Characterization of ECF42, a novel group of extracytoplasmic function (ECF) σ factors with C-terminal regulatory extensions

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Qiang Liu

München

18.09.2018

Diese Dissertation wurde angefertigt

unter der Leitung von Prof. Dr. Thorsten Mascher

im Bereich von Fakultät für Biologie

an der Ludwig-Maximilians-Universität München

Erstgutachter/in: Prof. Dr. Thorsten Mascher

Zweitgutachter/in: Dr. Ralf Heermann

Tag der Abgabe: 18.09.2018

Tag der mündlichen Prüfung: 07.12.2018

ERKLÄRUNG

Ich versichere hiermit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist.

Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt.

Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen. München, den

München, den 18.09.2018

Qiang Liu

Abstrakt

Nach den Ein-und Zweikomponentensystemen, stellen die *extracytoplasmic function* (ECF) σ Faktoren die dritte große Gruppe innerhalb der bakteriellen Signaltransduktion dar. Eine kürzlich durchgeführte phylogenetische Analyse der ECF σ-Faktoren ergab eine große Anzahl neuer Gruppen, deren Funktionen und Regulationsmechanismen noch nicht aufgeklärt sind. Diese Arbeit beschäftigt sich mit der Untersuchung und Charakterisierung der ECF σ-Faktoren der Gruppe 42. ECF σ Faktoren der Gruppe 42 sind stark verbreitet, insgesamt in 11 verschiedenen Phyla vorkommend, jedoch am häufigsten in den Actinobakterien zu finden. Eine Analyse des genomischen Kontextes der ECF42 σ Faktoren in Actinobakterien ergab, dass, im Gegensatz zu den klassischen ECF σ Faktoren, diese keinen Anti-σ Faktor im gleichen Operon kodieren. Stattdessen liegen die ECF42 σ Gene gemeinsam mit Genen, die für ein konserviertes, sogenanntes DGPF Protein kodieren, in einer Transkriptionseinheit. Diese DGPF Proteine sind aber offenbar nicht an der Regulation der Aktivität der ECF42 σ Faktoren beteiligt. Der Modellorganismus, auf dem die hier in dieser Arbeit gezeigten Daten basieren, ist *Streptomyces venezuelae*, der drei ECF42 σ Faktoren enthält. Zunächst wurden die Promotoren der ECF42 σ Faktoren hinsichtlich ihrer Erkennungssequenz TGTCGA (-35 Region) und CGA/TC (-10 Region) untersucht und es wurde herausgefunden, dass die Expression der ECF42 σ Faktoren positiv autoreguliert ist. Phänotyp *Microarrays* und RNA-Sequenzierungen wurden durchgeführt, um einen Einblick in die zelluläre Funktion der ECF42 σ Faktoren in *S. venezuelae* zu erlangen und um die Regulons dieser zu identifizieren. Dabei wurde herausgefunden, dass die ECF42 σ Faktoren hauptsächlich Gene regulieren, die für DGPF Proteine mit unbekannter Funktion kodieren. Eine Besonderheit der ECF42 σ Faktoren ist ihre lange C-terminale Verlängerung, die eine *tetratricopeptide repeat* (TPR) Domäne enthält. Diese Domäne ist in den klassischen ECF σ Faktoren nicht enthalten. Es wird angenommen, dass die TRP-Domäne eine bedeutende Funktion bei Protein-Protein Interaktionen vermittelt. In dieser Arbeit konnte gezeigt werden, dass eine Wechselwirkung zwischen C-Terminus und N-terminaler σ-Domäne besteht, die sich positiv auf die regulatorische Aktivität des ECF42 σ Faktors auswirkt. Eine Verkürzung des C-Terminus hat hingegen einen vollständigen Aktivitätsverlust des ECF42 σ Faktors zu Folge, was ebenfalls auf eine enorme Bedeutung des C-Terminus für die Aktivität des ECF42 σ Faktors hinweist. Zusammenfassend liefert diese Arbeit erste Einsichten in den Mechanismus der Aktivierung der ECF42 σ Faktoren sowie deren zelluläre Funktion.

Abstract

Extracytoplasmic function (ECF) σ factors represent the third most abundant fundamental principle of bacterial signal transduction, outnumbered only by one- and two-component systems. A recent census of ECF σ factors revealed a large number of novel groups whose functions and regulatory mechanisms have not yet been elucidated. Here we report the characterization of members of the novel group ECF42. ECF42 σ factors are a highly abundant and widely distributed ECF group that is present in 11 phyla, but is predominantly found in *Actinobacteria*. Analysis of the genomic context conservation did not identify a putative anti-σ factor in the same operon of ECF σ factor as classical ones. Instead, ECF42 genes are co-transcribed with genes encoding a conserved so-called 'DGPF protein', which is not involved in the regulation of ECF42 σ factor activity. We have experimentally verified the target promoter of these ECF σ factors ("TGTCGA" in the -35 region and "CGA/TC" in the -10 region), which was found upstream of the ECF42-encoding operons in *Streptomyces venezuelae*, suggesting that ECF42 σ factors are positively auto-regulated. ECF42 triple deletion mutant of *S. venezuelae* (*Δsven_0747 Δsven_4377 Δsven_7131*) was generated and submitted for Phenotype Microarrays to identify ECF42-related phenotypes. RNA sequencing (RNA-seq) was performed to define the regulons of the three ECF42 proteins in *S. venezuelae*, which identified mostly genes encode DGPF proteins with unknown function. In contrast to typical ECF σ factors, ECF42 σ factors are characterized by a long C-terminal extension containing a tetratricopeptide repeat (TPR) domain, which is postulated to mediate protein-protein interactions. An interaction between the C-terminal extension and the N-terminal σ domains was hypothesized and supported by experimental evidence from a mutational analysis. The putative interaction, mediated by the co-variable residues, plays a positive regulatory role on the activity of the ECF42 σ factor. Truncations of the C-terminal extension of ECF42 abolished σ factor activity completely, suggesting that it is necessary for σ factor activity. In conclusion, this work provides the first insights into the function and mechanism behind ECF42 σ factors' activation.

Abbreviations

Anti-sigma factor domain	ASD
5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside	X-gal
Bacterial Adenylate Cyclase-based Two-Hybrid	BACTH
DGPF family protein containing a conserved YciI domain	DGPF
Diethylpyrocarbonate	DEPC
Dithiothreitol	DTT
Ethylenediaminetetraacetic acid	EDTA
Extracytoplasmic function	ECF
Horseradish peroxidase	HRP
Isopropyl β-D-1-thiogalactopyranoside	IPTG
One component system	1CS
Phenylmethanesulfonylfluoride or phenylmethylsulfonyl fluorid	PMSF
p-nitrophenyl-β-D-glucuronide	PNPG
Relative luminescence output	RLU
Ribonucleic acid polymerase	RNAP
Sodium dodecyl sulfate	SDS
Tetratricopeptide repeat	TPR
Two component system	2CS

List of figures

List of tables

Contents

1. Introduction

1.1 Bacterial σ factors

The initial step for controlling of gene expression is the transcription of DNA into messenger RNA by the RNA polymerase (RNAP). The core RNAP is a complex multi-subunit enzyme of five subunits: two identical α subunits, as well as unique β, β´ and ω subunits (Ishihama, 2000; Gruber and Gross, 2003). Another essential component of the RNAP is the σ factor, which binds to the core RNAP creating the RNAP holoenzyme, and is responsible for promoter recognition in the initiation phase of transcription (Helmann and Chamberlin, 1988; Gruber and Gross, 2003).

Bacterial σ factors can be classified into two large and phylogenetically unrelated protein families, σ^{54} and σ^{70} (Helmann and Chamberlin, 1988; Gross et al., 1998), of which the σ^{70} is by far the most abundant. This σ factor family includes the essential housekeeping σ factors (group 1) and three alternative σ factor groups (Helmann and Chamberlin, 1988; Lonetto et al., 1992; Helmann, 2002; Gruber and Gross, 2003): those including the non-essential paralogs of the housekeeping σ factors (group 2), the flagellar, heat shock and sporulation-specific σ factors (group 3) and the extracytoplasmic function (ECF) σ factors (group 4).

Primary σ factors (group 1). The group 1 σ factors include the *E. coli* σ^{70} and its orthologues (Lonetto et al., 1992). These σ factors are essential proteins responsible for most of the transcription in rapidly growing bacterial cells and are thus often called primary σ factors (Burgess and Anthony, 2001; Helmann, 2002; Murakami and Darst, 2003). Binding of the σ factor to the core RNAP occurs at amino acids 260 to 309 of the β´ subunit (Burgess and Anthony, 2001), with weaker binding also mediated by several alternate sites on the β and β´ subunits (Arthur and Burgess, 1998; Arthur et al., 2000; Katayama et al., 2000; Gruber et al., 2001; Anthony and Burgess, 2002). During transcription initiation, the holoenzyme, directed by the σ factor, binds to DNA thereby forming the closed complex, in which the DNA is still double-stranded. As the DNA is unwound and the initiation site exposed, the DNA/holoenzyme structure forms the open complex, allowing elongation to proceed. After transcription initiation, the σ factor is released from the RNAP/DNA complex

and can be recycled to bind additional core RNAP complexes to again mediate transcriptional initiation (Craig et al., 1998; Sen et al., 2000; Saecker et al., 2002).

The primary σ factors are usually between 40 to 70 kDa in size and contain four conserved protein domains (σ_1, σ_2, σ_3 and σ_4) as displayed in **Figure 1.1** (Lonetto et al., 1992; Wosten, 1998; Paget, 2015). Domain σ_1 acts to prevent promiscuous binding of the σ factor to DNA in the absence of RNAP, and is not conserved among all σ factors (Helmann and Chamberlin, 1988). Region σ_2, or specifically, $\sigma_{2.1}$ and $\sigma_{2.2}$, are required for binding to core RNAP, while region $\sigma_{2.3}$ facilitates DNA promoter melting at the "-10" promoter site, and region $\sigma_{2.4}$ recognizes the "-10" consensus sequence "TATAAT", which consists of weak A-T bonds to facilitate DNA melting (Aiyar et al., 1994; Shuler et al., 1995; Tatti and Moran, 1995; Joo et al., 1997; Sharp et al., 1999; Gruber and Gross, 2003). The σ_3 region of σ^{70} factors interacts with the DNA upstream of the "-10" region, which normally contain a "TG" motif located at approximately "-15/-14" position on the promoter (Helmann and Chamberlin, 1988; Helmann, 1995; Sabelnikov et al., 1995; McCracken and Timms, 1999; Bashyam and Hasnain, 2004). The region σ_4 contains a helix-turn-helix motif, which interacts with the "TTGACA" consensus at the "-35" sequence of the promoter (Campbell et al., 2002; Murakami et al., 2002).

Alternative σ factors (group 2-4). In addition to primary σ factors, most bacteria also harbor alternative σ factors, which are non-essential and recognize distinct promoter sequences. Alternative σ factors are usually involved in responding to changes in the environment and rapidly initiate changes in transcription of specific subsets of genes (Helmann, 2002; Paget, 2015). So far, the alternative σ factors are mainly classified into three distinct groups (group 2, group3 and group 4) based on structural and functional homology (Burgess and Anthony, 2001; Helmann, 2002; Murakami and Darst, 2003). Group 2 σ factors are structurally very similar to group 1 and contain domains σ_1 to σ_4 (**Fig. 1.1**); but they are not essential (Helmann, 2002; Gruber and Gross, 2003). Group 3 consists σ factors involved in niche-specific regulation and can be further divided based on their regulation of heat shock, flagellar or sporulation genes. Structurally, proteins in this group typically lack the σ_1 domain (**Fig. 1.1**) (Gruber and Gross, 2003). Group 4 consists of the extracytoplasmic function (ECF) σ factors, which were named due to the initial analysis of a small number of examples that respond to periplasmic stress and heat shock, iron transport or protein secretion (Lonetto et al., 1994). Structurally, proteins of group 4 contain only the

σ_2 and σ_4 domains (Helmann, 2002), which are necessary and sufficient for assembling with RNAP and promoter recognition.

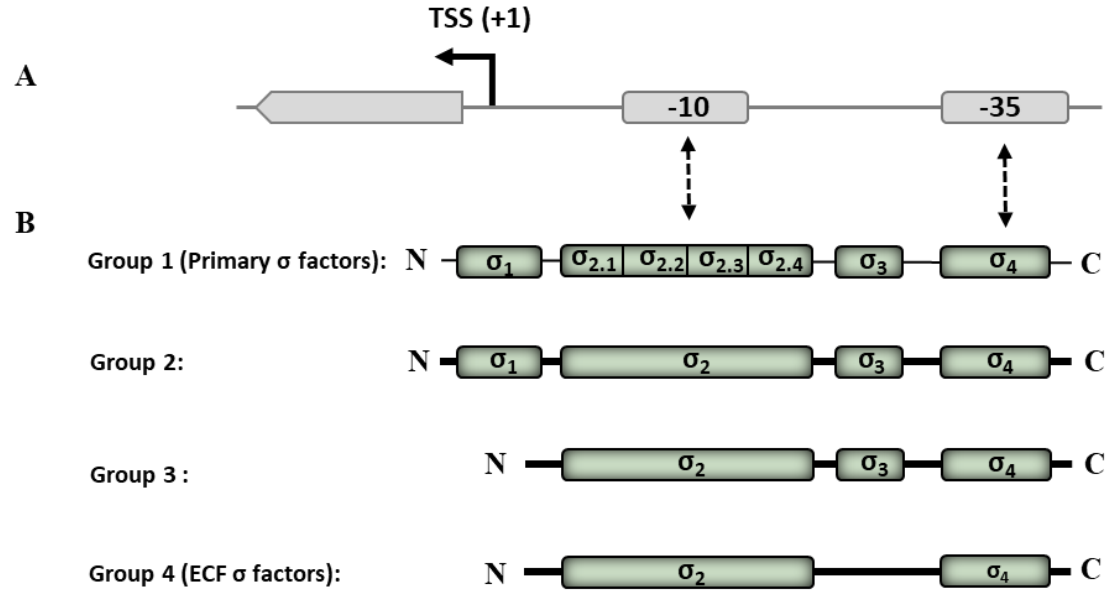


Figure 1. 1 Domain architecture of the σ^{70} protein family. (A) Schematic of a typical promoter, where +1 represent the transcriptional start site (TSS) and the bent arrow indicate the direction of transcription. The -35 and -10 sequences of the promoter are represented as grey boxes. These are recognized by the σ_4 and σ_2 domains, respectively. **(B)** Domain architecture (σ_1, σ_2, σ_3 and σ_4) in the four σ^{70} groups. The σ factors from group 1 (primary σ factors) and group 2 possess four domains (σ_1, σ_2, σ_3 and σ_4), while the σ factors from group 3 lack σ_1 but possess σ_2, σ_3 and σ_4 domains. The group 4 (ECF) σ factors only contain the σ_2 and σ_4 domains.

1.2 Extracytoplasmic function (ECF) σ factors

The ability of a cell to accurately respond to changing environments and lurking competitors is a prerequisite to survive the struggle for suitable habitats. In order to mediate such adaption processes, bacteria possess different means to connect an extracellular input with an appropriate cellular response. ECF σ factors are the third most abundant fundamental principle of bacterial signal transduction, outnumbered only by one- and two-component systems (1CS/2CS) (Staron et al., 2009; Huang et al., 2015). This group of alternative σ factors regulate diverse processes, such as stress response, differentiation, secondary metabolism and virulence (Helmann, 2002; Gicquel et al., 2013; Haines-Menges et al., 2014; Luo et al., 2014; Dou et al., 2018; Lopez-Garcia et al., 2018).

Their hallmark features are summarized in **Fig. 1.2** (Helmann, 2002; Mascher, 2013):

(i) ECF σ factors share a characteristic protein domain architecture with only two of the conserved regions of σ^{70} proteins remaining, namely σ_2 and σ_4 (**Fig. 1.1**), which are sufficient for promoter recognition and binding to RNAP.

(ii) The activity of ECF σ factors is regulated by their cognate anti-σ factors, which are often membrane-anchored proteins encoded in the same operon as the σ factor. In the absence of a signal, the anti-σ factor tightly binds the σ factor, thereby keeping it inactive. Once stimulated, the anti-σ factor releases the σ factor, which can then recruit the RNAP core enzyme to redirect transcription initiation to its target promoters.

(iii) ECF σ factors recognize alternative promoter sequences typically containing an 'AAC' motif in the -35 region and a 'CGT' in the -10 region.

(iv) The presence of this promoter motif upstream of the ECF-encoding operon leads to positive autoregulation of most ECF σ factors, thereby enhancing their activating effect as long as inducing conditions prevail. Once the stimulus cease, the simultaneous upregulation of the cognate anti-σ factor then ensures a swift shut-off of the σ factor's activity.

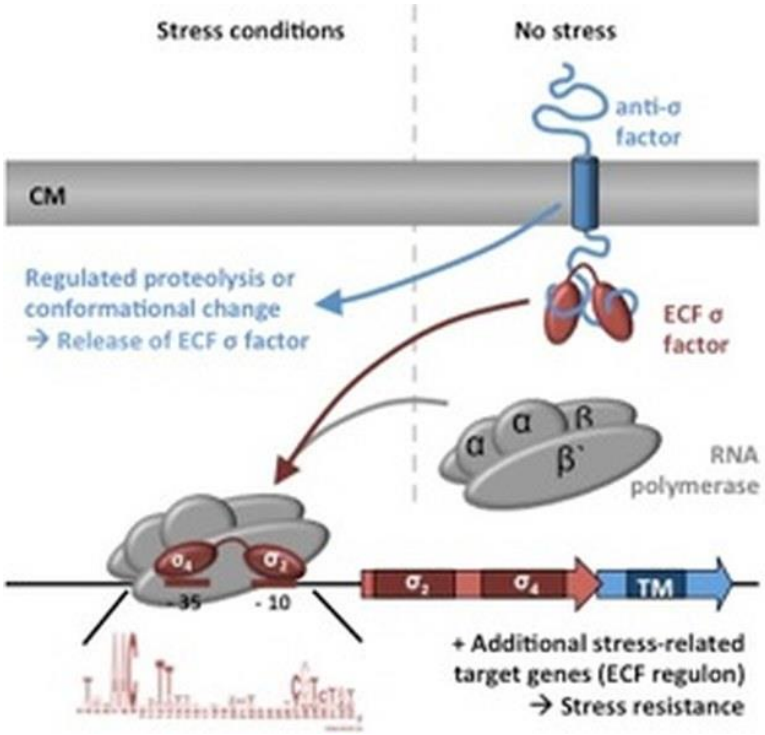


Figure 1. 2 Overview of the hallmark characteristics of ECF σ factors. The ECF σ factor is represented in green, and the cognate anti-σ factor and associated processes are highlighted in blue. The RNAP core enzyme with its subunits is shown in grey. The promoter is represented by the -35 and -10 boxes; a typical ECF-dependent promoter signature is shown below. R2/R4, conserved signature regions (regions σ_2 and σ_4, respectively) of σ^{70} proteins; CM, cytoplasmic membrane; TM, transmembrane region. Figure modified from (Mascher, 2013).

1.3 Classification of ECF σ factors

The first work on ECF classification was performed in 2009 and was based on the comprehensive phylogenetic analysis of more than 2700 ECF σ factors derived from 369 microbial genomes (Staron et al., 2009). This work defined 43 major groups (ECF01-ECF43, each containing over 10 sequences) (**Fig. 1.3**) and 24 minor groups (ECF101-ECF124, each containing less than 10 proteins). These groups have been defined based on the sequence similarity of the ECF σ factors and their cognate anti-σ factors, genomic context conservation as well as their group-specific target promoter signature. With an ever-increasing number and diversity of completed bacterial genome sequences becoming available, 28 additional new groups of ECF σ factors were subsequently defined and described (Gomez-Santos et al., 2011; Jogler et al., 2012; Huang et al., 2015). Thus, there are currently 94 ECF groups described, which include 62 major groups (ECF01-ECF63) and 32 minor groups (ECF101-ECF132). The main characteristics of each group regarding their genomic context conservation, taxonomic distribution, experimentally studied members, signaling and regulatory mechanism, and their (putative) physiological role have recently been summarized (Pinto and Mascher, 2016b).

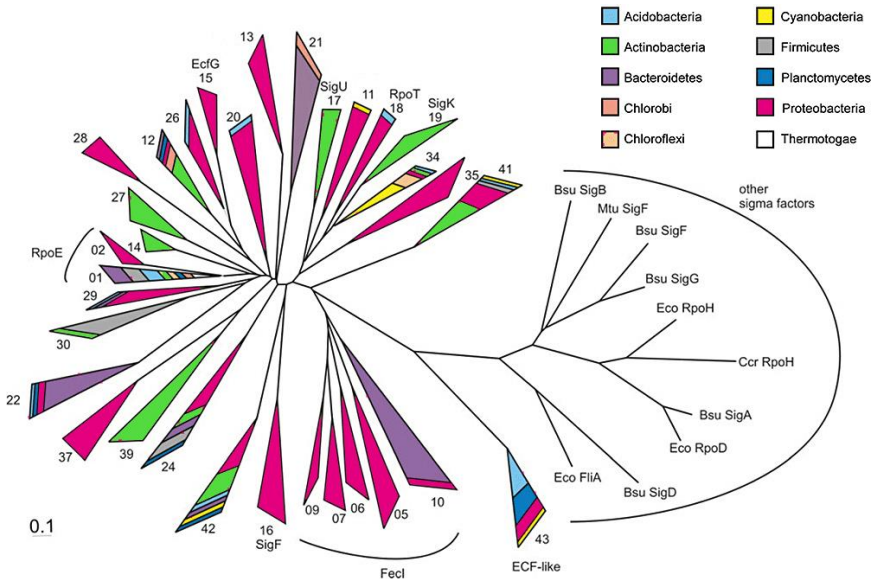


Figure 1. 3 Phylogenetic tree of ECF σ factors (Staron et al., 2009). The phylogenetic tree of ECF σ factors is based on a gapless multiple sequence alignment of regions σ_2 and σ_4, from the two most distant representatives from within each group. Each triangle represents one major group, color-coded according to the phylogenetic distribution of the ECFs factors. Thus, the length of the triangle's edges reflects the overall sequence diversity within the group (the longer the edges, the larger the sequence diversity within this group). CLUSTALW (Thompson et al., 1994) was used to construct the multiple sequence alignments. The tree was calculated by the Least Squares method of the Phylip (Felsenstein, 1989) programs PROTDIST and FITCH, which are implemented in the BioEdit Sequence Alignment Editor. For reasons of clarity, only the 32 most important groups are shown, which contain at least 20 proteins in the data set and/or show an extended genomic context conservation. Figure modified from (Staroń et al., 2009).

1.4 Physiological function of ECF σ factors

Although thousands of ECF σ factors have been identified and classified, only few of them have been characterized experimentally regarding their physiological functions. Based on the studies up to date, ECF σ factors in bacteria are involved in very diverse physiological processes, such as stress response, uptake process, secondary metabolism, virulence and the development or morphology of cells. The diverse functions of ECF σ factors that have already been characterized shall be briefly summarized in the next paragraphs. The ECF σ factors related functions from each ECF group have also been concluded in **Table 1.1**.

Envelope stress responses. The well-studied example of ECF σ factors involved in envelope stress responses is the *Bacillus subtilis* σ^W that belongs to group ECF01 (Helmann, 2006). In this case, the σ^W regulon was strongly induced by several antibiotics, most of which interfere with cell wall biosynthesis (Cao et al., 2002). Another example is the *E. coli* σ^E classified in group ECF02, which can be activated by heat or other stresses that generate unfolded envelope proteins thereby initiates the expression of genes encoding proteins for envelope synthesis (Alba and Gross, 2004).

Oxidative stress responses. Numerous groups of ECF σ factors are involved in oxidative stress responses (**Table 1.1**). Disruption of these ECF σ factors coding genes rendered the bacterial cells more sensitive to oxidative stress. This ECF-related phenotype was founded by the studies on the σ^H (group ECF12) of *Corynebacterium glutamicum* (Kim et al., 2005), σ^E (ECF14) of *Mycobacterium tuberculosis* (Wu et al., 1997), σ^T (ECF15) of *Caulobacter crescent* (Alvarez-Martinez et al., 2007), σ^F (ECF16) of *C. crescentus* (Alvarez-Martinez et al., 2006), σ^O (ECF21) of *Bacteroides fragilis* (Ndamukong et al., 2013), σ^M (ECF27) of *Corynebacterium glutamicum* (Nakunst et al., 2007), σ^F (ECF33) of *Bradyrhizobium japonicum* (Masloboeva et al., 2012), σ^R (ECF39) of *Streptomyces coelicolor* (Paget et al., 1998) and σ^J (ECF41) of *M. tuberculosis*.

Acquisition and secretion process. The well-characterized ECF σ factor involved in acquisition process is the *E. coli* σ^{FecI}, which control the transcription of the *fecA* operon coding a specific ferric citrate uptake system (Helmann, 2002). In this case, the presence of ferric citrate leads its binding to the outer membrane FecA protein under the Fe-limiting growth conditions. The signal generated by binding of ferric citrate with FecA is subsequently transmitted by FecA across the outer membrane into the periplasm to the anti-

σ factor FecR, which leads the release of σ^{FecI} from FecR resulting transcription of the *fecA* operon. Such an ECF σ factors involved iron uptake system has been demonstrated in *Pseudomonas putida, Pseudomonas aeruginosa, Serratia marcescens, Klebsiella pneumoniae, Aerobacter aerogenes, Bordetella pertussis, B. bronchseptica, B. avium, and Ralstonia solanacearum,* which has been reviewed by Braun and Mahren (Braun and Mahren, 2005). In contrast to acquisition, σ^{U} of *S. coelicolor* belongs to group ECF17, and has been found to be involved in protein secretion (Gordon et al., 2008).

Biosynthesis of antibiotics and other secondary metabolite. Biosynthesis of antibiotics is generally a complicated process and regulated by many factors in bacteria. As a regulator, ECF σ factors have been demonstrated that involved in the biosynthesis of antibiotics in many bacterial organisms. The σ^{Mibx} (ECF20) as a positive regulator is involved in the microbisporicin synthesis in *Microbispora coralline* (Foulston and Bibb, 2011). The σ^{T} (ECF27) of *S. coelicolor* (Mao et al., 2013) and σ^{Y} (ECF31) of *Bacillus subtilis* (Mendez et al., 2012) are involved in the production of actinorhodin and sublancin respectively. In addition, σ^{25} (ECF39) and σ^{6} (ECF41) of *Streptomyces avermitilis* are both involved in the production of avermectin (Jiang et al., 2011; Luo et al., 2014).

Pathogenicity and virulence. Virulence and virulence-associated genes are those that contribute to at least one aspect of bacterial disease transmission and infection processes. The expression of virulence and virulence-associated genes in response to particular stimuli is often regulated by ECF σ factors, which have been demonstrated in many cases. The σ^{L} (ECF17) of *Mycobacterium tuberculosis* regulates polyketide synthases and secretion of membrane proteins that are required for virulence (Hahn et al., 2005). While the σ^{D} (ECF40) of *M. tuberculosis* governs the expression of a small set of ribosomal genes typically expressed in stationary phase during *in vitro* growth and that loss of σ^{D} reduces macrophage TNF-alpha secretion as well as the lethality of *M. tuberculosis* infection in mice (Calamita et al., 2005). Other than *M. tuberculosis*, the virulence related ECF σ factors have been also found in other pathogens, such as σ^{E} of *Salmonella enterica Serovar Typhimurium* (Humphreys et al., 1999), σ^{HrpL} (ECF32) of *Pseudomonas syringae* (Ferreira et al., 2006) and σ^{PvdS} (ECF09) of *Pseudomonas aeruginosa* (Wilderman et al., 2001).

Table 1. 1 Physiological functions of characterized ECF σ factors.

ECF group	ECF name	Organism	Function	Reference
ECF01	Sig W	*B. subtilis*	Envelope stress response, antimicrobial compound resistance and detoxification	(Helmann, 2006)
ECF02	Sig E	*E. coli*	Envelope stress response	(Alba and Gross, 2004)
	Sig 22	*P. aeruginosa*	Cell wall stress response and alginate production	(Wood and Ohman, 2009)
ECF05	Sig FpvI	*P. aeruginosa*	Uptake of the siderophores ferrichrome and desferrioxamine	(Mettrick and Lamont, 2009)
ECF07	Sig HasI	*S. marcescens*	Haem uptake	(Biville et al., 2004)
ECF08	Sig VreI	*P. aeruginosa*	Adaptive response to phosphate-limiting conditions	(Faure et al., 2013)
ECF09	Sig PvdS	*P. aeruginosa*	Iron uptake	(Leoni et al., 2000)
ECF10	Sig	*B. thetaiotaomicron*	Mobilizing complex carbohydrates	(Xu et al., 2004)
ECF11	Sig RpoE	*R. sphaeroides*	Response to superoxide stress	(Anthony et al., 2005)
ECF12	Sig R	*S. coelicolor*	Disulphide and oxidative stress response, mycothiol metabolism	(Paget et al., 2001; Newton and Fahey, 2008; Feeney et al., 2017)
	Sig H	*C. glutamicum*	Oxidative and heat stress	(Kim et al., 2005)
ECF13	Sig MsrAB	*N. gonorrhoeae*	Oxidative damage response	(Gunesekere et al., 2006)
ECF14	Sig E	*M. tuberculosis*	Heat shock, acidic pH, detergent and oxidative stress	(Wu et al., 1997)
ECF15	Sig T	*C. crescentus*	Osmotic and oxidative stress responses	(Alvarez-Martinez et al., 2007)
	Sig PhyR	*M. extorquens*	Carbon starvation and heat shock	(Gourion et al., 2008; Francez-Charlot et al., 2009a)
	Sig RpoE	*S. meliloti*	Heat and salt, carbon and nitrogen starvation	(Sauviac et al., 2007)
ECF16	Sig F	*C. crescentus*	Oxidative and heavy metal stress response	(Alvarez-Martinez et al., 2006; Kohler et al., 2012)
ECF17	Sig U	*S. coelicolor*	Protein secretion	(Gordon et al., 2008)
	Sig L	*M. tuberculosis*	Polyketide synthases and virulence	(Hahn et al., 2005)

ECF group	ECF name	Organism	Function	Reference
ECF18	Sig RpoT	*P. putida*	Tolerance of toluene and other organic solvents	(Duque et al., 2007)
ECF19	Sig K	*M. bovis*	Production of antigenetic proteins	(Charlet et al., 2005)
ECF20	Sig CnrH	*C. metallidurans*	Nickel resistance	(Grosse et al., 2007)
ECF20	Sig Mibx	*M. coralline*	Antibiotic biosynthesis and immunity	(Foulston and Bibb, 2011)
ECF21	Sig O	*B. fragilis*	Oxidative stress response	(Ndamukong et al., 2013)
ECF26	Sig E	*S. novella*	Involving thiosulfate-oxidizing pathway	(Kappler et al., 2001)
ECF26	Sig Prtl	*P. fluorescens*	Production of the germination-arrest factor	(Okrent et al., 2014)
ECF27	Sig M	*C. glutamicum*	Heat, cold and thiol-oxidant diamide stress	(Nakunst et al., 2007)
ECF27	Sig T	*S. coelicolor*	Cell differentiation and actinorhodin production	(Feng et al., 2011; Mao et al., 2013)
ECF30	Sig V	*B. subtilis*	Resistance to lysozyme	(Hastie et al., 2013)
ECF31	Sig Y	*B. subtilis*	Sublancin producing and resistance	(Mendez et al., 2012)
ECF32	Sig HrpL	*P. syringae*	Hypersensitive response and pathogenicity	(Ferreira et al., 2006)
ECF33	Sig F	*B. japonicum*	Oxidative stress response	(Masloboeva et al., 2012)
ECF36	Sig C	*M. tuberculosis*	Pathogenesis and adaptive survival in the host	(Karls et al., 2006)
ECF39	Sig R	*S. coelicolor*	Cell envelope stress response	(Kang et al., 1999)
ECF39	Sig 25	*S. avermitilis*	Biosynthesis of avermectin and aligomycin	(Luo et al., 2014)
ECF40	Sig D	*M. tuberculosis*	Pathogenesis	(Calamita et al., 2005)
ECF41	Sig J	*M. tuberculosis*	Oxidative stress response	(Hu et al., 2004)
ECF41	Sig 6	*S. avermitilis*	Production of avermectin	(Jiang et al., 2011)
ECF42	Sig 10	*P. putida*	Antibiotic stress response and biofilm formation	(Tettmann et al., 2014)
ECF44	Sig CorE	*M. xanthus*	Copper homeostasis	(Gomez-Santos et al., 2011)
ECF52	Sig 52	*S. coelicolor*	Cell morphogenesis and secondary metabolism	(Lopez-Garcia et al., 2018)

Taken together, ECF σ factors are involved in very diverse physiological processes and functions in bacteria as we enumerated here, which include (i) cell wall stress response; (ii) specific oxidative stress; (iii) Acquisition and secretion process; (iv) biosynthesis of antibiotics and (v) pathogenicity and virulence. In addition to the known physiological functions of the group-define ECF σ factors captured by this classification, novel physiological roles of ungrouped ECF σ factors are unraveled continuously. For example, *Myxococcus xanthus* σ^{DdvS} have been revealed that involved in regulating the expression of the CRISPR-Cas system, thereby against invading genetic elements (Gaballa et al., 2018a). As more bacterial complete genomes are available, more and more diverse groups of ECF σ factors will be defined, meanwhile, more unknown physiological functions of ECF σ factors will be disclosed.

1.5 Mechanism of ECF σ factor activation

Most of ECF σ factors require a negative regulatory function to keep them inactive as long as non-inducing conditions prevail, thus, the activation of ECF σ factor is the first step for ECF σ factors to respond to a stimulus. Generally, the activity of ECF σ factors is inhibited by the association with their cognate anti-σ factors in the absence of an inducing stimulus (**Fig. 1.2**). However, the ECF classification already indicated that ECF-based signal transduction is far more complex and diverse than has been recognized and appreciated so far. This diversity in activating ECF σ factors shall be described in the following sections.

1.5.1 Regulated proteolysis of membrane-anchored anti-σ factors

Regulated proteolysis of membrane-anchored anti-σ factors is the best-understood mechanism of ECF σ factor activation. This activation mechanism was described or proposed for the ECF σ factors from groups ECF01-ECF04, ECF17, ECF30 and ECF40 (Ellermeier and Losick, 2006; Wood and Ohman, 2009; Hastie et al., 2013; Jaiswal et al., 2013) (**Fig. 1.4 A**). The well-studied example is that of *E. coli* σ^{RpoE} and its inhibitory regulator RseA, which is a membrane protein functions as an anti-σ factor (Helmann, 2002). In this case, the activity of σ^{RpoE} is primarily determined by the ratio of RseA to σ^{RpoE}. RseA can be rapidly degraded by the putative inner membrane serine protease (DegS) in response to extracytoplasmic stress of accumulation of misfolded or unfolded protein. This leads an increase in the free pool of σ^{RpoE} and initiates the transcription of σ^{RpoE} from its target

promoter (Ades et al., 1999). Another characterized example is that of *B. subtilis* σ^w (Helmann, 2002; Heinrich and Wiegert, 2009). Briefly, the membrane-anchored anti-σ factor binds σ^w and inhibits its activity in the absence of an inducing stimulus. Once the specific stimulus is perceived, the anti-σ factor release the σ^w from its inhibitory grip due to its three successive proteolytic steps: First, the cleavage of the extracellular domain (site-1) of the anti-σ factor. Second, the cleavage of the transmembrane helix (site-2) regulated by intramembrane proteolysis. Third, the degradation of the cytoplasmic portion of the anti-σ factor mediated by the cytoplasmic Clp-protease complex. This leads to the release of σ^w its anti-σ factor, allowing it to assemble with the RNAP and initiate the transcription of its target genes.

1.5.2 Conformational change of the anti-σ factor

ECF σ factors can be activated by disassociation from their cognate anti-σ factor not only by proteolytic degradation of the anti- σ factor but also by conformational changes (**Fig. 1.4 B**). This mechanism of ECF activation has been experimentally verified or proposed for members of the groups ECF11-ECF14, ECF101, ECF117 and ECF126 (Kang et al., 1999; Zdanowski et al., 2006; Dufour et al., 2008; Barik et al., 2010; Greenwell et al., 2011). The best-understood paradigms are RpoE (σ)-ChrR (anti-σ) from *Rhodobacter sphaeroides* (ECF11). In this case, the soluble anti-σ factor ChrR senses redox stress through a mechanism involving disulfide bridge formation of the conserved cysteine residues in its C-terminal region. This leads to the conformational change of ChrR and release of RpoE from the RpoE-ChrR complex. This allows for initiation of transcription from the target promoters of RpoE (Greenwell et al., 2011). The similar activation mechanism of ECF σ factor has also been demonstrated for SigR (σ)-RsrA (anti-σ) of *Streptomyces coelicolor* (ECF12), which is involved in mounting the disulfide stress response. The RsrA consist of 105 amino acids including 7 cysteines. In its reduced state, RsrA binds a single zinc atom (Zn-RsrA) and forms a 1:1 complex with SigR that suppresses its transcriptional activity (Kang et al., 1999). Once a "trigger disulfide" is formed in RsrA in the presence of suitable stimulus, it will drive the expulsion of the bound metal ion and to cause a change in protein conformation (Li et al., 2003). This in turn disrupts the SigR complex, releasing SigR to bind RNA polymerase.

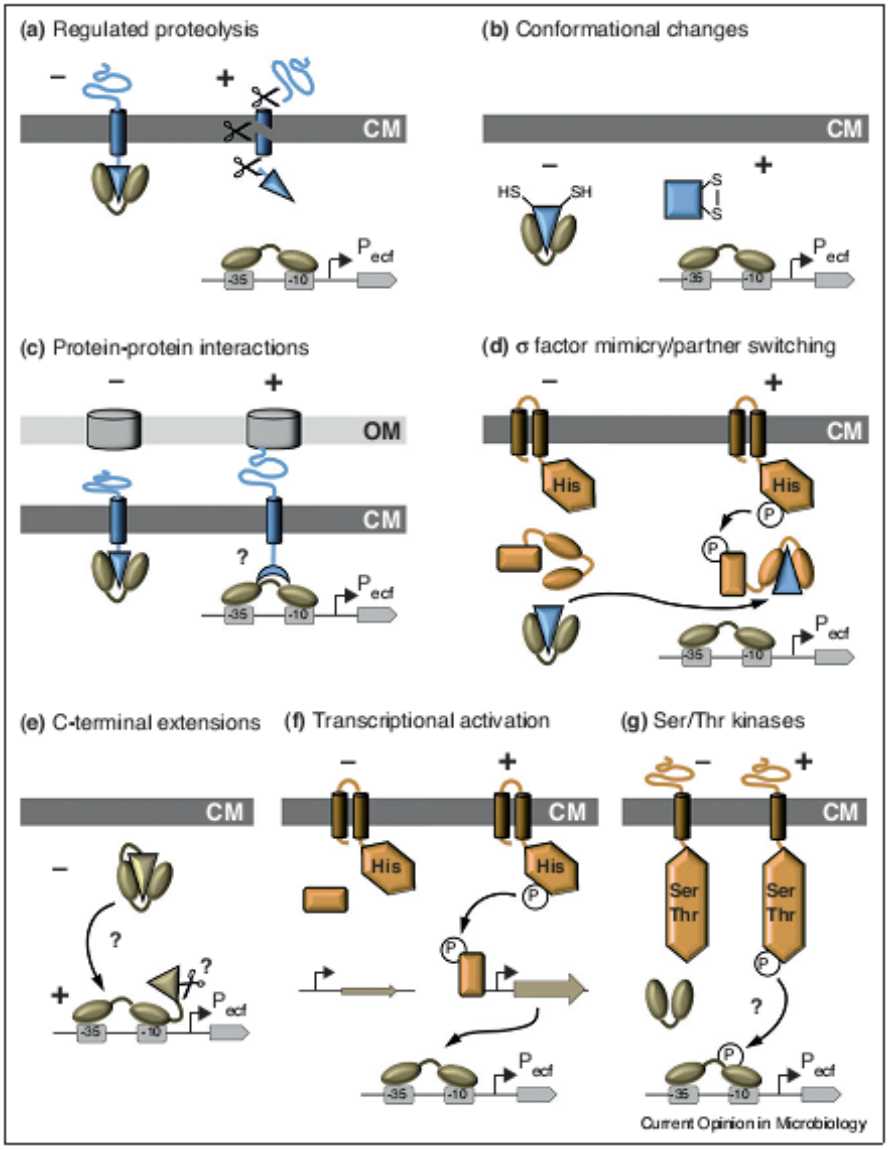


Figure 1. 4 Activation mechanisms of ECF σ factors. ECF σ factors and anti-σ factors are shown in gray and blue, respectively. The histidine kinases, serine/threonine kinases and response regulators are shown in brown. "+" or "-" indicates the presence or absence of a trigger stimulus. Figure from (Mascher, 2013)**.**

1.5.3 Activation of ECF σ factors by protein-protein interaction

Activation of ECF σ factors through protein-protein interaction cascades is the third well-described mechanism of ECF σ factor activation, which normally involves additional outer membrane proteins along with the σ and anti-σ proteins (**Fig. 1.4 C**). Such a mechanism was proposed for ECF σ factors of groups ECF05-ECF08, ECF10, ECF19 and ECF46 (Biville et al., 2004; Braun and Mahren, 2005; Mettrick and Lamont, 2009; Shukla et al., 2014). The best-described example is the FecI (σ)-FecR (anti-σ) pair from *E. coli*. In the absence of a stimulus, FecI is kept inactive by its cognate membrane-associated anti-σ factor, which

contains a N-terminal anti-σ domain and a C-terminal FecR domain. Once the inducing stimulus, iron-citrate, is bound to the outer membrane protein FecA. The FecA will then interact with the periplasmic domain of the anti-σ FecR. This interaction leads the release of the σ factor FecI, bound to the cytoplasmic domain of FecR, and resulting active FecI then initiate the transcription of the *fecA* operon coding a specific citrate uptake system (Helmann, 2002).

1.5.4 Activation of ECF σ factors through partner-switching mechanism

The activation of ECF σ factors based on the partner-switching mechanism is achieved by a third partner protein, which competitively binds the anti-σ factor, thereby extricating the σ factor from the inhibitory grip of anti-σ factor (**Fig. 1.4 D**). The remarkable example of such a mechanism is demonstrated by σ^{EcfG} (ECF15), anti-σ (NepR) and a third partner (PhyR) in *Alphaproteobacteria* (Campagne et al., 2012). PhyR is a response regulator of a bacterial 2CS and is unique in that it not only contains a classic receiver domain with the conserved phosphorylation site and catalytical motifs, but also contains an output domain that shows homology to ECF σ factors. The ECF σ factor-like output domain retains an ECF σ factor fold but lacks the $\sigma_{2.4}$ region and degenerate in $\sigma_{4.2}$ region (Herrou et al., 2010), thereby without having the DNA-binding ability. In the absence of an inducing stimulus, the σ^{EcfG} is sequestered by its cognate anti-σ (NepR), and the anti-σ antagonist, PhyR, remains in an unphosphorylated inactive state. Once triggered by a suitable stimulus, PhyR becomes phosphorylated and binds to NepR competitively via its ECF σ factor-like output domain. This leads the release of σ^{EcfG} from NepR and directs transcription of genes toward the stress response (Francez-Charlot et al., 2009b; Campagne et al., 2012).

1.5.5 Transcriptional activation of ECF σ factors

In addition to anti-σ factors, which link the input to the ECF σ factor, other regulators, such as two-component systems (2CS) can also play a role on the connection between signal input and ECF regulation, in this case at the level of transcription (**Fig. 1.4 F**). Such a mechanism was described for ECF σ factors from group ECF30, ECF32 and ECF39 (Paget et al., 1999a; Nizan-Koren et al., 2003; Hastie et al., 2013). In the case of σ^{E} (ECF39) from *S. coelicolor*, its expression is induced by the 2CS CseBC upon cell envelope stress. As a result of *sigE* expression, the σ^{E}-target genes – encoding the enzymes necessary for cell wall biosynthesis

– are upregulated (Paget et al., 1999b; Hong et al., 2002). Another example for such an ECF activation mechanism was also reported for σ^{HrpL} (ECF32) from *Pseudomonas syringae*. In this case, the HrpXY-like 2CS regulates the transcription of gene encoding σ^{HrpL} indirectly, via control of the expression of two homologous DNA-binding proteins, which induce the expression of σ^{HrpL} directly (Merighi et al., 2003; Lan et al., 2006).

1.5.6 Activation of ECF σ factors by Ser/Thr protein kinases

Based on comparative genomics analysis, a number of conserved ECF groups (ECF22, ECF42 and ECF43) were found to lack a recognizable anti-σ factor but instead seem to be co-expressed with neighboring genes encoding serine/threonine protein kinases (Staron et al., 2009; Jogler et al., 2012). These proteins are predominantly found in eukaryotic signal transducing cascade, but are also involved in regulating a number of processes in bacteria through phosphorylation of downstream regulators or enzymes (Dworkin, 2015). The conserved genomic context for these ECF groups indicates that serine/threonine 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. 1.4 G**). Indeed, such a mechanism was most recently described for a member of group ECF43 from *Xanthomonas citri* (Bayer-Santos et al., 2018).

1.5.7 Regulation of ECF σ factor activity by its C-terminal extension

In contrast to the classical ECF σ factors, which only harbor the σ^2 and σ^4 domains, a number of ECF σ factor groups (ECF41, ECF42, ECF44, ECF45, ECF48, ECF52, ECF53, ECF56 and ECF57) contain an additional C-terminal extension (Pinto and Mascher, 2016b) (**Fig. 1.4 E**). It has been demonstrated that small truncations of the C-terminal extension of ECF41 from *Bacillus licheniformis* increases its activity, suggesting that the C-terminal extension of ECF41 play an anti-σ factor-like role on the activity of ECF41. However, truncation of the whole C-terminal extension results in the complete loss of ECF41 activity, which contrarily indicates that the C-terminal extension is necessary for the activity of ECF41 (Wecke et al., 2012). In the case of CorE-like proteins (group ECF44), it has been demonstrated that binding of Cu^{2+} or other divalent metal ions to the Cys-rich domain (CRD) of the C-terminal extension of ECF44 is essential for its activity, while binding of Cu^{1+} inhibits the σ factor (Marcos-Torres et al., 2016). These studies suggest that such a C-

terminal extension may play a dual role, contributing both for the activation and inhibition of ECF σ factors. However, the exact regulatory mechanism of the C-terminal extension is still poorly understood.

1.6 *Streptomyces venezuelae* and its ECF σ factors

Streptomycetes are high GC Gram-positive bacteria predominantly found in soil. *Streptomyces* have received tremendous scientific attention since these microorganisms are characterized by a complex secondary metabolism, providing over two-thirds of the clinically useful antibiotics of natural origin (Liu et al., 2013). In addition, *Streptomyces* is also characterized by remarkably complex developmental features, which usually grows as branching hyphal filaments to form a mat of fungus-like mycelium, from which emerge serial branches that bear chains of spores (**Fig. 1.5**). In brief, a germ tube emerges from the spore and grows by tip extension and branch formation to form the substrate mycelium. In response to nutrient depletion and other signals, aerial hypha breaks the surface tension and grow into the air to form the aerial hypha, which will further extend to a long chain that contains many tens of nucleoids. When aerial hypha growth stops, multiple septa subdivide the apical compartment into single-nucleoid pre-spore compartments. After complex morphological and metabolic processes such as remodeling and thickening of the cell wall, condensation of the chromosome and production of the spore pigment, the pre-spore compartments develop into mature spores. The life cycle of the most widely studied model organism of *Streptomyces*, *S. coelicolor*, is shown in **Fig. 1.5**.

It is the complexity of their life cycle and the environment they live in that requires *Streptomyces* requires multitudinous signaling devices such as ECF σ factors. Two species of *Streptomyces* – *S. coelicolor* and *S. griseusi* (Worthen, 2008) are mostly chosen organisms for investigation of genetics and cell biology of *Streptomyces*. Besides, *S. venezuelae*, the producer of chloramphenicol, has recently been also established as a new model organism for studying Streptomyces development (Bibb et al., 2012), since both *S. coelicolor* and *S. griseusi* fail to produce spores in liquid medium, the aerial hyphae constitute only about 10% of the total biomass and cannot be separated from the non-differentiating vegetative mycelium. By contrast, *S. venezuelae* sporulates rapidly, synchronously and comprehensively in liquid culture (Glazebrook et al., 1990), which makes *S. venezuelae*

especially suitable for sensitive biochemical, cytological and molecular studies of developmental states.

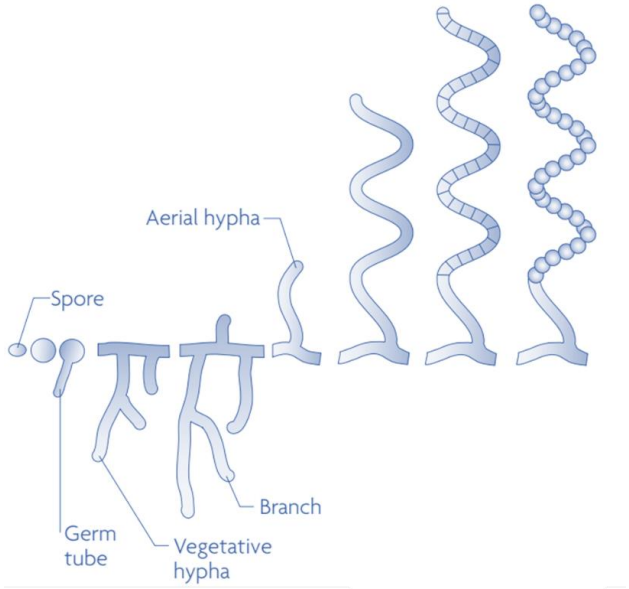


Figure 1. 5 Development life cycle of *S. coelicolor*. Under favorable conditions, the germ tube emerges from the spore and grows by tip extension and branch formation to form the vegetative hypha. In response to nutrient depletion and other signals, aerial hypha breaks the surface tension and grow into the air to form the aerial hypha. The aerial hypha further extends to a long chain that contains many tens of nucleoids. When the aerial hypha stops growing, multiple septa subdivide the apical compartment into single-nucleoid pre-spore compartments. After complex morphological and metabolic processes such as remodeling and thickening of the cell wall, condensation of the chromosome and production of the spore pigment, the pre-spore compartments develop into mature spores. Figure from (Flardh and Buttner, 2009)

The genome sequence of *S. venezuelae* ATCC10712 was released in 2011 (GenBank Accession No. NC_018750). It has a size of 8.2 Mb and encodes 7318 proteins, including 40 ECF σ factors from 23 distinct ECF groups (**Table 1.2**). This is far more than the 6 ECF σ factors per genome that are found on average in bacteria (Staron et al., 2009). This high number of ECF σ factors in the streptomycetes in general may reflect its rapid and precise regulation of intricate development process and diverse stress responses for adaptation in complex environments. To date, only ECF121 σ^{BldN} (Sven_3158) has been investigated in *S. venezuelae* and it has been shown to play a pivotal role in the developmental cascade that directly activates expression of the chaplin and rodlin genes encoding the hydrophobic sheath proteins of *S. venezuelae* (Bibb et al., 2012). The functions of the rest ECF σ factors investigated in other species of *Streptomyces* are listed in **Table 1.2**, which may provide useful information for future research on *S. venezuelae*.

Table 1. 2 ECF σ factors in *S. venezuelae* and other streptomycetes.

Locus	ECF group	Functions in *Streptomyces genus*	Name	References
SVEN_4513	ECF02			
SVEN_4870	ECF12	Global regulator of redox homeostasis in *S. coelicolor*	SigR	(Feeney et al., 2017)
SVEN_4793	ECF14			
SVEN_0063	ECF17	Protein secretion in *S. coelicolor*	SigU	(Gordon et al., 2008)
SVEN_0399	ECF19			
SVEN_6501	ECF20			
SVEN_3668	ECF27	Actinorhodin production in *S. coelicolor*	SigT	(Feng et al., 2011)
SVEN_6961	ECF34			
SVEN_0176	ECF36			
SVEN_2914	ECF38			
SVEN_3369	ECF38			
SVEN_6611	ECF38			
SVEN_3215	ECF39	Cell envelope stress response in *S. coelicolor*; Biosynthesis of avermectin and oligomycin in *S. avermitilis*	SigE Sig25	(Paget et al., 1999b) (Luo et al., 2014)
SVEN_3278	ECF39			
SVEN_3293	ECF39			
SVEN_3759	ECF39			
SVEN_4575	ECF39			
SVEN_4454	ECF40			
SVEN_0136	ECF41	Negative regulator of avermectin production in *S. avermitilis*	Sig6	(Jiang et al., 2011)
SVEN_0858	ECF41			
SVEN_3295	ECF41			
SVEN_3475	ECF41			
SVEN_3480	ECF41			
SVEN_3821	ECF41			
SVEN_3859	ECF41			
SVEN_1176	ECF41			
SVEN_0747	ECF42			
SVEN_4377	ECF42			
SVEN_7131	ECF42			
SVEN_0980	ECF50			
SVEN_0015	ECF51			
SVEN_3871	ECF52	Secondary metabolism and morphogenesis in *S. coelicolor*	Sig52	(Lopez-Garcia et al., 2018)
SVEN_0434	ECF53			
SVEN_6745	ECF53			
SVEN_4562	ECF56			
SVEN_4974	ECF118			
SVEN_3185	ECF121	Cell development and morphogenesis in *S. venezuelae*	SigBldN	(Bibb et al., 2012)
SVEN_4540	ECF123			
SVEN_4229	ECF126			
SVEN_4487	Ungrouped			

1.7 Aims of this study

This study focuses on a novel group of ECF σ factors, ECF42, which is characterized by a long C-terminal extension. These extensions contain a tetratricopeptide repeat (TPR) domain (Staron et al., 2009), which is postulated to be important for protein-protein interaction (D'Andrea and Regan, 2003). Genes encoding ECF42 proteins are not genomically associated with obvious anti-σ factor-encoding genes. Instead, the large majority is associated with genes encoding DGPF proteins of unknown function. So far, none of the σ factors belonging to ECF42 has been experimentally studied, with the exception of ECF-10 from *Pseudomonas putida*, which was found to be involved in stress resistance and biofilm formation (Tettmann et al., 2014). ECF42 proteins are particularly widely distributed in the *Actinobacteria* and the genome of *S. venezuelae* encodes three σ factors of this group (**Table 1.1**).

The aims and approaches of my thesis are summarized below:

Aim 1 (Target promoter determination): ECF σ factors perform their functions by regulating transcription of their target genes, which normally harbor an ECF group- specific target promoter. Identification of ECF target promoters enables the prediction of their target regulons, thereby providing a direct access to the physiological role of ECF-dependent regulation. As part of this thesis, the target promoter of ECF42 shall be predicted by bioinformatics analysis and subsequently experimentally verified.

Aim 2 (Physiological roles): The phylogenetic distribution analysis shows that ECF42 σ factors are highly abundant in *Actinobacteria*, especially in the genus *Streptomyces***.** The physiological role of ECF42 in *S. venezuelae* shall be investigated by phenotypically screening ECF42-mutants and by defining the target regulons of ECF42 σ factors.

Aim 3 (Regulatory mechanism): ECF42 genes lack obvious anti-σ factor-encoding genes in their vicinity, as do classical ECF σ factors, but instead, are associated with genes encoding DGPF proteins. Additionally, ECF42 σ factor harbor a conserved C-terminal extension domain (TPR), which may be involved in mediating protein-protein interactions. Thus, the regulatory role of the DGPF protein and the C-terminal extension on the activity of ECF42 σ factor shall be investigated in this study by *in vivo* and *in vitro* approaches, in order to unravel the signaling mechanism of ECF42 σ factors.

2. Materials and methods

2.1 Bioinformatics analysis

2.1.1 Phylogenetic distribution of ECF42 σ factors

The protein sequence of ECF42 σ factor from *S. venezuelae* (Sven_4377) was submitted to NCBI blastp and run against the non-redundant protein sequences database. The sequences of the complete 10,010 hits were extracted in December 2017. False positives (*i.e.*, proteins that did not belong to the ECF42 group) and proteins from more than one sequenced strain per species were removed leaving 2661 protein sequences for further analysis. Multiple sequence alignments were performed using ClustalW (Thompson et al., 1994) and the phylogenetic tree was generated from the gapless multiple alignments using the Neighbor-Joining method and Jukes-Cantor protein distance model, as implemented in CLC Main Workbench (Qiagen).

2.1.2 Genomic context analysis of ECF42 encoding genes

Genomic context analysis of the ECF42 family (COG4941) was performed by applying the multidendrogram approach to create the context tree using the database MicrobesOnline (Alm et al., 2005) at http://www.microbesonline.org/.

2.1.3 Domain architecture analysis of ECF42 σ factors

Protein domain architecture and alignment of ECF42 σ factors with classical ECF σ factors was analyzed using the DNAMAN software package (Lynnon BioSoft, Vaudreuil, Quebec, Canada). Classical ECF σ factors were included: RpoE from *Streptomyces coelicolor A3(2)* and *Pseudomonas putida KT2440*; SigE from *Streptomyces avermitilis* and *S. coelicolor A3(2)*; SigM from *S. coelicolor A3(2)*; SigL from *Mycobacterium bovis* and *S. avermitilis*; SigK from *Mycobacterium sp. JLS* and *S. avermitilis*. The protein sequences of ECF42 σ factors were extracted from *Xanthomonas campestris pv. Campestris, Rhodopirellula baltica, Bacillus cereus, Planctopirus limnophila, S. coelicolor A3(2), S. avermitilis MA*,

Streptomyces lydicus, *Streptomyces griseus*, *Streptomyces albus* and *Streptomyces venezuelae*.

2.1.4 Identification of the putative target promoter motif of ECF42 σ factors

Eighteen operons encoding ECF42 σ factors and DGPF proteins were selected and the 250 base pairs upstream of the starting codon of the first gene of the operon were extracted for promoter motif analysis. The presence of a putative promoter motif was investigated with the MEME motif discovery tool of the MEME Suite (Bailey et al., 2009), which is available at http://meme-suite.org/. Settings are listed in **Table 2.1**.

Table 2. 1 Settings used in MEME discovery tool for the identification of ECF42 putative target promoter motifs.

Parameter	Setting
Select the motif discovery mode:	Normal model
Select the sequence alphabet:	DNA, RNA or Protein
Select the site distribution:	Zero or one occurrence per sequence
Select the number of motifs:	3
How wide can motif be:	60 to 50
Can motif sites be on both strands:	Search given strand only

2.2 Medium, supplements and antibiotics

2.2.1 Medium and supplements

Lysogeny broth (LB) broth:

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
H_2O	Ad 1 L
(Agar)	15 g (for LB agar plates)

MYM:

Maltose	4 g
Yeast extract	4 g
Malt extract	10 g
H_2O	Ad 1 L
Agar	20 g (for MYM agar plates)

R2 trace elements (500x):

$ZnCl_2$	20 g
$FeCl_3 \cdot 6H_2O$	100 g
$CuCl_2 \cdot 2H_2O$	5 g
$MnCl_2 \cdot 4H_2O$	5 g
$Na_2B_4O_7 \cdot 10H_2O$	5 g
$(NH_4)_6Mo_7O_{24} \cdot 6H_2O$	5 g
H_2O	Add to 1 L

(When needed 2 ml of R2 trace elements was added per liter of MYM medium)

MS (SFM):

Mannitol	20 g
Soya flour	4 g
Malt extract	10 g
Tap H_2O	Ad 1 L
Agar	20 g (for MYM agar plates)

III'-Salts:

$MnSO_4 \cdot 4\ H_2O$	0.232 g
$MgSO_4 \cdot 7\ H_2O$	12.3 g
H_2O	Add to 1 L

10 x MOPS solution (1 L), adjusted to pH 7 with 10 M KOH:

MOPS	83.72 g
KH_2PO_4 (1M)	3.85 ml
K_2HPO_4 (1M)	6.15 ml
$(NH_4)_2SO_4$	33 g
H_2O	Add to 1 L

MOPS-based chemically defined medium (MCSE):

10× MOPS solution	10 ml
Tryptophan (5 mg/ml)	1 ml
Ammonium ferric citrate (2,2 mg/ml)	1 ml
III'-Salts	1 ml
Potassium glutamate (40%)	2 ml
Sodium succinate (30%)	2 ml
Fructose (20%)	1 ml
H_2O	Add to 100 ml

10 x MN-Medium:

$K_2HPO_4 \cdot 3H_2O$	136 g
KH_2PO_4	60 g
$Na_3C_6H_5O_7 \cdot 2\ H_2O$	10 g
H_2O	Add to 1 L

MNGE-Medium (10 ml):

1 x MN-Medium in H2O	9.2 ml
Glucose (20%)	1 ml
KC5H9NO4 (40%)	50 µl
C6H8O7·Fe·NH3 (2,2 mg/ml)	50 µl
Tryptophan (5 mg/ml)	100 µl
MgSO4 (1M)	30 µl

Expression Mix:

Yeast extract (5%)	500 µl
Casamino-acids (CAA) (10%)	250 µl
Tryptophan (5 mg/ml)	50 µl
H_2O	250 µl

BioLog metal ion cocktail:

$ZnCl_2 \cdot 7H_2O$	68 mg
$FeCl_2 \cdot 6H_2O$	135 mg
$MnCl_2 \cdot 4H_2O$	99 mg
$CaCl_2 \cdot 2H_2O$	74 mg
H_2O	Add to 100 ml

BioLog inoculating fluid:

IF-0a medium (Biolog):	6 ml
Dye mix D, G (Biolog):	72 µl
Glucose (500 mM):	72 µl
Metal ion cocktail:	72 µl
Cell suspension:	696 µl
H_2O:	288 µl

2.2.2 Antibiotics used for bacterial selection

Antibiotics used for selection of mutant strains are listed in **Table 2.2**.

Table 2. 2 Antibiotics used for selection of mutant strains.

Strain	Antibiotic	Working concentration (μg/ml)
B. subtilis	Chloramphenicol	5
	MLS selection:	
	Erythromycin	1
	Lincomycin	25
E. coli	Ampicillin	100
	Kanamycin	50
	Chloramphenicol	50
	Apramycin	50
	Hygromycin	50
S. venezuelae	Apramycin	50
	Thiostrepton	25
	Hygromycin	50
	Nalidixic acid	25

2.3 Bacterial strains, plasmids and oligonucleotides

Bacterial strains used in this study are listed in **Table 2.3.** Used vectors and generated plasmids are listed in **Table 2.4.** Oligonucleotides used in this study are listed in appendices **Table A1**.

Table 2. 3 Strains used in this study.

Strains	Genotype/Description	Reference
E. coli		
DH10ß	F-, *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX*74 *recA1 endA1 araD139* Δ(*ara leu*) 7697 *galU galK rpsL nupG* λ–	Invitrogen
ET12567	F-, *dam*-13::Tn9, *dcm*-6, *hsdM*, *hsdR*	(MacNeil et al., 1992)
ET12567/pUZ8002	*E. coli* ET12567 harboring pUZ8002, a not self-transmissible plasmid which can mobilize oriT-containing plasmids by conjugation	(Flett et al., 1997)
MK01	F-, Δ(*araD-araB*)567, Δ*lacZ*4787(::rrnB-3), λ-, Δ(*araH-araF*)570(::FRT), Δ*araE*p532::FRT, φPcp8-*araE*535, *rph-1*, Δ(*rhaD-rhaB*)568, *hsdR*514, *lacI*::*cat*	(Kogenaru and Tans, 2014)
BL21 (DE3) pLys	F-, *omp*T *hsd*SB (rB–, mB–) *gal dcm* (DE3) pLysS (Cam^R)	Invitrogen
BTH101	F-, *cya-99*, *araD139*, *galE15*, *galK16*, *rpsL1 (Str r)*, *hsdR2*, *mcrA1*, *mcrB1*	(Karimova et al.
TME3007	MK01 Level_M_QL001 (*3xflag_ ecf42_4454* + P_{ecf42_4454}*_luxCDABE*)	This work

Table 2.3 Strains used in this study (continued).

Strains	Genotype/Description	Reference
E. coli		
TME3008	MK01 Level_M_QL004 (*3xflag _ecf42_4454* M01 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3009	MK01 Level_M_QL005 (*3xflag _ecf42_4454* M02 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3010	MK01 Level_M_QL006 (*3xflag _ecf42_4454* M03 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3011	MK01 Level_M_QL007 (*3xflag _ecf42_4454* M04 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3012	MK01 Level_M_QL008 (*3xflag _ecf42_4454* M05 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3013	MK01 Level_M_QL010 (*3xflag _ecf42_4454* T1 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3014	MK01 Level_M_QL010 (*3xflag _ecf42_4454* T2 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3015	MK01 Level_M_QL010 (*3xflag _ecf42_4454* T3 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3016	MK01 Level_M_QL010 (*3xflag _ecf42_4454* T4 + P_{ecf42_4454}_*luxCDABE*)	This work
TME3017	MK01 Level_M_QL004 (*3xflag _ecf42_4454* M07 + P_{ecf42_4454}_*luxCDABE*)	This work
S. venezuelae		
TMS0001	*S. venezuelae* ATCC10712 wild type	Lab strain
TMS0039	Δ*sven_7131*	This work
TMS0044	Δ*sven_7131 Δsven_4377*	This work
TMS0050	Δ*sven_0747* Δ*sven_4377* Δ*sven_7131*	This work
TMS0112	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_0747*	This work
TMS0113	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_4377*	This work
TMS0114	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_7131*	This work
TMS0174	*attBϕC31*::pGUS	This work
TMS0175	Δsven_0747 Δ*sven_4377*Δ*sven_7131 attBϕC31*::pGUS	This work
TMS0176	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_0747 attBϕC31*::pGUS	This work
TMS0177	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_4377* attBϕC31::pGUS	This work
TMS0178	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_7131 attBϕC31*::pGUS	This work
TMS0179	*attBϕC31*::pGUS_*NP_0747*	This work
TMS0180	Δ*sven_0747* Δ*sven_4377* Δ*sven_7131* *attBϕC31*::pGUS_*NP_0747*	This work
TMS0181	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_0747 attBϕC31*::pGUS_*NP_0747*	This work
TMS0182	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_4377 attBϕC31*::pGUS_*NP_0747*	This work
TMS0183	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_7131 attBϕC31*::pGUS_*NP_0747*	This work
TMS0184	ATCC10712 *attBϕC31*::pGUS_*NP_4377*	This work
TMS0185	Δ*sven_0747* Δ*sven_4377* Δ*sven_7131 attBϕC31*::pGUS_*NP_4377*	This work
TMS0186	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257_*N_3xflag_sven_0747 attBϕC31*::pGUS_*NP_4377*	This work

Table 2.3 Strains used in this study (continued).

Strains	Genotype/Description	Reference
S. venezuelae		
TMS0187	*Δsven_0747* Δ*sven_4377* Δ*sven_7131* *attBϕBT1*::pIJ10257*_N_3xflag_sven_4377* *attBϕC31*::pGUS*_NP_4377*	This work
TMS0188	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257*_N_3xflag_sven_7131* attBϕC31::pGUS*_NP_4377*	This work
TMS0189	*attBϕC31*::pGUS*_NP_7131*	This work
TMS0190	Δ*sven_0747* Δ*sven_4377* Δ*sven_7131* *attBϕC31*::pGUS*_NP_7131*	This work
TMS0191	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257*_N_3xflag_sven_0747 attBϕC31*::pGUS*_NP_7131*	This work
TMS0192	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257*_N_3xflag_sven_4377* *attBϕC31*::pGUS*_NP_7131*	This work
TMS0193	Δ*sven_0747* Δ*sven_4377*Δ*sven_7131* *attBϕBT1*::pIJ10257*_N_3xflag_sven_7131 attBϕC31*::pGUS*_NP_7131*	This work
TMS0036	Δ*sven_4377*::*apr* pGUS *_NP_4377*	This work
TMS0051	Δ*sven_4377*::*apr* attBϕC31::pGUS *_NP_4377 attBϕBT1*::pIJ10257*_3xflag_ sven_4377_T1* (1-179aa)	This work
TMS0052	Δ*sven_4377*::*apr attBϕC31*::pGUS _NP_4377 *attBϕBT1*::pIJ10257*_3xflag_ sven_4377_T2* (1-230aa)	This work
TMS0053	Δ*sven_4377*::*apr attBϕC31*::pGUS _NP_4377 *attBϕBT1*::pIJ10257*_3xflag_ sven_4377_T3* (1-287aa)	This work
TMS0054	Δ*sven_4377*::*apr attBϕC31*::pGUS _NP_4377 *attBϕBT1*::pIJ10257*_3xflag_sven_4377_T4* (1-353aa)	This work
***B. subtilis* strains**		
W168	*B. subtilis* wild-type *trpC*	Lab strain
1A774	JH642 *rpoC*::(6xHis-tag) $Spec^R$	BGSC (C. Moran)
TMB2572	W168 *sacA*::pQL3Clux01 (pBS3Clux-*NP_sven_0747*)	This work
TMB2573	W168 *sacA*::pQL3Clux01 (pBS3Clux-*AP_sven_0747*)	This work
TMB2574	W168 *sacA*::pQL3Clux01 (pBS3Clux-*NP_sven_4377*)	This work
TMB2575	W168 *sacA*::pQL3Clux01 (pBS3Clux-*AP_sven_4377*)	This work
TMB2576	W168 *sacA*::pQL3Clux01 (pBS3Clux-*NP_sven_7131*)	This work
TMB2577	W168 *sacA*::pQL3Clux01 (pBS3Clux-*AP_sven_7131*)	This work
TMB2650	W168 *lacA*::erm P_{xylA}*_3xflag-sven_4377 sacA::cat NP_sven_4377 luxABCDE*	This work
TMB2653	W168 *lacA*::erm P_{xylA}*_3xflag-sven_4377 sacA::cat NP_sven_7131 luxABCDE*	This work
TMB2656	W168 *lacA*::erm P_{xylA}*_3xflag-sven_4377 sacA::cat AP_sven_4377 luxABCDE*	This work
TMB2659	W168 *lacA*::erm P_{xylA}*_3xflag-sven_4377 sacA::cat AP_sven_7131 luxABCDE*	This work
TMB2727	1A774 *rpoC*::(6xHis-tag) $Spec^R$ *lacA*::P_{xylA}*_3xflag_ sven_4377*)	This work
TMB2728	1A774 *rpoC*::(6xHis-tag) $Spec^R$ *lacA*::P_{xylA}*_3xflag_ sven_7131*)	This work

Table 2. 4 Plasmids used in this study.

Plasmids	Description	Reference
Plasmids for *S. venezuelae*		
pIJ10257	Protein overexpression vector, *Hyg*^{R}, integrates into the ΦBT1 attachment site of *Streptomyces* spp. ermEp* promoter cloned into pMS82	(Hong et al., 2005)
pGUS	*Apr*^{R}, conjugative and Φ31-integrative vector, *gus* reporter gene without promoter	(Myronovskyi et al., 2011)
pIJ12738	pKC1132 with multiple cloning site and I-*Sce*I site from pUC57-Simple_*Sce*I	(Fernandez-Martinez and Bibb,
pIJ12742	pGM1190 with ermE*p-I-*Sce*I gene, *Thi*^{R}	(Fernandez-Martinez and Bibb,
pQLS001	pIJ12738 derivative carrying up- and downstream regions of *sven_0747*	This work
pQLS002	pIJ12738 derivative carrying up- and downstream regions of *sven_4377*	This work
pQLS003	pIJ12738 derivative carrying up- and downstream region of *sven_7131*	This work
pQLS004	pIJ10257 derivative carrying *3xflag_sven_0747*	This work
pQLS005	pIJ10257 derivative carrying *3xflag_sven_4377*	This work
pQLS006	pIJ10257 derivative carrying *3xflag_sven_7131*	This work
pQLS007	pGUS derivative carrying putative target promoter of *sven_0747*	This work
pQLS008	pGUS derivative carrying putative target promoter of *sven_4377*	This work
pQLS009	pGUS derivative carrying putative target promoter of *sven_7131*	This work
pQLS010	pIJ10257 derivative carrying *3xflag_sven_4377_T1* (1-537 bp)	This work
pQLS011	pIJ10257 derivative carrying *3xflag_sven_4377_T2* (1-690 bp)	This work
pQLS012	pIJ10257 derivative carrying *3xflag_sven_4377_T3* (1-861 bp)	This work
pQLS013	pIJ10257 derivative carrying *3xflag_sven_4377_T4* (1-1059 bp)	This work
Plasmids for *E. coli*		
pETDuet-1	Vector designed for co-expression of two target genes	Invitrogen
Level 0-1	MoClo cloning vector	(Werner et al., 2012)
Level 0-9	MoClo cloning vector	(Werner et al., 2012)
Level 0-15	MoClo cloning vector	(Werner et al.. 2012)
Level 0-1 L	MoClo cloning vector	(Werner et al., 2012)
Level 1-1 R	MoClo cloning vector	(Werner et al., 2012)
Level 0-11_001	0-11 derivative carrying terminator 01	(Pinto et al., 2018)
Level 0-11_002	0-11 derivative carrying terminator 02	(Pinto et al., 2018)
Level 1-3 L	MoClo cloning vector	(Pinto et al., 2018)
Level 1-2 L insulator	MoClo cloning vector	(Pinto et al., 2018)

Table 2.4 Plasmids used in this study (continued).

Plasmids	Description	Reference
Plasmids for *E. coli*		
Level M end-linker 03	MoClo cloning vector	(Pinto et al., 2018)
Level M medium copy	MoClo cloning vector	(Pinto et al., 2018)
Level 0-1_P_{Bad}	0-1 derivative carrying arabinose inducible promoter P_{Bad}	(Pinto et al., 2018)
Level 0_1_P_{ecf42}	0-1 derivative carrying the target promoter of ECF42_4454	(Pinto et al., 2018)
Level 0-9_RBS	0-9 derivative carrying ribosome binding site	(Pinto et al., 2018)
Level 0-15_lux	0-15 derivative carrying *luxCDABE* operon	(Pinto et al., 2018)
Level 0-15 ECF42_4454	0-15 derivative carrying *ecf42_4454* gene (Locus tag: XCC3792)	(Pinto et al., 2018)
pQLE001	0-15 derivative carrying *3xflag_ecf42_4454* gene	This work
pQLE002	0-15 derivative carrying *3xflag* truncated *ecf42_4454* gene T1 (1-180aa)	This work
pQLE003	0-15 derivative carrying *3xflag* truncated *ecf42_4454* gene T2 (1-232aa)	This work
pQLE004	0-15 derivative carrying *3xflag* truncated *ecf42_4454* gene T3 (1-294aa)	This work
pQLE005	0-15 derivative carrying *3xflag* truncated *ecf42_4454* gene T4 (1-355aa)	This work
pQLE006	0-15 derivative carrying gene coding for a mutated ECF42_4454 M01 (S191A)	This work
pQLE007	0-15 derivative carrying gene coding for a mutated ECF42_4454 M02 (V195A)	This work
pQLE008	0-15 derivative carrying gene coding for a mutated ECF42_4454 M03 (I199A)	This work
pQLE009	0-15 derivative carrying gene coding for a mutated ECF42_4454 M04 (N201A)	This work
pQLE010	0-15 derivative carrying gene coding for a mutated ECF42_4454 M05 (R214A)	This work
pQLE011	0-15 derivative carrying gene coding for a mutated ECF42_4454 M07 (S191A, V195A, I199A, N201A)	This work
pQLE012	1-1 R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454*	This work
pQLE013	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* T1 (1-182aa)	This work
pQLE014	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* T2 (1-235aa)	This work
pQLE015	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* T3 (1-295aa)	This work
pQLE016	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* T4 (1-370aa)	This work
pQLE017	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* M01 (S191A)	This work
pQLE018	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* M02 (V195A)	This work
pQLE019	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* M03 (I199A)	This work
pQLE020	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* M04 (N201A)	This work
pQLE021	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* M05 (R214A)	This work
pQLE022	1-1R derivative carrying P_{BAD} promoter driving the expression of *3xflag ecf42_4454* M07 (S191A, V195A, I199A, N201A)	This work

Table 2.4 Plasmids used in this study (continued).

Plasmids	Description	Reference
Plasmids for *E. coli*		
pQLE023	1-1R derivative carrying P_{ecf42_4454} promoter driving *luxCDABE*	This work
pQLE024	Level M derivative carrying P_{BAD}_ *ecf42_4454*_T1 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE025	Level M derivative carrying P_{BAD}_ *ecf42_4454*_T1 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE026	Level M derivative carrying P_{BAD}_ *ecf42_4454*_T2 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE027	Level M derivative carrying P_{BAD}_ *ecf42_4454*_T3 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE028	Level M derivative carrying P_{BAD}_ *ecf42_4454*_T4 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE029	Level M derivative carrying P_{BAD}_ *ecf42_4454*_M01 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE030	Level M derivative carrying P_{BAD}_ *ecf42_4454*_M02 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE031	Level M derivative carrying P_{BAD}_ *ecf42_4454*_M03 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE032	Level M derivative carrying P_{BAD}_ *ecf42_4454*_M04 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE033	Level M derivative carrying P_{BAD}_ *ecf42_4454*_M05 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE034	Level M derivative carrying P_{BAD}_ *ecf42_4454*_M07 and P_{ecf42_4454}_*luxCDABE*	This work
pQLE035	pUT18C derivative carrying *sven_4376*	This work
pQLE036	pUT18C derivative carrying *sven_7130*	This work
pQLE037	pUT18C derivative carrying *sven_7131*	This work
pQLE038	pKT25 derivative carrying *sven_4376*	This work
pQLE039	pKT25 derivative carrying *sven_4377*	This work
pQLE040	pKT25 derivative carrying *sven_7130*	This work
pQLE041	pKT25 derivative carrying *sven_7131*	This work
pQLE042	pKT25 derivative carrying *ecf42_4454* sigma region (1-543 bp)	This work
pQLE043	pKNT25 derivative carrying *ecf42_4454* sigma region (1-543 bp)	This work
pQLE044	pUT18 derivative carrying *ecf42_4454* sigma region (1-543 bp)	This work
pQLE045	pUT18C derivative carrying *ecf42_4454* sigma region (1-543 bp)	This work
pQLE046	pKT25 derivative carrying *ecf42_4454* C-terminal extension (544-1269 bp)	This work
pQLE047	pKNT25derivative carrying *ecf42_4454* C-terminal extension (544-1269 bp)	This work
pQLE048	pUT18 derivative carrying *ecf42_4454* C-terminal extension (544-1269 bp)	This work
pQLE049	pUT18C derivative carrying *ecf42_4454* C-terminal extension (544-1269 bp)	This work
pQLE050	0-15 derivative carrying gene encoding DGPF protein (Locus tag: XCC3790) protein	This work
pQLE051	1-1R derivative carrying P_{Tet} promoter driving the expression of *3xflag_DGPF*	This work
pQLE052	pETDuet-1 derivative carrying *ecf42_4454* σ region (1 to 543 bp)	This work
pQLE053	pETDuet-1 derivative carrying *ecf42_4454* C-terminal extension (544-1269 bp)	This work

Table 2.4 Plasmids used in this study (continued).

Plasmids	Description	Reference
Plasmids for *E. coli*		
pQLE054	pETDuet-1 derivative carrying *ecf42_4454* σ region (1 to 543 bp) and C-terminal extension (544 to 1269 bp) for separately expression	This work
pQLE055	pETDuet-1 derivative carrying gene encoding RibD (Sven_1041)	This work
pQLE056	pETDuet-1 derivative carrying gene encoding DGPF (Sven_3806)	This work
pQLE057	pETDuet-1 derivative carrying genes encoding RibD and DGPF for co-expression	This work
Plasmids for *B. subtilis*		
pBS3Clux	*B. subtilis* BioBrick vector	(Radeck et al., 2013)
pBS2E_P_{xylA}	Integration at *lacA*, amp^R, mls^R, P_{xylA}	Daniela Pinto, unpublished
pQL3Clux01	pBS3Clux-*NP_sven_0747*	This work
pQL3Clux02	pBS3Clux-*AP_sven_0747*	This work
pQL3Clux03	pBS3Clux-*NP_sven_4377*	This work
pQL3Clux04	pBS3Clux-*AP_sven_0747*	This work
pQL3Clux05	pBS3Clux-*NP_sven_7131*	This work
pQL3Clux06	pBS3Clux-*AP_sven_7131*	This work
pQL2E-Pxyl 02	pBS2E_P_{xylA}_*3xflag-tag_sven_4377*	This work
pQL2E-Pxyl 05	pBS2E_P_{xylA}_*3xflag-tag_sven_7131*	This work

2.4 Nucleic acid manipulations

2.4.1 Synthesis of genes and oligonucleotides

Oligonucleotides were synthesized by Sigma (International AG, Germany). The lyophilized pellets were resuspended to a final concentration of 100 μM in H_2O and stored at -20 °C. Genes for heterologous expression were codon optimized and synthesized through the GeneArt Strings DNA fragments service from ThermoFisher (International AG, Germany).

2.4.2 Polymerase chain reaction (PCR)

PCR reactions were performed by using DNA polymerases from New England Biolabs (NEB) according to the manufacturer's instructions. OneTaq DNA Polymerase (M0480) was

used for confirming insertions, deletions or colony PCR, Q5 High-Fidelity DNA Polymerases (M0491) was used for creating fragments for cloning.

2.4.3 Agarose gel electrophoresis of DNA

Electrophoresis was performed using the 1% agarose gel containing ethidium bromide (Carl Roth) in 1 × TAE buffer (40 mM Tris, 20 mM Acetic acid, 1 mM EDTA). Samples were mixed with 6 × loading dye prior to loading. The voltage used for electrophoresis was 6 volts per cm.

2.4.4 Restriction endonuclease digestion

The restriction endonuclease digestion of DNA fragments and plasmids was performed by using the restriction enzymes from NEB according to their manufacturer's instructions.

2.4.5 Ligation of DNA fragments

The ligation of DNA fragments was performed by using the T4 DNA ligase (M0202) from NEB according to the recommended protocol and the reaction was incubated for 8 hours at 16 °C.

2.4.6 Purification of DNA fragments and preparation of plasmids

The DNA fragments from PCR products or DNA agarose gel were cleaned and concentrated by using the Hi-Yield Gel/PCR DNA Fragment Extraction Kits from Süd-Laborbedarf GmbH (SLG, Germany) according to the manuals. The plasmids extraction from *E. coli* was carried out by using the Hi-Yield Plasmid Mini DNA Isolation Kit (SLG, Germany) according to the manufacturer's instruction.

2.4.7 DNA sequencing

Sequencing of DNA fragments or plasmids was performed by Eurofins Genomics (Germany) using the Mix2Seq Kits according to the manufacturer's protocol.

2.4.8 Fusion PCR for the site mutagenesis of ECF42 σ factor

The single or multiple amino acids mutation at the C-terminal extension of ECF42 σ factor (Locus tag: XCC3792) from *Xanthomonas campestris* were generated by fusion PCR. A DNA fragment encoding N-terminus of ECF42_4454 was amplified from plasmid Level 0_15_ECF42_4454 by using primers TM5670 and TM5671. The DNA fragments encoding different mutated versions of the C-terminal extension were amplified from plasmid Level 0-15 _ECF42_4454 by using primers TM5672 and TM5620 for M01 (S191A), TM5673 and TM5620 for M02 (V195A), TM5674 and TM5620 for M03 (I199A) and TM5675 and TM5620 for M04 (N201A). Then, DNA fragments encoding the N-terminus and the C-terminal extension were joined together by fusion PCR using Q5 High-Fidelity DNA Polymerases to generate different full *ecf42_4454* genes encoding different mutated ECF42 σ factors M01 (S191A), M02 (V195A), M03 (I199A) and M04 (N201A). For M05 (R214A), The N-terminal σ and the C-terminal extension encoding regions were amplified from plasmid Level 0-15_*ecf42_4454* using the primer pairs TM5670-TM5676, TM5677-TM5620, respectively. For M07 (S191A, V195A, I199A, N201A), fragments encoding the N-terminal σ regions and the C-terminal extensions were amplified from plasmid pQLE010 using the primer pairs TM5670-TM5856 and TM5857-TM5620, respectively. All fused fragments of mutated ECF42 coding genes were cloned into MoClo vector 0-15 using the MoClo cloning strategy.

2.4.9 MoClo cloning assemble

For the analysis of the interaction between the N-terminal σ domain and C-terminal extension of the ECF42 σ factor (XCC3792) from *Xanthomonas campestris*, all genetic constructs (**Table 2.4**) were generated using the MoClo cloning method, which is based on the golden gate cloning strategy (Werner et al., 2012). The general assembly was performed accordingly to the following protocol: 50 ng of backbone vector, 50 ng of insert or insert-

containing plasmids, 1 µl of BbsI or BsaI, 1 µl of T4 DNA ligase and 2 µl 10 × T4 DNA ligase buffer were mixed in a final reaction volume of 20 µl. The mixture was incubated at 37 °C for 3 hours; at 50 °C for 10 min and at 80 °C for 10 min. Then, 10 µl of the reaction mix was used for *E. coli* transformation. The scheme for DNA assembly from level 0 to level M constructs is illustrated (**Fig. 2.1**).

Level 0 plasmids containing the arabinose inducible promoter P_{BAD} (0-1_P_{BAD}), ECF42 target promoter (0-1_P_{ecf42_4454}), the anhydrotettracycline inducible promoter P_{Tet} (0-1_P_{Tet}), the RBS (0-9_RBS), the *luxCDABE* reporter cassette (0-15_*luxCDABE)* and the terminators 01, 02 and 03 (0-11_01, 0-11_02, and 0-11_03) (**Fig. 2.1 A**) were a kind gift from Georg Fritz. DNA fragments encoding different C-terminally truncated ECF42 alleles - T1 (1-180 aa), T2 (1-232 aa), T3 (1-294 aa) and T4 (1-355 aa) were amplified from plasmid pQLE001 by using the primer pairs TM5590 and TM5714, TM5590 and TM5715, TM5590 and TM5716, TM5590 and TM5717, respectively. The amplified fragments were cloned into 0-15 vector to generate the plasmids pQLE002, pQLE003, pQLE004 and pQLE005. The fragments encoding mutated *ecf42* genes (see section 2.4.8) encoding M01 (S191A), M02 (V195A), M03 (I199A), M04 (N201A), M05 (R214A) and M07 (S191A, V195A, I199A, N201A) were cloned into 0-15 vector to generate plasmids pQLE006, pQLE007, pQLE008, pQLE009, pQLE010 and pQLE011. The DGPF coding gene was synthesized and cloned into 0-15 vector to generate the plasmid pQLE050. A *3xflag* tag coding sequence was introduced at the N-terminal position of all C-terminally truncated ECF42 or DGPF proteins by PCR.

For level 1 plasmids (**Fig. 2.1 B)**, different truncated or mutated *ecf42* genes were placed under the control of the arabinose inducible promoter P_{BAD} in the MoClo vector 1-1R. The target promoter sequences of ECF42_4454 σ factor was placed upstream of the *luxCDABE* reporter in vector 1-3L. The gene encoding the DGPF protein (*XCC3790*) was placed under the control of the anhydrotettracycline inducible promoter P_{Tet} in the vector 1-5L to generate the plasmid pQLE051. The level M plasmids were generated by combining different transcriptional units from level 1 plasmids (**Fig. 2.1 C)**. Each transcriptional unit was insulated to avoid transcriptional interference. The generated level M plasmid was transformed into *E. coli* MK01 for further assessment.

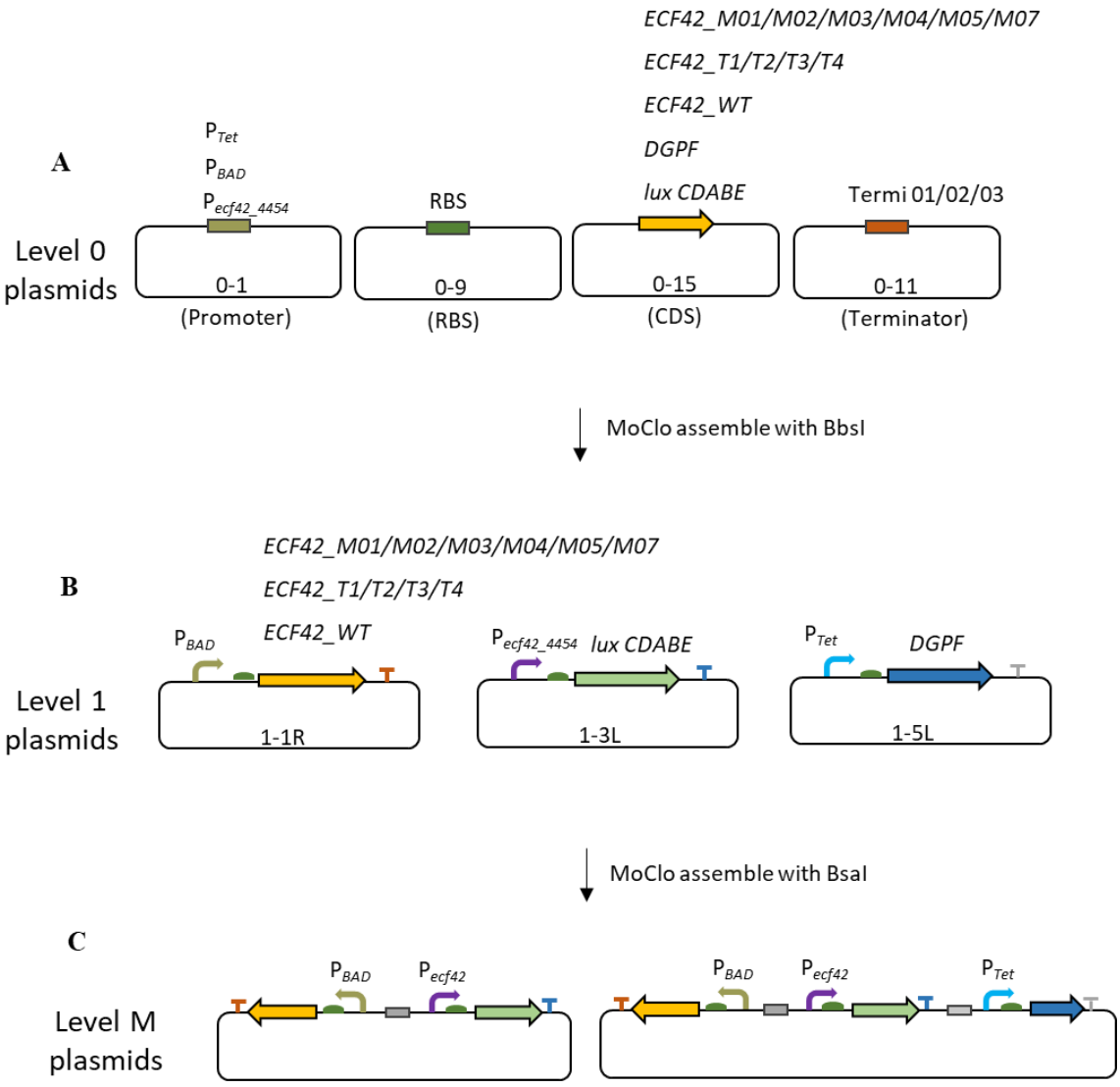


Figure 2. 1 Illustration of MoClo cloning and assemble strategy. (A) Level 0 plasmids. The anhydrotetracycline-inducible promoter P_{Tet}, the arabinose-inducible promoter P_{BAD} and the target promoter of ECF42 P_{ecf42_4454} were cloned into the 0-1 vector to generate the promoter- containing level 0-1 plasmids. The ribosome binding site (RBS) was cloned into the 0-9 plasmid. Genes encoding ECF42 σ factor wild type (WT), C-terminally truncated alleles (T1, T2, T3 and T4), point mutated ECF42 (M01, M02, M03, M04, M05 and M07) and the DGPF protein were cloned into the 0-15 plasmid to generate the coding sequence containing level 0-15 plasmids. Three different terminators (01, 02 and 03) were cloned into 0-11 vectors. **(B)** Level 1 plasmids. All fragments encoding ECF42 alleles and mutants were assembled in 1-1R vector under the control of the arabinose inducible promoter P_{BAD}. The *luxCDABE* reporter operon was assembled in the 1-3L vector under the control of the target promoter of the ECF42 σ factor. The gene encoding the DGPF protein was assembled in 1-5L vector under the control of the anhydrotetracycline inducible promoter P_{Tet}. **(C)** Level M plasmids. Transcriptional units containing each ECF42 version were assembled with the *luxCDABE* transcriptional unit or both *luxCDABE* and DGPF transcriptional units.

2.5 Protein methods

2.5.1 SDS PolyAcrylamide Gel Electrophoresis (SDS-PAGE)

SDS-polyacrylamide gels at a concentration of 12.5% were prepared by using the Mini-PROTEAN® Tetra Handcast System (Bio-Rad, USA) as described in **Table 2.5**. Electrophoresis was performed in 1 × SDS-PAGE running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS) at 220 V until the dye-front arrived at the bottom of the gel.

Table 2. 5 Composition of the SDS-PAGE gels.

For 2 separating gels:		For 2 stacking gels:	
30% acrylamide	4.2 ml	30% acrylamide	0.67 ml
1.5 M Tris, pH8.8	2.5 ml	1.5 M Tris, pH6.8	0.4 ml
10% SDS	100 µl	10% SDS	50 µl
dH_2O	3.2 ml	dH_2O	3.85 ml
10% APS	75 µl	10% APS	50 µl
TEMED	10 µl	TEMED	8 µl

2.5.2 Coomassie staining

The proteins in the SDS-PAGE gels were stained by Coomassie Brilliant blue R-250 staining solution (45% methanol, 10% acetic acid, 0.25% Coomassie Brilliant Blue R-250) at room temperature with gentle shaking for 2 hours. The gel was destained by using the destaining solution (45% methanol, 10% acetic acid) for 30 minutes. The destaining solution was removed and new one was added three or four times until the bands of proteins were clearly distinguishable.

2.5.3 Overexpression of N-terminal σ region and C-terminal extension of an ECF42 σ factor in *E. coli*

The DNA fragment encoding the N-terminal sigma domain (181 aa) of ECF42 σ factor (XCC3792) was amplified from plasmid pQLE001 using primers TM6068 and TM6069 and cloned into vector pETDuet-1 with *Nco*I and *Bam*HI to generate the overexpression plasmid

pQLE052. The fragments encoding the C-terminal extension was amplified from plasmid pQLE001 with primers TM6070 and TM6071 and cloned into vector pETDuet-1 with *Nde*I *and Kpn*I to generate the overexpression plasmid pQLE053. Additionally, both fragments encoding the N- and C-terminal domains were simultaneously cloned into the positions 1 and 2 of pETDuet-1 to generate the co-expression plasmid pQLE054. For purification purposes of the N- and C-terminal regions of ECF42 σ factor, the N-terminal of ECF42 was tagged with a 10 × His tag at the N-terminus and the C-terminal extension of ECF42 was tagged with a Strep tag at the C-terminus.

Constructed overexpression plasmids were transformed into *E. coli* BL21 (DE3) pLysE to generate the overexpression strains. Day cultures of the overexpression strains were prepared by diluting the overnight cultures 1:1000 in fresh LB medium with antibiotics and incubated at 37 °C with agitation. Once the optical density at 600 nm (OD_{600}) of the culture reached 0.6, the production of the proteins was induced by addition of 0.5 mM IPTG and incubation was continued at 37 °C for 3 hours. The cells were harvested by centrifugation and resuspended in lysis buffer (100 mM, Tris-HCl, pH8.0, 100 mM NaCl) before sonication with 50% power for 3 minutes with pulse durations of 5 sec on and 5 sec off (GM70 sonicator from BANDELIN, Germany). Expression and solubility of the proteins was checked by SDS-PAGE and Western blot.

2.5.4 Refolding of ECF42 fragments *in vitro*

Since both N-terminal 10×His_ and C-terminal_Strep-tagged ECF42 proteins in *E. coli* were expressed as insoluble inclusion bodies, they needed to be refolded *in vitro* before purification as described previously (Eiamphungporn and Helmann, 2008). The production of the N- and C-terminal domain of ECF42 σ factor was induced by 1 mM IPTG at 37 °C for 3 hours when the optical density at 600nm (OD_{600}) of the culture reached 0.6. Cells from 500 ml of culture were collected by centrifugation and resuspended in 20 ml of disruption buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.1 mM DTT, 1mM β-mercaptoethanol, 233 mM NaCl, 10% (v/v) glycerol] for sonication as described in section 2.5.3. The pellet was collected and washed twice with 10 ml TEDG buffer (10 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.1 mM DTT, 10% glycerol) containing 0.5% (v/v) Triton X-100 and was then dissolved in 10 ml of the same buffer with 1% (v/v) of Sarkosyl. After centrifugation to remove the insoluble fraction, the supernatant was gradually diluted (1 ml per min) with

TEDG buffer containing 0.01% Triton X-100 to a final volume of 100 ml, to allow renaturation of the N- or C-terminus of the ECF42 σ factor. Purification of the refolded proteins was carried out using the ÄKTA start protein purification system (GE Healthcare).

2.5.5 Purification of ECF42 σ factor fragments

HisTrap HP columns (5ml, GE Healthcare) was used for the purification of 10 × His_sigma region. Briefly, the solution of refolded N-terminal σ region was loaded onto the column equilibrated with TEDG-0.01% (v/v) Triton X-100. After washing with 50 ml of TEDG containing 0.15 M NaCl and 0.01% Triton X-100, 10 × His_sigma region was eluted using 20 ml elution buffer (20 mM HEPES pH8.0, 0.15 M NaCl, 1 mM β-mercaptoethanol, 10% glycerol) with a linear gradient concentration of imidazole from 20 mM to 500 mM. Fractions containing the target protein were collected and stored at -80 °C until use. Protein concentration was determined by using the Pierce™ BCA Protein Assay Kit (ThermoFisher, Germany).

Purification of the C-terminal extension_Strep-tag of ECF42 σ factor was performed by using the 1 ml Strep-Tactin XT Superflow cartridge (IBA, Germany) according to the manufacturer's instructions. Briefly, the refolded C-terminal extension_Strep-tag protein solution was loaded onto the column equilibrated with TEDG-0.01% (v/v) Triton X-100. After washing with 50 ml of TEDG containing 0.15 M NaCl and 0.01% Triton X-100, C-terminus_Strep-tag was eluted using 20 ml elution buffer (20 mM HEPES pH8.0, 150 mM NaCl, 1 mM EDTA, 50 mM biotin, 10% glycerol). Fractions containing the C-terminus_Strep-tag were collected and stored at -80 °C until use.

2.5.6 Pull-down assay of DGPF protein with RibD

Genes encoding RibD (*sven_1041*) and DGPF (*sven_3806*) from *S. venezuelae* were codon optimized for *E. coli* by using the GeneArt Strings DNA Fragments service from ThermoFisher. The DNA fragment encoding RibD was amplified using primers TM6134 and TM6135 and cloned into vector pETDuet-1 with *Nco*I and *Bam*HI to generate the plasmid pQLE055 and the fragment encoding DGPF protein was amplified using primers TM6136 and TM6137 and cloned into plasmid pQLE055 with *Nco*I and *Bam*HI to generate

the plasmid pQLE057. A 10×His and Strep tags were introduced into the N-terminus of RibD and C-terminus of DGPF, respectively. The plasmid pQLE057, co-expressing RibD and DGPF under the control of the IPTG inducible promoter, was transformed into *E. coli* BL21 (DE3) pLysE and the expression was induced by 0.4 mM IPTG at 25 °C for 4 hours, when the optical density at 600 nm (OD_{600}) of the culture reached 0.6. The cells were harvested by centrifugation and resuspended in lysis buffer (20 mM Tris pH 8.0, 5 mM EDTA, 50 mM NaCl, 1× EDTA-free protease inhibitors (Roche)] for sonication as described in section 2.5.3. The Strep tagged DGPF protein was purified from the cell lysate as described before, and the 10× His tagged RibD protein potentially interacted and co-purified with Strep tagged DGPF protein was confirmed by Western blot using an anti-His tag antibody.

2.5.7 Protein crosslinking

The interaction between the N-terminal σ region and the C-terminal extension of ECF42 was analyzed by crosslinking using glutaraldehyde as the crosslinking agent. Ten microgram of purified 10×His_σ region and equal molar amount of C-terminal extension_Strep-tag were mixed in the final volume of 50 µl 20 mM HEPES buffer (pH8.0). After incubated at room temperature for 5 min, the mixture was treated with 5 µl of 2.3% freshly prepared solution of glutaraldehyde and incubated at room temperature for 10 minutes. The reaction was terminated by addition of 10 µl of 1 M Tris-HCl pH8.0. Then, 10 µl of 6× SDS protein loading buffer (ThermoFisher, Germany) was added into the reaction mixture and the cross-linked proteins were checked by electrophoresis conducted in 12% SDS-polyacrylamide gels. The alcohol dehydrogenase from *Thauera aromatic* (ThaADH), which is known to form homodimers, was used as positive control.

2.5.8 Immunoprecipitation of RNAP from *E. coli*

The RNAP immunoprecipitation (IP) was carried out to investigate the interaction of the RNAP with the C-terminally truncated ECF42-T1 (1-180 aa) in *E. coli* by using the anti-*E. coli* RNAP ImmunoAffinity Resin (BioLegend). The C-terminally truncated ECF42-T1 (1-180 aa) was overexpressed in *E. coli* (TME3013) by induction with 0.2% arabinose at 30 °C for 4 hours. Cells were collected by centrifugation at 5000 g for 10 min at 4 °C. The

supernatant was discarded and immediately 800µl of ice-cold lysis buffer [20 mM Tris pH 8.0, 5 mM EDTA, 50 mM NaCl, 1× EDTA-free protease inhibitors (Roche)] was added to the cells. After incubation on ice for 30 min, the cells were subjected to sonication as described in the section 2.5.3. Lysates were centrifuges at 12,000 g in a pre-cooled centrifuge for 10 min and the supernatant was kept for further analysis.

The anti-*E. coli* RNAP ImmunoAffinity Resin beads were washed three times with cell lysis buffer as used above. The antibody concentration was adjusted to 5 µg/ml in PBS buffer (pH8.0) and 500 µl of diluted antibody were transferred to a new 1.5 ml tube for each sample. The protein concentration of the cell lysates was adjusted to 1 mg/ml with lysis buffer. Five hundred microliters of cell extract was added to the tube containing 500 µl of diluted antibody resin beads and incubated at 4 °C for 2 hours with gentle agitation.

The agarose beads were collected by centrifugation and washed 3 times with ice-cold cell lysis buffer. After the final wash step, the supernatant was removed and 20 µl of 6× SDS protein loading buffer (ThermoFisher) were added. The samples were incubated for at 95 °C 5 min and centrifuged at 14,000 g in for 5 min at room temperature. The immunoprecipitated RNAP was checked by SDS-PAGE by loading 20 µl of the sample, and the C-terminally truncated 3×Flag_ECF42-T1 (1-180 aa) potentially co-immunoprecipitated with RNAP was checked by Western blot using anti-FLAG antibody.

2.5.9 Western blot analysis

2.5.9.1 Western blotting analysis of ECF42 σ factor in *S. venezuelae*.

ECF42 overexpression strains TMS0112, TMS0113 and TMS0114 (**Table 2.3**) were inoculated in 50 ml MYM medium supplemented with R2 trace elements and incubated at 30 °C with agitation. Cells from 10 ml of culture were harvested by centrifugation after growing for 8, 10 and 12 hours and resuspended in 1 ml ice-cold sonication buffer [20 mM Tris pH 8.0, 5 mM EDTA, 50 mM NaCl, 1 x EDTA-free protease inhibitors (Roche)] for sonication. Lysates were then centrifuged at 14,000 g for 15 min at 4 °C to remove cell debris. The protein concentration of the clear lysate was determined using Pierce™ BCA Protein Assay Kit (ThermoFisher, Germany). Twenty microgram of proteins from each sample was submitted for 12.5% SDS-PAGE. After electrophoresis, transfer was carried out to a

polyvinylidene difluoride (PVDF) membrane (Sigma) using a wet transfer cell from Bio-Rad according to the manufacture instruction. The membrane was then incubated in blocking buffer [2.5% milk powder in TBS buffer (50 mM Tris PH 7.6, 150 mM NaCl)] for 1 hour at room temperature. Afterwards, the membrane was incubated with the primary Monoclonal ANTI_FLAG M2 antibody produced in mouse (Sigma, F3165) at room temperature for 1 hour. After four washing steps (10 min per step) with blocking buffer, the membrane was incubated with a secondary Goat Anti-Mouse IgG (H+L) HRP conjugated antibody at room temperature for 1 hour. After four washing steps with block buffer, the membrane was washed with TBS buffer before the signals were detected with a LumiImager (Alpha Innotech) using AceGlow (VWR, Germany) as the chemiluminescence substrate.

2.5.9.2 Western blot analysis of ECF42 σ factor in *E. coli*.

The expression of different C-terminally truncated or mutated ECF42 σ factors was detected by Western blot analysis by using an anti-FLAG antibody. An overnight culture of *E. coli* MK01 strain harboring ECF42 overexpression plasmid (**Table 2.4**) was diluted (1:100) into LB medium (10 ml) and grown at 37 °C with shaking until OD_{600} reached 0.5. Arabinose to a final concentration of 0.2% was then added to the culture to induce the expression of the *ecf42* genes and incubated for another 2 hours.

Cells were then harvested by centrifugation with 5,000 g at 4 °C for 10 min and resuspended in 1.5 ml of lysis buffer (100 mM Tris-HCl pH8.0, 100 mM NaCl) and lysed by sonication as described in the section 2.5.3. Protein concentration was quantified by using Pierce™ BCA Protein Assay Kit (ThermoFisher), and 20 µg of protein from each sample was used for SDS-PAGE analysis. Proteins were then transferred to a PVDF membrane as described above. The membrane was then incubated in blocking buffer for 1 hour at room temperature after which it was incubated with Direct-Blot™ HRP anti-DYKDDDDK Tag Antibody (BioLegend) diluted in block buffer at room temperature for 1 hour. After four washing steps with block buffer, the membrane was washed with TBS buffer before the signal was detected by a LumiImager (Alpha Innotech) using AceGlow (VWR) as the chemiluminescence substrate.

2.6 Manipulation of *S. venezuelae*

2.6.1 Preparation of *S. venezuelae* spores

S. venezuelae was plated out for a confluent lawn on a MYM agar plate and grown at 30 °C for 4 days, when sporulation has already occured. Two microliters of sterile 20% glycerol solution was pipetted onto the surface of the plate and gently rub with a sterile cotton bud to dislodge the spores. The surface of the plate was washed with another 1 ml of 20% glycerol solution then the spore solution was pipetted into a sterile 2 ml microfuge tube. The spores were stored at -20 °C until use.

2.6.2 Total DNA preparation of *S. venezuelae*

Fifty microliters of *S. venezuelae* spores were inoculated into 10 ml MYM liquid medium supplemented with R2 trace elements and grown at 30 °C with agitation for 14 hours. Cells were harvested by centrifugation and the pellet was washed with 10% sucrose and resuspended in 2 ml SET buffer (75 mM NaCl, 25 mM EDTA pH8.0, 20 mM Tris pH7.5). Two microliters of RNase (10 mg/ml) and 40 μl of lysozyme (50 mg/ml) were added to the cell suspension and incubated at 37 °C for 1 hour. Sixty microliters of proteinase K (20 mg/ml) and 240 μl of 10% SDS were added and incubated with cell lysate at 55 °C for 2 hours. Eight hundred microliters of 5M NaCl and 2 ml of chloroform were added after the cell lysate was cooled down to room temperature and then mixed gently for 30 min. After centrifugation at 6,000 g, 4 °C for 15 min, the supernatant was transferred to a new tube to be mixed with 0.6 volumes of isopropanol gently for 30 min. After centrifugation at 12,000 g, 4 °C for 20 min, the supernatant was removed and the DNA pellet was collected and washed with 70% ethanol. Total DNA was dissolved in 200 μl TE buffer (10 mM Tris, pH 8.0, 1.0 mM EDTA) and stored at -20 °C until use.

2.6.3 Conjugation of *S. venezuelae*

Plasmids were transferred to *S. venezuelae* by conjugation using *E. coli* ET12567/ pUZ8002. An overnight culture of *E. coli* ET12567/pUZ8002 harboring the target plasmid was prepared in LB liquid medium supplemented with the appropriate antibiotics. On the next day, the overnight culture was diluted 1:100 in 10 ml of fresh LB medium supplemented with the appropriate antibiotics. When the OD_{600} of the culture reached a value between 0.4 and 0.6, cells were harvested by centrifugation and washed twice with fresh LB medium. Cells were then resuspended in 100 µl LB medium and mixed with approximately 10^8 spores of *S. venezuelae*. The mixed suspension was plated on a MS (SFM) plate supplemented with 50 mM $MgCl_2$ without antibiotics and incubated overnight at 30 °C. Nalidixic acid (0.5 mg) and other appropriate antibiotics to select for the conjugated plasmid in *S. venezuelae* were mixed in 1 ml water and overlaid on the plate. The plate was incubated at 30 °C for 3 to 5 days until *S. venezuelae* colonies were visible. The exconjugants of *S. venezuelae* were then streaked on a fresh DNA agar plate supplemented with nalidixic acid and additional antibiotics for selection of the transferred plasmid three times and the colonies were then tested by colony PCR for the presence of the transferred plasmid.

2.6.4 Generation of ECF42 σ factors triple deletion mutant

S. venezuelae harbors three genes encoding ECF42 σ factors (*sven_0747, sven_4377 and sven_7131*). The markerless triple deletion mutant (Δ*sven_0747* Δ*sven_4377* Δ*sven_7131*) was generated based on the I-SceI Meganuclease-mediated method (**Fig. 2.2**) (Fernandez-Martinez and Bibb, 2014). Approximately 2 kb fragments upstream of genes *sven_0747, sven_4377* and *sven_7131* were amplified using the primer pairs TM4364-TM4365, TM4368-TM4369 and TM4452-TM4453 (**Table 2.5**), respectively. The 2 kb fragments downstream of genes *sven_0747, sven_4377* and *sven_7131* were amplified using the primer pairs TM4366-TM4367, TM4370-TM4371 and TM4454-TM4455, respectively (**Table A1**). The up- and down fragments of each ECF42 gene were pairwise cloned into the delivery vector pIJ12738 (that contains the I-SceI recognition site) with HindIII and KpnI resulting in the constructs pQLS001, pQLS002 and pQLS003 (**Table 2.4**).

As shown in **Fig. 2.2**, plasmid pQLS003 was introduced into *S. venezuelae* wild type by conjugation to generate the single cross-over mutant of *sven_7131*. Single cross-over

mutants were selected on the basis of apramycin resistance and confirmed by colony PCR. The plasmid encoding the I-SceI Meganuclease (pIJ2742) was then conjugated into the single cross-over strain and selected on the basis of thiostrepton resistance.

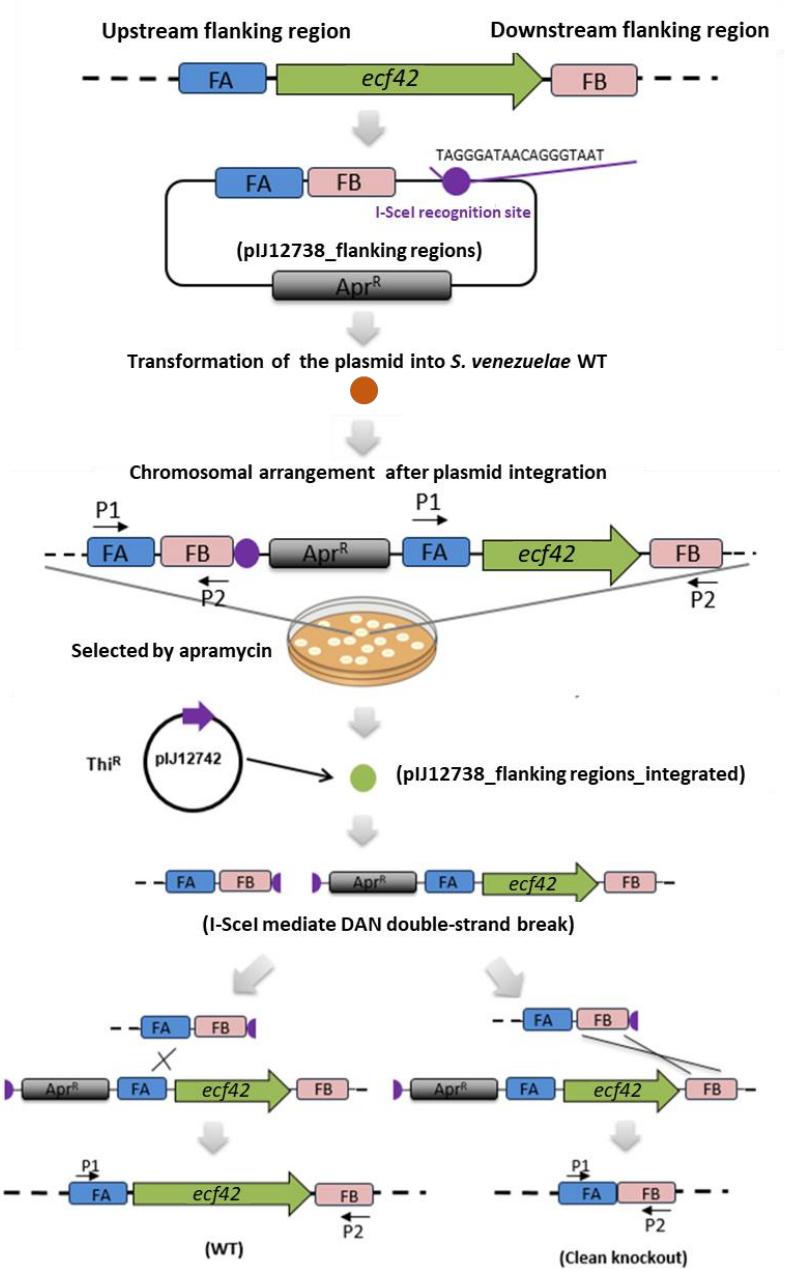


Figure 2. 2 Illustration of the meganuclease I-SceI mediated ecf42 gene deletion. Approximately 2 kb fragments, upstream (FA) and downstream (FB) of the genes encoding Sven_7131 ECF42 σ factor, were amplified and cloned into the vector pIJ12738 containing the I-SceI recognition site. The resulting plasmid was transformed into *S. venezuelae* WT by conjugation. The plasmid integrated into the chromosome DNA at the FA or FB region (FA used as the example here). Single cross-over mutants were selected by apramycin resistance, which was on the basis of the resistance of the introduced plasmid pIJ12738. The correct genotype was confirmed by colony PCR by using primers located up- (P1) and downstream (P2) of the ECF42 coding gene. The plasmid pIJ12742 encoding I-SceI Meganuclease was then conjugated into the single cross-over strain and conjugants were selected on the basis of thiostrepton resistance from pIJ12742. The constitutively

expressed I-SceI caused the double strand breaks at its introduced recognition site (purple dot) in the chromosome. After homologous recombination at FA or FB regions, the double cross-overs were counter-selected on the basis of apramycin sensitivity. If the homologous recombination happens at FA region (left panel on the bottom), the genotype of the strain reverts to WT. If the homologous recombination happens at FB region (right panel on the bottom), the *ecf42* gene is successfully deleted form the chromosome. Deletion was confirmed by colony PCR and sequencing.

Exconjugants were were grown on media containing 50 µg/ml thiostrepton at 30 °C. The constitutively expressed I-SceI generated double strand breaks at its recognition site in the chromosome. After homologous recombination, the double cross-over mutants were selected on the basis of apramycin sensitivity before confirmation by colony PCR. Since pIJ2742 has a temperature sensitive origin of replication, confirmed Δ*sven_7131* mutants were therefore grown at 37 °C, to promote loss of the plasmid. Once a markerless ECF42 single deletion mutant TMS0039 (Δ*sven_7131*) was obtained, the ECF42 double deletion mutant TMS0044 (Δ*sven_7131* Δ*sven_4377*) was subsequently obtained by deleting *sven_4377* gene using the same method of the single deletion mutant (Δ*sven_7131*) background. The triple deletion mutant of ECF42 - TMS0050 (*Δsven_0747 Δsven_4377 Δsven_7131*) - was then generated by deletion of the third ECF42 gene *sven_0747* based on the double deletion mutant TMS0044 in the same way.

2.6.5 Phenotype Microarrays for *S. venezuelae*

Biolog Phenotype MicroArrays for phenotypic characterization of microbial cells (Shea et al., 2012; Mackie et al., 2014), which is a commercial service provided by Biolog (Biolog, Hayward, California), was performed to compare *S. venezuelae* wild-type and ECF42 triple deletion mutant to find the ECF42 related phenotypes. Briefly, *S. venezuelae* was grown overnight on MYM agar medium at 30 °C. After subculture for a second time, cells from the agar plate were transferred into a sterile capped tube containing 25 ml of IF-0a medium (Biolog, Hayward, California) to obtain a uniform suspension with transmittance of OD_{590} value 0.05. One hundred microliters of cell suspension with Biolog metal ion cocktail was then inoculated into the 96 well microplates PM1 to PM20 with different media composition in each well, such as different carbon sources, nitrogen sources, chemical reagents and antibiotics (more information concerning the plate layouts and media composition can be obtained at http://www.biolog.com/products-static/phenotype_ microbial_cells_use.php). All PM plates were incubated at 30 °C for 48 hours and metabolic activity was continuously monitored in all wells of the arrays.

2.6.6 Overexpression of ECF42 σ factors in *S. venezuelae*

Three genes (*sven_0747*, *sven_4377* and *sven_7131*) encoding ECF42 σ factors were amplified with primer pairs TM5631 and TM4056, TM5632 and TM4059 and TM5633 and TM4062, respectively, and cloned into the vector pIJ10257 (Hong et al., 2005) with *Nde*I and *Hind*III to generate ECF42 overexpression plasmids pQLS004, pQLS005 and pQLS006 respectively (**Table 2.4**). The expression of ECF42 σ factors in *S. venezuelae* was controlled by the constitutive promoter P_{ermE*} from vector pIJ10257. A fragment encoding the 3xFLAG tag was introduced to the N-terminus of each ECF42 encoding gene through the forward primer. Each generated overexpression plasmid was conjugated into *S. venezuelae* ECF42 triple deletion mutant TMS0050 to obtain ECF42 overexpression strains TMS0112 (*3xflag_sven*_0747), TMS0113 (*3xflag_sven*_4377) and TMS0114 (*3xflag_sven*_7131). The overexpression of ECF42 σ factors in *S. venezuelae* was confirmed by Western blot using the anti-FLAG antibody as described in the section 2.5.9.

2.6.7 RNA isolation from *S. venezuelae*

For RNA isolation, three independent 30 ml cultures of each *S. venezuelae* strain were grown in MYM supplemented with trace element solution at 30 °C with shaking (initial optical density at 450 nm of 0.01). After 10 hours of incubation, cultures were mixed with 20 ml of precooled (-20 °C) solution of 60% glycerol and 34% of sodium chloride. The cells were then harvested by centrifugation at 8,000 g for 30 min at -20 °C. Cells were resuspended in 1 ml of RNA TRI reagent (ZYMO RESEARCH) and transferred to a beat-beater tube containing glass beads and homogenized with the help of a bead-beater (FastPrep FP120, Thermo) for 3 cycles of 30 sec. Two hundred microliters of chloroform were then added to the homogenized samples, thoroughly mixed by vortexing and centrifuged at 4 °C, 12,000 g for 20 min. The supernatant was transferred to a new tube and mixed with 2.5 volumes of pure ethanol and kept at -20 °C for 12 hours.

Then, the sample was centrifuged at 4 °C, 12,000 g for 20 min. The pellet was washed twice with 80% ethanol. The prepared crude nucleic acids were then dissolved in 200 µl diethylpyrocarbonate-treated water (DEPC-water) and further purified with Direct-zol RNA MiniPrep Plus kit (ZYMO RESEARCH). After that, ribosomal RNA was removed with Ribo-Zero rRNA Removal Kit (Gram-Positive Bacteria, Illumina). The remaining RNA was

cleaned and concentrated in RNAase-free water by using the RNA Clean & Concentrator-25 kit (ZYMO RESEARCH).

2.6.9 RNA-seq and data analysis

The RNAseq and data analysis was performed at the biotechnology center of TU Dresden. Briefly, the mRNA was quantified with Tape Station and 100 ng were chemically fragmented using the Ultra II Directional RNA Library Prep Kit (NEB). Next, reverse transcription was performed with random hexamers and second strand synthesis was performed using dUTP-primer mix. After a XP-bead purification (1.8x) step, the cDNA was subjected to end repair and A-tailing before custom adapters were ligated. After ligation, adapters were removed by an XP bead purification (Beckman Coulter) adding beads in a ratio of 0.9:1. Indexing was done by PCR enrichment (with 6 cycles) using custom amplification primers carrying the index sequence. After two more XP bead purifications (0.9x), libraries were quantified using the Fragment Analyzer (AATI). For Illumina flow cell production, samples were equimolarly pooled and distributed on all lanes used for 75 bp single read sequencing on an Illumina NextSeq 500.

After sequencing, FastQC (v0.11.3; http://www.bioinformatics.babraham.ac.uk/) was used to perform a basic quality control check on the sequence data. Reference genome (ASM25323v1) and annotation (v36) for *S. venezuelae* ATCC 10712 were obtained from Ensembl Bacteria. Reads were mapped to the *S. venezuelae* genome using GSNAP (v2017-08-15). Further quality control on mapped reads, rRNA content and coverage of coding genes was performed with RNA-SeQC (v1.1.8). A table of raw readcounts per gene was created based on the overlap of the uniquely mapped reads with the *S. venezuelae* gene annotation using featureCounts (v1.5.3). Normalization of the raw readcounts based on the library size was performed with the DESeq2 R package (v1.16.1). Principle component analysis, sample-to-sample Euclidean distance as well as Pearson's and Spearman's correlation coefficients were computed based on the normalized gene expression level. For testing differential expression with DESeq2, the count data were fitted to the negative binomial distribution and the p-values for the statistical significance of the fold change were adjusted for multiple testing with the Benjamini-Hochberg correction for controlling the false discovery rate accepting a maximum of 5% false discoveries (adjusted p-value ≤ 0.05).

2.6.10 β-Glucuronidase (GusA) assay

S. venezuelae was grown in liquid MYM medium with shaking for 48 hours at 30 °C. Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM phosphate buffer at pH 7.0, 0.1% triton X-100 and 0.27% β-mercaptoethanol) for sonication as described in section 2.5.3. Cell debris were removed by centrifugation and the protein concentration of the supernatant was measured using the Pierce™ BCA Protein Assay Kit (ThermoFisher). The activity of β-glucuronidase (GusA) in the cell lysate was measured as described previously (Sherwood and Bibb, 2013; Fernandez-Martinez et al., 2015). Briefly, 100 µl of cell lysate and 100 µl of Z-buffer (60 mM Na_2HPO_4, 60 mM NaH_2PO_4 , 10 mM KCl, 10 mM $MgSO_4$, pH7.0) containing PNPG at a final concentration of 4 mg/ml were mixed and distributed in a 96-well plate. The plate was immediately incubated in a Synergy™ NEOALPHAB multi-mode microplate reader from BioTek® for 2 hours at 30°C. The OD_{450} and OD_{550} were measured in each well. The GusA activity was calculated and expressed as miller units per milligram of protein using the following equation: $1000 \times (OD_{420} - 1.75 \times OD_{550})$/time of reaction (min) × volume of culture assayed × protein concentration (mg/ml).

2.7 Manipulation of *B. subtilis*

2.7.1 Transformation of *B. subtilis*

An overnight culture was prepared by inoculating a *B. subtilis* colony into 4 ml of LB medium supplemented with the appropriate antibiotics and incubated at 37 °C with agitation. One hundred microliters of the overnight culture was inoculated into 10 ml of MNGE medium and incubated at 37 °C with agitation. When the OD_{600} value of the culture reached 1.1 to 1.2, four hundred microliters of the culture were incubated with 2 µg of the relevant linearized plasmid at 37 °C for 1 hour with agitation. After that, 100 µl of expression mix was added and incubation was continued for another hour. Finally, 100 µl of the culture was plated on the selective plates. The linearization of pBS3Clux or pBS2E derived plasmids (Radeck et al., 2013) was performed by using the restriction enzymes ScaI and BsaI, respectively.

2.7.2 Luciferase assay with *B. subtilis*

For the implementation of ECF42 from *S. venezuelae* in *B. subtilis*, ECF42 σ factor and its cognate promoter from *S. venezuelae* were moved to the heterologous host. Luciferase activities of *B. subtilis* strains harboring P_{ecf42}-*lux* and P_{xylA}-*ecf42* fusions were measured using a Synergy™ NEOALPHAB multi-mode microplate reader from BioTek® (Winooski, VT, United States). The Bio-reader was controlled using the software Gen5™ (version 2.06). One hundred microliter cultures were used per well in 96-well plates (black wall, clear bottom, clear lid, Greiner Bio-One). Incubation in the reader occurred at 37 °C with linear agitation and luminescence and OD_{600} were measured every 5 min.

Briefly, an overnight culture was prepared by inoculating a single colony into 4 ml of MCSE medium supplemented with the appropriate antibiotics. Day culture was prepared by diluting 20 µl of the overnight culture into 10 ml pre-heated MCSEC medium without antibiotics and incubated at 37 °C with agitation. When the OD_{600} reached 0.2, it was diluted to an OD_{600} of 0.05 in pre-warmed MCSE medium and distributed into 96-well plate. After incubation in the Bio-reader for 1 hour, production of the ECF42 σ factor was induced by the addition of xylose to a final concentration of 0.5% (w/v), and the incubation and measurement continued for 18 hours. Specific luminescence activity is given in relative luminescence units (RLU) normalized by the background corrected cell density (RLU/OD_{600}).

2.7.3 Pull-down assay of *B. subtilis* RNAP

Since the ECF42 σ factors from *S. venezuelae* were inactive in the heterologous host (*B. subtilis*), a pull-down assay of *B. subtilis* RNAP was performed to check if ECF42 σ factors from *S. venezuelae* could interact with the *B. subtilis* RNAP. Two *ecf42* genes from *S. venezuelae* (*sven_4377* and *sven_7131*) were codon optimized as described in 2.4.1 for expression in *B. subtilis* and separately cloned into the integrative vector pBS2E_P_{xylA} to generated plasmids pQL2E-Pxyl 02 and pQL2E-Pxyl 05 (**Table 2.4**). The resulting plasmid was transformed into *B. subtilis* 1A774, which contains a 6xHis-tag β′ subunit of the RNAP, to generate strains TMB2727 (*Sven_4377*) and TMB2728 (*Sven_7131*). The expression of genes encoding ECF42 σ factors was controlled by the xylose inducible promoter P_{xylA} and both ECF42 σ factors were N-terminally tagged with a 3xFLAG tag. The RNAP pull-down assays were performed as described previously (MacLellan et al., 2008). In brief, overnight

cultures of TMB2727 or TMB2728 were diluted into 100 mL of fresh LB medium supplemented with the appropriate antibiotics and grown at 37 °C with agitation until an OD_{600} of 0.4. Expression of *ecf42* genes was induced with 0.5% xylose for 2 hours. Cells were harvested by centrifugation at 5,000 g for 10 min and resuspended in lysis buffer (50 mM phosphate buffer, 100 mM NaCl, 0.1mM PMSF, 5m M imidazole, pH 8.0) for sonication as described in the section 2.5.3. The cleared lysate was collected by centrifugation (12,000 g, 4 °C for 30 min) and loaded into a column containing 1 mL Ni-NTA metal affinity beads (Qiagen, Germany). The beads were washed with 10 column volumes of lysis buffer containing 5, 10, and 20 mM imidazole. Elution was carried out using 0.5 ml lysis buffer with increasing imidazole concentration (50, 100, 250 and 500 mM). Protein concentration was quantified by using the Pierce™ BCA Protein Assay Kit (ThermoFisher). Twenty microgram samples of cleared lysate, washing steps, and elution fractions were submitted to SDS-PAGE for checking the presence of RNAP (by coomassie staining) and the 3xFLAG_ECF42 σ factors (by Western blot).

2.8 Manipulation of *E. coli*

2.8.1 Preparation of *E. coli* competent cells

E. coli cells were inoculated in 4 ml of LB medium supplemented with the appropriate antibiotics when necessary and incubated at 37 °C with shaking. This overnight culture was diluted 1:100 in fresh LB medium and incubated at 37 °C with shaking until the OD_{600} reached 0.4. The culture was then split into 4 falcon tubes and incubated on ice for 20 min. The cells were harvested by centrifugation at 6,000 g for 5 min at 4 °C and resuspended in 20 ml of 100 mM $CaCl_2$. After 20 min of incubation on ice, cells were harvested by centrifugation as before and resuspended in 4 ml of 15% glycerol containing 100 mM of $CaCl_2$. One hundred microliter aliquots were split into 1.5 ml tubes and stored at -80 °C until use.

2.8.2 Transformation of *E. coli*

Frozen competent cells were incubated on ice for 5 min after which 5 to 10 µl of a ligation reaction or 20 ng of pure plasmid DNA was added and mixed gently. After incubation on ice for 20 min, cells were subjected to heat shock (90 sec at 42 °C), and placed back on ice

for 5 min. After the addition of 600 µl of fresh LB medium the cells were incubated for 1 hour at 37 °C with shaking. Finally, the cells were plated on LB agar plates supplemented with selective antibiotics and incubated overnight at 37 °C.

2.8.3 Two-hybrid assay in *E. coli*

To investigate the interaction between ECF42 σ factors and DGPF proteins or the interaction between the N-terminal σ domain and the C-terminal extension of ECF42 σ factors, bacterial two-hybrid was performed based on the BACTH System kit (Karimova et al., 1998). For the interaction between ECF42 σ factors and DGPF proteins, genes encoding ECF42 σ factors (*sven_4377 or sven_7131*) and DGPF proteins (*sven_4376* or *sven_7130*), were cloned into vectors pKT25 or pUT18C to generate the plasmids pQLE035, pQLE035, pQLE036, pQLE037, pQLE038, pQLE039, pQLE040 and pQLE041 (**Table 2.4**). For the interaction between the N-terminal and C-terminal of ECF42 (XCC3792), genes encoding the N-terminal (1 to 178aa) and the C-terminal (179 to 265aa) fragments were cloned into pKT25, pKT25N, pUT18 or pUT18C to generate the plasmids pQLE042, pQLE043, pQLE044, pQLE045, pQLE046, pQLE047, pQLE048 and pQLE049 (**Table 2.4**). Plasmids encoding the ECF42 σ factor was co-transformed into *E. coli* BTH101 with plasmids encoding the DGPF protein. Plasmids encoding the N-terminal σ region were co-transformed into *E. coli* BTH101 with plasmids carrying genes encoding the C-terminal extension. The pKT25-zip - pUT18C-zip plasmid pair was used as positive control. Twenty microliters of co-transformed cells were plated on LB plates supplemented with IPTG (0.5 mM), X-gal (40 µg/ml) and the appropriate antibiotics for selection. The plate was incubated at 30 °C for 2 days before imaging.

2.8.4 Luciferase assay of *E. coli*

Luciferase activity of *E. coli* MK01 strains harboring level M parts-containing plasmids were assayed using Synergy™ NEOALPHAB multi-mode microplate reader from BioTek® (Winooski, VT, USA). The reader was controlled by using the Gen5™ software. The day cultures were prepared by diluting 40 µl of the overnight cultures into 4 ml of fresh LB medium without antibiotics. After incubation at 37 °C with agitation for 2 hours, the cultures were diluted to an OD_{600} of 0.01 in fresh LB medium and transferred to 96-well plates. The

plate was incubated in the Bio-reader at 30 °C with agitation. After incubation for 1 hour, arabinose to a final concentration of 0.2% was added to induce the expression of the truncated or mutated ECF42 σ factor. Cell growth (OD_{600}) and luminescence was monitored every 10 min for 6 hours. The relative luminescence output (RLU) was normalized to cell density (OD_{600}). The OD_{600} and RLU values were background-corrected by subtracting the respective values measured for wells containing non-inoculated LB medium.

3. Results

3.1 Bioinformatics analysis on ECF42 σ factors

The initial analysis of genetic phylogenetic distribution and genomic context on group ECF42 σ factors in 2009 was based on the dataset containing only 111 ECF42 protein sequences (Staron et al., 2009). To account for the huge increase in bacterial genomes sequenced within recent decade, we reanalyzed group ECF42 based on 2661 ECF42 protein sequences that were extracted from NCBI database.

3.1.1 Phylogenetic distribution of ECF42 σ factors

An unrooted phylogenetic tree of ECF42 σ factors in bacteria was constructed based on the gapless multiple sequence alignment of 2661 ECF42 protein sequences (**Fig. 3.1**). ECF42 σ factors show a wide taxonomic distribution and proteins of this group can be found in 11 different phyla, but are predominant in the *Actinobacteria* (75.5%) and *Proteobacteria* (19.4%) (**Table 3.1**), which form eight and four separated clusters on the tree, respectively. Except one branch of ECF42 σ factors from *Proteobacteria* that is included in a branch otherwise formed by ECF42 σ factors of *Actinobacteria*, the remaining ECF42 σ factors from *Bacteroidetes*, *Acidobacteria*, *Firmicutes*, *Planctomycetes*, *Cyanobacteria*, *Verrucomicrobia*, *Spirochaetes*, *Chloroflexi* and *Gemmatimonadetes* form separated branches (**Fig. 3.1**). The 2661 ECF42 σ factors derive from 1003 different species, which suggest several of these species harbor more than one copy of the ECF42 gene in their genome. This is particularly obvious in the phylum *Actinobacteria*, which harbors an average of more than three ECF42 σ factors per species. However, members of other phyla harbor only one or two ECF42 σ factors per species (**Table 3.1**).

Based on the predominant distribution of the ECF42 σ factors in the *Actinobacteria*, we chose ECF42 σ factors from *S. venezuelae*, which is the novel model organism to study actinobacterial development, to investigate physiological role, regulatory mechanism and target regulons of ECF42 σ factors.

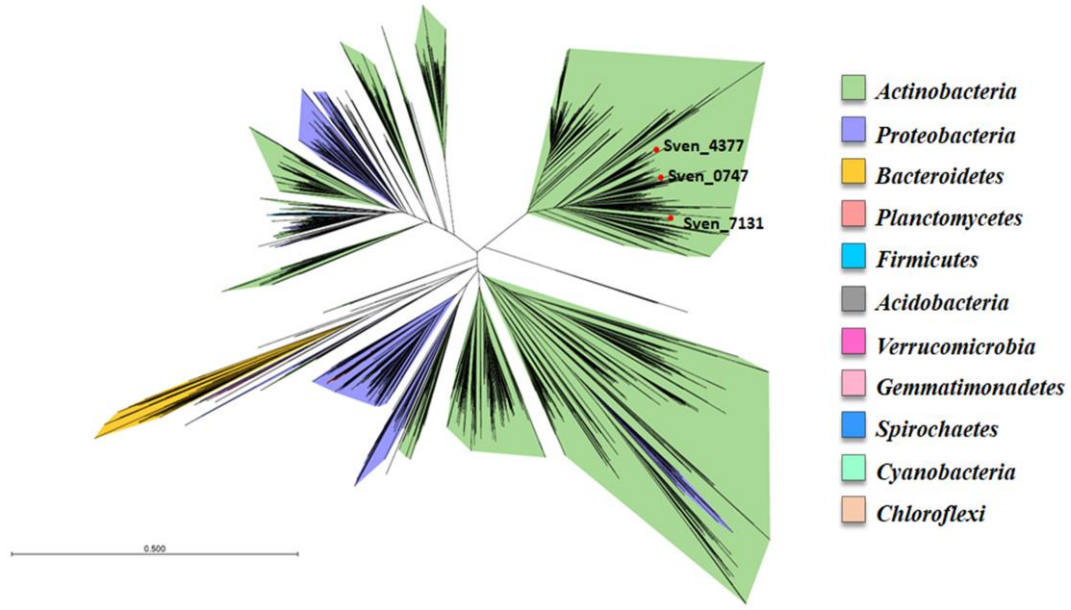


Figure 3. 1 Phylogenetic distribution of ECF42 σ factors. The phylogenetic tree of ECF42 σ factors is based on a gapless multiple sequence alignment of 2661 ECF42 protein sequences obtained using ClustalW as implemented in CLC Main Workbench. The resulting tree was generated using the neighbor-joining method and Jukes-Cantor protein distance model. Assignment of sequences to bacterial phyla is indicated in color. The position of the three ECF42 σ factors encoded in *S. venezuelae* (Sven_0747, Sven_4377 and Sven_7131) were highlighted with a red dot.

Table 3. 1 Phylogenetic distribution of ECF42 σ factors

Phyla	ECF42 proteins in each phylum	Percentage of ECF42 proteins	Characterized species with ECF42 proteins	ECF42 proteins per species
Actinobacteria	2011	**75.5**	544	3.7
Proteobacteria	517	19.4	353	1.5
Bacteroidetes	36	1.4	33	1.1
Acidobacteria	20	0.8	11	1.8
Firmicutes	20	0.8	19	1.1
Planctomycetes	18	0.7	15	1.2
Cyanobacteria	12	0.5	11	1.1
Verrucomicrobia	9	0.3	7	1.3
Spirochaetes	7	0.3	3	2.3
Chloroflexi	6	0.2	4	1.5
Gemmatimonadetes	5	0.2	3	1.7

3.1.2 Genomic context of ECF42

In contrast to most ECF groups defined to date (Pinto and Mascher, 2016b), ECF42 coding genes lack neighboring anti-σ factor coding genes. Instead, more than 90% of *ecf42* genes

are genomically associated with genes encoding DGPF proteins (YciI superfamily, COG3795) (**Fig. 3.2**). In almost 90% of these cases, DGPF coding genes are located closely upstream of those encoding ECF42 σ factors, oriented in the same direction and with a short intergenic region, which is suggestive of co-transcription. In addition to the DGPF proteins, in a few cases proteins from other families are also encoded in the neighborhood: glyoxalase/bleomycin resistance protein/dioxygenase (COG-PhnB), acyl-CoA dehydrogenases (COG-CaiA), an uncharacterized conserved protein (COG-3832) and a ketosteroid isomerase homolog (COG-4319) (**Fig. 3.2**).

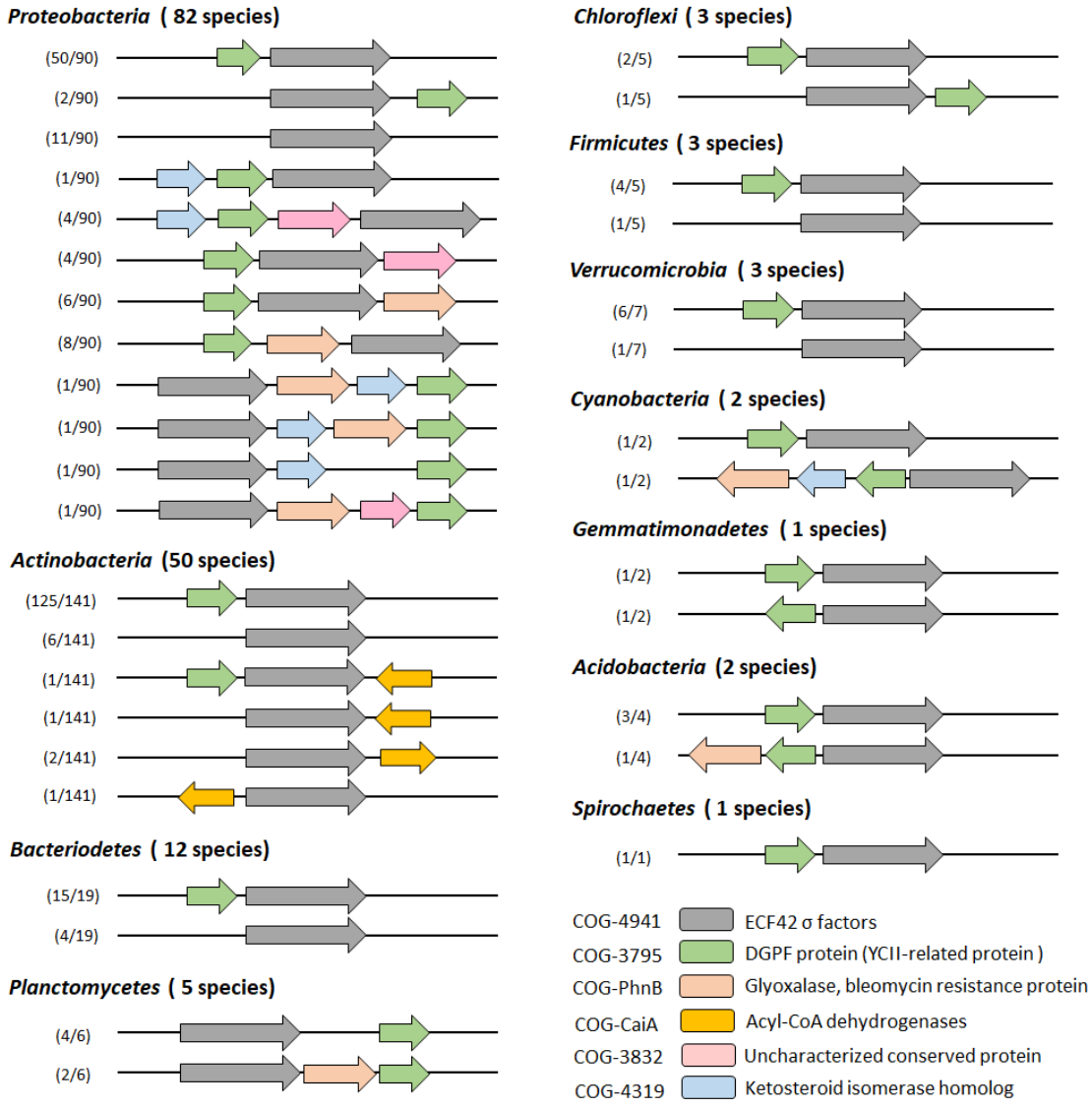


Figure 3. 2 Conserved genomic context of ECF42-encoding genes. Genes are color coded according to their COG family number. The number of analyzed species per phylum is indicated in parenthesis. On the left side of each genomic context illustration is indicated the ratio between the number of genes with that conserved genomic context and the number of ECF42 coding genes identified in all the species of that phylum.

3.1.3 Domain architecture of ECF42 σ factors

ECF σ factors contain only the σ_2 and σ_4 conserved domains characteristic of the σ^{70} protein family. These are sufficient to ensure interaction with the RNA polymerase and promoter recognition. A multiple protein sequence alignment of classical ECF σ factors from different organisms and those of group ECF42 revealed a large C-terminal extension containing a tetratricopeptide repeat (TPR) conserved protein domain (**Fig. 3.3**). Such TRP domains are known to mediate protein-protein interactions (D'Andrea and Regan, 2003).

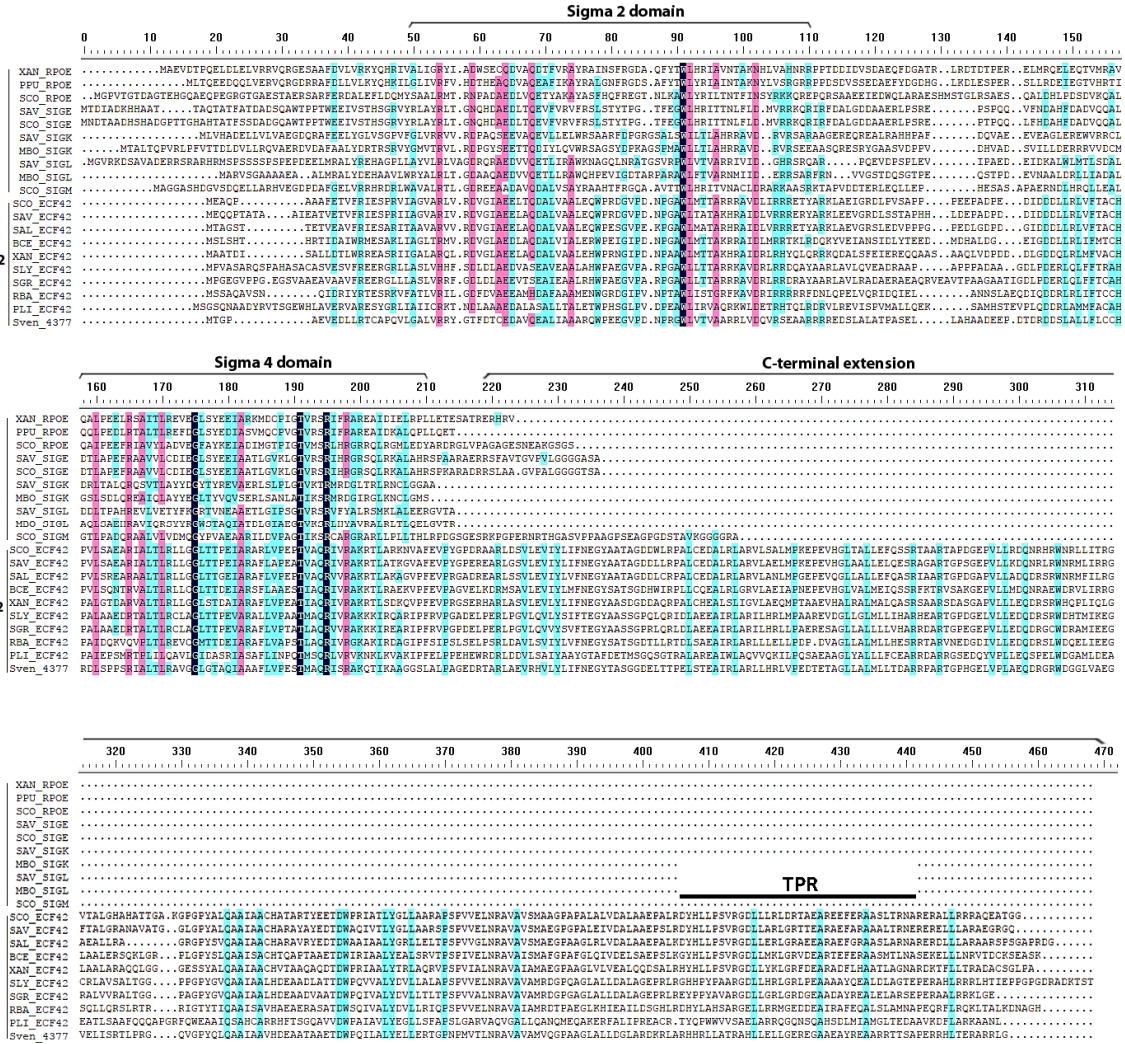


Figure 3. 3 Domain architecture of ECF42 σ factors. The multiple sequence alignment of selected classical and ECF42 σ factors was constructed using DNAMAN. Identical amino acids at the same position are shaded in black, similar amino acids are shaded in pink or in light blue. The σ_2 and σ_4 domains and the C-terminal extension containing the TPR domain are marked on top of the alignment.

3.1.4 Prediction of conserved target promoter motif of ECF42 σ factors

ECF σ factors select promoters with high stringency by combining specific interactions with the -10 and -35 promoter elements and hence each ECF group has their own group-specific target promoter. Identification of the target promoters of ECF σ factors enables us to predict their target regulons, thereby providing a direct access to the physiological role of ECF-dependent regulation. Generally, the ECF-target promoters can be found at the intergenic regions directly upstream of the operon encoding the ECF σ factors, which leads to the positive auto-regulation of most them. Therefore, the putative target promoters of ECF42 σ factors were predicted using the MEME Suite (Bailey et al., 2009) to search for a conserved promoter motif upstream of the operons encoding ECF42 σ factors from different bacterial species belonging to different phyla. A conserved putative promoter motif was identified with "TGTCGA" in -35 region and "CGTC" in -10 region (**Fig. 3.4**).

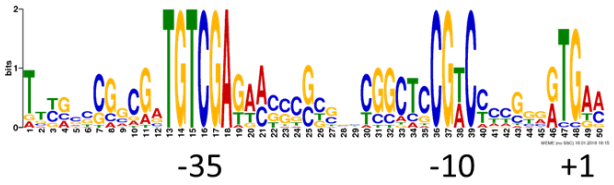


Figure 3. 4 Logo of the putative ECF42 target promoter motif. The logo was generated using the MEME motif discovery tool in MEME Suite available at http://meme-suite.org/. The logo graphically represents the ECF42 promoter position weight matrix and illustrates the degree of sequence conservation for each nucleotide position. The matrix is based on 18 putative promoter sequences found upstream of DGPF- and ECF42 encoding operons.

3.2 ECF42 target promoter determination

3.2.1 Verification of the target promoter of ECF42 σ factors in *S. venezuelae*

Since a conserved promoter motif has been found upstream of ECF42 genes, the upstream sequences of the operons encoding ECF42 σ factors (*sven_0747*, *sven_4377* and *sven_7131*) in *S. venezuelae* were run against the ECF42 target promoter motif to identify the presence of the same promoter sequence. As expected, sequences matching such bipartite promoter motif were found upstream of the three DGPF- and ECF42-encoding operons of *S. venezuelae* (**Fig. 3.5 A**).

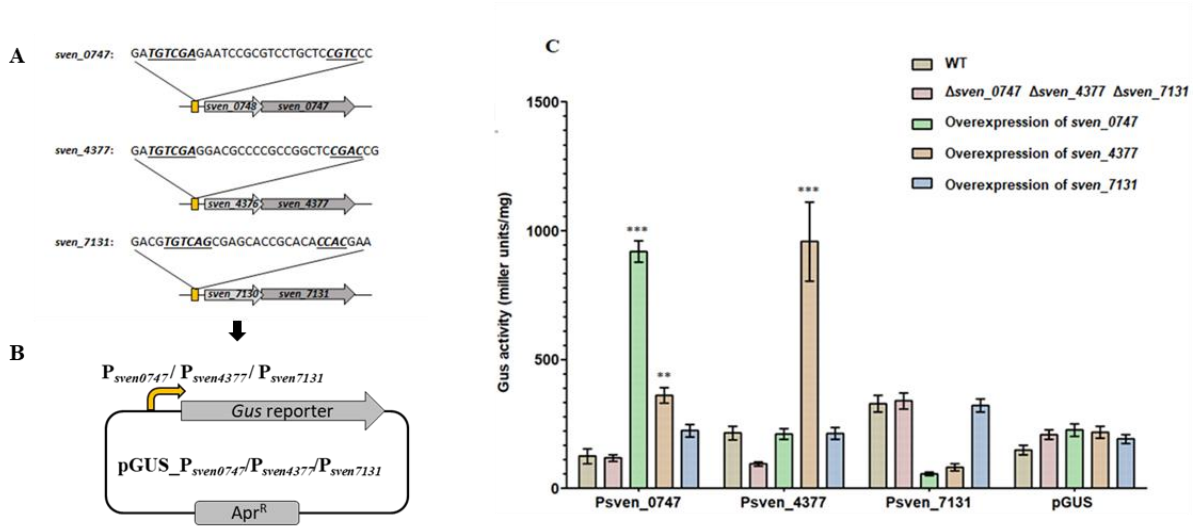


Figure 3. 5 ECF42 promoter determination and crosstalk in *S. venezuelae*. (A) Genomic context of the three ECF42 coding genes of *S. venezuelae*: *sven_0747*, *sven_4377 and sven_7131*. DGPF and ECF42 genes are shown by light gray and dark gray respectively. The orange square represents the location of the putative ECF42 target promoter. The putative -35 and -10 elements are highlighted. **(B)** DNA fragments containing the putative promoters, which start 80 bp upstream of the putative -35 region and end 20 bp downstream of the putative -10 region, were inserted into pGUS, which carries a *gus* reporter gene, generating the plasmids pQLS007 (pGUS_$P_{sven0747}$), pQLS008 (pGUS_$P_{sven4377}$) and pQLS009 (pGUS_$P_{sven7131}$). **(C)** Evaluation of the activity of putative ECF42 target promoters in strains overexpressing different ECF42 σ factors. The integrative reporter plasmid was transformed into *S. venezuelae* strains individually overexpressing the ECF42 σ factors Sven_0747 (TMS0112), Sven_4377 (TMS0113) and Sven_7131 (TMS0114) as well as in the WT (WT) and ECF42 triple deletion mutant (TMS0050) strains. Activity of each promoter and of the empty pGUS vector was measured. The activity was expressed as miller units per microgram proteins. (***P < 0.001, **P < 0.01 as compared to the WT).

Subsequently, the conserved promoter sequences presented upstream of the three DGPF- and ECF42-encoding operons was investigated to disclose whether they are in fact the target promoters of the ECF42 σ factors. A DNA fragment containing the predicted promoter sequence of each ECF42 (**Fig. 3.5 A**) was fused to the β-glucoronidase (Gus) reporter gene in the vector pGUS to generate the plasmids pQLS007 (pGUS_$P_{sven0747}$), pQLS008 (pGUS_$P_{sven4377}$) and pQLS009 (pGUS_$P_{sven7131}$) (**Fig. 3.5 B**). Each plasmid was conjugated into *S. venezuelae* and the activity of the promoter was assessed by measuring the Gus activity in strains individually overexpressing the three ECF42 σ factors (TMS0112, TMS0113 and TMS0114) from the constitutive promoter *ermE**p. As shown (**Fig. 3.5 C**), the activity of the *sven_0747* promoter (P_{sven_0747}) is approximately ten-fold higher in the strain overexpressing Sven_0747 (TMS0112) than in the WT strain, and the *sven_4377* promoter (P_{sven_4377}) has a 5-fold higher activity in the strain overexpressing Sven_4377 (TMS0113) than in the wild type (**Fig. 3.5 C**). The activity of the putative *sven_7131* promoter has not been detected in any of the ECF42 overexpression strains (**Fig. 3.5 C**).

However, these results confirm that two of the predicted promoters are targets of ECF42 σ factors in *S. venezuelae*.

As mentioned above, the genome of *S. venezuelae* encodes three putative ECF42 σ factors. To investigate if there is crosstalk between different ECF42 σ factors and their target promoters, the activity of each ECF42 target promoter was also evaluated in strains overexpressing the other two ECF42 σ factors. No crosstalk was observed between any of the non-cognate ECF-promoter pairs except for the Sven_4377-P_{sven_0747} pair (**Fig. 3.5 C**).

3.2.2 Implementation of ECF42 σ factors from *S. venezuelae* in *B. subtilis*

The genetic manipulations in *S. venezuelae* is a time-consuming process. Thus, for more conveniently investigate the regulatory mechanism of ECF42 σ factors, the σ factors and their cognate promoters from *S. venezuelae* were implemented in a heterologous host, *B. subtilis* 168, which is the Gram-positive model organism characterized by strong genetics based on natural competence and the availability of numerous genetic tools including standardized vectors suites based on the BioBrick cloning standard (Radeck et al., 2013; Popp et al., 2017) developed by our lab.

Genes encoding ECF42 σ factors (*sven_4377* and *sven_7131*) were codon adapted for optimized expression in *B. subtilis* and cloned into the *B. subtilis* integrative vector pBS2E_P_{xylA}, in which the expression of the *ecf42* genes is controlled by the xylose inducible promoter P_{xylA} (**Fig. 3.6 A**). The target promoter sequences of ECF42 σ factors were cloned into the integrative vector pBS3C*lux* upstream of the luciferase reporter gene cluster (**Fig. 3.6 A**). *B. subtilis* was transformed with both pBS2E_P_{xylA} derived ECF42 overexpression plasmid and pBS3C*lux* derived reporter plasmid, then the promoter activation by the overexpressed ECF42 σ factors was assessed by monitoring luminescence over time. As shown, no activation was not observed in strains overexpressing ECF42 σ factors (Sven_4377 or Sven_7131) for neither their natural (NP) nor the adjusted promoters (AP), in which the GC content was adapted to *B. subtilis*. (**Fig. 3.6 B**).

Since the lack of activity of these σ factors in *B. subtilis* has been observed, the expression of ECF42 σ factors (Sven_4377 or Sven_7131) was checked by Western blot using an anti-FLAG antibody and were able to confirm the expression of the codon adapted *ecf42* genes (**Fig. 3.6 C**). In addition, the interaction between these ECF42σ factors with the RNAP of

B. subtilis was checked by performing a RNAP pull-down assay in *B. subtilis* strain 1A774, which expresses a 6×His-tagged RNAP β´ subunit. The ECF42 σ factors (Sven_4377 or Sven_7131) were overexpressed and the RNAP complex from each strain was purified (**Fig. 3.6 D**) through the 6×His-tagged RNAP β´ subunit. The presence of FLAG-tagged ECF42 σ factors in the purified complexes was detected by Western blot by using an anti-FLAG antibody (**Fig. 3.6 E**). As can be seen by comparing **Fig. 3.6 D** and **Fig. 3.6 E**, Sven_7131 is pulled-down with the RNAP in a specific way, *i.e.*, higher amounts of RNAP seen in the SDS-PAGE loosely correlate with higher amounts of Sven_7131 detected in the Western blot. However, Sven_4377 is pulled-down in a non-specific way, *i.e.*, higher amounts of RNAP seen in the SDS-PAGE do not correlate with higher amounts of Sven_4377 detected in the Western blot; this observation can explain the lack of activity of the Sven_4377-P_{Sven_4377} pair in *B. subtilis*.

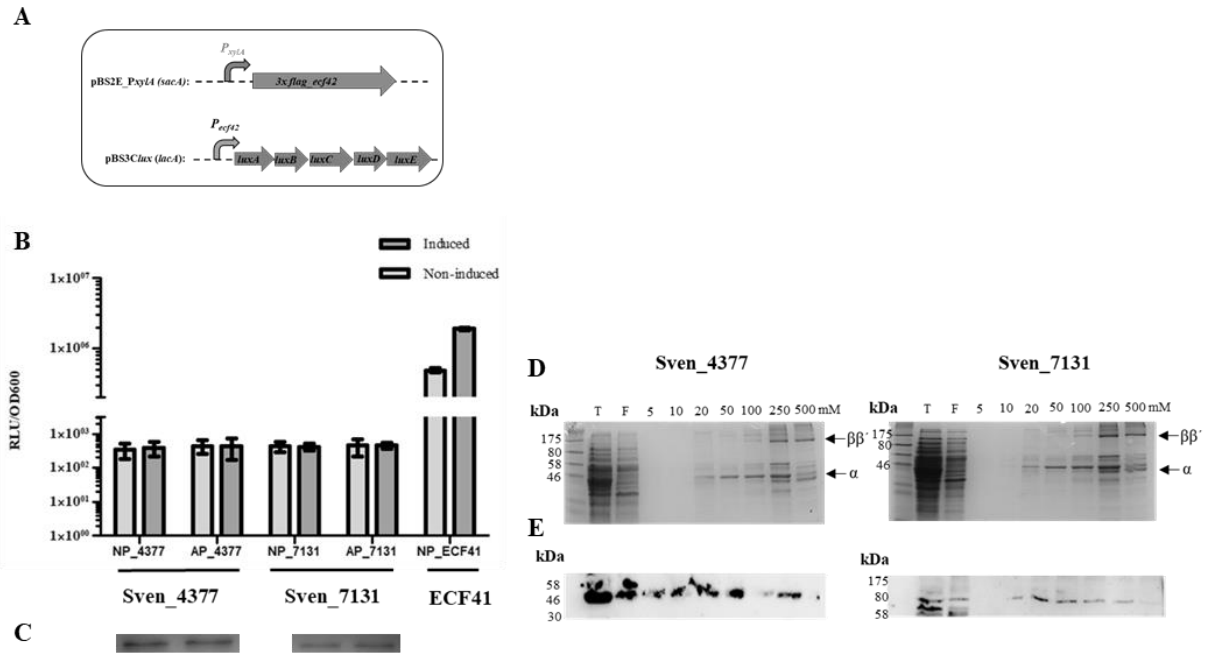


Figure 3. 6 Implementation of ECF42 σ factors from *S. venezuelae* in *B. subtilis*. (A) Genes (*sven_4377* and *sven_7131*) coding ECF42 σ factors were introduced into *B. subtilis* via the integrative vector pBS2E_P_{xylA}, in which the expression of the ECF42 is slaved to the xylose inducible promoter P_{xylA}. The target promoters of ECF42 σ factors were cloned into the integrative vector pBS3C*lux* and drive the expression of the luciferase reporter gene cluster. **(B)** Activity of the natural (NP) and adjusted promoters (AP) was checked by measuring the activity of luciferase. ECF41 with its target promoter were constructed in the same way and tested simultaneously as positive control. **(C)** The expression of ECF42 σ factors in *B. subtilis* was checked by using an anti-FLAG antibody. **(D)** SDS-PAGE analysis of the total (T) protein extract of *B. subtilis* cells overexpressing FLAG-tagged ECF42 σ factors, flow-through (F) from the Ni-NTA column and samples of the elution fractions obtained with different concentrations of imidazole (5 to 500 mM). **(E)** Western Blot detection of the FLAG-tagged ECF42 σ factors (Sven_4377 and Sven_7131) co-purified with the RNAP.

Collectively, these results indicated that ECF42 σ factors from *S. venezuelae* are expressed in *B. subtilis* and at least one of them is able to associate with the RNAP of the host. Although these ECF σ factor-promoter pairs functions in *S. venezuelae*, they do not function properly in *B. subtilis*, for reasons that remain elusive at this point.

3.3 Physiological roles of ECF42 σ factors in *S. venezuelae*

3.3.1 Generation of ECF42 triple deletion mutant

ECF42 σ factors are highly abundant in *Actinobacteria*, especially in the genus *Streptomyces*, with up to 4 copies per genome. This suggests an important physiological role in these organisms. Based on that, the biological role of ECF42 σ factors in *S. venezuelae* was investigated in this study.

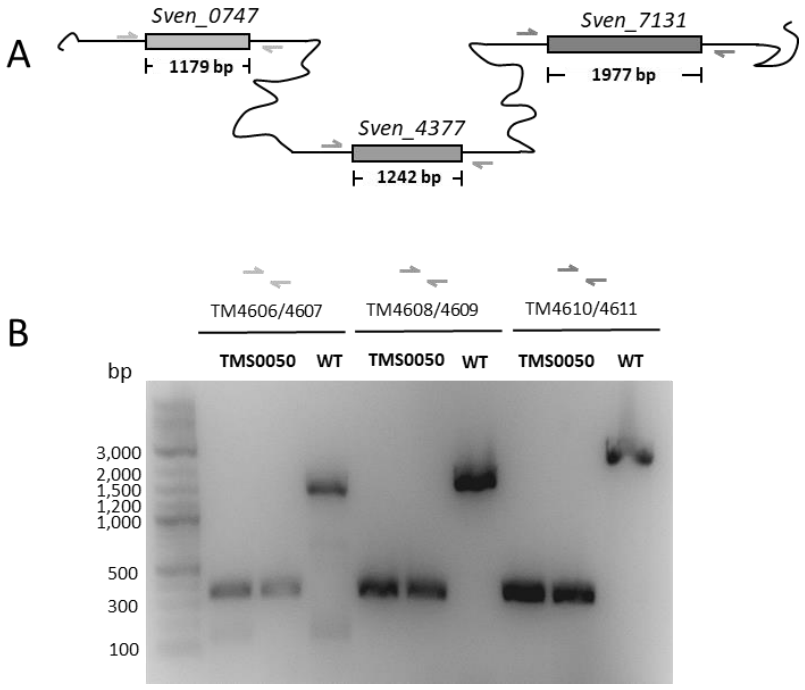


Figure 3. 7 Generation of ECF42 triple deletion mutant. (A) Total DNA of the ECF42 triple deletion strain (TMS0050) was extracted and analyzed by PCR using three different primer pairs (yellow for *sven_0747*, green for *sven_4377* and brown for *sven_7131*), which bind up- and downstream of the three ECF42 coding genes. **(B)** The amplified fragments encompassing the ECF42 coding genes (1179 bp of *sven_0747*, 1242 bp of *sven_4377 and 1977* bp of *sven_7131*) were obtained by PCR using total DNA from the WT strain in three independent PCR reactions with the primer pairs (TM4606 and TM4607 for *sven_0747*, TM4608 and TM4609 for *sven_4377* and, TM4610 and TM4611 for *sven_7131*). The amplified fragments were smaller in triple deletion mutant TMS0050, confirming the absence of the deleted genes.

First, a triple deletion mutant of *S. venezuelae* (TMS0050) was created by deleting three genes (*sven_0747*, *sven_4377* and *sven_7131*) encoding ECF42 σ factors employing the I-SceI Meganuclease mediated method (Fernandez-Martinez and Bibb, 2014). The genotype of TMS0050 was confirmed by PCR. In brief, total DNA from the strain TMS0050 was extracted and analyzed by PCR using three primer pairs (TMS4606 and TM4607 for *sven_0747*, TMS4608 and TM4609 for *sven_4377* and, TMS4610 and TM4611 for *sven_7131*), which bind up- and downstream of each of the ECF42 loci (**Fig. 3.7 A**). In the WT of *S. venezuelae*, these primer pairs generate fragments of 1179, 1242 and 1977 bp, respectively (**Fig. 3.7 B**). In the TMS0050 strain, the same primer pairs generate smaller fragments that correspond to the lack of the ECF42 coding genes, confirming their deletion. The three loci were sequenced to confirm deletion (**Fig. 3.7 B**).

3.3.2 Identification of ECF42-dependent phenotypes by Phenotype Microarrays

The physiological role of ECF42 σ factors is currently unknown. Identification of the ECF42-related phenotype will be helpful to disclose the function of ECF42 σ factors in *S. venezuelae*. Hence, Phenotype Microarrays (PM) analysis, which is a high-throughput approach that allows testing hundreds of different physiological conditions in parallel in order to identify phenotypes associated with genetic alterations (Bochner, 2003), was performed for *S. venezuelae* and TMS0050 to identify ECF42-dependent phenotypes.

The PM analysis, commercially provided by Biolog (Hayward CA, USA), on *S. venezuelae* WT and TMS0050 included 960 assays for carbon, nitrogen, phosphorus, and sulfur utilization, nutrient stimulation, pH and osmotic stress as well as chemical sensitivity test covering 240 different compounds. More information concerning the plate layouts and media composition can be obtained at http://www.biolog.com/products-static/phenotype_microbial_cells_use.php. Careful analysis of the PM results shown that the growth of *S. venezuelae* WT and TMS0050 behave in a similar way in all tested conditions (**Fig. 3.8 A**) except for one: in the presence of the dipeptide methionine-histidine (Met-His) as the sole nitrogen source the triple deletion mutant showed a slightly higher metabolic activity than the WT (**Fig. 3.8 A**). However, this behavior was not observed in the presence of other dipeptides containing methionine (**Fig. 3.8 B**) or histidine (**Fig. 3.8 C**) or in the presence of the single amino acids as sole nitrogen source. We were therefore not able to identify any ECF42-related phenotype.

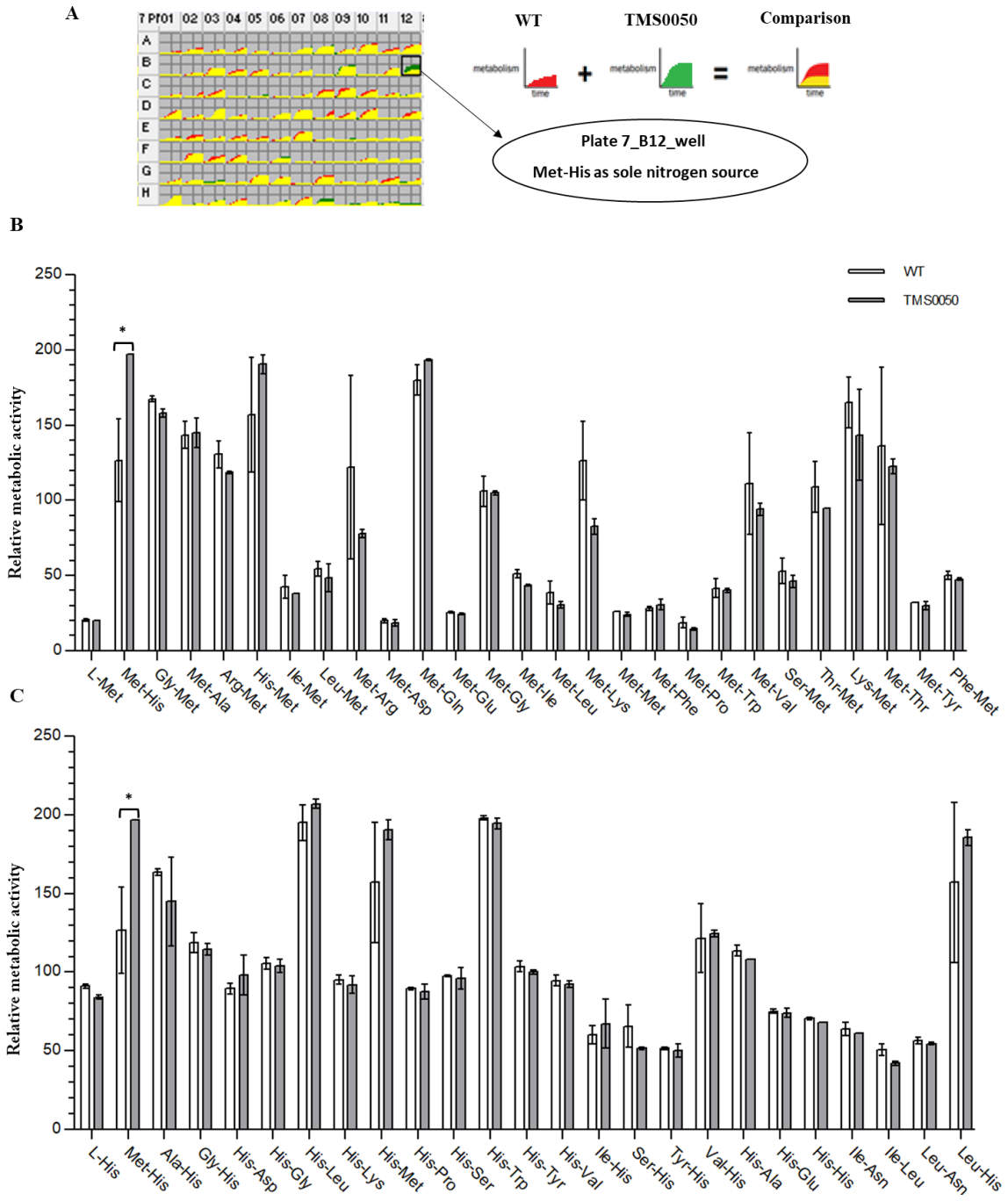


Figure 3. 8 Investigation of ECF42-related phenotype by PM analysis. (A) *S. venezuelae* WT and ECF42 triple deletion mutant (TMS0050) were subjected to PM analysis. Each strain was treated with different conditions in each well of 20 distinct plates (PM1 to PM20). A statistically significant difference of metabolic activity between *S. venezuelae* WT and TMS0050 was observed in well B12 of PM7 plate, in which growth medium had the dipeptide methionine-histidine (Met-His) as the sole nitrogen source. The green area represents the metabolic activity of the ECF42 triple deletion mutant (TMS0050) while the red area represents the metabolic activity of the WT over time. The overlapped regions are represented in yellow. **(B and C)** Metabolic activity of *S. venezuelae* WT and TMS0050 in the presence of methionine or methionine containing dipeptides as sole nitrogen source **(B)**, and in the presence of histidine or histidine containing dipeptides as sole nitrogen source (**C**).

3.3.3 Defining the target regulons of ECF42 σ factors in *S. venezuelae* by RNAseq

Since the phenotypic analysis was non-conclusive, we next aimed at unraveling the physiological role of ECF42s by identifying their target genes in *S. venezuelae*. Towards this goal, RNAseq analyses were performed on strains overexpressing each of the ECF42 σ factors in comparison to the ECF42 triple deletion mutant.

Overexpression of ECF42 σ factors. Strains overexpressing ECF42 σ factors from the constitutive promoter *ermE**p in the ECF42 triple deletion mutant background were constructed. The growth of *S. venezuelae* WT, ECF42 triple deletion mutant (TMS0050) and of the different ECF42 overexpression strains - TMS0112 (*sven_0747*), TMS0113 (*sven_4377*) and TMS0114 (*sven_7131*) – was evaluated in MYM medium (**Fig. 3.9 A**). The production of the ECF42 σ factors (Sven_0747, Sven_4377 and Sven_7131) after 8 hours (lag phase), 10 hours (log phase) and 12 hours (stationary phase) of growth was investigated by Western blot using an anti-FLAG antibody (**Fig. 3.9 B**). All three ECF42 σ factors were successfully produced in all growth phases.

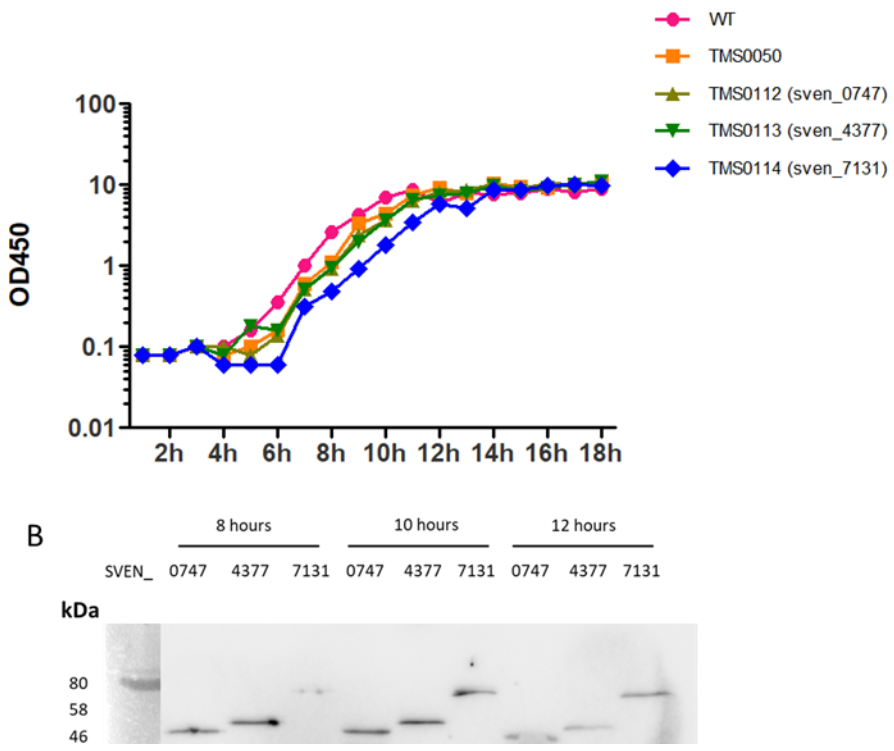


Figure 3. 9 Overexpression of ECF42 σ factors in *S. venezuelae*. (**A**) Growth curves of *S. venezuelae* strains in MYM trace element (start cell density of OD_{450}=0.01) at 30 °C with shaking. OD_{450} was measured every hour. (**B**) The overexpression of the ECF42 σ factors (Sven_0747, Sven_4377 and Sven_7131) at 8, 10 and 12 hours was checked by Western Blot using an anti-FLAG antibody.

Defining the target regulons by RNA-seq. After the ECF42 overexpression strains were constructed, total RNA of cells from mid-exponential cultures (10 hours of incubation), containing the highest level of ECF42 proteins (**Fig. 3.9 B**), was extracted by using Direct-zol RNA MiniPrep Plus kit. Then, the mRNA was purified from total RNA by depleting Ribosomal RNA (rRNA) using Ribo-Zero rRNA Removal Kit. After the mRNA was fragmented and quantified, it was submitted for reverse transcription and synthesis of the second DNA strand to generate the libraries using the Ultra II Directional RNA Library Prep Kit (NEB). Then, the generated libraries were quantified and pooled for 75 bp single read sequencing on an Illumina NextSeq 500. The transcriptional profiles from the ECF42 σ factor overexpression strains were analyzed and compared to that of the ECF42 triple deletion mutant. The up-regulated genes (≥ 2-fold, $P \leq 0.005$, reads ≥ 300) in each of the ECF42 overexpression strains can be found in the appendix (**Table A2**) and the upregulated genes (≥ 2-fold) harboring a target promoter motif of ECF42 are listed in **Table 3.2**. The DGPF protein coding genes *sven_0748* and *sven_4376*, encoded immediately upstream of the ECF42-encoding genes *sven_0747* and *sven_4377* were upregulated more than 16-fold (**Table 3.2**). These results obviously indicated DGPF genes are the target of ECF42 σ factors in *S. venezuelae.*

Table 3. 2 Genes up-regulated in the ECF42 overexpression strains

Gene locus	Fold-change	P value	Description
Strain overexpressing *sven_0747*			
sven_0748	16.34	1.00945E-23	DGPFAETKE family protein
Strain overexpressing *sven_4377*			
sven_3806	643.59	0	DGPFAETKE family protein
sven_4376	20.97	0	DGPFAETKE family protein
sven_6046	7.21	5.51978E-25	Molybdate-binding domain of ModE
sven_3126	2.91	1.98672E-07	Hypothetical protein
sven_4575	2.83	2.09087E-05	Putative RNA polymerase ECF39 group σ factor
sven_1017	2.36	1.52519E-12	Hypothetical protein
sven_0690	2.33	1.53116E-10	Cysteine synthase
sven_4614	2.19	5.28467E-16	Hypothetical protein
sven_1219	2.04	4.78931E-25	Iron utilization protein
Strain overexpressing *sven_7131*			
sven_2864	4.86	1.75073E-05	Phage major capsid protein
sven_1613	3.94	1.76361E-08	Aspartate or tyrosine or aromatic aminotransferase
sven_3169	2.28	1.94133E-14	Molecular chaperone IbpA, HSP20 family
sven_5840	2	4.73082E-06	Hypothetical protein

In addition to genes encoding DGPF proteins or those whose products have no assigned function, genes whose products are involved in multi-functions were also upregulated in the ECF42 overexpression strains (**Table 3.2**). For the targets of Sven_0747, the associated DGPF proteins Sven_0748 is the only target. For the targets of Sven_4377, (i) Sven_6046 harbors a Molybdate-binding domain found in ModE protein of *E. coli*, which is a repressor of modABCD operon coding molybdate-specific transport system (Grunden et al., 1996); (ii) Sven_4575 belongs to the group of ECF 39 and members in this group are involved in cell envelope stress response (Kang et al., 1999) and antibiotics synthesis (Luo et al., 2014); (iii) Sven_0690 shows a homolog domain with cysteine synthase, which involved in amino acid transport and metabolism; (iv) Sven_1219 is a siderophore interacting protein, which shares the domain stricture of the ferredoxin reductase like family, thus involved in iron utilization (Catalano-Dupuy et al., 2006). For the targets of Sven_7131, (i) Sven_2864 is a phage major capsid protein; (ii) Sven_1613 belongs to the aspartate aminotransferase superfamily, members of which are involved in amino acid metabolism; (iii) Sven_3169 is a molecular chaperone, which belongs to heat shock protein (HSP) superfamily, members in this family are found to be involved in resistances to heat and superoxide stress in *E. coli* (Kitagawa et al., 2000). Although most of the target genes of ECF42 σ factors in *S. venezuelae* are with lower fold changes, putative ECF42 target promoters were identified in the upstream regions of all of these genes, suggesting that they are direct targets of ECF42 σ factors (**Fig. 3.10 AB**).

Since the results of RNA-seq had shown DFPG genes upstream of ECF42 genes were the significant targets of ECF42 σ factors, we then focused on the expression levels of all the DGPF coding genes, identified in the genome of *S. venezuelae*, to gain further insight into their expression patterns (**Fig. 3.10 C**). Regarding the three DGPF proteins associated with ECF42 σ factors, only two of them were overexpressed in the strain producing their associated ECF42 σ factors: Sven_0748 in TMS0112 (overexpression of Sven_0747) and Sven_4376 in TMS0113 (overexpression of Sven_4377) (**Fig. 3.10 C**). This is in agreement with our data concerning the identification of ECF42 target promoters in *S. venezuelae*, since *sven_0748* and *sven_4376*, but not *sven_7130*, were active upon overproducing the cognate ECF42 σ factor. Additionally, we also observed that the DGPF protein-encoding gene *sven_0748* is overexpressed in the strain overexpressing Sven_4377. Again, this observation is in agreement with our previous observation that the promoter of *sven_0748* is active in the strain overproducing Sven_4377 (**Fig. 3.10 C**). Other than the ECF42 associated DGPF

genes, one of the three additional genes coding DGPF proteins that is not genomically associated with any ECF42 coding gene (*sven_3806*), was strongly upregulated (643-fold) in the Sven_4377 overexpression strain (**Fig. 3.10 C**). Collectively, our data suggests that regulating the expression of the DGPF proteins is the major role of the ECF42 σ factors in *S. venezuelae*.

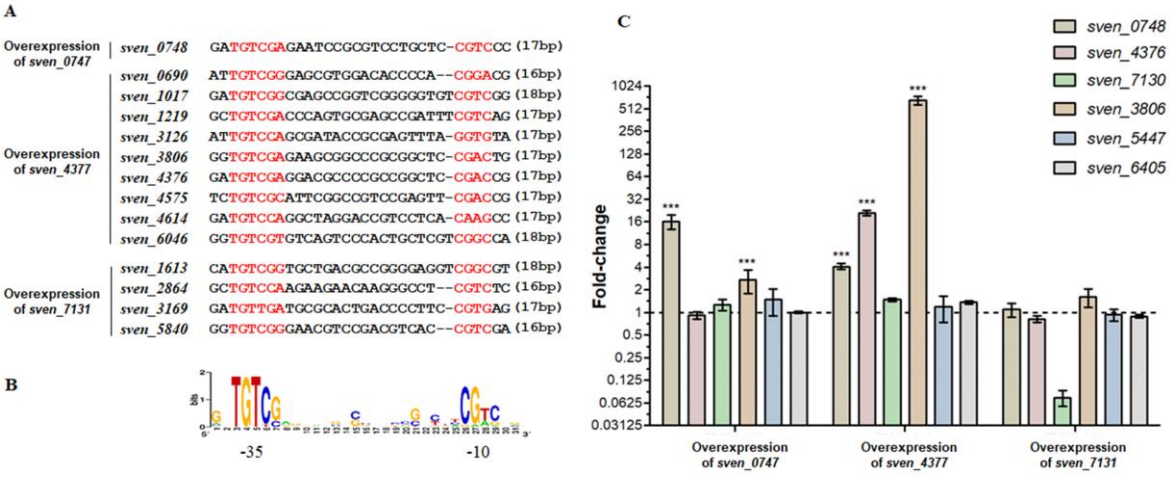


Figure 3. 10 Identification of ECF42 target regulons in *S. venezuelae*. (**A**) Putative promoter sequences in the upstream region of the upregulated genes. The promoter sequences harbor the ECF42 target promoter motif and are shown accordingly to individual ECF42 overexpression strains. The length of the spacer between -35 and -10 regions is indicated for each promoter sequence. (**B**) Logo generated based on the promoter sequences listed in panel A. We note that this motif is very similar to the conserved ECF42 promoter motif shown in Figure 3.4 although the conservation in positions 7, 8 and 29 is lower. (**C**) Expression pattern of all DGPF protein coding genes identified in the genome of *S. venezuelae*. In the graph are shown the fold-change on the expression of the DGPF protein coding genes (*sven_0748*, *sven_4376*, *sven_7130*, *sven_3806*, *sven_5447* and *sven_6405*) in the *S. venezuelae* ECF42 overexpression strains TMS0012 (Sven_0747), TMS0113 (Sven_4377) and TMS0114 (Sven_7131) in reference to the triple deletion mutant TMS0050. (***P < 0.001)

3.3.5 Interaction between DGPF and RibD

DGPF genes were identified as the main target of ECF42 regulation in *S. venezuelae*, thus elucidation of the function of DGPF might be important to disclose the physiological roles of ECF42 σ factors. Currently, the exact function of DGPF proteins is unknown, although the structure of members of the same protein family has been determined (Willis et al., 2005) and one of them has been characterized as a dehydrochlorinase (Hayes et al., 2013). Additionally, the *Escherichia coli* DGPF protein (YciI) was found to interact with RibD, a deaminase/reductase of the riboflavin biosynthetic pathway (Hu et al., 2009). Based on this information, the interaction between DGPF protein (Sven_3806) and RibD (Sven_1041)

from *S. venezuelae* was initially investigated by pull-down in *E. coli* to further explore the physiological role of ECF42 σ factors.

Both codon-optimized genes encoding DGPF protein (Sven_3806) and RibD protein (Sven_1041) were cloned into the overexpression vector pETduet1 separately or in tandem (**Fig. 3.11 A**). A 10xHis tag was introduced into the N-terminus of RibD (37.8 kDa) and a Strep tag was introduced into the C-terminus of the DGPF protein (15.9 kDa). The capture of the DGPF-Strep protein was performed using the magnetic Strep beads (**Fig. 3.11 B**) and the co-purification of 10xHis_RibD protein was investigated by Western blot using an anti-His antibody (**Fig. 3.11 C**).

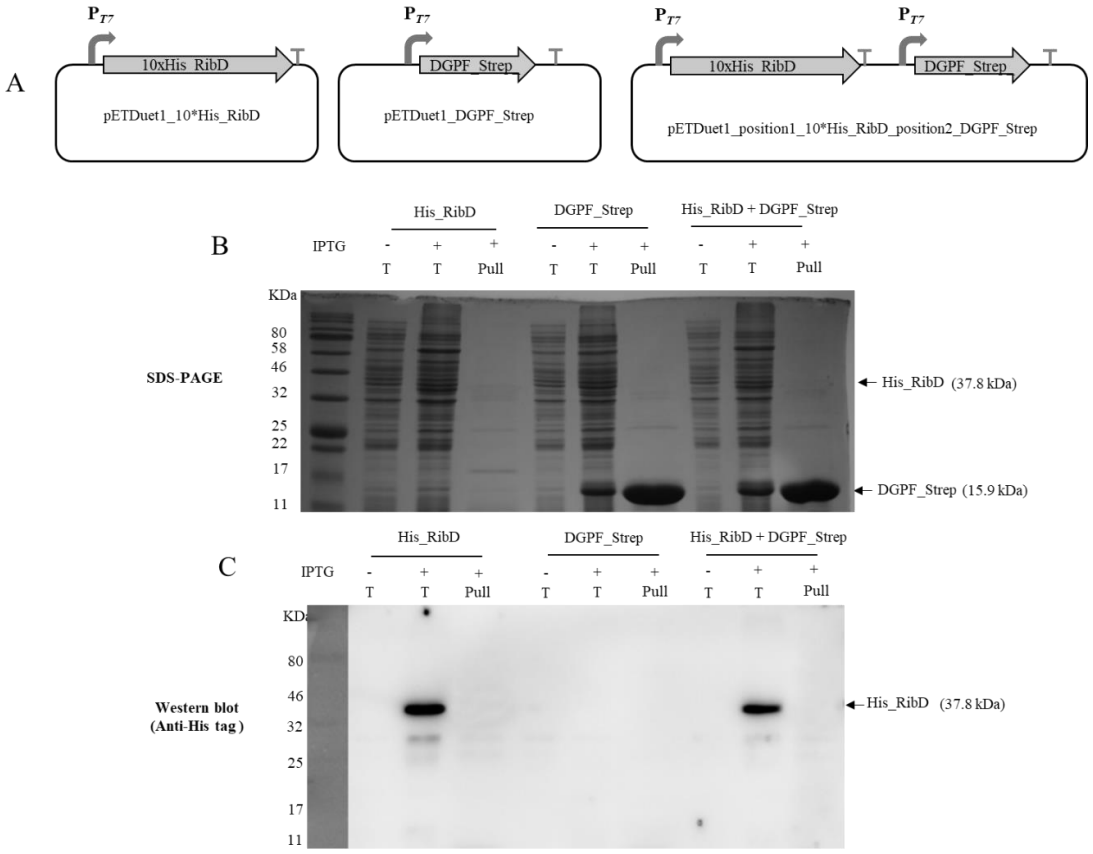


Figure 3. 11 Pull-down analysis of the interaction between DGPF and RibD protein. (**A**) Codon optimized genes encoding RibD (*sven_1041*) and DGPF (*sven_3806*) proteins from *S. venezuelae* were cloned into the pETDuet1 vector separately or in tandem to generate these three overexpression plasmids. Expression of the proteins was independently driven by the IPTG inducible T7 promoter. A 10xHis tag was introduced into the N-terminus of RibD (37.8 kDa) and a Strep tag was introduced into the C-terminus of the DGPF protein (15.9 kDa). (**B**) SDS-PAGE of the total soluble proteins (T) from the strains overexpressing 10xHis_RibD, DGPF_Strep or both with (+) or without (-) IPTG induction. "Pull" represents the protein fractions recovered from the pull-down samples using the magnetic Strep beads. (**C**) Western blot analysis of the same samples as in (**B**) using an anti His antibody. The protein samples and the order in each lane are as the same as in (**B**).

The 10xHis_RibD protein was detected in the total soluble extract of strains overexpressing 10xHis_RibD and 10xHis _RibD and DGPF_Strep. However, the 10xHis_RibD was not detected in the DGPF_Strep pull-down sample from the strains co-overexpressing 10xHis_RibD and DGPF_Strep (**Fig. 3.11 C**). These results show that despite being successfully expressed in *E. coli*, both 10xHis_RibD and DGPF_Strep proteins do not interact with each other.

3.4 Regulation of ECF42 σ factor activity

3.4.1 Regulatory role of DGPF proteins

Generally, the activity of ECF σ factors are controlled by their cognate anti-σ factor located in the their same operons. However, the conserved features of the genetic context in the ECF42 group (**Fig. 3.2**) indicate that ECF42 lack a cognate anti-σ factor. Instead, they are associated and putatively co-expressed with genes encoding DGPF proteins of unknown function (**Fig. 3.5**). The crystal structure of a member of this protein family (COG3795), YciI from *Haemophilus influenzae*, was solved and contains a putative active site, but no enzymatic activity was observed in a subsequent screen (Willis et al., 2005).

Based on the available data, we hypothesized that DGPF protein plays a regulatory role on ECF42 σ factors (**Fig. 3.12**). In the absence of stimulus, ECF42 σ factors are inactive, either by intramolecular interactions or by oligomerization mediated by the TPR-repeat containing C-terminal extension. The YciI-like (DGPF) proteins might function as sensors in a partner-switching mechanism: upon activation, DGPF might change their confirmation and bind the TPR-domain, thereby releasing the ECF σ domain from inhibition (**Fig. 3.12 A**). Alternatively, DGPF proteins and ECF42 σ factors might form a protein-protein interaction complex reminiscent of a typical σ and anti-σ pair in the absence of a suitable stimulus. Once activated, this interaction would then be disrupted, for example by a conformational change of the DGPF proteins, thereby releasing ECF42 σ factors to initiate gene expression from their target promoters (**Fig. 3.12 B**).

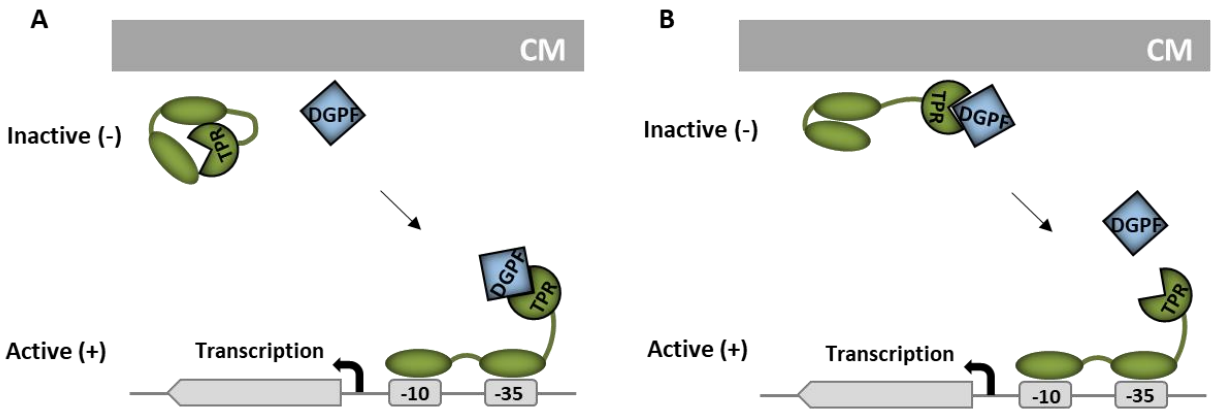


Figure 3. 12 Putative regulation mechanism of DGPF proteins on ECF42 σ factors activation. (**A**) ECF42 σ factors are inactive in the absence of stimulus. Once the stimulus is perceived by the DGPF protein, it induces a conformational change and it is then bound by the TPR-domain, thereby releasing the ECF σ domain from inhibition. (**B**) In the absence of an inducible stimulus, the DGPF protein and the ECF42 σ factor form a complex reminiscent of a σ and anti-σ complex. Under activating conditions, this interaction is disrupted, thereby releasing the ECF42 σ factor to initiate transcription.

3.4.1.1 Interaction between ECF42 and DGPF proteins in two-hybrid assay

No matter the DGPF protein plays the negative or positive regulatory role on the activity of ECF42 σ factor as we hypothesized, DGPF needs to interact with ECF42 σ factor to execute its regulatory functions. Thus, the interaction between the DGPF protein and the ECF42 σ factor was initially investigated by two-hybrid (BACTH) assays in *E. coli*.

The BACTH system is based on the interaction-mediated reconstitution of the adenylate cyclase (Karimova et al., 1998). It exploits the fact that the catalytic domain of adenylate cyclase (CyaA) from *Bordetella pertussis* (Goyard et al., 1993) consists of two complementary fragments – T25 and T18 – that are not active when physically separated. When these two fragments are brought together by interacting proteins (here, DGPF and ECF42 σ factor), it results in functional complementation between T25 and T18 fragments and, therefore leads the generation of cyclic AMP (cAMP). The generated cAMP then binds to the catabolite activator protein (CAP) to form the cAMP/CAP complex, which is a pleiotropic regulator of gene transcription in *E. coli*. Amongst other targets, it induces the expression of the β-galactosidase gene. In this way, blue colonies on LB-IPTG-X-gal plate are indicative of protein-protein interactions.

The codon-optimized DGPF genes (*sven_4376* and *sven_7130*) and the genes encoding ECF42 σ factors (*sven_4377* and *sven_7131*) were cloned into pTK25 and pUT18C vectors. pTK25 and pUT18C derivatives were co-transformed into *E. coli* BTH101. The transformants were plated on LB-IPTG-X-gal plates and incubated at 30 °C for 2 days. Only an interaction between the positive control plasmids – pTK25-Zip and pUT18C-Zip – was detected, but no additional interactions were observed for any other pair (**Fig. 3.13**). These results suggested that there is no direct interaction between ECF42 σ factors and their cognate DGPF proteins.

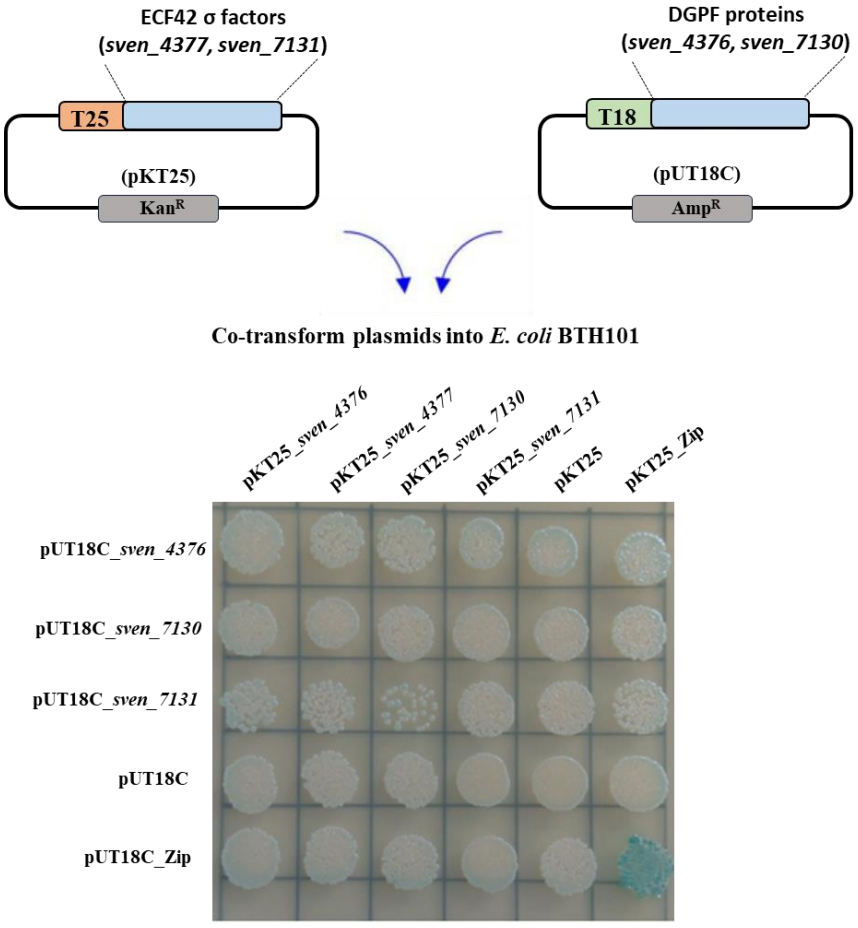


Figure 3. 13 Interaction between ECF42 σ factors and DGPF proteins. Genes encoding ECF42 σ factors (*sven_4377* or *sven_7131*) and DGPF proteins (*sven_4376* or *sven_7130*) were cloned into pKT25 and pUT18C vectors to generate two sets of plasmids. Each pKT25 derived plasmid (ECF42 σ factor or DGPF protein) was co-transformed with each of pUT18C derived plasmid (ECF42 σ factor or DGPF protein) into *E. coli* BTH101. Twenty microliters of co-transformed cells were dropped on a LB-agar plate supplemented with IPTG (0.5 mM), X-gal (40 μg/ml), ampicillin (100 μg/ml) and kanamycin (50 μg/ml). The plate was incubated at 30 °C for 2 days.

3.4.1.2 Effect of DGPF on promoter activation of ECF42 σ factor

Other than the interaction between DGPF and ECF42 σ factors, the effect of the DGPF protein on the activity of the ECF42 σ factors was investigated subsequently. The genes encoding the ECF42 σ factor (*XCC3792*) and its cognate DGPF proteins (*XCC3790*) of *Xanthomonas campestris* were placed under the control of the arabinose inducible promoter (P_{BAD}) and the anhydrotetracycline inducible promoter (P_{Tet}), respectively (**Fig. 3.14 A**). The *luxCDABE* reporter cassette was placed under the control of the ECF42-target promoter. These three transcription units were assembled into one plasmid using the Moclo system (**Fig. 3.14 A)**. *E. coli* was transformed with this generated plasmid and the activity of the ECF42 σ factor was evaluated by measuring the luminescence output. Induction of ECF42 σ factor expression by arabinose results in a 10-fold increase in the P_{ecf42} promoter activity (**Fig. 3.14 B**). However, the activity of P_{ecf42} in the presence of ECF42 σ factor is not influenced by the DGPF protein (**Fig. 3.14 B)**. Expression of both ECF42 σ factor and DGPF proteins was confirmed by Western blot. Together with the results of the two-hybrid assay, the collective data indicates that the DGPF protein is not involved directly in the regulation of ECF42 activity.

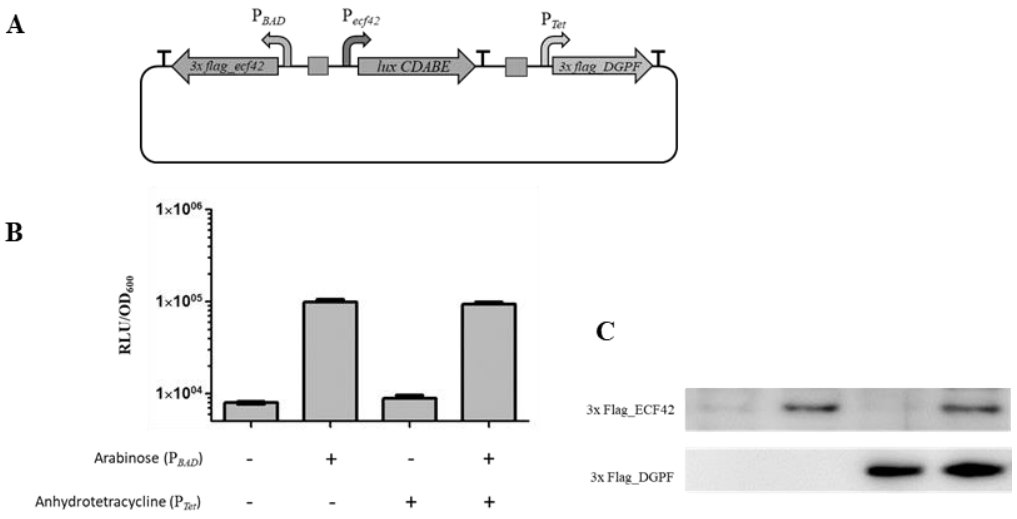


Figure 3. 14 Influence of the DGPF protein on the activity of the ECF42 σ factor. (**A**) Genes encoding the ECF42 σ factor and its cognate DGPF proteins of *X. campestris* were placed under the control of the arabinose inducible promoter (P_{BAD}) and the anhydrotetracycline inducible promoter (P_{Tet}), respectively. The *luxCDABE* reporter cassette was placed under the control of the ECF42 σ factor target promoter. All these three transcription units were assembled using the MoClo cloning method. T shapes represent terminators and the gray squares represent the genetic insulators, preventing the transcriptional interference of the transcription units. (**B**) The activity of P_{ecf42} was evaluated by measuring luminescence in the presence (+) or absence (-) of the ECF42 σ factor and the DGPF protein. (**C**) Western blot analysis of FLAG-tagged ECF42 σ factors and DGPF proteins from the strains induced with the different inducers as the same order as shown above.

3.4.2 Regulatory role of the C-terminal extension of ECF42 σ factor

In addition to being putatively co-expressed with a conserved neighboring gene encoding a DGPF protein, ECF42 σ factors are characterized by the presence of a large C-terminal extension containing a conserved TPR domain (**Fig. 3.3**). Generally, ECF σ factors require their cognate anti-σ factors, locating in the same operon of ECF σ factors, to keep them inactive as long as non-inducing conditions prevail. However, given that no anti-σ factors was identified for ECF42 σ factors, this negative function might reside either on the C-terminal extension and/or on the partner proteins (DGPF). Our data shows that the DGPF protein does not interact with the σ factor (**Fig. 3.13**) and that its activity is not affected by the DGPF protein (**Fig. 3.14**). We therefore propose a regulatory role of the C-terminal extension of ECF42 σ factors, which might serve as an anti-σ like function that directly interacts with the N-terminal σ domain to block the activity of the ECF42 σ factor (**Fig. 3.15**).

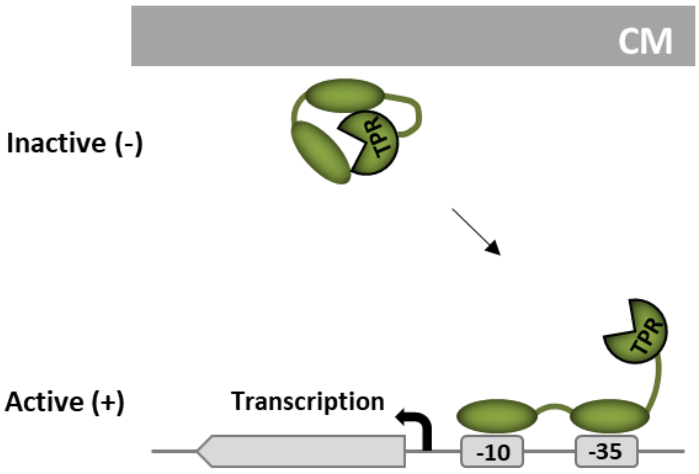


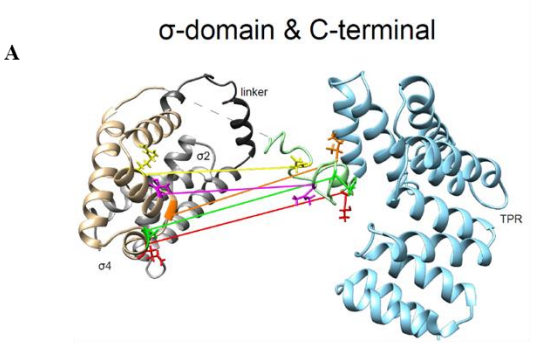
Figure 3. 15 Putative regulatory role of the C-terminal extension of ECF42 σ factors. The C-terminal extension interacts with the N-terminal σ domain and acts as an anti-σ factor in absence of a suitable stimulus. Once triggered, this interaction is then disrupted (probably by a conformational change), thereby releasing ECF42 σ factor to initiate transcription.

3.4.2.1 Prediction of the interaction surfaces between the σ region and the C-terminal extension of ECF42 σ factor

As we hypothesized, the C-terminal extension of ECF42 σ factor might interact with its N-terminal σ domains and play the role of anti-σ factor. Thus, the direct coupling analysis (DCA) (Weigt et al., 2009; Ekeberg et al., 2013) was initially applied to identify the direct residue contacts between the σ region and the C-terminal extension of ECF42 σ factor. DCA

is a statistical inference framework used to infer direct co-evolutionary couplings among residue pairs in multiple sequence alignments. In this case, the DCA was employed for the conserved N-terminal σ domain and the C-terminal extension of ECF42 σ factors and five co-variable amino acid pairs were identified between the σ_4 domain and the C-terminal extension of ECF42 σ factors (XCC3790) from *X. campestris*: R132 - S191, T136 - V195, G141 - I199, A151 - N201 and A147 - R214 (**Fig. 3.16**). This direct-coupling-analysis between N- and C- terminal domains of ECF42 σ factors was performed by our collaborators Hao and Georg (unpublished data).

A



B

N-terminus	C-terminus
R132	S191
T136	V195
G141	I199
A151	N201
A147	R214

C

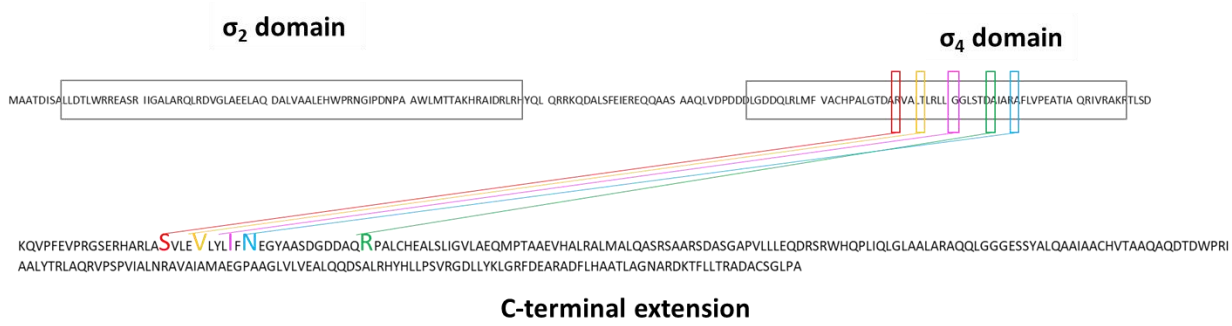


Figure 3. 16 Prediction of interaction between the N-terminal σ domain and the C-terminal extension of *X. campestris* ECF42 σ factor. (**A**) Five potential interacting residues between the C-terminal extension and the N-terminal domain of the ECF42 σ factor from *X. campestris* were identified by DCA. (**B**) The positions and residues of the co-variable amino acid pairs between the N-terminal σ region and the C-terminal extension. (**C**) Position of the co-variable amino acids in the sequence of the σ_4 region and in the C-terminal extension.

3.4.2.2 Alanine-scanning mutagenesis of the co-variable residues in the C-terminal extension decreases the activity of the ECF42 σ factor

Since we hypothesized that the C-terminal extension plays a regulatory role and five co-variable amino acid pairs were identified between the C-terminal extension and σ_4 domain

by DCA, the influence of these amino acids to the activity of ECF42 σ factor was subsequently investigated by site-directed mutagenesis (**Fig. 3.17**).

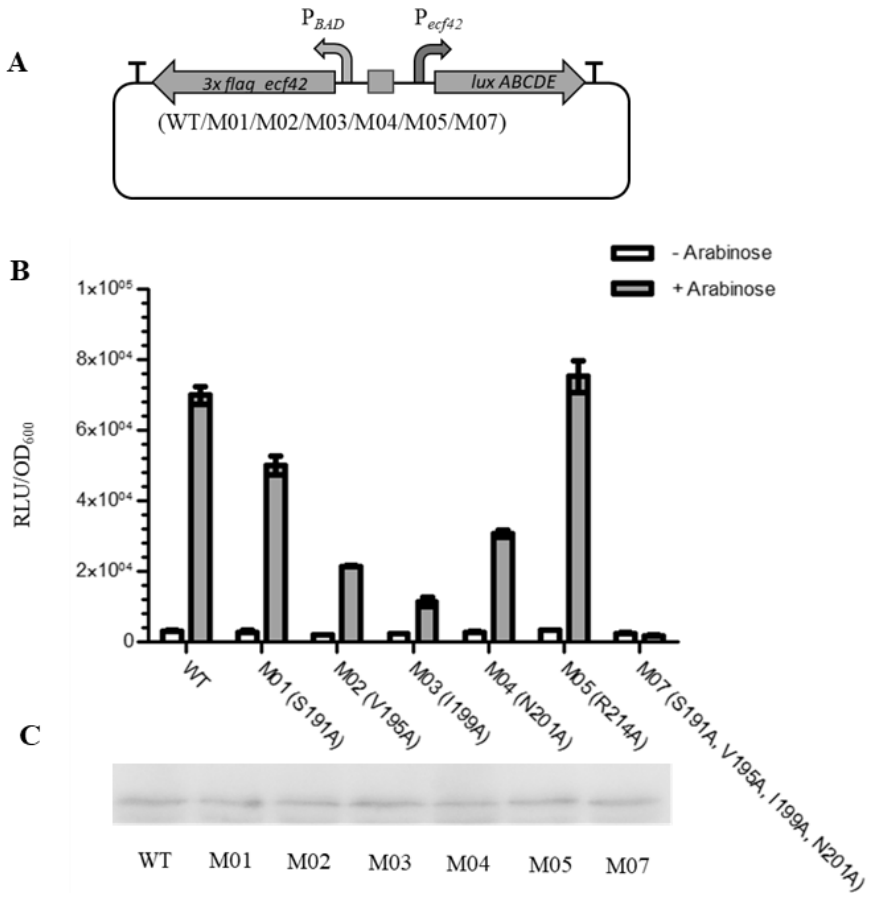


Figure 3. 17 Mutation on the predicted co-variable residues on the C-terminal extension decrease the activity of ECF42 σ factor. (A) The genes encoding ECF42 σ factor WT and the different mutants M01 (S191A), M02 (V195A), M03 (I199A), M04 (N201A) and M05 (R214A), were placed under the control of the arabinose inducible promoter P_{BAD}, while the target promoter of ECF42 σ factor is controlling the expression of the *luxCDABE* reporter cassette. Both transcription units were carried in the same plasmid. T shapes represent terminators and the gray squares represent the genetic insulators, preventing the transcriptional interference of the transcription units from their both sides. **(B)** Activity of different ECF42 σ factor variants as measured by their ability to activate their target promoter. **(C)** Western blot detection of the FLAG-tagged ECF42 variants.

Each residue in the C-terminal extension was substituted by alanine to generate different ECF42 alleles: M01 (S191A), M02 (V195A), M03 (I199A), M04 (N201A) and M05 (R214A). Genes encoding the different ECF42 variants were placed under the control of the arabinose-inducible promoter P_{BAD} and the target promoter of the ECF42 σ factor was used to drive the expression of the *luxCDABE* reporter gene (**Fig. 3.17 A**). The activity of the different variants of the ECF42 σ factor was evaluated by measuring luminescence in *E. coli*. All variants – except for M05 (R214A) – show a reduction in activity of 20 to 80% relative to the wild type (**Fig. 3.17 B**).

Next, another variant was generated in which all four relevant mutations were combined (M07 - S191A V195A I199A N201A), to investigate if the effects of the single mutations on the activity of ECF42 σ factor are cumulative. Indeed, the activity of the M07 variant is dramatically reduced almost to non-induced levels (**Fig. 3.17 B**). This effect was not due to a reduced production level of the mutated alleles of ECF42 σ factor, as demonstrated by Western blot (**Fig. 3.17 C**). These results indicate that the postulated contact between the N-terminal σ region and the C-terminal extension, most likely mediated by the co-variable residues, plays a positive regulatory role on the activity of ECF42 σ factor.

3.4.2.3 Interaction of the N-terminal σ domain and the C-terminal extension of ECF42 σ factors in two-hybrid assay

Although the mutations on the contact sites of the C-terminal reduced the activity of ECF42 σ factor, the contact between the C-terminal extension and N-terminal σ domain was still a prediction of DCA analysis. Thus, two-hybrid assay was performed to experimentally verify the interaction between the N-terminal σ domain and the C-terminal extension of the ECF42 σ factor (XCC3792).

DNA fragments encoding N-terminus or C-terminus were cloned into the pKT25, pKNT25, pUT18 and pUT18C vectors (**Fig. 3.18 A**). *E. coli* BTH101 was co-transformed with each generated T25 fragment containing plasmid and T18 fragment containing plasmid. The empty vectors and the plasmids containing DNA fragment coding leucine zipper (Zip), which are known to form the homodimer (O'Shea et al., 1991), were used as negative and positive control, respectively. The transformants were plated on LB-X-Gal-IPTG plates for screening. As shown (**Fig. 3.18 B**), except for the positive control (pUT18C_zip and pKT25_zip), no other blue colonies were observed. These results argue against an interaction between the different domains of the ECF42 σ factor.

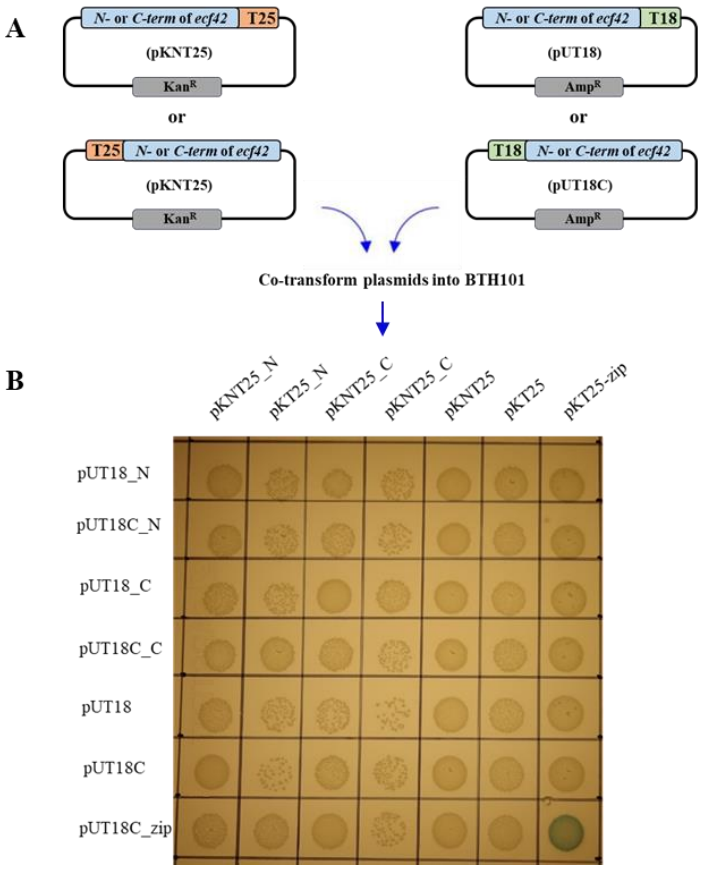


Figure 3. 18 Investigation of the interaction between the N-terminal σ domain and the C-terminal extension of the ECF42 σ factor. (**A**) Gene fragments encoding the N-terminal (N) or C-terminal (C) domains were cloned into T25 and T18 fragments containing vectors. The empty vector and the plasmids (pUT18C_zip and pKT25_zip) containing Zip, which are known to form a homodimer, were used as negative and positive controls, respectively. (**B**) Twenty µl of co-transformed cell were dropped on LB plates supplemented with IPTG (0.5 mM) and X-gal (40 µg/ml). Blue color represents interaction between the hybrid proteins.

3.4.2.4 *In vitro* cross-linking assays to study the interaction between the N-terminal σ region and the C-terminal extension of ECF42 σ factor

Since no interaction between N-terminal σ domain and C-terminal extension was observed in two-hybrid assay, which might be caused by the misfolding of the each part of proteins in *E. coli*, the *in vitro* cross-linking assay thus was performed to further investigate the potential interaction between N-terminus and C-terminus of ECF 42 σ factor subsequently. Both fragments were overexpressed in *E. coli* under the control of an IPTG inducible promoter, however, they were both expressed insolubly (**Fig. 3.19 A**). Therefore, both of the protein fragments were refolded *in vitro* before submitting for purification. The N-terminus and C-terminus of ECF42 σ factor were purified after refolding based on their introduced tags respectively (10×His-σ_domain and C-terminal_extension-Strep; **Fig. 3.19 BC**). Equal molar amounts of the purified N-terminal σ domain and C-terminal extension were mixed

together and crosslinked by the addition of formaldehyde. The crosslinked samples were analyzed by SDS-PAGE (**Fig. 3.19 D**). The alcohol dehydrogenase (ThaADH), which is known to form a homodimer, was used as positive control (**Fig. 3.19 E**). As the results shown, unfortunately, we were again unable to observe any interaction between the different domains of ECF42 proteins in the cross-linking assay (**Fig. 3.19 D**).

Taken together, some potentially contact sites between N-terminal σ domain and C-terminal extension were identified by the DCA prediction, and mutation on the contact sites of C-terminal extension significantly reduced the activity of ECF42 σ factor. However, interaction between N and C-terminal domains were not observed in either two-hybrid assay or *in vitro* cross-linking assay.

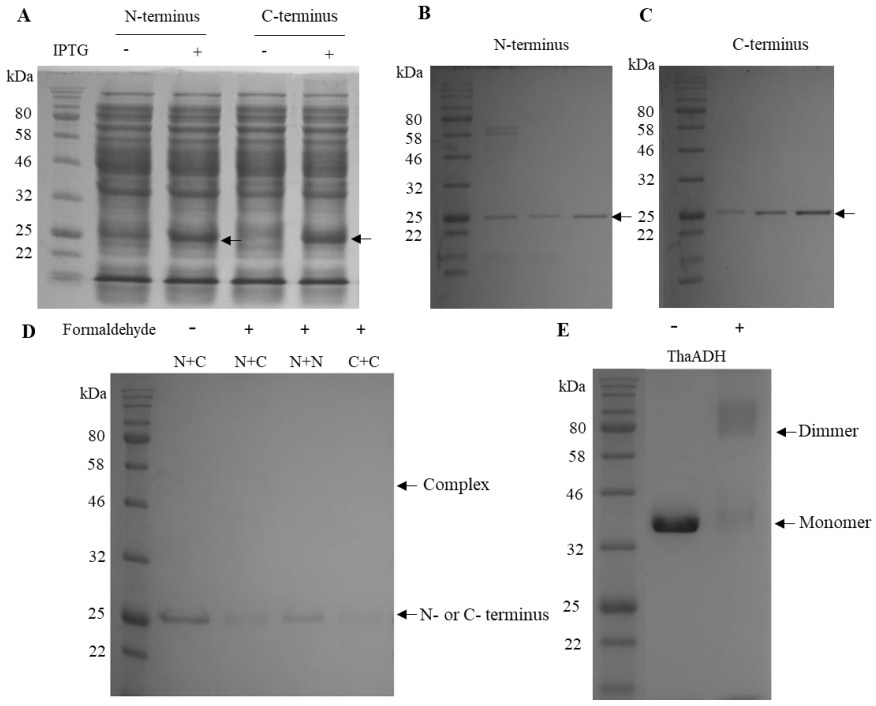


Figure 3. 19 Investigation of the interaction between the N-terminal σ domain and the C-terminal extension of ECF42 σ factor by in vitro crosslinking. (**A**) Fragments encoding the N-terminal σ domain and the C-terminal extension of ECF42 σ factor were cloned into pETDuet-1 overexpression vector, under the control of an IPTG inducible promoter. The expression of both proteins in *E. coli* with (+) or without (-) induction by IPTG was analyzed by SDS-PAGE. The arrows indicate the over-expressed proteins. (**B**) The 10 × His tagged N-terminal σ domain of ECF42 σ factor (arrow) was purified by using His-Trap HP column. Different fractions eluted from column were analyzed by SDS-PAGE. (**C**) The Strep tagged C-terminal extension of ECF42 σ factor (arrow) was purified by using Strep-Tactin XT Superflow cartridge. Different fractions eluted from column were analyzed by SDS-PAGE. (**D**) Crosslinking of the N-terminal σ domain (N) and the C-terminal extension (C) of ECF42 σ factor. Equal amounts of the proteins were mixed as indicated above each lane. The mixture was incubated at room temperature for 5 min and treated with formaldehyde (+) or 20 mM HEPES buffer (-). The arrow indicated the expected position of the dimers. (**E**) Positive control of the crosslinking experiment. The alcohol dehydrogenase (ThaADH), which is known to form a homodimer, was successfully crosslinked.

3.4.2.5 The regulatory role of the C-terminal extension on the activity of ECF42 σ factor

Generally, ECF σ factor harbors only the σ2 domains and σ4 domains, which is sufficient for their functions: assembling with RNAP and initiating the transcription from their target promoters. Interestingly, proteins in ECF42 group are characterized by a long C-terminal extension in addition to σ domains. Such a C-terminal extension with group specificity has also been found in some other groups of ECF σ factors (Pinto and Mascher, 2016b). The function and regulatory role of the C-terminal extension in ECF σ factors has rarely been investigated so far. However, the C-terminal extension of ECF41 has been reported plays both a positive and negative role on the activity of ECF41, which means a shorter truncation of the C-terminal extension increase the activity of ECF41, but the whole truncation of the C-terminal extension results in the complete loss of ECF41 activity (Wecke et al., 2012). Based on this information, the regulatory role of the C-terminal extension of ECF42 was investigated by the sequential deletion analysis.

First, the activity of different C-terminal truncations of the ECF42 σ factor (XCC3790) of *X. campestris* was analyzed in *E. coli* (**Fig. 3.20 ABC**). Genes coding for the full-length (WT) and C-terminally truncated ECF42 σ factor were cloned under the control of the arabinose inducible promoter P_{BAD} and the ECF42 promoter sequence P_{ecf42} was responsible to drive the expression of the *luxCDABE* reporter cassette. Both transcription units were assembled into one plasmid (**Fig. 3.20 B**), which was then transformed into *E. coli* for the evaluation.

After induction with arabinose, the WT ECF42 σ factor shows a seven-fold higher activity than in the absence of arabinose (**Fig. 3.20 C**). Compared to the WT, all C-terminally truncated versions of ECF42 σ factor significantly decreased the activity on the promoter activation. The longest version (T4, 1-355aa) retained some activity showing a 5-fold increase in activity in relation to the non-induced sample. However, the T3 (1-294aa), T2 (1-232aa) and T1 (1-180aa) have completely lost activity (**Fig. 3.20 C**). This behavior was not due to a reduced production level of the truncated alleles, as demonstrated by Western blot (**Fig. 3.20 D**).

Second, the activity of C-terminally truncated ECF42 from *S. venezuelae* was also evaluated *in vivo*. Gene encoding WT and a series of C-terminally truncated ECF42 σ factor (*ven_4377*) were cloned into the integrative vector pIJ10257 under the control of the constitutive promoter P_{ermE*} (**Fig. 3.20 E**). The target promoter of ECF42 σ factor was cloned upstream

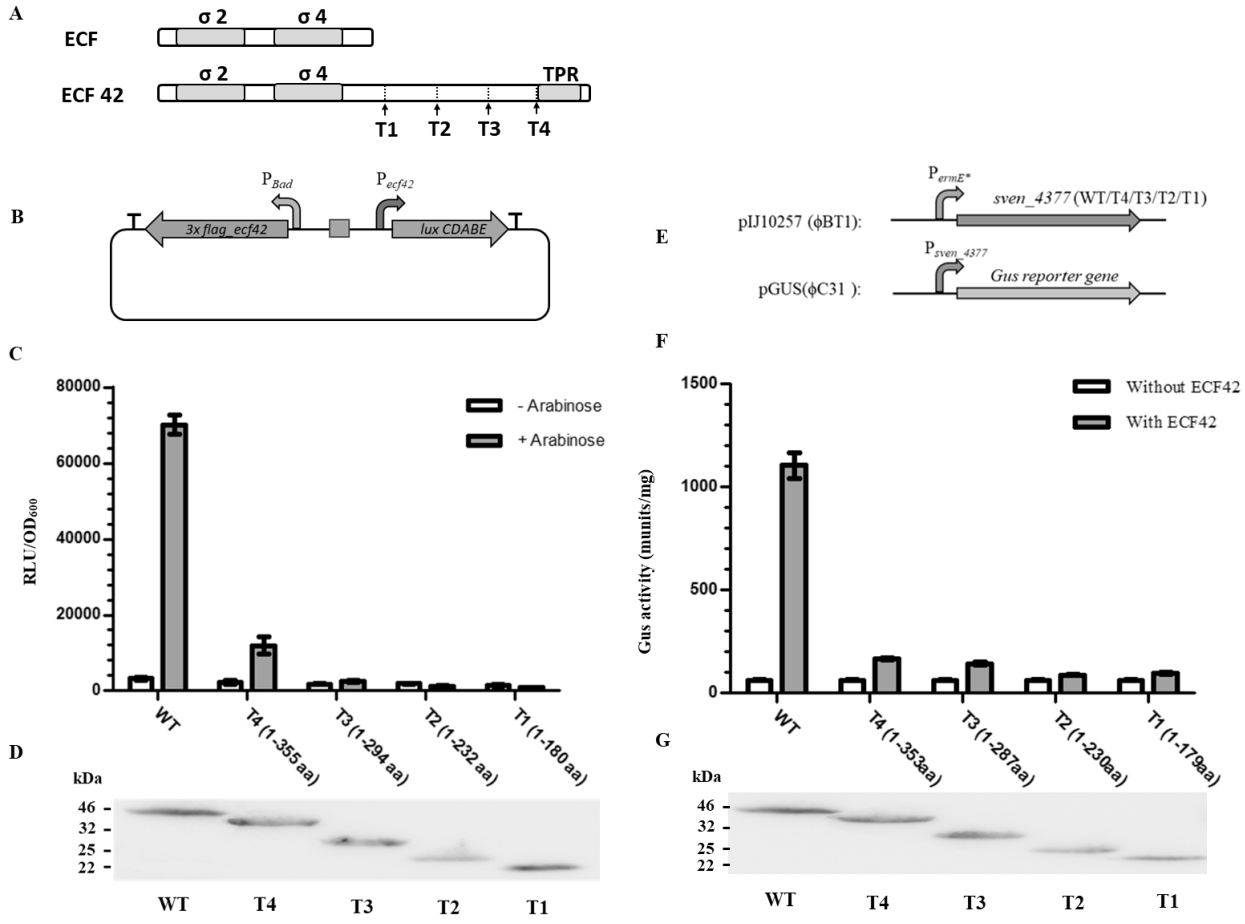


Figure 3. 20 Investigation of the regulatory role of the C-terminal extension of ECF42 σ factor. (**A**) Schematic representation of the protein domain architecture of classical and ECF42 σ factors and positions of the constructed truncations. (**B**) Genes encoding different C-terminally truncated and WT versions of ECF42 σ factor (XCC3790) from *X. campestris* were expressed from an arabinose inducible promoter. The target promoter of ECF42 σ factor drives the *luxCDABE* reporter cassette. Both transcription units were assembled in one plasmid. (**C**) The activity of different C-terminally truncated and WT ECF42 σ factors from *X. campestris* were evaluated in *E. coli* with or without induction by arabinose. (**D**) Western blot detection of the FLAG-tagged ECF42 alleles with an anti-FLAG antibody. (**E**) Genes encoding WT and different C-terminally truncated ECF42 σ factor (Sven_4377) from *S. venezuelae* were cloned into an integrative vector pIJ10257. Production of WT or different truncated ECF42 σ factor was achieved from a strong constitutive promoter P_{ermE*}. The target promoter of ECF42 σ factor was responsible for the expression of the *Gus* reporter gene placed in another integrative vector pGUS. Both generated plasmids were transformed into *S. venezuelae* and integrated into the phage φBT1 and φC31 integration sites. (**F**) Activity of the *sven_4377* promoter in strains expressing the WT and different alleles of ECF42 σ factors. The *sven_4377* promoter was cloned into the integrative pGUS reporter vector and transformed into *S. venezuelae* strains overexpressing truncated FLAG-tagged ECF42 σ factors T1 (1-179 aa), T2 (1-230 aa), T3 (1-287 aa) and T4 (1-353 aa) (X-axis). The activity of the promoter in the different strains was determined by β-glucoronidase assays (gray bars). The activity of the *sven_4377* promoter was also evaluated in the ECF42 triple deletion mutant of *S. venezuelae* as the negative control (white bars). The activity was expressed as miller units per microgram of protein (munits/mg) (Y-axis). (**G**) Production of all ECF42 truncated alleles in *S. venezuelae* was verified by Western blot using a FLAG-tag specific antibody.

of the gene coding for the β-glucuronidase (Gus) in the vector pGUS (**Fig. 3.20 E**). Both integrative plasmids were transformed into *S. venezuelae* by conjugation. The activity of WT

and the C-terminally truncated ECF42 σ factor was evaluated by measuring the Gus activity. All our truncations resulted in a severe decrease in the activity of the ECF42 σ factor relative to its full-length allele (WT), as judged by its target promoter activity (**Fig. 3.20 F**). This behavior was not due to a reduced production level of the truncated alleles, as demonstrated by Western blot (**Fig. 3.20 G**).

Taken together, the results obtained with all C-terminal truncations of ECF42 of *X. campestris* and *S. venezuelae* suggest that the C-terminal extension of ECF42 is necessary for its activity, as seen previously for ECF41 (Wecke et al., 2012). However, in contrast to the situation for ECF41 σ factor, the complete ECF42 C-terminal extension is necessary for activity, since even the shortest truncation of the C-terminal extension (T4) leads to a significant loss of σ factor activity.

3.4.2.6 The role of C-terminal extension on the interaction with the RNAP

As demonstrated above, the C-terminal extension is necessary for ECF42 σ factor activity, as its truncation abolish ECF activity. One possible mechanism to explain this behavior is that the deletion of the C-terminal extension impaired the interaction between ECF42 σ factor and RNAP, which would prevent transcription initiation from ECF42 target promoter. Therefore, immunoprecipitation (IP) of RNAP in *E. coli* was performed to investigate the effect of C-terminal truncation on the interaction between ECF42 σ factor and RNAP.

The shortest truncated allele T1 (1-133aa), in which the whole C-terminus was truncated and that only contain σ_2 and σ_4 regions of ECF42 σ factor (XCC3790) (**Fig. 3.21 A**), was overexpressed in *E. coli* under the control of P_{BAD} promoter. After production of ECF42-T1 induced by arabinose, the RNAP complex was isolated by using the anti-*E. coli* RNA polymerase (β subunit) immune affinity resin. The protein fractions eluted from the affinity resin were analyzed for the presence of RNAP by SDS-PAGE. As the result shown, the RNAP was successfully purified from the cell lysate (**Fig. 3.21 B**). Meanwhile, the presence of the FLAG-tagged ECF42-T1 in the RNAP complex was investigated by Western blot using an anti-FLAG antibody (**Fig. 3.21 C**). Both ECF42 WT and ECF42-T1 were detected in the purified RNAP fractions, suggesting that the C-terminal extension of the ECF42 σ factor is not necessary to mediate the interaction with RNAP.

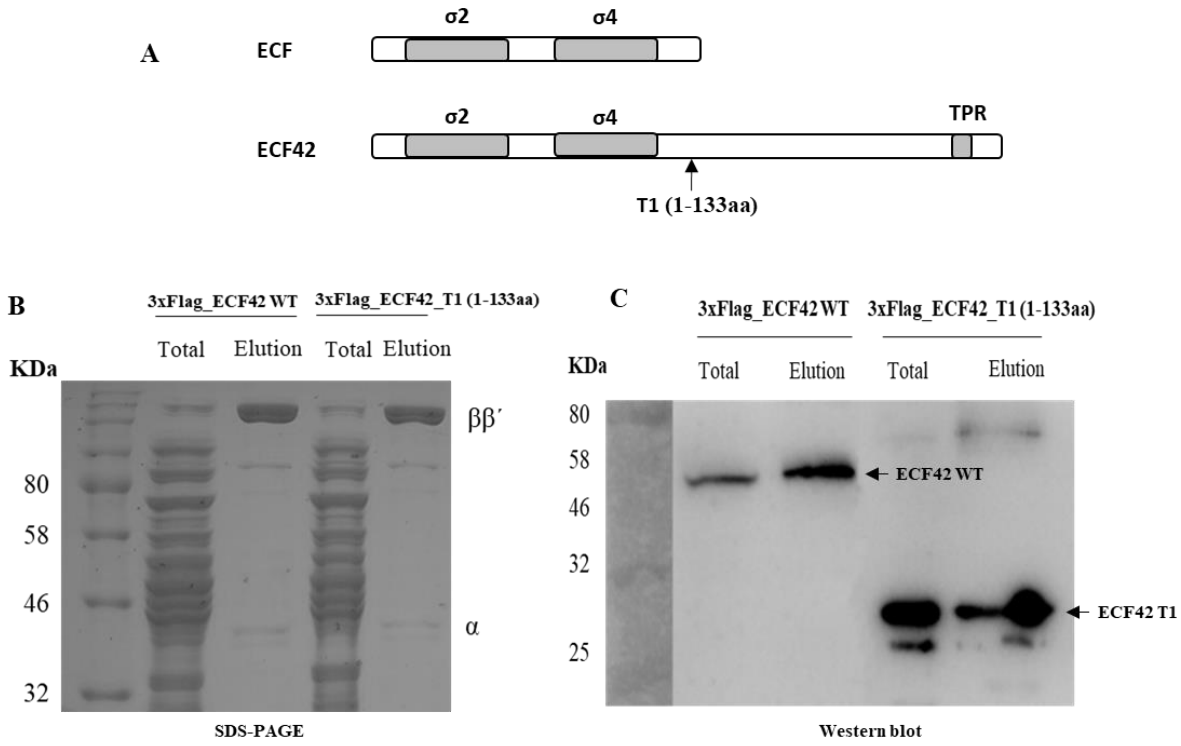


Figure 3. 21 C-terminal extension is not necessary for ECF42 σ factor interaction with RNAP. (**A**) Schematic representation of the protein domain architecture of classical and ECF42 σ factors and the relative position of the ECF42-T1 truncation. (**B**) SDS-PAGE of total soluble cell lysate (Total) or purified RNAP (Elution) from *E. coli* overexpressed FLAG-tagged ECF42 WT or ECF42-T1. ββ' and α subunits of the RNA polymerase are indicated on the SDS-PAGE gel. (**C**) Western blot detection of the FLAG-tagged ECF42 WT or ECF42-T1 in total soluble cell lysate (Total) or in purified RNAP protein fractions (Elution) using anti-FLAG antibody.

4. Discussion

4.1 Phylogenetic distribution and ECF42 abundance

In this study, we provide the very first detailed characterization of members of the novel group of ECF42 proteins with regard to their phylogenetic distribution, physiological role, target promoter sequence, regulons, and the regulatory role of the C-terminal extension and of the conserved co-transcribed DGPF proteins.

The novel group of ECF42 σ factors was initially defined and described in the first comprehensive analysis on ECF σ factors in 2009 (Staron et al., 2009). As a huge increase in available bacterial genome sequences occurred over the last decade, the characteristics of this novel group of ECF42 σ factors, regarding phylogenetic distribution and genomic context, were reanalyzed based on more ECF42 protein sequences. Our analysis showed that proteins in the group of ECF42 show a wide taxonomic distribution, being present in 11 different phyla but predominantly in the *Actinobacteria* (**Fig. 3.1**) with about four copies per genome (**Table 3.1**), which is in accordance with the initial analysis (Staron et al., 2009).

The number of ECF σ factors positively correlates with genome size, thus larger genomes harbor more ECF σ factors (Huang et al., 2015). Another hypothesis for the large number of ECF σ in a genome is that they provide a regulatory advantage for bacteria having complex developmental features or living in complex and rapidly changing environments. These might contribute to an increase ability to perceive their environment and integrate stimulus from different sources, allowing for more complex cognitive behavior (Pinto and Mascher, 2016a).

Following this line of thought, it is not surprising that ECF42 σ factors are so abundant in *Actinobacteria*, which are characterized by large genomes and a remarkably complex developmental process (Flardh and Buttner, 2009). Alternatively or additionally, the high numbers of ECF42 σ factors in *Actinobacteria* also suggests that they might play an important physiological role in these bacteria.

On the other hand, the over-representation of these ECF σ factors in *Actinobacteria*, which encode several ECF42 copies per genome on average (**Table 3.1**) might indicate a functional redundancy. But this interpretation is in contrast to our data that indicates they might work

independently, since we have observed little cross-talk between the ECF-promoter pairs, with the minor exception of the non-cognate pair Sven_4377 - P_{sven_0747} (**Fig. 3.5 C**). This is in agreement with our RNAseq data, showing that each of the three ECF42 proteins activates a different set of genes without any overlap between them (**Table 3.2**). The analysis of the putative ECF42 binding sites in the upstream regions of the genes controlled by each ECF42 σ factor did not reveal any obvious sequence determinants in the core -10 and -35 promoter regions (**Fig. 3.5 A**). However, we cannot exclude that less conserved residues in the spacer region will contribute or determine this specificity, as demonstrated for other ECF promoters in *B. subtilis* (Gaballa et al., 2018b). Unfortunately, the limited number of ECF42 target genes in *S. venezuelae* – indicative of small regulons – currently prevents the identification of such residues.

4.2 The physiological role of ECF42 σ factors in *S. venezuelae*

The genome of *S. venezuelae* encodes three ECF42 σ factors, to avoid the complementarily function to each other, the ECF42 triple deletion mutant of *S. venezuelae* was generated and used in a MicroArray Phenotype assay to identify ECF42-related phenotypes. In addition, RNAseq was performed to identify the regulons of ECF42 σ factors, thereby hoping for some indication as deduced by the (predicted) function of ECF42 target genes. Both of these approaches unfortunately gave very little information concerning their functions in *S. venezuelae*.

For MicroArray Phenotype assay, the slight increase in metabolic activity of the ECF42 triple deletion mutant in the presence of the Met-His dipeptide as sole nitrogen source (**Fig. 3.8**) was surprising, since it implies a negative effect of these ECF σ factors on metabolism. There are two ways in which this mechanism can envisioned:

First, a direct effect of the ECF42 σ factors on genes whose products are involved in the import or utilization of this dipeptide could be postulated. Indeed, a conserved ECF42 promoter signature can be found in the antisense orientation of the 5'-untranslated region of the operon encoding the methionine ABC transporter (*sven_1158*, *sven_1157* and *sven*_1156) as shown in **Fig. 4.1**. This would suggest that ECF42 σ factors negatively regulate the expression of this operon by transcriptional interference. However, the RNAseq data show only a slight downregulation of genes in this operon in the strains overexpressing Sven_0747 or Sven_4377 (between 0.8- and 0.57-fold). In addition, the activity of the promoter of this

gene cluster (P_{Met}) was checked in individual *S. venezuelae* strains overexpressing ECF42 σ factors and compared to that in wild type and ECF42 triple deletion mutant. However, the activity of P_{Met} under normal growth conditions was undetectable in all strains when using the mCherry fluorescent protein as a reporter (data not shown). This observation might be due to the low transcriptional level of methionine ABC transporter gene cluster – a situation reminiscent to *E. coli*, where this operon was only activated if methionine or methionine-related substrates were present as sole nitrogen sources (Kadner and Watson, 1974). It might therefore be worthy to look for suitable growth conditions for *S. venezuelae*, in which the activity of P_{Met} is strong and then compare the behavior of the wild type and the ECF42 triple mutant. Nevertheless, the putative negative effect of ECF42 σ factors on methionine utilization through downregulation of the methionine ABC transporter gene cluster still requires further investigations.

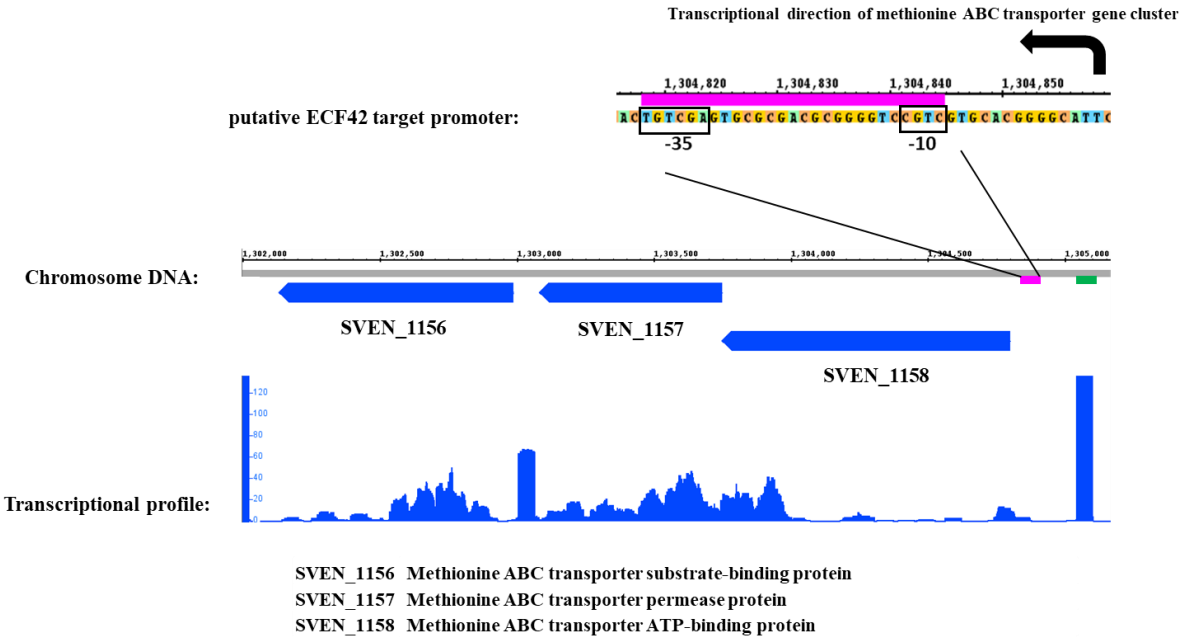


Figure 4. 1 Putative binding site of ECF42 σ factor promoter motif in methionine ABC transporter gene cluster of *S. venezuelae*. A putative promoter motif of ECF42 σ factor indicated as a purple square in the chromosome has been found in the methionine ABC transporter gene cluster of *S. venezuelae*. The ECF42 putative promoter region is zoomed in from the chromosomal DNA and the core -35 and -10 regions are highlighted by a black frame. The putative ECF42 target promoter is located downstream of the transcriptional start site (green square in the chromosome) of the methionine ABC transporter gene cluster. The direction of the ECF42 dependent transcription is contrary to that of the ABC transporter gene cluster.

Second, we could hypothesize an indirect effect of the ECF42 σ factors mediated by their target, the DGPF proteins, as identified by the RNAseq analysis. Proteins of this family have been shown to interact with RibD, a deaminase/reductase of the riboflavin biosynthetic pathway (Hu et al., 2009). Riboflavin is involved in the production of methionine from homocysteine via folate. The interaction of the DGPF with RibD could to some extent interfere with this process. Unfortunately, the interaction between DGPF protein (Sven_3806) and RibD protein (Sven_1041) could not be verified by pulldown assays (**Fig. 3.11**).

In addition to the Phenotype Microarray assays, RNAseq was performed to identify the regulons of ECF42 σ factors in *S. venezuelae* and in this way disclosing their physiological roles. However, few significant target genes were identified, except for the DGPF genes located in the same operon as the ECF42 σ factor coding genes (**Table A2**). This observation suggests that they are to some extent independently self-controlled systems in *S. venezuelae*. Such kind of self-contained ECF regulons that work independently have already been reported in *S. coelicolor* (Kao et al., 2018). In the reported case, the artificial overexpression of the ECF σ factor activates the transcription of genes in the same ECF-contained genomic island. The authors claim that the ECF σ factor and its regulon are a self-contained transcriptional unit, which was presumably acquired by horizontal gene transfer because of the typical characteristics of a genomic island (Kao et al., 2018). Indeed, the transcription units of an ECF σ factor and its contiguous regulon has been reported to be transferrable by horizontal gene transfer in bacteria (Grass et al., 2000; Tibazarwa et al., 2000). Taken together, we were not able to identify a solid ECF42-related phenotype. RNAseq analysis at least identified the solid target protein (DGPF) of ECF42 σ factors in *S. venezuelae* despite their function being unknown. Thus, further research on the function of DGPF proteins will be necessary to explore the physiological role of ECF42 σ factors.

4.3 Target promoter determination of ECF42 σ factors in *S. venezuelae*

Each group of ECF σ factors has their own group-specific target promoter. Hence, identification of the target promoter of ECF σ factors enables us to predict their target genes, thereby providing a direct access to the physiological role of ECF-dependent regulation. Generally, ECF σ factors are positively auto-regulated, which means the target promoter of

the ECF σ factors are usually found upstream of their encoding operons and can be activated and transcribed by the ECF σ factors themselves (Mascher, 2013). In agreement with previous analyses of the ECF42 target promoter motif of ECF42 (Rhodius et al., 2013; Pinto and Mascher, 2016b), the search for over-represented motifs in the upstream regions of operons encoding ECF42 σ factors identified a conserved promoter motif with "TGTCGA" in the -35 region and "CGTC" in the -10 region (**Fig. 3.4**). This finding strongly suggests this conserved motif is most likely the target promoter motif of ECF42 σ factors in *S. venezuelae* as well.

Following this identification of the conserved promoter motif, the activity of the upstream conserved promoters (P_{sven_0747}, P_{sven_4377} and P_{sven_7131}) from three ECF42 encoding operons of *S. venezuelae* were tested in *S. venezuelae* overexpressing their cognate ECF42 σ factor. Two ECF42 σ factor/promoter pairs (Sven_0747-P_{sven_0747} and Sven_4377-P_{sven_4377}) showed significant activity (**Fig. 3.5**), strongly supporting the notion that the predicted promoter motif indeed is the target promoter of ECF42 σ factor and that they are positively regulated as described for other ECF σ factors. This conclusion was also supported by the RNAseq data, in which the DGPF-encoding genes *sven_0748* and *sven_4376* – located in the ECF42 operons of *sven_0747* and *sven_4376* – were highly upregulated in *S. venezuelae* overexpressing *sven_0747* and *sven_4376*, respectively (**Table 3.2**).

However, one ECF42 σ factor/promoter pair – Sven_7131-P_{sven_7131} – was not active. This result made sense in light of the observation that the most likely promoter sequence of ECF42 σ factor we can find upstream of *sven_7131* (**Fig. 3.5 A**) was not very similar to the conserved ECF42 promoter motif (**Fig. 3.4**). Additionally, the ECF42 σ factor encoded by *sven_7131* is much longer than the other two and contains an additional phage integrase N-terminal SAM-like domain (PF02899), which is known to be involved in DNA recombination (Kwon et al., 1997). Remarkably, overproduction of Sven_7131 only resulted in reduced fold-change differences in gene expression (**Table A2**), which suggests that Sven_7131 might not act as a σ factor at all. In conclusion, we experimentally verified the target promoter motif of ECF42 σ factors in *S. venezuelae* and the presence of the target promoter upstream of the ECF42 σ factors encoding operons suggests they are auto-regulated.

4.4 Genomic context conservation and domain architecture of ECF42 σ factors

ECF42 σ factors deviate in a number of characteristics from classical ECF σ factors. Genes encoding ECF σ factors are usually co-transcribed with their cognate anti-σ factors, which are required as negative regulators of the activity of ECF σ factors and keep them inactive in the absence of inducing conditions. However, the genomic context analysis of ECF42 coding genes did not reveal any obvious anti-σ factor coding genes but instead genes encoding DGPF (COG3795) proteins of unknown function (**Fig. 3.2**). Additionally, ECF42 σ factors are characterized by the presence of long C-terminal extensions containing TPR domains that is often involved in mediating protein-protein interactions (D'Andrea and Regan, 2003). In contrast, classical ECF σ factors normally contain only the two σ_2 and σ_4 conserved domains (**Fig. 1.1**).

Based on this conserved genomic context and architectures of ECF42 proteins, we initially proposed that both DGPF proteins and the C-terminal extension of ECF42 σ factors are possibly involved in the regulation of ECF42-dependent gene expression. This assumption led to three central hypotheses:

(i) In the absence of inducing conditions, ECF42 σ factors might be kept inactive due to the formation of a protein-protein interaction complex between the DGPF protein and the ECF42 σ factor, similarly to the σ - anti-σ pair of typical ECF σ factors. Once triggered by a suitable stimulus, the ECF42 σ factor would then be released from the DGPF protein to initiate gene expression from its target promoters (**Fig 3.12 B**).

(ii) Alternatively, ECF42 σ factors might be kept inactive by intramolecular interactions or by oligomerization mediated by their own C-terminal TPR domains in the absence of suitable stimulus. In this case, the DGPF protein might function as a sensor, which changes its conformation in the presence of an inducing stimulus and binds the TPR domain located in the C-terminal of ECF42 σ factor, thereby releasing the N-terminal σ domain of ECF42 σ factor from inhibition (**Fig 3.12 A**).

(iii) The third hypothesis postulated a regulatory mechanism independent of the DGPF protein: The activity of ECF42 σ factor might be inhibited by its C-terminal extension in the absence of a stimulus, while the ECF42 σ factor could be activated by self-cleavage of the C-terminal extension once the suitable stimulus occurs. This hypothesis is in line with a

previous study on ECF41 σ factors (Wecke et al., 2012), which also lack an anti-σ factor but harbor a conserved C-terminal extension. Here, partial truncation of the C-terminal extension increases the activity of the ECF41 σ factor.

Taken together, the bioinformatics analysis, provide us with the first testable hypothesis for the regulatory mechanism of ECF42 σ factors.

4.5 Regulation mechanism of ECF42 σ factors

As described above, the regulatory function governing the activity of ECF42 σ factors might reside either in the associated DGFP proteins and/or in the conserved C-terminal extension of these proteins. To verify the role of the DGPF protein as either a repressor or an activator of ECF42 σ factor, interaction between DGPF proteins and ECF42 σ factors from *S. venezuelae* was initially investigated by two-hybrid assay in *E. coli* and no interaction was observed (**Fig. 3.13**). In agreement with this, the activity of ECF42 σ factor of *X. campestris* (XCC3792) was not affected by the presence of the cognate DGPF protein (XCC3792) (**Fig. 3.14**). All results from our study suggested that the DGPF protein are not involved in the regulation of ECF42 proteins. Given that the ECF42 operons in *S. venezuelae*, which include the DGPF-encoding gene, are positively auto-regulated (**Fig. 3.5**) and DGPF proteins do not seem to harbor a regulatory function, they seem to be just the main target of ECF42 σ factors. This is in agreement with the results from the RNAseq analysis, in which an additional DGPF-encoding gene (*sven_3806*) located separately from ECF42 genes was significantly upregulated by 643-fold in the *S. venezuelae* overexpressing Sven_4377 (**Fig. 3.10**).

For classical ECF σ/anti-σ pairs, the regulation is based on specific protein-protein interactions between the partner proteins that prevent the activity of the ECF σ factor in the absence of inducing conditions (Helmann, 2002). The presence of TPR domains, known to be responsible for protein-protein interactions (D'Andrea and Regan, 2003), in the C-terminal extension of ECF42 σ factors suggested a regulatory role of this domain, in line with recent evidences on the regulatory role of other C-terminal extensions of ECF groups, ECF41 (Wecke et al., 2012) and ECF44 (Marcos-Torres et al., 2016). In the case of ECF41 σ factors, it was demonstrated that the extension has both an activating and a repressing role on σ factor activity, while the extension of ECF44 coordinates a metal ion and determines the specificity of the σ factor (Marcos-Torres et al., 2016).

Initially, we hypothesized that the C-terminal extension plays the role of an anti-σ factor binding to the σ domain of the ECF42 σ factor in the absence of inducing conditions. In line with this hypothesis, five interacting amino acid pair between the σ domain and the C-terminal extension were identified using a direct coupling analysis (DCA). Surprisingly, the activity of ECF42 σ factor was reduced or completely lost when the putative interaction residues located in the C-terminal extension were changed to alanine (**Fig. 3.17**). These results suggest that the interaction between N-terminal σ domain and the C-terminal extension is necessary for ECF42 σ factor's activity. Unfortunately, we were unable to verify the interaction both in an *in vivo* two-hybrid assay (**Fig. 3.18**) and in an *in vitro* cross-linking assay (**Fig. 3.19**). Whether this suggests that the N- and C-terminal portions of the σ factor do not interact directly and that those residues might be important for other interactions or that their mutations disturb the correct folding of the σ factor remains to be investigated. Alternatively, experimental limitations might also lead to the observed result: both the N-terminal and the C-terminal portions of the σ factor are insoluble when overexpressed in *E. coli* (**Fig. 3.19 A**). This behavior might contribute to the negative results in bacterial two-hybrid assay; the absence of interaction observed in cross-linking assay could be caused by protein misfolding *in vitro*; or the interaction is too transient or not strong enough once the two portions are no linker connected within the some protein to be detected by these assays. Hence, the interaction between N- and C-terminal domains will require further investigations.

The TPR domain-containing proteins (Cdc16p, Cdc23p and Cdc27p) of *Saccharomyces cerevisiae* have been reported to be able to form homodimers (Lamb et al., 1994). Thus, it is possible that the activity of ECF42 σ factor is inhibited by its self-dimerization mediated by its C-terminal TPR domains under non-induced conditions. Once triggered by suitable stimuli, the interaction between the TPR domains could be broken, e.g. by the competitive binding of the TPR domain to an unknown signal molecules, thereby releasing the ECF42 σ factors. Although the hypothesis that the TPR domain could play an inhibitory role sounds reasonable, our data from the sequential deletion analysis suggested that the TPR containing C-terminal extension of ECF42 σ factors is instead essential for its activity (**Fig. 3.20**). One possible explanation for this behavior would be to postulate a role of the C-terminal extension of ECF42 σ factor in the assemble of the RNAP-ECF42 holoenzyme. In this case, deletion or disturbance of the C-terminal domain would disable the ECF42 σ factor assembly with RNAP, as has been demonstrated previously for ECF41 σ factors, another ECF group

containing a C-terminal extension (Wecke et al., 2012). But such a mechanism contradicts our RNAP pull-down results, which demonstrated that even a complete deletion of the C-terminal extension did not decrease the affinity of ECF42 σ factor for RNAP (**Fig. 3.21**).

Another possible explanation might be that the TPR-containing C-terminal extension is involved in promoter recognition via an unknown mechanism, in which case truncation of the C-terminal extension would reduce or abolish the ability of ECF42 σ factor to recognize and/or bind its target promoter, thereby preventing transcription initiation. Further studies on the affinity of wild type or C-terminally truncated ECF42 σ factor to the target promoter need to be performed in order to challenge this hypothesis.

Collectively, our data suggests that the full C-terminal extension of ECF42 σ factor is essential for its activity. We hypothesize that any disturbance of the TPR protein-protein interaction domain will render the ECF σ factor inactive. But the exact mechanism by which this effect is exerted will require further investigations. While this question needs to be answered in the future, this work – as presented in this thesis – provides a first comprehensive analysis of a novel and widely distributed ECF group.

We have identified the ECF42 target promoters in *S. venezuelae* and shown that the main target genes of these σ factors are encoding DGPF proteins. Together, these results provide the first insights into the function and mechanism of one of the most abundant, but previously untouched groups of ECF σ factors. I hope that these results will inspire subsequent studies on ECF42 σ factors from different bacterial species to ultimately unravel both the physiological role as well as the exact mechanism of ECF42-dependent stimulus perception and σ factor activation.

5. References

Ades, S.E., Connolly, L.E., Alba, B.M., and Gross, C.A. (1999). The Escherichia coli sigma(E)-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor. *Genes & Development* 13(18)**,** 2449-2461.

Aiyar, S.E., Juang, Y.L., Helmann, J.D., and deHaseth, P.L. (1994). Mutations in sigma factor that affect the temperature dependence of transcription from a promoter, but not from a mismatch bubble in double-stranded DNA. *Biochemistry* 33(38)**,** 11501-11506.

Alba, B.M., and Gross, C.A. (2004). Regulation of the *Escherichia coli* sigma-dependent envelope stress response. *Mol Microbiol* 52(3)**,** 613-619.

Alm, E.J., Huang, K.H., Price, M.N., Koche, R.P., Keller, K., Dubchak, I.L., et al. (2005). The MicrobesOnline Web site for comparative genomics. *Genome Res* 15(7)**,** 1015-1022.

Alvarez-Martinez, C.E., Baldini, R.L., and Gomes, S.L. (2006). A *caulobacter crescentus* extracytoplasmic function sigma factor mediating the response to oxidative stress in stationary phase. *J Bacteriol* 188(5)**,** 1835-1846.

Alvarez-Martinez, C.E., Lourenco, R.F., Baldini, R.L., Laub, M.T., and Gomes, S.L. (2007). The ECF sigma factor sigma(T) is involved in osmotic and oxidative stress responses in *Caulobacter crescentus*. *Mol Microbiol* 66(5)**,** 1240-1255.

Anthony, J.R., Warczak, K.L., and Donohue, T.J. (2005). A transcriptional response to singlet oxygen, a toxic byproduct of photosynthesis. *Proc Natl Acad Sci U S A* 102(18)**,** 6502-6507.

Anthony, L.C., and Burgess, R.R. (2002). Conformational flexibility in sigma70 region 2 during transcription initiation. *J Biol Chem* 277(48)**,** 46433-46441.

Arthur, T.M., Anthony, L.C., and Burgess, R.R. (2000). Mutational analysis of beta '260-309, a sigma 70 binding site located on *Escherichia coli* core RNA polymerase. *J Biol Chem* 275(30)**,** 23113-23119.

Arthur, T.M., and Burgess, R.R. (1998). Localization of a sigma70 binding site on the N terminus of the *Escherichia coli* RNA polymerase beta' subunit. *J Biol Chem* 273(47)**,** 31381-31387.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37(Web Server issue)**,** W202-208.

Barik, S., Sureka, K., Mukherjee, P., Basu, J., and Kundu, M. (2010). RseA, the SigE specific anti-sigma factor of *Mycobacterium tuberculosis*, is inactivated by phosphorylation-dependent ClpC1P2 proteolysis. *Mol Microbiol* 75(3)**,** 592-606.

Bashyam, M.D., and Hasnain, S.E. (2004). The extracytoplasmic function sigma factors: role in bacterial pathogenesis. *Infect Genet Evol* 4(4)**,** 301-308.

Bayer-Santos, E., Lima, L.D., Ceseti, L.D., Ratagami, C.Y., de Santana, E.S., da Silva, A.M., et al. (2018). *Xanthomonas citri* T6SS mediates resistance to Dictyostelium predation and is regulated by an ECF sigma factor and cognate Ser/Thr kinase. *Environ Microbiol* 20(4)**,** 1562-1575.

Bibb, M.J., Domonkos, A., Chandra, G., and Buttner, M.J. (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(6)**,** 1033-1049.

Biville, F., Cwerman, H., Letoffe, S., Rossi, M.S., Drouet, V., Ghigo, J.M., et al. (2004). Haemophore-mediated signalling in *Serratia marcescens*: a new mode of regulation for an extra cytoplasmic function (ECF) sigma factor involved in haem acquisition. *Mol Microbiol* 53(4)**,** 1267-1277.

Bochner, B.R. (2003). New technologies to assess genotype-phenotype relationships. *Nat Rev Genet* 4(4)**,** 309-314.

Braun, V., and Mahren, S. (2005). Transmembrane transcriptional control (surface signalling) of the *Escherichia coli* Fec type. *FEMS Microbiol Rev* 29(4)**,** 673-684.

Burgess, R.R., and Anthony, L. (2001). How sigma docks to RNA polymerase and what sigma does. *Curr Opin Microbiol* 4(2)**,** 126-131.

Calamita, H., Ko, C., Tyagi, S., Yoshimatsu, T., Morrison, N.E., and Bishai, W.R. (2005). The *Mycobacterium tuberculosis* SigD sigma factor controls the expression of ribosome-associated gene products in stationary phase and is required for full virulence. *Cell Microbiol* 7(2)**,** 233-244.

Campagne, S., Damberger, F.F., Kaczmarczyk, A., Francez-Charlot, A., Allain, F.H., and Vorholt, J.A. (2012). Structural basis for sigma factor mimicry in the general stress response of *Alphaproteobacteria*. *Proc Natl Acad Sci U S A* 109(21)**,** E1405-1414.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., et al. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* 9(3)**,** 527-539.

Cao, M., Wang, T., Ye, R., and Helmann, J.D. (2002). Antibiotics that inhibit cell wall biosynthesis induce expression of the Bacillus subtilis sigma(W) and sigma(M) regulons. *Mol Microbiol* 45(5)**,** 1267-1276.

Catalano-Dupuy, D.L., Orecchia, M., Rial, D.V., and Ceccarelli, E.A. (2006). Reduction of the pea ferredoxin-NADP(H) reductase catalytic efficiency by the structuring of a carboxyl-terminal artificial metal binding site. *Biochemistry* 45(46)**,** 13899-13909.

Charlet, D., Mostowy, S., Alexander, D., Sit, L., Wiker, H.G., and Behr, M.A. (2005). Reduced expression of antigenic proteins MPB70 and MPB83 in *Mycobacterium bovis* BCG strains due to a start codon mutation in sigK. *Mol Microbiol* 56(5)**,** 1302-1313.

Craig, M.L., Tsodikov, O.V., McQuade, K.L., Schlax, P.E., Jr., Capp, M.W., Saecker, R.M., et al. (1998). DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J Mol Biol* 283(4)**,** 741-756.

D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem Sci* 28(12)**,** 655-662.

Dou, Y., Rutanhira, H., Chen, X., Mishra, A., Wang, C., and Fletcher, H.M. (2018). Role of extracytoplasmic function sigma factor PG1660 (RpoE) in the oxidative stress resistance regulatory network of *Porphyromonas gingivalis*. *Mol Oral Microbiol* 33(1)**,** 89-104.

Dufour, Y.S., Landick, R., and Donohue, T.J. (2008). Organization and evolution of the biological response to singlet oxygen stress. *J Mol Biol* 383(3)**,** 713-730.

Duque, E., Rodriguez-Herva, J.J., de la Torre, J., Dominguez-Cuevas, P., Munoz-Rojas, J., and Ramos, J.L. (2007). The RpoT regulon of *Pseudomonas putida* DOT-T1E and its role in stress endurance against solvents. *J Bacteriol* 189(1)**,** 207-219.

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

Eiamphungporn, W., and Helmann, J.D. (2008). The *Bacillus subtilis* sigma(M) regulon and its contribution to cell envelope stress responses. *Mol Microbiol* 67(4)**,** 830-848.

Ekeberg, M., Lovkvist, C., Lan, Y.H., Weigt, M., and Aurell, E. (2013). Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Physical Review E* 87(1).

Ellermeier, C.D., and Losick, R. (2006). Evidence for a novel protease governing regulated intramembrane proteolysis and resistance to antimicrobial peptides in *Bacillus subtilis*. *Genes Dev* 20(14)**,** 1911-1922.

Faure, L.M., Llamas, M.A., Bastiaansen, K.C., de Bentzmann, S., and Bigot, S. (2013). Phosphate starvation relayed by PhoB activates the expression of the *Pseudomonas aeruginosa* sigma vreI ECF factor and its target genes. *Microbiology* 159(Pt 7)**,** 1315-1327.

Feeney, M.A., Chandra, G., Findlay, K.C., Paget, M.S.B., and Buttner, M.J. (2017). Translational Control of the SigR-Directed Oxidative Stress Response in *Streptomyces* via IF3-Mediated Repression of a Noncanonical GTC Start Codon. *MBio* 8(3).

Felsenstein, J. (1989). Mathematics vs. Evolution: Mathematical Evolutionary Theory. *Science* 246(4932)**,** 941-942.

Feng, W.H., Mao, X.M., Liu, Z.H., and Li, Y.Q. (2011). The ECF sigma factor SigT regulates actinorhodin production in response to nitrogen stress in *Streptomyces coelicolor*. *Appl Microbiol Biotechnol* 92(5)**,** 1009-1021.

Fernandez-Martinez, L.T., and Bibb, M.J. (2014). Use of the meganuclease I-SceI of *Saccharomyces cerevisiae* to select for gene deletions in actinomycetes. *Sci Rep* 4**,** 7100.

Fernandez-Martinez, L.T., Gomez-Escribano, J.P., and Bibb, M.J. (2015). A relA-dependent regulatory cascade for auto-induction of microbisporicin production in *Microbispora corallina*. *Mol Microbiol* 97(3)**,** 502-514.

Ferreira, A.O., Myers, C.R., Gordon, J.S., Martin, G.B., Vencato, M., Collmer, A., et al. (2006). Whole-genome expression profiling defines the HrpL regulon of *Pseudomonas syringae pv. tomato* DC3000, allows de novo reconstruction of the Hrp cis clement, and identifies novel coregulated genes. *Mol Plant Microbe Interact* 19(11)**,** 1167-1179.

Flardh, K., and Buttner, M.J. (2009). Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* 7(1)**,** 36-49.

Flett, F., Mersinias, V., and Smith, C.P. (1997). High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* 155(2)**,** 223-229.

Foulston, L., and Bibb, M. (2011). Feed-forward regulation of microbisporicin biosynthesis in *Microbispora corallina*. *J Bacteriol* 193(12)**,** 3064-3071.

Francez-Charlot, A., Frunzke, J., Reichen, C., Ebneter, J.Z., Gourion, B., and Vorholt, J.A. (2009a). Sigma factor mimicry involved in regulation of general stress response. *Proceedings of the National Academy of Sciences of the United States of America* 106(9)**,** 3467-3472.

Francez-Charlot, A., Frunzke, J., Reichen, C., Ebneter, J.Z., Gourion, B., and Vorholt, J.A. (2009b). Sigma factor mimicry involved in regulation of general stress response. *Proc Natl Acad Sci U S A* 106(9)**,** 3467-3472.

Gaballa, A., Guariglia-Oropeza, V., Durr, F., Butcher, B.G., Chen, A.Y., Chandrangsu, P., et al. (2018a). Modulation of extracytoplasmic function (ECF) sigma factor promoter selectivity by spacer region sequence. *Nucleic Acids Res* 46(1)**,** 134-145.

Gaballa, A., Guariglia-Oropeza, V., Durr, F., Butcher, B.G., Chen, A.Y., Chandrangsu, P., et al. (2018b). Modulation of extracytoplasmic function (ECF) sigma factor promoter selectivity by spacer region sequence. *Nucleic Acids Res.* 46(1)**,** 134-145.

Gicquel, G., Bouffartigues, E., Bains, M., Oxaran, V., Rosay, T., Lesouhaitier, O., et al. (2013). The extra-cytoplasmic function sigma factor sigX modulates biofilm and virulence-related properties in *Pseudomonas aeruginosa*. *PLoS One* 8(11)**,** e80407.

Glazebrook, M.A., Doull, J.L., Stuttard, C., and Vining, L.C. (1990). Sporulation of *Streptomyces venezuelae* in submerged cultures. *J Gen Microbiol* 136(3)**,** 581-588.

Gomez-Santos, N., Perez, J., Sanchez-Sutil, M.C., Moraleda-Munoz, A., and Munoz-Dorado, J. (2011). CorE from *Myxococcus xanthus* is a copper-dependent RNA polymerase sigma factor. *Plos Genetics* 7(6).

Gordon, N.D., Ottaviano, G.L., Connell, S.E., Tobkin, G.V., Son, C.H., Shterental, S., et al. (2008). Secreted-protein response to sigmaU activity in *Streptomyces coelicolor*. *J Bacteriol* 190(3)**,** 894-904.

Gourion, B., Francez-Charlot, A., and Vorholt, J.A. (2008). PhyR is involved in the general stress response of *Methylobacterium extorquens* AM1. *J Bacteriol* 190(3)**,** 1027-1035.

Goyard, S., Sebo, P., D'Andria, O., Ladant, D., and Ullmann, A. (1993). *Bordetella pertussis* adenylate cyclase: a toxin with multiple talents. *Zentralbl Bakteriol* 278(2-3)**,** 326-333.

Grass, G., Grosse, C., and Nies, D.H. (2000). Regulation of the cnr cobalt and nickel resistance determinant from *Ralstonia sp* strain CH34. *J Bacteriol* 182(5)**,** 1390-1398.

Greenwell, R., Nam, T.W., and Donohue, T.J. (2011). Features of *Rhodobacter sphaeroides* ChrR required for stimuli to promote the dissociation of sigma(E)/ChrR complexes. *J Mol Biol* 407(4)**,** 477-491.

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

Grosse, C., Friedrich, S., and Nies, D.H. (2007). Contribution of extracytoplasmic function sigma factors to transition metal homeostasis in *Cupriavidus metallidurans* strain CH34. *J Mol Microbiol Biotechnol* 12(3-4)**,** 227-240.

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

Gruber, T.M., Markov, D., Sharp, M.M., Young, B.A., Lu, C.Z., Zhong, H.J., et al. (2001). Binding of the initiation factor sigma(70) to core RNA polymerase is a multistep process. *Molecular Cell* 8(1)**,** 21-31.

Grunden, A.M., Ray, R.M., Rosentel, J.K., Healy, F.G., and Shanmugam, K.T. (1996). Repression of the Escherichia coli modABCD (molybdate transport) operon by ModE. *J Bacteriol* 178(3)**,** 735-744.

Gunesekere, I.C., Kahler, C.M., Ryan, C.S., Snyder, L.A.S., Saunders, N.J., Rood, J.I., et al. (2006). ECF, an alternative sigma factor from *Neisseria gonorrhoeae*, controls expression of msrAB, which encodes methionine sulfoxide reductase. *J Bacteriol* 188(10)**,** 3463-3469.

Hahn, M.Y., Raman, S., Anaya, M., and Husson, R.N. (2005). The *Mycobacterium tuberculosis* extracytoplasmic-function sigma factor SigL regulates polyketide synthases and secreted or membrane proteins and is required for virulence. *J Bacteriol* 187(20)**,** 7062-7071.

Haines-Menges, B., Whitaker, W.B., and Boyd, E.F. (2014). Alternative aigma factor RpoE is important for vibrio parahaemolyticus cell envelope stress response and intestinal colonization. *Infect Immun* 82(9)**,** 3667-3677.

Hastie, J.L., Williams, K.B., and Ellermeier, C.D. (2013). The activity of sigmaV, an extracytoplasmic function sigma factor of *Bacillus subtilis*, is controlled by regulated proteolysis of the anti-sigma factor RsiV. *J Bacteriol* 195(14)**,** 3135-3144.

Hayes, R.P., Lewis, K.M., Xun, L., and Kang, C. (2013). Catalytic mechanism of 5-chlorohydroxyhydroquinone dehydrochlorinase from the YCII superfamily of largely unknown function. *J Biol Chem* 288(40)**,** 28447-28456.

Heinrich, J., and Wiegert, T. (2009). Regulated intramembrane proteolysis in the control of extracytoplasmic function sigma factors. *Res Microbiol* 160(9)**,** 696-703.

Helmann, J.D. (1995). Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res* 23(13)**,** 2351-2360.

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

Helmann, J.D. (2006). Deciphering a complex genetic regulatory network: the *Bacillus subtilis* sigmaW protein and intrinsic resistance to antimicrobial compounds. *Sci Prog* 89(Pt 3-4)**,** 243-266.

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

Herrou, J., Foreman, R., Fiebig, A., and Crosson, S. (2010). A structural model of anti-anti-sigma inhibition by a two-component receiver domain: the PhyR stress response regulator. *Mol Microbiol* 78(2)**,** 290-304.

Hong, H.J., Hutchings, M.I., Hill, L.M., and Buttner, M.J. (2005). The role of the novel Fem protein VanK in vancomycin resistance in *Streptomyces coelicolor*. *J Biol Chem* 280(13)**,** 13055-13061.

Hong, H.J., Paget, M.S.B., and Buttner, M.J. (2002). A signal transduction system in *Streptomyces coelicolor* that activates the expression of a putative cell wall glycan operon in response to vancomycin and other cell wall-specific antibiotics. *Mol Microbiol* 44(5)**,** 1199-1211.

Hu, P., Janga, S.C., Babu, M., Diaz-Mejia, J.J., Butland, G., Yang, W., et al. (2009). Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biol* 7(4)**,** e96.

Hu, Y., Kendall, S., Stoker, N.G., and Coates, A.R. (2004). The *Mycobacterium tuberculosis* sigJ gene controls sensitivity of the bacterium to hydrogen peroxide. *FEMS Microbiol Lett* 237(2)**,** 415-423.

Huang, X., Pinto, D., Fritz, G., and Mascher, T. (2015). Environmental sensing in *Actinobacteria*: a comprehensive survey on the signaling capacity of this phylum. *J Bacteriol* 197(15)**,** 2517-2535.

Humphreys, S., Stevenson, A., Bacon, A., Weinhardt, A.B., and Roberts, M. (1999). The alternative sigma factor, sigma(E), is critically important for the virulence of Salmonella typhimurium. *Infection and Immunity* 67(4)**,** 1560-1568.

Ishihama, A. (2000). Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* 54**,** 499-518.

Jaiswal, R.K., Prabha, T.S., Manjeera, G., and Gopal, B. (2013). *Mycobacterium tuberculosis* RsdA provides a conformational rationale for selective regulation of sigma-factor activity by proteolysis. *Nucleic Acids Res* 41(5)**,** 3414-3423.

Jiang, L., Liu, Y., Wang, P., Wen, Y., Song, Y., Chen, Z., et al. (2011). Inactivation of the extracytoplasmic function sigma factor Sig6 stimulates avermectin production in *Streptomyces avermitilis*. *Biotechnol Lett* 33(10)**,** 1955-1961.

Jogler, C., Waldmann, J., Huang, X.L., Jogler, M., Glockner, F.O., Mascher, T., et al. (2012). Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in planctomycetes by comparative genomics. *J Bacteriol* 194(23)**,** 6419-6430.

Joo, D.M., Ng, N., and Calendar, R. (1997). A sigma32 mutant with a single amino acid change in the highly conserved region 2.2 exhibits reduced core RNA polymerase affinity. *Proc Natl Acad Sci U S A* 94(10)**,** 4907-4912.

Kadner, R.J., and Watson, W.J. (1974). Methionine transport in *Escherichia coli*: physiological and genetic evidence for two uptake systems. *J Bacteriol* 119(2)**,** 401-409.

Kang, J.G., Paget, M.S., Seok, Y.J., Hahn, M.Y., Bae, J.B., Hahn, J.S., et al. (1999). RsrA, an anti-sigma factor regulated by redox change. *EMBO J* 18(15)**,** 4292-4298.

Kao, C.M., Karoonuthaisiri, N., Weaver, D., Vroom, J.A., Gai, S.A., Ho, M.-L., et al. (2018). A genomic island of *Streptomyces coelicolor* harbors the self-contained regulon of an ECF sigma factor. *bioRxiv***,** 247056.

Kappler, U., Friedrich, C.G., Truper, H.G., and Dahl, C. (2001). Evidence for two pathways of thiosulfate oxidation in *Starkeya novella* (formerly *Thiobacillus novellus*). *Arch Microbiol* 175(2)**,** 102-111.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* 95(10)**,** 5752-5756.

Karls, R.K., Guarner, J., McMurray, D.N., Birkness, K.A., and Quinn, F.D. (2006). Examination of *Mycobacterium tuberculosis* sigma factor mutants using low-dose aerosol infection of guinea pigs suggests a role for SigC in pathogenesis. *Microbiology* 152(Pt 6)**,** 1591-1600.

Katayama, A., Fujita, N., and Ishihama, A. (2000). Mapping of subunit-subunit contact surfaces on the beta' subunit of *Escherichia coli* RNA polymerase. *J Biol Chem* 275(5)**,** 3583-3592.

Kim, T.H., Kim, H.J., Park, J.S., Kim, Y., Kim, P., and Lee, H.S. (2005). Functional analysis of sigH expression in *Corynebacterium glutamicum*. *Biochem Biophys Res Commun* 331(4)**,** 1542-1547.

Kitagawa, M., Matsumura, Y., and Tsuchido, T. (2000). Small heat shock proteins, IbpA and IbpB, are involved in resistances to heat and superoxide stresses in Escherichia coli. *FEMS Microbiol Lett* 184(2)**,** 165-171.

Kogenaru, M., and Tans, S.J. (2014). An improved *Escherichia coli* strain to host gene regulatory networks involving both the AraC and LacI inducible transcription factors. *J Biol Eng* 8.

Kohler, C., Lourenco, R.F., Avelar, G.M., and Gomes, S.L. (2012). Extracytoplasmic function (ECF) sigma factor sigma(F) is involved in *Caulobacter crescentus* response to heavy metal stress. *BMC Microbiol* 12.

Kwon, H.J., Tirumalai, R., Landy, A., and Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* 276(5309)**,** 126-131.

Lamb, J.R., Michaud, W.A., Sikorski, R.S., and Hieter, P.A. (1994). Cdc16p, Cdc23p and Cdc27p form a complex essential for mitosis. *EMBO J* 13(18)**,** 4321-4328.

Lan, L., Deng, X., Zhou, J., and 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(9)**,** 976-987.

Leoni, L., Orsi, N., de Lorenzo, V., and Visca, P. (2000). Functional analysis of PvdS, an iron starvation sigma factor of *Pseudomonas aeruginosa*. *J Bacteriol* 182(6)**,** 1481-1491.

Li, W., Bottrill, A.R., Bibb, M.J., Buttner, M.J., Paget, M.S.B., and Kleanthous, C. (2003). The role of zinc in the disulphide stress-regulated anti-sigma factor RsrA from Streptomyces coelicolor. *Journal of Molecular Biology* 333(2)**,** 461-472.

Liu, G., Chater, K.F., Chandra, G., Niu, G., and Tan, H. (2013). Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol Mol Biol Rev* 77(1)**,** 112-143.

Lonetto, M., Gribskov, M., and Gross, C.A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* 174(12)**,** 3843-3849.

Lonetto, M.A., Brown, K.L., Rudd, K.E., and Buttner, M.J. (1994). Analysis of the *Streptomyces coelicolor* sigE gene reveals the existence of a subfamily of eubacterial RNA-polymerase sigma-factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci U S A* 91(16)**,** 7573-7577.

Lopez-Garcia, M.T., Yague, P., Gonzalez-Quinonez, N., Rioseras, B., and Manteca, A. (2018). The SCO4117 ECF sigma factor pleiotropically controls secondary metabolism and morphogenesis in *Streptomyces coelicolor*. *Front Microbiol* 9**,** 312.

Luo, S., Sun, D., Zhu, J., Chen, Z., Wen, Y., and Li, J. (2014). An extracytoplasmic function sigma factor, sigma(25), differentially regulates avermectin and oligomycin biosynthesis in *Streptomyces avermitilis*. *Appl Microbiol Biotechnol* 98(16)**,** 7097-7112.

Mackie, A.M., Hassan, K.A., Paulsen, I.T., and Tetu, S.G. (2014). Biolog Phenotype Microarrays for phenotypic characterization of microbial cells. *Methods Mol Biol* 1096**,** 123-130.

MacLellan, S.R., Wecke, T., and Helmann, J.D. (2008). A previously unidentified sigma factor and two accessory proteins regulate oxalate decarboxylase expression in *Bacillus subtilis*. *Mol Microbiol* 69(4)**,** 954-967.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., and MacNeil, T. (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* 111(1)**,** 61-68.

Mao, X.M., Sun, N., Wang, F., Luo, S., Zhou, Z., Feng, W.H., et al. (2013). Dual positive feedback regulation of protein degradation of an Extra-cytoplasmic function sigma factor for cell differentiation in *Streptomyces coelicolor*. *J Biol Chem* 288(43)**,** 31217-31228.

Marcos-Torres, F.J., Perez, J., Gomez-Santos, N., Moraleda-Munoz, A., and Munoz-Dorado, J. (2016). In depth analysis of the mechanism of action of metal-dependent sigma factors: characterization of CorE2 from *Myxococcus xanthus*. *Nucleic Acids Res* 44(12)**,** 5571-5584.

Mascher, T. (2013). Signaling diversity and evolution of extracytoplasmic function (ECF) sigma factors. *Curr Opin Microbiol* 16(2)**,** 148-155.

Masloboeva, N., Reutimann, L., Stiefel, P., Follador, R., Leimer, N., Hennecke, H., et al. (2012). Reactive oxygen species-inducible ECF sigma factors of *Bradyrhizobium japonicum*. *PLoS One* 7(8)**,** e43421.

McCracken, A., and Timms, P. (1999). Efficiency of transcription from promoter sequence variants in *Lactobacillus* is both strain and context dependent. *J Bacteriol* 181(20)**,** 6569-6572.

Mendez, R., Gutierrez, A., Reyes, J., and Marquez-Magana, L. (2012). The extracytoplasmic function sigma factor SigY is important for efficient maintenance of the Sp beta prophage that encodes sublancin in *Bacillus subtilis*. *DNA Cell Biol* 31(6)**,** 946-955.

Merighi, M., Majerczak, D.R., Stover, E.H., and Coplin, D.L. (2003). The HrpX/HrpY two-component system activates hrpS expression, the first step in the regulatory cascade controlling the Hrp regulon in *Pantoea stewartii subsp. stewartii*. *Mol Plant Microbe Interact* 16(3)**,** 238-248.

Mettrick, K.A., and Lamont, I.L. (2009). Different roles for anti-sigma factors in siderophore signalling pathways of *Pseudomonas aeruginosa*. *Mol Microbiol* 74(5)**,** 1257-1271.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol* 13(1)**,** 31-39.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296(5571)**,** 1285-1290.

Myronovskyi, M., Welle, E., Fedorenko, V., and Luzhetskyy, A. (2011). Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. *Appl Environ Microbiol* 77(15)**,** 5370-5383.

Nakunst, D., Larisch, C., Huser, A.T., Tauch, A., Puhler, A., and Kalinowski, J. (2007). The extracytoplasmic function-type sigma factor SigM of *Corynebacterium glutamicum* ATCC 13032 is involved in transcription of disulfide stress-related genes. *J Bacteriol* 189(13)**,** 4696-4707.

Ndamukong, I.C., Gee, J., and Smith, C.J. (2013). The extracytoplasmic function sigma factor EcfO protects Bacteroides fragilis against oxidative stress. *J Bacteriol* 195(1)**,** 145-155.

Newton, G.L., and Fahey, R.C. (2008). Regulation of mycothiol metabolism by sigma(R) and the thiol redox sensor anti-sigma factor RsrA. *Mol Microbiol* 68(4)**,** 805-809.

Nizan-Koren, R., Manulis, S., Mor, H., Iraki, N.M., and Barash, I. (2003). The regulatory cascade that activates the hrp regulon in *Erwinia herbicola pv. gypsophilae*. *Mol Plant Microbe Interact* 16(3)**,** 249-260.

O'Shea, E.K., Klemm, J.D., Kim, P.S., and Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254(5031)**,** 539-544.

Okrent, R.A., Halgren, A.B., Azevedo, M.D., Chang, J.H., Mills, D.I., Maselko, M., et al. (2014). Negative regulation of germination-arrest factor production in *Pseudomonas fluorescens* WH6 by a putative extracytoplasmic function sigma factor. *Microbiology* 160(Pt 11)**,** 2432-2442.

Paget, M.S. (2015). Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules* 5(3)**,** 1245-1265.

Paget, M.S., Leibovitz, E., and Buttner, M.J. (1999a). A putative two-component signal transduction system regulates sigmaE, a sigma factor required for normal cell wall integrity in Streptomyces coelicolor A3(2). *Mol Microbiol* 33(1)**,** 97-107.

Paget, M.S., Molle, V., Cohen, G., Aharonowitz, Y., and Buttner, M.J. (2001). Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol Microbiol* 42(4)**,** 1007-1020.

Paget, M.S.B., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. (1999b). Evidence that the extracytoplasmic function sigma factor sigma(E) is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* 181(1)**,** 204-211.

Paget, M.S.B., Kang, J.G., Roe, J.H., and Buttner, M.J. (1998). sigma(R), an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in Streptomyces coelicolor A3(2). *Embo Journal* 17(19)**,** 5776-5782.

Pinto, D., and Mascher, T. (2016a). (Actino)Bacterial "intelligence": using comparative genomics to unravel the information processing capacities of microbes. *Current Genetics* 62(3)**,** 487-498.

Pinto, D., and Mascher, T. (2016b). The ECF Classification: A Phylogenetic Reflection of the Regulatory Diversity in the Extracytoplasmic Function σ Factor Protein Family.

*Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria***,** 64-96.

Pinto, D., Vecchione, S., Wu, H., Mauri, M., Mascher, T., and Fritz, G. (2018). Engineering orthogonal synthetic timer circuits based on extracytoplasmic function sigma factors. *Nucleic Acids Res*.

Popp, P.F., Dotzler, M., Radeck, J., Bartels, J., and Mascher, T. (2017). The Bacillus BioBrick Box 2.0: expanding the genetic toolbox for the standardized work with *Bacillus subtilis*. *Sci Rep* 7(1)**,** 15058.

Radeck, J., Kraft, K., Bartels, J., Cikovic, T., Durr, F., Emenegger, J., et al. (2013). The Bacillus BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. *J Biol Eng* 7(1)**,** 29.

Rhodius, V.A., Segall-Shapiro, T.H., Sharon, B.D., Ghodasara, A., Orlova, E., Tabakh, H., et al. (2013). Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Molecular Systems Biology* 9.

Sabelnikov, A.G., Greenberg, B., and Lacks, S.A. (1995). An extended -10 promoter alone directs transcription of the DpnII operon of *Streptococcus pneumoniae*. *J Mol Biol* 250(2)**,** 144-155.

Saecker, R.M., Tsodikov, O.V., McQuade, K.L., Schlax, P.E., Jr., Capp, M.W., and Record, M.T., Jr. (2002). Kinetic studies and structural models of the association of *E. coli* sigma(70) RNA polymerase with the lambdaP(R) promoter: large scale conformational changes in forming the kinetically significant intermediates. *J Mol Biol* 319(3)**,** 649-671.

Sauviac, L., Philippe, H., Phok, K., and Bruand, C. (2007). An extracytoplasmic function sigma factor acts as a general stress response regulator in *Sinorhizobium meliloti*. *J Bacteriol* 189(11)**,** 4204-4216.

Sen, R., Nagai, H., and Shimamoto, N. (2000). Polymerase arrest at the lambda P-R promoter during transcription initiation. *J Biol Chem* 275(15)**,** 10899-10904.

Sharp, M.M., Chan, C.L., Lu, C.Z., Marr, M.T., Nechaev, S., Merritt, E.W., et al. (1999). The interface of sigma with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes Dev* 13(22)**,** 3015-3026.

Shea, A., Wolcott, M., Daefler, S., and Rozak, D.A. (2012). Biolog phenotype microarrays. *Methods Mol Biol* 881**,** 331-373.

Sherwood, E.J., and Bibb, M.J. (2013). The antibiotic planosporicin coordinates its own production in the actinomycete *Planomonospora alba*. *Proc Natl Acad Sci U S A* 110(27)**,** E2500-E2509.

Shukla, J., Gupta, R., Thakur, K.G., Gokhale, R., and Gopal, B. (2014). Structural basis for the redox sensitivity of the *Mycobacterium tuberculosis* SigK-RskA sigma-anti-sigma complex. *Acta Crystallogr D Biol Crystallogr* 70(Pt 4)**,** 1026-1036.

Shuler, M.F., Tatti, K.M., Wade, K.H., and Moran, C.P., Jr. (1995). A single amino acid substitution in sigma E affects its ability to bind core RNA polymerase. *J Bacteriol* 177(13)**,** 3687-3694.

Staron, A., Sofia, H.J., Dietrich, S., Ulrich, L.E., Liesegang, H., and Mascher, T. (2009). The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. *Mol Microbiol* 74(3)**,** 557-581.

Tatti, K.M., and Moran, C.P., Jr. (1995). sigma E changed to sigma B specificity by amino acid substitutions in its -10 binding region. *J Bacteriol* 177(22)**,** 6506-6509.

Tettmann, B., Dotsch, A., Armant, O., Fjell, C.D., and 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(16)**,** 4911-4919.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22)**,** 4673-4680.

Tibazarwa, C., Wuertz, S., Mergeay, M., Wyns, L., and van Der Lelie, D. (2000). Regulation of the cnr cobalt and nickel resistance determinant of *Ralstonia eutropha (Alcaligenes eutrophus)* CH34. *J Bacteriol* 182(5)**,** 1399-1409.

Wecke, T., Halang, P., Staron, A., Dufour, Y.S., Donohue, T.J., and Mascher, T. (2012). Extracytoplasmic function sigma factors of the widely distributed group ECF41 contain a fused regulatory domain. *Microbiologyopen* 1(2)**,** 194-213.

Weigt, M., White, R.A., Szurmant, H., Hoch, J.A., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A* 106(1)**,** 67-72.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012). Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioengineered* 3(1)**,** 38-43.

Wilderman, P.J., Vasil, A.I., Johnson, Z., Wilson, M.J., Cunliffe, H.E., Lamont, I.L., et al. (2001). Characterization of an endoprotease (PrpL) encoded by a PvdS-regulated gene in Pseudomonas aeruginosa. *Infect Immun* 69(9)**,** 5385-5394.

Willis, M.A., Song, F., Zhuang, Z., Krajewski, W., Chalamasetty, V.R., Reddy, P., et al. (2005). Structure of YciI from *Haemophilus influenzae* (HI0828) reveals a ferredoxin-like alpha/beta-fold with a histidine/aspartate centered catalytic site. *Proteins* 59(3)**,** 648-652.

Wood, L.F., and Ohman, D.E. (2009). Use of cell wall stress to characterize sigma(22) (AlgT/U) activation by regulated proteolysis and its regulon in *Pseudomonas aeruginosa*. *Mol Microbiol* 72(1)**,** 183-201.

Worthen, D.B. (2008). *Streptomyces* in nature and medicine: The antibiotic makers. *Journal of the History of Medicine and Allied Sciences* 63(2)**,** 273-274.

Wosten, M.M. (1998). Eubacterial sigma-factors. *FEMS Microbiol Rev* 22(3)**,** 127-150.

Wu, Q.L., Kong, D., Lam, K., and Husson, R.N. (1997). A mycobacterial extracytoplasmic function sigma factor involved in survival following stress. *J Bacteriol* 179(9)**,** 2922-2929.

Xu, J., Chiang, H.C., Bjursell, M.K., and Gordon, J.I. (2004). Message from a human gut symbiont: sensitivity is a prerequisite for sharing. *Trends Microbiol* 12(1)**,** 21-28.

Zdanowski, K., Doughty, P., Jakimowicz, P., O'Hara, L., Buttner, M.J., Paget, M.S., et al. (2006). Assignment of the zinc ligands in RsrA, a redox-sensing ZAS protein from *Streptomyces coelicolor*. *Biochemistry* 45(27)**,** 8294-8300.

Appendices

Table A 1. Oligonucleotides used in this study

Name	Sequence (5′ → 3′)
TM4364	AAATTAAGCTTCCGAGCCGCCGCCCGCCGGG
TM4365	AAATTGGATCCTCCGTGACGGTGGGCGGCACGCCG
TM4366	AAATTGGATCCCCCGCCCGCGGCCCCCGCTCCGC
TM4367	AAATTGGTACCCGGGCAGCGGGGCCGGTGCGGC
TM4368	AAATTAAGCTTAGATGGACCCGGCCGTCGGGGGCCG
TM4369	AAATTGGATCCACGTCCGCCGCGGACTCGTGCAGCAC
TM4370	AAATTGGATCCGCCGGAGGGGGCGGGCGACGGCTC
TM4371	AAATTGGTACCAAGGACTCGAGGGCGGCCTTGGTGAAGACC
TM4452	AAATTAAGCTTCCAAGCCCGACCTGCGGTCGCTGCC
TM4453	AAATTGAATTCGCTTGCCTCGCTGCTCCTCGGCTCTGG
TM4454	AAATTGAATTCGGGGAGCTTCCGCGGGCCCAGGTG
TM4455	AAATTGGTACCAGGGGCCCTCGCCGAAGAAGCCGGGG
TM4606	GATGGTCATCGACGTCGACGGCTAC
TM4607	CGATGAACCACTGCCTGGCCCAGATC
TM4608	GAGTACCTGGCCAGCTTCTACCTG
TM4609	TGAGAAGGAGCGACGAGTACTGTG
TM4610	CTGATCGAGGTCCGGTCCAAGGAGG
TM4611	CAGGGACGAACGCACGCGTCTGTG
TM5631	AAATTCATATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACAT CGACTACAAGGACGATGACGACAAGACCCCTTCGATGGACGAGGTCCTGCTC
TM4056	AAATTAAGCTTTCAGCGCCCCTCCTCGTCGCCGCCGCCGC
TM5632	AAATTCATATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACAT CGACTACAAGGACGATGACGACAAGACCGGCCCCGCCGAGGTCGAGGACCT
TM4059	AAATTAAGCTTTCAGCCCAACCGCCGCGCCCGCTCGGTG
TM5633	AAATTCATATGGACTACAAGGACCACGACGGCGACTACAAGGACCACGACAT CGACTACAAGGACGATGACGACAAGCCCGAGCCGGACACCCGCCGTACGGT
TM4062	AAATTAAGCTTTCAGGGCCGGGGCTTGAGCTGACGCAAGG
TM5631	AAATTAAGCTTTCAGTCCTCCCCCGCGGGCAGCGCG
TM4056	AAATTAAGCTTTCAGTCCTCGGGCACCAGCCGGTGG
TM5632	AAATTAAGCTTTCACCCCCGGGGCAGGGTGCGGCTG
TM4059	AAATTAAGCTTTCAGAGCAGCGCGAGCCCGGCCGCGG
TM6090	AAATTTCTAGACGGCTCCCGCTTCGCCACCGACGG
TM6091	AAATTGGTACCGGTGTCTCTCCTCGGTTCCGTGC
TM6092	AAATTTCTAGAAGCGGCCGGCCGGGCCGGAAG
TM6093	AAATTGGTACC CCGTGGGGTGACTGCCCGCCTC
TM4718	AAATTTCTAGAGTGGGCGGTCCGTCCCCTCGAATC

TM4719	AAATTGGTACCCTCGGGTGTCCTCCCCGTCCCG
TM4732	ACCGCGATGCTGTTGTGGGC
TM4733	GCGACCTTGCCCCTCCAACG
TM4734	CGATCACGGCACGATCATCGTG
TM4735	CGGTCGAGGCTGAACGCCCAC
TM4530	GTGCTGCAAGGCGATTAAGTTG
TM4531	CACAGGAAACAGCTATGACATG
TM5590	AAATTGAAGACAAAATGGATTATAAGGATCATGATGGTGATTATAAGGATCAT GATATCGACTACAAAGACGATGACGACAAGATGGCAGCAACCGATATTAGCG
TM5591	AAATTGAAGACAACGAATTATGCAGGCAGACCGCTACAT
TM5714	AAATTGAAGACAAAAGCTTAACCACGCGGAACTTCAAACGGAAC
TM5715	AAATTGAAGACAAAAGCTTATGCGGTCGGCATCTGCTCTGCCAG
TM5716	AAATTGAAGACAAAAGCTTATTCACCGCCACCCAGCTGCTGTGC
TM5717	AAATTGAAGACAAAAGCTTAGTGGCGCAGTGCACTATCCTGCTG
TM5611	CTTTGAGTGAGCTGATACCG
TM5612	CCACCTGACGTCTAAGAAAC
TM5613	GTTTACCCGCCAATATATCCTG
TM5614	CAAAGGAGATCCTGATCTGAC
TM5619	AAATT GAAGACAAAATGGATTATAAGGATCATG
TM5620	AAATTGAAGACAAAAGCTTATGCAGGCAGACCGCTACAT
TM5670	AAATTGAAGACAAAATGGCAGCAACCGATATTAG
TM5671	TGCCAGACGTGCATGACG
TM5672	CGTCATGCACGTCTGGCAGCAGTTCTGGAAGTTCTGTATC
TM5673	CGTCATGCACGTCTGGCAAGCGTTCTGGAAGCACTGTATCTGATTTTTAACGA AGGC
TM5674	CGTCATGCACGTCTGGCAAGCGTTCTGGAAGTTCTGTATCTGGCATTTAACGA AGGCTATGCAGCC
TM5675	CGTCATGCACGTCTGGCAAGCGTTCTGGAAGTTCTGTATCTGATTTTTGCAGAA GGCTATGCAGCCAGTG
TM5676	CTGGGCATCATCACCATCAC
TM5677	GTGATGGTGATGATGCCCAGGCACCGGCTCTGTGTCATGAAGC
TM5856	AAATTGAAGACGCTGCCAGACGTGCATGACGTTC
TM5857	AAATTGAAGACCTGGCAGCAGTTCTGGAAGCACTGTATCTGGCATTTGCAGAA GGCTATGCAGCCAGTGATGGTG
TM3707	AATTCGCGGCCGCTTCTAGAGCGCCGGAAAAAAGCTCGGCCGGTGATGTCGA GAATCCGCGTCCTGCTCCGTCCCCGGGGAGATA
TM3708	CTAGTATCTCCCCGGGGACGGAGCAGGACGCGGATTCTCGACATCACCGGCCG AGCTTTTTTCCGGCGCTCTAGAAGCGGCCGCG
TM3709	AATTCGCGGCCGCTTCTAGAGAAGCACATTAAAATATAACGTTTTTTGTCGAA ACTCCGTTTTCTCTCTCGTCATATAAGTGAATA
TM3710	CTAGTATTCACTTATATGACGAGAGAGAAAACGGAGTTTCGACAAAAAACGTT ATATTTTAATGTGCTT CTCTAGAAGCGGCCGCG
TM3711	AATTCGCGGCCGCTTCTAGAGAACTTTCGAAGATCGCCGTCGGGGATGTCGAG GACGCCCCGCCGGCTCCGACCGAGGGGTGA TA
TM3712	CTAGTATCACCCCTCGGTCGGAGCCGGCGGGGCGTCCTCGACATCCCCGACGG CGATCTTCGAAAGTT CTCTAGAAGCGGCCGCG

TM3713	AATTCGCGGCCGCTTCTAGAGAAGCACATTAAAATATAACGTTTTTTGTCGAA ACTCCGTTTTCTCTCTCGACATATAAGTGAA TA
TM3714	CTAGTATTCACTTATATGTCGAGAGAGAAAACGGAGTTTCGACAAAAAACGTT ATATTTTAATGTGCTT CTCTAGAAGCGGCCGCG
TM3715	AATTCGCGGCCGCTTCTAGAGATCGAACTGGCCCGCCGCTTCCGCAGGATCGT CGGCGACGGCGAGAGCGTGGTCCAGCAGG TA
TM3716	CTAGTACCTGCTGGACCACGCTCTCGCCGTCGCCGACGATCCTGCGGAAGCGG CGGGCCAGTTCGAT CTCTAGAAGCGGCCGCG
TM3717	AATTCGCGGCCGCTTCTAGAGAAGCACATTAAAATATAACGTTTTTGGATCGT CGGCGACGGCGAGAGCGTG ATATAAGTGAA TA
TM3718	CTAGTATTCACTTATATCACGCTCTCGCCGTCGCCGACGATCCAAAAACGTTAT ATTTTAATGTGCTT CTCTAGAAGCGGCCGCG
TM3731	AAATTGAATTCGCGGCCGCtTCTAGAAAGGAGGTGGCCGGCATGACAGGTCCA GCTGAAGTTGAAG
TM3732	AAATTCTGCAGCGGCCGCTACTAGTATTAACCGGTTCAACCCAAACGACGAGC ACGTTCTG
TM3735	AAATTGAATTCGCGGCCGCtTCTAGAAAGGAGGTGGCCGGCATGCCAGAACCT GATACACGTCGTA
TM3736	AAATTCTGCAGCGGCCGCTACTAGTATTAACCGGTTCATGGACGTGGTTTAAG TTGACGCAA
TM6134	AAATTCCATGGGCCATCATCATCATCATCATCATCATCATCACGTTGCACAGC AGGCAGATATTAC
TM6135	AAATTGGATCCTTATGCACCTTTCAGTGTGGCGG
TM6136	AAATTCATATGAAGTATCTGGTGATGGTGCAAG
TM6137	AAATTGGTACCTTATTTTTCGAACTGCGGGTGGCTCCA

Table A 2. Upregulated genes (fold $\geqslant$ 2, P $\leqslant$ 0.005, reads $\geqslant$ 300) in ECF42 overexpression *S. venezuelae* strains compared to ECF42 triple deletion mutant.

Locus	**log2-fold Change**	***P*-value**	**Description**
S. venezuelae* overexpressing *sven_0747			
SVEN_0747	18.79	8.62E-54	Putative RNA polymerase sigma factor
SVEN_0748	4.03	1.12E-90	Hypothetical protein
SVEN_4522	3.5	0	Putative integral membrane protein
SVEN_4523	1.67	4.63E-62	Oligopeptide binding protein
SVEN_3779	1.49	0.0073	Transcriptional regulator, MarR family
SVEN_t8	1.31	4.78E-12	tRNA
SVEN_t59	1.15	2.26E-07	tRNA
SVEN_t46	1.11	5.60E-09	tRNA
SVEN_4219	1.08	1.07E-10	Putative integral membrane protein
SVEN_t7	1.07	5.90E-08	tRNA
SVEN_t37	1	0.0033	tRNA
SVEN_t37	1	0.0033	tRNA
S. venezuelae* overexpressing *sven_4377			
SVEN_4377	15.97	8.4E-107	RNA polymerase sigma-70 factor, ECF subfamily
SVEN_3806	9.33	0	Hypothetical protein
SVEN_3805	6.15	0	GCN5-related N-acetyltransferase

SVEN_4376	4.39	0	DGPFAETKE family protein
SVEN_4522	3.16	0	Putative integral membrane protein
SVEN_6046	2.85	5.52E-25	Molybdate-binding domain of ModE
SVEN_6659	1.74	5.11E-06	Xylulose kinase
SVEN_4040	1.72	3.63E-19	Hypothetical protein
SVEN_3036	1.68	2.49E-22	Hypothetical protein
SVEN_4476	1.59	5.09E-11	Putative secreted lipase
SVEN_3356	1.55	7.37E-17	Hypothetical protein
SVEN_3924	1.55	4.71E-22	Putative integral membrane protein
SVEN_3126	1.54	1.99E-07	Hypothetical protein
SVEN_6100	1.53	2.42E-19	Hypothetical protein
SVEN_4523	1.52	3.21E-54	Oligopeptide binding protein
SVEN_4575	1.5	2.09E-05	Putative RNA polymerase ECF-subfamily sigma factor
SVEN_3534	1.5	7.29E-16	ABC-type multidrug transport system, ATPase component
SVEN_6957	1.45	6.37E-32	MarR-family transcriptional regulator
SVEN_3714	1.44	5.88E-34	Cytochrome d ubiquinol oxidase subunit II
SVEN_0961	1.44	9.71E-12	Hypothetical protein
SVEN_2147	1.41	2.33E-15	HNH endonuclease
SVEN_3526	1.4	3.14E-19	Hypothetical protein
SVEN_0718	1.39	1.14E-16	Possible sugar kinase
SVEN_2257	1.37	1.59E-13	Ferredoxin reductase
SVEN_4229	1.3	3.05E-16	RNA polymerase ECF-subfamily sigma factor
SVEN_3531	1.28	6.93E-14	Universal stress protein family
SVEN_1867	1.26	8.87E-27	Arsenate reductase
SVEN_0260	1.26	3.76E-14	GCN5-related N-acetyltransferase
SVEN_2571	1.26	6.01E-13	Desferrioxamine E biosynthesis protein DesB
SVEN_5883	1.25	7.15E-13	Glyoxalase family protein
SVEN_1017	1.24	1.53E-12	Hypothetical protein
SVEN_0690	1.22	1.53E-10	Putative lyase
SVEN_0322	1.19	3.26E-09	Hypothetical protein
SVEN_1503	1.18	2.18E-13	Penicillin amidase family protein
SVEN_6081	1.18	0.000184	LPXTG-motif cell wall anchor domain protein
SVEN_3715	1.17	1.17E-25	Transport ATP-binding protein CydCD
SVEN_2672	1.15	3.49E-17	ATP-dependent Clp protease adaptor protein ClpS
SVEN_4981	1.15	1.71E-20	Transcriptional regulator, MarR family
SVEN_0273	1.15	3.15E-10	Putative regulatory protein
SVEN_4614	1.13	5.28E-16	Hypothetical protein
SVEN_2241	1.12	3.29E-23	Organic hydroperoxide resistance transcriptional regulator
SVEN_3697	1.12	3.08E-09	Hypothetical protein
SVEN_2058	1.12	2.75E-06	Putative integral membrane protein
SVEN_1504	1.11	2.67E-12	Putative acetyltransferase
SVEN_2406	1.08	9.16E-18	Putative membrane protein
SVEN_2487	1.08	3.73E-07	Hypothetical protein
SVEN_1911	1.08	5.27E-08	Hypothetical protein
SVEN_4062	1.07	2.08E-10	Hypothetical protein

SVEN_0838	1.07	2.77E-16	Hypothetical protein
SVEN_2188	1.06	8.99E-06	LrgA-associated membrane protein LrgB
SVEN_1241	1.05	8.14E-15	Putative ferredoxin
SVEN_2607	1.05	2.56E-09	Transcriptional regulator, IclR family
SVEN_1177	1.05	2.02E-15	Prolyl endopeptidase
SVEN_0837	1.05	5.49E-12	Putative membrane protein
SVEN_1787	1.04	2.96E-15	Hypothetical protein
SVEN_3532	1.04	3.14E-07	Lyase
SVEN_4689	1.04	1.71E-11	Hypothetical protein
SVEN_1219	1.03	4.79E-25	Iron utilization protein
SVEN_1401	1.03	2.39E-16	Ribosyl nicotinamide transporter, PnuC
SVEN_4751	1.03	1.65E-10	Hypothetical protein
SVEN_0001	1.03	0.000393	Hypothetical protein
SVEN_3361	1.03	1.72E-13	Hypothetical protein
SVEN_4295	1.02	3.39E-08	Hypothetical protein
SVEN_3415	1.01	7.20E-08	Transcriptional regulator, GntR family
S. venezuelae* overexpressing *sven_7131			
SVEN_4522	3.86	0	Putative integral membrane protein
SVEN_2864	2.28	1.75E-05	Phage major capsid protein
SVEN_1613	1.98	1.76E-08	Aspartate or tyrosine or aromatic aminotransferase
SVEN_2991	1.88	1.14E-08	Hypothetical protein
SVEN_2992	1.66	2.31E-18	L-lysine permease
SVEN_4523	1.59	4.16E-59	Oligopeptide binding protein
SVEN_1612	1.58	6.47E-17	Transcriptional regulator, HxlR family
SVEN_4977	1.42	6.01E-16	ATP-dependent protease La Type I
SVEN_1615	1.32	4.13E-23	Hypothetical protein
SVEN_5837	1.3	1.28E-07	Secreted protein
SVEN_1614	1.29	3.01E-14	Hypothetical protein
SVEN_3169	1.19	1.94E-14	Hsp18_3
SVEN_5838	1.18	1.44E-05	Hypothetical protein
SVEN_5836	1.17	8.44E-08	Hypothetical protein
SVEN_5839	1.11	7.25E-07	Hypothetical protein
SVEN_3298	1.05	5.4E-17	Hypothetical protein
SVEN_1595	1.05	1.08E-14	Integral membrane protein TerC
SVEN_3297	1.04	1.21E-15	Hypothetical protein
SVEN_0065	1.01	2.24E-09	Extracellular alkaline serine protease 2
SVEN_5840	1	4.73E-06	Hypothetical protein

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Thorsten Mascher for the continuous support on both living and studying in Germany. I would like to thank you for giving me such a chance to come to Munich (amazing place!) where I have realized one my wishes - experiencing the Octoberfest and tasting the best beers and sausages in the world. Thank you for your encouragement and patient advice when I was feeling lost on my project. Thank you for supporting me to do some research on side project of my own personal interests. I have to say you are really a very nice and inspiring supervisor.

The second person I would like to thank is Dr. Daniela Pinto. Like most of us say, you are not only a scientist but also an artist. I was shocked the first time I saw your lab book. The writing just like it was printed – amazing! Thank you for teaching me many skills for both working on the bench and on the computer. You are always patient and professional to give me suggestions and answers when I have problems and ask stupid questions. Thank you so much for your support on the writing of my thesis and our manuscript. I am sure I will keep asking you some stupid questions no matter where I stay in the future, because you are a problem solver.

In addition, I would like thank Chong Fang, Xiaoluo Huang and Dayane Araujo, who help me a lot to adapt the lab environment in the beginning when I started to work in the lab. Great gratitude also goes to all other people in our group (Stephan Mauersberger, Diana Wolf, Ainhoa, Julia Bartels, Jara Radeck, Philipp Popp, Fransiska Dürr, Qian …), who helped me with my living in Germany or scientific work in the lab. Thanks also go to my friends Wenlong Zhang, Yuehui Tian and Duorong Xu, with whom I had a great and unforgettable time in Munich.

I also greatly acknowledge the financial support of China Scholarship Council (CSC). Thanks to all the people that contribute to the completion of this thesis.

Finally, a special thanks to my parents, who respect and support my choices all the time. I would also like to thank my beloved wife Juan Huang - thank you for understanding and supporting me for everything. Thanks for all your support and encouragement.

Curriculum vitae

Place of birth:

China, September 7th, 1987.

Education:

10.2014 - Present	Ph. D. candidate at Faculty of Biology, Ludwig-Maximilians-University Munich (LMU), Munich, Germany.
10.2015 - Present	Guest Ph. D. student at Institut für Mikrobiologie, Technische Universität Dresden (TU Dresden), Dresden, Germany.
09.2011 -06.2014.	Master student at Institute of Preventive Veterinary Medicine, Sichuan Agricultural University, Chengdu, China
09.2007 -06.2011.	Bachelor, Animal quarantine inspection, Sichuan Agricultural University, Chengdu, China

Conferences:

- **Qiang Liu**, Daniela Pinto and Thorsten Mascher. Characterization of a novel group of C-terminal extension containing extracytoplasmic function (ECF) ECF42 sigma factor. Conference "How microorganisms view their world", 23-25 September 2018, Marburg, Germany. Poster.

- **Qiang Liu**, Daniela Pinto and Thorsten Mascher. Regulation mechanism and physiological role of novel Extracytoplasmic Function (ECF) sigma factors from *Streptomyces venezuelae*. Annual Conference 2016 of the Association for General and Applied Microbiology (VAAM), 13-16 March 2016, Jena, Germany. Poster.

- Daniela Pinto, Franziska Dürr, **Qiang Liu** and Thorsten Mascher. Optimization of an Extracytoplasmic Function sigma factor based switch. CAS Conference Synthetic Biology II, 27-29 July 2015, Munich, Germany. Poster.

Publications:

- **Qiang Liu**, Daniela Pinto and Thorsten Mascher. (2018). Characterization of the widely distributed novel ECF42 group of extracytoplasmic function σ factors in *Streptomyces venezuelae*. *Journal of Bacteriology*. DOI: 10.1128/JB.00437-18.

- Daniela Pinto, Franziska Dürr, Dayane Araújo, **Qiang Liu**, and Thorsten Mascher. Modulation of the behavior of an Extracytoplasmic Function sigma factor (ECF)-based genetic switch. Research article close to submission.

- **Qiang Liu***, Franziska Dürr*, Hao Wu*, Daniela Pinto, Georg Fritz and Thorsten Mascher. On the role of C-terminal extensions in the control of ECF σ factor activity. Research article in preparation. (*co-first author)

Email: liuqiangdyy@163.com