking phosphatidylethanolamine. J. Bacteriol. **179:**1029-1034.
22. **Hunke S, Betton JM.** 2003. Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli.* Mol. Microbiol. **50:**1579-1589.
23. **Hal Jones C, Danese PN, Pinkner JS, Silhavy TJ, Hultgren SJ.** 1997. The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems. EMBO J. **16:**6394-6406.
24. **Danese PN, Silhavy TJ.** 1997. The sigma (E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli.* Genes Dev. **11:**1183-1193.
25. **Danese PN, Snyder WB, Cosma CL, Davis L, Silhavy TJ.** 1995. The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. Genes Dev. **9:**387-398.
26. **Pogliano J, Lynch AS, Belin D, Lin E, Beckwith J.** 1997. Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. Genes Dev. **11:**1169-1182.
27. **Raivio TL, Silhavy TJ.** 2001. Periplasmic stress and ECF sigma factors. Annu. Rev. Microbiol. **55:**591-624.
28. **Missiakas D, Betton JM, Raina S.** 1996. New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. Mol. Microbiol.**21:**871-884.
29. **Rhodius VA, Suh WC, Nonaka G, West J, Gross CA.** 2005. Conserved and variable functions of the σ^E stress response in related genomes. PLoS Biol. **4:**e2.
30. **Dartigalongue C, Missiakas D, Raina S.** 2001. Characterization of the *Escherichia coli* σ^E Regulon. J. Biol. Chem.**276:**20866-20875.
31. **Johansen J, Rasmussen AA, Overgaard M, Valentin-Hansen P.** 2006. Conserved small non-coding RNAs that belong to the σ^E regulon: role in down-regulation of outer membrane proteins. J. Mol. Biol. **364:**1-8.
32. **Thompson KM, Rhodius VA, Gottesman S.** 2007. σ^E regulates and is regulated by a small RNA in *Escherichia coli*. J. Bacteriol.**189:**4243-4256.
33. **Udekwu KI, Wagner EGH.** 2007. SigmaE controls biogenesis of the antisense RNA MicA. Nucleic Acids Res. **35:**1279-1288.
34. **Mascher T, Margulis NG, Wang T, Ye RW, Helmann JD.** 2003. Cell wall stress responses in *Bacillus subtilis*: the regulatory network of the bacitracin stimulon. Mol. Microbiol.**50:**1591-1604.
35. **Thackray PD, Moir A.** 2003. SigM, an extracytoplasmic function sigma factor of *Bacillus subtilis*, is activated in response to cell wall antibiotics, ethanol, heat, acid, and superoxide stress. J. Bacteriol. **185:**3491-3498.
36. **Kingston AW, Liao X, Helmann JD.** 2013. Contributions of the σ^W, σ^M and σ^X regulons to the lantibiotic resistome of *Bacillus subtilis*. Mol. Microbiol. **90:**502-518.

37. **Eiamphungporn W, Helmann JD.** 2008. The *Bacillus subtilis* σ^M regulon and its contribution to cell envelope stress responses. Mol. Microbiol. **67:**830-848.
38. **Flärdh K, Buttner MJ.** 2009. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. Nat. Rev. Microbiol. **7:**36-49.
39. **Hutchings MI, Hoskisson PA, Chandra G, Buttner MJ.** 2004. Sensing and responding to diverse extracellular signals? Analysis of the sensor kinases and response regulators of *Streptomyces coelicolor* A3 (2). Microbiology **150:**2795-2806.
40. **Ulrich LE, Zhulin IB.** 2009. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. Nucleic Acids Res. **38**:D401-D407.
41. **Hong HJ, Hutchings MI, Neu JM, Wright GD, Paget MS, Buttner MJ.** 2004. Characterization of an inducible vancomycin resistance system in *Streptomyces coelicolor* reveals a novel gene (*vanK*) required for drug resistance. Mol. Microbiol. **52:**1107-1121.
42. **Hutchings MI, Hong HJ, Buttner MJ.** 2006. The vancomycin resistance VanRS two-component signal transduction system of *Streptomyces coelicolor*. Mol. Microbiol. **59:**923-935.
43. **Paget MS, Leibovitz E, Buttner MJ.** 1999. A putative two-component signal transduction system regulates σ^E, a sigma factor required for normal cell wall integrity in *Streptomyces coelicolor* A3 (2). Mol. Microbiol. **33:**97-107.
44. **Hong HJ, Paget MS, Buttner MJ.** 2002. A signal transduction system in *Streptomyces coelicolor* that activates the expression of a putative cell wall glycan operon in response to vancomycin and other cell wall-specific antibiotics. Mol. Microbiol. **44:**1199-1211.
45. **Paget MS, Chamberlin L, Atrih A, Foster SJ, Buttner MJ.** 1999. Evidence that the extracytoplasmic function sigma factor σ^E is required for normal cell wall structure in *Streptomyces coelicolor* A3 (2). J. Bacteriol. **181:**204-211.
46. **Hutchings MI, Hong H-J, Leibovitz E, Sutcliffe IC, Buttner MJ.** 2006. The σ^E cell envelope stress response of *Streptomyces coelicolor* is influenced by a novel lipoprotein, CseA. J. Bacteriol. **188:**7222-7229.
47. **Ferreira AO, Myers CR, Gordon JS, Martin GB, Vencato M, Collmer A, Wehling MD, Alfano JR, Moreno-Hagelsieb G, Lamboy WF.** 2006. Whole-genome expression profiling defines the HrpL regulon of *Pseudomonas syringae* pv. tomato DC3000, allows de novo reconstruction of the Hrp cis element, and identifies novel coregulated genes. Mol. Plant Microbe. Interact. **19:**1167-1179.
48. **Lan L, Deng X, Zhou J, Tang X.** 2006. Genome-wide gene expression analysis of *Pseudomonas syringae* pv. tomato DC3000 reveals overlapping and distinct pathways regulated by hrpL and hrpRS. Mol. Plant Microbe. Interact. **19:**976-987.
49. **Merighi M, Majerczak DR, Stover EH, Coplin DL.** 2003. The HrpX/HrpY two-component system activates *hrpS* expression, the first step in the regulatory cascade controlling the Hrp regulon in *Pantoea stewartii* subsp. stewartii. Mol. Plant Microbe. Interact. **16:**238-248.
50. **Vencato M, Tian F, Alfano JR, Buell CR, Cartinhour S, DeClerck GA, Guttman DS, Stavrinides J, Joardar V, Lindeberg M.** 2006. Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of *Pseudomonas syringae* pv. phaseolicola 1448A. Mol. Plant Microbe. Interact. **19:**1193-1206.

51. **Buttner M, Chater K, Bibb M.** 1990. Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3 (2). J. Bacteriol. **172:**3367-3378.
52. **Bush MJ, Bibb MJ, Chandra G, Findlay KC, Buttner MJ.** 2013. Genes required for aerial growth, cell division, and chromosome segregation are targets of WhiA before sporulation in *Streptomyces venezuelae*. MBio **4:**e00684-00613.
53. **Al-Bassam MM, Bibb MJ, Bush MJ, Chandra G, Buttner MJ.** 2014. Response regulator heterodimer formation controls a key stage in *Streptomyces* Development. PLoS Genet. **10:**e1004554.
54. **Teytelman L, Özaydın B, Zill O, Lefrançois P, Snyder M, Rine J, Eisen MB.** 2009. Impact of chromatin structures on DNA processing for genomic analyses. PloS one **4:**e6700.
55. **Mascher T, Hachmann A-B, Helmann JD.** 2007. Regulatory overlap and functional redundancy among *Bacillus subtilis* extracytoplasmic function σ factors. J. Bacteriol. **189:**6919-6927.
56. **Moreno-Hagelsieb G, Collado-Vides J.** 2002. A powerful non-homology method for the prediction of operons in prokaryotes. Bioinformatics **18:**S329-S336.
57. **Burger A, Sichler K, Kelemen G, Buttner M, Wohlleben W.** 2000. Identification and characterization of the *mre* gene region of *Streptomyces coelicolor* A3 (2). Mol. Gen. Genet. **263:**1053-1060.
58. **Kleinschnitz EM, Heichlinger A, Schirner K, Winkler J, Latus A, Maldener I, Wohlleben W, Muth G.** 2011. Proteins encoded by the mre gene cluster in *Streptomyces coelicolor* A3 (2) cooperate in spore wall synthesis. Mol. Microbiol. **79:**1367-1379.
59. **Krogh A, Larsson B, Von Heijne G, Sonnhammer EL.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567-580.
60. **Sonnhammer EL, Von Heijne G, Krogh A.** A hidden Markov model for predicting transmembrane helices in protein sequences, Proc. Int. Conf. Intell. Syst. Mol. Biol. **6**:175-82.
61. **Letunic I, Doerks T, Bork P.** 2015. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. **43:** D257-D260.
62. **Schultz J, Milpetz F, Bork P, Ponting CP.** 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA **95:**5857-5864.
63. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J.** 2014. Pfam: the protein families database. Nucleic Acids Res.**42:** D222-D230.
64. **Liu X, Brutlag DL, Liu JS.** BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes, Pac. Symp. Biocomput. 127-138.
65. **Münch R, Hiller K, Grote A, Scheer M, Klein J, Schobert M, Jahn D.** 2005. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. Bioinformatics **21:**4187-4189.
66. **Grote A, Klein J, Retter I, Haddad I, Behling S, Bunk B, Biegler I, Yarmolinetz S, Jahn D, Münch R.** 2009. PRODORIC (release 2009): a database and tool platform for the analysis of gene regulation in prokaryotes. Nucleic Acids Res. **37:**D61-D65.

67. **Hesketh A, Hill C, Mokhtar J, Novotna G, Tran N, Bibb M, Hong H-J.** 2011. Genome-wide dynamics of a bacterial response to antibiotics that target the cell envelope. BMC genomics **12:**226.
68. **Crooks GE, Hon G, Chandonia J-M, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome Res. **14:**1188-1190.
69. **Wade JT, Roa DC, Grainger DC, Hurd D, Busby SJ, Struhl K, Nudler E.** 2006. Extensive functional overlap between σ factors in *Escherichia coli*. Nat. Struct. Mol. Biol. **13:**806-814.
70. **Qiu J, Helmann JD.** 2001. The -10 region is a key promoter specificity determinant for the *Bacillus subtilis* extracytoplasmic-function σ factors σ^X and σ^W. J. Bacteriol. **183:**1921-1927.
71. **Bentley S, Chater K, Cerdeno-Tarraga A-M, Challis G, Thomson N, James K, Harris D, Quail M, Kieser H, Harper D.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). Nature **417:**141-147.
72. **Strakova E, Zikova A, Vohradsky J.** 2014. Inference of sigma factor controlled networks by using numerical modeling applied to microarray time series data of the germinating prokaryote. Nucleic Acids Res. **42:**748-763.
73. **Macheboeuf P, Contreras-Martel C, Job V, Dideberg O, Dessen A.** 2006. Penicillin binding proteins: key players in bacterial cell cycle and drug resistance processes. FEMS Microbiol. Rev. **30:**673-691.
74. **Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P.** 2008. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol. Rev. **32:**234-258.
75. **Denome SA, Elf PK, Henderson TA, Nelson DE, Young KD.** 1999. *Escherichia coli* mutants lacking all possible combinations of eight penicillin binding proteins: viability, characteristics, and implications for peptidoglycan synthesis. J. Bacteriol. **181:**3981-3993.
76. **Hoskins J, Matsushima P, Mullen DL, Tang J, Zhao G, Meier TI, Nicas TI, Jaskunas SR.** 1999. Gene disruption studies of penicillin-binding proteins 1a, 1b, and 2a in *Streptococcus pneumoniae*. J. Bacteriol. **181:**6552-6555.
77. **Dasgupta A, Datta P, Kundu M, Basu J.** 2006. The serine/threonine kinase PknB of *Mycobacterium tuberculosis* phosphorylates PBPA, a penicillin-binding protein required for cell division. Microbiology **152:**493-504.
78. **Mercer KL, Weiss DS.** 2002. The *Escherichia coli* cell division protein FtsW is required to recruit its cognate transpeptidase, FtsI (PBP3), to the division site. J. Bacteriol. **184:**904-912.
79. **Chambers HF.** 1997. Methicillin resistance in *Staphylococci*: molecular and biochemical basis and clinical implications. Clin. Microbiol. Rev. **10:**781-791.
80. **Katayama Y, Zhang H-Z, Chambers HF.** 2004. PBP 2a mutations producing very-high-level resistance to beta-lactams. Antimicrob. Agents Chemother. **48:**453-459.
81. **Lim D, Strynadka NC.** 2002. Structural basis for the β-lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. Nat. Struct. Mol. Biol. **9:**870-876.
82. **Fontana R, Aldegheri M, Ligozzi M, Lopez H, Sucari A, Satta G.** 1994. Overproduction of a low-affinity penicillin-binding protein and high-level ampicillin resistance in *Enterococcus faecium*. Antimicrob. Agents Chemother. **38:**1980-1983.

83. **Sauvage E, Kerff F, Fonze E, Herman R, Schoot B, Marquette J-P, Taburet Y, Prevost D, Dumas J, Leonard G.** 2002. The 2.4-Å crystal structure of the penicillin-resistant penicillin-binding protein PBP5fm from *Enterococcus faecium* in complex with benzylpenicillin. Cell. Mol. Life Sci. **59:**1223-1232.
84. **Jiang H, Kendrick KE.** 2000. Cloning and characterization of the gene encoding penicillin-binding protein A of *Streptomyces griseus*. FEMS Microbiol. Lett. **193:**63-68.
85. **Biarrotte-Sorin S, Hugonnet J-E, Delfosse V, Mainardi J-L, Gutmann L, Arthur M, Mayer C.** 2006. Crystal structure of a novel β-lactam-insensitive peptidoglycan transpeptidase. J. Mol. Biol. **359:**533-538.
86. **Lavollay M, Arthur M, Fourgeaud M, Dubost L, Marie A, Veziris N, Blanot D, Gutmann L, Mainardi J-L.** 2008. The peptidoglycan of stationary-phase *Mycobacterium tuberculosis* predominantly contains cross-links generated by L, D-transpeptidation. J. Bacteriol. **190:**4360-4366.
87. **Magnet S, Bellais S, Dubost L, Fourgeaud M, Mainardi J-L, Petit-Frère S, Marie A, Mengin-Lecreulx D, Arthur M, Gutmann L.** 2007. Identification of the L, D-transpeptidases responsible for attachment of the Braun lipoprotein to *Escherichia coli* peptidoglycan. J. Bacteriol. **189:**3927-3931.
88. **Yem DW, Wu HC.** 1978. Physiological characterization of an *Escherichia coli* mutant altered in the structure of murein lipoprotein. J. Bacteriol. **133:**1419-1426.
89. **Hayashi S, Wu HC.** 1990. Lipoproteins in bacteria. J. Bioenerg. Biomembr. **22:**451-471.
90. **Kawai Y, Marles-Wright J, Cleverley RM, Emmins R, Ishikawa S, Kuwano M, Heinz N, Bui NK, Hoyland CN, Ogasawara N.** 2011. A widespread family of bacterial cell wall assembly proteins. EMBO J. **30:**4931-4941.
91. **Chan YG, Frankel MB, Dengler V, Schneewind O, Missiakas D.** 2013. *Staphylococcus aureus* mutants lacking the LytR-CpsA-Psr family of enzymes release cell wall teichoic acids into the extracellular medium. J. Bacteriol.**195:**4650-4659.
92. **Chan YG-Y, Kim HK, Schneewind O, Missiakas D.** 2014. The capsular polysaccharide of *Staphylococcus aureus* is attached to peptidoglycan by the LytR-CpsA-Psr (LCP) family of enzymes. J. Biol. Chem. **289:**15680-15690.
93. **Zilla ML, Chan YG, Lunderberg JM, Schneewind O, Missiakas D.** 2015. LytR-CpsA-Psr enzymes as determinants of *Bacillus anthracis* secondary cell wall polysaccharide assembly. J. Bacteriol.**197:**343-353.
94. **Brown S, Santa Maria Jr JP, Walker S.** 2013. Wall teichoic acids of gram-positive bacteria. Annu. Rev. Microbiol. **67**: 313-336.
95. **Rossi J, Bischoff M, Wada A, Berger-Bächi B.** 2003. MsrR, a putative cell envelope-associated element involved in *Staphylococcus aureus* sarA attenuation. Antimicrob. Agents Chemother. **47:**2558-2564.
96. **Over B, Heusser R, McCallum N, Schulthess B, Kupferschmied P, Gaiani JM, Sifri CD, Berger-Bächi B, Stutzmann Meier P.** 2011. LytR-CpsA-Psr proteins in *Staphylococcus aureus* display partial functional redundancy and the deletion of all three severely impairs septum placement and cell separation. FEMS Microbiol. Lett. **320:**142-151.
97. **Wu C, Huang I, Chang C, Reardon-Robinson ME, Das A, Ton-That H.** 2014. Lethality of sortase depletion in *Actinomyces oris* caused by excessive membrane accumulation of a surface glycoprotein. Mol. Microbiol. **94:**1227-1241.

98. **Carballido-López R, Formstone A.** 2007. Shape determination in *Bacillus subtilis*. Curr. Opin. Microbiol. **10:**611-616.
99. **Figge RM, Divakaruni AV, Gober JW.** 2004. MreB, the cell shape-determining bacterial actin homologue, co-ordinates cell wall morphogenesis in *Caulobacter crescentus.* Mol. Microbiol.**51:**1321-1332.
100. **Osborn MJ, Rothfield L.** 2007. Cell shape determination in *Escherichia coli*. Curr. Opin. Microbiol. **10:**606-610.
101. **Mazza P, Noens EE, Schirner K, Grantcharova N, Mommaas AM, Koerten HK, Muth G, Flärdh K, Van Wezel GP, Wohlleben W.** 2006. MreB of *Streptomyces coelicolor* is not essential for vegetative growth but is required for the integrity of aerial hyphae and spores. Mol. Microbiol.**60:**838-852.
102. **Flärdh K.** 2003. Essential role of DivIVA in polar growth and morphogenesis in *Streptomyces coelicolor* A3 (2). Mol. Microbiol.**49:**1523-1536.
103. **Davis N, Chater K.** 1992. The *Streptomyces coelicolor whiB* gene encodes a small transcription factor-like protein dispensable for growth but essential for sporulation. Mol. Gen. Genet. **232:**351-358.
104. **Soliveri J, Gomez J, Bishai W, Chater K.** 2000. Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene *whiB* are present in *Streptomyces* and other actinomycetes. Microbiology **146:**333-343.
105. **Jakimowicz P, Cheesman MR, Bishai WR, Chater KF, Thomson AJ, Buttner MJ.** 2005. Evidence that the *Streptomyces* developmental protein WhiD, a member of the WhiB family, binds a [4Fe-4S] cluster. J. Biol. Chem. **280:**8309-8315.
106. **Singh A, Guidry L, Narasimhulu K, Mai D, Trombley J, Redding KE, Giles GI, Lancaster JR, Steyn AJ.** 2007. *Mycobacterium tuberculosis* WhiB3 responds to O2 and nitric oxide via its [4Fe-4S] cluster and is essential for nutrient starvation survival. Proc. Natl. Acad. Sci. USA **104:**11562-11567.
107. **Stapleton MR, Smith LJ, Hunt DM, Buxton RS, Green J.** 2012. *Mycobacterium tuberculosis* WhiB1 represses transcription of the essential chaperonin GroEL2. Tuberculosis (Edinb) **92:**328-332.
108. **Burian J, Yim G, Hsing M, Axerio-Cilies P, Cherkasov A, Spiegelman GB, Thompson CJ.** 2013. The mycobacterial antibiotic resistance determinant WhiB7 acts as a transcriptional activator by binding the primary sigma factor SigA (RpoV). Nucleic Acids Res. **41**:10062–10076.
109. **Geiman DE, Raghunand TR, Agarwal N, Bishai WR.** 2006. Differential gene expression in response to exposure to antimycobacterial agents and other stress conditions among seven *Mycobacterium tuberculosis whiB*-like genes. Antimicrob. Agents Chemother. **50:**2836-2841.
110. **Kang J-G, Hahn M-Y, Ishihama A, Roe J-H.** 1997. Identification of sigma factors for growth phase-related promoter selectivity of RNA polymerases from *Streptomyces coelicolor* A3 (2). Nucleic Acids Res.**25:**2566-2573.
111. **Soliveri J, Brown K, Buttner M, Chater K.** 1992. Two promoters for the *whiB* sporulation gene of *Streptomyces coelicolor* A3 (2) and their activities in relation to development. J. Bacteriol.**174:**6215-6220.
112. **Ernst CM, Staubitz P, Mishra NN, Yang S-J, Hornig G, Kalbacher H, Bayer AS, Kraus D, Peschel A.** 2009. The bacterial defensin resistance protein MprF consists of separable domains for lipid lysinylation and antimicrobial peptide repulsion. PLoS Pathog. **5:**e1000660.

113. **Ruzin A, Severin A, Moghazeh SL, Etienne J, Bradford PA, Projan SJ, Shlaes DM.** 2003. Inactivation of *mprF* affects vancomycin susceptibility in *Staphylococcus aureus* Biochim. Biophys. Acta. **1621:**117-121.
114. **Nishi H, Komatsuzawa H, Fujiwara T, McCallum N, Sugai M.** 2004. Reduced content of lysyl-phosphatidylglycerol in the cytoplasmic membrane affects susceptibility to moenomycin, as well as vancomycin, gentamicin, and antimicrobial peptides, in *Staphylococcus aureus*. Antimicrob. Agents Chemother. **48:**4800-4807.
115. **Friedman L, Alder JD, Silverman JA.** 2006. Genetic changes that correlate with reduced susceptibility to daptomycin in *Staphylococcus aureus*. Antimicrob. Agents Chemother. **50:**2137-2145.
116. **Vrancken K, Van Mellaert L, Anné J.** 2008. Characterization of the *Streptomyces lividans* PspA response. J. Bacteriol.**190:**3475-3481.
117. **Joly N, Engl C, Jovanovic G, Huvet M, Toni T, Sheng X, Stumpf MP, Buck M.** 2010. Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. FEMS Microbiol. Rev. **34:**797-827.
118. **Darwin AJ.** 2005. The phage-shock-protein response. Mol. Microbiol.**57:**621-628.
119. **Campbell EA, Muzzin O, Chlenov M, Sun JL, Olson CA, Weinman O, Trester-Zedlitz ML, Darst SA.** 2002. Structure of the bacterial RNA polymerase promoter specificity σ subunit. Mol. Cell **9:**527-539.
120. **Murakami KS, Darst SA.** 2003. Bacterial RNA polymerases: the wholo story. Curr. Opin. Struct. Biol. **13:**31-39.
121. **Lane WJ, Darst SA.** 2006. The structural basis for promoter -35 element recognition by the group IV σ factors. PLoS Biol. **4:**e269.
122. **Rhodius VA, Segall-Shapiro TH, Sharon BD, Ghodasara A, Orlova E, Tabakh H, Burkhardt DH, Clancy K, Peterson TC, Gross CA.** 2013. Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. Mol Syst. Biol. **9: 702**.
123. **Cao M, Kobel PA, Morshedi MM, Wu MFW, Paddon C, Helmann JD.** 2002. Defining the *Bacillus subtilis* σ^W regulon: a comparative analysis of promoter consensus search, run-off transcription/macroarray analysis (ROMA), and transcriptional profiling approaches. J. Mol. Biol. **316:**443-457.
124. **Notredame C, Higgins DG, Heringa J.** 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302:**205-217.
125. **Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS.** 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. **38:**D396-D400.
126. **Hesketh A, Deery M, Hong H-J.** 2015. High-resolution mass spectrometry based proteomic analysis of the response to vancomycin-induced cell wall stress in *Streptomyces coelicolor* A3 (2). J. Proteome Res. **14:** 2915-2928.
127. **Kieser T, Bibb M, Buttner M, Chater K, Hopwood D.** 2000. Practical Streptomyces Genetics (The John Innes Foundation, Norwich, United Kingdom).
128. **Gregory MA, Till R, Smith MC.** 2003. Integration site for *Streptomyces* phage φBT1 and development of site-specific integrating vectors. J. Bacteriol.**185:**5320-5323.
129. **MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T.** 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene **111:**61-68.

130. **Hesketh A, Chen WJ, Ryding J, Chang S, Bibb M.** 2007. The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3 (2). Genome Biol **8:**R161.
131. **Buttner MJ, Fearnley IM, Bibb MJ.** 1987. The agarase gene (*dagA*) of *Streptomyces coelicolor* A3 (2): nucleotide sequence and transcriptional analysis. Mol. Gen. Genet. **209:**101-109.
132. **Gust B, Challis GL, Fowler K, Kieser T, Chater KF.** 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc. Natl. Acad. Sci. USA. **100:**1541-1546.
133. **Gust B, Chandra G, Jakimowicz D, Yuqing T, Bruton CJ, Chater KF.** 2004. λ Red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. Adv. Appl. Microbiol. **54:**107-128.

Supplementary files for Chapter 4

The following supplementary files could be found in the attached CD.

Fig. S1. Western blot analysis of *S. coelicolor ΔsigE attB$_{\Phi BT1}$::3xFLAG-sigE* grown in NMMP liquid cultures and sampled after 0, 15, 30, 45 and 60 mins treatment with 10 µg/ml (w/v)vancomycin.

Fig. S2. Chromosome-wide distribution of σ^E binding sites in *S. coelicolor* identified by ChIP-seq analysis.

Fig. S3. Results from selected *in vitro* transcription assays (*sco3396* [A] *and sco4471* [B]), from promoters that show complete dependence on σ^E for transcription.

Fig. S4. ChIP-seq and microarray transcriptional profiling data for five selected σ^E-targets (*sco4471, sco4263, sco7657, sco3396* and *sco3397* (*mprF*) that have a single promoter and are completely dependent upon σ^E for their transcription (group 1, as defined by S1 mapping).

Fig. S5. ChIP-seq and microarray transcriptional profiling data for selected σ^E-targets (*sco2334, sco4134, sco5030, sco2897* and *sco4847, sco5358* [A] and *sco3044, cwg, sco7233, sco3712* [B]) that have a single promoter and are partially dependent upon σ^E for their transcription (group 2, as defined by S1 mapping).

Fig. S6. ChIP-seq and microarray transcriptional profiling data for selected σ^E-targets (*mreB, hrdD, sco3194, sco4934*) that have multiple promoters with one at least partially dependent upon σ^E for transcription (group 3, as defined by S1 mapping).

Fig. S7. Conservation of σ^E orthologues across 19 *Streptomyces* genomes.

Table S1. Complete ChIP-seq data set for *S.coelicolor* M600 (anti-FLAG) and *ΔsigE attBΦBT1*:: 3xFLAG-*sigE* (anti-FLAG) (ordered by significance).

Table S2. Selected 19 *Streptomyces* genomes used for core-regulon analysis.

Table S3. Putative σ^E binding sites of 19 *Streptomyces* genomes identified by genome-wide bioinformatics prediction.

Table S4. Predicted σ^E targets from 19 *Streptomyces* genomes.

Table S5. Number of targets identified by genome-wide bioinformatics prediction.

Table S6. Orthologues of targets belonging to the extended core σ^E regulon.

Table S7. Bacterial strains, plasmids and oligonucleotides.

Chapter 5

Discussion

5 Discussion

5.1 Comparative genomics of ECF σ factors in *Planctomycetes* and *Actinobacteria*

ECF σ factors represent a major type of signal transduction proteins in bacteria (Staroń & Mascher, 2010). They are involved in multiple physiological processes including stress response (Eiamphungporn & Helmann, 2008; Kim et al, 2012), transportation (Braun & Mahren, 2005) and cell differentiation (Bibb et al, 2012; Bibb et al, 2000). In the post-genomic era, a huge number of ECF sequences are annotated and deposited in the public database. This provides a possibility to phylogenetically group them, and subsequently identify conserved features related to their physiological roles and signal transduction mechanisms. This has already been demonstrated by a previous comparative genomic study in which approximately 1800 ECF σ factors from 369 genomes were assigned into 67 ECF groups with various putative physiological roles and signal transduction mechanisms (Staroń et al, 2009). In this study, through a comparative genomic study of ECF σ factors from eight planctomycetal genomes and 119 actinobacterial genomes, we identified 26 novel ECF groups. Based on the analysis of the conserved features, signal transduction mechanisms and physiological roles of many ECF groups were proposed (Chapter 2 and Chapter 3).

5.1.1 Comparative genomics, a powerful tool to identify novel ECF groups

The first comparative genomic study of ECF σ factors was conducted by Staroń *et al.* almost six years ago. In that study, 2708 ECF σ factors from 369 genomes were retrieved and 1873 ECF σ factors were classified into 67 groups (43 major ECF groups, ECF01-43, and 24 minor ECF groups, ECF101-124). The remaining 835 ECF σ factors could not be classified. This study also led to the classification of 22 out of 105 ECF σ factors from three sequenced planctomycetal genomes, and 363 out of 512 ECF σ factors mainly from approximately 35 sequenced actinobacterial genomes (Staroń et al, 2009). Later, in 2011, a novel ECF group (later, termed ECF44, (Mascher, 2013)) was proposed from the analysis of 21 ECF sequences that resemble a copper-dependent ECF σ factor, CorE (Gómez-Santos et al, 2011).

In our study, we retrieved 361 ECF σ factors from eight complete planctomycetal genomes (Chapter 2) and 2203 ECF σ factors from 119 complete actinobacterial genomes (Chapter 3). This dataset indeed reflect an approximately three-fold increase of the number of both the sequenced genomes and the ECF sequences from *Planctomycetes* and *Actinobacteria* compared to that collected by Staroń *et al*. Based on the re-classification of the ECF sequences that are not assigned into previous groups, we finally identified eight novel ECF groups consisting of 202 ECF sequences from *Planctomycetes* (Chapter 2) and 18 novel ECF groups consisting of 427 ECF sequences from *Actinobacteria* (Chapter 3). This leads to the increase of ECF group number from 68 to 94 in the ECF family.

Taken together, our study demonstrates that comparative genomics is a powerful tool to identify novel ECF groups from the rapidly increasing genome sequences publically available. While our work only covers two bacterial phyla, the future work should identify more novel ECF groups from the genome sequences of other bacterial phyla.

5.1.2 The ECF-dependent regulation in the distinct bacterial phylum, *Planctomycetes*

Planctomycetes is a distinct bacterial phylum. Bacteria belonging to this phylum show many unusual properties such as harboring complex endomembrane systems (Acehan et al, 2014; Lindsay et al, 1997; Lindsay et al, 2001; Sagulenko et al, 2014; Santarella-Mellwig et al, 2013) and being capable of taking up proteins by eukaryotic-like endocytosis(Lonhienne et al, 2010). In the past decades, although much knowledge has been known about their life styles and differentiation processes (Fuerst, 1995; Fuerst & Sagulenko, 2011), the regulatory cascades mediate these life styles and differentiation processes remain largely unknown. In this study, along with the classification of ECF σ factors, comparative genomics provided an initial picture of regulation governed by ECF σ factors in this distinct bacterial phylum.

As shown by the Fig. 7 and Table S5 in Chapter 2, the vast majority of 361 ECF σ factors from the eight planctomycetal genomes were assigned into 11 ECF groups. This includes three previously identified groups (ECF01, ECF42 and ECF43) (Staroń et al, 2009) and eight novel groups (ECF01-Gob, ECF45, ECF46, ECFSTK1, ECFSTK-2, ECFSTK-3 and ECFSTK-4) identified in our study.

All the eight planctomycetal genomes contain ECF σ factors belonging to the group ECF01. On average, each genome contains approximately 10 ECF σ factors belonging to the group ECF01 (Table S5, Chapter 2). The wide distribution and abundance of ECF01 σ factors in

the planctomyceteal genomes suggest that they might exert important regulatory roles in *Planctomycetes*. In the previous classification study, it was shown that ECF01 contained the well characterized ECF σ factor, σ^W from *B. subtilis* (Staroń et al, 2009). σ^W is involved in responding to multiple cell envelope stresses caused by alkaline shock (Wiegert et al, 2001), salt shock (Steil et al, 2003), and antibiotics (Cao et al, 2002). Given that ECF σ factors belonging to the same group tend to have similar physiological roles (Staroń et al, 2009; Wecke et al, 2012), it is tempting to envision that ECF01 σ factors in *Planctomycetes* might show a σ^W-like regulatory role. Nevertheless, further experimental studies are necessary for the elucidation of the detailed roles of ECF01 σ factors in *Planctomycetes*.

Only five planctomyceteal genomes (*Planctomyces maris* DSM 8797, *Gemmata obscuriglobus* UQM 2246, *Pirellula staleyi* DSM 6068, *Planctomyces brasiliensis* DSM 5305 and *Rhodopirellula baltica* SH 1) contain ECF42 σ factors. While the genome of *Rhodopirellula baltica* SH 1 contains two ECF42 σ factors, each of the remaining four genomes contain only one ECF42 σ factor (Table S5, Chapter 2). The absence of ECF42 σ factors in the other three analyzed planctomyceteal genomes (*Blastopirellula marina* DSM 3645, *Planctomyces limnophilus* DSM 3776 and *Isosphaera pallida* ATCC 43644) suggest that the regulatory cascade governed by ECF42 σ factors might not be universal in *Planctomycetes*. Recently, one ECF42 σ factor from *Pseudomonas putida* KT2440 was characterized to be implicated in antibiotic resistance and biofilm formation. Deletion of this ECF σ factor from *Pseudomonas putida* KT2440 significantly influences the transcription of 12 genes, including three (*ttgA*, *ttgB*, and *ttgC*) encode a major multidrug efflux pump (Tettmann et al, 2014). This information should provide a reference for the future studies on the physiological roles of ECF42 σ factors in *Planctomycetes*.

Among the eight planctomycetal genomes, only the genome of *Gemmata obscuriglobus* UQM 2246 was found to encode ECF01-Gob σ factors (Table S5, Chapter 2). This suggests that the function of ECF01-Gob σ factors might be only important for *Gemmata obscuriglobus* UQM 2246, but not for other planctomycetal species. The ECF σ factors belonging to ECF01-Gob contain a long C-terminal extension with up to 1000 amino acids. Approximately one third of these C-terminal extensions contain multiple numbers of WD40 domains, which might be implicated in the assembly of protein complexes. Most of these C-terminal extensions also contain one to three trans-membrane helices, suggesting ECF01-Gob σ factors are mainly located in the membrane (Chapter 2). However, based on their

sequence properties alone, the function of ECF01-Gob σ factors is hard to be deduced. Further experimental studies should help to unravel their physiological roles.

Seven planctomycetal genomes (from eight) contain ECF46 σ factors. On average, each of these seven genomes contains nine ECF46 σ factors (Table S5, Chapter 2). These suggest that ECF46 σ factors might exert important regulatory roles in *Planctomycetes*. ECF46 was suggested to be a FecI-like group (Chapter 2). In the previous study, five FecI-like groups (ECF05-10) were identified. The ECF σ factors belonging to these groups were suggested to be implicated in iron uptake or polysaccharide metabolism (Mascher, 2013; Staroń et al, 2009). It would be interesting to know whether ECF46 σ factors in *Planctomycetes* have a similar function with ECF σ factors belonging to other FecI-like groups.

The eight planctomycetal genomes together contain 80 ECF σ factors with their encoding genes that are genomically adjacent to genes encoding Ser/Thr kinases. *Rhodopirellula baltica* SH 1 is rich in this type of ECF σ factors, with 19 encoded by its genome. *Planctomyces brasiliensis* DSM 5305 and *Planctomyces limnophilus* DSM 3776 yet contain the smallest number of this type of ECF σ factors, with only three encoded by each of these two genomes (Table S5, Chapter 2). These 80 ECF σ factors were assigned into five ECF groups, ECF43, ECFSTK1, ECFSTK-2, ECFSTK-3 and ECFSTK-4 (Table S5, Chapter 2). It has been recently proposed these five ECF groups might share a common regulatory mechanism in which the activity of ECF σ factors is regulated by the Ser/Thr kinases (Mascher, 2013). Nevertheless, due to the absence of experimental examples of these groups, the detailed regulatory mechanisms and the physiological roles of ECF σ factors belonging to these groups remain hard to be deduced. Further experimental studies using *Planctomycetes* as the workhorse might help to elucidate their regulatory mechanisms and physiological roles.

5.1.3 The ECF-dependent regulation in *Actinobacteria*

Actinobacteria is a bacterial phylum with biotechnological and medical importance (Ventura et al, 2007). Although some ECF σ factors from bacteria belonging to this phylum have been well characterized (Paget et al, 2002; Rodrigue et al, 2006), a large number of remaining ECF σ factors are uninvestigated. In this thesis, using comparative genomics, we did a comprehensive study of ECF σ factors from 119 actinobacterial genomes and provided much knowledge related to ECF-dependent regulation in this phylum.

As shown in Table S5 and Table S6 in the Chapter 3, ECF σ factors from the 119 actinobacterial genomes were mainly assigned into 39 ECF groups. This includes 21 ECF groups identified in the previous studies (ECF01, 12, 14, 17-20, 24, 26, 27, 30, 34, 36, 38-42, 118, 122 and 123) (Staroń et al, 2009) and 18 ECF groups (ECF47-56, ECF125-132) identified in our study.

Approximately 90% actinobacterial genomes contain at least one ECF12 σ factor (Fig. 1 and Table S5, Chapter 3). The wide distribution of ECF12 σ factors in actinobacterial genomes suggests that they might exert important regulatory roles in *Actinobacteria.* The group ECF12 includes the well characterized ECF σ factor, σ^R from *S. coelicolor* (Staroń et al, 2009). Due to the similar sequence properties of other ECF12 σ factors with σ^R, ECF12 is also called a σ^R-like group. σ^R is involved in the response to thiol-oxidative stress responses in *S. coelicolor* (Kang et al, 1999; Paget et al, 2001; Paget et al, 2002; Paget et al, 1998). It directs the transcription of over 100 target genes in *S.coelicolor*, which encode proteins involved in thiol-redox homeostasis, Fe-S delivery, flavin-mediated redox reactions, protein quality control and regulatory functions (Kim et al, 2012). Further studies might help to show whether all the σ^R-like proteins in *Actinobacteria* are implicated in the thiol-oxidative stress responses.

ECF14 σ factors are also widely distributed in actinobacterial genomes. One or two ECF14 σ factors are found to present in each of approximately 80 (of the 119) actinobacterial genomes. (Fig. 1 and Table S5, Chapter 3). This suggests that the functions of ECF14 σ factors might be important for the actinobacterial speices. ECF14 is an actinobacterial-specific group (Staroń et al, 2009). One experimentally validated example of this group is σ^E from *Mycobacterium smegmatis*, which is implicated in the adaptive responses to multiple stresses such as oxidative regents, detergents and heat (Wu et al, 1997). It is likely that other ECF14 σ factors in *Actinobacteria* show a σ^E-like regulatory role. But still, this awaits future experimental validations.

Most (approximately 60%) actinobacterial genomes also contain multiple numbers of ECF41 σ factors. Several species are quite abundant in ECF41 σ factors. For example, *Streptomyces coelicolor* A3(2) has 13 ECF41 σ factors. *Streptomyces bingchenggensis* BCW-1 has 17 ECF41 σ factors. *Catenulispora acidiphila* DSM 44928 contains 12 ECF41 σ factors (Fig. 1 and Table S5, Chapter 3). The abundance of ECF41 σ factors in these actinobacterial species implies that they might exert important regulatory roles. It has been

shown that ECF41 σ factors contain a conserved C-terminal extension with about 100 amino acids (Staroń et al, 2009). Previous study of ECF41 members, $Ecf41_{Bli}$ (from *Bacillus licheniformis*) and $Ecf41_{Rsp}$ (from *Rhodobacter sphaeroides*) showed that their activities were regulated by their C-terminal extension domains. Nevertheless, although multiple phenotype-searching experiments were done, this study failed to obtain a physiological role associated with $Ecf41_{Bli}$ and $Ecf41_{Rsp}$ (Wecke et al, 2012). Likewise, the physiological roles of ECF41 in *Actinobacteria* remain hard to be deduced.

In contrast to ECF12, ECF 14 and ECF41, other ECF groups showed a relatively narrow distribution in the actinobacterial genomes (Fig. 1 and Table S5, Chapter 3). In particular, several ECF groups are only distributed in a limited number of actinobacterial genomes. For example, ECF24 σ factors are only present in *Amycolicicoccus subflavus* DQS3-9A1, *Mycobacterium vanbaalenii* PYR-1, *Rhodococcus* jostii RHA1 and *Rhodococcus opacus* B4. Only the genome of *Eggerthella* sp. YY7918 encodes an ECF30 σ factor (Table S5, Chapter 3). Therefore, the regulation governed by these groups of ECF σ factors might not be universal in *Actinobacteria*.

5.2 Defining the regulon of genes controlled by σ^E in *Streptomyces*

σ^E is a conserved ECF σ factor in the genus *Streptomyces* (Fig. S7, Chapter 4). Although its physiological role of cell envelope stress response has been well characterized in *S. coelicolor* (Hong et al, 2002; Paget et al, 1999a), the regulon of genes that it controls remained largely unknown. In this study, using a combination of ChIP-seq, DNA microarray and bioinformatic analysis, we obtained 91 σ^E targets from *S. coelicolor*. Almost half of these targets were predicted to have cell envelope-related functions. Many other targets were also predicted to be involved in cell regulation and cell metabolism. In addition, through bioinformatics prediction, we defined all the putative σ^E regulons from 19 *Streptomyces* species and established a σ^E-core regulon (Chapter 4). These results provided a detailed picture of the regulatory network governed by σ^E in *Streptomyces* genus.

5.2.1 How many genes are controlled by σ^E in *S. coelicolor*?

Despite the fact that 91 genes (Table 1, Chapter 4) have been assigned to the σ^E regulon in *S. coelicolor*, the real gene pool that is controlled by σ^E could be much bigger. Indeed, over 200 promoters were found to be reliable (P-value below 10^{-4}) σ^E targets from our ChIP-seq

experiments (Table S1, Chapter 4). Those targets, which were not included into our final σ^E regulon (91 genes), were either located too far away from the closest predicted protein encoding sequence (over 400 bp from the start codon) or located in the reverse transcriptional direction of the closed predicted protein encoding sequence (Table S1, Chapter 4). Nevertheless, although these genes were excluded, the possibility that they control the transcription of a distant downstream gene (resulting in a long 5' untranslated region (UTR)) or indeed control the expression of a regulatory anti-sense RNA could not be ruled out. Indeed, not only mRNAs with long 5' UTR have been found to be transcribed by an ECF σ factor (Eiamphungporn & Helmann, 2008), but also anti-sense mRNAs that show a repressing function (Eiamphungporn & Helmann, 2009).

In addition, the genome-wide bioinformatics prediction identified over 5000 σ^E binding sites (Table S3, Chapter 4) in *S. coelicolor*. This number is much higher than that we obtained from ChIP-seq experiments. Although it is unlikely that all of these binding sites represent true σ^E targets *in vivo*, this number provides a possibility that there are more σ^E targets in *S. coelicolor* than that we have already identified from our ChIP-seq experiment.

Indeed, in our ChIP-seq data analysis, we employed a P-value below 10^{-4} as a cut off to identify the reliable targets (Chapter 4). This might rule out some targets with higher P-value. It is likely that if we increased the P-value from the current cut-off value of 10^{-4} to 10^{-3} or even higher in our ChIP-seq data analysis, we might identify more targets from our ChIP-seq data. Taken together, although 91 genes were finally assigned to the σ^E regulon, it should be always kept in mind that there might be more σ^E targets in *S. coelicolor*.

5.2.2 Genes that are differentially transcribed in the *S. coelicolor ΔsigE* mutant

σ^E has been characterized to be a key regulator involved in the cell envelope stress response (Hong et al, 2002; Paget et al, 1999a). Compared to wild type *S. coelicolor*, *ΔsigE* mutant shows increased sensitivity to cell envelope disruptors (e.g., lysozyme). On the medium deficient in Mg^{2+}, *ΔsigE* mutant also displays development related defects such as poor sporulation (Paget et al, 1999a). Despite that the phenotypic changes in the *ΔsigE* mutant have been identified for over a decade, the molecular mechanisms underlying these changes remain unknown. Here, it was found that 46 genes were differentially transcribed greater than fourfold in the *ΔsigE* mutant compared to the wild type either before vancomycin treatment or after vancomycin treatment for 30 minutes (Table S1). Given their significant

transcriptional changes in the *ΔsigE* mutant, these genes should contribute to the phenotypic changes of *ΔsigE* mutant.

Of these 46 genes, nine (*sco1875, sco3044, sco3194, sco3396, mprF, sco4261, sco4263, sco4471* and s*co7657*) were identified to be potential direct σ^E targets in our ChIP-seq experiment (Table S1). These genes were clearly down-regulated in the *ΔsigE* mutant compared to the wild type both before vancomycin treatment and after vancomycin treatment, suggesting they are dependent on σ^E for their transcription. As discussed in Chapter 4, *sco1875* encodes a penicillin-binding protein, which might be involved in the synthesis of peptidoglycan. *sco3044* encodes a LytR_cpsA_psr family protein, which is possibly involved in cell wall teichoic acid deposition. *sco3194* encodes a L, D transpeptidase which may be implicated in the synthesis of peptidoglycan via catalyzing the formation of a L-Lys_3/D-Asx-L-Lys_3 (3-3) bridge between two penta-peptide chains of the peptidoglycan precursor. *mprF* encodes a lysylphosphatidylglycerol (LPG) synthase that might be implicated in the resistance towards cationic antimicrobial peptides and antibiotics. Deletion of *sco4471* in *S. coelicolor* resulted in a four times increased sensitivity to lysozyme and an almost two fold increased sensitivity towards the antibiotics targeting the cell envelope. *sco3396, sco4263* and s*co7657* encode the proteins, the functions of which are unknown.

The remaining 37 genes could not be identified as potential direct σ^E targets by our ChIP-seq experiment (Table S1, Chapter 4). It is likely that the transcription of these genes is indirectly influenced by σ^E (e.g., the transcription of these genes could be controlled by one of the σ^E direct targets with regulatory function (Table 1, Chapter 4)). These genes show two different transcriptional patterns. 26 genes show a transcriptional up-regulation in the *ΔsigE* mutant compared to in the wild type both before vancomycin treatment and after vancomycin treatment. Others, however, show a four-fold transcriptional down-regulation in the *ΔsigE* mutant compared to the wild type either before vancomycin treatment or after vancomycin treatment (Table S1). The vast majority of these genes are predicted to associate with cell metabolism (11 genes), cell regulation (three genes) and cell envelope related function (13 genes) (Table S1). Among them, two genes, *bldKC* and *ramS* have been characterized to be key genes in the differentiation process of *S. coelicolor* (Flärdh & Buttner, 2009; Kodani et al, 2004; Nodwell & Losick, 1998; Nodwell et al, 1996).

bldKC encodes a putative permease unit of the BldK oligopeptide ABC transporter. The *bldK* locus contains five genes, *bldKA*, *bldKB*, *bldKC*, *bldKD* and *bldKE*. Similar to *bldKC*, *bldKA* also encodes a putative permease unit, whereas *bldKD* and *bldKE* encode the putative ATPase subunits and BldKB is a putative ABC transport system lipoprotein. Together, these proteins constitute an ABC transporter (Nodwell et al, 1996; information retrieved from the StrepDB database http://strepdb.streptomyces.org.uk). BldK is involved in the import of an extracellular signal that controls aerial mycelium formation in *S. coelicolor* (Nodwell & Losick, 1998; Nodwell et al, 1996). Disruption of *bldK* locus causes a clear "Bld" phenotype in which formation of aerial mycelium is severely inhibited while growing on rich medium (Nodwell et al, 1996). As shown in Table S1, *bldKC* shows more than four-fold decreased transcription in the *ΔsigE* mutant compared to in the wild type before vancomycin treatment. Indeed, *bldKA*, *bldKB*, *bldKD* and *bldKE* also show a significant decreased transcription compared to the wild type before vancomycin treatment (more than two fold, data not shown). Given the importance of *bldK* locus in the formation of aerial mycelium, the decreased transcription of the *bldK* locus in the *ΔsigE* mutant might decrease the formation of aerial mycelium and further contribute to the phenotype of poor sporulation observed in the *ΔsigE* mutant (Paget et al, 1999a). It should be pointed out that, after vancomycin treatment for 30 minutes, the transcriptional differences of *bldK* locus in between the *ΔsigE* mutant and the wild type are not significant (Table S1 and data not shown). This suggests these genes might not contribute to the phenotypic changes of *ΔsigE* mutant compared to wild type under stress conditions.

ramS encodes the precursor of SapB in *S. coelicolor*, a lantibiotic-like peptide, which functions as an important biosurfactant to facilitate the growth of the aerial hyphae into the air (Flärdh & Buttner, 2009; Kodani et al, 2004). *ramS* is located in a convergently transcribed operon *ramCSAB*, which is activated by the response regulator RamR (Keijser et al, 2002; Nguyen et al, 2002; O'Connor et al, 2002; O'Connor & Nodwell, 2005). RamC contains a domain that is similar to a lantibiotic synthetase and is suggested to be involved in processing the RamS into mature SapB (Kodani et al, 2004). RamAB encodes an ABC transporter and is likely to be involved in the export of the SapB precursor (Kodani et al, 2004; Ma & Kendall, 1994; Willey et al, 2006). It has been shown that introducing extra copies of the *ram* genes from *S. coelicolor* into *S. lividans* results in accelerated mycelium formation (Ma & Kendall, 1994). Moreover, deletion of the *ram* genes in *S. coelicolor* leads to a "Bld" phenotype on rich medium (Capstick et al, 2007; Nguyen et al, 2002; O'Connor

et al, 2002). In stark contrast with *bldK*, *ramS* is strongly upregulated (more than 6 fold) in the *ΔsigE* mutant compared to the wild type both before and after vancomycin treatment (Table S1). Indeed, based on the analysis of our transcriptomic data, *ramC, ramA, ramB* and *ramR* are also upregulated in the *ΔsigE* mutant to a certain extent compared to in the wild type both before and after vancomycin treatment (data not shown). Given the importance of *ramCSAB* in the formation of aerial mycelium, it is likely that the up-regulation of these genes in the *ΔsigE* mutant partly counteract the morphological defects caused by the down-regulation of the developmental related genes such as *bldK*.

Taken together, the comparative transcriptomic analysis led to the identification of both transcriptionally up-regulated genes and transcriptionally down-regulated genes in the *ΔsigE* mutant. The vast majority of these genes might be regulated by σ^E through an indirect effect. The phenotypic change in the *ΔsigE* mutant should result from overlapping effects of all the differentially transcribed genes.

5.3 Outlook

This thesis focused on the classification and functional characterization of ECF σ factors from two bacterial phyla, *Planctomycetes* and *Actinobacteria*. Much knowledge about ECF-dependent regulation in these two phyla of bacteria was obtained. In line with the aim of this thesis, the future work could concern the following points.

5.3.1 Assigning the remaining unclassified ECF σ factors into novel ECF groups

In the first two parts of this study, comparative genomics was used to classify the ECF σ factors from *Planctomycetes* and *Actinobacteria* (Chapter 2 and Chapter 3). Eight novel ECF groups consisting of 202 protein members from *Planctomycetes* (Chapter 2) and 18 novel ECF groups consisting of 427 protein members from *Actinobacteria* (Chapter 3) were identified. This study together with previous classification study thus led to the classification of approximately 80% ECF σ factors from *Planctomycetes* (Chapter 2) and approximately 96% ECF σ factors from *Actinobacteria* (Chapter 3). For the remaining 20% unclassified ECF σ factors from *Planctomycetes* and 4% unclassified ECF σ factors from *Actinobacteria*, future studies can help to assign them into newly identified ECF groups. It is likely that the rapidly increasing genome sequences would give rise to the identification

of many ECF σ factors that resemble these unclassified ECF σ factors, further allowing the classification of these ECF σ factors into novel ECF groups.

5.3.2 Validation of the hypothesis obtained from comparative genomics of ECF σ factors

Our study identified 26 novel ECF groups in total (Chapter 2 and Chapter 3). Based on the conserved properties of these groups, many novel ECF-dependent signal transduction mechanisms were proposed. For example, the activity of ECF σ factors belonging to ECFSTK1-STK4 might be regulated by the Ser/Thr kinases that are gnomically adjacent to them (Chapter 2). The activity of ECF σ factors belonging to ECF48 might be regulated by their extended C-terminal domains (Chapter 3). The conserved properties of some ECF groups also enabled us to envision their possible physiological roles. For example, based on the predicted functions of their conserved domains, ECF53 σ factors were suggested to be implicated in regulating carbohydrate metabolism (Chapter 3). Experimental validation of these hypotheses in the future should enable us to obtain new knowledge about ECF σ factors and further understand ECF-dependent regulation in *Planctomycetes* and *Actinobacteria*.

5.3.3 Further elucidation of the regulatory network governed by σ^E in *Streptomyces*

In the third part of this study, both experimental work and bioinformatics analysis were used to identify the regulon controlled by a conserved ECF σ factor σ^E in *Streptomyces*. This study enabled us to obtain a large number of targets that controlled by σ^E. Nevertheless, much work is still necessary to elucidate the complete regulatory network governed by σ^E in *Streptomyces*.

Using a combination of ChIP-seq, DNA microarray and bioinformatics analysis, we assigned 91 genes into the σ^E regulon of *S. coelicolor*. Among them, 30 genes encode proteins without any known functional domains (not including transmembrane helices) (Table 1, Chapter 4). Therefore, the functional roles of these σ^E targets are hard to be deduced. These hamper us to elucidate the full regulatory roles of σ^E in *S. coelicolor*. Future studies can focus on the investigation of functions of these current unknown σ^E targets and further elucidate the regulatory role of σ^E. Genetic experiments such as

generating mutants of these targets in *S. coelicolor* might help to define their physiological roles.

Of the 91 genes under σ^E-control in *S. coelicolor*, 15 genes encode proteins involved in cell regulation (Table 1, Chapter 4). These genes might further regulate the transcription of a number of downstream genes. In order to fully understand the regulatory network governed by σ^E, the identification of downstream genes controlled by these 15 targets with regulatory function would be necessary.

Of the 19 *in vitro* validated σ^E targets (by S1 mapping or *in vitro* transcription) from *S. coelicolor*, 14 are partially dependent on σ^E for their transcription (Fig. 3, Chapter 4). This suggests that some of the other approximately 60 σ factors in *S. coelicolor* (Bentley et al, 2002) might also be involved in their transcription. In the future work, it would be interesting to know which σ factors cooperate with σ^E to transcribe the same genes and whether the regulation governed by σ^E is involved or affected by these other σ factors.

The genome-wide bioinformatics prediction and the following analysis enabled us to establish a σ^E core regulon consisting of 118 targets across 19 *Streptomyces* genomes (Fig. 5, Table 1, Table S6, Chapter 4). This core-regulon represents a conserved regulatory network governed by σ^E in *Streptomyces*. However, only 21 of these core-σ^E targets belong to the validated σ^E regulon (91 genes) in *S. coelicolor* (Table 1, Fig. 5A, Chapter 4). For the remaining core-σ^E targets that solely identified by genome-wide bioinformatics prediction, future experimental work can help to validate them.

Supplementary material for Chapter 5

Table S1 Genes that are differentially transcribed greater than fourfold between *ΔsigE* mutant and *S. coelicolor* wild type strain M600[1].

Locus	G[2]	Annotation[3]	Δ*sigE*-T0/M600-T0[4]	Δ*sigE*-T30/M600-T30[5]
sco1875		putative secreted penicillin binding protein	-0,50	-2,05
sco3044		conserved hypothetical protein	-1,27	-2,44
sco3194		putative lipoprotein	-2,17	-0,92
sco3396		putative membrane protein	-0,48	-2,55
sco3397		putative integral membrane lysyl-tRNA synthetase	-2,40	-5,44
sco4261		putative response regulator	-2,30	-0,93
sco4263		putative transcriptional regulatory protein	-0,33	-2,54
sco4471		putative secreted protein	-1,72	-4,53
sco7657		putative secreted protein	-4,58	-6,02
sco0179	1	putative zinc-containing dehydrogenase	3,90	3,55
sco0199	1	putative alcohol dehydrogenase	4,84	2,66
sco0213	1	putative nitrate/nitrite transporter protein	3,43	1,17
sco0216	1	nitrate reductase alpha chain NarG2	3,97	1,61
sco0217	1	nitrate reductase beta chain NarH2	4,44	1,71
sco0218	1	putative nitrate reductase delta chain NarJ2	5,04	3,04
sco0219	1	putative nitrate reductase delta chain NarI2	4,68	2,92
sco0382	1	UDP-glucose/GDP-mannose family dehydrogenase (putative secreted protein)	1,80	2,19
sco3945	1	cytochrome oxidase subunit I cydA	-2,56	0,44
sco4157	1	putative protease	-0,92	-3,05
sco6102	1	putative nitrite/sulphite reductase	-2,02	-0,36
sco0168	2	possible regulator protein	4,25	1,84
sco0204	2	putative luxR family two-component response regulator	2,40	1,40
sco4677	2	putative regulatory protein	1,81	3,50
sco0201	3	putative integral membrane protein SCJ12.13c	3,99	2,02
sco0607	3	hypothetical lipoportein SCF55.31	2,07	0,14
sco1557	3	putative lipoprotein	-3,03	-0,60
sco4472	3	putative secreted protein	-0,61	-2,28
sco6509	3	hydrophobic protein	2,74	0,64
sco7186	3	putative integral membrane protein	1,77	2,40
sco6682	3	ramS	3,76	3,82
sco1558	3	putative ABC transporter permease protein	-2,81	-0,11
sco1559	3	putative ABC transporter ATP-binding protein	-2,34	-0,17
sco1620	3	glycine betaine transport system	-2,29	-1,04

		permease protein opuABC		
sco4498	3	putative proton transport protein	-2,46	-0,29
sco5114	3	BldKC, putative ABC transport system integral membrane protein	-2,27	-0,71
sco6011	3	probable ABC-type transmembrane transport protein	-2,13	0,50
sco0169	4	conserved hypothetical protein SCJ1.18	3,05	0,95
sco0197	4	conserved hypothetical protein	2,69	0,87
sco0198	4	conserved hypothetical protein SCJ12.10c	2,50	0,85
sco0200	4	conserved hypothetical protein SCJ12.12c	3,16	1,67
sco0212	4	hypothetical protein SCJ12.24c	3,17	1,22
sco0220	4	hypothetical protein SCJ12.32	2,55	1,10
sco0268	4	hypothetical protein	1,09	3,20
sco0682	4	hypothetical protein SCF15.03c	2,13	2,45
sco0684	4	hypothetical protein SCF15.05c	1,58	2,24
sco5389	4	hypothetical protein 2SC6G5.33	1,81	2,16

1. 46 genes, which are differentially transcribed greater than fourfold between *ΔsigE* mutant and *S. coelicolor* wild type strain M600 either before vancomycin treatment or after vancomycin treatment for 30 minutes, were selected to show here. Nine potential σ^E direct targets were shaded in light brown color. A gene was defined as a potential σ^E direct target if it is a downstream gene of an identified ChIP peak, shown in Table S1 of Chapter 4. Among the transcriptional data of the remaining 37 genes, the positive values, which indicate a transcriptional up-regulation, were shaded in light blue. The transcriptomic data of *S. coelicolor* wild type used in this analysis is retrieved from Hesketh *et al.* (Hesketh *et al.* 2011). The transcriptomic data of *S. coelicolor ΔsigE* mutant used in this analysis is the same as described in Chapter 4.
2. 37 genes were assigned to four functional groups: (1) cell metabolism; (2) cell regulation; (3) envelope related protein; (4) hypothetical proteins.
3. The annotation of each gene was based on the information documented in the StrepDB database (http://strepdb.streptomyces.org.uk/).
4. "Δ*sigE*-T0/M600-T0"=the log-fold change ($\log_2$ scale) in transcription of *ΔsigE* mutant compared to the wild-type *S. coelicolor* M600 before vancomycin treatment.
5. "Δ*sigE*-T30/M600-T30"=the log-fold change ($\log_2$ scale) in transcription of *ΔsigE* mutant compared to the wild-type *S. coelicolor* M600 after vancomycin treatment for 30 minutes.

References for Chapter 1 and Chapter 5

Abbadi S, El Hadidy G, Gomaa N, Cooksey R (2009) Strain differentiation of *Mycobacterium tuberculosis* complex isolated from sputum of pulmonary tuberculosis patients. *Int J Infect Dis* **13:** 236-242

Acehan D, Santarella-Mellwig R, Devos DP (2014) A bacterial tubulovesicular network. *J Cell Sci* **127:** 277-280

Ades SE, Connolly LE, Alba BM, Gross CA (1999) The *Escherichia coli* σ^E-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-σ factor. *Genes Dev* **13:** 2449-2461

Akiyama Y, Kanehara K, Ito K (2004) RseP (YaeL), an *Escherichia coli* RIP protease, cleaves transmembrane sequences. *EMBO J* **23:** 4434-4442

Al-Bassam MM, Bibb MJ, Bush MJ, Chandra G, Buttner MJ (2014) Response regulator heterodimer formation controls a key stage in *Streptomyces* development. *PLoS Genet* **10:** e1004554

Alba BM, Leeds JA, Onufryk C, Lu CZ, Gross CA (2002) DegS and YaeL participate sequentially in the cleavage of RseA to activate the σ^E -dependent extracytoplasmic stress response *Genes Dev* **16:** 2156-2168

Angerer A, Enz S, Ochs M, Braun V (1995) Transcriptional regulation of ferric citrate transport in *Escherichia coli* K-12. Fecl belongs to a new subfamily of σ^{70}-type factors that respond to extracytoplasmic stimuli. *Mol Microbiol* **18:** 163-174

Angert ER (2005) Alternatives to binary fission in bacteria. *Nat Rev Microbiol* **3:** 214-224

Aravind L, Ponting CP (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

Arvind L, Ponting C (1997) The GAF domain: an evolutionary link between diverse phototransduction protiens. *Trends Biochem Sci* **22:** 458-459

Atsushi M, Soon-Kwang H, Hiroshi I, Sueharu H, Teruhiko B (1994) Phosphorylation of the AfsR protein involved in secondary metabolism in *Streptomyces* species by a eukary otic-type protein kinase. *Gene* **146:** 47-56

Bae J-B, Park J-H, Hahn M-Y, Kim M-S, Roe J-H (2004) Redox-dependent changes in RsrA, an anti-sigma factor in *Streptomyces coelicolor*: zinc release and disulfide bond formation. *J Mol Biol* **335:** 425-435

Becker J, Wittmann C (2012) Bio-based production of chemicals, materials and fuels–*Corynebacterium glutamicum* as versatile cell factory. *Curr Opin Biotechnol* **23:** 631-640

Běhal V (2000) Bioactive products from *Streptomyces*. *Adv Appl Microbiol* **47:** 113-156

Beilharz K, Nováková L, Fadda D, Branny P, Massidda O, Veening J-W (2012) Control of cell division in *Streptococcus pneumoniae* by the conserved Ser/Thr protein kinase StkP. *Proc Natl Acad Sci U S A* **109:** E905-E913

Bentley S, Chater K, Cerdeno-Tarraga A-M, Challis G, Thomson N, James K, Harris D, Quail M, Kieser H, Harper D (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature* **417:** 141-147

Bibb MJ, Domonkos Á, Chandra G, Buttner MJ (2012) Expression of the chaplin and rodlin hydrophobic sheath proteins in *Streptomyces venezuelae* is controlled by σ^{BldN} and a cognate anti-sigma factor, RsbN. *Mol Microbiol* **84:** 1033-1049

Bibb MJ, Molle V, Buttner MJ (2000) σ^{BldN}, an extracytoplasmic function RNA polymerase sigma factor required for aerial mycelium formation in *Streptomyces coelicolor* A3 (2). *J Bacteriol* **182:** 4606-4616

Bilwes AM, Alex LA, Crane BR, Simon MI (1999) Structure of CheA, a signal-transducing histidine kinase. *Cell* **96:** 131-141

Bondoso J, Albuquerque L, Nobre MF, Lobo-da-Cunha A, da Costa MS, Lage OM (2011) *Aquisphaera giovannonii* gen. nov., sp. nov., a planctomycete isolated from a freshwater aquarium. *Int J Syst Evol Microbiol* **61:** 2844-2850

Bontemps C, Toussaint M, Revol P-V, Hotel L, Jeanbille M, Uroz S, Turpault M-P, Blaudez D, Leblond P (2013) Taxonomic and functional diversity of *Streptomyces* in a forest soil. *FEMS Microbiol Lett* **342:** 157-167

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

Brooks BE, Buchanan SK (2008) Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. *Biochim Biophys Acta* **1778:** 1930-1945

Buckley DH, Huangyutitham V, Nelson TA, Rumberger A, Thies JE (2006) Diversity of *Planctomycetes* in soil in relation to soil history and environmental heterogeneity. *Appl Environ Microbiol* **72:** 4522-4531

Bush MJ, Bibb MJ, Chandra G, Findlay KC, Buttner MJ (2013) Genes required for aerial growth, cell division, and chromosome segregation are targets of WhiA before sporulation in *Streptomyces venezuelae*. *MBio* **4:** e00684-00613

Buttner M, Chater K, Bibb M (1990) Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3 (2). *J Bacteriol* **172:** 3367-3378

Campagne S, Damberger FF, Kaczmarczyk A, Francez-Charlot A, Allain FH-T, Vorholt JA (2012) Structural basis for sigma factor mimicry in the general stress response of Alphaproteobacteria. *Proc Natl Acad Sci U S A* **109:** E1405-E1414

Campbell EA, Greenwell R, Anthony JR, Wang S, Lim L, Das K, Sofia HJ, Donohue TJ, Darst SA (2007) A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria. *Mol Cell* **27:** 793-805

Campbell EA, Tupy JL, Gruber TM, Wang S, Sharp MM, Gross CA, Darst SA (2003) Crystal structure of *Escherichia coli* σ^E with the cytoplasmic domain of its anti-σ RseA. *Mol Cell* **11:** 1067-1078

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

Capra EJ, Laub MT (2012) The evolution of two-component signal transduction systems. *Annu Rev Microbiol* **66:** 325

Capstick DS, Willey JM, Buttner MJ, Elliot MA (2007) SapB and the chaplins: connections between morphogenetic proteins in *Streptomyces coelicolor*. *Mol Microbiol* **64:** 602-613

Casino P, Rubio V, Marina A (2010) The mechanism of signal transduction by two-component systems. *Curr Opin Struct Biol* **20:** 763-771

Cavicchioli R, Chiang RC, Kalman LV, Gunsalus RP (1996) Role of the periplasmic domain of the *Escherichia coli* NarX sensor-transmitter protein in nitrate-dependent signal transduction and gene regulation. *Mol Microbiol* **21:** 901-911

Cezairliyan BO, Sauer RT (2007) Inhibition of regulated proteolysis by RseB. *Proc Natl Acad Sci U S A* **104:** 3771-3776

Chaba R, Grigorova IL, Flynn JM, Baker TA, Gross CA (2007) Design principles of the proteolytic cascade governing the σ^E-mediated envelope stress response in *Escherichia coli*: keys to graded, buffered, and rapid signal transduction. *Genes Dev* **21:** 124-136

Champness WC (1988) New loci required for *Streptomyces coelicolor* morphological and physiological differentiation. *J Bacteriol* **170:** 1168-1174

Chandra G, Chater KF (2014) Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol Rev* **38:** 345-379

Chater K (1972) A morphological and genetic mapping study of white colony mutants of *Streptomyces coelicolor*. *J Gen Microbiol* **72:** 9-28

Cook GM, Berney M, Gebhard S, Heinemann M, Cox RA, Danilchanka O, Niederweis M (2009) Physiology of *Mycobacteria*. *Adv Microb Physiol* **55**: 81-319

Cowley S, Ko M, Pick N, Chow R, Downing KJ, Gordhan BG, Betts JC, Mizrahi V, Smith DA, Stokes RW (2004) The *Mycobacterium tuberculosis* protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth *in vivo*. *Mol Microbiol* **52:** 1691-1702

Cruz-Morales P, Vijgenboom E, Iruegas-Bocardo F, Girard G, Yáñez-Guerra LA, Ramos-Aboites HE, Pernodet J-L, Anné J, Van Wezel GP, Barona-Gómez F (2013) The genome sequence of *Streptomyces lividans* 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island. *Genome Biol Evol* **5:** 1165-1175

De Las Penas A, Connolly L, Gross CA (1997) The σ^E-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of σ^E. *Mol Microbiol* **24:** 373-385

de Lima Procópio RE, da Silva IR, Martins MK, de Azevedo JL, de Araújo JM (2012) Antibiotics produced by *Streptomyces*. *Braz J Infect Dis* **16:** 466-471

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

Dworkin J (2015) Ser/Thr phosphorylation as a regulatory mechanism in bacteria. *Curr Opin Microbiol* **24:** 47-52

Ehrlich J, Bartz QR, Smith RM, Joslyn DA, Burkholder PR (1947) Chloromycetin, a new antibiotic from a soil actinomycete. *Science* **106:** 417-417

Ehrlich J, Gottlieb D, Burkholder PR, Anderson LE, Pridham T (1948) *Streptomyces venezuelae*, n. sp., the source of chloromycetin. *J Bacteriol* **56:** 467

Eiamphungporn W, Helmann JD (2008) The *Bacillus subtilis* σ^M regulon and its contribution to cell envelope stress responses. *Mol Microbiol* **67:** 830-848

Eiamphungporn W, Helmann JD (2009) Extracytoplasmic function σ factors regulate expression of the *Bacillus subtilis yabE* gene via a cis-acting antisense RNA. *J Bacteriol* **191:** 1101-1105

Elliot MA, Buttner MJ, Nodwell JR (2008) Multicellular development in *Streptomyces*. In: *Myxobacteria*: Multicellularity and Differentiation (Whitworth, D. E., ed). *American Society for Microbiology Press* 419-438

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

Enz S, Mahren S, Stroeher UH, Braun V (2000) Surface signaling in ferric citrate transport gene induction: interaction of the FecA, FecR, and FecI regulatory proteins. *J Bacteriol* **182:** 637-646

Felsenstein J (1989) phylip– Phylogeny Inference Package (Version 3.2). Cladistics **5**: 164–166

Ferguson AD, Chakraborty R, Smith BS, Esser L, van der Helm D, Deisenhofer J (2002) Structural basis of gating by the outer membrane transporter FecA. *Science* **295:** 1715-1719

Flärdh K (2003) Growth polarity and cell division in *Streptomyces*. *Curr Opin Microbiol* **6:** 564-571

Flärdh K, Buttner MJ (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* **7:** 36-49

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

Francez-Charlot A, Kaczmarczyk A, Fischer H-M, Vorholt JA (2015) The general stress response in Alphaproteobacteria. *Trends Microbiol* **23:** 164-171

Franzmann P, Skerman V (1984) *Gemmata obscuriglobus*, a new genus and species of the budding bacteria. *Antonie van Leeuwenhoek* **50:** 261-268

Fuerst JA (1995) The planctomycetes: emerging models for microbial ecology, evolution and cell biology. *Microbiology* **141:** 1493-1506

Fuerst JA, Sagulenko E (2011) Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. *Nat Rev Microbiol* **9:** 403-413

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

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

Gómez-Santos N, Pérez J, Sánchez-Sutil MC, Moraleda-Muñoz A, Muñoz-Dorado J (2011) CorE from *Myxococcus xanthus* is a copper-dependent RNA polymerase sigma factor. *PLoS Genet* **7:** e1002106

Gourion B, Rossignol M, Vorholt JA (2006) A proteomic study of *Methylobacterium extorquens* reveals a response regulator essential for epiphytic growth. *Proc Natl Acad Sci U S A* **103:** 13186-13191

Greenwell R, Nam T-W, Donohue TJ (2011) Features of *Rhodobacter sphaeroides* ChrR required for stimuli to promote the dissociation of σ^E/ChrR complexes. *J Mol Biol* **407:** 477-491

Hadfield TL, McEvoy P, Polotsky Y, Tzinserling VA, Yakovlev AA (2000) The pathology of diphtheria. *J Infect Dis* **181:** S116-S120

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser (Oxf) 41: 95–98.

Härle C, Kim I, Angerer A, Braun V (1995) Signal transfer through three compartments: transcription initiation of the *Escherichia coli* ferric citrate transport system from the cell surface. *EMBO J* **14:** 1430

Hastie JL, Williams KB, Ellermeier CD (2013) The activity of σ^V, an extracytoplasmic function σ factor of *Bacillus subtilis*, is controlled by regulated proteolysis of the anti-σ factor RsiV. *J Bacteriol* **195:** 3135-3144

Hastie JL, Williams KB, Sepúlveda C, Houtman JC, Forest KT, Ellermeier CD (2014) Evidence of a bacterial receptor for lysozyme: binding of lysozyme to the anti-σ factor RsiV controls activation of the ECF σ factor σ^V. *PLoS Genet* **10:** e1004643

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

Heinrich J, Wiegert T (2009) Regulated intramembrane proteolysis in the control of extracytoplasmic function sigma factors. *Res Microbiol* **160:** 696-703

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

Herrou J, Rotskoff G, Luo Y, Roux B, Crosson S (2012) Structural basis of a protein partner switch that regulates the general stress response of α-proteobacteria. *Proc Natl Acad Sci U S A* **109:** E1415-E1423

Hesketh A, Hill C, Mokhtar J, Novotna G, Tran N, Bibb M, Hong H-J (2011) Genome-wide dynamics of a bacterial response to antibiotics that target the cell envelope. BMC genomics **12:** 226

Hong HJ, Hutchings MI, Neu JM, Wright GD, Paget MS, Buttner MJ (2004) Characterization of an inducible vancomycin resistance system in *Streptomyces coelicolor* reveals a novel gene (*vanK*) required for drug resistance. *Mol Microbiol* **52:** 1107-1121

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

Hopwood D, Wildermuth H, PALMER HM (1970) Mutants of *Streptomyces coelicolor* defective in sporulation. *J Gen Microbiol* **61:** 397-408

Hughes KT, Mathee K (1998) The anti-sigma factors. *Annu Rev Microbiol* **52:** 231-286

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

Hutchings MI, Hong HJ, Buttner MJ (2006b) The vancomycin resistance VanRS two-component signal transduction system of *Streptomyces coelicolor*. *Mol Microbiol* **59:** 923-935

Hutchings MI, Hoskisson PA, Chandra G, Buttner MJ (2004) Sensing and responding to diverse extracellular signals? Analysis of the sensor kinases and response regulators of *Streptomyces coelicolor* A3 (2). *Microbiology* **150:** 2795-2806

Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Ōmura S (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21:** 526-531

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

Kanehara K, Ito K, Akiyama Y (2002) YaeL (EcfE) activates the σ^E pathway of stress response through a site-2 cleavage of anti-σ^E, RseA. *Genes Dev* **16:** 2147-2155

Kang JG, Paget MS, Seok YJ, Hahn MY, Bae JB, Hahn JS, Kleanthous C, Buttner MJ, Roe JH (1999) RsrA, an anti-sigma factor regulated by redox change. *EMBO J* **18:** 4292-4298

Keijser BJ, van Wezel GP, Canters GW, Vijgenboom E (2002) Developmental regulation of the *Streptomyces lividans ram* genes: involvement of RamR in regulation of the *ramCSAB* operon. *J Bacteriol* **184:** 4420-4429

Kieser T, Bibb M, Buttner M, Chater K, Hopwood D. (2000) Practical *Streptomyces* Genetics (The John Innes Foundation, Norwich, United Kingdom).

Kim I, Stiefel A, Angerer A, Braun V (1997) Transcription induction of the ferric citrate transport genes via the N-terminus of the FecA outer membrane protein, the Ton system and the electrochemical potential of the cytoplasmic membrane. *Mol Microbiol* **23:** 333-344

Kim MS, Dufour YS, Yoo JS, Cho YB, Park JH, Nam GB, Kim HM, Lee KL, Donohue TJ, Roe JH (2012) Conservation of thiol-oxidative stress responses regulated by SigR orthologues in actinomycetes. *Mol Microbiol* **85:** 326-344

Kodani S, Hudson ME, Durrant MC, Buttner MJ, Nodwell JR, Willey JM (2004) The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. *Proc Natl Acad Sci U S A* **101:** 11448-11453

Kolb A, Busby S, Buc I, Garges S, Adhya S (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem* **62:** 749-797

Koster M, van Klompenburg W, Bitter W, Leong J, Weisbeek P (1994) Role for the outer membrane ferric siderophore receptor PupB in signal transduction across the bacterial cell envelope. *EMBO J* **13:** 2805

Kulichevskaya IS, Ivanova AA, Detkova EN, Rijpstra WIC, Damsté JSS, Dedysh SN (2015) *Planctomicrobium piriforme* gen. nov., sp. nov., a stalked planctomycete from a littoral wetland of a boreal lake. *Int J Syst Evol Microbiol* **65:** 1659-1665

Kulichevskaya IS, Ivanova AO, Belova SE, Baulina OI, Bodelier PL, Rijpstra WIC, Damsté JSS, Zavarzin GA, Dedysh SN (2007) *Schlesneria paludicola* gen. nov., sp. nov., the first acidophilic member of the order Planctomycetales, from Sphagnum-dominated boreal wetlands. *Int J Syst Evol Microbiol* **57:** 2680-2687

Lambert D, Loria R (1989) *Streptomyces scabies* sp. nov., nom. rev. *Int J Syst Evol Microbiol* **39:** 387-392

Laub MT, Goulian M (2007) Specificity in two-component signal transduction pathways. *Annu Rev Genet* **41:** 121-145

LeDeaux JR, Yu N, Grossman AD (1995) Different roles for KinA, KinB, and KinC in the initiation of sporulation in *Bacillus subtilis*. *J Bacteriol* **177:** 861-863

Lee H-J, Whang K-S (2014) *Streptomyces graminilatus* sp. nov., isolated from bamboo litter. *Int J Syst Evol Microbiol* **64:** 528-532

Lee K-C, Webb RI, Fuerst JA (2009) The cell cycle of the planctomycete *Gemmata obscuriglobus* with respect to cell compartmentalization. *BMC Cell Biol* **10:** 4

Leonardo MR, Forst S (1996) Re-examination of the role of the periplasmic domain of EnvZ in sensing of osmolarity signals in *Escherichia coli*. *Mol Microbiol* **22:** 405-413

Lindsay MR, Webb RI, Fuerst JA (1997) Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus *Pirellula*. *Microbiology* **143:** 739-748

Lindsay MR, Webb RI, Strous M, Jetten MS, Butler MK, Forde RJ, Fuerst JA (2001) Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. *Arch Microbiol* **175:** 413-429

Liu G, Chater KF, Chandra G, Niu G, Tan H (2013) Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol Mol Biol Rev* **77:** 112-143

Lonetto MA, Brown KL, Rudd KE, Buttner MJ (1994) Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A* **91:** 7573-7577

Lonhienne TG, Sagulenko E, Webb RI, Lee K-C, Franke J, Devos DP, Nouwens A, Carroll BJ, Fuerst JA (2010) Endocytosis-like protein uptake in the bacterium Gemmata obscuriglobus. *Proc Natl Acad Sci U S A* **107:** 12883-12888

Ma H, Kendall K (1994) Cloning and analysis of a gene cluster from *Streptomyces coelicolor* that causes accelerated aerial mycelium formation in *Streptomyces lividans*. *J Bacteriol* **176:** 3800-3811

Mahren S, Braun V (2003) The FecI extracytoplasmic-function sigma factor of *Escherichia coli* interacts with the β′ subunit of RNA polymerase. *J Bacteriol* **185:** 1796-1802

Mahren S, Enz S, Braun V (2002) Functional interaction of region 4 of the extracytoplasmic function sigma factor FecI with the cytoplasmic portion of the FecR transmembrane protein of the *Escherichia coli* ferric citrate transport system. *J Bacteriol* **184:** 3704-3711

Mansouri K, Piepersberg W (1991) Genetics of streptomycin production in *Streptomyces griseus*: nucleotide sequence of five genes, *strFGHIK*, including a phosphatase gene. *Mol Gen Genet* **228:** 459-469

Mascher T (2013) Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. *Curr Opin Microbiol* **16:** 148-155

Mascher T, Helmann JD, Unden G (2006) Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev* **70:** 910-938

Mascher T, Zimmer SL, Smith T-A, Helmann JD (2004) Antibiotic-inducible promoter regulated by the cell envelope stress-sensing two-component system LiaRS of *Bacillus subtilis*. *Antimicrob Agents Chemother* **48:** 2888-2896

Medema MH, Trefzer A, Kovalchuk A, van den Berg M, Müller U, Heijne W, Wu L, Alam MT, Ronning CM, Nierman WC (2010) The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol Evol* **2:** 212-224

Merrick M (1976) A morphological and genetic mapping study of bald colony mutants of *Streptomyces coelicolor*. *J Gen Microbiol* **96:** 299-315

Miller VL, Taylor RK, Mekalanos JJ (1987) Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. *Cell* **48:** 271-279

Missiakas D, Mayer MP, Lemaire M, Georgopoulos C, Raina S (1997) Modulation of the *Escherichia coli* σ^E (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. *Mol Microbiol* **24:** 355-371

Möglich A, Ayers RA, Moffat K (2009) Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure* **17:** 1282-1294

Nádvorník R, Vomastek T, Janeček J, Techniková Z, Branny P (1999) Pkg2, a novel transmembrane protein Ser/Thr kinase of *Streptomyces granaticolor*. *J Bacteriol* **181:** 15-23

Nariya H, Inouye S (2005) Identification of a protein Ser/Thr kinase cascade that regulates essential transcriptional activators in *Myxococcus xanthus* development. *Mol Microbiol* **58:** 367-379

Newman JD, Falkowski MJ, Schilke BA, Anthony LC, Donohue TJ (1999) The *Rhodobacter sphaeroides* ECF sigma factor, σ^E, and the target promoters *cycA* P3 and *rpoE* P1. *J Mol Biol* **294:** 307-320

Nguyen KT, Willey JM, Nguyen LD, Nguyen LT, Viollier PH, Thompson CJ (2002) A central regulator of morphological differentiation in the multicellular bacterium *Streptomyces coelicolor*. *Mol Microbiol* **46:** 1223-1238

Nodwell JR, Losick R (1998) Purification of an extracellular signaling molecule involved in production of aerial mycelium by *Streptomyces coelicolor*. *J Bacteriol* **180:** 1334-1337

Nodwell JR, McGovern K, Losick R (1996) An oligopeptide permease responsible for the import of an extracellular signal governing aerial mycelium formation in *Streptomyces coelicolor*. *Mol Microbiol* **22:** 881-893

O'Connor TJ, Kanellis P, Nodwell JR (2002) The *ramC* gene is required for morphogenesis in *Streptomyces coelicolor* and expressed in a cell type-specific manner under the direct control of RamR. *Mol Microbiol* **45:** 45-57

O'Connor TJ, Nodwell JR (2005) Pivotal roles for the receiver domain in the mechanism of action of the response regulator RamR of *Streptomyces coelicolor*. *J Mol Biol* **351:** 1030-1047

Ochs M, Veitinger S, Kim I, Weiz D, Angerer A, Braun V (1995) Regulation of citrate-dependent iron transport of *Escherichia coli*: FecR is required for transcription activation by FeeI. *Mol Microbiol* **15:** 119-132

Ohki R, Tateno K, Masuyama W, Moriya S, Kobayashi K, Ogasawara N (2003) The BceRS two-component regulatory system induces expression of the bacitracin transporter, BceAB, in *Bacillus subtilis*. *Mol Microbiol* **49:** 1135-1144

Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* **190:** 4050-4060

Ōmura S (1992) The expanded horizon for microbial metabolites—a review. *Gene* **115:** 141-149

Paget M, Helmann JD (2003) The sigma70 family of sigma factors. *Genome Biol* **4:** 203

Paget MS, Bae JB, Hahn MY, Li W, Kleanthous C, Roe JH, Buttner MJ (2001) Mutational analysis of RsrA, a zinc-binding anti-sigma factor with a thiol–disulphide redox switch. *Mol Microbiol* **39:** 1036-1047

Paget MS, Chamberlin L, Atrih A, Foster SJ, Buttner MJ (1999a) Evidence that the extracytoplasmic function sigma factor σ^E is required for normal cell wall structure in *Streptomyces coelicolor* A3 (2). *J Bacteriol* **181:** 204-211

Paget MS, Hong H-J, Bibb MJ, Buttner MJ (2002) The ECF sigma factors of *Streptomyces coelicolor* A3 (2). SGM symposium 61: Signals, switches, regulons and cascades: control of bacterial gene expression. (D. A. Hodgson, C. M. Thomas, ed). Cambridge University Press. 105-126.

Paget MS, Kang JG, Roe JH, Buttner MJ (1998) σ^R, an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress *in Streptomyces coelicolor* A3 (2). *EMBO J* **17:** 5776-5782

Paget MS, Leibovitz E, Buttner MJ (1999b) A putative two-component signal transduction system regulates σ^E, a sigma factor required for normal cell wall integrity in *Streptomyces coelicolor* A3 (2). *Mol Microbiol* **33:** 97-107

Papavinasasundaram K, Chan B, Chung J-H, Colston MJ, Davis EO, Av-Gay Y (2005) Deletion of the *Mycobacterium tuberculosis pknH* gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. *J Bacteriol* **187:** 5751-5760

Parkinson JS (1976) *cheA*, *cheB*, and *cheC* genes of *Escherichia coli* and their role in chemotaxis. *J Bacteriol* **126:** 758-770

Parkinson JS (2010) Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. *Annu Rev Microbiol* **64:** 101-122

Pearson A, Budin M, Brocks JJ (2003) Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci U S A* **100:** 15352-15357

Pereira SF, Goss L, Dworkin J (2011) Eukaryote-like serine/threonine kinases and phosphatases in bacteria. *Microbiol Mol Biol Rev* **75:** 192-212

Perraud A-L, Weiss V, Gross R (1999) Signalling pathways in two-component phosphorelay systems. *Trends Microbiol* **7:** 115-120

Picard C, Fioramonti J, Francois A, Robinson T, Neant F, Matuchansky C (2005) Review article: bifidobacteria as probiotic agents–physiological effects and clinical benefits. *Aliment Pharmacol Ther* **22:** 495-512

Raivio TL, Silhavy TJ (1999) The σ^E and Cpx regulatory pathways: overlapping but distinct envelope stress responses. *Curr Opin Microbiol* **2:** 159-165

Ramos JP, Campos CED, de Souza Caldas PC, Ferreira NV, da Silva MVB, Redner P, Campelo CL, Vale SF, Barroso EC, de Melo Medeiros RF (2013) *Mycobacterium fragae* sp. nov., a non-chromogenic species isolated from human respiratory specimens. *Int J Syst Evol Microbiol* **63:** 2583-2587

Rodrigue S, Provvedi R, Jacques P-E, Gaudreau L, Manganelli R (2006) The σ factors of *Mycobacterium tuberculosis*. *FEMS Microbiol Rev* **30:** 926-941

Ruggiero A, De Simone P, Smaldone G, Squeglia F, Berisio R (2012) Bacterial cell division regulation by Ser/Thr kinases: a structural perspective. *Curr Protein Pept Sci* **13:** 756

Ruiz N, Silhavy TJ (2005) Sensing external stress: watchdogs of the *Escherichia coli* cell envelope. *Curr Opin Microbiol* **8:** 122-126

Sagulenko E, Morgan GP, Webb RI, Yee B, Lee K-C, Fuerst JA (2014) Structural studies of planctomycete *Gemmata obscuriglobus* support cell compartmentalisation in a bacterium. *PloS one* **9:** e91344

Santarella-Mellwig R, Pruggnaller S, Roos N, Mattaj IW, Devos DP (2013) Three-dimensional reconstruction of bacteria with a complex endomembrane system. PLoS Biol **11:**e1001565

Sasaki S, Takeshita F, Okuda K, Ishii N (2001) *Mycobacterium leprae* and leprosy: a compendium. *Microbiol Immunol* **45**: 729-736

Schatz A, Bugle E, Waksman SA (1944) Streptomycin, a substance exhibiting antibiotic activity against Gram-Positive and Gram-Negative Bacteria. *Exp Biol Med (Maywood)* **55:** 66-69

Shibazaki A, Omoto Y, Kudo T, Yaguchi T, Saito A, Ando A, Mikami Y, Gonoi T (2011) *Streptomyces coacervatus* sp. nov., isolated from the intestinal tract of *Armadillidium vulgare*. *Int J Syst Evol Microbiol* **61:** 1073-1077

Shinnick T, Good R (1994) Mycobacterial taxonomy. *Eur J Clin Microbiol Infect Dis* **13:** 884-901

Shu Q, Jiao N (2008) Different *Planctomycetes* diversity patterns in latitudinal surface seawater of the open sea and in sediment. *J Microbiol* **46:** 154-159

St-Onge R, Goyer C, Coffin R, Filion M (2008) Genetic diversity of *Streptomyces* spp. causing common scab of potato in eastern Canada. *Syst Appl Microbiol* **31:** 474-484

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

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

Steil L, Hoffmann T, Budde I, Völker U, Bremer E (2003) Genome-wide transcriptional profiling analysis of adaptation of *Bacillus subtilis* to high salinity. *J Bacteriol* **185:** 6358-6370

Stephenson K, Hoch JA (2001) PAS-A domain of phosphorelay sensor kinase A: a catalytic ATP-binding domain involved in the initiation of development in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **98:** 15251-15256

Stewart V (2003) Nitrate-and nitrite-responsive sensors NarX and NarQ of proteobacteria. *Biochem Soc Trans* **31:** 1-10

Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. *Annu Rev Biochem* **69:** 183-215

Tekniepe BL, Schmidt JM, Starr P (1981) Life cycle of a budding and appendaged bacterium belonging to morphotype IV of the *Blastocaulis-Planctomyces* group. *Curr Microbiol* **5:** 1-6

Tettmann B, Dötsch A, Armant O, Fjell CD, Overhage J (2014) Knockout of extracytoplasmic function sigma factor ECF-10 affects stress resistance and biofilm formation in *Pseudomonas putida* KT2440. *Appl Environ Microbiol* **80:** 4911-4919

Turroni F, Duranti S, Bottacini F, Guglielmetti S, Van Sinderen D, Ventura M (2014) *Bifidobacterium bifidum* as an example of a specialized human gut commensal. *Front Microbiol.* **5:**437

Turroni F, Van Sinderen D, Ventura M (2011) Genomics and ecological overview of the genus *Bifidobacterium*. *Int J Food Microbiol* **149:** 37-44

Tutar Y (2008) Syn, anti, and finally both conformations of cyclic AMP are involved in the CRP-dependent transcription initiation mechanism in *E. coli* lac operon. *Cell Biochem Funct* **26:** 399-405

Ulrich LE, Koonin EV, Zhulin IB (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol* **13:** 52-56

Van Hove B, Staudenmaier H, Braun V (1990) Novel two-component transmembrane transcription control: regulation of iron dicitrate transport in *Escherichia coli* K-12. *J Bacteriol* **172:** 6749-6758

Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, van Sinderen D (2007) Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71:** 495-548

Veyisoglu A, Sahin N (2014) *Streptomyces hoynatensis* sp. nov., isolated from deep marine sediment. *Int J Syst Evol Microbiol* **64:** 819-826

Walsh NP, Alba BM, Bose B, Gross CA, Sauer RT (2003) OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* **113:** 61-71

Wanner LA (2006) A survey of genetic variation in *Streptomyces* isolates causing potato common scab in the United States. *Phytopathology* **96:** 1363-1371

Wecke T, Halang P, Staroń A, Dufour YS, Donohue TJ, Mascher T (2012) Extracytoplasmic function σ factors of the widely distributed group ECF41 contain a fused regulatory domain. *MicrobiologyOpen* **1:** 194-213

Welz D, Braun V (1998) Ferric citrate transport of *Escherichia coli*: functional regions of the FecR transmembrane regulatory protein. *J Bacteriol* **180:** 2387-2394

Wiegert T, Homuth G, Versteeg S, Schumann W (2001) Alkaline shock induces the *Bacillus subtilis* σ^W regulon. *Mol Microbiol* **41:** 59-71

Wiley DJ, Nordfeldth R, Rosenzweig J, DaFonseca CJ, Gustin R, Wolf-Watz H, Schesser K (2006) The Ser/Thr kinase activity of the Yersinia protein kinase A (YpkA) is necessary for full virulence in the mouse, mollifying phagocytes, and disrupting the eukaryotic cytoskeleton. *Microb Pathog* **40:** 234-243

Wilken C, Kitzing K, Kurzbauer R, Ehrmann M, Clausen T (2004) Crystal structure of the DegS stress sensor: How a PDZ domain recognizes misfolded protein and activates a protease. *Cell* **117:** 483-494

Willey JM, Willems A, Kodani S, Nodwell JR (2006) Morphogenetic surfactants and their role in the formation of aerial hyphae in *Streptomyces coelicolor*. *Mol Microbiol* **59:** 731-742

Wilson C, Zhan H, Swint-Kruse L, Matthews K (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci* **64:** 3-16

Woo HM, Park J-B (2014) Recent progress in development of synthetic biology platforms and metabolic engineering of *Corynebacterium glutamicum*. *J Biotechnol* **180:** 43-51

Wu Q-L, Kong D, Lam K, Husson RN (1997) A mycobacterial extracytoplasmic function sigma factor involved in survival following stress. *J Bacteriol* **179:** 2922-2929

Yue WW, Grizot S, Buchanan SK (2003) Structural evidence for iron-free citrate and ferric citrate binding to the TonB-dependent outer membrane transporter FecA. *J Mol Biol* **332:** 353-368

Zhang W, Shi L (2005) Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. *Microbiology* **151:** 2159-2173

Zhu H, Jiang S, Yao Q, Wang Y, Chen M, Chen Y, Guo J (2011) *Streptomyces fenghuangensis* sp. nov., isolated from seawater. *Int J Syst Evol Microbiol* **61:** 2811-2815

Zhulin IB, Taylor BL, Dixon R (1997) PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends Biochem Sci* **22:** 331-333

Acknowledgements

Time flies and it soon comes to the end of my thesis. At this point, vivid memories of these four years come to my mind and could never be stopped. There have been so much fun and unforgettable time rooting in my heart and every cell of my body since I first arrived at Munich airport in September, 2011. I can never forget the time when I first met people in Thorsten's lab. I can never forget the time when I got every small achievement from my project. I can also never forget the time when I flied between Germany and England in order to learn *Streptomyces* and do *Streptomyces* project. To be honest, it is really sad to recall all of these since many of you who create this great time surrounding me are currently living far away from Munich and it is not easy for us to get together with each other so often. But, this would never affect my great gratitude to you.

The first person I would like to thank a lot is Thorsten Mascher. Without your support, I should not have a chance to come to Munich, such a beautiful city. Alternatively, I probably went to some corners in the world where I might never be as happy as I am in Munich. It is you who is always so nice to give me suggestions on my project. It is you who gives me all kinds of supports for my living in Munich. It is you who are so patient to give me so many encouragements when I met problems. I would bring all of these unforgettable memories to every place I would stay in the future.

The second person I would like to thank a lot is Mark Buttner. It was October, 2012 when I first met you on the "29th signal transduction conference". We got to find so many interesting topics to talk about. Then after a long period applying for a UK visa, I finally arrived at Norwich and started to learn *Streptomyces.* I have to say, it is such an unforgettable time. You are so nice in all the aspects and I would never forget all the help and advices you gave me during the time while I stayed there.

The third person I want to thank a lot is Daniela Pinto. I would say you are not only a science researcher, but also an artist. The work you contributed to our projects is not only science itself, but also beautiful artistic design from your amazing heart. You are also so

nice, patient and careful in supervising me in solving so many scientific questions. I would never forget the knowledge you teach me about how to do academic writing.

The fourth person I want to thank a lot is Matt Bush. Without your supervision, I might not be able to finish my part in SigE project so smoothly. I would never forget the "strict" requirements you put on me while I did the SigE project, which really enables me to know much important knowledge about science.

I would like to also give my big thanks to Susanne Gebhard, Tina Wecke, Sebastian Dintner, Maureen Bibb and Mahmoud Al-Bassam for your kind supervision while I was fresh for many other experiments. I also greatly acknowledge the help of Georg Fritze and Govind Chandra for your kind helps in much data analysis. Big thanks also go to three girls (Chong Fang, Carolin Höfler, Dayane Araujo) in the office E02. 029. I would never forget the fun you brought into the office all the time. Great gratitude also goes to all other people in the Thorsten Mascher's Lab (Carolin Kobras, Fransiska Dürr, Qiang Liu, Jara Radeck, Julia Bartels, Karen Schrecke, Nikolai Peschek, Philipp Popp, Mona Dotzler…) and big *Streptomyces* Lab in Norwich (Neil Holmes, Susan Schlimpert, Morgan Feeney, Martin Schafer, Anyarat Thanapipatsiri, Jean Franco Castro, Valeria Razmilic, Dave Widdick, Zhiwei Qin …) while I stayed there. You were the people who helped me in so many aspects and created such a nice working environment around me during my thesis. You were the people who comforted me when I got stuck into my experiment.

I also greatly acknowledge the China Scholarship Council to provide me living expenses for my study these years. Big thanks also go to those people who help me to apply for this scholarship and take care of all other things for my study.

Finally, I want to dedicate this thesis to my wife, my brother and my parents. During these four years, I seldom went home and got together with you. Thanks a lot for all your support and encouragement for my study here.

Curriculum Vitae -Xiaoluo Huang

Gender / Birth: Male

Birth: March 1988

Birthplace: Hubei, China

Nationality: Chinese

Education:

09.2011-present	Ludwig-Maximilians-University (LMU) Munich (Germany), Ph.D study
09.2013-08.2014	John Innes Centre, University of East Anglia, UK, visiting Ph.D student
09.2008-06.2011	Master in Biochemistry and Molecular Biology, State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University (China)
09.2004-06.2008	B.S in Biological Science, Huazhong Agricultural University (China)

International conferences

1. 29th Symposium “Mechanisms of Gene Regulation”. 2012. Ammersee, Germany. Title “Identification and classification of novel extracytoplasmic function (ECF) sigma factors from the phyla *Planctomycetes* and *Actinobacteria*”. Poster.

2. Annual Conference of the Association for General and Applied Microbiology (VAAM). 2015. Marburg, Germany. Title “Defining the regulon of genes controlled by σ^E, a key regulator of cell envelope stress in *Streptomyces coelicolor*”. Talk.

Selected publications

1. Tran NT[#], **Huang XL**[#], Bush MJ, Chandra G, Pinto D, Hong HJ, Hutchings MI, Mascher T and Buttner MJ. Defining the regulon of genes controlled by σ^E, a key regulator of cell envelope stress in *Streptomyces coelicolor*. (**[#]co-first author**, in preparation)
2. **Huang X**[#]**,** Pinto D[#], Fritz G, Mascher T. 2015. Environmental sensing in *Actinobateria*: A comprehensive survey on the signaling capacity of this phylum. Journal of bacteriology 197:2517-35. (**[#]co-first author**)
3. Jogler C, Waldmann J, **Huang X**, Jogler M, Glöckner FO, Mascher T, Kolter R. 2012. Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in *Planctomycetes* by comparative genomics. Journal of bacteriology 194:6419-6430.
4. Lin L, Meng X, Liu P, Hong Y, Wu G, **Huang X**, Li C, Dong J, Xiao L, Liu Z. 2009. Improved catalytic efficiency of Endo-β-1, 4-glucanase from *Bacillus subtilis* BME-15 by directed evolution. Applied microbiology and biotechnology 82:671-679.