U.PORTO

3º CICLO

DOUTORAMENTO EM BIOLOGIA MOLECULAR E CELULAR

Role and molecular mechanisms of a novel sigma factor

Rute Patrícia Martins Barbosa de Oliveira

D

2020

U.PORTO

Rute Patrícia Martins Barbosa de Oliveira.

Role and molecular mechanisms of a novel sigma factor

Role and molecular mechanisms of a novel sigma factor

Rute Patrícia Martins Barbosa de Oliveira

D.ICBAS **2020**

SEDE ADMINISTRATIVA

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

FACULDADE DE CIÊNCIAS

RUTE PATRÍCIA MARTINS BARBOSA DE OLIVEIRA

Role and molecular mechanisms of a novel sigma factor

Tese de Candidatura ao grau de Doutor em Biologia Molecular e Celular;

Programa Doutoral da Universidade do Porto (Instituto de Ciências Biomédicas de Abel Salazar e Faculdade de Ciências)

Orientador:

Marta Vaz Mendes

Professor Auxiliar Convidado
Instituto de Ciências Biomédicas de Abel Salazar, Universidade do Porto, Portugal

Investigador Auxiliar
i3S - Instituto de Investigação e Inovação em Saúde, IBMC - Instituto de Biologia Molecular e Celular da Universidade do Porto, Portugal

Co-orientador:

Mark Jeremy Buttner

Head, Department of Molecular Microbiology
John Innes Centre, Nowich, United Kingdom

Este trabalho foi financiado pela

Fundação para a Ciência e a Tecnologia

SFRH/BD/107862/2015

This work was partially funded by National Funds through FCT- Fundação para a Ciência e a Tecnologia, I.P., under the project ERA-IB-2/0001/2015. It was further supported by FEDER - Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation (POCI), Portugal 2020; and by Portuguese funds through FCT Fundação para a Ciência e a Tecnologia, I.P/Ministério da Ciência, Tecnologia e Ensino Superior POCI-01-0145-FEDER-007274 and NORTE-01-0145-FEDER-000012- Structured program on bioengineered therapies for infectious diseases and tissue regeneration, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). Part of the work from Chapter II was funded of the European Molecular Biology Organisation award.

‘Begin at the beginning,’ the King said, very gravely, ‘and go on till you come to the end: then stop.’

Lewis Carroll, Alice in Wonderland

Agradecimentos/Acknowledgements

Before starting, I would to like to acknowledge the many people that have contributed to the success of this work. My sincere gratitude to all of you that guided, encouraged, or even warmly supported me and made it all a lot more fun.

À Professora Paula Tamagnini, envio um agradecimento especial por me ter recebido no seu grupo, e por ter tido sempre a porta do seu gabinete aberta para me receber. Por tudo o que me ensinou, e que levo agora comigo na minha jornada científica.

Marta, obrigada por teres aceite orientar este projeto. Por teres tido uma postura *hands-on* e me teres dado as ferramentas necessárias para me tornar uma investigadora autónoma.

To my co-supervisor Mark Buttner, I am deeply grateful that you accepted to be my supervisor and hosted me in the outstanding Molecular Microbiology department. To work at the core of *Streptomyces* research was essential for the project, but more importantly, it was a constant inspiration and kept my motivation up at all steps of the way. I thank you for your kindness during my time in Norwich, and for teaching me so much more than you will ever understand.

Ao Professor Cláudio Sunkel, diretor do programa doutoral em Biologia Molecular e Celular, agradeço o apoio logístico que me abriu tantas oportunidades durante estes 4 anos. Pelo aconselhamento, e todo o tempo disponibilizado, especialmente na última fase do meu doutoramento. Agradeço ainda ao Professor Vítor Costa e à Professora Salomé Gomes por terem aceite constituir a minha comissão de acompanhamento científico, e pelo tempo disponibilizado para acompanhar e discutir este trabalho.

To my dearest friends from the MolMicro family, I am grateful for the loving and caring environment that you built during my stay in Norwich. Smelly Kelley, thank you for gravitating my walk through this process, even from the far side of the world. To you I am deeply grateful for hosting me in Norwich, for making it so simple and easy to change, for sharing every happy and sad moments. Mighty Matt, I could never have enough words to express my gratitude for all that you have done since all this started. You are such an amazing scientist and I thank you for sharing part of that with me. Thank you for being on call for all these years. This, truly, would not have been possible without your wise technical and spiritual advice. Felix, you are another scientist and wonderful heart that I look up to. Thank you so much for your constant interest in my project, for your amazing scientific clarity, but especially for your kindness, and friendship. To Daniel and Tom Scott, thank you for all the interventions! Thank you for the bonds we so easily created. Eleni, Natalia and Oscar, I am so grateful that I had the chance to share so

many moments with you. Thank you for hosting me, for taking me to ER and staying there for hours, with continuous entertaining, as if we had been friends for a long time. Thank you for tapas Saturday evenings, greek/mexican food and sushi festivals. Thank you for being there.

I must also acknowledge the co-authors that contributed to this work. Govind Chandra, Delia Casas-Pastor and Georg Fritz, my sincere gratitude for the sharing of your knowledge and expertise.

A toda a família BSM, o meu muito obrigada por me acolherem. Tive a sorte de trabalhar num ambiente incrível, que só vocês poderiam ter construído. Em especial, ao Paulo, o meu eterno obrigada por estar sempre lá. Ao Carlos Flores e à Eunice Ferreira, os meus girassóis, obrigada por me ensinarem a ser um "pequeno girassol". Por todos os momentos musicais, e por todos os sorrisos, sempre disponíveis. Obrigada por, juntamente com o Steeve e o Zé Pedro, tornarem o *ghetto* o nosso cantinho. Obrigada a todos por isso. Zé Pedro, a ti uma palavra de apreço especial, um "mega obrigado" por partilhares o teu tempo dedicando-te também a este projeto. Pela honestidade. Por seres o meu esperto em cristalografia, e estares sempre disponível para os meus pedidos (até os de recolha de dados a 7Å). Ao Luís Gales pelo interesse em fazer parte desta jornada. Obrigada ao Frederico Lourenço por toda a ajuda com as produções de proteína! Ainda, dispensar algum tempo para agradecer à Cátia e ao João que foram também muito importantes neste processo.

Não podia deixar de agradecer também a três pessoas que dedicaram muito do seu tempo para me ouvir e tranquilizar. Obrigada Pedro Martins, Bárbara Amorim e Joana Correia! À Joana Gomes por todos os almoços, cafés, etc. Obrigada pelo teu constante apoio. À minha INEB *crew* particular. Cati, Dani, Estrelinha, Luísa, Bia, Rúben e Zé Henrique, o meu obrigado por serem pessoas incríveis e pelo vosso suporte emocional, que tanta diferença fizeram nos meus dias.

Aos Binos da minha vida. Todo este processo foi mais simples e, ao mesmo tempo, mais rico porque, mesmo na distância, vocês estiveram por perto. Joana, Ana, Aida, Rita, Rodrigo, Tiago, Alexandre, João, obrigada pela torcida! Ao Carlos Amorim e à Eunice Sousa, obrigada por manterem o espírito e o corpo ativo, pelas refeições em conjunto ou a caminhada anual, que sempre me ajudaram a fazer este caminho. Ao Luís Melo e à Graça pelo imenso carinho e, também, pelas palavras duras quando foi preciso. Por isso, estou-vos sempre grata. À Martinha e à Soninha, por acreditarem no meu sucesso. Por terem ajudado a construir o meu lado positivo e a minha resiliência. À Si, por estar sempre aqui comigo. Por me teres guiado nos primeiros passos em *Streptomyces*, mas em especial pela duradoura amizade, mil milhões de obrigados! Ao

Luís Flores e à Francisca por assegurar a minha saúde nos períodos mais difíceis. À Guida, ao Gaspar, ao Cenourinha, e ao Queirós por estarem sempre no *backstage*.

Um agradecimento também aos meus sogros e cunhados por equilibrarem a balança, sem que nunca lhes fosse solicitado. Ao Gui, ao Kiko e ao Gonçalo, pela sua fé incondicional nas minhas "coisas científicas (!)". E dentro dos mimos da família, obrigada à minha avó, um exemplo de força, e o meu mais importante agradecimento aos meus pais e ao Bruno, por serem os meus maiores exemplos de trabalho e dedicação. Uma palavra especial aos meus pais, por esperarem sempre o meu melhor. Por acreditarem que era possível. Por dedicarem uma vida de proteção e apoio para me trazer até aqui. Ao Bruno, por seres o meu *role model* na ciência, e por me ensinares, todos os dias, que "no meio de duas pessoas que se encontram, como no meio de alguém que se encontre a si próprio, há sempre um espaço qualquer que, nem por poder ser muito pequeno, deixa por isso de ser muito importante. Tão importante que se não existisse não havia o Mundo".

A todos, muito obrigada!

List of publications

O autor esclarece que na elaboração desta tese foram incluídos dados das publicações abaixo indicadas, e declara ter participado ativamente na conceção e execução das experiências que estiveram na origem da mesma, assim como na sua interpretação, discussão e redação.

The author declares that this thesis includes data from the publications listed below, and that she participated actively in the conception and execution of the experiments that produced such data, as well as in their interpretation, discussion and writing.

Results obtained were included in the publication below.

Oliveira R, Bush MJ, Pires SDS, Chandra G, Casas-Pastor D, Fritz G and Mendes MV (2020) *Streptomyces tsukubaensis* differentiation is aided by the redox responsive ECF56σ/anti-σ pair SigG1/RsfG. *Accepted for publication in Applied and Environmental Microbiology*

Future publications including the results obtained:

Oliveira R, Leite JP, Lourenço F, Gales L, Casas-Pastor D, Fritz G, Buttner MJ, Mendes MV. Crystal structure of the SnoaL_2 ECF56 SigG1 and its cognate anti-sigma RsfG. *Manuscript in preparation*

Pires SDS, Oliveira R, Barreiro C, Moradas-Ferreira P, Mendes MV. OxyR-dependent regulation of the oxidative stress response and iron metabolism in *S. tsukubaensis*. *Manuscript in preparation*

Pires SDS, Oliveira R, Moradas-Ferreira P, Mendes MV. The importance of redox balance in tacrolimus production. *Manuscript in preparation*

Table of contents

Table of abreviations

HRP: horseradish peroxidase
3’ UTR: three prime untranslated region
5’ UTR: three prime untranslated region
A: absorbance
ACP: acyl carrier protein
AT: aminotransferase
ATP: adenosine-5'-triphosphate
ATPase: adenosine triphosphatase
bp: base pairs
BLAST: Basic local alignment search tool
BSA: bovine serum albumin
c-di-GMP: cyclic diguanylate monophosphate
CAS: chrome azurol S
cDNA: complementary DNA
CFU: colony-forming units
DAB: diaminobenzidine
ddH_2O: double deionized water
DHE: dihydroethidium
DHR: dihydrorhodamine 123
DIG: digoxygenin
DNA: Deoxyribonucleic acid
DNase: deoxyribonuclease
DTT: dithiothreitol
EDTA: ethylenediamine tetracetic acid
EMSA: electrophoretic mobility shift assay
h: hour
H_2O_2: hydrogen peroxide
HEPES: N-2-hydroxyethylpiperazine-N′-2-ethanesulfonic acid
IPTG: isopropyl β-D-1-thiogalactopyranoside
LB: Luria-Bertani
LC: liquid chromatography
mL: mililitres
min: minutes
MAD: multi-wavelength anomalous dispersion
MS: mass spectrometry
MTBE: methyl tert-butyl ether
mRNA: messenger RNA
MW: molecular weight
nt: nucleotide

O_2^{-}: superoxide anion

OH^{-}: Hydroxyl radical

OD: optical density

ONPG: ortho-nitrophenyl-β-galactoside

PAGE: polyacrylamide gel electrophoresis

PCR: polymerase chain reaction

PDB: protein data bank

PEG: polyethylene glycol

pI: isoelectric point

Pi: inorganic phosphate

PMFS: phenylmethylsulfonyl fluoride

PNPG: p-nitrophenyl-d-glucuronide

poly(dC): polydeoxycytidylic acid

PPP: pentose phosphate pathway´

RBS: ribosome binding site

RNA: ribonucleic acid

RNAP: RNA polymerase

RNase: ribonuclease

ROS: reactive oxygen species

rpm: rotations per minute

rRNA: ribosomal RNA

RT-PCR: reverse transcriptase-polymerase chain reaction

RT-qPCR: reverse transcriptase-quantitative polymerase chain reaction

RTT: relative retention time

SAD: single-wavelength anomalous diffraction

SDS: sodium dodecylsulphate

SeMet: labelled with selenomethionines

SOD: superoxide Dismutase

TCA: tricarboxylic acid

TFA: trifluoroacetic acid

Tris: tris(hydroxymethyl)aminomethane

UV: ultraviolet

v/v: volume per volume

vol: volumes

w/v: weight per volume

X-Gluc: 5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid

Abstract

Actinobacteria are ubiquously distributed throughout diverse ecosystems and are exposed to highly distinct environmental conditions. *Streptomyces* are soil-dwelling obligate aerobes of the Actinobacteria phylum that harbour an arsenal of regulators to circumvent stressful signals, and thrives in highly oxidizing environments. To do so, they re-arrange their molecular machinery that controls transcription, mainly through the dynamic trading of sigma subunits of the RNA polymerase, the sigma factors. Among the plethora of sigma factors, the extracytoplasmic sigma factors (ECF) are the key responders to inputs from the environment. *Streptomyces* genomes harbour dozens of ECFs, which is significantly above the average of 10 ECF/genome encountered in other well-studied bacteria, like *Escherichia coli* or *Bacillus subtilis*, and confers a high degree of plasticity to transcriptional regulation in these cells. Recently, several ECFs with domain topologies that go beyond the σ2-σ4 core have been discovered. Despite the knowledge gathered in 25 years of ECF research, the biology of these distinct ECFs is poorly understood.

In this work, we characterised the biological role of a novel ECF that, in addition to the minimal σ2-σ4 core, exhibits a long extension at the C-terminus that codes for a SnoaL_2 domain, SigG1. Although ECF with CT extensions have been proposed to be isolated regulators, and that CT extensions play an inhibitory role through intramolecular regulation, in this work, we report the existence of a SigG1 antagonist, RsfG, which binds to SigG1 to prevent its runaway activity during the differentiation process.

Using strains with a disrupted σ/anti-σ pair we integrated next generation sequencing (NGS) RNA-seq and ChIP-seq experiments and identified the role of SigG1 in the morphological differentiation programme. The RsfG-SigG1 system is active upon H_2O_2-induced stress generated by the fading vegetative mycelia, to minimize the damage caused by oxidative stress and help treading through the various differentiation stages that will build its complex multicellular structure. SigG1 initiates the expression of determinants of morphogenesis and the control of metal imbalance. Hence, unlike classical ECFs that are regulators of extracytoplasmic cues, SigG1 uses intracellular ROS as signalling molecules to trigger the activation of the energetic support that fuels the metabolic shift from vegetative to aerial differentiation.

To the best of our knowledge, we propose for the first time that an ECF with a linked SnoaL_2 extension might be modulated by two layers of posttranscriptional regulation, that include interdomain dependent regulation mediated by the SnoaL_2 domain, and the coordination with a cognate anti-σ factor, RsfG. We propose that the conformation adopted by members of the ECF56 family distinguishes these proteins from their phylogenetic neighbours in ECF41 family, which introduces a new mode to control

gene expression from SnoaL_2-containing ECFs. To prove this hypothesis, we are currently addressing the three-dimensional structure of SigG1 and RsfG. These crystalographic studies will also reveal how RsfG binds to SigG1, and what dissociates the sigma factor from the cognate anti-sigma factor. Moreover, we are studying the SigG1 orthologue in a *Streptomyces venezuelae* model, SigG2. Although SigG2 might not follow SigG1 transcriptional landscape, our studies indicate numerous layers of ECF regulation, which generate a stringent ECF56-dependent transcription inhibition. Overall, our results show that the timely activation of a SnoaL_2 ECF might require four responsive switches: 1) a redox sensing mechanism, 2) a small molecule binding, 3) intramolecular interactions and 4) the sequestration by an anti-sigma factor. Therefore, we propose the use of SigG1 as a synthetic biology tool, suitable for engineering pathways of interest in bacteria that do not express ECF56 sigma factors.

In sum, this work advances the studies on the progression of *Streptomyces* differentiation modulated by oxidative stress signals. It demonstrates how the expression of a sigma factor with a unique architecture is essential to ameliorate stress and maintain the reductive-oxidative balance, during the differentiation of the multicellular mycelia. Furthermore, with this work we were able to assign a cognate anti-sigma factor to an ECF56, hence opening new avenues towards unveiling the function of novel ECFs that evolved with complex domain organisations that go beyond the traditional ECF.

Resumo

As espécies de actinobactérias seguem uma distribuição ubíqua por diversos ecossistemas estando, por isso, expostas a condições ambientais muito distintas. *Streptomyces, é* um género do filo Actinobacteria constituído por bactérias aeróbias obrigatórias que habitam no solo. São, por isso, bactérias que reúnem um arsenal de proteínas reguladoras, para superar os sinais de stress a que estão expostas, e assim prosperar em ambientes de elevado potencial oxidante. Para este propósito, proporcionam um rearranjo da maquinaria molecular que controla a transcrição, sobretudo através da permuta dinâmica de subunidades sigma da ARN polimerase, os fatores sigma. De entre a diversidade de fatores sigma que existem, os fatores sigma extra-citoplasmáticos (ECF) constituem as principais moléculas de resposta aos estímulos provenientes do ambiente. Os genomas de *Streptomyces* incluem um número de ECFs significativamente superior à média de ECFs em genomas de outras bactérias bastante estudadas, como *Escherichia coli* ou *Bacillus subtilis*, o que confere um elevado grau de plasticidade à regulação da transcrição nestas bactérias. Recentemente, foram descobertos vários ECFs com uma topologia da proteína que vai para além dos domínios conservados σ2-σ4. Apesar de todo o conhecimento reunido em 25 anos de investigação nestas proteínas, ainda pouco se sabe sobre estes novos ECFs.

Neste trabalho, caracterizámos o papel biológico de um novo ECF que para além dos domínios σ2-σ4, contém uma longa extensão C- terminal, que codifica para um domínio SnoaL_2. Os ECFs com extensões C-terminais são por vezes indicados como sendo reguladores que actuam isolados, e a sua extensão C-terminal poderá favorecer interações intramoleculares com um papel inibitório para o ECF. Contudo, neste trabalho verificámos a existência de uma proteína antagonista que se liga ao SigG1, para prevenir uma actividade descontrolada deste ECF durante o processo de diferenciação, a RsfG.

Com recurso à integração de experiências de sequênciação de nova-geração (NGS) como RNA-seq e ChIP-seq, e utilizando estirpes com o sistema sigma/anti-siga corrompido, fomos capazes de indentificar o papel do SigG1 no programa de diferenciação morfológica. Durante o desenvolvimento, um micélio vegetativo enfraquecido produz H_2O_2, que vai constituir o estímulo de stress que activa o sistema RsfG-SigG1 e, deste modo, atenuar o dano oxidativo gerado e dar suporte ao avanço pelo diferentes estadios de desevolvimento que darão origem à estrutura pluricelular. A proteína SigG1 inicia a expressão de genes determinantes da morfogénese e de controlo da homeostasia de metais. Assim, e ao contrário dos tradicionais ECFs que respondem a estímulos extracitoplasmáticos, é capaz de usar espécies reactivas de

oxigénio intracelulares como moléculas de sinalização, que vão activar as reservas energéticas da célula. Isto permitirá o suporte para a transição metabólica que se dá durante a diferenciação do micelio vegetativo em micélio aéreo.

Pela primeira vez propõe-se que um ECF com uma extensão SnoaL_2 poderá estar sujeito a duas camadas de regulação pós-tradução, que incluem a regulação dependente de ligações interdomínios mediada pelo SnoaL_2, e a coordenação com um fator anti-sigma, a proteína RsfG. Sugerimos que são as alterações conformacionais adoptadas por membros da família ECF56 que os distingue dos membros da família filogenética mais próxima, a família ECF41, e que isto introduz um novo mecanismo de regulação da expressão genética dependente de ECFs com domínios SnoaL_2. Para demonstrar esta hipótese, estamos actualmente a estudar a estrutura tridimensional do SigG1 e da RsfG. Estes estudos cristalográficos têm como objetivo compreender melhor a interação entre a RsfG e o SigG1, bem com o mecanismo de dissociação entre as duas proteínas e libertação do SigG1. Adicionalmente, estamos a estudar o papel biológico de um ortólogo do SigG1 num modelo de *Streptomyces venezuelae,* o SigG2. Apesar da regulação dependente do SigG2 parecer não resultar no mesmo quadro transcricional que a regulação por SigG1, os nossos estudos reúnem evidências da existênica de várias camadas de regulação, que providenciam uma restringente inibição da transcrição dependente de proteínas ECF56. No seu conjunto, os nossos resultados mostram que a activação atempada de um ECF com uma extensão SnoaL_2 poderá depender de 4 mecanismos de resposta: 1) um mecanismo sensor do estado redox da célula, 2) a ligação de um ligando de baixo peso molecular, 3) interacções intramoleculares e 4) o sequestro do fator sigma pelo seu anti-sigma correspondente. Por isto mesmo, propomos o uso do fator sigma SigG1 como uma ferramenta de biología sintética, capaz de modelar vias metabólicas ou de regulação de interesse, em bactérias que não expressam fatores sigma do tipo ECF56 naturalmente.

Em suma, os resultados descritos neste trabaho representam um significativo avanço científico no estudo da progressão da diferenciação em *Streptomyces* que é modulada pelo stress oxidativo. Este trabalho demonstra como a expressão de um fator sigma com uma arquitectura única, pode ser essencial para atenuar o stress e manter o equilibro oxidante-redutor dentro da célula, durante a diferenciação do micélio pluricelular. Adicionalmente, com este trabalho fomos capazes de atribuir um fator anti-sigma a um ECF do tipo ECF56, abrindo assim novos caminhos que poderão contribuir para descobrir a função de novos ECFs que evoluiram com uma organizaão em termos de domínios proteicos mais complexas do que o tradicional ECF.

Chapter 1

Introduction

1.1. *Streptomyces*

1.1.1. *Streptomyces*, an untapped source of bioactive compounds

'It turns out it's not just gratitude that makes rain smell so appealing after a long period of dry weather' (Halton, 2018). Animals, and specially humans, are highly sensitive to the earthy fragrance of the soil after a rainy period. The odour comes from a particular volatile organic compound called geosmin produced by *Streptomyces* (Gerber and Lechevalier, 1965), one of the numerous bacterial genera that colonize the soil.

Streptomyces are Gram-positive bacteria of the Actinobacteria phylum that have evolved ubiquitously in soil, freshwater and marine ecosystems. It comprises bacteria with high GC content genomes, ranging from 2.5 to 9.7Mb in size (Miao and Davies, 2010). Among other well-studied genera, Actinobacteria includes opportunistic human pathogens like the members of *Mycobacterium*, *Gordonia*, *Corynebacterium* or *Nocardia* genera, which cause infectious disease of the lung, skin, bone and soft tissues (Barksdale, 1970; Hemmersbach-Miller et al., 2018; Miao and Davies, 2010; Sowani et al., 2017). A significant number of actinobacterial species have been highly valuable in medicine and life sciences research as they offer an infinite number of specialized compounds produce by the secondary metabolism. Pharmacologically active compounds produced by Actinobacteria have been widely used as antibacterial, antifungal, immunosuppressant or anti-tumour agents (Table 1). *Streptomyces,* stands as the most drug-prolific genus.

1.1.2. An intricate filamentous life cycle

Alongside with its unique biosynthetic capacities, *Streptomyces* bacteria exhibit a multicellular-based particular shape and are perpetuated in the environment through the dispersion of spores. Differentiation of streptomycetes may occur in solid or liquid cultures, although most *Streptomyces* strains fail to sporulate in liquid culture and full differentiation in liquid fermentations is rarely observed. Nevertheless, a significant part of the differentiation process is comparable to that observed in solid cultures, in particular at the transcriptional level (Manteca et al., 2010). They undergo a complex filamentous life cycle, which initiates with the germination of a spore when the environment conditions are favourable (Figure 1). Spore germination originates branched hyphae that grow by tip extension into the substrate to form dense filamentous vegetative mycelia (MI). MI is characterised by highly septate and cross-wall compartmentalized hyphae and an active primary metabolism (Yagüe et al., 2016) that facilitates nutrient scavenging. Upon nutrient starvation, part of the MI mycelium undergoes cell death and viable MI compartments give rise to a multinucleated secondary vegetative mycelium (MII).

Table 1. Pharmacologically active metabolites produced by *Streptomyces*

Compound	Species	Biological activity	Ref. (DOI/PubMedID)
Abyssomicins	*Streptomyces sp.*	Antibacterial	DOI: 10.1021/OL0705999
Actinomycin D	*S. parvullus*	Antitumoral, antibacterial	DOI: 0021-9193/81/110670-08$02.00/0
Amphotericin B	*S. nodosus*	Antifungal	DOI: 10.1016/S1074-5521(01)00046-1
Apramycin	*S. tenebrarius*	Antibacterial	DOI: 10.1021/jo00874a003
Acivicin	*S. sviceus*	Antitumoral	PMID: 7026076
Avermectin	*S. avermitilis*	Antiparasitic	DOI: 10.1128/aac.15.3.361
Benzastatin C	*S. nitrosporeus*	Antiviral	DOI: 10.1248/bpb.30.795
Blasticidin	*S. griseochromogenes*	Antibacterial	PMID: 4953866
Bleomycin	*S. verticillus*	Antibacterial	DOI: 10.7164/antibiotics.19.200
Bottromycin	*Streptomyces sp.*	Antitumoural	DOI:10.1016/j.chembiol.2012.08.013
Candicidin	*S. griseus*	Antifungal	DOI: 10.1099/00221287-148-1-51
Chalcomycin	*Streptomyces B7064*	Antitumoral	DOI: 10.7164/antibiotics.55.893
Chlorohydroquinones	*Streptomyces sp.*	Anticancer	DOI: 10.1021/np058011z
Cicloheximide	*S. griseus*	Antifungal	PMID: 808159
Clavulanic acid	*S. clavuligerus*	Antibacterial	DOI: 10.1007/BF00270593
Concanamycin	*S. neyagawaensis*	Antifungal	DOI: 10.1099/mic.0.28194-0
Daryamides	*Streptomyces CNQ085*	Antitumoral	DOI: 10.1021/np0603828
Doxorubicin	*S. peucetius*	Antitumoral	DOI: 10.1111/bph.12202
Hyaluromycin	*S. hyaluromycini*	Antitumoral	DOI: 10.3390/md12010491
Hygromycin B	*S. hygroscopicus*	Antibacterial, antifungal	DOI: 10.7164/antibiotics.29.685
Kanamycin	*S. kanamyceticus*	Antibacterial	DOI: 0021-9193/84/010079-05$02.00/0
Mitomycin C	*S. lavendulae*	Antitumoral	DOI: 10.1016/S1074-5521(99)80040-4
Neomycin	*S. fradiae*	Antibacterial	DOI: 10.1021/ja01175a516
Nogalamycin	*S. nogalater*	Antitumoral	DOI: 10.1007/s004380100554
Nystatin	*S. noursei*	Antifungal	DOI: 10.1007/s00253-004-1802-4
Pimaricin	*S. natalensis*	Antifungal	DOI: 10.1136/thx.23.5.519
Puromycin	*S. alboniger*	Antibacterial	DOI: 10.1016/S0044-4057(73)80089-9
Rapamycin	*S. rapamycinicus*	Immunosupressive	DOI: 10.7164/antibiotics.37.1231
Simocyclinone D8	*S. antibioticus*	Antibacterial	DOI:10.1128/AAC.49.3.1093-1100.2005
Spectinomycin	*S. spectabilis*	Antibacterial	DOI: 10.1126/science.149.3688.1096
Staurosporinone	*Streptomyces sp.*	Antitumoral	DOI: 10.1038/ja.2006.46
Streptomycin	*S. flavopersicus*	Antibacterial	DOI: 10.1073/pnas.51.5.883
Tetracyclines	*Streptomyces spp.*	Antibacterial	DOI: 10.1002/jps.2600600413
Thiostreptone	*S. azureus*	Antibacterial	DOI: 10.1016/0006-291X(70)90956-3
Vancomycin	*S. orientalis*	Antibacterial	DOI: 10.1038/1841894b0
Viomycin	*S. floridae*	Antibacterial	PMID: 14799786

As the colony matures, it activates transcription of proteins required for the formation of aerial mycelium. The transition from vegetative to aerial mycelia has had more attention in the recent years. In previous work, we have identified one of the triggers of *Streptomyces* differentiation into aerial hyphae.

We proposed that the highly active primary metabolism during aerobic growth generates high amounts of reactive oxygen species (ROS), and that the oxidative environment leads to an initial phase of controlled cell death (Figure 1, (Beites et al., 2015)). An intracellular redox imbalance requires significant rearrangement of the cell metabolism that often results in growth inhibition. To counteract this stress, cells adapt by upregulating iron storage sinks, to control iron availability and prevent the occurrence of Fenton chemistry (Figure 2, (Imlay et al., 1988)), where H_2O_2 reacts with labile iron, damaging molecules such as DNA or protein iron-sulphur clusters among others. Due to this connection, iron and other metals are integrated in the morphological differentiation

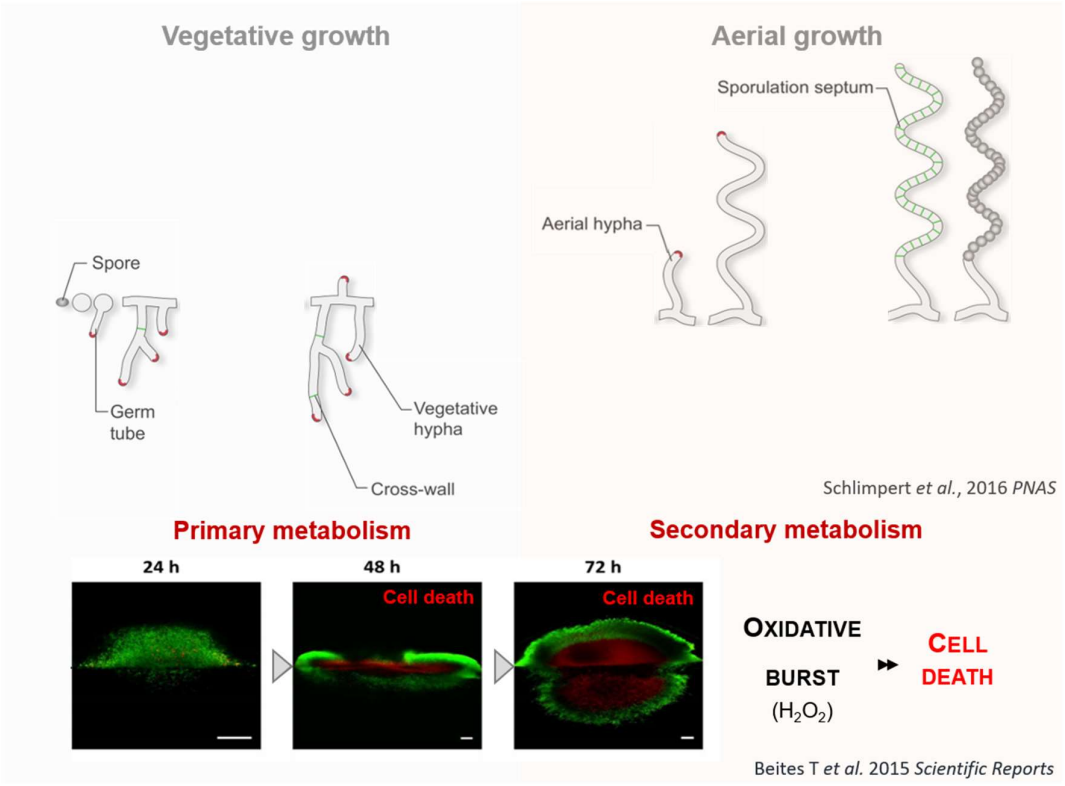


Figure 1. Streptomycetes life cycle. Spore germination triggers well-defined steps of hyphae differentiation that culminate in the septation and cell wall disruption of aerial hyphae, to generate new reproductive structures. The vegetative and aerial stages of growth are characterised by active processes of the primary and secondary metabolisms, respectively. The transition between these two stages is accompanied by cell death phenomena that re-routes nutrients for the progression of differentiation. Adapted from (Beites et al., 2015; Schlimpert et al., 2017).

network. In fact, iron has been reported to be directly involved in development at different levels of these cascades, for instance, with the presence of exposed [4Fe–4S] clusters in WhiB-like regulators (Bush, 2018) or the downregulation of *bldN* transcription in conditions of poor availability of iron (Traxler et al., 2012). We suggested that the efforts made in re-establishing the redox balanced environment, include the regulation of iron homeostasis, so that the correct levels of iron ensure the expression of the *bldN* developmental determinant (Beites et al., 2015). This brought additional evidence to the paradigm shift that ROS can act as signalling molecules and not only as a killing mechanism.

It is during the differentiation from vegetative to aerial mycelium that the synthesis of secondary metabolites is triggered. Secondary metabolites are small molecules that are not essential for basic cell growth and primary metabolism, but that the bacteria produce, presumably, to control the signalling and communication for the protection of resources in a harsh natural habitat. In *S. coelicolor,* the suicidal behaviour in the transition to a more complex morphology is associated to the activity of genes that

$$Fe^{2+} + H_2O_2 \rightarrow [FeO^{2+}] + H^+ + OH^- \rightarrow Fe^{3+} + OH^- + HO^{\bullet}$$

Figure 2. Fenton reaction. H_2O_2 oxidizes soluble intracellular iron, which results in the production of hydroxyl radicals that react with biomolecules causing damages in DNA, proteins and lipids (Imlay, 2015).

determine the onset of secondary metabolism (Tenconi et al., 2018). The cycle culminates in the timely differentiation of the aerial hyphae into chains of new resistant spores – each containing one copy of the genome – that will ensure the reproductive legacy of the bacteria. Mature spores produce pigmented compounds that grant *Streptomyces* colonies their characteristic colour.

1.1.2.1. Developmental regulatory cascades

To cope with such a complex life cycle *Streptomyces* depend on the contribution of large set of landmark genes that tightly control the different stages of the polar growth (Figure 3, reviewed by (Bush et al., 2015)). During the vegetative stage, differentiation is strictly dependent on cascades of proteins produced upon the activation of the '*bald'* (*bld*) genes. The Bld regulators orchestrate the transition to the aerial lifestyle. Disruption of *bld* genes causes a dramatic effect on the colony, as it blocks aerial mycelium formation and sporulation. The colony fails to develop aerial hyphae and exhibits a bald and shining phenotype. Most of the differentiation determinants are under the control of a DNA-binding repressor regulator BldD, identified in the genome of at least 93 *Streptomyces* strains (Al-Bassam et al., 2018). The activity of BldD is modulated by the

specific signal of the cyclic dinucleotide c-di-GMP that promotes protein dimerization (Tschowri et al., 2014). As the intracellular levels of c-di-GMP increase, these small molecules mediate the formation of a BldD dimeric complex that will bind DNA and repress the transcription of downstream genes. Binding to target promoter regions occurs hierarchically and is characteristic to each developmental stage.

Hyphae efficient growth carries on through a polarized distribution polarisome-like complex led by DivIVA that drives apical growth and hyphal branching by polarisome-splitting (Flardh et al., 2012). Later on the development process, a family of proteins encoded by the '*white'* (*whi*) genes regulates aerial mycelia differentiation and its conversion into spores. Absence of the *whi* genes results in a white and fuzzy appearance of the colony that arrests its development at an early aerial non-septate hyphae stage (Bush et al., 2015; Flärdh and Buttner, 2009; McCormick and Flärdh, 2012). In fact, its activity seems to be more prevalent at the level of transcription, with the important participation of WhiB-like (*wbl*) genes. The *wbl* genes are frequently associated to gene expression for virulence, antibiotic resistance and morphogenesis (reviewed in (Bush, 2018)). In particular, Wbl regulators are essential to initiate septation during Stre*ptomyces* sporulation. Some of these proteins are responsive to stress conditions like nutrient availability or oxidative stress (Geiman et al., 2006; Lee et al., 2013).

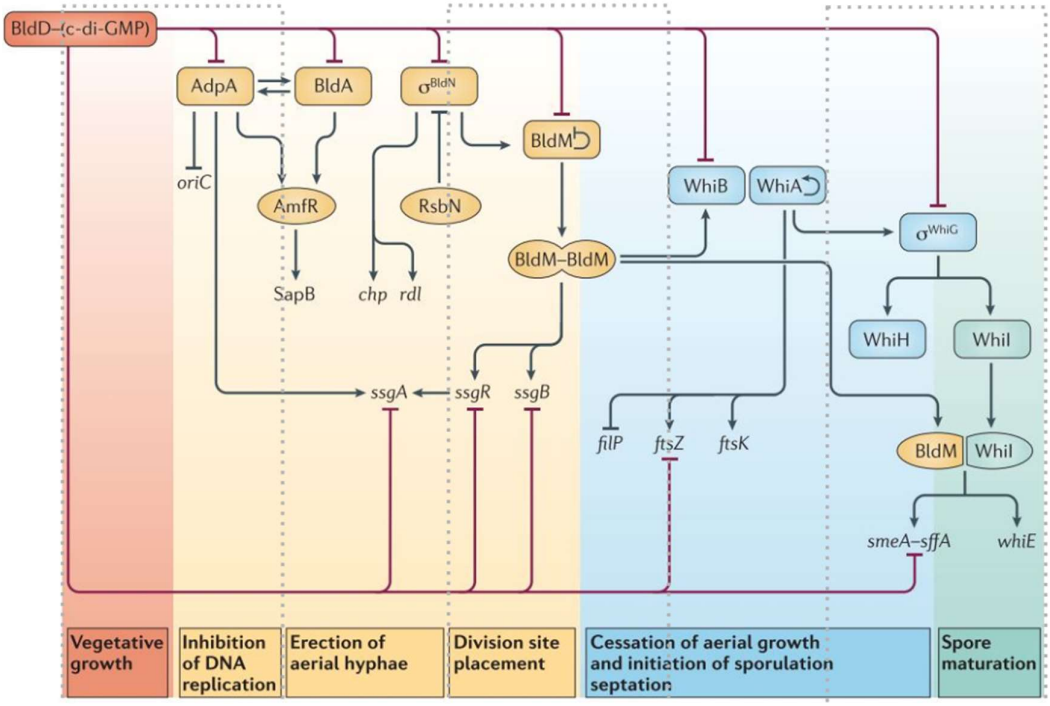


Figure 3. Gene regulatory cascades that determine *Streptomyces* differentiation. The growth stages that give rise to new reproductive units are illustrated. Adapted from (Bush et al., 2015)

During the transition from aerial mycelia formation to sporulation, expression of cell division and chromosome segregation determinants conveys the conditions for synchronous sporulation. *Streptomyces* differentiated hyphae are coated with a hydrophobic layer of proteins that protects the bacteria from desiccation. Two families of proteins are in charge of building this protective layer, the rodlins and the chaplins (Claessen et al., 2004; Elliot et al., 2003). BldD controls the expression of these genes through the regulation of the expression of a specific sigma factor, BldN and its cognate anti-sigma factor, RsbN (Bibb et al., 2012) that act upstream of the rodlins and chaplins. Fully mature spores acquire a strain-specific colour derived from a polyketide spore pigment produced during the later stages of development.

1.1.2.2. Stress-induced lifestyles

Recently, a breakthrough described that some *Streptomyces* species deviate from this typical cycle under nutrient starvation, and engage into an exploratory growth. This is characterised by the rapid growth of non-branching vegetative hyphae (reviewed in (Jones and Elliot, 2017)), that never reach aerial mycelium or spores formation, and by a significant alkalinisation of the substrate where bacteria are growing. The so-called explorer cells are linked to conditions of low glucose, high amino acid concentrations and low iron availability (Jones et al., 2019), and are formed independently of the presence of the typical developmental regulators, suggesting a growth programme that diverges from the canonical life cycle. Other types of differentiated cells were observed for *Streptomyces*. Paradoxically, despite living shielded by a complex cell wall that confers a good level of protection, *Streptomyces* cell may respond to environmental cues by shedding and rebuilding the cell wall to intertwine between vegetative hyphal cells and wall-deficient cells, the S-forms (Stress-induced cells). Strikingly, some cells adapt a permanent cell wall-free form of living, the L-forms, that persist without a cell envelope, for example, in response to cell wall targeting antibiotics (Claessen and Errington, 2019).

1.1.3. *Streptomyces venezuelae* as a model organism to study differentiation

The most amazing traits and molecular processes of adaptation to extreme environs have been discovered in exquisite bacterial strains, however, working with less studied strains often carry major challenges, in particular in what concerns to the genetic manipulation for molecular studies. To overcome this, the research community chose to work with *Streptomyces coelicolor* A3(2) as a model organism to study morphological development of Actinobacteria, through the use of its *bld* and *whi* mutants (Hoskisson and van Wezel, 2019). The main advantage of these bacteria is that they produce pigmented antibiotics that are used a markers for genetics experiments, the red-

pigmented prodiginine antibiotic, undecylprodigiosin, and the blue-pigmented polyketide actinorhodin.

Whilst *S. coelicolor* is the classical model for development and antibiotic production, other *Streptomyces* strains have emerged as good genetic tools. *S. venezuelae* has been routinely used as a model to study the differentiation process primarily due to its fast growth rate comparing to other streptomycetes. Not very far from what is observed in *S. coelicolor* it has a doubling time of approximately 40 min and other accessible genetic tractability due to its high transformation efficiencies. Furthermore, it exhibits sensitivity to antibiotics routinely used in molecular cloning. Consequently, researchers have been able to develop a good number of genetic manipulation tools, which facilitates routine molecular work. More importantly, the most notorious trait of *S. venezuelae* cultures is that they lack mycelial clumping, and complete the differentiation process – producing viable spores – in liquid media. Altogether, these features makes it an attractive strain to use for physiological studies.

1.1.4. *Streptomyces tsukubaensis*

In this work, we use an industrial sttreptomyces strain called *Streptomyces tsukubaensis* NRRL 18488. It is a soil-dwelling strain with a 7.62Mbp linear chromosome strain with a 71.52% G+C content, that contains two circular plasmids, pSTS1 and pSTS2, of 24.7 and 31.1kbp, respectively (Barreiro et al., 2012). In *S. tsukubaensis*, spores exhibit a brownish-grey pigmentation. Besides producing other bioactive compounds, it is the industrial producer of tacrolimus, an immunosuppressant extensively used in the clinical practice to prevent organ rejection in transplantation patients (Krentz et al., 1994). Tacrolimus, also known as FK506, is a macrolide polyketide first isolated in 1987 from a culture of *S. tsukubaensis* 9993 (Kino et al., 1987). It was approved by the FDA in 1994, and since then it has gradually replaced cyclosporine in immunosuppressive therapies due to its improved therapeutic benefits. In addition, it has been effective in nerve regeneration, osteoblast differentiation, treatment of eczema, bladder disorders and cancer (Chang et al., 2019; Davis et al., 2019; Jiang et al., 1993; Nagrani and Zito, 2019; Tang et al., 2002). FK506 is naturally produced in 18 *Streptomyces* strains. Although the regulatory mechanisms for this production are less conserved, highly similar FK506 biosynthetic clusters seem to be acquired by these strains, possibly due to horizontal gene transfer (reviewed in Barreiro and Martinez-Castro (2014)).

Recently, we have identified key molecular mechanisms that can modulate the production of FK506 in *Streptomyces tsukubaensis* NRRL 18488. We have shown that FK506 biosynthesis is highly dependent on an optimal redox state, mainly achieved by

the activation of the primary antioxidant responders, like peroxidases and catalases. By detoxifying the excess of H_2O_2, cells reach a low oxidizing environment that is ideal for the biosynthesis of the immunosuppressant (Pires et al., unpublished results).

1.2. Environmental sensing and bacterial intelligence

Living in the soil represents a challenge in coping with constant fluctuations in nutritional supplies and ensure survival. By developing intricate and sophisticated mechanisms of sensing and responding to environmental stress, bacterial cells have a so-called 'intelligence' to respond to intra- and extracellular stimuli. Resembling transversal traits to other forms of life, bacteria have numerous ways of processing information and strive for cellular robustness. In response to information received from different stressful conditions, bacteria will fine-tune molecular adaptations at different levels. At the DNA level, for instance, it modulates gene expression by altering nucleoid and chromatin state with environmentally mediated epigenetic regulation (Casadesús and Low, 2013). At the RNA level, cells harbour diverse species of RNA that, together with signal-dependent levels of its synthesis and folding, act as effective activating/silencing devices (Güell et al., 2011). While protein production and post-translational regulation are obvious checkpoints of molecular stress responses, bacterial cells react to environment and adapt through the direct activity of a large extent of transcriptional regulators (Huang et al., 2015). In bacterial transcription, three fundamental pillars orchestrate external cues signal sensing and amplification: one-component systems (OCS), two-component systems (TCS) and extracytoplasmic function (ECF) sigma factors. One-component systems are the simplest response regulators, which act as auto regulatory DNA-binding transcription factors (Ulrich et al., 2005). In this type of system, the perceived input signal is usually responsible for the alteration of protein conformation that results in an output response. TCS and ECFs convey another level of refinement when compared to OCS, as they dependent on a partner that senses the stress injury and transmits the signal to the effective regulator. TCS rely on phosphorelay phenomena, where histidine kinases perceive the stimuli and transfer the signal to DNA-binding response regulators, which will programme gene transcription and trigger the molecular cascades of response (Francis and Porter, 2019). In ECF systems, gene expression is directly controlled through the discrete recruitment of the RNA polymerase (RNAP) to promoters by its sigma subunit, the sigma (σ) factor (Helmann, 2019).

1.2.1 Sigma factor, the versatile RNAP subunit

To carry out transcription the RNAP, assembles five highly conserved core subunits that constitute the heteromultimeric catalytic active core enzyme (α2ββ'ω). To initiate transcription, the core RNAP must associate a dissociable σ subunit that confers promoter specificity, directing the RNAP holoenzyme to promoter sequences to initiate the necessary transcription. This association is reversible to promote the dynamic interchange between different sigma factors according to the target promoter that will be activated (Österberg et al., 2011).

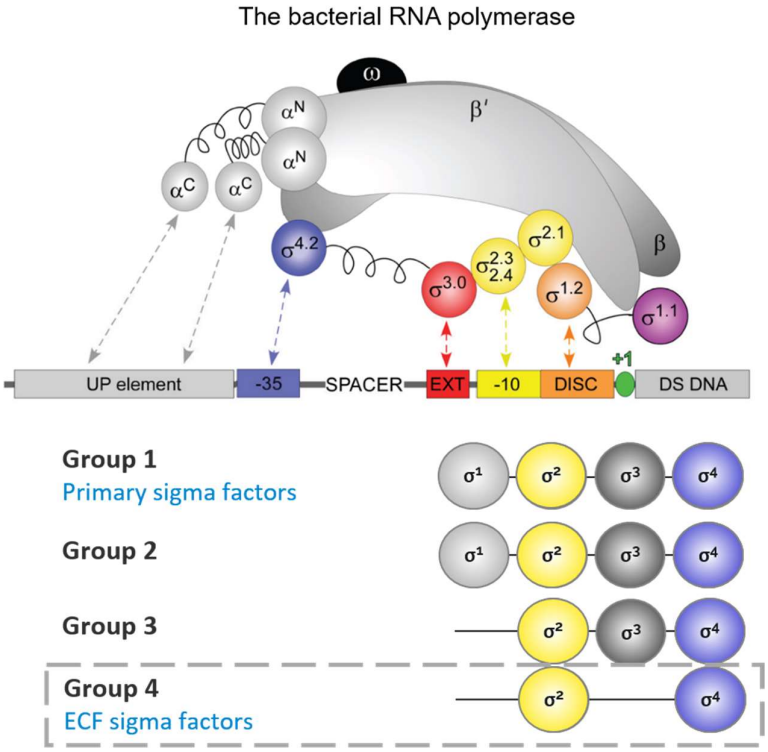


Figure 4. The bacterial RNA polymerase. Schematic representation of the RNAP subunits and the promoter DNA. The core subunits are in grey. The sigma factor σ^{70} regions are highlighted in colour: σ^{70}_r1 (σ_1): orange; σ^{70}_r2 (σ_2): yellow; σ^{70}_r3 (σ_3): red; and σ^{70}_r4 (σ_4): blue. Linker regions in α and σ subunits are shown as springs. Adapted from (Ruff et al., 2015).

1.2.1.1 Primary and alternative bacterial σ factors

Bacteria harbour a variable number of ECF-sigmas (ECFs) to compete and direct the transcription of specific classes of stress-responsive genes. The majority of bacterial sigma factors belong to the σ^{70} protein family, and are classified in four separate phylogenetic groups according to sequence similarity, protein domain architecture and function (Helmann, 2002; Lonetto et al., 1992).

Members of Group I are more abundant in the cell and comprise proteins responsible for the initiation of the expression of the majority of the housekeeping genes. They are designated as primary sigma factors and have a complex protein architecture with four highly conserved regions: σ^{70}_r1, σ^{70}_r2, σ^{70}_r3 and σ^{70}_r4 (Figure 4). Within group II we find proteins that share domain organisation with the primary sigma factors, but that are not considered essential for growth like, for example, *Streptmoyces coelicolor* σ^{S} (Buttner and Lewis, 1992) or σ^{D} (Jishage and Ishihama, 1999). Members of Group III are significantly smaller alternative sigma factors, as they are sequentially distant from groups I and II. They usually activate regulons of developmental checkpoint or stress responses. Well-known sigmas from this group include flagellar, heat shock, general stress response and sporulation regulators (Boylan et al., 1993; Eichelberg and Galán, 2000; Gallagher et al., 2019; Grossman et al., 1984).

Group IV is the largest group of alternative sigma factors, also known as ECF sigma factors. Members of group IV harbour a core region composed of the σ^{70}_r2 (σ2) and σ^{70}_r4 (σ4) regions and are responsible for transcribing a wide on-demand array of genes in response to specific stress signals, providing the cells with a precise way of sensing environmental fluctuations. *S. coelicolor* σ^{R}, which is involve in the oxidative stress response (Paget et al., 1998), or *S. coelicolor* σ^{E} involved in cell envelope stress are largely studied examples of these proteins (Paget et al., 1999).

1.2.1.2 The Extra Cytoplasmic Function (ECF) σ factors

ECFs were first discovered by Michael Lonetto and Mark Buttner in 1994, after analysing a particularly short sigma factor from *Streptomyces*, sigma E (Lonetto et al., 1994). By the time, the impact of their breakthrough was not so evident, but with their discoveries, they opened to the community a plethora of sigma factors that later became the largest and most diverse family of sigma factors.

ECFs are critical coordinators of the stress response that are recruited for specific pathways with higher stringency than the observed for the σ^{70} housekeeping proteins. As specific transcriptional regulators, they have a few hallmarks traditionally recognized in these sigma factors. Proteins domains σ2 and σ4 bind to the -10 and -35 regions, respectively and are separated by optimal spacer lengths of 16-17bp (Helmann, 2002). In addition, the linker region between σ2 and σ4 protein domains has been described to have an active regulatory role in both the recognition of the bipartite sequence motif and the RNAP binding mechanism (Lane and Darst, 2006; Lin et al., 2019). Classical ECF target promoters share a conserved AAC motif in the -35 box, and are activated upon perception of a signal at the cell surface by a cognate sensor protein, the anti-sigma factor. The anti-sigma factor is often co-transcribed by an *ecf* neighbouring gene.

The dynamics of ECF promoter recognition has been extensively characterised in recent works. Transcription may be initiated from more than one start site in the promoter in order to adapt the mRNA synthesis stability or efficiency, or to create alternative sites for translation initiation (Wade and Grainger, 2014). In accordance, in many cases, ECFs recognize variable promoter elements in a mix-and-match fashion that enables transcription to initiate from different elements depending on environmental conditions (Guzina and Djordjevic, 2017). While the -10 sequence is mostly invariant, the -35 box often accommodates some variability in nucleotide bases (Bibb et al., 2012; Feeney et al., 2017). Therefore, different ECFs may overlap and promote initiation from multiple holoenzymes at the same nucleotide (Typas et al., 2007).

Protein structural data also indicates high conservation among ECF features that assemble transcription initiation complex. In brief, protein structure determination works show that the attachment of σ_2 to the RNAP-β' subunit is similar to other σ_2 domains. In addition, the ECF σ_2-σ_4 linker behaves like a σ_3 domain in sigma factors from group III and inserts into the active centre cleft of the RNAP. However, although the residues within this linker are important for maintaining the bond between σ_2 and σ_4 and stabilize the binding of these regions to the RNAP, they do not seem to play a significant role directly in RNAP binding (Fang et al., 2019; Gaballa et al., 2018; Lin et al., 2019). Altogether, this knowledge suggests that the ECF-RNAP interfaces are conserved and that there is room for overlapping of different ECFs to the same promoters. Therefore, the maintenance of specificity of gene expression is dictated by the simultaneous conservation of the sequence within the -35 and -10 boxes of the ECF.

1.2.1.2.1 ECF σ Factor Group phylogenetic classification

Since 1994, significant knowledge has been gathered on ECF sigma factors, but detailed information on how these proteins evolved is still sparse. Actinobacteria ECFs were classified in 157 well-defined phylogenetic groups, according to their conserved domain organisation, genomic neighbourhood context and predicted putative regulatory partners (Casas-Pastor et al., 2019; Huang et al., 2015; Jogler et al., 2012; Staron et al., 2009). They are ubiquitously distributed and are more abundant in *Streptomyces* genomes than in other bacteria. In well-studied model bacteria, a confined number of ECFs assist stress-responsive transcription. For instance, we find 7 ECFs *E. coli,* 13 ECFs in *Mycobacterium tuberculosis* and 18 ECFs in *Bacillus subtilis*. In other bacteria like, *Streptomyces*, we find more than 30 ECF sigma factors, which is very distinctive from the average 10 ECF/genome described (Casas-Pastor et al., 2019) and reflects a higher level of investment in transcription. In *S. tsukubaensis,* 42 ECFs coordinate stress-responsive transcription, in harmony to what is observed for the model organisms,

S. venezuelae (34 ECF) or *S. coelicolor* (50 ECF). This level of investment in signalling complexity, suggest novel mechanisms of ECF-dependent transduction and has resulted in the large number of ECF groups in the classification provided. Many ECFs within different families co-exist with cognate regulatory partners like anti-sigma factors, which seem to be specific of their own groups. This might be indicative of the co-evolution between ECF and anti-sigma factor sequences. Besides anti-sigmas, two-component systems, serine/threonine kinases and other regulators such as TetR repressors are often found in the vicinity of ECFs (Casas-Pastor et al., 2019).

1.2.1.2.2 Stress responsive ECF σ sigma factors

Traditionally, ECFs mediate the response to extracellular signals, and proteins with these role can be found in different ECF families. For instance, they can respond to alkaline shock (*e.g. B. subtilis* σ^W), nitrogen starvation (*e.g. S. coelicolor* σ^T), heat shock (*e.g. M. tuberculosis* σ^H), or cell wall-targeted antibiotics (*e.g. B. subtilis* σ^v) (Feng et al., 2011; Ho and Ellermeier, 2019; Raman et al., 2001; Wiegert et al., 2001).

With the identification of hundreds of new ECFs, nowadays, it is well accepted that ECFs are not only operators from the extracytoplasmic sensing, but they are also responsible for mediating gene expression in response to intracellular cues. These include regulators of morphological transitions (*e.g. S. venezuelae* BldN), virulence (*e.g. Pseudomonas aeruginosa* VreI), the thiol-disulfide redox balance (*e.g. S. coelicolor* σ^R) and metal resistance and homeostasis (*e.g. Cupriavidus metallidurans* RpoE) (Bibb et al., 2012; Grosse et al., 2007; Llamas et al., 2009; Paget et al., 1998).

1.2.1.2.3 Posttranscriptional regulation of ECFs

ECFs are often auto-regulators of their own expression, by positive feedback loops, but their runaway activity can be prevented post-transcriptionally by interacting with accessory proteins, in most cases located at the same operon. While some are autonomous regulators and lack a regulatory partner – the so called 'orphan' ECFs – others require a dedicated repressor, usually co-expressed with the ECF, that prevents the polar activity of the ECF under standard conditions, the anti-sigma factor.

1.2.1.2.3.1 The anti-sigma factor

ECF antagonists do not have typical protein conserved domains and usually share low sequence similarity. With the advent of X-ray crystallography in the last decade, crystal structure information on sigma/anti-sigma complexes helped to characterise a number of sigma antagonists. These studies were fundamental to define distinct classes of anti-sigma factors, since they unveiled that well conserved structural features determine the anti-sigma mode of action, despite the low sequence identity.

One important preserved feature is the anti-sigma factor domain (ASD), characterised by the specific folding of alpha-helix bundles that will enable the anti-sigma factor to contact and confine the ECF core domains, to prevent their binding to the RNAP or to the promoter DNA. ASDs of class I were found in well studied anti-sigma factors like *E. coli* RseA, *Rhodobacter sphaeroids* ChrR or *B. subtillis* RsiW (Campbell et al., 2007; Campbell et al., 2003; Devkota et al., 2017) and are characterised by a three-helix bundle fold followed by a short helix (Figure 5A). Class II ASDs, such as the domains in *C. metallidurans* CnrY (Maillard et al., 2014) or NepR from *Sphingomonas sp*. (Campagne et al., 2012), exhibit a small two helix clamp (Figure 5B). Class III ASDs were identified in *S. venezuelae* RsbN (Figure 5C) which comprises three conserved helices with less organised linker regions (Schumacher et al., 2018). This shows that despite the class of the ASD involved in the interaction with the ECF, these complexes adopt a compact shape that renders the ECF inactive, as it requires an extended conformation of the σ_2-σ_4 region to bind to promoter DNA (Feklistov et al., 2014). More recently, studies on *Clostridium thermocellum* SigI-RsgI complexes, described a new class of ASDs (Figure 5D). Unlike the other classes in which the ASD contacts the σ2 and σ4 region, RsgI binds to SigI C-terminal region to prevent the sigma factor binding to the -35 region in the DNA (Wei et al., 2019).

In order to be activated, the ECF is released from the ASD which can be accomplished through different mechanisms (Mascher, 2013). Anti-sigma factors often have CT transmembrane domains that promote its attachment to the cell membrane, leaving the NT part in the cytosol to enable its binding to anti-sigma factor (Figure 6). Upon a trigger signal, the anti-sigma factor can suffer either regulated proteolysis (e.g. RsbN σ^W-RsiW (Ho and Ellermeier, 2012) or conformational changes (e.g. CnrH-CnrY (Trepreau et al., 2011) that prompt the release of the cytoplasmic sigma factor from the inhibitory grip.

A recent breakthrough on σ/anti-σ complex structures described a novel mechanism of modulation of the interaction between the two proteins, which is dependent on the presence of a small molecule linking the two proteins that promotes a tight contact interface. In *S. venezuelae* WhiG-RsiG complexes, the inhibition of WhiG is controlled by the levels of c-di-GMP required to form a c-di-GMP dimer that mediates this interaction (Gallagher et al., 2019). WhiG is not an ECF sigma factor, since it comprises σ2, σ3 and σ4 domains. Strikingly, the mode of action of this complex described in this work seems to be confined to the ECF core region of this protein. The interface between WhiG σ2 and RsiG alpha helices forms a loop that encircles the c-di-GMP dimer. This unique conformation constitutes a three-layered inhibitory grip that 1) depends on the availability of a ligand, 2) increases the distance between the σ2 and σ4

domains - which prevents binding to the RNAP- and 3) generates points of clash with the promoter DNA. Overall, studies on ASDs have brought new avenues on the regulation of sigma factor activity, which is ensured by several layers of processes that are intimately dependent on the biochemical nature of each anti-sigma factor.

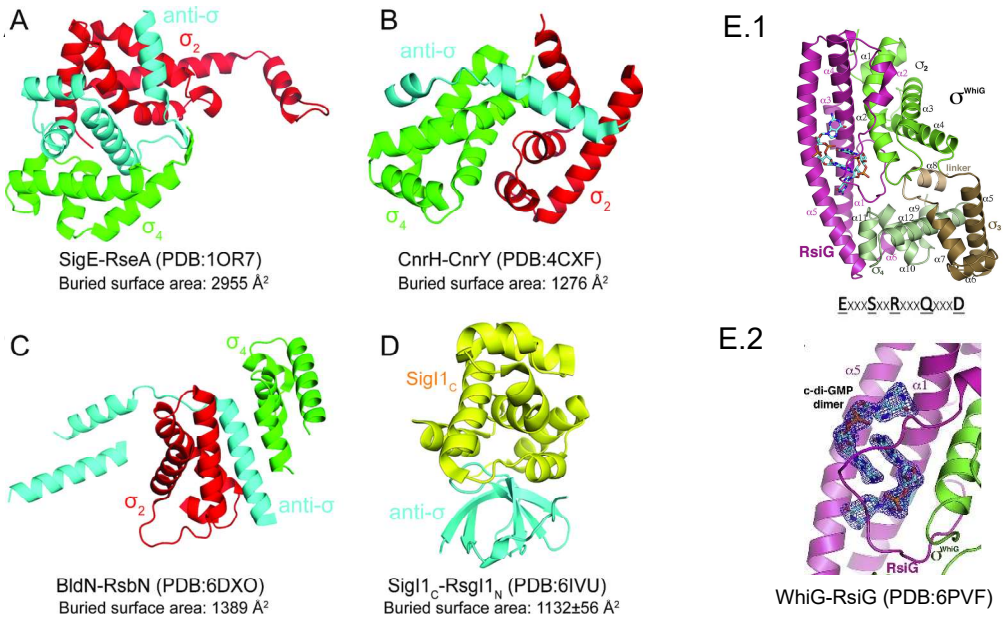


Figure 5. The anti-sigma factors. Crystal structural information on σ/anti-σ protein complexes identified four classes of σ cognate antagonists. In each complex, the anti-sigma factor is labelled in cyan. The σ_2 domain is in red and the σ_4 is in green, except for SigI, which is represented by its C-terminal region in yellow. *E.coli* SigE-RseA (**A**); *C. metallidurans* CnrH-CnrY (**B**); *S. venezuelae* BldN-RsbN (**C**); *C. thermocellum* Sig-RsgI (**D**) and *S. venezuelae* WhiG-RsiG (**E.1**). Detail of *S. venezuelae* WhiG-RsiG showing the c-di-GMP dimer-binding region (**E.2**). Adapted from (Wei et al., 2019) and (Gallagher et al., 2019).

1.2.1.2.3.2 Other regulatory partners

The switch from inactive to active forms of sigma factors has been extensively investigated. Different studies clearly show that various signal transduction processes, with interconnected mechanisms, control of the sigma factor release to achieve a more stringent inactivation of the sigma factor.

While some systems exclusively depend on the σ/anti-σ proteins, other systems require the participation of accessory proteins that work together with the ECF and the anti-sigma through direct or indirect interactions. The oxidative stress responsive σ^R-RsrA is an example of mechanism strictly dependent on the ECF and its antagonist, in which σ^R is activated by changes on the redox state and metal availability in the cytoplasm. Metal ions (Zn^{2+}) specifically bind to zinc-containing anti-sigma (ZAS) conserved residues in the anti-σ^R, RsrA. Upon exposure to thiol oxidizing agents zinc is

released from RsrA and the two cysteines in RsrA are trapped in a disulfide bond, which promotes the inactive conformation of RsrA, and releases σ^R to activate its redox stress-responsive regulon (Park et al., 2019).

One interesting example of control of σ/anti-σ unbinding by additional interacting partners is the programmed proteolysis found in *E. coli* σ^E, characterised by a three-step proteolytic degradation of the membrane associated RseA (reviewed in (Mitchell and Silhavy, 2019). An inner membrane site-1-protease DegS starts the process by cleaving the periplasmic domain of RseA. Then, the site-2-protease RseP cleaves RseA within the transmembrane domain, exposing the soluble fragment of RseA into the cytoplasm to be degraded by cytoplasmic proteases and free σ^E. In *B. subtilis*, there is a second interesting example of regulated proteolysis, the destruction of the anti-sigma factor for σ^V, RsiV. Binding of lysozyme to RsiV, triggers a three-step programmed proteolytic degradation of RsiV and thus activation of σ^V (Ho and Ellermeier, 2019). A first degradation step is carried out by a major signal peptidase (Sip) that acts as a site-1 peptidase for this anti-sigma factor. The site-1 cleaved RsiV serves as a substrate for the RasP site-2 protease and, finally, the extracted cytoplasmic portion of RsiV becomes exposed to cytosolic proteases, which release σ^V. Binding of RsiV to the lysozyme active site pocket inactivates the enzyme, and ensures the stringency of the lysozyme resistance system (Figure 6).

Another well-studied regulation mechanism that requires σ accessory proteins is the partner switch regulation (Figure 6). Canonical systems are composed of an anti-sigma factor with serine kinase activity, a phosphorylatable anti-σ antagonist (Sulphate Transporter and Anti-sigma antagonist - STAS - domains), a phosphatase, and a target sigma factor. In *Pseudomonas aeruginosa*, for instance, the phosphorylation of the anti-sigma factor antagonist HsbA promotes its binding the response regulator HsbR (the anti-sigma for RpoS). On the other hand, the non-phosphorylated HsbA binds to FlgM, the anti-sigma factor of FliA. The switch between the activation of the two sigma factors decides entering into biogenesis of biofilms (RpoS-dependent) or swimming (regulated by FliA). The phosphorylation state of HsbA is determined by an alternate activity of the HsbR serine kinase/anti-sigma (K/AS) and phosphatase domains (PP2C) (Bouillet et al., 2019).

More complex systems provide the inhibitory grip by combining classic ECF regulation with phosphorelay events traditionally observed in TCS (Figure 6). While serine/threonine kinases, like the HsbR K/AS domain, tend to produce a dynamic but stable phosphorylation event, usually associated to transitions between developmental stages, histidine kinases generate a more liable phosphorylation that favours a temporal control of a response regulator (Pompeo et al., 2016). The PhyR-dependent regulation

of σ^T in *Caulobacter crescentus* is an example of a histidine kinase-dependent phosphorelay. PhyR is a regulator composed of paralogues of ECF sigma factors core domains, the σ-like domain. NepR is the anti-sigma factor that sequesters σ^T when PhyR is inactive. Upon a stress condition, PhyR is phosphorylated by the sensor histidine kinase PhyK and competes with σ^T, binding to NepR with higher affinity, to release σ^T and activate the stress response (Herrou et al., 2012; Herrou et al., 2015).

Earlier this year, Iyer and collaborators uncovered a mode of ECF control that is independent of an anti-sigma factor in *Vibrio parahaemolyticus* (Iyer et al., 2020). In these bacteria, EcfP is inactive until the threonine kinase PknT phosphorylates Thr63 within the EcfP σ2 domain, to provide the negative charge required for the ECF to contact the RNAP and drive transcription of a polymyxin-resistant regulon (Figure 6).

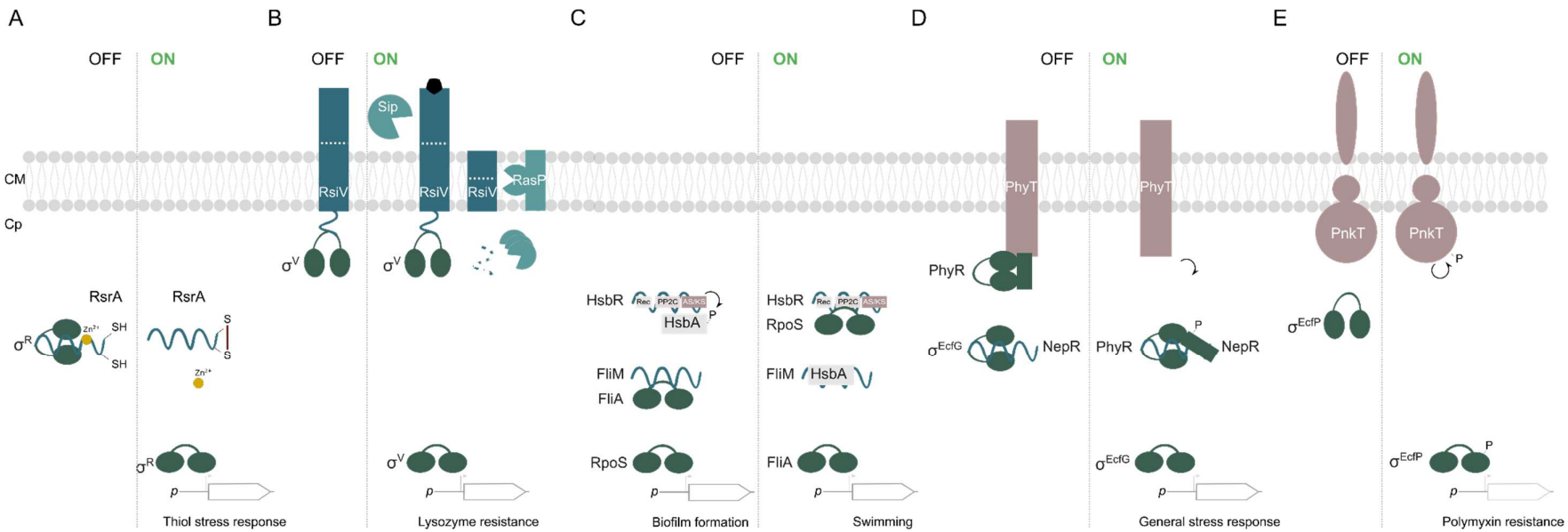


Figure 6. Schematic illustration of the different modes of regulating ECF σ activity. Individual conformational change, *S. coelicolor* RsrA-σ^R (**A**). Regulated proteolysis*, B. subtillis* RsiV-σ^V (**B**). Sigma factor mimicry/partner-switch, *P. aeruginosa* HsbR-RpoS and FliM-FliA (**C**). Ser/Thr kinases, *Vibrio parahaemolyticus* EcfP (**D**) Combined regulatory mechanisms, *C. crescentus* NepR-σ^T (**E**). ECF σ factors are coloured dark green (note that PhyR contains ECF orthologue domains but it is not an ECF); anti-σ factors are represented by blue helices, or blue rectangles if they have transmembrane domains; accessory proteases are in light green; proteins/domains with kinase activity are coloured pink. Phosphorylation-dependent signalling is depicted with a curved arrow and the letter 'P'. OFF/ON indicates absence or presence of inducing conditions. CM: cell membrane; Cp: cytoplasm; AS/SK: anti-sigma factor/serine threonine kinase domain; PP2C: phosphatase domain; S-S: disulfide bond formation, S-S or SR-SR. Adapted from the following studies: (Bouillet et al., 2019; Herrou et al., 2015; Ho and Ellermeier, 2019; Iyer et al., 2020; Park et al., 2019).

1.2.1.2.3.3 C-terminal (CT) extensions

Throughout time, biochemical studies have been showing that the discovery of ECF antagonists involves more challenges than initially thought, mainly because classical sequence homology analyses against known anti-sigma factors have proven to be unsuccessful. This trait has made the ECF-ASF classification efforts more difficult, and has been the basis for researchers to propose that ECFs with no obvious anti-sigma factors could be auto-regulatory through internal σ domains, replacing the anti-sigma inhibitory function (Casas-Pastor et al., 2019; Huang et al., 2015; Staron et al., 2009). In fact, this classification has brought to light a number of ECFs that have complex domain compositions, with additional functional domains fused to the ECF σ2-σ4 core. Many of these extra domains are present at the C-terminus region of the protein (found in 19 ECF families) but only a few recent studies have been focusing on describing and assigning them an auto-regulatory function.

Experimental data on ECFs with CTs have shed light on the auto-regulatory role these domains, which promote specific conformations on the ECF to inhibit its activity. A well-studied example is ECF41 (σ^{ECF41}) from *Bacillus licheniformis,* which contains a NTF2-like domain that binds to the ECF core to prevent its own activity (Wecke et al., 2012; Wu et al., 2019). Other types of CT extensions encoding Pfam domains are able to perform this flexible protein folding, like as the tetratricopeptide repeat (TPR) at *B. licheniformis* ECF42 CT (Liu et al., 2018). Nonetheless, not all compact conformations of the ECF are favourable to an inhibitory function. For instance, the cysteine rich domain (CRD) at *Myxococcus xanthus* CorE and CorE2 or the SnoaL-like domain in the *M. tuberculosis* sigma J (σ^{J}_{Mtub}) were shown to promote DNA binding and therefore a more effective ECF activity (Gómez-Santos et al., 2011; Goutam et al., 2017; Marcos-Torres et al., 2016).

Further ECF CT extensions of different natures have been identified, but the lack of experimental data on the activity of these ECFs leaves their physiological roles elusive. Some examples are the presence of redox-sensitive ZAS domain in ECF48, several transmembrane helices in ECF264, the conserved CGxxGxGxCxC identified in ECF288 or the fused carbohydrate-binding domain (CBD) in ECF53 (Casas-Pastor et al., 2019; Huang et al., 2015).

1.3 Hypothesis and aims

This study aims at consolidating our understanding on how bacteria respond to inputs from the environment and re-arranges the molecular machinery to ensure cellular homeostasis and strive. In *S. tsukubaensis* the number of ECFs deviates from the average, which is indicative of a large investment in regulating transcription. Within this

array of sigma factors, we find complex proteins that differ from the traditional ECF domain topology, by harbouring conserved extensions, a lot of them with functions that remain uncovered. In particular, these studies focused on the biological role of STSU_11560, an ECF that contains a SnoaL_2 CT extension, of which molecular significance had not been addressed.

The work in this thesis combined genome-wide approaches with biochemical studies to unlock the mechanisms behind STSU_11560-dependent regulatory cascades in *Streptomyces*. In particular, we established two major objectives:

1) We investigated if STSU_11560 could be controlled post-translationally, by autoregulation through the CT extension, or by a cognate anti-sigma factor.
2) We sought to identify the stimulus/stimuli to which STSU_11560 responds and further define its role by unravelling which genes are under the control STSU_11560.

1.4 References

Al-Bassam, M.M., Haist, J., Neumann, S.A., Lindenberg, S., and Tschowri, N. (2018). Expression Patterns, Genomic Conservation and Input Into Developmental Regulation of the GGDEF/EAL/HD-GYP Domain Proteins in Streptomyces. Frontiers in Microbiology *9*, 2524-2524.

Barksdale, L. (1970). Corynebacterium diphtheriae and its relatives. Bacteriological reviews *34*, 378-422.

Barreiro, C., and Martinez-Castro, M. (2014). Trends in the biosynthesis and production of the immunosuppressant tacrolimus (FK506). Appl Microbiol Biotechnol *98*, 497-507.

Barreiro, C., Prieto, C., Sola-Landa, A., Solera, E., Martínez-Castro, M., Pérez-Redondo, R., García-Estrada, C., Aparicio, J.F., Fernández-Martínez, L.T., Santos-Aberturas, J.*, et al.* (2012). Draft genome of Streptomyces tsukubaensis NRRL 18488, the producer of the clinically important immunosuppressant tacrolimus (FK506). Journal of Bacteriology *194*, 3756-3757.

Beites, T., Oliveira, P., Rioseras, B., Pires, S.D.S., Oliveira, R., Tamagnini, P., Moradas-Ferreira, P., Manteca, Á., and Mendes, M.V. (2015). Streptomyces natalensis programmed cell death and morphological differentiation are dependent on oxidative stress. Scientific Reports *5*, 12887.

Bibb, M.J., Domonkos, Á., 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. Molecular Microbiology *84*, 1033-1049.

Bouillet, S., Ba, M., Houot, L., Iobbi-Nivol, C., and Bordi, C. (2019). Connected partner-switches control the life style of Pseudomonas aeruginosa through RpoS regulation. Scientific Reports *9*, 6496-6496.

Boylan, S.A., Redfield, A.R., Brody, M.S., and Price, C.W. (1993). Stress-induced activation of the sigma B transcription factor of Bacillus subtilis. Journal of Bacteriology *175*, 7931-7937.

Bush, M.J. (2018). The actinobacterial WhiB-like (Wbl) family of transcription factors. Molecular Microbiology *110*, 663-676.

Bush, M.J., Tschowri, N., Schlimpert, S., Flärdh, K., and Buttner, M.J. (2015). c-di-GMP signalling and the regulation of developmental transitions in streptomycetes. Nature Reviews Microbiology *13*, 749-760.

Buttner, M.J., and Lewis, C.G. (1992). Construction and characterization of Streptomyces coelicolor A3(2) mutants that are multiply deficient in the nonessential hrd-encoded RNA polymerase sigma factors. Journal of Bacteriology *174*, 5165-5167.

Campagne, S., Damberger, F.F., Kaczmarczyk, A., Francez-Charlot, A., Allain, F.H.-T., and Vorholt, J.A. (2012). Structural basis for sigma factor mimicry in the general stress response of Alphaproteobacteria. Proceedings of the National Academy of Sciences *109*, E1405-E1414.

Campbell, E.A., Greenwell, R., Anthony, J.R., Wang, S., Lim, L., Das, K., Sofia, H.J., Donohue, T.J., and Darst, S.A. (2007). A Conserved Structural Module Regulates Transcriptional Responses to Diverse Stress Signals in Bacteria. Molecular Cell *27*, 793-805.

Campbell, E.A., Tupy, J.L., Gruber, T.M., Wang, S., Sharp, M.M., Gross, C.A., and Darst, S.A. (2003). Crystal Structure of Escherichia coli σ^E with the Cytoplasmic Domain of Its Anti-σ RseA. Molecular Cell *11*, 1067-1078.

Casadesús, J., and Low, D.A. (2013). Programmed Heterogeneity: Epigenetic Mechanisms in Bacteria. The Journal of Biological Chemistry *288*, 13929-13935.

Casas-Pastor, D., Müller, R.R., Becker, A., Buttner, M., Gross, C., Mascher, T., Goesmann, A., and Fritz, G. (2019). Expansion and re-classification of the extracytoplasmic function (ECF) σ factor family. bioRxiv, 2019.2012.2011.873521.

Chang, C., Li, K., Jiang, S., Li, B., Cao, L., and Wang, P. (2019). Downregulation of TRPC6 expression is a critical molecular event during FK506 treatment for overactive bladder. Cell Calcium *77*, 8-19.

Claessen, D., and Errington, J. (2019). Cell Wall Deficiency as a Coping Strategy for Stress. Trends in Microbiology *27*, 1025-1033.

Claessen, D., Stokroos, I., Deelstra, H.J., Penninga, N.A., Bormann, C., Salas, J.A., Dijkhuizen, L., and Wösten, H.A.B. (2004). The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlins and chaplins. Molecular Microbiology *53*, 433-443.

Davis, B., Hilgart, D., Erickson, S., Labroo, P., Burton, J., Sant, H., Shea, J., Gale, B., and Agarwal, J. (2019). Local FK506 delivery at the direct nerve repair site improves nerve regeneration. Muscle & Nerve *60*, 613-620.

Devkota, S.R., Kwon, E., Ha, S.C., Chang, H.W., and Kim, D.Y. (2017). Structural insights into the regulation of Bacillus subtilis SigW activity by anti-sigma RsiW. PLOS ONE *12*, e0174284.

Eichelberg, K., and Galán, J.E. (2000). The Flagellar Sigma Factor FliA σ28 Regulates the Expression of *Salmonella* Genes Associated with the Centisome 63 Type III Secretion System. Infection and immunity *68*, 2735-2743.

Elliot, M.A., Karoonuthaisiri, N., Huang, J., Bibb, M.J., Cohen, S.N., Kao, C.M., and Buttner, M.J. (2003). The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in Streptomyces coelicolor. Genes Dev *17*, 1727-1740.

Fang, C., Li, L., Shen, L., Shi, J., Wang, S., Feng, Y., and Zhang, Y. (2019). Structures and mechanism of transcription initiation by bacterial ECF factors. Nucleic Acids Research *47*, 7094-7104.

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*, e00815-00817.

Feklistov, A., Sharon, B.D., Darst, S.a., and Gross, C.a. (2014). Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. Annual review of microbiology, 357-376.

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. Applied Microbiology and Biotechnology *92*, 1009-1021.

Flärdh, K., and Buttner, M.J. (2009). Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. Nature Reviews Microbiology *7*, 36-49.

Flardh, K., Richards, D.M., Hempel, A.M., Howard, M., and Buttner, M.J. (2012). Regulation of apical growth and hyphal branching in Streptomyces. Curr Opin Microbiol *15*, 737-743.

Francis, V.I., and Porter, S.L. (2019). Multikinase Networks: Two-Component Signaling Networks Integrating Multiple Stimuli. Annual Review of Microbiology *73*, 199-223.

Gaballa, A., Guariglia-Oropeza, V., Dürr, F., Butcher, B.G., Chen, A.Y., Chandrangsu, P., and Helmann, J.D. (2018). Modulation of extracytoplasmic function (ECF) sigma factor promoter selectivity by spacer region sequence. Nucleic Acids Research *46*, 134-145.

Gallagher, K.A., Schumacher, M.A., Bush, M.J., Bibb, M.J., Chandra, G., Holmes, N.A., Zeng, W., Henderson, M., Zhang, H., Findlay, K.C.*, et al.* (2019). c-di-GMP Arms an Anti-σ to Control Progression of Multicellular Differentiation in *Streptomyces*. Molecular Cell *77*, 586-599.e586.

Geiman, D.E., Raghunand, T.R., Agarwal, N., and Bishai, W.R. (2006). Differential Gene Expression in Response to Exposure to Antimycobacterial Agents and Other Stress Conditions among Seven Mycobacterium tuberculosis whiB-Like Genes. Antimicrobial agents and chemotherapy *50*, 2836-2841.

Gerber, N.N., and Lechevalier, H.A. (1965). Geosmin, an Earthy-Smelling Substance Isolated from Actinomycetes Downloaded from.

Gómez-Santos, N., Pérez, J., Sánchez-Sutil, M.C., Moraleda-Muñoz, A., and Muñoz-Dorado, J. (2011). CorE from Myxococcus xanthus Is a Copper-Dependent RNA Polymerase Sigma Factor. PLoS Genetics *7*, e1002106-e1002106.

Goutam, K., Gupta, A.K., and Gopal, B. (2017). The fused SnoaL-2 domain in the *Mycobacterium tuberculosis* sigma factor σjmodulates promoter recognition. Nucleic Acids Research *45*, 9760-9772.

Grosse, C., Friedrich, S., and Nies, D.H. (2007). Contribution of Extracytoplasmic Function Sigma Factors to Transition Metal Homeostasis in *Cupriavidus metallidurans* Strain CH34. J Mol Microbiol Biotechnol *12*, 227-240.

Grossman, A.D., Erickson, J.W., and Gross, C.A. (1984). The htpR gene product of E. coli is a sigma factor for heat-shock promoters. Cell *38*, 383-390.

Güell, M., Yus, E., Lluch-Senar, M., and Serrano, L. (2011). Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nat Rev Micro *9*, 658-669.

Guzina, J., and Djordjevic, M. (2017). Mix-and-matching as a promoter recognition mechanism by ECF σ factors. BMC Evolutionary Biology *17*, 12-12.

Halton, M. (2018). Petrichor: why does rain smell so good? BBC News.

Helmann, J.D. (2002). The extracytoplasmic function (ECF) sigma factors. Advances in microbial physiology *46*, 47-110.

Helmann, J.D. (2019). Where to Begin? Sigma factors and the selectivity of transcription initiation in Bacteria. Molecular Microbiology, mmi.14309-mmi.14309.

Hemmersbach-Miller, M., Stout, J.E., Woodworth, M.H., Cox, G.M., and Saullo, J.L. (2018). Nocardia infections in the transplanted host. Transplant infectious disease : an official journal of the Transplantation Society *20*, e12902-e12902.

Herrou, J., Rotskoff, G., Luo, Y., Roux, B., and Crosson, S. (2012). Structural basis of a protein partner switch that regulates the general stress response of α-proteobacteria. Proceedings of the National Academy of Sciences *109*, E1415–E1423.

Herrou, J., Willett, J.W., and Crosson, S. (2015). Structured and Dynamic Disordered Domains Regulate the Activity of a Multifunctional Anti-σ Factor. mBio *6*, e00910-00915.

Ho, T.D., and Ellermeier, C.D. (2019). Activation of the Extracytoplasmic Function σ Factor σV by Lysozyme. Molecular Microbiology, mmi.14348-mmi.14348.

Hoskisson, P.A., and van Wezel, G.P. (2019). Streptomyces coelicolor. Trends in Microbiology.

Huang, X., Pinto, D., Fritz, G., and Mascher, T. (2015). Environmental Sensing in Actinobacteria: a Comprehensive Survey on the Signaling Capacity of This Phylum. Journal of Bacteriology *197*, 2517-2535.

Imlay, J., Chin, S., and Linn, S. (1988). Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. Science *240*, 640-642.

Imlay, J.A. (2015). Diagnosing oxidative stress in bacteria: not as easy as you might think. Current Opinion in Microbiology *24*, 124-131.

Iyer, S.C., Casas-Pastor, D., Kraus, D., Mann, P., Schirner, K., Glatter, T., Fritz, G., and Ringgaard, S. (2020). Transcriptional regulation by σ factor phosphorylation in bacteria. Nature Microbiology.

Jiang, H., Yamamoto, S., Nishikawa, K., and Kato, R. (1993). Anti-tumor-promoting action of FK506, a potent immunosuppressive agent. Carcinogenesis *14*, 67-71.

Jishage, M., and Ishihama, A. (1999). Transcriptional Organization and In Vivo Role of the *Escherichia coli rsd* Gene, Encoding the Regulator of RNA Polymerase Sigma D. Journal of Bacteriology *181*, 3768-3776.

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

Jones, S.E., and Elliot, M.A. (2017). Streptomyces Exploration: Competition, Volatile Communication and New Bacterial Behaviours. Trends in Microbiology *xx*, 522-531.

Jones, S.E., Pham, C.A., Zambri, M.P., McKillip, J., Carlson, E.E., and Elliot, M.A. (2019). Streptomyces Volatile Compounds Influence Exploration and Microbial Community Dynamics by Altering Iron Availability. mBio *10*, e00171-00119.

Kino, T., Hatanaka, H., Hashimoto, M., Nishiyama, M., Goto, T., Okuhara, M., Kohsaka, M., Aoki, H., and Imanaka, H. (1987). FK-506, a novel immunosuppressant isolated from a Streptomyces. I. Fermentation, isolation, and physico-chemical and biological characteristics. J Antibiot (Tokyo) *40*, 1249-1255.

Krentz, A.J., Dmitrewski, J., Mayer, D., McMaster, P., Buckels, J., Smith, J.M., and Nattrass, M. (1994). Tacrolimus (FK506) versus cyclosporin in prevention of liver allograft rejection. Lancet (London, England) *344*, 948-949.

Lane, W.J., and Darst, S.A. (2006). The structural basis for promoter -35 element recognition by the group IV σ factors. PLoS biology *4*, 1491-1500.

Lee, J.-Y., Kim, H.-J., Kim, E.-S., Kim, P., Kim, Y., and Lee, H.-S. (2013). Regulatory interaction of the Corynebacterium glutamicum whc genes in oxidative stress responses. Journal of biotechnology *168*, 149-154.

Lin, W., Mandal, S., Degen, D., Cho, M.S., Feng, Y., Das, K., and Ebright, R.H. (2019). Structural basis of ECF-σ-factor-dependent transcription initiation. Nature Communications *10*, 710-710.

Liu, Q., Pinto, D., and Mascher, T. (2018). Characterization of the Widely Distributed Novel ECF42 Group of Extracytoplasmic Function σ Factors in Streptomyces venezuelae. Journal of Bacteriology *200*, e00437-00418.

Llamas, M.A., van der Sar, A., Chu, B.C.H., Sparrius, M., Vogel, H.J., and Bitter, W. (2009). A Novel Extracytoplasmic Function (ECF) Sigma Factor Regulates Virulence in Pseudomonas aeruginosa. PLoS pathogens *5*, e1000572.

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

Lonetto, M.A., Brown, K.L., Rudd li, K.E., and Buttner, M.J. (1994). Analysis of the Streptomyces coelicolor sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase or factors involved in the regulation of extracytoplasmic functions. Microbiology *91*, 7573-7577.

Maillard, A.P., Girard, E., Ziani, W., Petit-Härtlein, I., Kahn, R., and Covès, J. (2014). The Crystal Structure of the Anti-σ Factor CnrY in Complex with the σ Factor CnrH Shows a New Structural Class of Anti-σ Factors Targeting Extracytoplasmic Function σ Factors. Journal of Molecular Biology *426*, 2313-2327.

Manteca, A., Sanchez, J., Jung, H.R., Schwammle, V., and Jensen, O.N. (2010). Quantitative proteomics analysis of Streptomyces coelicolor development demonstrates that onset of secondary metabolism coincides with hypha differentiation. Mol Cell Proteomics *9*, 1423-1436.

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

Mascher, T. (2013). Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. Current Opinion in Microbiology *16*, 148-155.

McCormick, J.R., and Flärdh, K. (2012). Signals and regulators that govern Streptomyces development. FEMS Microbiology Reviews *36*, 206-231.

Miao, V., and Davies, J. (2010). Actinobacteria: the good, the bad, and the ugly. Antonie van Leeuwenhoek *98*, 143-150.

Mitchell, A.M., and Silhavy, T.J. (2019). Envelope stress responses: balancing damage repair and toxicity. Nature Reviews Microbiology *17*, 417–428.

Nagrani, N.K., and Zito, P.M. (2019). Topical Tacrolimus (FK506, Protopic) in the Treatment of Atopic Dermatitis. Journal of the Dermatology Nurses' Association *11*, 41-44.

Österberg, S., Peso-Santos, T.d., and Shingler, V. (2011). Regulation of Alternative Sigma Factor Use. Annual Review of Microbiology *65*, 37-55.

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

Paget, M.S.B., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. (1999). Evidence that the Extracytoplasmic Function Sigma Factor σE Is Required for Normal Cell Wall Structure in Streptomyces coelicolor A3. *3*, 204-211.

Park, J.H., Lee, J.H., and Roe, J.H. (2019). SigR, a hub of multilayered regulation of redox and antibiotic stress responses. Molecular Microbiology, mmi.14341-mmi.14341.

Pompeo, F., Foulquier, E., and Galinier, A. (2016). Impact of Serine/Threonine Protein Kinases on the Regulation of Sporulation in Bacillus subtilis. Frontiers in Microbiology *7*.

Raman, S., Song, T., Puyang, X., Bardarov, S., Jacobs, W.R., and Husson, R.N. (2001). The Alternative Sigma Factor SigH Regulates Major Components of Oxidative and Heat Stress Responses in *Mycobacterium tuberculosis*. Journal of Bacteriology *183*, 6119-6125.

Ruff, E.F., Record, M.T., and Artsimovitch, I. (2015). Initial Events in Bacterial Transcription Initiation. Biomolecules *5*, 1035-1062.

Schlimpert, S., Wasserstrom, S., Chandra, G., Bibb, M.J., Findlay, K.C., Flärdh, K., and Buttner, M.J. (2017). Two dynamin-like proteins stabilize FtsZ rings during Streptomyces sporulation. Proceedings of the National Academy of Sciences, 201704612-201704612.

Schumacher, M.A., Bush, M.J., Bibb, M.J., Ramos-León, F., Chandra, G., Zeng, W., and Buttner, M.J. (2018). The crystal structure of the RsbN–σBldN complex from Streptomyces venezuelae defines a new structural class of anti-σ factor. Nucleic Acids Research *46*, 7405–7417.

Sowani, H., Kulkarni, M., Zinjarde, S., and Javdekar, V. (2017). Gordonia and Related Genera as Opportunistic Human Pathogens Causing Infections of Skin, Soft Tissues, and Bones. The Microbiology of Skin, Soft Tissue, Bone and Joint Infections, 105-121.

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*, 557-581.

Tang, L., Ebara, S., Kawasaki, S., Wakabayashi, S., Nikaido, T., and Takaoka, K. (2002). FK506 enhanced osteoblastic differentiation in mesenchymal cells. Cell Biology International *26*, 75-84.

Tenconi, E., Traxler, M.F., Hoebreck, C., van Wezel, G.P., and Rigali, S. (2018). Production of Prodiginines Is Part of a Programmed Cell Death Process in Streptomyces coelicolor. Frontiers in Microbiology *9*, 1742-1742.

Traxler, M.F., Seyedsayamdost, M.R., Clardy, J., and Kolter, R. (2012). Interspecies modulation of bacterial development through iron competition and siderophore piracy. Mol Microbiol *86*, 628-644.

Tschowri, N., Schumacher, M.A., Schlimpert, S., Chinnam, N.B., Findlay, K.C., Brennan, R.G., and Buttner, M.J. (2014). Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. Cell *158*, 1136-1147.

Typas, A., Becker, G., and Hengge, R. (2007). The molecular basis of selective promoter activation by the σS subunit of RNA polymerase. Molecular Microbiology *63*, 1296-1306.

Ulrich, L.E., Koonin, E.V., and Zhulin, I.B. (2005). One-component systems dominate signal transduction in prokaryotes. Trends in Microbiology *13*, 52-56.

Wade, J.T., and Grainger, D.C. (2014). Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. Nature Reviews Microbiology *12*, 647-653.

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*, 194-213.

Wei, Z., Chen, C., Liu, Y.-J., Dong, S., Li, J., Qi, K., Liu, S., Ding, X., Ortiz de Ora, L., Muñoz-Gutiérrez, I.*, et al.* (2019). Alternative σI/anti-σI factors represent a unique form of bacterial σ/anti-σ complex. Nucleic Acids Research *47*, 5988-5997.

Wiegert, T., Homuth, G., Versteeg, S., and Schumann, W. (2001). Alkaline shock induces the Bacillus subtilis σW regulon. Molecular Microbiology *41*, 59-71.

Wu, H., Liu, Q., Casas-Pastor, D., Dürr, F., Mascher, T., and Fritz, G. (2019). The role of C-terminal extensions in controlling ECF σ factor activity in the widely conserved groups ECF41 and ECF42. Molecular Microbiology *112*, 498-514.

Yagüe, P., Willemse, J., Koning, R.I., Rioseras, B., López-García, M.T., Gonzalez-Quiñonez, N., Lopez-Iglesias, C., Shliaha, P.V., Rogowska-Wrzesinska, A., Koster, A.J.*, et al.* (2016). Subcompartmentalization by cross-membranes during early growth of Streptomyces hyphae. Nature Communications *7*, 12467.

Chapter 2

Streptomyces tsukubaensis differentiation is aided by the redox responsive ECF56σ/anti-σ pair, SigG1/RsfG

Rute Oliveira, [a, b, c] Matthew J. Bush,[e] Sílvia Pires,[b*] Govind Chandra,[e] Delia Casas-Pastor,[f] Georg Fritz,[f**] and Marta V. Mendes,[a,b,d#]

[a]Bioengineering and Synthetic Microbiology Group, i3S- Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

[b]IBMC, Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

[c]Programa Doutoral em Biologia Molecular e Celular (MCBiology), ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Rua de Jorge Viterbo Ferreira n. 228, 4050-313 Porto, Portugal

[d]ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Rua de Jorge Viterbo Ferreira n. 228, 4050-313 Porto, Portugal

[e]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[f]LOEWE-Zentrum für Synthetische Mikrobiologie, Philipps-Universität Marburg, 35032 Marburg, Germany

[*] Present address: Weill Cornell University, New York, NY 10065, United States of America

[**] Present address: School for Molecular Sciences, University of Western Australia, Perth 6009, Australia

Accepted for publication in Applied and Environmental Microbiology

Streptomyces tsukubaensis differentiation is aided by the redox responsive ECF56σ/anti-σ pair, SigG1/RsfG

Rute Oliveira,[a,b,c] Matthew J. Bush,[e] Sílvia Pires,[b*] Govind Chandra,[e] Delia Casas-Pastor,[f] Georg Fritz,[f**] and Marta V. Mendes,[a,b,d#]

[a]Bioengineering and Synthetic Microbiology Group, i3S- Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

[b]IBMC, Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

[c]Programa Doutoral em Biologia Molecular e Celular (MCBiology), ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal

[d]ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Rua de Jorge Viterbo Ferreira n. 228, 4050-313 Porto, Portugal

[e]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[f]LOEWE -Zentrum für Synthetische Mikrobiologie, Philipps -Universität Marburg, 35032 Marburg, Germany

Running Head: Biological role of a SnoaL2-ECFσ and its anti-σ factor

Address correspondence to Marta V. Mendes, mvm@ibmc.up.pt

* Present address: Weill Cornell University, New York, NY 10065, United States of America

** Present address: School for Molecular Sciences, University of Western Australia, Perth 6009, Australia

Word count for the abstract: 219

Word count for the text: 5442

2.1 ABSTRACT

Extracytoplasmic function (ECF) sigma factors are key transcriptional regulators that prokaryotes have evolved to respond to environmental challenges. Unlike other model bacteria, *Streptomyces tsukubaensis* harbours 42 ECFs to reprogram stress-responsive gene expression. SigG1 features minimal conserved ECF σ_2-σ_4 architecture and an additional C-terminal extension that encodes a SnoaL_2 domain, which is characteristic for ECF σ factors of group ECF56. Although proteins with such domain organisation are widely found among Actinobacteria, the functional role of ECFs with a fused SnoaL_2 domain (ECF-SnoaL_2) remains unknown. Traditionally, ECFs detect and respond to extracellular stimuli, but many ECFs transduce specific intracellular inputs. We performed next generation sequencing (NGS) RNA-seq and ChIP-seq experiments using strains with the disrupted SigG1/anti-σ pair. By integrating both datasets, we identified the role of ECF-SnoaL_2 SigG1 in the morphological differentiation programme. SigG1 regulates the expression of alanine dehydrogenase, *ald1* and the WhiB-like regulator C, *wblC* required for differentiation, in addition to iron and copper trafficking systems. Furthermore, we show that in addition to having an autoregulatory conformation mediated by the SnoaL_2 domain, SigG1 is controlled by a cognate anti-sigma protein RsfG, coded by a co-transcribed *sigG1*-neighbouring gene. Overall, our work establishes a model in which the activity of a σ factor of group ECF56 regulates morphogenesis and metal-ions homeostasis during development to ensure the timely progression of multicellular differentiation.

IMPORTANCE

Many regulators that participate in *Streptomyces* morphogenesis have been extensively characterised, but specific players that are expressed to fine-tune the redox state during developmental transitions are less understood. During electron are transfer in aerobic conditions, oxygen is converted into reactive oxygen species (ROS). This effect is unwelcome due to the damage it causes to biomolecules. Our results corroborate that maintaining a low, non-detrimental levels of intracellular ROS is beneficial for developmental progression and minimizes harmful effects on various cell components. We characterise a new ECF/anti-sigma system that governs a stress response to cytoplasmic oxidative signals. We propose a model whereby an ECF bearing a SnoaL_2 extension could be modulated by two layers of posttranslational regulation. While the presence of the SnoaL_2 extension indicates inter-domain dependent regulation, we show that coordination with a putative anti-σ factor might be essential to fine-tune the activity of this novel ECF.

INTRODUCTION

Streptomycetes are unique soil-dwelling bacteria with a prolific biosynthetic capacity responsible for hundreds of widely used bioactive compounds. In particular, *Streptomyces*

tsukubaensis NRRL18488 naturally produces FK506, also known as tacrolimus, which is frequently used as an immunosuppressant in clinical practice, for instance, to prevent organ rejection in transplantation patients, eczema or cancer (1-3). *Streptomyces* species follow a complex multicellular filamentous lifecycle that starts with the germination of a spore when the environmental conditions are favourable. Hyphae emerge that grow by tip-extension and branching, to form a dense filamentous vegetative mycelium that facilitates nutrient scavenging. When nutrients are depleted, a highly regulated process of differentiation is initiated. Vegetative growth is halted before the erection of aerial hyphae that extend upwards, away from the vegetative mycelium. Pivotal to this transition from the substrate into the air is the production of a hydrophobic coat that consists of two families of proteins, the rodlins and the chaplins (4, 5). Subsequently, cell division and chromosome segregation generate chains of uni-genomic spores. These spores can be dispersed, allowing the colonisation of new environmental niches in which nutrients are replete. Fully mature spores frequently acquire a strain-specific colour derived from a polyketide spore pigment produced during the late stages of development. In *S. tsukubaensis*, spores exhibit a grey pigmentation (6). The process of differentiation in *Streptomyces* has been largely studied in strains used as model organisms such as *S. coelicolor* and *S. venezuelae.* It is strictly dependent on gene networks which are controlled by the Bld (bald) and the Whi (white) family of regulators (7-9). The Bld proteins orchestrate the transition to the aerial lifestyle and so disruption of *bld* genes blocks the erection of aerial mycelium (a “bald” phenotype). The Whi proteins mediate the conversion of aerial hyphae into spores and so disruption of *whi* genes results in a block at the aerial growth stage (a “white” phenotype). The morphological differentiation process overlaps with the onset of *Streptomyces* specialized metabolism and the production of secondary metabolites.

In order to survive complex niches, *Streptomyces*, as many other bacteria genera, have developed sophisticated response mechanisms to environmental stress signals. Many such mechanisms culminate in the regulation of transcription to provide the necessary molecular adaptation. In *Streptomyces*, the main mechanisms that orchestrate external signal-sensing and amplification include the autoregulatory DNA-binding transcription factors (10), two-component systems (11) and the extracytoplasmic function (ECF) sigma factors (12). In ECF-dependent systems, the sigma factor directly controls gene expression by recruitment of the RNA polymerase (RNAP) to specific promoters. Control is mediated via the binding of the ECF-sigma by its cognate anti-sigma factor, usually coded by a gene that is co-transcribed with the ECF-sigma. In response to environmental stimuli, the anti-sigma releases the ECF-sigma, allowing the cell to moderate gene expression and so respond to environmental fluctuations (reviewed in (13)).

Bacteria harbour a variable number of ECF-sigmas (ECFs) to compete and direct the transcription of specific classes of stress-responsive genes. ECFs form the largest family of sigma factors, the group IV, which derive from *Escherichia coli* σ^{70} and display the minimal architecture required to bind DNA, the sigma σ^{70}_region 2 (σ_2) and σ^{70}_ region 4 (σ_4) conserved domains, which bind to the -10 and -35 promoter regions, respectively (14, 15). The size of the linker region between σ2 and σ4 has been described to have an active regulatory role in both the recognition of the bipartite sequence motif and the binding to the RNAP (16, 17). Canonical ECFs recognise conserved motifs in the -10 and the -35 boxes, often separated by an optimal spacer of 16-17bp (14) to accommodate the binding of σ_2 and σ_4 to the promoters.

Despite the discovery of ECFs in 1994 (13), detailed information regarding how these proteins evolved is still sparse. ECFs show great diversity and have been classified in 157 well-defined phylogenetic families (18-21). They are ubiquitously distributed and are more abundant in *Streptomyces* genomes than in other well-studied bacteria. In some of these families, ECFs harbour long extensions at the N and/or C terminal (NT and CT) encoding other conserved domains, besides the σ_2 and σ_4 DNA-binding domains (19, 22). A number of ECFs with CT extensions have been experimentally characterised (reviewed in (23)), and in three of them the CT has been shown to perform a self-regulatory role that has replaced the anti-σ inhibitory function (24, 25). The best studied case is the ECF44 CorE from *Myxococcus xanthus* that harbours a CT cysteine rich domain (CRD), which promotes a compact conformation that is favourable for DNA binding (26, 27). ECF41 (σ^{ECF41}) from *Bacillus licheniformis* is rendered inactive by the CT NTF2-like domain (IPR032710), which contains a distal consensus (NPDKL) that links the SnoaL_2 domain to a signature motif in the σ2-σ4 linker (YVGPWLPEP). This promotes a compact conformation of the protein, but reducing the protein to its ECF core is not sufficient for effective transcription (24). *Mycobacterium tuberculosis* sigma J (σ^{J}_{Mtub}) is another ECF in family ECF41 with a long CT extension that encodes the SnoaL-like domain (IPR03740, PF12680: SnoaL_2). The SnoaL_2 domain is an extension with approximately 100 residues that is a member of the NTF2-like superfamily. It shares identity with other conserved domains in this family such as SnoaL (PF07366) domain, the NTF2 domain (PF02136), or the PHZA_PHZB domain (PF03284). Previously described SnoaL_2 domains, are important players in the biosynthesis of specialised compounds, like the antibiotic nogalamycin in *Streptomyces nogalacter*, acting as polyketide cyclases to catalyses an intermediate hydroxylation step of nogalamycin biosynthesis (28). Both Direct Coupling Analyses (DCA) and the determination of the σ^{J}_{Mtub} crystal structure have shown that the SnoaL_2 is a flexible region of the protein that promotes a compact conformation (25, 29). Although SnoaL_2 residues contact the ECF core (25) and partial truncation of the SnoaL_2 domain lead to hyperactive sigma activity (24), the full deletion of the CT domain abolished

sigma factor activity and it was suggested that the compact state does not serve as a σ^{J}_{Mtub} antagonist (29). A similar observation has been made for *B. licheniformis* ECF42 where the tetratricopeptide repeat (TPR) CT extension is necessary for its activity (30). Even though the information collected suggests that, where present, SnoaL-like domains are important in regulating ECF-sigma activity, the physiological role of this extension remains unknown.

In *S. tsukubaensis,* stress-responsive transcription is coordinated by 42 ECFs, showing a level of investment in signalling that may reflect novel mechanisms of ECF-dependent transduction. Of the total ECFs, 8 harbour CT extensions in addition to the ECF core σ_2 and σ_4 conserved domains. In this work, we provide the first functional characterisation of an ECF that contains a SnoaL_2 CT extension, SigG1. Although SigG1 might regulate itself through the SnoaL_2 domain, we identify RsfG as its cognate anti-sigma factor. Finally, we identify genes under SigG1 control and hypothesize how SigG1 orchestrates the timely progression of morphogenesis and secondary metabolism.

2.2 RESULTS

***sigG1* encodes an ECF sigma factor with a SnoaL_2 extension**. A gene encoding an ECF, *STSU_11560*, was identified in *Streptomyces tsukubaensis,* upstream of the gene that codes for OxyR - the major bacterial regulator of H_2O_2-induced stress response (31). STSU_11560 harbours the amino terminal $\sigma^{70}\_r_2$ (PF04542, σ_2) and $\sigma^{70}\_r_4$ (PF04545, σ_4) conserved domains, and a C-terminal (CT) extension, encoding a SnoaL_2 domain (PF12680), which makes it distinctive from other typical ECF σ factors (Figure 1A). Domains σ_2 and σ_4 are separated by an unusually long (53 aa) spacer. The presence of the CT SnoaL_2 domain in ECF sigma factors is distributed among different phyla (18, 19). They are predominantly found in the Actinobacteria and Firmicutes and, in particular, they are included within groups ECF41, ECF56, ECF205, ECF294 and ECF295 of the general ECF group classification (Figure S1) (18). Besides STSU_11560, we identified three additional ECFs with a SnoaL_2 extension in the *S. tsukubaensis* genome: STSU_14518, and STSU_17474*,* that belong to the ECF41 group; and STSU_12530, a closer homologue to STSU_11560. These last two ECFs are enclosed within group ECF56, which is exclusively composed of ECFs with SnoaL_2 domains at the CT extension. STSU_11560 belongs to subgroup3 of the ECF56 family (Figure S1).

From the collection of sigma factors annotated in actinobacterial genomes , the closest homologous protein to STSU_11560 found in *Streptomyces* was SigG from *Streptomyces clavuligerus* ATCC 27064 (67% identity, E value: $3e^{-137}$). The pool of SnoaL_2-containing ECFs in *S. tsukubaensis* resembles the organisation of these ECFs in the *Mycobacterium* genus. Sequence homology revealed that the full-length proteins encoded within loci

STSU_11560 and *STSU_12530* are highly similar to SigG from *Mycobacterium tuberculosis* H37Rv (40% identity, and 47% identity, respectively), while STSU_14518 and STSU_17474 showed identity to SigJ from *M. tuberculosis* H37Rv (43% identity and 38% identity, respectively). Following these analyses, we named the gene *STSU_11560* as *sigG1* and *STSU_12530* as *sigG2*. This work focuses on the physiological characterisation of *sigG1.*

Direct Coupling Analyses (DCA) predicts intramolecular contacts between SigG1 ECF core domains and SnoaL_2. ECFs with additional domains have been proposed to promote intramolecular contacts that regulate ECF activity (26, 30). Making use of the statistical method called direct-coupling analysis (DCA) we assessed whether the Snoal_2 domain is folded towards the ECF domains σ_2 and σ_4. DCA is based on the covariation observed between pairs of residues that interact in large families of homologous proteins. When two residues interact, mutations in one of them need to be compensated by mutations in the second in order to preserve the contact. DCA is able to find pairs of amino acids that covariate due to their direct interaction (32). DCA has been successfully used for the prediction of contacts between the core ECF and the SnoaL-like CT of ECF41 and the TPR-containing CT of ECF42, which later were proven to have an important regulatory role on ECF activity (25). Here we found that DCA predicted a SigG1 tertiary structure that suggests contacts between the first half of SnoaL_2 CT and the linker that separates σ_2 and σ_4 (Figure 1B, areas B and C). Additionally, DCA revealed a high probability of contacts between the N-terminal part of SnoaL_2 and the σ_4 region (Figure 1B, area A). These results suggest a potential self-regulatory function for SnoaL_2 (Figure 1C).

SigG1 interacts with a putative anti-sigma factor, RsfG. Downstream from *sigG1* we identified an additional uncharacterized gene, *STSU_11555*. Many known ECFs are bound by their respective cognate anti-sigma factor (ASF), an interaction that renders the ECF inactive until it is released to direct transcription from stress-responsive promoters. The anti-sigma is usually encoded by a gene adjacent to the *ecf* and can be transcribed independently or co-transcribed in a single mRNA like for example, the *bldN-rsbN* pair in *S. venezuelae* (33) or the *sigI-rsgI* pair in *Bacillus subtillis* (34). To investigate whether this is true for *sigG1* and *STSU_11555*, we analysed the transcription of both genes. Transcription analyses by RT-PCR revealed a longer transcript that included both sequences, indicating that *sigG1* and *STSU_11555* are co-transcribed (Figure 2A). The transcriptional start site (TSS) of *sigG1* was mapped by 5' RACE 118 nucleotides upstream of the annotated start codon. This result was confirmed by analysis of existing RNA sequencing (RNA-seq) data (35). In addition, we identified a TSS for *STSU_11555*, 74 nucleotides upstream of its coding region, indicating that it can also be transcribed from its own dedicated promoter, independently of *sigG1.*

STSU_11555 encodes an hypothetical protein with no conserved Pfam domains. To assess if SigG1 could be regulated posttranslationally by interacting with STSU_11555 we performed bacterial adenylate cyclase two-hybrid experiments (BACTH) and found that the two proteins physically interact strongly in *E. coli* (Figure 2B). These results support the hypothesis that STSU_11555 could act as the SigG1 cognate anti-sigma factor. We therefore named this gene *rsfG* (regulator of sigma factor G1). Additionally, BACTH results show dimerization of RsfG that might function in homomultimeric complexes. Truncated versions of SigG1 and RsfG - where the σ_2-σ_4 region and the SnoaL2 domain were individually tested against the full length RsfG; and the RsfG N-terminal part was tested against the full length SigG1 - were not sufficient to promote interaction between the two proteins in the BACTH experiments, possibly because specific residues required for complex assembly were missing, or due to the loss of stable conformations of the proteins. The formation of a SigG1-RsfG complex was confirmed via co-expression of the two coding sequences in *E. coli* using the pRSFDuet™-1 system. The coding sequences for SigG1 and RsfG were cloned into the IPTG-inducible vector to co-produce RsfG with an amino terminal (NT) histidine tag (6His-RsfG) and an untagged version of SigG1 in *E. coli*. By immobilizing the 6His-RsfG on a nickel agarose affinity resin (Ni-NTA) it was possible to pull-down the untagged sigma factor, indicating that RsfG and SigG1 form a complex (Figure 2C). Elution of the purified proteins revealed two isoforms of 6His-RsfG, as identified by Peptide Mass Fingerprinting (PMF) and Western blot: one protein with the predicted 17 kDa and a second isoform with approximately 19 kDa. Size exclusion chromatography displayed a single peak corresponding to the SigG1/RsfG complex (data not shown).

SigG1 binds to target promoters *in vivo*. To identify the set of genes under the direct control of SigG1 we generated a polyclonal antibody against SigG1 to be used in Chromatin Immunoprecipitation and Sequencing (ChIP-seq). Prior to ChIP-seq, we first replaced the *sigG1* and *rsfG* coding regions with an apramycin resistance cassette and generated single deletion mutants for these genes (Δ*sigG1* and Δ*rsfG*, respectively). The ChIP-seq experiments were performed in the *rsfG* null mutant to ensure SigG1 was available (i.e not bound to the anti-sigma) to allow detection when bound to its target promoters. Analysis of *sigG1* transcript levels by RT-qPCR confirmed expression of *sigG1* in the *rsfG* mutant, and SigG1 protein was similarly detected by automated western blot (WES Simple) (Figure S2) confirming that the polyclonal antibody was suitable for use in ChIP-seq experiments. The *sigG1* null mutant was used as a negative control. DNA purified following immunoprecipitation using α-SigG1 was sequenced. The genomic regions with a fold-change enrichment of at least

1.5 (Δ*rsfG* vs Δ*sigG1*) were considered significant for peak calling. Using this approach we identitifed six peaks dispersed throughout the genome (Figure 3A; Table 1).

Consistent with its function as a sigma factor, with the exception of peak 3, the remaining peaks are located in promoter regions (Figure 3B). Strikingly, no clear peak was identified near the *sigG1* promoter suggesting that this sigma factor might not be subject to auto-regulation via a positive feedback loop, a common feature of ECFs. SigG1 binds at the promoters of two genes related with morphological differentiation in *Streptomyces* - alanine dehydrogenase, *ald1* (36) and the WhiB-like regulator C, *wblC* (37). We also identified three additional SigG1 target genes involved in metal ions-dependent homeostasis, *STSU_23024*, *STSU_32197* (Figure S3B) and aconitate hydratase, *acn*. Furthermore, peak 3 is located in the intergenic region between *STSU_22045* and *STSU_22050* with no apparent promoter region (Figure 3B). However, a careful inspection of RNA-seq data suggests that SigG1 binds to the upstream region of a putative antisense RNA of unknown function at the *STSU_22050* locus (Figure S3B).

To investigate which genes were dependent on *sigG1* regulation, we assessed genome-wide transcription by total RNA sequencing (RNA-seq). We compared transcript levels in the wild-type and in Δ*sigG1* by analysing gene expression in MGm-2.5 liquid cultures at mid-exponential phase of growth. A combination of the data obtained from the RNA-seq and the ChIP-seq experiments indicates that at least four target genes are members of the SigG1 regulon (Table 1). Surprisingly, inspection of both ChIP-seq and RNA-seq NGS data unveiled an unexpected genomic rearrangement in the Δ*sigG1* mutant. Genomic instability resulting in excision of large fragments of DNA has previously been described in *Streptomyces* and other bacteria (38-42). Through NGS re-sequencing of the full genome of this strain, we confirmed a 228 Kb deletion in the terminal region of the chromosome from position 349601 to 577663 bp (Figure 3A).

Considering the central core of the ChIP-seq peaks, and the identification of the TSSs for each gene (35), we were able to isolate and align promoter sequences harbouring the six identified peaks. By inputting the sequences for the *sigG1*-dependent genes into the MEME Suite tool (43), we identified a putative consensus sequence for SigG1 DNA-binding, defined by the conserved residues CCG in the -35 element and the residues CCCTCC in the -10 element (Figure 3C), separated by 17bp non-conserved residues.

***sigG1* is important for progression of morphological differentiation.** To validate the *sigG1*-dependent recruitment of RNAP to the target promoters we have functionally characterised the deletion mutants for the newly identified ECF/ASF pair. One of the most significant peaks obtained with ChIP-seq was found in the promoter region of *ald1* that codes for an alanine dehydrogenase enzyme, Ald. Ald converts L-alanine residues, generated by

protein turnover, into extra pools of pyruvate that help fuel metabolism for growth progression in *Streptomyces tsukubaensis* (36). On ISP4 media, the *sigG1* null mutant exhibited impaired morphological development of sporogenic hyphae. After 6 days of growth on ISP4 agar, the wild-type strain had sporulated, whereas Δ*sigG1* lacked the spore-associated pigmented appearance of the wild-type (Figure 4A). This phenotype was fully restored on agar plates by the introduction of a single copy of the *sigG1* wild-type allele under the control of its native promoter expressed from the ϕC31 integration site. The delay in differentiation was further studied by scanning electron microscopy (SEM). SEM revealed that despite their initial white appearance, colonies of the *sigG1* mutant eventually fully differentiate leading to long chains of spores (Figure 4B). We conclude that SigG1 is required for the timely differentiation of *S. tsukubaensis*. Strikingly, Δ*rsfG* grown on solid media exhibited a darker pigmentation that is usually associated with spores, whereas a double mutant for both *sigG1* and *rsfG* exhibited a partial reversal of the phenotype, as it resembled the wild-type morphology. Although some influence of the genomic gap in Δ*sigG1* strain would be expected, the phenotypic rescue observed in the complemented strain (that also contains the genomic deletion) shows that the morphological phenotype was mostly due to the absence of SigG1.

Using the RNA-seq approach we further evaluated transcription in the Δ*rsfG* mutant and compared it to transcript levels in the wild-type. The identification of 18 significantly deregulated genes (Table 2; FDR pvalue < 0.05) revealed that *rsfG* plays a targeted role in the cell that is confined to the regulation of small operons. These encode for instance different types of proteases or, more interestingly, an operon encoding a putative ECF/ASF system of unknown function. Moreover, Pre-ranked Gene Set Enrichment Analysis ((GSEA, (44)) of Δ*rsfG* mRNA-seq profile showed an activation of genes involved in the TCA cycle (Figure S4B). Gene expression results were validated by RT-qPCR on a selection of candidate genes and the pattern of expression correlated with the one observed in the RNA-seq data. The level of transcripts for *hrdB* an *rpsP* reference genes, encoding the primary essential sigma factor and the 30S ribosomal protein S16, HrdB and RspP, respectively, was monitored as an internal normalization control.

***sigG1* is required for the maintenance of metal-ions homeostasis.** The SigG1 regulon includes two genes whose products are involved in the maintenace of metal ions availabilty. *STSU_23024* shares homology to genes that have been linked to metal ions-dependent drug efflux systems (45). It encodes a Major Facilitator Superfamily (MFS) transporter with 76% identity to the multidrug resistance protein Bmr3_2 from *Streptomyces sp. AVP053U2. STSU_32197*, is a newly identified gene that had not been annotated before (Figure S3B). It codes for an hypothetical protein and is part of an operon encoding a copper

transport system. Hence, and given the high identity shared by the product of *STSU_23024* with metal ions-dependent transporter systems, we further asked if metal ions homeostasis was compromised in the *sigG1* mutants. Using antiSMASH 4.0 (46) we identified at least three siderophore biosynthetic clusters in *S. tsukubaensis* predicted to encode Fe-enterobactin (STSU_33190 - STSU_33135 FecCD), Fe-desferrioxamine B (STSU_23636- STSU_23676) and a second Fe-hydroxamate (STSU_16507 - STSU_16442) siderophores. We assessed the ability of *S. tsukubaensis* to release siderophores, high affinity chelators that sequester ferric iron at very low concentrations to assist in metal ions internalisation from the substrate (47). These small molecules are therefore secreted in iron-limiting conditions. Culture supernatants from iron-limited conditions exhibited substantially higher levels of siderophores when *sigG1* was absent (as shown by the CAS assay in Figure 5A) and, more strikingly, Δ*sigG1* produced siderophores in iron-replete cultures, revealing that these bacteria were under iron starvation. These results were corroborated by a 30% decrease in the intracellular iron content, as measured by flame atomic spectrometry (Figure 5B). Noticeably, the intracellular iron content was restored upon complementation of the mutant strain (Figure 5B), corroborating the importance of *sigG1* in the maintenace of iron levels.

The emergence of a copper transporter system as a putative SigG1 target prompted us to assess *sigG1*-dependent copper tolerance. Serial dilutions of spores of the wild-type, Δ*sigG1*, Δ*rsfG*, Δ*sigG1-rsfG* and the Δ*sigG1* strain complemented with *sigG1*, were inoculated on ISP4 copper-replete media (ISP4 supplemented with 100μM $CuSO_4$). After 6 days of growth, all *sigG1*/*rsfG*-related mutant strains exhibited an impaired growth when compared to the wild-type (Figure 5C). Strikingly, Δ*sigG1* spores were more sensitive towards copper than any of the other strains, in which spores fail to grow only in the highest dilutions tested, due to the toxicity of detrimetal concentrations of $CuSO_4$. The dramatic arrest of development of aerial mycelium observed in Δ*sigG1* was reverted by expressing the *sigG1* wild-type allele in this mutant.

The Streptomyces* oxidative stress response is dependent on *sigG1. The disruption of metal ions homeostasis is frequently associated with concurrent impairment of the cell redox state. A major cause of toxicity is the cross-reaction of intracellular iron with H_2O_2 - through the Fenton reaction (48) - that results in harmful levels of oxidative stress. To examine the impact of the fluctuations in iron availability upon *sigG1* deletion we assessed the maintenance of the redox balance by addressing the response to H_2O_2. The OxyR regulon is at the core of the bacterial defence against H_2O_2 stress. The location of *sigG1* relative to the *oxyR-ahpCD* operon in *S. tsukubaensis* supports the possibility that SigG1 has a role in the oxidative stress response.

In order to investigate the role of SigG1 in the oxidative stress response, we exposed the cells to exogenous H_2O_2 to trigger the activation of OxyR. After 15 minutes of H_2O_2 exposure, we examined *sigG1* transcript levels by RT-qPCR. Upon disruption of the major regulator of H_2O_2, *oxyR*, we observed an upregulation of *sigG1* transcripts in standard conditions, indicating that *sigG1* is expressed during the response to the oxidative stress damage (Figure 6). Although transcription in the wild-type was not affected by the treatment, *sigG1* levels were strongly induced by exogenous H_2O_2 in the absence of OxyR (Figure 6), indicating that *sigG1* is activated in response to H_2O_2-mediated oxidative stress. The transcript levels of *rsfG* were unchanged upon the ROS insult in these strains.

Deletion of *sigG1* impairs secondary metabolism in *S. tsukubaensis.* Oxidative stress responses are tightly linked to secondary metabolism in *Streptomyces* (49). We have previously observed that FK506 production in *S. tsukubaensis* is dependent on the intracellular redox balance (Pires *et al*., unpublished data). Upon *sigG1* deletion, we detected a delayed onset of FK506 biosynthesis, which resulted in a significant reduction (>30%) in the total production of this immunosuppressant when compared to the wild-type (Figure S5).

2.3 DISCUSSION

The occurrence of several copies of SnoaL_2-containing ECFs in some *Streptomyces* species might reflect their evolutionary significance. Although we could not define a clear phylogenetic relationship between *Streptomyces* genomes containing more than one copy of SnoaL_2-containing ECF encoding gene, we speculate that they could emerge in a proportional fashion to the complexity of the metabolism in each strain. Recently, ECFs with additional domains have been predicted to rely on the autoregulatory activity of the ECF conserved core through interaction with domains at the C-termini. SigG1 lacks the CT part of the σ^{ECF41} NTF2-like extension containing the conserved NPDKL motif predicted to interact with the YVGPWLPEP motif in the linker between σ_2 and σ_4 (Figure S6) (24, 25). DCA analysis further suggested that despite the conservation of the SnoaL_2 domain, SigG1 conformation is established by pairs of amino acid residues that are distinct from the ones that are in contact in σ^{ECF41}. Typically, genes encoding ECF41 sigma factors are located nearby genes that code for carboxymuconolactone decarboxylases, oxidoreductases, or epimerases (COE) but no conserved genomic context was ever assigned to ECF56 (19). Nearby *sigG1*, we found an adjacent gene encoding the RsfG protein, which was co-transcribed with *sigG1*. We observed that RsfG interacts with SigG1 acting as the cognate sigma antagonist to control SigG1 activity until the oxidative trigger produced during differentiation emerges. Hence, we propose that intramolecular regulation in SigG1 might not be sufficient to fully inhibit transcription, as it requires an anti-sigma factor to prevent its runaway activity. SigG1 binds to the promoters of

genes involved in morphological differentiation and metal ions homeostasis. When we analysed the phenotype of the double *sigG1*/*rsfG* mutant, it resembled the wild-type morphology throughout development. With this observation, one can speculate that upon disruption of the *sigG1*/*rsfG* system, there is a mechanism of compensation, similarly to what is often opted for by bacteria upon disruption of two component systems (11, 50-53). For instance, in the so called *many-to-one* two-component systems, the sensor kinases interact with different effectors to regulate the signalling pathway at different levels and alternate among different outputs (11).

In previous studies, we demonstrated that differentiation into the air in *Streptomyces* is tightly controlled by oxidative stress (54). We illustrated that the transition from first mycelium stage (MI) to secondary metabolite producing differentiated mycelium (MII) is preceded by phenomena of cell death of the vegetative mycelia, which reroutes nutrients to differentiate MII hyphae into nascent aerial mycelium. We further showed that cell death at the colony core resulted from an increase of the H_2O_2 intracellular pool. In *S. tsukubaensis*, the LysR-type regulator OxyR perceives the H_2O_2 signal and drives transcription of alkylhydroperoxidases to detoxify the excess of these reactive oxygen species (ROS) (Pires et al, unpublished data). In parallel, the OxyR regulon includes other players that maintain metal ions homeostasis. In this study, we provide evidence that the antioxidant response in *Streptomyces* harbours novel players that compose an ECF sigma/anti-sigma system. In the 147 *Streptomyces* genomes within ECF56 family in our database, we found a gene encoding an OxyR-like LysR-type transcriptional regulator in the proximity of a SigG1 homologue-encoding gene. We hypothesize that SigG1 is likely to be an ECF that responds to H_2O_2 intracellular signals, generated by developmental transitions, to activate the cascades of genes that regulate the metabolic shift during growth. *M. tuberculosis* SigJ has also been implicated in the response to oxidative stress. Despite the susceptibility to H_2O_2 exhibited by a Δ*sigJ* mutant, *sigJ* mRNA levels are not responsive to H_2O_2-induced stress (55).

The control of intracellular H_2O_2 homeostasis is particularly critical. It accompanies the homeostasis of metal ions content as the cells face emergent toxicity due to the Fenton reaction (48), by which intracellular iron reacts with H_2O_2 to generate free radicals that damage DNA, proteins or lipid molecules. SigG1 deletion indicates that SigG1 participates in the control of fluctuations in iron availability. The observation that *rsfG*-depleted cells did not show alterations in siderophore production suggests that it is unlikely that SigG1 directly regulates iron flux, and reinforces the indirect effect of SigG1-RsfG system in metal ions homeostasis. We propose that the effect is rather on other iron binding regulators and iron-containing enzymes that promote the intracellular electron transfer during oxidative respiration. The SigG1 regulon includes genes encoding proteins with iron-sulphur centres that are known regulators of metal ions homeostasis, including AcnA and WblC. In *Streptomyces*, WblC is

involved in developmental cascades and, moreover, it regulates the expression of the gene encoding SigR, a well-studied sigma factor responsible for coordinating the response to oxidative stress induced by thiol agents (56). Furthermore, SigG1 regulates the expression of two genes encoding putative metal ions-dependent efflux systems described in other bacteria. Together with a disrupted iron balance, we show that strains defective in *sigG1* were significantly more sensitive to exogenous copper than the wild type. The mechanisms controlling copper uptake or secretion in *Streptomyces* were recently identified (57-60). The function of the product of *STSU_32197* remains unknown since no clear homologues were identified for this protein. However, one possibility is SigG1 mediates copper tolerance by activating an operon for the coding sequences for the copper chaperone/P-type ATPase system CopA/CopZ responsible for the secretion of these metal ions in *Streptomyces* (*STSU_32197-STSU_32205*). Overall, we show that SigG1 is involved in maintaining iron and copper homeostasis, probably in an effort to counteract the metal ions imbalance caused by oxidative damage.

The SigG1-specific motif displays the 17bp optimal spacer size between the -35 and -10 promoter elements defined for ECFs. In addition, the residues in the -10 box resembled the -10 conserved motif proposed for ECFs in the ECF56 group families (19). However, neither the SigG1 consensus, nor -10 and -35 boxes predicted for promoters recognized by σ factors of the ECF56 group are present upstream of *sigG1*. Instead, *sigG1* promoter harbours a conserved motif described for the sigma-factor BldN, a developmental determinant regulated by iron availability at the interface of vegetative to aerial growth (61, 62). The potential BldN-dependent regulation of *sigG1* could be one explanation for the iron limitation and the impaired aerial growth observed for *S. tsukubaensis* Δ*sigG1.* As described for operons with an intergenic region between the sigma-factor and the cognate anti-sigma (61), it seems likely that the *sigG1* promoter drives the expression of a readthrough transcript but, additional *rsfG* transcripts arise from the activity of a *rsfG* dedicated promoter.

These findings provide an explanation as to how *Streptomyces* protects itself from oxidative stress generated endogenously at the colony core. We hypothesize that as the rate of primary metabolism slows down upon completion of the earlier processes of development, nutrient depletion provides the oxidative trigger required to activate SigG1. SigG1 responds by activating the expression of Ald1, which catalyses the alanine from protein turnover, leading to increased pyruvate production during sporulation. AcnA will then promote pyruvate conversion into energy through the TCA cycle. This extra energy, together with a balance in intracellular metal ions promoted by other SigG1 targets, will support the metabolic processes responsible for the developmental shift to aerial differentiation. The presence of a SigG1-dependent promoter upstream of the aconitate hydratase encoding gene and the upregulation

of the genes that composed the TCA cycle molecular signature, observed when deleted *rsfG,* support this idea (Figure S4B). Moreover, we observed the enhanced transcription of genes encoding other proteins that are directly involved in metal ions homeostasis (*sodA*, *bfr*) in the Δ*rsfG* mutant ($p > 0.05$).

Transitions between mycelia differentiation stages are frequently associated with the onset of secondary metabolism. Another important finding in our study using the industrial antibiotic producing conditions was that disrupting the sigma factor significantly decreased the production of FK506 (Figure S5). This further reinforces SigG1 as an important regulator for the timely progression of growth in *S. tsukubaensis* and, by association, for the production of FK506.

We have here presented a model for the action of a novel ECF-anti-sigma pair in the hierarchy of signalling cascades that leads to differentiation in *Streptomyces*. To the best of our knowledge, this is the first work describing the physiological role of an ECF with a SnoaL_2 extension in bacteria. Interestingly, catalytic residues that enable the hydroxylase function of SnoaL-like hydroxylases are absent in SigG1 SnoaL_2 sequence. Since this might indicate that this domain does not act as a functional enzyme, further work on the role of SnoaL_2 in SigG1 would elucidate if this could act as a sensor, as proposed for other domains with catalytic residues that are most likely inactive (63, 64). The mechanism of SigG1 activation through release from its antagonist is still unclear. The presence of two species of RsfG in the purified complex lead us think that this is dependent on a posttranslational modification of the anti-sigma factor as described, for instance, for *M. tuberculosis* oxidative stress-responsive SigH, which is released upon phosphorylation of its cognate anti-sigma RshA (65). Future determination of the tertiary structures of SigG1 and RsfG will provide more information on intra or intermolecular interactions and shed more light on the mechanisms that adjust the SigG1 activity.

Overall, the insights gained in this work provide a link between morphological development, metal ions homeostasis and oxidative stress in *Streptomyces tsukubaensis,* in which, via the activation of a unique SnoaL_2-containing ECF, *S. tsukubaensis* can respond to stress to thrive in the complex environment of the soil.

2.4 MATERIALS AND METHODS

Bacterial strains and growth conditions. Bacterial strains used in this study are described in Table S1. *Escherichia coli* strains were grown under aeration in LB (Luria-Bertani) liquid medium or on LB agar at 30 °C or 37°C. *Streptomyces tsukubaensis* NRRL 18488 strains were grown at 30°C on ISP4 solid medium (Difco™) for spores production. Liquid cultures were grown in MGm-2.5 medium at 28˚C, under aeration at 220 rpm (6). Ampicillin at 100 $\mu g.mL^{-1}$, kanamycin 50 $\mu g.mL^{-1}$, hygromycin 25 $\mu g.mL^{-1}$, apramycin 50 $\mu g.mL^{-1}$,

chloramphenicol at 25 μg.mL^{-1} or thiostreptone at 40 μg.mL^{-1} where added to the media when required. Phenotypic analysis of *S. tsukubaensis* and derived strains was conducted on ISP4 solid media or in MGm-2.5 liquid cultures. Samples were harvested throughout the growth curve according to the following key: mid-exponential phase, 72h; late exponential phase and FK506 production onset, 96h; early stationary phase, 120h; late stationary phase, 144-192h; FK506 production end-point, 192h.

Construction of deletion mutant strains. In-frame deletions Δ*sigG1*::apr, Δ*rsfG*::apr and Δ*sigG1-rsfG*::apr were generated using REDIRECT™ PCR targeting technology (66). *E. coli* BW25113 containing a λ RED plasmid, pIJ790, was used to replace the native alleles on the 15C1 cosmid with the apr-oriT cassette amplified from the pIJ773 template plasmid, using specific oligonucleotides red_sigG1_F/Rand red_rsfG_F/R listed in Table S1, to create mutations in the *sigG1* and *rsfG* loci. *E. coli* ET12567 containing pUZ8002 was used for λ RED-mediated recombinering upon conjugation experiments. Conjugations between *E. coli* and *S. tsukubaensis* were carried out as described in (67). In-frame deletions were confirmed by PCR analysis (with primers red_conf_sigG1_F/R and red_conf_rsfG_F/R) and by Southern hybridization. For complementation of the *sigG1* deletion mutant strain with the native allele, *sigG1* and its promoter region, were amplified using oligonucleotides SigG1_SpeI_F and SigG1_SpeI_R cut with *SpeI* and cloned into the *XbaI* site of pIJ12333. The construct was introduced into *S. tsukubaensis* Δ*sigG1* by conjugation.

Scanning electron microscopy. Samples for scanning electron microscopy were obtained by growing strains on ISP4 agar for 15 days. Isolated colonies were mounted on an aluminium stub using Tissue TekR (BDH Laboratory Supplies, Poole, England). The stub was then immediately plunged into liquid nitrogen slush at approximately -210˚C to cryo-preserve the material. The frozen sample was transferred, in vacuo, onto the cryostage of an ALTO 2500 cryo-transfer system (Gatan, Oxford, England) attached to an FEI Nova NanoSEM 450 (FEI, Eindhoven, The Netherlands). Sublimation of surface frost was performed at -95˚C for 3½ minutes before sputter coating the sample with platinum for 2½ minutes at 10mA, at colder than -110˚C. After sputter-coating, the sample was moved onto the cryo-stage in the main chamber of the microscope, held at -125˚C, and imaged at 2.6 to 3kV.

***Streptomyces* genomic DNA isolation.** Genomic DNA from *Streptomyces* was obtained with the Master™ pure Gram-positive DNA purification kit (Epicentre) or with the optimized procedure for Gram-positive bacteria from the GeneJET Genomic DNA purification kit (Fermentas) according to the protocol provided by the manufacturers. *S. tsukubaensis* genome pair-end resequencing was performed in an Illumina (NovaSeq 6000) platform using a 350 bp PCR-free library (Novogene, Hong Kong).

Polymerase chain reaction (PCR) and oligonucleotides used. DNA fragments used in this study were obtained by PCR using the G2 GoTaq® (Promega) or Q5 High fidelity polymerase (New England Biolabs). All oligonucleotides used are listed in supplementary Table S2.

RNA isolation and RT-qPCR analyses. Samples for gene expression studies were harvested at 72h, 96h or 120h of growth. For H_2O_2 induced stress experiments, samples were collected at 72 h (t_0) and 15 min after the addition of 5 mM H2O2 (t_1). Culture aliquots were mixed with two volumes of RNA Protect Bacteria Reagent (Qiagen) and maintained for 5 min at room temperature. Cells were harvested by centrifugation and immediately frozen by immersion in liquid nitrogen. The total RNA was isolated using the RNeasy® Mini kit (Quiagen) according to manufacturer instructions with modifications described in (Beites et al., 2011). Total RNA concentration was determined with a NanoDrop ND-1000 spectrophotometer (Thermo Scientific), and RNA quality and integrity were evaluated in an Experion™ Automated Electrophoresis System (Bio-Rad). For cDNA synthesis, 1µg of DNase I-treated total RNA was transcribed with the iScript™ Select cDNA synthesis Super mix kit (Bio-Rad). RT-qPCR amplifications were performed in an iCycler iQ5 Real-Time PCR detection system (Bio-Rad) using 0.2 µM of each primer of the pairs listed in Table S. and using, 10 µL of KAPA SYBR FAST® RT-qPCR Master Mix (KAPA Biosystems) and 2 µL of template cDNA. Standard serial dilutions of the cDNA were used to check the relative efficiency and quality of each primer pair. Non-template controls were included. A melting curve analysis was performed at the end of each RT-qPCR to exclude the formation of nonspecific products. Analysis included three biological replicates and technical triplicates for each cDNA. The data obtained was analysed using the method described by M. W. Pfaffl (68). For each analysis *rpsP* and *hrdB* mRNAs were used for normalization. The identity of each amplified product was confirmed by sequencing.

Mapping of the 5' terminus of mRNA by Rapid Amplification of cDNA Ends (5' RACE). Transcriptional Start Site (TSS) identification was performed using the Invitrogen kit 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0, following the manufacturer's instructions. First strand cDNA synthesis was carried out using the gene-specific primer GSP1 (Table S1). PCR amplification of tailed cDNA was carried out using the 5'-RACE abridged anchor primer (AAP) with the GSP2 nested primer (Table S1). Specificity of the PCR products was confirmed by re-amplification using the AUAP primer and a GSP3 nested primer, and by sequencing.

RNA sequencing. For the genome-wide transcriptomics experiments, mycelia samples were harvested from liquid cultures at 72h, and total RNA was isolated as described above. Quality control of the total RNA was assessed through the RNA integrity number (RIN). The library construction of cDNA molecules from total RNA samples was carried out using

TruSeq Stranded Total RNA with Ribozero Library Preparation Kit. The generated DNA fragments (DNA library) were sequenced in the Illumina Hiseq 4000 platform, using 150bp paired-end sequencing reads (Stab-Vida, Portugal).

Preparation of cell-free protein extracts. Cells were suspended in lysis buffer containing 50mM potassium phosphate buffer (pH 6.8) supplemented with Complete EDTA-free protease inhibitor cocktail (Roche Applied Science) and lysed on ice using a sonifier. Protein concentration was determined using the BCA Protein Assay Reagent (Thermo Scientific). Bovine serum albumin was used to determine standard curves.

Secreted siderophores detection. Siderophore production was addressed using the Chrome Azurol S (CAS) assay according (69). Cultures were grown in MGm-2.5 media with or without ferrous iron supplementation. Samples were harvested at 96h and supernatants placed on CAS agar plates at 30°C.

Determination of total iron content. Extracellular iron levels were determined in samples harvested throughout growth in MGm-2.5. Cells were centrifuged and the supernatant was recovered to determine total extracellular iron content, using the QuantiChromTM Iron Assay Kit (DIFE-250) according to the manufacturer's instructions. Intracellular iron levels were measured by flame atomic absorption spectrometry (F-AAS). Samples were collected from liquid cultures by centrifugation and cell pellets were washed for three times with TE buffer (20mM Tris-HCl, 5mM EDTA, pH 7.7), followed by one wash with metal-free double distilled water to remove salts. Cells were suspended in 65% (v/v) HNO_3, lysed at 75°C and the supernatant was analysed for Fe content by F-AAS at 248.3 nm, using a PU 9200X spectrophotometer (Philips). Fe content was normalized to protein concentration.

FK506 quantification. FK506 production by *S. tsukubaensis* strains was quantified by HPLC. Extraction of FK506 was carried out by mixing 1 mL of culture with 1mL of 100% methanol, for 1h at 30˚C. The mixture was centrifuged and the supernatant was recovered and analysed using a SunFire™ C18 column (4.6 x 150 mm, 3.5 µm; Waters) in a HPLC system (Merck-Hitachi). The UV detector was set at 210 nm and the oven was set at 55 ˚C. Elution was performed with a gradient mobile phase of 0.1% (v/v) trifluoroacetic acid (TFA) and 20% (v/v) Methyl tert-butyl ether (MTBE) in acetonitrile.

Overexpression and purification of recombinant 6xHis-tagged SigG1 by IMAC. The full-length coding sequence of *sigG1* was cloned into the *NdeI* and *XhoI* restriction sites of the pET15b vector (Novagen) and protein overexpression was induced with 0.2mM IPTG, at 16˚C, 180rpm. The N-terminal hexa-histidine tagged SigG1 (6His-SigG1) was isolated using metal ion affinity chromatography (IMAC). Cells were recovered by centrifugation, resuspended in lysis buffer (20 mM sodium phosphate, 0.3M NaCl, 10µg/mL DnaseI, 200µg/mL lysozyme, protease inhibitor cocktail pH 7.4) and lysed using a French press

homogenizer at 8000psi. After centrifugation, the clear extract was loaded onto a 1 mL HisTrap niquel sepharose high performance column (HisTrap, GE healthcare) pre-charged with Ni^{2+} and equilibrated with binding buffer (20 mM sodium phosphate, 300 mM NaCl, 20mM imidazole, pH 7.4). The column was attached to a BioLP fast protein liquid chromatography (FPLC) system (Bio-Rad) and further washed with binding buffer at a flow rate of 1 $mL.min^{-1}$. Elution was carried out to an adequate level of purity with approximately 120mM imidazole, and fractions were collected and analyzed by SDS-PAGE. Peptide mass fingerprinting (PMF) confirmed the identity of the protein. Buffer exchange to 20 mM sodium phosphate, 300 mM NaCl was carried out. The protein was dialysed and concentrated using an Amicon Ultra-15 30Kda (Milipore).

Production of SigG1-specific polyclonal IgY antibodies. The 6His-SigG1 **purified protein** was used to induce immunization of quail for the production of the specific antibody in egg yolks (HenBiotech). Due to high level of conservation of amino acid sequences among ECF sigma factors, and the possibility of cross-reaction with other sigma factors in protein extracts, antibodies against specific epitopes of SigG1 were raised. The V16-A39 epitope ($SigG1^{V16\text{-}A39}$) was selected for quail immunization.

Immunoblot detection of SigG1 and 6His-RsfG. Total protein extracts were run on 10% SDS-PAGE and transferred to a nitrocellulose membrane Hybond™–C Extra (Amerhsam Pharmacia Biotech) using a Mini-PROTEAN Mini Trans-Blot® system (Bio-Rad). The membrane was blocked with 5 % (w/v) dried milk in TPBS/TBST as required, rinsed twice in TPBS/TBST and incubated with appropriate primary antibodies. After incubation, the membrane was rinsed twice in TPBS/TBST, and incubated with the secondary antibody. Signals were revealed with Prime Plus ECL detection kit (Bio-Rad). SigG1 was detected using the polyclonal antibodies raised in quail. For detection of 6His-RsfG we used the tetra-his mouse antibody (Qiagen). Proteins were visualised via labelling with an anti-chicken IgY (IgG) coupled to peroxidase (A9046, Sigma) or an anti-mouse secondary antibody conjugated to HRP (Santa Cruz Biotechnology).

Recombinant expression and purification of SigG1-RsfG complex. For the co-expression studies, the full-length coding sequences of *rsfG* and *sigG1* were cloned into the MCS-I (*EcoRI* and *HindIII* sites) and MCS-II (*NdeI* and *KpnI* sites), respectively, of the pRSFDuet™-1 (Novagen) vector to generate a poly-histidine tag RsfG recombinant protein and a SigG1 untagged recombinant protein. Expression of soluble proteins in *E. coli* Nico 21 cells (New England Biolabs) was achieved after 20h incubation with 1mM IPTG, at room temperature, with aeration. The molecular weights of the recombinant proteins were verified by nanoLC–MS/MS, with an Ultimate 3000 liquid chromatography system coupled to a Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific). SigG1-6xHis-RsfG complex was purified from the soluble cell lysate by Ni-NTA affinity chromatography.

Cells were suspended in lysis buffer (20 mM sodium phosphate, 0.15M NaCl, 10µg/mL DnaseI, 200µg/mL lysozyme, protease inhibitor cocktail pH 7.4) and disrupted by mechanical lysis through a FRENCH Press (Thermo Scientific) at 4000psi. The soluble co-expressed proteins were batch purified in 0.15M NaCl, 20 mM sodium phosphate buffer, pH 7.4 using a Ni-NTA agarose affinity chromatography matrix (Qiagen). Elution was performed in 250 mM imidazole and analyzed by SDS-PAGE. The protein fractions were concentrated in Amicon 10kDa cutoff Amicon filters. Analytical size exclusion chromatography (SEC) was used to probe the molecular weight of the SigG1-6xHis-RsfG complex using a Superose12 10/300 GL analytical grade column (GE Healthcare) connected to an ÄKTA Purifier 10 system (GE Healthcare). The total protein was 0.9 mg/ml was loaded onto the column and the protein was eluted in 20mM sodium phosphate buffer containing 150 mM NaCl, pH 7.4. The experimental molecular weight of the complex was determined by plotting the elution volume against a standard curve with the following standards: ChemotrypsinogenA (25 kDa), BSA (66 kDa) and Aldolase (158 kDa).

Automated western blot (Wes). Analysis of the native SigG1 protein in crude *S. tsukubaensis* extracts was performed using the quantitative Wes capillary electrophoresis and blotting system (ProteinSimple, San Jose, CA) with the Wes No Secondary Detection (12 to 230 kDa) Master kit. Protein samples were prepared in accordance with the manufacturer's directions. Labelling was achieved using the anti-SigG1$^{V16\text{-}A39}$ antibody and a rabbit anti-chicken IgY (IgG) coupled to peroxidase antibody (A9046, Sigma). Results were analysed using the Protein Simple software Compass (version 2.6.7).

ChIP-sequencing. For the Chromatin Immunoprecipitation (ChIP) assays *S. tsukubaensis* strains were grown in MGm-2.5 and samples were harvested and prepared as described in (70) with the following modifications. For immunoprecipitation of SigG1 cross-linked DNA, the total extract was incubated overnight with polyclonal anti-SigG1$^{V16\text{-}A39}$ antibody at 4°C. Immunoprecipitation was carried out using a goat anti-chicken IgY agarose (ab76444, Abcam) for 4 hours. Genomic DNA libraries enriched for SigG1 binding were produced from these samples, size selected to ~100-500bp, and sequenced on an Illumina NovaSeq 6000 platform (Novogene, Hong Kong), using 150 bp paired-end reads. These experiments were performed in two biological replicates.

Bacterial two-hybrid assays. Bacterial adenylate cyclase two-hybrid assays (BACTH, Euromedex) were performed as described in the literature (71). Briefly, the DNA fragments that encode the full-length RsfG and SigG1 proteins were cloned into BACTH T18 and T25 containing vectors obtained from Euromedex. The empty plasmids and the zip plasmids were used as negative control and positive control, respectively. Plasmids were co-transformed into *E. coli* BTH101 and incubated at 30°C for 2–3 days. Transformants were

grown on M63/MacConkey agar at 30°C for 3–4 days. The positive clones were assessed using the β-galactosidase assay (72). Single colonies of the co-transformed bacteria were expanded overnight in LB broth containing the β -galactosidase chromogenic indicator 5-bromo-4-chloro-3-indolyl-b-D-galacto-pyranoside (X-Gal) at 100 mg/ml, 0.5 mM Isopropyl-b-D-1-thiogalactopyranoside (IPTG) and antibiotics for selection. Cells were harvested and resuspended in Z buffer (60 mM Na2HPO4,40 mM NaH2PO4, 10 mM KCl, 1 mM MgSO4,50 mM b-mercaptoethanol pH 7.0). Cultures optical density were determined at OD_{600nm}. Cells were permeabilized by mixing with toluene. Permeabilized cells were mixed with 10 mg/ml ortho-nitrophenyl-β-galactoside (ONPG) and OD_{405nm} was measured over time. β-galactosidase activity units were calculated as previously described by Miller JH (72).

Bioinformatics procedures. Phylogenetic trees. ECF phylogenetic trees were built as described in (18). Direct Coupling Analysis. Protein sequences of members of groups ECF41 and ECF56 were retrieved from the most recent ECF classification (18) and aligned using Clustal Omega 1.2.3 (73). DCA was performed using Gaussian DCA (74) as described in (25). Calculation of differentially expressed genes. After trimming (ambiguous limit= 2nt, quality limit= 0.01), sequence reads generated from RNA-seq were mapped onto the concatenated version of the reference genome (see *ChIP-sequencing* part in this section). The result of mapping served to determine the gene expression levels based on the trancripts per Million (TPM). Expression with log2 fold change ≥ log2(1.5) and q-value ≤ 0.05 or log2fold change ≤ –log2(1.5) and q-value ≤ 0.05 was considered as differentially expressed. Additional analyses of the generated sequence raw data were carried out using CLC Genomics Workbench 11 (CLC Bio, Denmark). Functional enrichment analyses. Regulons were categorized according to their associated gene ontology IDs (GO) retrieved from the information available at Uniprot database. GSEA analysis was performed using the Broad Institute GSEA software (44) using Gene set collections obtained by searching the *S.tsukubaensis* genome for homologues of proteins involved in *Streptomyces* sporulation and TCA through NCBI BLASTp analyses (https://blast.ncbi.nlm.nih.gov). The Δ*sigG1* vs WT full transcript list was ranked according to the log2 fold-change expression values and probed against the indicated gene signatures using the GSEA Pre-ranked mode with the following parameters: 10000 permutations, classic scoring schem and meandiv normalization. ChIP-seq data analysis. The reads in fastq files resulting from paired-end sequencing were aligned to the genome of *Streptomyces tsukubaensis* using the bowtie2 (75) and further treated as described in (70). Enrichment for the Δ*sigG1* control samples was subtracted from the enrichment in Δ*rsfG* samples. Significance of enrichment values were calculated assuming normal distribution of the enrichment values. Results were visualized in the Integrated Genome Browser (76), or in the CLC Genomics Workbench 11 (**CLC Bio, Denmark).** The

reference genome was the concatenation of all contigs available at NCBI (genome assembly ASM2971v2; WGS project AJSZ01).

Statistical analyses. For each experiment, we assayed at least three independent biological replicates. Statistical significance was addressed through the GraphPad 6 software, according to the requirements of each data set.

Data availability. Sequencing data was submitted in GEO under the following accession codes: GSE144815 for RNAseq data and GSE144907 for ChIPseq data.

ACKNOWLEDGEMENTS

This work was partially funded by National Funds through FCT- Fundação para a Ciência e a Tecnologia, I.P., under the project ERA-IB-2/0001/2015. It was further supported by FEDER - Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation (POCI), Portugal 2020; and by Portuguese funds through FCT Fundação para a Ciência e a Tecnologia, I.P/Ministério da Ciência, Tecnologia e Ensino Superior POCI-01-0145-FEDER-007274 and NORTE-01-0145-FEDER-000012. BBSRC supported this work through the Institute Strategic Programme grant BB/J004561/1 to the John Innes Centre. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. R.O. was supported by the FCT fellowship SFRH/BD/107862/2015 and by the EMBO fellowship ASTF438-2015, M.V.M. was supported by the FCT fellowship SFRH/BPD/95683/2013 and the FCT contract DL57/2016/CP1355/CT0023 and D.C.P. and G.F. were supported through the IMPRS-Mic and the ERASynBio project ECFexpress (BMBF grant 031L0010B). The authors are grateful to Kim Findlay at the Bioimaging platform of the John Innes Centre (JIC, UK) for performing the SEM imaging of *Streptomyces* samples, Mervyn Bibb (JIC, UK) for the pIJ12333 plasmid , Mark Buttner (JIC, UK) for his comments and discussion regarding the work and Paula Tamagnini (i3S, PT) for comments on the manuscript. The authors acknowledge the support of the i3S Scientific Platforms *Cell Culture and Genotyping*, *Biochemical and Biophysical Technologies* and *Proteomics*. The authors have no conflict of interest to declare.

2.5 REFERENCES

1. Krentz AJ, Dmitrewski J, Mayer D, McMaster P, Buckels J, Smith JM, Nattrass M. 1994. Tacrolimus (FK506) versus cyclosporin in prevention of liver allograft rejection. Lancet (London, England) 344:948-949.

2. Tang L, Ebara S, Kawasaki S, Wakabayashi S, Nikaido T, Takaoka K. 2002. FK506 ENHANCED OSTEOBLASTIC DIFFERENTIATION IN MESENCHYMAL CELLS. Cell Biology International 26:75-84.

3. Chang C, Li K, Jiang S, Li B, Cao L, Wang P. 2019. Downregulation of TRPC6 expression is a critical molecular event during FK506 treatment for overactive bladder. Cell Calcium 77:8-19.

4. Jiang H, Yamamoto S, Nishikawa K, Kato R. 1993. Anti-tumor-promoting action of FK506, a potent immunosuppressive agent. Carcinogenesis 14:67-71.

5. Nagrani NK, Zito PM. 2019. Topical Tacrolimus (FK506, Protopic) in the Treatment of Atopic Dermatitis. Journal of the Dermatology Nurses' Association 11:41-44.

6. Davis B, Hilgart D, Erickson S, Labroo P, Burton J, Sant H, Shea J, Gale B, Agarwal J. 2019. Local FK506 delivery at the direct nerve repair site improves nerve regeneration. Muscle & Nerve 60:613-620.

7. Claessen D, Stokroos I, Deelstra HJ, Penninga NA, Bormann C, Salas JA, Dijkhuizen L, Wösten HAB. 2004. The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlins and chaplins. Molecular Microbiology 53:433-443.

8. Elliot MA, Karoonuthaisiri N, Huang J, Bibb MJ, Cohen SN, Kao CM, Buttner MJ. 2003. The chaplins: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in Streptomyces coelicolor. Genes & Development 17:1727-1740.

9. Martinez-Castro M, Salehi-Najafabadi Z, Romero F, Perez-Sanchiz R, Fernandez-Chimeno RI, Martin JF, Barreiro C. 2013. Taxonomy and chemically semi-defined media for the analysis of the tacrolimus producer *Streptomyces tsukubaensis*. Appl Microbiol Biotechnol 97:2139-52.

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

11. McCormick JR, Flardh K. 2012. Signals and regulators that govern *Streptomyces* development. FEMS Microbiol Rev 36:206-31.

12. Bush MJ, Tschowri N, Schlimpert S, Flärdh K, Buttner MJ. 2015. c-di-GMP signalling and the regulation of developmental transitions in streptomycetes. Nature Reviews Microbiology 13:749-760.

13. Ulrich LE, Koonin EV, Zhulin IB. 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol 13:52-6.

14. Francis VI, Porter SL. 2019. Multikinase Networks: Two-Component Signaling Networks Integrating Multiple Stimuli. Annual Review of Microbiology 73:199-223.

15. Helmann JD. 2019. Where to Begin? Sigma factors and the selectivity of transcription initiation in Bacteria. Molecular Microbiology doi:10.1111/mmi.14309:mmi.14309-mmi.14309.

16. Lonetto MA, Donohue TJ, Gross CA, Buttner MJ. 2019. Discovery of the Extracytoplasmic Function σ Factors. Molecular Microbiology 112:mmi.14307-mmi.14307.

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

18. Lonetto M, Gribskov M, Gross CA. 1992. The sigma 70 family: sequence conservation and evolutionary relationships. Journal of Bacteriology 174:3843-3849.

19. Lane WJ, Darst SA. 2006. The structural basis for promoter -35 element recognition by the group IV σ factors. PLoS Biology 4:1491-1500.

20. Lin W, Mandal S, Degen D, Cho MS, Feng Y, Das K, Ebright RH. 2019. Structural basis of ECF-σ-factor-dependent transcription initiation. Nature Communications 10:710-710.

21. Casas-Pastor D, Müller RR, Becker A, Buttner M, Gross C, Mascher T, Goesmann A, Fritz G. 2019. Expansion and re-classification of the extracytoplasmic function (ECF) σ factor family. bioRxiv doi:10.1101/2019.12.11.873521:2019.12.11.873521.

22. Huang X, Pinto D, Fritz G, Mascher T. 2015. Environmental Sensing in Actinobacteria: a Comprehensive Survey on the Signaling Capacity of This Phylum. Journal of Bacteriology 197:2517-2535.

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

24. Jogler C, Waldmann J, Huang X, Jogler M, Glöckner FO, Mascher T. 2012. Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in planctomycetes by comparative genomics. Journal of Bacteriology 194:6419-6430.

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

26. Pinto D, Liu Q, Mascher T. 2019. ECF σ factors with regulatory extensions: The one-component systems of the σ universe. Molecular Microbiology doi:10.1111/mmi.14323:mmi.14323-mmi.14323.

27. Wecke T, Halang P, Staron A, Dufour YS, Donohue TJ, Mascher T. 2012. Extracytoplasmic function sigma factors of the widely distributed group ECF41 contain a fused regulatory domain. Microbiologyopen 1:194-213.

28. Wu H, Liu Q, Casas-Pastor D, Dürr F, Mascher T, Fritz G. 2019. The role of C-terminal extensions in controlling ECF σ factor activity in the widely conserved groups ECF41 and ECF42. Molecular Microbiology 112:498-514.

29. Gómez-Santos N, Pérez J, Sánchez-Sutil MC, Moraleda-Muñoz A, Muñoz-Dorado J. 2011. CorE from Myxococcus xanthus Is a Copper-Dependent RNA Polymerase Sigma Factor. PLoS Genetics 7:e1002106-e1002106.

30. Marcos-Torres FJ, Pérez J, Gómez-Santos N, Moraleda-Muñoz A, Muñoz-Dorado J. 2016. In depth analysis of the mechanism of action of metal-dependent sigma factors: characterization of CorE2 from Myxococcus xanthus. Nucleic Acids Research 44:5571-5584.

31. Goutam K, Gupta AK, Gopal B. 2017. The fused SnoaL-2 domain in the Mycobacterium tuberculosis sigma factor σjmodulates promoter recognition. Nucleic Acids Research 45:9760-9772.

32. Liu Q, Pinto D, Mascher T. 2018. Characterization of the Widely Distributed Novel ECF42 Group of Extracytoplasmic Function σ Factors in Streptomyces venezuelae. Journal of Bacteriology 200.

33. Antelmann HaH, John D. 2011. Thiol-Based Redox Switches and Gene Regulation. Antioxidants & Redox Signaling 14:1049-1063.

34. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. 2009. Identification of direct residue contacts in protein-protein interaction by message passing. Proceedings of the National Academy of Sciences 106:67-72.

35. Bauer JS, Fillinger S, Förstner K, Herbig A, Jones AC, Flinspach K, Sharma C, Gross H, Nieselt K, Apel AK. 2017. dRNA-seq transcriptional profiling of the FK506 biosynthetic gene cluster in Streptomyces tsukubaensis NRRL18488 and general analysis of the transcriptome. RNA Biology 14:1617-1626.

36. Siranosian KJ, Ireton K, Grossman AD. 1993. Alanine dehydrogenase (ald) is required for normal sporulation in Bacillus subtilis. Journal of Bacteriology 175:6789-6796.

37. Bush MJ. 2018. The actinobacterial WhiB-like (Wbl) family of transcription factors. Molecular Microbiology 110:663-676.

38. Birch A, Häusler A, Hütter R. 1990. Genome rearrangement and genetic instability in Streptomyces spp. Journal of Bacteriology 172:4138-4142.

39. Cullum J, Altenbuchner J, Flett F, Piendl W. 1986. DNA amplification and genetic instability in Streptomyces. Biotechnol Genet Eng Rev 4:59-78.

40. Leblond P, Decaris B. 1994. New insights into the genetic of streptomyces instability. FEMS Microbiology Letters.

41. Volff J-N, Altenbuchner J. 1998. Genetic instability of the Streptomyces chromosome. Molecular Microbiology 27:239-246.

42. Zhang Z, Du C, de Barsy F, Liem M, Liakopoulos A, van Wezel GP, Choi YH, Claessen D, Rozen DE. 2020. Antibiotic production in Streptomyces is organized by a division of labor through terminal genomic differentiation. Science Advances 6:eaay5781.

43. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202-8.

44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.

45. Neyfakh AA, Bidnenko VE, Chen LB. 1991. Efflux-mediated multidrug resistance in Bacillus subtilis: similarities and dissimilarities with the mammalian system. Proceedings of the National Academy of Sciences 88:4781-4785.

46. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos Emmanuel LC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Research 45:W36-W41.

47. Andrews SC, Robinson AK, Rodriguez-Quinones F. 2003. Bacterial iron homeostasis. FEMS Microbiol Rev 27:215-37.

48. Imlay J, Chin S, Linn S. 1988. Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. Science 240:640-642.

49. Beites T, Pires SDS, Santos CL, Osório H, Moradas-Ferreira P, Mendes MV. 2011. Crosstalk between ROS homeostasis and secondary metabolism in *S. natalensis* ATCC 27448: modulation of pimaricin production by intracellular ROS. PLoS ONE 6:e27472.

50. Carpenter BM, West AL, Gancz H, Servetas SL, Pich OQ, Gilbreath JJ, Hallinger DR, Forsyth MH, Merrell DS, Michel SLJ. 2015. Crosstalk between the HpArsRS two-component system and HpNikR is necessary for maximal activation of urease transcription. Frontiers in Microbiology 6.

51. Lazar Adler NR, Allwood EM, Deveson Lucas D, Harrison P, Watts S, Dimitropoulos A, Treerat P, Alwis P, Devenish RJ, Prescott M, Govan B, Adler B, Harper M, Boyce JD. 2016. Perturbation of the two-component signal transduction system, BprRS, results in attenuated virulence and motility defects in Burkholderia pseudomallei. BMC Genomics 17:331.

52. Brenot A, Weston BF, Caparon MG. 2007. A PerR-regulated metal transporter (PmtA) is an interface between oxidative stress and metal homeostasis in Streptococcus pyogenes. Molecular Microbiology 63:1185-1196.

53. Caby M, Bontemps-Gallo S, Gruau P, Delrue B, Madec E, Lacroix J-M. 2018. The EnvZ-OmpR Two-Component Signaling System Is Inactivated in a Mutant Devoid of Osmoregulated Periplasmic Glucans in Dickeya dadantii. Frontiers in Microbiology 9.

54. Beites T, Oliveira P, Rioseras B, Pires SDS, Oliveira R, Tamagnini P, Moradas-Ferreira P, Manteca Á, Mendes MV. 2015. Streptomyces natalensis programmed cell death and morphological differentiation are dependent on oxidative stress. Scientific Reports 5:12887.

55. Hu Y, Kendall S, Stoker NG, Coates ARM. 2004. The Mycobacterium tuberculosis sigJ gene controls sensitivity of the bacterium to hydrogen peroxide. FEMS Microbiology Letters 237:415-423.

56. Yoo J-S, Oh G-S, Ryoo S, Roe J-H. 2016. Induction of a stable sigma factor SigR by translation-inhibiting antibiotics confers resistance to antibiotics. Scientific Reports 6:28628-28628.

57. González-Quiñónez N, Corte-Rodríguez M, Álvarez-Fernández-García R, Rioseras B, López-García MT, Fernández-García G, Montes-Bayón M, Manteca A, Yagüe P. 2019. Cytosolic copper is a major modulator of germination, development and secondary metabolism in Streptomyces coelicolor. Scientific Reports 9:4214-4214.

58. Chaplin AK, Tan BG, Vijgenboom E, Worrall JAR. 2015. Copper trafficking in the CsoR regulon of Streptomyces lividans. Metallomics 7:145-155.

59. Worrall JAR, Vijgenboom E. 2010. Copper mining in Streptomyces: enzymes, natural products and development. Natural Product Reports 27:742-756.

60. Dwarakanath S, Chaplin AK, Hough MA, Rigali S, Vijgenboom E, Worrall JAR. 2012. Response to Copper Stress in Streptomyces lividans Extends beyond Genes under Direct Control of a Copper-sensitive Operon Repressor Protein (CsoR). Journal of Biological Chemistry 287:17833-17847.

61. Bibb MJ, Domonkos A, Chandra G, Buttner MJ. 2012. Expression of the chaplin and rodlin hydrophobic sheath proteins in *Streptomyces venezuelae* is controlled by sigma(BldN) and a cognate anti-sigma factor, RsbN. Mol Microbiol 84:1033-49.

62. Traxler MF, Seyedsayamdost MR, Clardy J, Kolter R. 2012. Interspecies modulation of bacterial development through iron competition and siderophore piracy. Mol Microbiol 86:628-44.

63. Hwang S, Cordova B, Abdo M, Pfeiffer F, Maupin-Furlow JA. 2017. ThiN as a Versatile Domain of Transcriptional Repressors and Catalytic Enzymes of Thiamine Biosynthesis. Journal of Bacteriology 199:e00810-16.

64. Beinker P, Lohkamp B, Peltonen T, Niemi J, Mäntsälä P, Schneider G. 2006. Crystal Structures of SnoaL2 and AclR: Two Putative Hydroxylases in the Biosynthesis of Aromatic Polyketide Antibiotics. Journal of Molecular Biology 359:728-740.

65. Park ST, Kang C-M, Husson RN. 2008. Regulation of the SigH stress response regulon by an essential protein kinase in Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences 105:13105-13110.

66. Gust B, Challis GL, Fowler K, Kieser T, Chater KF. 2003. PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proceedings of the National Academy of Sciences 100:1541-1546.

67. Kieser T, Bibb M, Buttner M, Chater K, Hopwood DA. 2000. Practical *Streptomyces* Genetics. John Innes Centre, Norwich, United Kigdom.

68. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research 29.

69. Shin SH, Lim Y, Lee SE, Yang NW, Rhee JH. 2001. CAS agar diffusion assay for the measurement of siderophores in biological fluids. Journal of Microbiological Methods 44:89-95.

70. Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proceedings of the National Academy of Sciences 95:5752-5756.

71. Miller JH. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

72. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539.

73. Baldassi C, Zamparo M, Feinauer C, Procaccini A, Zecchina R, Weigt M, Pagnani A. 2014. Fast and Accurate Multivariate Gaussian Modeling of Protein Families: Predicting Residue Contacts and Protein-Interaction Partners. PLOS ONE 9:e92721.

74. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.

75. Liao Y, Smyth GK, Shi W. 2013. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923-930.

76. Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.

77. Sun J, Nishiyama T, Shimizu K, Kadota K. 2013. TCC: an R package for comparing tag count data with robust normalization strategies. BMC Bioinformatics 14:219.

78. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JGR, Korf I, Lapp H, Lehväslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E. 2002. The Bioperl Toolkit: Perl Modules for the Life Sciences. Genome Research 12:1611-1618.

79. Nicol JW, Helt GA, Blanchard SG, Jr., Raja A, Loraine AE. 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25:2730-2731.

80. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: A Sequence Logo Generator. Genome Research 14:1188-1190.

Table 1. SigG1-binding regions identified by ChIP-seq analysis

Peak	Upstream gene ID	Downstream gene ID	Product	Location in genome	Distance (bp)*	Δ*rsfG*/Δ*sigG1* (LogFC)	*sigG1*-dependent**
1	*STSU_t23051*	*STSU_23024*	MFS transporter	2553980-2554012	44	1,04	N
2	*STSU_28250*	*ald* (*STSU_28255*)	Alanine dehydrogenase, Ald1	1133124-1133156	30	0,97	Y
3	*STSU_22045*	*STSU_22050*	Putative antisense RNA	2781241-2781273	na	0,73	Y
4	*STSU_10736*	*wblC* (*STSU_10741*)	WhiB-like transcriptional regulator, WblC	5272304-5272336	117	0,71	Y
5	*STSU_32195*	*STSU_32197*	Hypothetical protein	375398-375430	187	0,62	Y
6	*STSU_06343*	*acnA* (*STSU_06338*)	Aconitate hydratase, AcnA	6313866-375430	1002	0,61	N

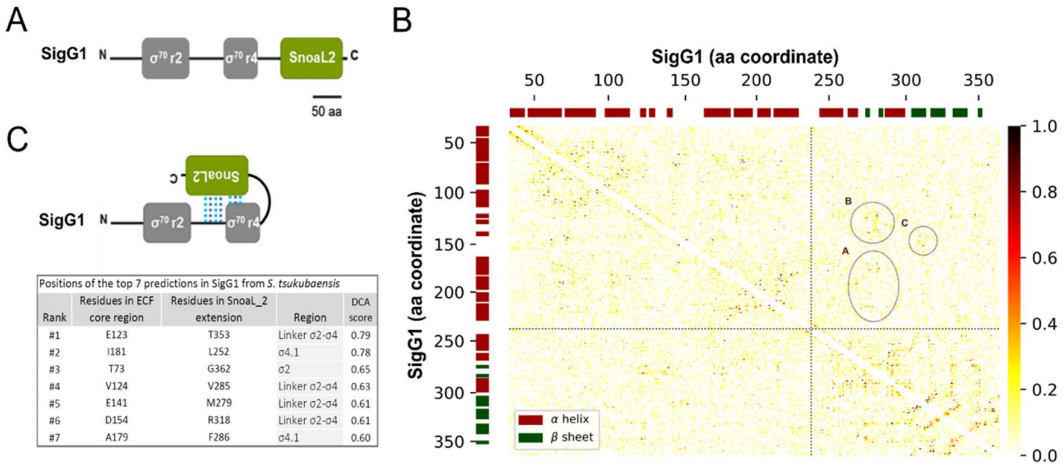
* Distance to START codon

** Causal relationship upon *sigG1* deletion, RNA-seq

Table 2. *rsfG*-dependent genes identified by RNA-seq analysis (FDR<0,05)

Gene ID	Product	Δ*rsfG*/WT (logFC)	*pFDR*
STSU_03649	Hypothetical protein	-3,3	1,2E-02
STSU_03654	Trypsin-like serine protease	-5,6	2,3E-04
STSU_11545	Hypothetical protein	1,8	2,6E-02
STSU_11550	Hypothetical protein	4,3	3,0E-04
STSU_11560	RNA polymerase factor sigma-70, SigG1	3,8	3,2E-03
STSU_12320	Subtilisin-like serine protease	-3,2	1,5E-02
STSU_16452	Amino acid permease-associated protein	-2,0	2,6E-02
STSU_16977	Hypothetical protein	-2,9	1,4E-02
STSU_16987	Hypothetical protein	-6,7	4,0E-05
STSU_19470	Transpeptidase	-2,1	2,5E-02
STSU_19475	Hypothetical protein	-4,9	2,9E-04
STSU_19480	ECF subfamily protein RNA polymerase sigma-24 subunit	-6,5	4,0E-05
STSU_25317	Phospholipase	-8,9	4,0E-05
STSU_25322	Protein phosphatase	-6,7	2,7E-04
STSU_25327	Hypothetical protein	-2,6	3,9E-03
STSU_31170	ErfK/YbiS/YcfS/YnhG family protein	-8,8	2,7E-04
STSU_31175	Cytochrome P450 family protein	-6,2	2,7E-04
STSU_32005	Rifamycin polyketide synthase, FkbB	-1,8	2,1E-02

Fig. 1



Positions of the top 7 predictions in SigG1 from *S. tsukubaensis*

Rank	Residues in ECF core region	Residues in SnoaL_2 extension	Region	DCA score
#1	E123	T353	Linker σ2-σ4	0.79
#2	I181	L252	σ4.1	0.78
#3	T73	G362	σ2	0.65
#4	V124	V285	Linker σ2-σ4	0.63
#5	E141	M279	Linker σ2-σ4	0.61
#6	D154	R318	Linker σ2-σ4	0.61
#7	A179	F286	σ4.1	0.60

Figure 1. Predicted intramolecular interactions for ECF56. (**A**) Schematic representation of the Pfam domain organisation of the *sigG1* encoded protein. SigG1 harbours the σ2 (PF04542) and σ4 (PF04545) conserved domains, and an additional long C-terminal extension, coding for a SnoaL_2 domain (PF12680). (**B**) Contact map for predicted contacts between ECF core regions and the SnoaL2 extension, using Direct Coupling Analyses (DCA). Each axis corresponds to the amino acids of SigG1 and the heatmap represents the DCA output, where darker spots correspond to higher scores. Scores >1 are set to 1 in order to allow for the observation of smaller scores. Dashed lines split the core ECF regions from the C-terminal extension. Significant contacts between SnoaL2 and core ECF regions are indicated by circles. (**C**) Model for compact conformation of SigG1 as predicted by DCA. DCA rank and scores of the significant contacts between ECF core regions and the SnoaL_2 domain are shown.

Fig. 2

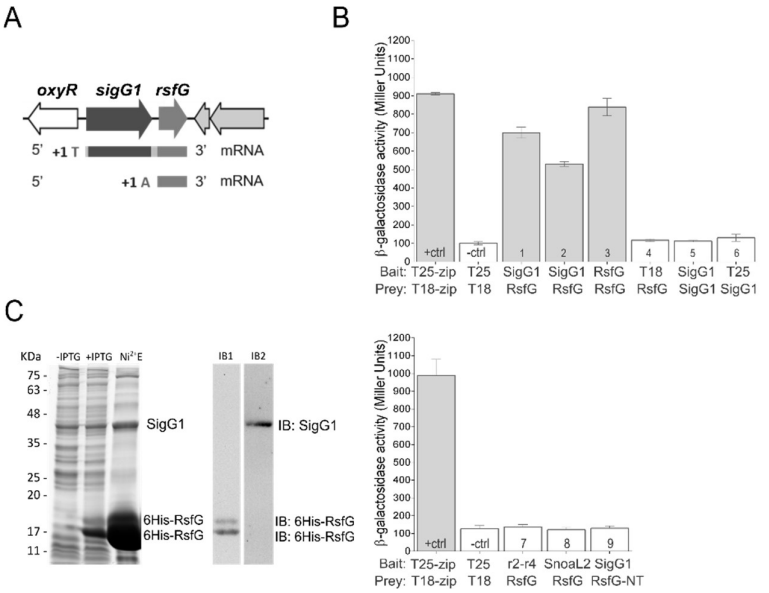


Figure 2. SigG1 interaction with RsfG evaluated *in vivo* (**A**) Co-transcription of *sigG1* and *rsfG* was observed by RT-PCR. Transcription start sites (TSS) were mapped by 5'RACE PCR. (**B**) BACTH assays for SigG1-RsfG binding *in vivo*. Positive controls pUT18Czip and pKT25-zip (+ctrl); Negative control with empty vectors pUT18C and pKT25 (-ctrl); T18-SigG1 and T25-RsfG (1) SigG1-T18 and T25-RsfG (2); T18-RsfG and T25-RsfG (3); pUT18 empty vector and T25-RsfG (4); pUT18 empty vector and T25-RsfG (4); T18-SigG1 and T25-SigG1 (5); pUT18 empty vector and T25-SigG1 (6); T25-RsfG and truncated forms of SigG1 – T18-SigG1_r2-r4 region (7); T25-RsfG and truncated forms of of SigG1 – T18-SigG1_SnoaL_2 domain (8); T25-RsfG_N-terminal and T18-SigG1 (9). Proteins were overexpressed in *E. coli* BTH101. Results are the average of three independent experiments. Grey bars indicate positive interaction between bait and prey. (**C**) Co-expression and co-purification of SigG1 and RsfG *in E. coli Nico21* cells. After incubation with 1 mM IPTG, cells were recovered and supernatants were incubated with a Ni-NTA agarose for Ni^{2+} affinity chromatography. SigG1 bound to the hexahistidine tagged RsfG was eluted with 250mM imidazole (Ni^{2+}E). Immunoblot detection using an antibody against the His-tag (IB1) and a polyclonal antibody against SigG1 (IB2).

Fig. 3

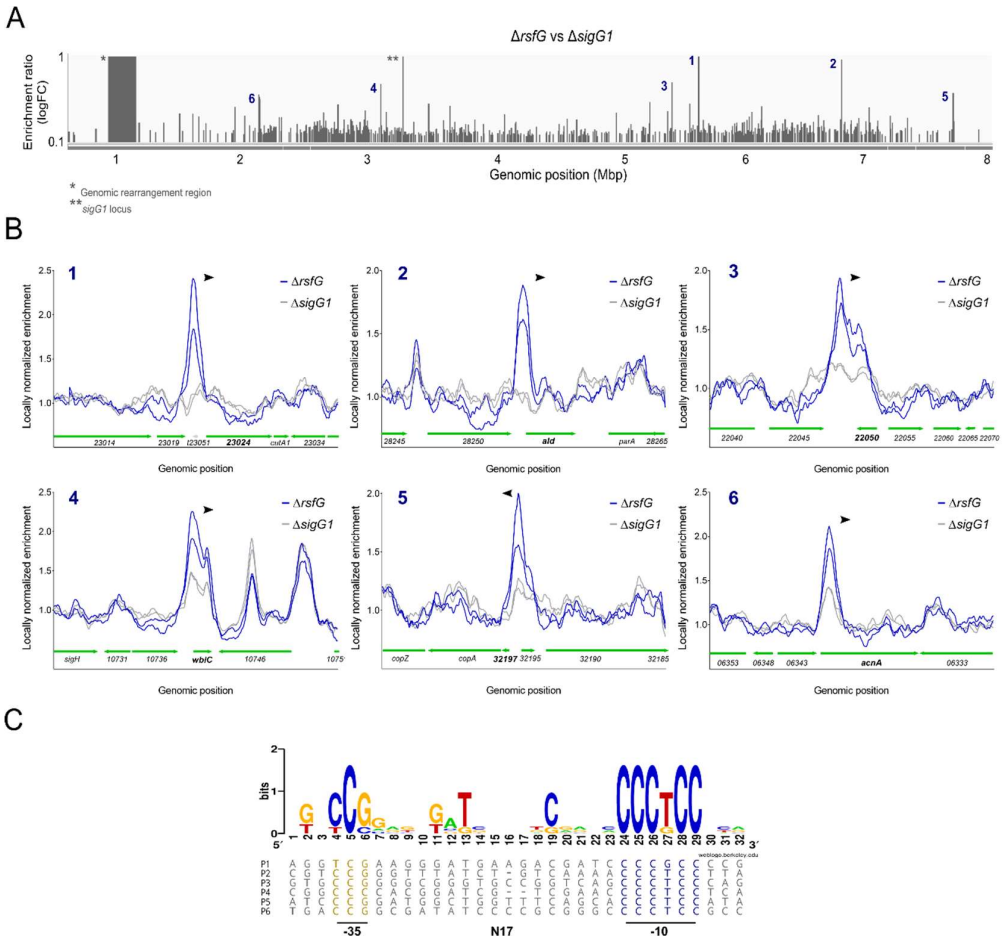


Figure 3. SigG1 binding to target promoters *in vivo* and *in vitro* (**A**) Genome-wide distribution of SigG1 binding sites identified by ChIP-seq analyses using the anti-SigG1 polyclonal antibody. Peak calling was performed from two independent biological replicates. (**B**) Close-up of a 2kb region around the target genes. *S. tsukubaensis* Δ*rsfG:* blue. Negative control (Δ*sigG1*): grey. Genes in each genomic context are depicted in green (**C**) MEME consensus for SigG1 DNA-binding motif obtained by aligning the *sigG1*-dependent ChIP-seq enriched sequences with the MUSCLE algorithm. The consensus logo was obtained using the WebLogo platform (1). 1: *pSTSU_23024*; 2: *pald*; 3: *STSU_22050*; 4: *pwblC*; 5: *pSTSU_32197*; 6: *pacnA*.

Fig. 4

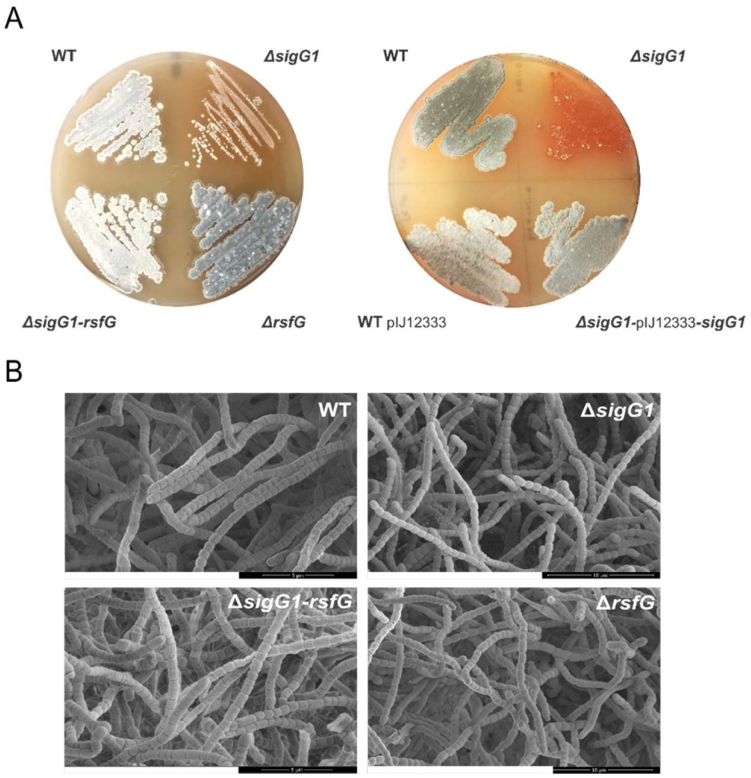


Figure 4. Phenotypes of *S. tsukubaensis* wild-type, *sigG1* and *rsfG null* mutants. (**A**) 10^6 spores were streak on ISP4 and grown for 14 days. To evaluate the phenotype of the complemented strains, 10^4 spores were plated on ISP4. The pIJ12333 vector was used to express *sigG1* coding sequence under the activity of its own promoter. The WT expressing the empty pIJ12333 is depicted as a control. (**B**) Differentiation into spores chains was examined by scanning electron microscopy (SEM), after for 14 days of growth on ISP4.

Fig. 5

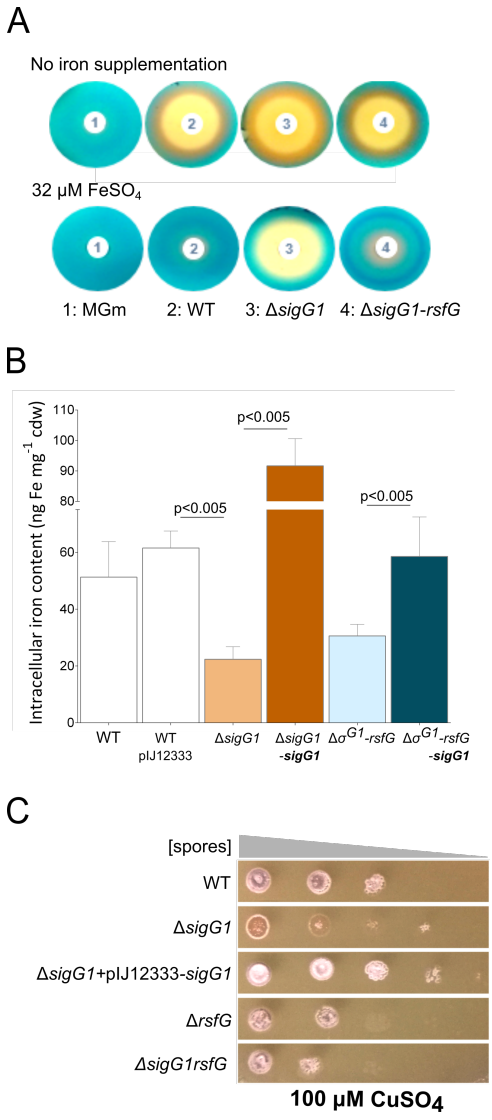


Figure 5. Evaluation of metal homeostasis in *sigG1 mutants*. (**A**) Chrome azurol S (CAS) assay to determine siderophore secretion in the supernatants of strains grown in iron limiting and iron-sufficient conditions ($FeSO_4$ 32μM) conditions. (**B**) Intracellular iron content in *S. tsukubaensis* wild-type, Δ*sigG1*, Δ*rsfG-sigG1* and respective complemented strains measured by flame atomic spectrometry (F-AAS). Values are representative of at least three independent experiments. (**C**) Phenotypes of *the* wild-type, *sigG1* and *rsfG-related* strains grown on solid ISP4 supplemented with 100 μM $CuSO_4$ medium for 6 days. Serial dilutions used were 10^5, 10^4, 10^3, 10^2, 10^1 spores.

Fig. 6

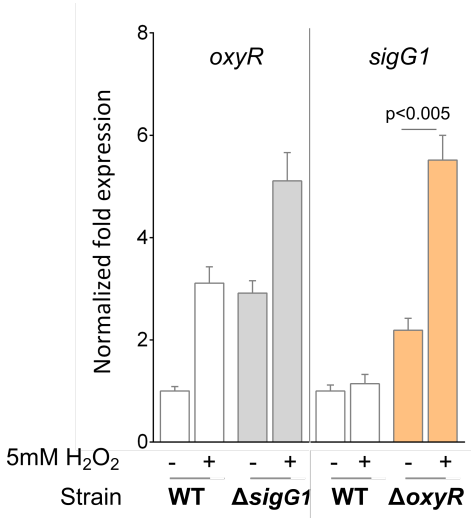


Figure 6. Stress-responsive gene expression. Transcription analyses of *oxyR* and *sigG1* expression in *S. tsukubaensis* mutant strains, upon a H_2O_2 stimulus evaluated by RT-qPCR. The Mean Normalised Fold Expression of the target genes was calculated relatively to the transcription of the reference genes (*rpsP* and *hrdB*) and internal normalisation was performed using the wild-type as the control. Results are representative of three biological replicates.

References

1. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: A Sequence Logo Generator. Genome Research 14:1188-1190.

Supplementary material I

Fig. S1

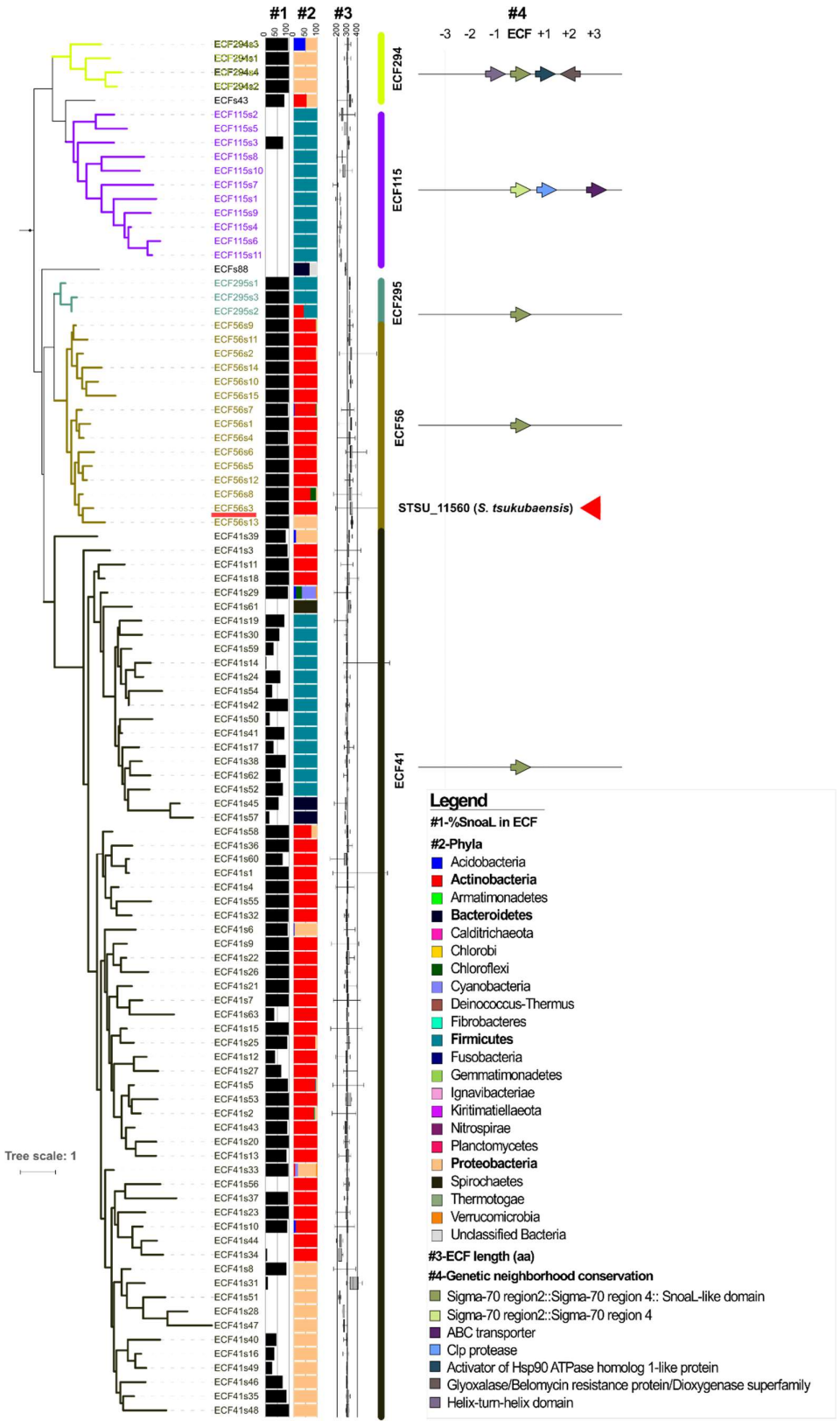


Figure S1. SigG1 phylogenetic distribution within the bacteria ECF group classification. The most recent classification distributes ECFs throughout a total of 157 families (see main text for references). Each family is labeled with a different colour. Phylogenetic analyses indicate that SnoaL2 containing ECFs (#1) are present in various phyla (#2 and indicated on the right) and are distributed throughout 5 ECF families. SigG1 is included in the ECF56 family (highlighted in dark gold), which includes proteins with the conserved σ2 and σ4 domains followed by a CT SnoaL2 domain, exclusively. ECF sigma factors were collected from 14483 bacterial genomes. The conserved position of the ECF relatively to other known proteins is depicted as *Genomic Neighbourhood Conservation* (#4).

Fig. S2

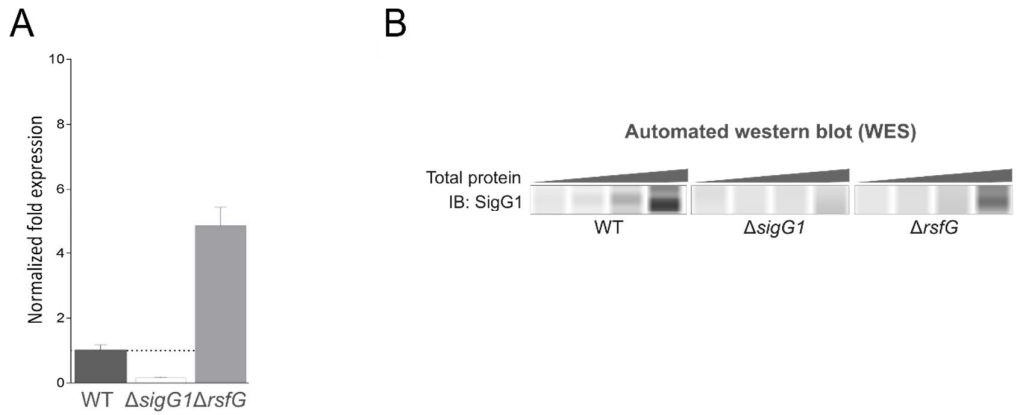


Figure S2. *SigG1* gene expression in *S. tsukubaensis*. (**A**) Transcript levels in *sigG1*-related mutant strains evaluated by RT-qPCR. (**A**) SigG1 protein expression in the wild-type and the *rsfG null* mutant was detected by automated western blot (WES Simple) using a polyclonal anti-SigG1 antibody (**B**).

Fig. S3

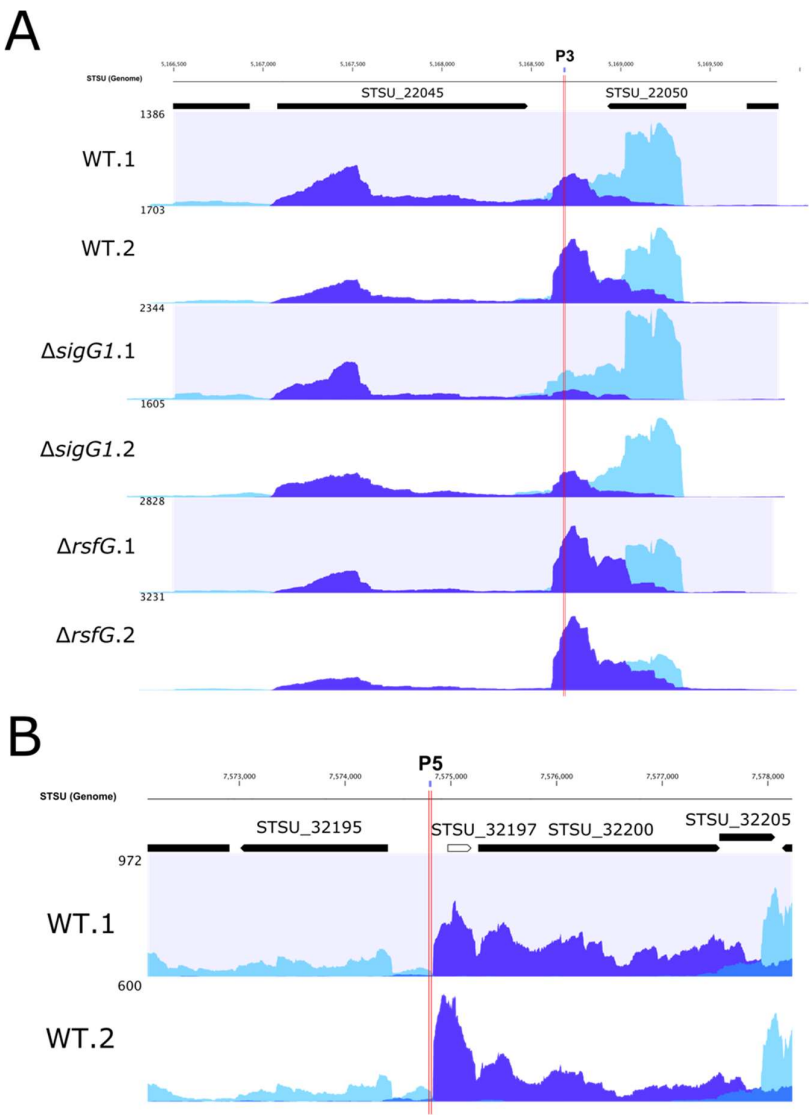


Figure S3. Visualisation of RNA-seq coverage in selected genomic regions. (**A**) Reads coverage of the genomic region that includes the ChIP-seq peak 3 (P3) showing a putative antisense RNA for *STSU_22050*, regulated by SigG1. *STSU_22050* codes for a hypothetical protein. (**B**) Reads coverage of the ChIP-seq peak 5 (P5) region showing the presence of a putative a new gene (*STSU_32197*), encoding a glycine-rich transmembrane hypothetical protein, upstream of Cu^{2+}-exporting ATPase (STSU_32200). The y-axis shows the read count. Dark blue indicates coverage of the forward reads and light blue reads mapping in the reverse orientation. NGS data was obtained by RNA-seq and SigG1 native ChIP-seq experiments.

Fig. S4

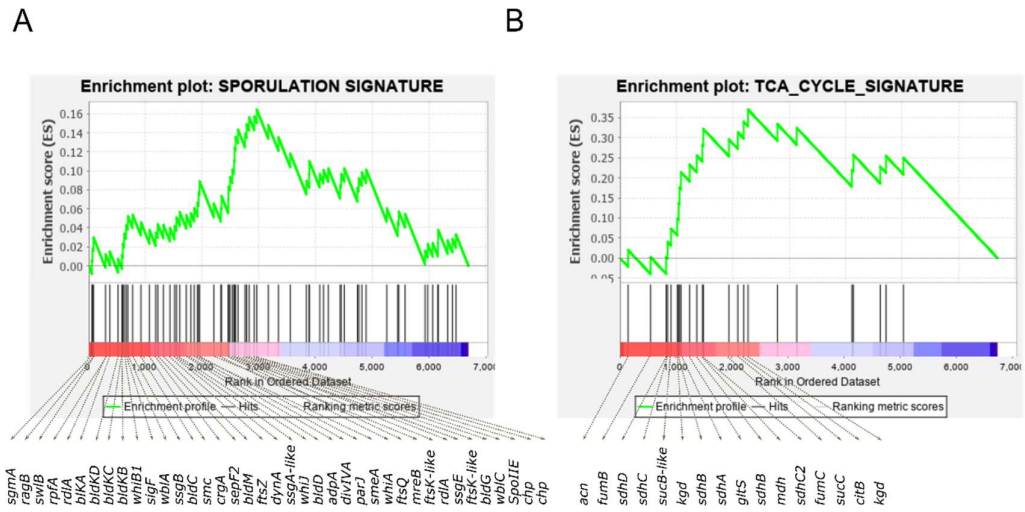


Figure S4. Gene set enrichment analysis (GSEA) in the Δ*rsfG* mutant. Molecular signatures for specific cellular pathways were scanned against the whole transcriptome of the wild-type and the Δ*rsfG* obtained by RNA-sequencing. Enrichment of candidate gene signatures in Δ*rsfG* as compared to the wild-type are depicted. (**A**) GSEA Pre-rank analysis using the molecular signature of known regulators of sporulation in *Streptomyces* Normalized Enrichment Score (NES)= 1.62; FDR q-value= 0.039. (**B**) GSEA Pre-rank analysis of the molecular signature of proteins involved in the TCA cycle in *Streptomyces*. Normalized Enrichment Score (NES)= 2.17; FDR q-value= 0.002.

Fig. S5

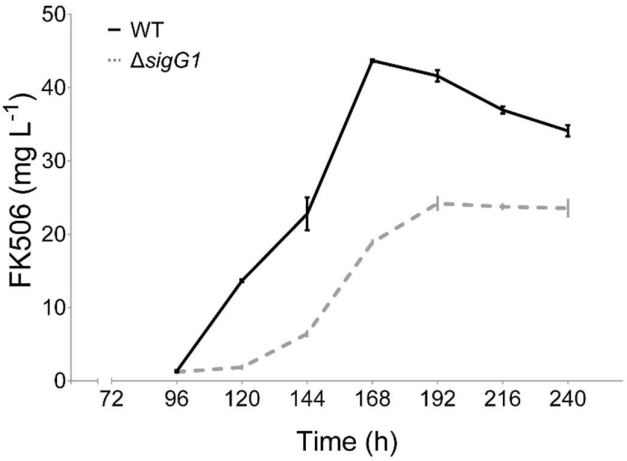


Figure S5. Analysis of FK506 biosynthesis of wild-type (black line) and Δ*sigG1* strains (dashed line) throughout growth. (A) FK506 production measured by high performance liquid chromatography (HPLC). Values are representative of at least three independent experiments.

Fig. S6

Figure S6. Multiple sequence alignment of ECF41 activity inhibiting motif. Selected sequences for sigma factors representatives of the ECF41 and ECF56 families were aligned as described in (2). The figure depicts the consensus NPDKL motif of the NTF2-like CT region required to interact with the σ_2 and σ_4 linker and inhibit the ECF activity (3). The conserved motif in ECF41 is not present in ECF56. ECF domains and secondary structure are shown. DCA predictions are mapped for selected residues. Pairs of contacting residues are highlighted in colour.

References

2. Casas-Pastor D, Müller RR, Becker A, Buttner M, Gross C, Mascher T, Goesmann A, Fritz G. 2019. Expansion and re-classification of the extracytoplasmic function (ECF) σ factor family. bioRxiv doi:10.1101/2019.12.11.873521:2019.12.11.873521.

3. Wecke T, Halang P, Staron A, Dufour YS, Donohue TJ, Mascher T. 2012. Extracytoplasmic function sigma factors of the widely distributed group ECF41 contain a fused regulatory domain. Microbiologyopen 1:194-213.

Table S1. Strains, plasmids and primers used in this study

Strains	Relevant genotype	Reference
S. tsukubaensis		
NRRL 18488	Wild type	(1)
Δ*sigG1*	Δ*sigG1*::*apr*; Apr^R	This study
Δ*rsfG*	Δ*rsfG*::*apr*; Apr^R	This study
Δ*sigG1-rsfG*	Δ *sigG1-rsfG*::apr; Apr^R	This study
WT pIJ12333	Wild type expressing pIJ12333 empty vector	This study
Δ*sigG1* pIJ12333*+sigG1*	Δ*sigG1*::apr expressing pIJ12333-*sigG1*; Tsr^R Apr^R (ϕC31-transduction)	This study
Δ*sigG1-rsfG1* pIJ12333 *+sigG1*	Δ*STSU_11555-60*::apr expressing pIJ12333-*STSU_11560*; Tsr^R Apr^R (ϕC31-transduction)	This study
Δ*oxyR*	Δ*oxyR::apr*; Apr^R	(Pires et al, unpublished results)
E.coli		
BTH101	F- *cya-99 araD139 galE15 galK16 rpsL1 (Strr) hsdR2 mcrA1 mcrB1*	(2)
BW25113 [pIJ790]	Δ(*araD-araB*)*567*, Δ*lacZ4787*(::*rrnB-4*), *lacIp-4000*(*lacI^Q*), λ-, *rpoS369*(Am), *rph-1*, Δ(*rhaD-rhaB*)*568*, *hsdR514*; Cm^R	(3)
DH5α	F- ϕ80*lacZ*ΔM15 Δ(*lacZYA-argF) U*169 *recA1 endA1 hsd*R17*(*rK-, mK+*) pho*A *sup*E44 λ– *thi-1 gyr*A96 *relA1*	
ET12567 [pUZ8002]	*dam, dcm, hsd*, Kan^R , Cm^R	(4)
Nico21	*can*::CBD *fhuA2* [*lon*] *ompT gal* (λ DE3) [*dcm*] *arnA*::CBD *slyD*::CBD *glm*S6Ala Δ*hsdS* λ DE3 = λ *sBamHIo* Δ*EcoRI-B int*::(*lacI*::P*lac*UV5::T7 gene1) *i21* Δ*nin5*	New England Biolabs

Plasmids	Comments	Source
pIJ12333	Vector for conjugal transfer of DNA from *E. coli* to *Streptomyces*; integrative (ϕC31-transduction), Hyg^R Tsr^R	(5)
pIJ12333-*SigG1*	pIJ12333 carrying *sigG1* preceded by its native promoter, Tsr^R	This study
pIJ773	Plasmid template for amplification of the *apr-oriT* cassette for Redirect™ PCR-targeting	(6)
pIJ790	Modified l RED recombination plasmid [oriR101] [repA101(ts)] araBp-gam-be-exo, Cm^R	(6)
pUZ8002	RP4 derivative with defective oriT, Kan^R	(4)
pET15b	T7 expression vector, Amp^R	Novagen
pET15b-SigG1	pET15b-*sigG1* full length coding sequence	This study
pRSFDuet-1™	T7 co-expression vector. Contains two multiple cloning sites (MCS), Kan^R	Novagen

pRSFDuet3	pRSFDuet-1 carrying *rsfG* cloned into MCS1 and *sigG1* cloned into MCS2 of pRSFDuet-1	This study
pUT18C	Two-hybrid vector, N-terminal cyAT18fusion, Amp^R	(2)
pUT18	Two-hybrid vector, C-terminal cyAT18fusion, Amp^R	(2)
pKT25	Two-hybrid vector, N-terminal cyAT25fusion, Kan^R	(2)
pKNT25	Two-hybrid vector, C-terminal cyAT25fusion, Kan^R	(2)
pKT25-zip	Derivative of pKT25 carrying the leucine zipper of GCN4 fused in frame to T25	(2)
pUT18C-zip	Derivative of pUT18C carrying the leucine zipper of GCN4 fused in frame to T18	(2)

Table S2. Oligonucleotides used in this study

Primer	**Sequence 5' - 3'**	**Use**
Red sigG1 F	GTGTTTCCCGGTGGTGCTGCTGATACGAAGAGAGGTATGATTCCGGGGATCCGTCGACC	Redirect™ gene KO
Red_sigG1_R	CACGGGAAAGGAATCGGCGGACTGTTGTGACCCGGGTCATGTAGGCTGGAGCTGGTTC	Redirect™ gene KO
Red_rsfG_F	GCGCACCCTCACCACCACGGCAAAGGACGCAGAACCATGATTCCGGGGATCCGTCGACC	Redirect™ gene KO
Red_rsfG_R	ACGGGACGTCCGGGCCGCCGTCGCTCCGGCCGGTGTTCATGTAGGCTGGAGCTGCTTC	Redirect™ gene KO
Red_conf_sigG1_F	CGGAAGTGCAGATGCTCGGC	Redirect™ gene KO
Red_conf_sigG1_R	AACCGAGCAAAGGCATGACC	Redirect™ gene KO
Red_conf_rsfG_F	TGCGGCTCCCACGGGACGT	Redirect™ gene KO
Red_conf_rsfG_R	GCACCCTCACCACCACGGCA	Redirect™ gene KO
RT_sigG1_F	GACGGTCATATCGAGGGCGT	RT-qPCR analyses
RT_sigG1_R	GTGATCCAGCATCTGCCGCC	RT-qPCR analyses
RT_rsfG_F	CCGAACTGCTCCAGGGCGAT	RT-qPCR analyses
RT_rsfG_R	GGACAGGCTGATCGGCACCT	RT-qPCR analyses
RT_oxyR_F	CAGCCCAGCCTGTCGCAACT	RT-qPCR analyses
RT_oxyR_F	AGAGCGCGGGCTGACTCATG	RT-qPCR analyses

RT_hrdB_F	GCGGCACTGACCATCAGCGT	RT-qPCR reference gene
RT_hrdB_F	GATTCCGCCAACCCAGTGGA	RT-qPCR reference gene
RT_rpsP_F	GCGCCGACGGAAAGCCAGTA	RT-qPCR reference gene
RT_rpsP_R	CCATCGAGGAGATCGGCCTG	RT-qPCR reference gene
GSP1_sigG1	CAGGGCGAGTGCGGACAGGT	TSS mapping
GSP3_sigG1	GATGCTCGGCTCCTTCACC	TSS mapping
GSP1_rsfG	CAGATGGTGAGGGTGTCGCC	TSS mapping
GSP2_rsfG	CGTAGTAGCGGGAGCGGATGT	TSS mapping
5'RACE AAP	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	TSS mapping
5´RACE AUAP	GGCCACGCGTCGACTAGTAC	TSS mapping
KpnI_sigG1_F	CGGGTACCTATGAGCGACCAGGGTGG	Bacterial two-hybrid
EcoRI_sigG1_stop_R	AGGAATTCGGCGGACTGTTGTGACC	Bacterial two-hybrid
EcoRI_sigG1_R	CTGAATTCTCCGCCAGGGTGGGG	Bacterial two-hybrid
KpnI_rsfG_F	CTGGTACCCATGAACGACACCACCGC	Bacterial two-hybrid
EcoRI_rsfG_stop_R	ACGAATTCACGGGGTGATCGGGAC	Bacterial two-hybrid
EcoRI_rsfG_R	ATGAATTCACCGCGACCCGGGTCA	Bacterial two-hybrid
KpnI_SigG1SnoaL2	GTGGTACCTGCGGCCACCGAGCCCAGCG	Bacterial two-hybrid
EcoRI_SigG1r2r4	TAGAATTCCGCACCCCACTCCGA	Bacterial two-hybrid
EcoRI_rsfGNT	ATGAATTCGGTGCCGATCAGCCTGTC	Bacterial two-hybrid
T18seq_F	GTGTGGAATTGTGAGCGGAT	Bacterial two-hybrid
T18seq_R	TTCCACAACAAGTCGATGCG	Bacterial two-hybrid
T25seq_F	CGGTGACCAGCGGCGATT	Bacterial two-hybrid
T25seq_R	GGCGATTAAGTTGGGTAACGCC	Bacterial two-hybrid
NT25seq_F	CCCCAGGCTTTACACTTTATGC	Bacterial two-hybrid
NT25seq_R	TTGATGCCATCGAGTACGGCT	Bacterial two-hybrid
T18Cseq_F	GTGCCGAGCGGACGTTCGA	Bacterial two-hybrid
T18Cseq_R	CTTAACTATGCGGCATCAGAGC	Bacterial two-hybrid
rsfG_MCS1_F	ACACAGAATTCATGAACGACACCACCGCA	Protein co-expression

rsfG_MCS1_R	CATAAGCTTTCACCGGTAGTAGGCCGGGGA	Protein co-expression
sigG1_MCS2_F	GAACCCATATGAGCGACCAGGGTGGC	Protein co-expression
sigG1_MCS2_R	CAGGTACCATCGGCGGACTGTTGTGA	Protein co-expression
SigG1_SpeI_F	GCACTAGTTCACGGGAAAGGAATCGG	*psigG1-sigG1* expression
SigG1_SpeI_R	GCACTAGTAGAGCAGCACCTTCCGGG	*psigG1-sigG1* expression

References

1. Barreiro C, Prieto C, Sola-Landa A, Solera E, Martinez-Castro M, Perez-Redondo R, Garcia-Estrada C, Aparicio JF, Fernandez-Martinez LT, Santos-Aberturas J, Salehi-Najafabadi Z, Rodriguez-Garcia A, Tauch A, Martin JF. 2012. Draft genome of *Streptomyces tsukubaensis* NRRL 18488, the producer of the clinically important immunosuppressant tacrolimus (FK506). J Bacteriol 194:3756-7.
2. Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proceedings of the National Academy of Sciences 95:5752-5756.
3. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640-5.
4. Paget MSB, Chamberlin L, Atrih A, Foster SJ, Buttner MJ. 1999. Evidence that the Extracytoplasmic Function Sigma Factor çE Is Required for Normal Cell Wall Structure in Streptomyces coelicolor A3(2). Journal of Bacteriology 181:204-211.
5. Sherwood EJ, Hesketh AR, Bibb MJ. 2013. Cloning and Analysis of the Planosporicin Lantibiotic Biosynthetic Gene Cluster of Planomonospora alba. Journal of Bacteriology 195:2309-2321.
6. Gust B, Challis GL, Fowler K, Kieser T, Chater KF. 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci U S A 100:1541-6.

Supplementary material II

SnoaL2- containing ECF sigma factors.

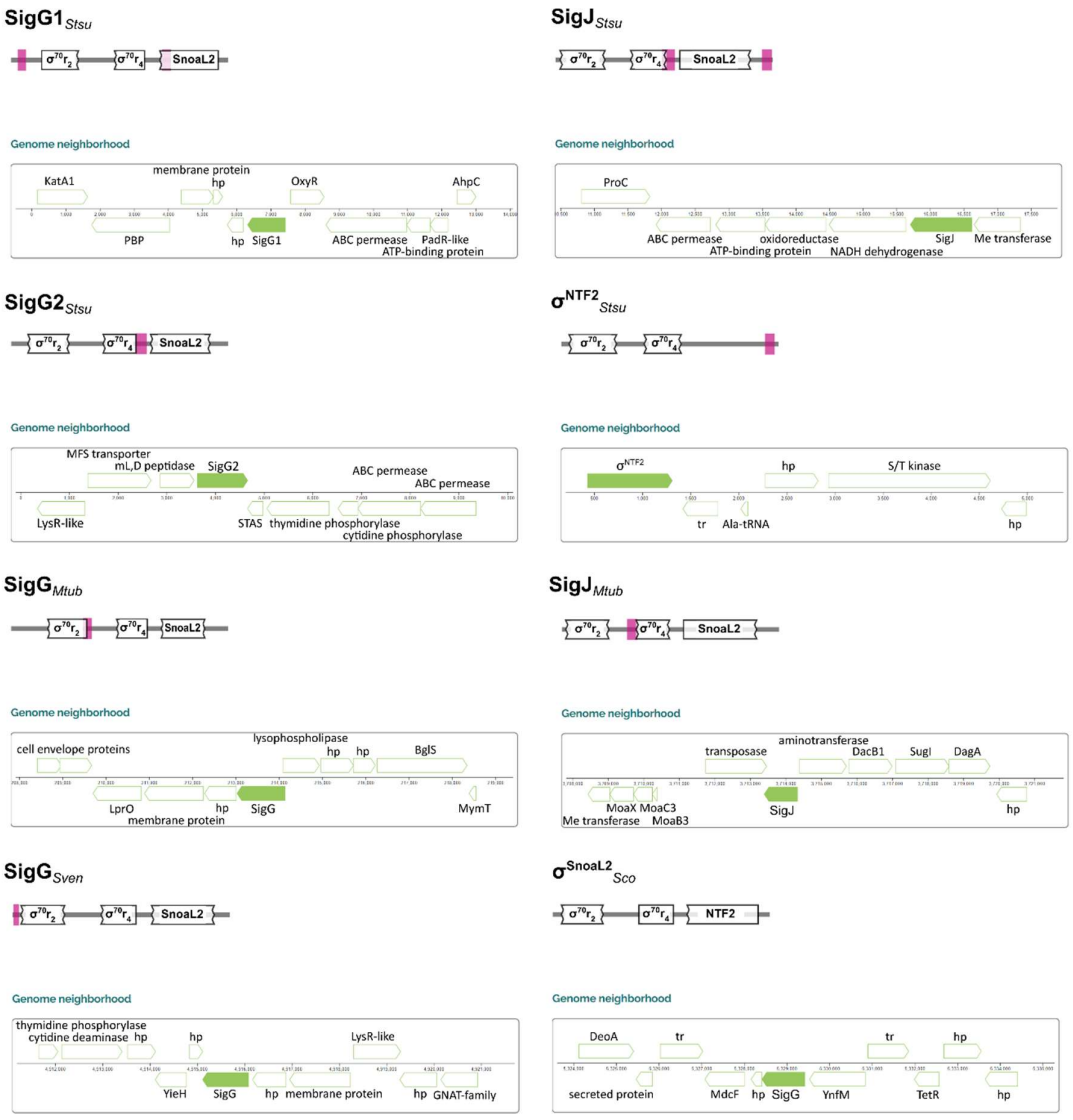


Figure SII1. Illustration of the *Streptomyces* SnoaL2-containing ECFs and their closest homologues in *M. tuberculosis*. Schematic representation of the conserved domain organisation in SnoaL2 ECF proteins, with disordered regions shown in pink. The genomic neighbourhood for each ECF is indicated. Adapted from the MIST 3.0 database (https://mistdb.com/genomes).

Deletion of *sigG* and *rsfG* genes in *S. tsukubaensis.* To disrupt *sigG1* and *rsfG1* we replaced each coding sequence with an apramycin resistance cassette by homology recombination, using the REDIRECT™ assay (Figure SII2). Gene deletion was confirmed by PCR using primers for the flanking regions of both genes and by Southern blot hybridization with using a specific probe labelled with digoxigenin (DIG).

Genomic DNA of the deletion mutants was compared to the wild type, and the differences in the target region were visualized using an antibody against digoxigenin (DIG) (Figure SII3).

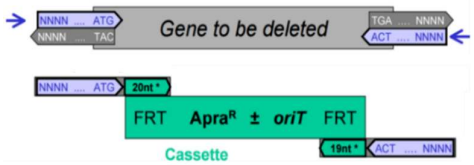


Figure SII2. Illustration of in-frame gene complete deletion by replacement of the chromosomal locus of interest with an apramycin resistance cassette. Blue arrows indicate the primers used to create in-frame deletions. For primer design, the sequence flanking the gene (NNNN) was fused to 19/20 nt of the apramycin coding sequence (apramycin resistance cassette flanked by FRT recombination sites is in green). Extracted from (Gust et al., 2003).

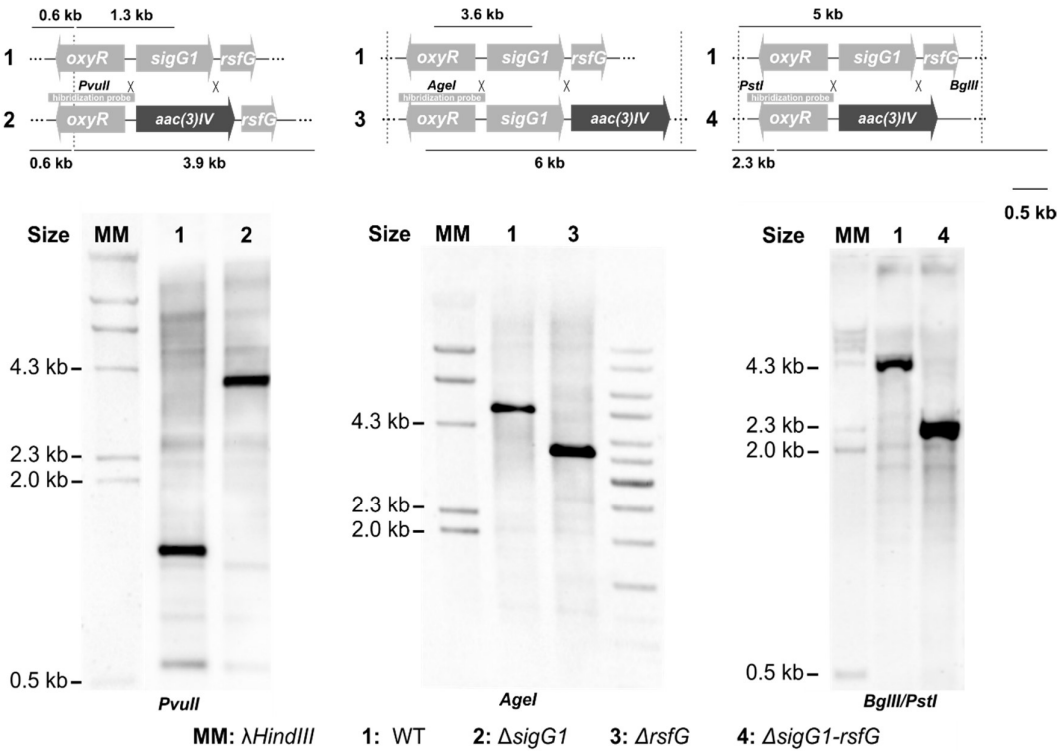


Figure SII3. Construction of Δ*sigG1*, Δ*rsfG* and Δ*sigG1-rsfG1* strains by gene replacement using the REDIRECT™ technology. Predicted restriction enzyme polymorphisms caused by gene replacement are depicted on the upper panels. The *PvuII*, *AgeI* or *BglII*+*PstI* restriction patterns before and after replacement are shown. The lower panels correspond to the confirmation of gene disruption by Southern blot hybridisation of the *PvuII*, *AgeI* or *BglII*+*PstI* digested chromosomal DNA. A gray bar indicates the probe used for southern hybridisation. A 1:1 mixture of λ*HindIII* with DIG-labelled DNA Molecular Weight Marker II (Roche) was used (MM).

Plasmid integration into *S. tsukubaensis* chromosome. Within the toolbox available for *Streptomyces* strains, we find a system that allows the incorporation of selected DNA features to manipulate gene expression. For instance, bacteriophage-derived integration systems has been routinely used to achieve site-specific recombination of alleles, under the activity of promoters or other regulatory features of interest. To select the best system to integrate and express *sigG1* and *rsfG* wild type alleles in complementation experiments (for phenotype recovery or for chromatin immunopurification using a FLAG-tagged version of SigG1), we picked a set of vectors that were available for the other well-studied *Streptomyces* strains and tested their integration into the *S. tsukubaensis* genome. Due to the presence of an apramycin cassette in the deletion mutants generated, we started by searching for a suitable resistance gene to be used in combination with apramycin for selection of the integrated vectors. As an industrial strain, there were not many studies available with molecular manipulation of *S. tsukubaensis*. To characterise the antibiotics resistance we grew wild type cells in the presence of a set of antibiotics widely used for plasmid selection (see Table SII1 and Figure SII4 for an example). *S. tsukubaensis* wild type cells showed natural resistance to hygromycin and viomycin.

Table SII1. Selection of antibiotic resistance cassette for plasmid integration in *S. tsukubaensis*

Strain	**Antibiotics commonly used to select for *Streptomyces***				
S.tsukubaensis	Viomycin^{R}	Hygromycin^{R}	Kanamycin^{S}	Thiostrepton^{S}	Apramycin^{S}

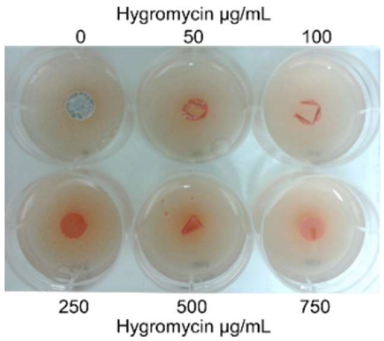


Figure SII4. *S. tsukubaensis* natural resistance to hygromycin. *S. tsukubaensis* wild type was grown on ISP4 solid media supplemented with hygromycin or 7 days. Cells survived to 0, 50, 100, 250, 500, 750 µg/mL concentrations of hygromycin.

In phage-encoded integration the attB site, usually adjacent to the integrase gene, is conserved among *Streptomyces* genomes. Examination of the *S. tsukubaensis* genome unveiled which of the most common phage attachment sites were conserved in this strain (Table SII2). By interspecies conjugation, we confirmed the effective integration of exogenous DNA in *S. tsukubaensis* ϕC31 and VWB integration sites.

Table SII2. Integration sites for plasmid integration into *S. tsukubaensis* genome

attB site / Strain	ϕC31	ϕBT1	TG1	SV1	pSam2	VWB	ϕJoe
S. tsukubaensis							
No integration (plasmid)		pRT802	pRF11	pBF3	[1]		[1]
Integration (plasmid)	pIJ12333				[1]	pSOK804.1	[1]

[1]: Not tested experimentally in this work
[2]: No attB site found in genome.

Physiological characterisation of deletion mutants. To investigate the physiological role of new molecular players we have routinely used two main growing media. For liquid cultures, we grew *Streptomyces* cells in a defined medium suitable for FK506 producing conditions (MGm-2.5, (Martinez-Castro et al., 2013)). This medium supports high FK506 production yields but also a dispersed growth, that has been well characterised. For mycelia growing on solid agar we used ISP4 (Difco), which is the only known defined medium where we observe full sporulation. These conditions were advantageous because they ensured the fitness of the cells until the completion of secondary metabolism. Growth in these media revealed that the new sigma/anti-sigma system was essential during developmental transitions, often characterised by nutrient limitation events (Figure SII5). Hence, we sought to carry out a deeper characterisation of growth in other carbon sources, to investigate the effect of using different nutritional sources and better characterise the morphological phenotype. Results showed that *S. tsukubaensis* strains failed to complete sporulation when grown in maltose (MYM), sucrose (R2, R5) or glucose (R2, R5), instead of having starch (ISP4) as the main carbon source. The developmental delay caused by the loss of *sigG1* was more evident in maltose growing conditions, but we could not discard that this was due to a polar effect of the genomic rearrangement observed for this strain (Figure SII6).

As described before, we addressed the maintenance of the redox balance metal homeostasis by measuring levels of intracellular iron. We observed that the low availability of iron in Δ*sigG1* is sustained throughout growth and is independent of the supplementation of the culture media with exogenous iron (Figure SII7A). We further observed that *sigG1*-depleted cells were able to increase the production of iron-chelating siderophores whereas *rsfG* deletion mutants did not (Figure SII7B). To investigate whether the iron-deficient environment was due to a reprogramming of iron acquisition or export, we tested of *sigG1*-related mutants in the presence of metals, which could complement the loss of *sigG1*, and restore the wild type phenotype. Complementation with iron ($FeSO_4$) or magnesium ($MgCl_2$) was not achieved with these experiments, since no clear restoring of the phenotype was observed. This was indicative that the metal

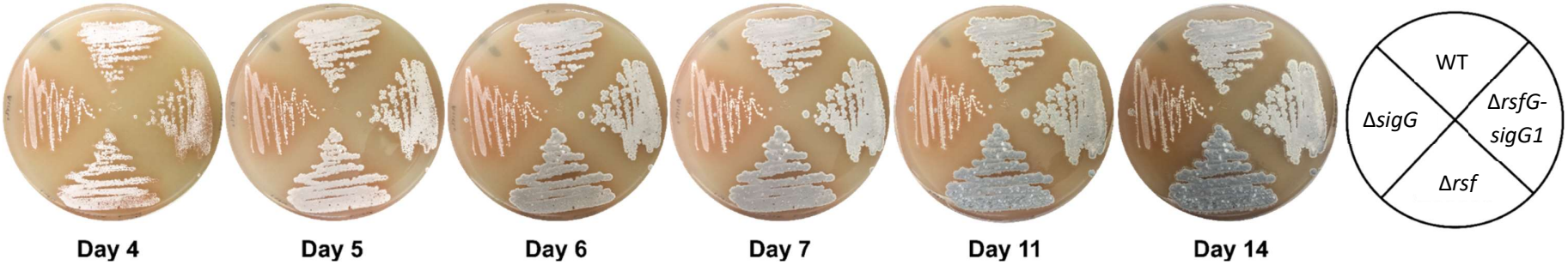


Figure SII5. Phenotypes of *sigG1* and *rsfG*-related strains on solid sporulation-substrate. Strains were grown fro 14 days at 30˚C in defined ISP4 (Difco) medium.

Figure SII6. Phenotypes of *sigG1* and *rsfG*-related strains on solid substrates with different carbon sources. Defined media with starch (ISP4); maltose, yeast extract and malt extract (MYM); sucrose and glucose (R2); sucrose, glucose, casamionacids and yeast extract (R5) carbon sources were used.

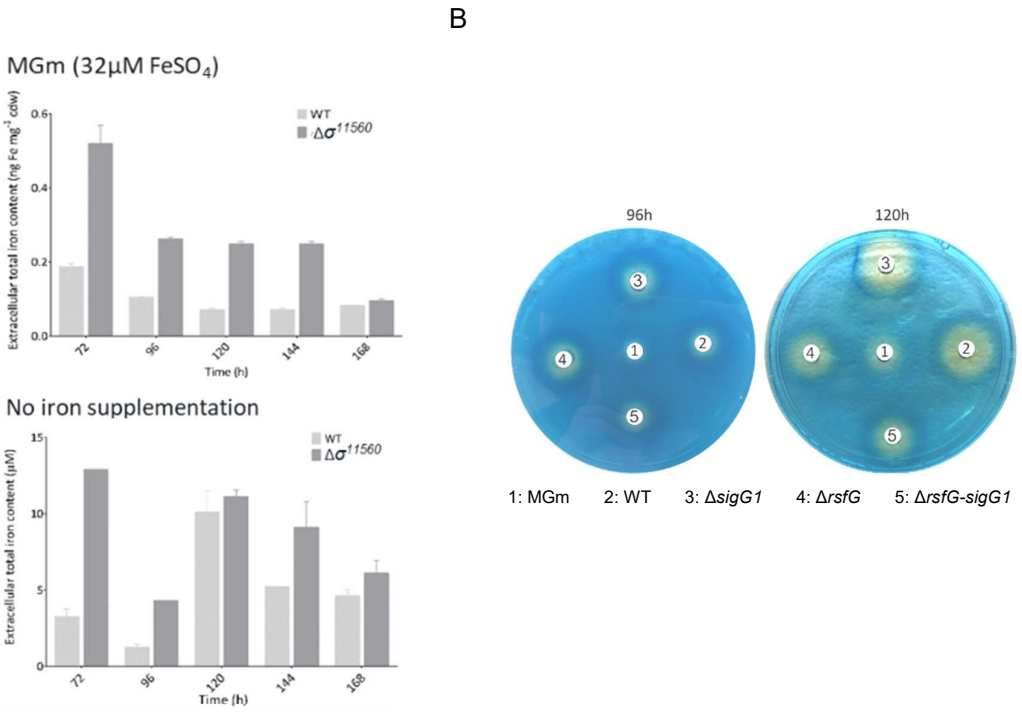
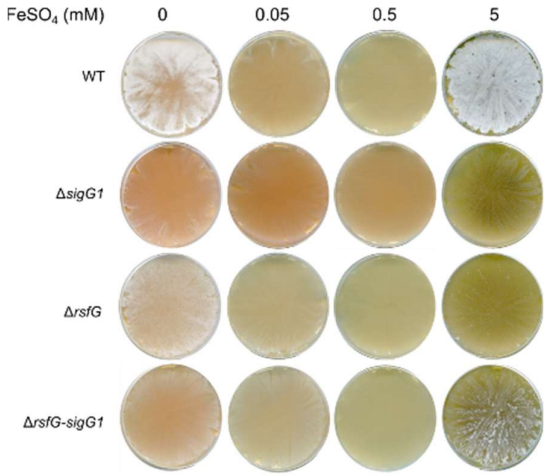


Figure SII7. SigG1 impacts iron homeostasis in *S. tsukubaensis*. Extracellular iron content was determined throughout growth of *S. tsukubaensis* strains in MGm-2.5, using the kit QuantiChrom™ Iron Assay Kit (DIFE-250) (**A**). Chrome azurol S (CAS) assay to determine siderophore secretion in the supernatants of *S. tsukubaensis* strains grown in iron-sufficient conditions. (**B**) Values are representative of at least three independent experiments. 1: MGm; 2: WT; 3: Δ*sigG1; 4:* Δ*rsfG; 5:* Δ*rsfG-sigG1.*

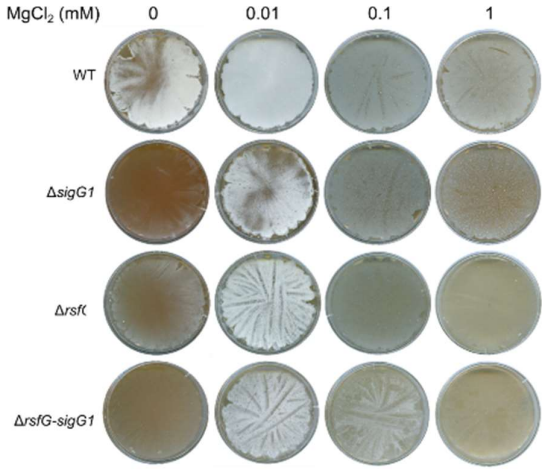
availability phenotype was a more complex indirect effect, and not exclusively due to a defective metal uptake (Figure SII8).

We further observed that loss of *sigG1* induced oxidative stress-responsive gene expression mainly through the activation of genes related to ROS degradation enzymes, as well as genes encoding iron-sulfur containing proteins (Figure SII9). Moreover, it increased intracellular reactive oxygen species (ROS) and activated essential ROS detoxifying enzymes (Figure SII9). The Δ*rsfG* and the double mutant did not show significant differences in oxidative stress when compared to the wild type. In addition, we characterised the oxidative stress response by exposing the cells to stress-inducing agents. The *sigG1* and *rsfG*-depleted strains showed increased resistance to both stress induced by exogenous H_2O_2 and to disulfide stress induced by diamide. In agreement with the previous findings, the *sigG*1 mutant strain shows no sensitivity to the presence of exogenous H_2O_2 (Figure SII10). An explanation for this could be the activation of antioxidant enzymes (Figure SII9). Examination of the mRNA-seq levels does not show alterations in the molecular signatures associated to the respiratory chain or oxidative

A



B



C

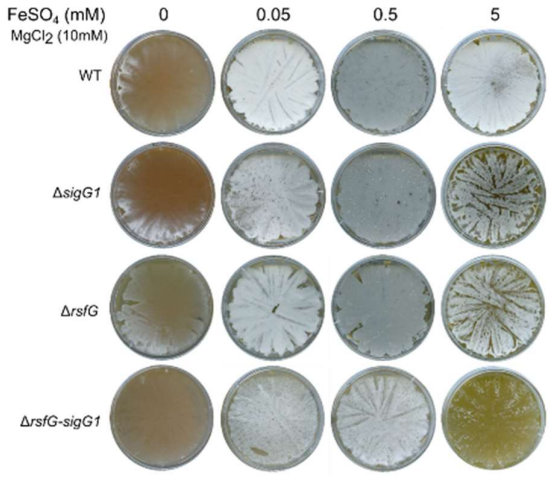


Figure SII8. Phenotypes of *sigG1* and *rsfG*-related strains on solid substrates supplemented with divalent metals. Strains were grown for 7 days on ISP4 medium supplementd with increasing concentrations of $FeSO_4$, $MgCl_2$ or a combination of $FeSO_4$ with $MgCl_2$.

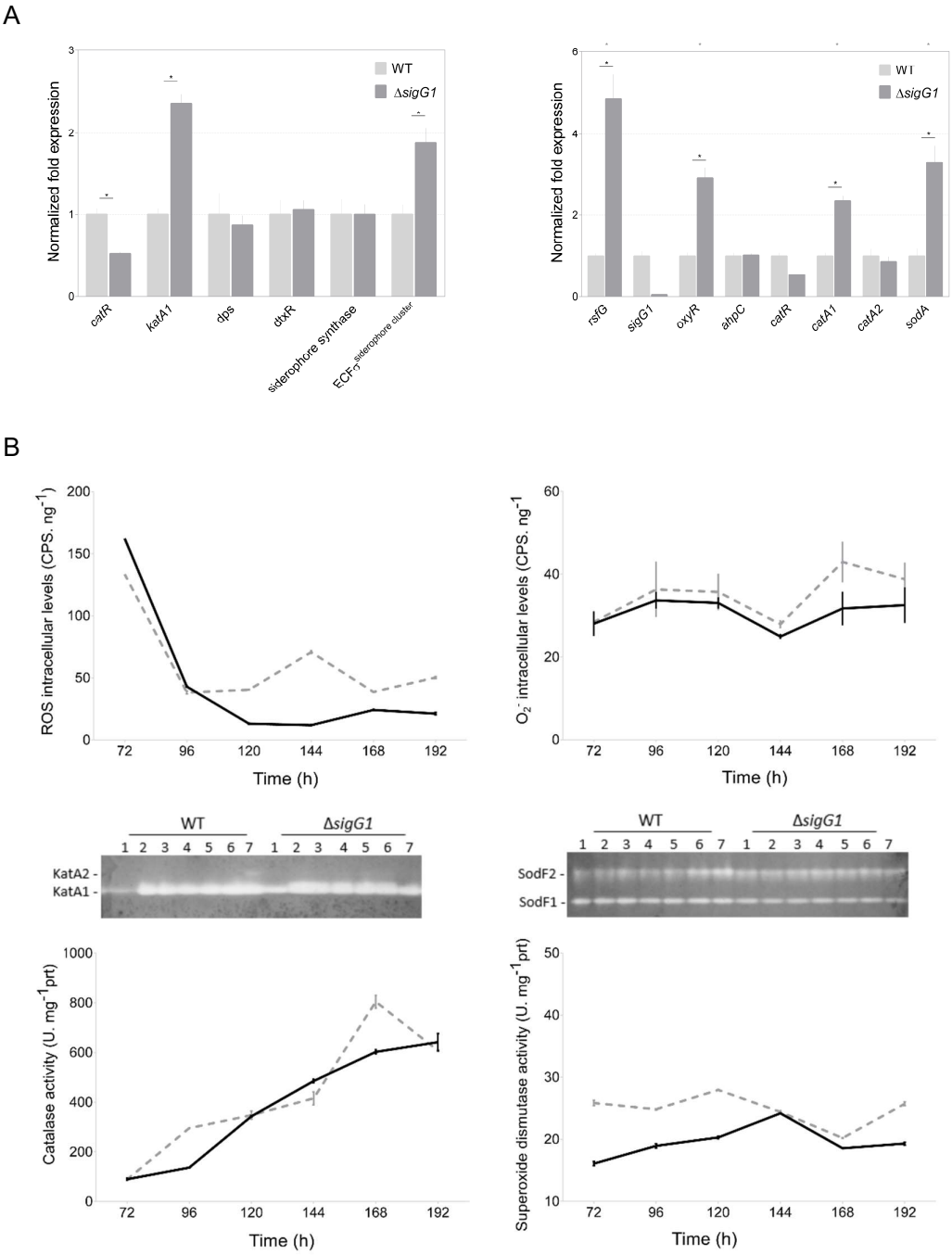


Figure SII9. *S. tsukubaensis* oxidative stress response mediated by *sigG1.* Gene expression of genes involved in the antioxidant defence and iron metabolism in Δ*sigG1,* as examined by RT-qPCR (**A**) Intracellular ROS levels of *S. tsukubaensis* strains measured via the oxidation of the Dihydroethidium (DHE: O_2^-) and the Dihydrorhodamine 123 (DHR: high affinity to H_2O_2) fluorescent probes. Specific activity of antioxidant enzymes were monitored spectrophotometrically throughout growth in MGm-2.5 liquid media. Black line: wild type. Dashed gray line: Δ*sigG1*. In-gel activity of catalase and superoxide dismutase enzymes throughout growth is depicted. (**B**) Values are representative of at least three independent experiments.

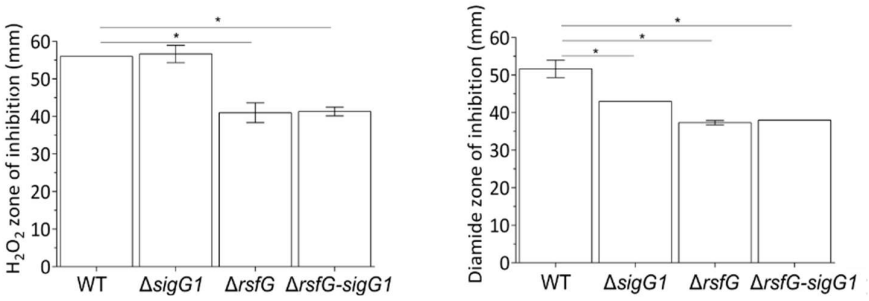


Figure SII10. Inhibition of growth by stress-inducing agents. Strains were grown in DNA agar in the presence of exogenous H_2O_2 and thiol stress by diamide. Results are the mean of three independent experiments.

phosphorylation, indicating that the activation of redox defences is originated at the posttranscriptional level.

Cell oxidative stress responses are tightly linked to secondary metabolism in *S. tsukubaensis* (Pires *et al*, unpublished results), since its biosynthesis is exclusively triggered when there is an intracellular redox balance. Meausuring FK506 production in liquid FK506-producing cultures we observed that FK506 production is significantly reduced in Δ*sigG1*. GSEA preranked analyses show that this phenotype was associated to a downregulation of the genes within FK506 biosynthetic cluster in both Δ*sigG1* and Δ*rsfG* (Figure SII11A). secondary metabolism deffects have been associated to rearragenments that occur in *Streptomyces* chromosomes, often to protect colony-wide fitness by discarding entire heavy burden biosynthetic clusters (Zhang et al., 2020). Further analyses showed of that in fact FK506 production is impaired in all *sigG1* and *rsfG* deletion mutants, indicating that this was a specific effect of disrupting the *sigG1*/*rsfG* system, rather than a polar effect of Δ*sigG1* genomic rearrangement. (Figure SII11B). A closer look at the HPLC spectra obtained, we found that not only FK506 was less abundant, but also other metabolites extracted with this method were absent or produced in very low amounts in the *sigG1*-deletion strain indicating that the overall secondary metabolism is affected in this mutant (Figure SII11C).

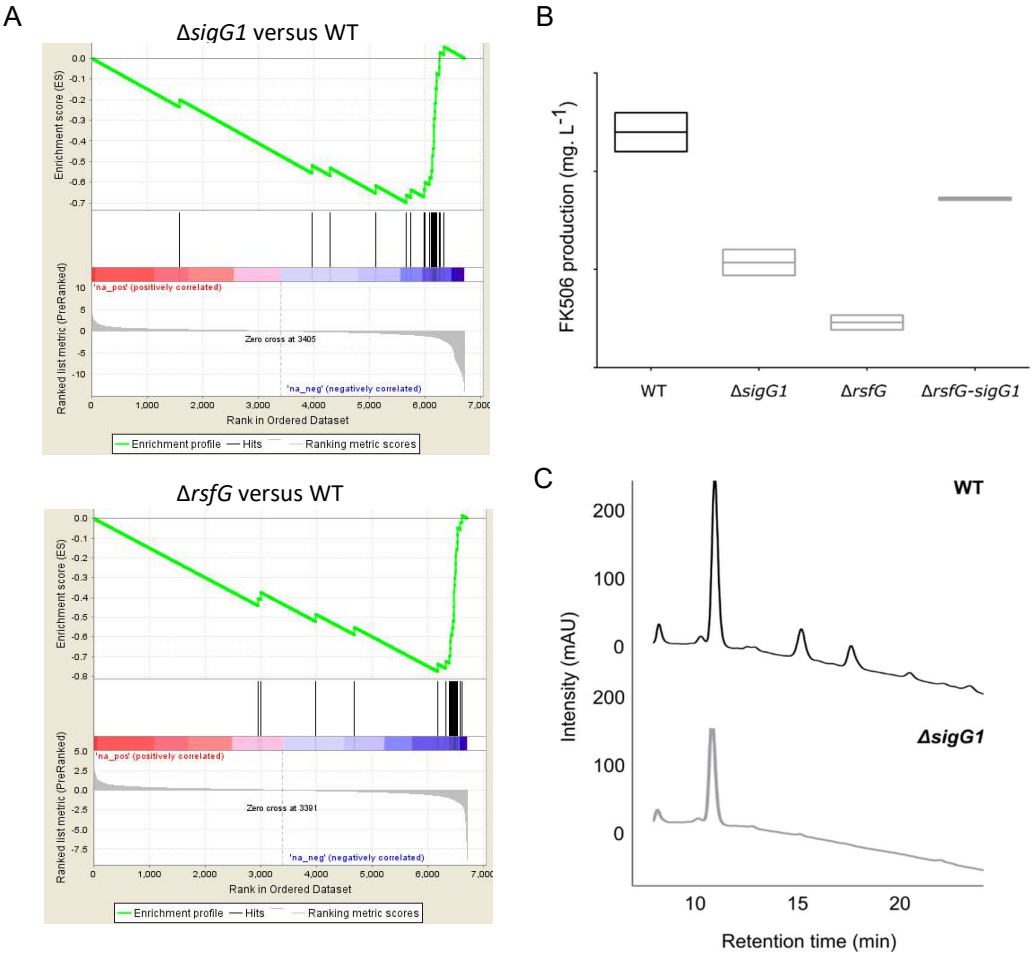


Figure SII11. FK506 production. Gene set enrichment analyses shows a downregulation of FK506 biosynthesis molecular signature in Δ*sigG1* and Δ*rsfG*. Expression of genes within the biosynthetic cluster of FK506 was examined against the whole transcriptome, pre-ranked according to the fold change by the GSEA Prerank platform (Broad Institute). Normalized Enrichment Score (NES): -4.31 (Δ*sigG1*), -4.82 (Δ*rsfG)*. Nominal p-value: 0.0 (**A**) Total FK506 was measured by HPLC at 192h. Values are representative of at least three independent experiments (**B**).

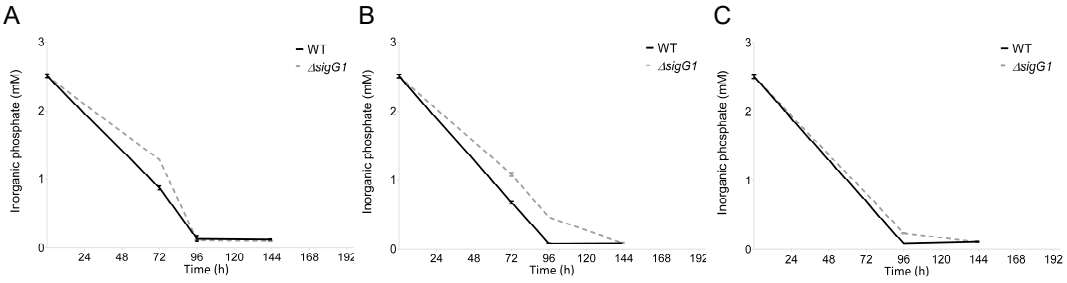


Figure SII12. Inorganic phosphate consumption (Pi). Pi levels were measured in samples taken throughout growth in phosphate-depleted medium (MGm-2.5) (**A**). The pattern of Pi consumption was impaired in the deletion mutant despite the supplementation of the MGm-2.5 medium with H_2O_2 (**B**) or the culture in iron depleted medium (**C**). WT: balck line; Δ*sigG1*: dashed gray line. Results are representative of three independent experiments.

As extensively characterised by our lab and others, secondary metabolism in *Streptomyces* and in particular in *S. tsukubaensis* (Martinez-Castro et al., 2013), is intimately related to phosphate consumption in the culture. We have measured Pi depletion in standard and stress conditions in our cultures, and found that phosphate is consumed at a slower rate in Δ*sigG1,* when compared to the wild type (Figure SII12A). As characterised before, phosphate depletion occurs between 80h and 90h, just before FK506 was detected (from 96h onwards). This result is thus in concordance with the reduced secondary metabolism observed for this mutant, and in particular with the lower FK506 yields detected at the end of the fermentation. We have previously proposed that oxidative stress impairs phosphate uptake and secondary metabolism in *Streptomyces* (Beites et al., 2014), and this could be one explanation for the Pi consumption pattern observed in Δ*sigG1* cultures. As observed in the graphs, the delay in Pi consumption was kept despite the depletion of iron in the culture medium or the addition of exogenous H_2O_2 (Figure SII12B and Figure SII12C).

In order to characterise *sigG1* expression and define the best time point to test the effect of *sigG1* deletion we assessed transcript levels by RT-qPCR throughout growth in liquid cultures. We examined transcription at 72h (early exponential phase, EEP) 96h (late exponential phase, LEP) and 120h (stationary phase, SP). Results show that *sigG1* activation occurs in early exponential growth (EEP). (Figure SII13). In later stages of growth *sigG1* transcription decreases and it is detected at the lowest levels in stationary phase (SP). Moreover as previously shown (see main text), *sigG1* levels are significantly more abundant in the absence of *oxyR* or *rsfG*.

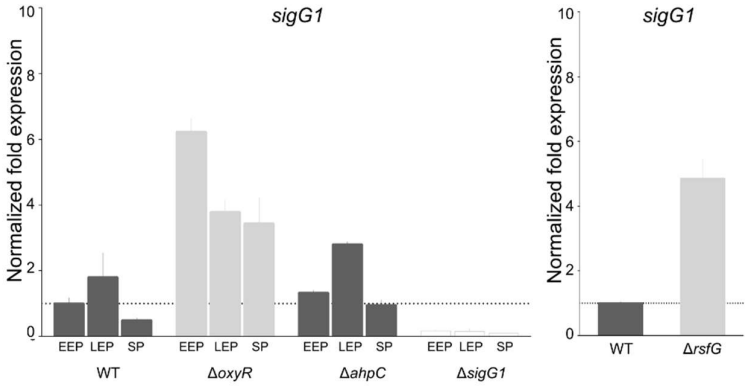


Figure SII13. Analysis of sigG1 expression throughout growth. We examined the transcript levels of *sigG1* by RT-qPCR, in *S. tsukubaensis* cells of different backgrounds, grown in MGm-2.5 liquid cultures. Results shown are representative of three independent biological replicates.

To get a more through overview of the effect of deleting the *sigG1*/*rsfG* system *in S. tsukubaensis*, we performed a gene enrichment analysis of the cluster of orthologous

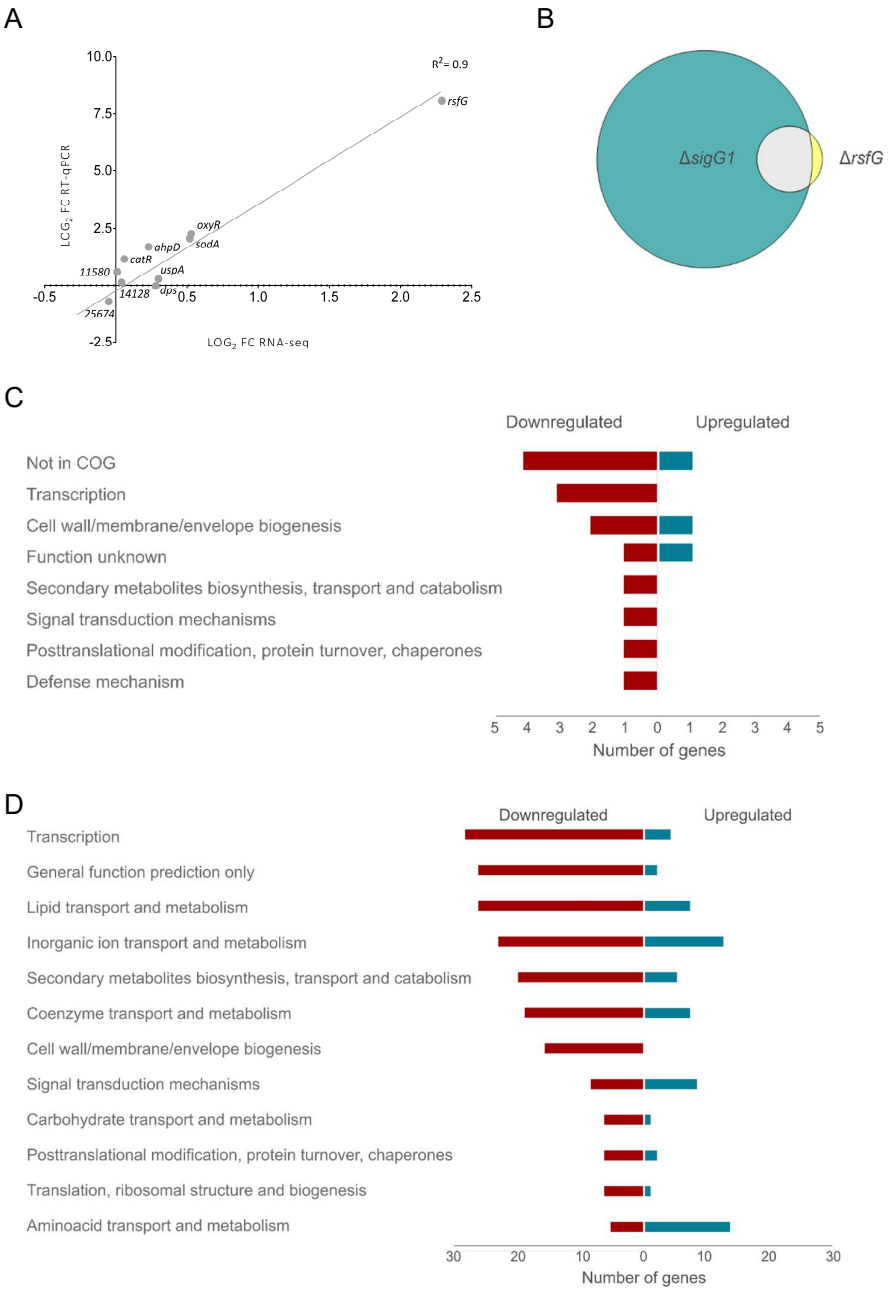
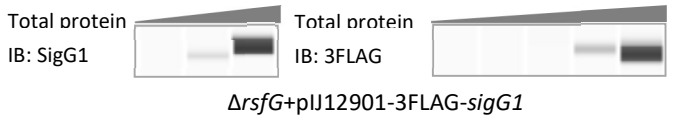


Figure SII4. Whole-transcriptome analyses upon *sigG1* and *rsfG* deletions. Validation of genes differently regulated in *sigG1* RNA-seq data by RT-qPCR (**A**). Venn diagram of genes differently regulated by 2-fold enrichment absolute value, with FDR p-values < 0.05 in Δ*sigG1* (577) and Δ*rsfG* (18) (**B**). Gene Enrichmenet Analysis of Differentially Expressed Genes (DEGs) in Δ*sigG1* (**C**) *and* Δ*rsfG* (**D**) according to in the COG annotation (Tatusov et al., 2000). COG terms are depicted according to the frequency of existance in the total deregulated trasncriptome. Differentially expressed genes concerning upregulated are represented by red bars and downregulated genes are shown in blue bars. X-axis, number of DEGs within each category. Y-axis, major functional COG categories. Genes within the 228nt Δ*sigG1* genomic rearrangement region were left out of this analysis). Poorly classified genes are not represented in Δ*sigG1* graphs.

Although we cannot exclude that Δ*sigG1* genomic rearrangements have influenced the overall gene expression, we found a significant overlap between DEGs in this strain when compared to Δ*rsfG,* which showed that the roles of both genes are relevant in the same biological processes (Figure SII14B). Several COGs were well represented in Δ*sigG1* like transcription, inorganic ion transport and metabolism, lipid transport and metabolism, inorganic ion metabolism or secondary metabolism biosynthesis, transport and catabolism (Figure SII14C; and two COG with at least three representatives in Δ*rsfG* in the transcription and cell wall/membrane/envelope biogenesis categories (Figure SII14D).

Production of a FLAG-tagged SigG1 protein for immunopurification. In order to define the SigG1 genome binding profile by ChIP-sequencing we prepared a SigG1 overexpression strain that lacked RsfG, to ensure the full availability of SigG1. We fused the coding sequence of a 3xFLAG tag to the N-terminal of SigG1 coding sequence and cloned it downstream of the SigG1 native promoter, in the replicative plasmid pIJ12901 (Δ*rsfG*+pIJ12901-3FLAG-*sigG1*), to produce a 3FLAG-SigG1 protein. We then assessed the expression of the full-length SigG1 in cell lysates of exponentially growing *S. tsukubaensis* cells, using both an anti-SigG1 and the MII anti-FLAG antibody (Figure SII15A). The wild type strain expressing non-FLAG-tagged SigG1 was used as a control for non-specific binding of the antibodies. Unexpectedly, this experiment produced a truncated version of SigG1 of about 15kDa, which was detected with both antibodies used.

A



B

DYKDHDGDYK DHDIDYKDDD DKMSDQGGGD STQTAAPVTA PGGAPEPASR

σ2

PGRSGLGEPS AETIRTGGEE VFAALAERYR HELRVHLYRM LGSFTDAEDL

VQETLLKAWR RRETFEGRAG FRAWLYRIAT NTALD

Figure SII15. Generation of a FLAG-tagged version of SigG1 in *S. tsukubaensis*. Automated Western blot analyses of the N-terminally FLAG-tagged SigG1, expressed from the native promoter in cell lysates of the *rsfG* mutant background. Increasing amounts of total protein (up to 40 μM) were tested. SigG1 was detected using the anti-SigG1^{A16V39} antibody. The anti-FLAG MII antibody was used to detect the 3xFLAG tag. Primary antibodies were detected the using the quantitative 'Wes' capillary electrophoresis and blotting system (ProteinSimple) (**A**) Putative amino acid sequence of the resulting FLAG tagged protein. The 3xFLAG tag preceding the N-terminal region of SigG1 is highlighted in gray (**B**).

the first 135 aa of 3FLAG-SigG1. This indicated that the overexpression of the fusion protein was unsuccessful, as it produced an incomplete version of this protein that appears to comprise the FLAG-tag fused to SigG1 σ2 domain (Figure SII15B). One possibility is that the excess of SigG1 produced by a replicative plasmid might be detrimental for cells, which potentiates the degradation of the protein as a defence mechanism.

Production of anti-SigG1 IgY antibodies in quail egg yolks. The recombinant histidine tagged full-length version of SigG1 (6H-SigG1) was expressed in *E. coli* and purified by immunoaffinity chromatography (IMAC). Despite the poor solubility of 6H-SigG1, we obtained a sufficient yield of pure protein after elution with 120mM imidazole. The purified protein was used for quail immunization for the production of IgY antibodies in quail egg yolks by HenBiotech (Figure SII16A).

Due to high level of conservation of amino acid sequences among ECF sigma factors, and the possibility of cross-reactions with other sigma factors in protein extracts, antibodies against specific epitopes of SigG1 were raised. We have carried out sequence alignments of *S. tsukubaensis* ECFs amino acid sequences to define regions that were less conserved among the four SnoaL2-containing ECFs previously identified (see main text in this Chapter). We selected epitopes V16-A39 (anti-SigG1^{A16V39}), E123-V142 (anti-SigG1^{E123V142}) and T353-E369 (anti-SigG1^{T353E369}) to serve as antigens in anti-SigG1 antibodies production (Figure SII16B). Structure homology modelling of SigG1 predicted ternary structure confirmed the availability of the selected epitopes for detection (Figure SII16C). Purified recombinant peptides were used in quail immunization, for the production of antibodies in quail egg yolks by HenBiotech.

The level of affinity of the newly generated antibodies to the SigG1 protein was evaluated by enzyme-linked immunosorbent assays (ELISA) using the purified 6H-SigG1 recombinant protein. Results revealed that all four antibodies were able to detect 6H-SigG1, and that the antibody with the highest affinity was the anti-SigG1^{A16V39} (Figure SII17A). The other three antibodies were less effective in detecting SigG1.

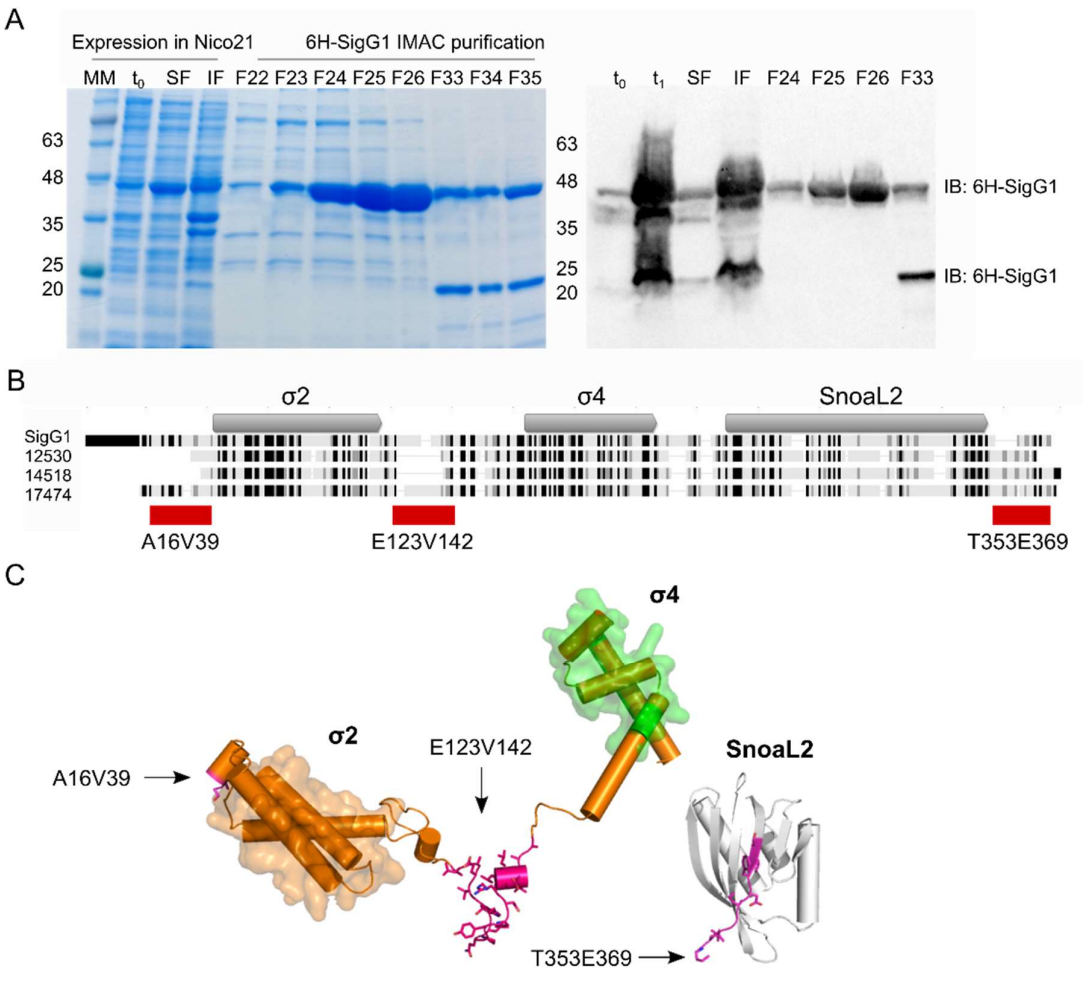
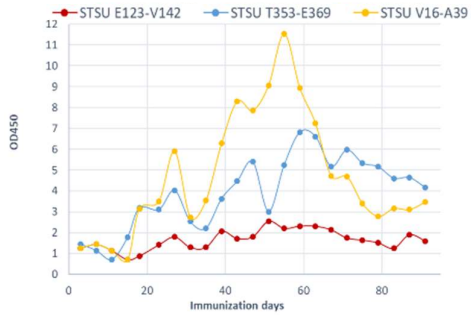


Figure SII16. Preparation of SigG1 peptides to serve as antigens in anti-SigG1 antibodies production. Production of the 6H-SigG1 recombinant protein to serve as antigen for the generation of the anti-SigG1 full-length (FL) antibody. SigG1 was detected by Western blot using a mouse anti-histidine antibody. The rabbit anti-mouse IgG HRP-conjugated was used as secondary antibody. The identity of SigG was further confirmed by peptide-mass fingerprinting (PMF). Results revealed a full length protein of about 40kDa and a second species that migrated near the 25kDa region of the gel, which was also identified as SigG1, the protein with a smaller mass is identified by the anti-histidine WB and could correspond to a truncated version of SigG1 that lacked the SnoaL2 domain (**A**). Sequence alignment of SnoaL2 containing sigma-factors in *S. tsukubaensis.* Less conserved regions among the four SnoaL2 ECFs are highlighted in red. Recombinant peptides of about 20 residues within these regions were used as antigens in quail immunization. Antibody A: anti-SigG1^{A16V39}; Antibody B: anti-SigG1^{E123V142}; Antibody C: Anti-SigG1^{T353E369} (**B**). Structure homology modelling of SigG1 predicted ternary structure to confirm the availability of the selected epitopes for detection (**C**).

A



B

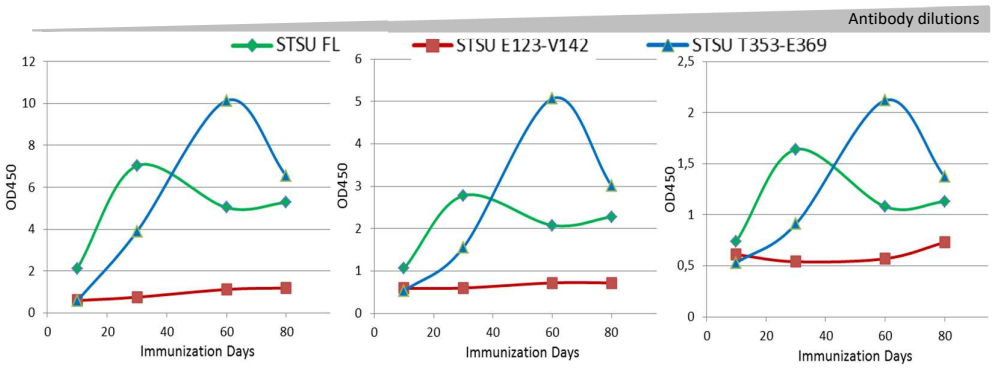


Figure SII17. Analysis of Anti-SigG1 antibody titer by ELISA. Plates coated with SigG1 recombinant antigens at 0.5µg per well serial dilutions were performed with anti-SigG1 and rabbit anti-chicken IgY HRP-conjugated at 1:10000 was used as secondary antibody. The antibody raised against the A16V39 epitope shows the highest affinity (**A**). The three remaininng antibodies were less effective (**B**). SigG1 is detected with a relative affinity as follows: anti-SigG1^{A16V39} > anti-SigG1^{T353E369} > anti-SigG1 FL > anti-SigG1^{E123V142}. Anti-SigG1 titer was estimated to be above 1:5000. Anti-SigG1 FL: 28-31 days of immunization; anti-SigG1^{A16V39}: 60-63 days of immunization; anti-SigG1^{E123V142}: 60-63 days of immunization; anti-SigG1^{T353E369}: 60-63 days of immunization.

The comparison among the affinities is as follows: anti-SigG1^{A16V39} > anti-SigG1^{T353E369} > anti-SigG1 FL > anti-SigG1^{E123V142} (Figure SII17B).

Detection of SigG1 with quail polyclonal antibodies. To test the effectivenes of the generated antibodies in recognising SigG1 we carried out a dot blot assay with increasing concentrations of the recombinant 6H-SigG1 purified protein (Figure SII18). We labeled the protein with each primary antibody at 1:200 or 1:500 by incubating for 1hour at room temperature. Detection was achieve with an anti-chicken antibody at 1:10000, after 1h incubation at room temperature. The four SigG1-specific antibodies showed high affinity for SigG1, since they were able to detect 6H-SigG1 at very low ammounts (>5ng of purified protein).

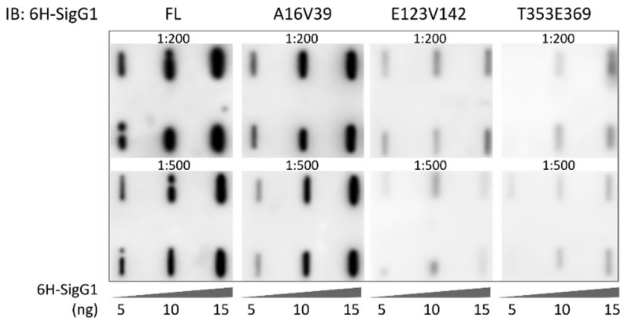


Figure SII18. Detection of the recombinant 6H-SigG1 purified from *E. coli* by Western blot. WB detection using anti-SigG1 polyclonal antibodies raised in quail egg yolks. SigG1 was detected using the antibodies at the dilutions of 1:200 and 1:500. Anti-chicken IgY HRP-conjugated at 1:10000 was used as a secondary antibody. The newly generated antibodies detect SigG to the lowest amount of 5ng. FL: antibody raised against the 6H-SigG1 full-length protein. A16V39: anti-SigG1^{A16V39}; E123V142: anti-SigG1^{E123V142}; T353E369: anti-SigG1^{T353E369}.

We further addressed the sensitivity of the quail antibodies to the native SigG1 in *S. tsukubaensis* total cell lysates. Previous transcriptomics analyses (by RNA-seq and RT-qPCR, see Chapter 2 main text) indicated that SigG1 was expressed in low levels in the liquid culture conditions used. Western blotting using exponentially growing cells (when SigG1 transcription throughout growth peaked) corroborated the low abundance of SigG1 in these conditions. Nevertheless, and despite the presence of non-specific binding. SigG1 was successfully detected in the Δ*sigG1* strain complemented with *sigG1* allele (which was shown to result in an overexpression of *sigG1* transcripts) (Figure SII19, lane O). Absence of these bands in the *sigG1* deletion background confirmed that

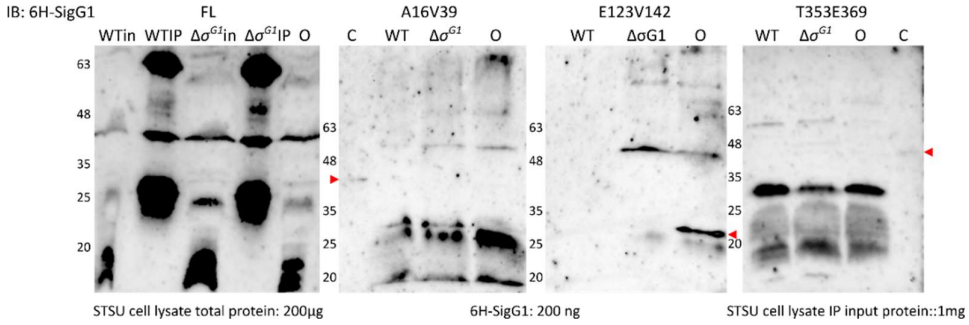


Figure SII19. Western blot analysis of SigG1 expression in *S. tsukubaensis* cell lysates. WB detection using anti-SigG1 polyclonal antibodies raised in quail egg yolks. SigG1 was detected using the antibodies ON at the dilution of 1:1000. Anti-chicken IgY HRP-conjugated at 1:10000 was used as a secondary antibody. WTin: IP input wild type; WTIP: SigG1-enriched wild type; Δσ^{G1}in: IP input Δ*sigG1*; Δσ^{G1}IP: SigG1-enriched Δ*sigG1*; O: SigG1 transcript overexpression obtained with Δσ^{G1}Δ*sigG1*+pIJ12333-*sigG1* complementation; C: recombinant 6H-SigG1 purified from *E. coli.* A red arrow indicates SigG1.

this was the SigG1-specific signal. Moreover, in an effort to increase the SigG1 signal we performed a previous immunopurification (IP) experiment, where we conducted a purification of SigG1 in the total extracts using the anti-SigG1 FL antibody, to enrich the cell lysates for the native SigG1 protein. These experiments were inconclusive. Using the high sensitive automated western blot system provided by Protein Simple (Wes) we were able to confirm detection of SigG1 with the selected antibody (see main text in this Chapter, Figure S2), which led us to select the SigG1^{A16V39} antibody to use in the subsequent chromatin pull-down experiments, despite the affinity to SigG1 observed with the anti-SigG1 FL antibody.

Genome instability in *S. tsukubaensis* Δ*sigG1* mutant. The NGS data from the ChIP-seq experiments unveiled an unexpected genomic rearrangement in the Δ*sigG1* mutant. Genomic instability resulting in excission of large fragments of DNA has previously been described in *Streptomyces* and other bacteria (Birch et al., 1990; Cullum et al., 1986; Leblond and Decaris, 1994; Volff and Altenbuchner, 1998; Zhang et al., 2020). Southern blot hybridization using this region of the chromosome confirmed the a 200 Kb deletion in terminal region of the Δ*sigG1* chromosome when compared to the wild type (Figure SII20). Through NGS re-sequencing of the full genome of this strain (Novogene, HK) we defined the exact sequence within the deleted region of the chromosome that goes from position 349601 bp to 577663 bp. Thorough examination of the deleted region revealed an array of proteins that were not expressed in this strains (Table SII3). We further analysed this seaquence using the antismash 2.0 platform, which indicated that Δ*sigG1* lacks the sequence corresponding to the T1-PKS biosynthtetic cluster responsible for the production of the bafilomycin antibiotic (77% identity).

Despite the phenotypic rescue achieved after complementation with *sigG1* wild type allele, we asked if the above-mentioned phenotype was exclusively due to the loss of *sigG* and independent of the genome rearrangement. Making use of the CRISPR technology we developed for *S. tsukubaensis* (Albuquerque et al. personal communication), we generated a new null mutant for *sigG1*, by deleting the *sigG1* coding sequence to generate Δ*sigG1_CR*. The new Δ*sigG1_CR* mutant exhibit a similar morphological phenotype of the Δ*sigG1* generated with the REDIRECT™ technology (Figure SII21), which indicate that deleting *sigG1* using a different procedure resulted in a delay in morphological differentiation in Δ*sigG1_CR*. Genomic rearrangements in the Δ*sigG1* strain could account for the severity of the developmental phenotype observed in this mutant.

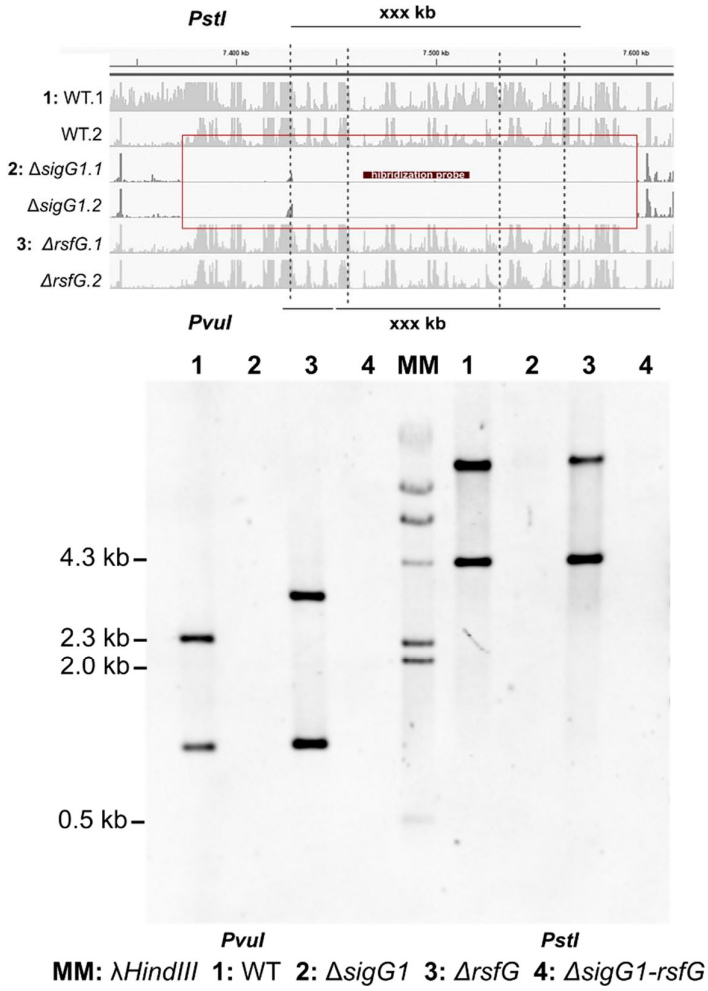


Figure SII20. Southern blot hybridization of *S. tsukubaensis* genomes to assess genomic rearrangement. The polar deletion of 200kb region was detected by Southern blotting hybridisation of *S. tsukubaensis* Δ*sigG1*, Δ*rsfG* and Δ*sigG1-rsfG1* chromosomal DNA digested with *PvuI* or *PstI* using an internal fragment as probe. A 1:1 mixture of λ*HindIII* with *DIG-labelled DNA Molecular Molecular Weight Marker II* (Roche) was used as molecular ladder (MM). The red bar represents the internal fragment used as probe in the Southern blot hybridisation.

Table SII3. List of proteins within the deleted region in the *S. tsukubaensis* Δ*sigG1* genome

Locus ID	Length (bp)	Product
STSU_01120	1302	hypothetical protein
STSU_01125	957	hypothetical protein
STSU_01130	1602	peptidase S8 and S53 subtilisin kexin sedolisin
STSU_01135	162	glycerophosphoryl diester phosphodiesterase
STSU_01140	2381	amino acid adenylation protein
STSU_01145	1816	amino acid adenylation protein

STSU_01150	795	hypothetical protein
STSU_01155	918	aminoglycoside phosphotransferase
STSU_01160	234	TetR family transcriptional regulator
STSU_01165	399	hydrolase
STSU_01170	690	hypothetical protein
STSU_01175	1242	hypothetical protein
STSU_01180	567	hypothetical protein
STSU_01185	921	heat shock protein HtpX
STSU_01190	1554	magnesium or manganese-dependent protein phosphatase
STSU_01195	633	hypothetical protein
STSU_01200	2706	metallophosphoesterase
STSU_01205	804	metallophosphoesterase
STSU_01210	102	hypothetical protein
STSU_01215	549	hypothetical protein
STSU_01220	201	hypothetical protein
STSU_01225	252	hypothetical protein
STSU_01230	801	hydroxylase
STSU_01235	1236	putative integral membrane protein
STSU_01240	1047	hypothetical protein
STSU_01245	363	hypothetical protein
STSU_01250	432	hypothetical protein
STSU_01255	330	peptidase C60 sortase A and B
STSU_01260	669	pentapeptide repeat-containing protein
STSU_01265	170	thioredoxin reductase
STSU_01270	121	FAD-dependent pyridine nucleotide-disulfide oxidoreductase
STSU_01275	1494	oxidoreductase FAD-binding protein
STSU_01280	1434	glycerol kinase
STSU_01285	777	DeoR family transcriptional regulator
STSU_01290	1221	hypothetical protein
STSU_01295	888	hypothetical protein
STSU_01300	470	TetR family transcriptional regulator
STSU_01305	3207	monooxygenase P450
STSU_01310	1506	amino acid transporter
STSU_01315	717	GntR family transcriptional regulator
STSU_01320	1206	N-isopropylammelide isopropylaminohydrolase
STSU_01325	1347	hypothetical protein
STSU_01330	129	G-D-S-L family lipolytic protein
STSU_01335	1284	hypothetical protein
STSU_01340	156	hypothetical protein
STSU_01345	603	isochorismatase hydrolase

STSU_01350	915	AraC family transcriptional regulator
STSU_01355	435	hypothetical protein
STSU_01360	282	hypothetical protein
STSU_01365	636	hypothetical protein
STSU_01370	615	hypothetical protein
STSU_01375	1326	glutamate-1-semialdehyde 2.1-aminomutase
STSU_01380	552	dTDP-4-dehydrorhamnose 3.5-epimerase
STSU_01385	1248	hypothetical protein
STSU_01390	957	glycosyl transferase family protein
STSU_01395	1287	putative integral membrane protein
STSU_01400	1341	hypothetical protein
STSU_01405	1026	nucleotide sugar epimerase/dehydratase
STSU_01410	648	hypothetical protein
STSU_01415	807	transferase
STSU_01420	1236	putative methyltransferase
STSU_01425	1275	transferase
STSU_01430	1302	putative integral membrane protein
STSU_01435	2028	heparinase II/III family protein
STSU_01440	2214	putative bi-domain oxidoreductase
STSU_01445	1929	asparagine synthetase
STSU_01450	1572	putative membrane-bound polysaccharide biosynthesis protein
STSU_01455	1404	lipopolysaccharide biosynthesis protein
STSU_01460	1293	hypothetical protein
STSU_01465	1317	GDP-mannose 6-dehydrogenase
STSU_01470	1404	glycosyl transferase family protein
STSU_01475	198	hypothetical protein
STSU_01480	2244	integral membrane protein
STSU_01485	771	TetR family transcriptional regulator
STSU_01490	543	deaminase-reductase domain-containing protein
STSU_01495	912	chitinase
STSU_01500	273	TetR family transcriptional regulator
STSU_01505	1047	hypothetical protein
STSU_01510	1068	hypothetical protein
STSU_01515	621	TetR family transcriptional regulator
STSU_01520	582	peptidase
STSU_01525	216	hypothetical protein
STSU_01530	285	hypothetical protein
STSU_01535	2433	penicillin amidase
STSU_01540	867	hypothetical protein
STSU_01545	4902	DNA-binding protein

STSU_01550	1161	hypothetical protein
STSU_01555	704	hypothetical protein
STSU_01560	527	hypothetical protein
STSU_01565	300	hypothetical protein
STSU_01570	235	hypothetical protein
STSU_01575	1852	hypothetical protein
STSU_01580	1863	hypothetical protein
STSU_01585	1137	hypothetical protein
STSU_01590	639	esterase
STSU_01595	699	TetR family transcriptional regulator
STSU_01600	942	dehydrogenase
STSU_01605	627	isochorismatase
STSU_01610	477	MarR family transcriptional regulator
STSU_01615	978	AsnC family transcriptional regulator
STSU_01620	330	endoribonuclease L-PSP
STSU_01625	861	alpha/beta hydrolase fold protein
STSU_01630	1164	alpha-hydroxyacid dehydrogenase. FMN-dependent L-lactate dehydrogenase
STSU_01635	1272	cytochrome P450
STSU_01640	2103	modular polyketide synthase
STSU_01645	420	cobalamin B12-binding domain-containing protein
STSU_01650	1587	methylmalonyl-CoA mutase large subunit
STSU_01655	933	3-oxoacyl-ACP synthase
STSU_01660	7134	putative type-I PKS
STSU_01665	585	N-acetyltransferase GCN5
STSU_01670	384	hypothetical protein
STSU_01675	966	pirin-like protein
STSU_01680	336	hypothetical protein
STSU_01685	1164	hypothetical protein
STSU_01690	1182	Methicillin resistance protein
STSU_01695	1388	hypothetical protein
STSU_01700	684	putative two-component system response regulator
STSU_01705	1449	putative two-component system sensor kinase
STSU_01710	549	hypothetical protein
STSU_01715	1227	peptidase S8 and S53 subtilisin kexin sedolisin
STSU_01720	1578	oxidoreductase
STSU_01725	723	trypsin-like protease
STSU_01730	1167	peptidase S8 and S53 subtilisin kexin sedolisin
STSU_01735	273	hypothetical protein
STSU_01740	1002	D-alanyl-D-alanine carboxypeptidase

STSU_01745	1251	two-component system sensor kinase
STSU_01750	768	two-component system response regulator
STSU_01755	981	hypothetical protein
STSU_01760	528	glucose-methanol-choline oxidoreductase
STSU_01765	1092	glucose-methanol-choline oxidoreductase
STSU_01770	1422	Xaa-Pro aminopeptidase II
STSU_01775	984	transcriptional regulator
STSU_01780	714	short-chain dehydrogenase/reductase SDR
STSU_01785	696	TetR family transcriptional regulator
STSU_01790	2319	5-methyltetrahydropteroyltriglutamate-homocysteineS-methyltransferase
STSU_01795	495	hypothetical protein
STSU_01800	582	TetR family transcriptional regulator
STSU_01805	1087	peptidyl-arginine deiminase
STSU_01810	801	hypothetical protein
STSU_01815	981	AraC family transcriptional regulator
STSU_01820	750	short-chain dehydrogenase/reductase SDR
STSU_01825	624	hypothetical protein
STSU_01830	1356	O-acetylhomoserine aminocarboxypropyltransferase
STSU_01835	861	homoserine O-acetyltransferase
STSU_01840	507	hypothetical protein
STSU_01845	507	hypothetical protein
STSU_01850	1008	oxidoreductase
STSU_01855	1167	glycosyl transferase
STSU_01860	804	hypothetical protein
STSU_01865	241	hypothetical protein
STSU_01870	1047	glycosyl transferase
STSU_01875	990	prenyltransferase
STSU_01880	1044	NAD-dependent epimerase/dehydratase
STSU_01885	1446	aminotransferase
STSU_01890	1047	hypothetical protein
STSU_01895	996	sugar transferase
STSU_01900	369	hypothetical protein
STSU_01905	926	LuxR family transcriptional regulator
STSU_01910	1722	hypothetical protein
STSU_01915	195	hypothetical protein
STSU_01920	639	TetR family transcriptional regulator
STSU_01925	1401	hypothetical protein
STSU_01930	1221	oxidoreductase
STSU_01935	960	carbon-nitrogen family hydrolase

STSU_01940	882	hypothetical protein
STSU_01945	1467	cytosine/purine. uracil. thiamine. allantoin family permease
STSU_01950	1089	agmatinase
STSU_01955	585	transcriptional regulator
STSU_01960	714	hypothetical protein
STSU_01965	651	two-component system response regulator
STSU_01970	1167	two-component system sensor kinase
STSU_01975	942	ABC transporter ATP-binding protein
STSU_01980	618	putative ABC transporter permease
STSU_01985	1083	hypothetical protein
STSU_01990	963	acyl-carrier-protein S-malonyltransferase
STSU_01995	366	LuxR family transcriptional regulator
STSU_02000	768	transcriptional regulator
STSU_02005	1113	transcriptional regulator
STSU_02010	663	methoxymalonate biosynthesis protein
STSU_02015	1086	methoxymalonate biosynthesis protein
STSU_02020	1116	methoxymalonate biosynthesis protein
STSU_02025	282	methoxymalonate biosynthesis protein
STSU_02030	939	methoxymalonate biosynthesis protein
STSU_02035	432	modular polyketide synthase BFAS5

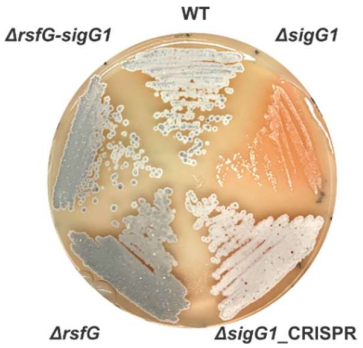


Figure SI21. Phenotypes of *S. tsukubaensis* wild type, *sigG1* and *rsfG null* mutants grown on solid ISP4. The phenotypes of the Δ*sigG1* mutant generated by Redirect™ (Δ*sigG1*) and CRISPR (Δ*sigG1_CR*) are depicted, 10^6 spores were streaked on ISP4 and grown for 6 days.

Materials and Methods

Bacterial strains, growth conditions and preparation of *Streptomyces* Spores. *S. tsukubaensis* strains were generally grown in MGm-2.5 liquid medium or on ISP4 solid agar, unless otherwise stated. Spores suspensions were prepared from single colonies in 20% glycerol, 0.20% Triton X-100. For the detailed procedures, see the Materials and Methods section in Chapter 2.

Construction of deletion mutant strains with the REDIRECT technology. *E. coli* BW25113 containing a λ RED plasmid, pIJ790, was used to replace the native alleles on the 15C1 cosmid with the apr-oriT cassette amplified from the pIJ773 template plasmid, using specific oligonucleotides XXX listed in Supplementary Table S2, to create mutations in the *sigG1* and *rsfG* loci. *E. coli* ET12567 containing pUZ8002 was used for λ RED-mediated recombineering (Gust 2002) upon conjugation experiments. Conjugations between *E. coli* and *S. tsukubaensis* were carried out as described in Kieser 2000. Deletions in the chromosome were confirmed by PCR analysis and by Southern hybridization. For complementation of the deletion mutant strains with the native alleles. *SigG1* or *rsfG* and its promoter region were amplified cut with SpeI/BamHI and cloned into the XbaI/NotI site of pIJ12333.1. The construct was introduced into *S. tsukubaensis* Δ*sigG1* by conjugation.

***Streptomyces* genomic DNA extraction and Polymerase chain reaction (PCR).** Genomic DNA was recurrently obtained with the Master™ pure Gram positive DNA purification kit (Epicentre) or with the protocol optimized for Gram-positive bacteria from the GeneJET Genomic DNA purification kit (Fermentas) according to the protocol provided by the manufacturers. The methods used for DNA amplification can be found in the Materials and Methods section in Chapter 2.

Southern blot hybridization. Four Southern analysis, genomic DNA was digested by restriction endonucleases as required and the resulting fragments were separated by gel electrophoresis. DIG labelled λphage DNA digested with *Hind*III (Roche Applied Science) was used as a ladder marker. DNA transfer from the agarose gel to a nylon membrane was carried out in a vacuum system (BIO-RAD) and following the manufacturer's instructions. DNA probes were labelled with digoxygenin using DIG-High prime kit (Roche Applied Science) and following the protocol provided by the manufacturer. Hybridisation and detection were performed using the alkaline phosphatase substrate CPD-Star kit (Roche).

RNA extraction, RT-qPCR analyses and RNA sequencing experiments were performed according to the methods described in see the Materials and Methods section in Chapter 2.

Catalase activity. The rate of degradation of H_2O_2 was determined spectrophotometrically at 240 nm as described by Beers and Sizer 1952. Assays were carried out at 25 ºC. Catalase activity was expressed in units per mg of total protein (U mg-1) in which one unit of enzyme activity is defined as the amount required for the conversion of 1 µmol substrate into product in 1 minute. Catalase activity was further monitored on 7.5% native-PAGE gels, using a specific negative staining (Clare et al, 1984). Gels were soaked into a solution of horseradish peroxidase, which catalyses the conversion of diaminobenzidine into a brown compound, using H_2O_2 as a co-factor. Clear areas of the gel indicate the catalyses of H_2O_2 by proteins with this specific activity.

Superoxide dismutase (SOD) activity. SOD mediated detoxification was assessed by measuring the rate of cytochrome C reduction by the superoxide anion as described in Beauchamp and Fridovich 1971. Total protein extracts were incubated with the generator of O2- radicals xanthine oxidase. The reduction of cytochrome C was measured at 550 nm and 25 ºC, 1 Unit of SOD activity corresponds to a 50 % decrease in the cytochrome c reduction rate in 1 min. The final SOD specific activity is indicated in SOD activity units per mg of protein. SOD activity was also monitored in 10% native-PAGE gels by a negative specific staining (Beauchamp and Fridovich 1971). The light induced reduction of riboflavin by TEMED was used to generate superoxide anion, which interacts with nitroblue tetrazolium (NBT) forming a blue precipitate. The absence of the blue precipitate in a particular gel area indicated the detoxification of the radical by proteins with SOD activity.

Fluorescence Measurement of Intracellular ROS. Intracellular H_2O_2 and O_2^- levels were detected with dihydrorhodamine 123 (DHR) and Dihydroethidium (DHE) (Invitrogen), respectively. DHR is a nonfluorescent molecule that upon oxidation by ROS is transformed into fluorescent rhodamine 123 (Royall and Ischiropoulos 1993). Dihydroethidium is a superoxide anion indicator that exhibits red fluorescence upon oxidation. Samples were collected throughout growth and cell pellets were suspended in solutions of DHR (15 µg. mL^{-1}) and DHE (5 µg. mL^{-1}) in 50 mM potassium phosphate buffer pH 6.8. Cells were incubated in the dark at 30 ˚C for 30 min (DHR) or 60 min (DHE). Cells were washed twice in 50 mM potassium phosphate buffer pH 6.8 and lysed by sonication. The amount of ROS was quantified with a spectrofluorometer (Fluoromax-4, Horiba) with excitation and emission at 504 and 534 nm (DHR) or at 355 nm and 420 nm (DHE). Total protein content of crude extracts was used as a normalisation factor.

Zone of growth inhibition assay. A suspension of 10^7 *S. tsukubaensis* spores was mixed with 5 ml pre-warmed soft agar and poured on Difco nutrient agar medium. Filter disks embedded in 15 µl of 1M Diamide or 15 µl 30% H_2O_2 were placed at the centre of

plates. The diameters of growth inhibition areas were measured after incubation at 30 °C for 2 days.

Inorganic phosphate depletion determination. For growth measurement and extracellular inorganic phosphate quantification, culture samples were harvested and centrifuged throughout growth in MGm-2.5 medium. The supernatants were collected and used in the malachite green assay (Lanzetta et al. 1979). The pellets were washed with 0.9% NaCl and dried at 80 °C for 48 h to monitor growth.

Secreted siderophores detection. Siderophore production was addressed using the Chrome Azurol S (CAS) assay according to …. Bagg and Neilands 1987 Schwyn and Neilands 1987??. The cultures were grown in MGm-2.5 liquid media supplemented with iron ferrous and ferric forms. Samples were harvested throughout growth and supernatants placed on CAS agar plates. The presence of orange halos after 48h of incubation at 30°C, indicated the production of hydroxamate siderophores.

FK506 quantification. FK506 production by *S. tsukubaensis* strains was quantified by HPLC. Briefly, 1 mL of culture was mixed with an equal volume of 100% methanol and incubated for 1h at 30 ˚C. The mixture was centrifuged for 10 min. The supernatant was recovered and analysed in a HPLC system (Merck-Hitachi) equipped with UV detector set at 210 nm, using a SunFireTM C18 column (4.6 x 150 mm. 3.5 µm; Waters) and the oven set at 55 ˚C. Elution was performed with a gradient mobile phase of 0.1% (v/v) trifluoroacetic acid (TFA) and 20% (v/v) Methyl tert-butyl ether (MTBE) in acetonitrile according to the following program: 40% 0-5 min, up to 80% 5-35 min, up to 90% 35-36 min, 90% 36-39 min down to 40% 39-40 min and 40% 40-43 min. Flow rate was maintained at 0.5 mL. min^{-1}. FK506 and FK520 chromatogram peaks were identified according to their Relative Retention Time (RRT) and using external standards (Sigma). RRT towards purified FK506 is 1 (RT/RTstd =1). In this case, RRT of FK520 was 0.96.

Statistical analyses

For each experiment, samples were assayed in at least three independent replicates. Statistical significance was addressed through the GraphPad software, according to the requirements of each data set.

Preparation of cell-free *S. tsukubaensis* protein extracts. Cells were suspended in lysis buffer containing 50mM potassium phosphate buffer (pH 6.8) supplemented with Complete EDTA-free protease inhibitor cocktail (Roche Applied Science) and lysed for 30 minutes on ice in a sonifier (3x10 seconds on/30 seconds off; 8 microns; 50% duty cycle). Lysates were centrifuged at 13000 rpm for 20 minutes at 4˚C to remove cell debris and the supernatant recovered. Protein concentration was determined using the BCA Protein Assay Reagent (Thermo Scientific). Bovine serum albumin was used to determine standard curves.

Western blot hybridisation. Total protein extracts were run on 10% SDS-PAGE and transferred to a nitrocellulose membrane Hybond^{TM}–C Extra (Amerhsam Pharmacia Biotech) using a Mini-PROTEAN Mini Trans-Blot® system (BIO-RAD). The membrane was blocked 16-18h at 4ºC with 5 % (w/v) dried milk in TPBS/TBST as required, rinsed twice in TPBS/TBST and incubated with appropriate primary antibodies. After incubation, the membrane was rinsed twice in TPBS/TBST and incubated with the secondary antibody. Signals were revealed with Prime Plus ECL detection kit (BIO-RAD). For detection of SigG1 we used the polyclonal antibodies raised in quail egg yolks. See figures legends for primary antibody labelling specifications. For detection of histidine-tagged proteins the tetra-his mouse antibody (Qiagen) was used. Detection of 3xFLAG-SigG1 was achieved with the mouse MII anti-FLAG antibody (F4725, Sigma). Proteins were visualised via labelling with an anti-chicken IgY (IgG) coupled to peroxidase (A9046, Sigma, 1:10000) or an anti-mouse secondary antibody conjugated to HRP (Santa Cruz Biotechnology, 1:10000).

Automated western blot (Wes). Analysis of the native SigG1 protein in crude *S. tsukubaensis* extracts was performed using the quantitative Wes capillary electrophoresis and blotting system (ProteinSimple, San Jose. CA) with the Wes No Secondary Detection (12 to 230 kDa) Master kit. Wes uses an automated, capillary-based method for immunodetection of proteins. Cell lysates were diluted to a final protein concentration of 4 mg/ml. Primary labelling was done with the anti-SigG1^{A16V39} (1:50) and the MII anti-FLAG (1:100). Detection was achieved using a rabbit anti-chicken IgY (IgG) coupled to peroxidase (A9046, Sigma, 1:100), or the Anti-Rabbit Detection Module (ProteinSimple, San Jose, CA). Data were analysed with the Compass Software**.**

Production and purification of recombinant 6H-SigG1 by IMAC. For the production of anti-SigG1 full-length polyclonal antibodies, we expressed and purified the recombinant 6H-SigG as previously described (REF). Fractions with the protein of the expected size were combined, dialysed against the binding buffer (20mM phosphate buffer pH 7.4, the 0.5M NaCl) and concentrated using an Amicon Ultra-15 30kDa (Milipore). Buffer was exchanged to 20 mM sodium phosphate, 300 mM NaCl. The total protein was used to induce immunization of quail for the production of the SigG1-specific IgY antibody in quail egg yolks at HenBiotech, Proteins & Antibodies (Coimbra, Portugal).

Generation of polyclonal IgY antibodies against SigG1-specific small epitopes. Sequence alignments of the SnoaL2-containing ECFs were performed using the Muscle algorithm (REF). Structure homology predictions were carried out using the Phyre2 software. Expression, purification and subsequent production of IgY the polyclonal antibodies against the selected epitopes was done at HenBiotech, Proteins & Antibodies (Coimbra, Portugal).

Construction of a markerless Δ*sigG1_CR* mutant. The markerless *sigG1* mutant was generated using pCRISPR-TT vector, a derivative of pCRISPR-Cas9 (Tong et al., 2015), where the *tipA* promoter controlling *cas9* expression was replaced for the constitutive *ermE*p1 promoter, and the theophylline riboswitch (Rudolph et al., 2013). The 2.3 kb *Xba*I-*Pae*I fragment containing the IS1 element and the thiostrepton resistance gene was replaced for a 103bp *Stu*I cloning cassette (Albuquerque et al., personal communication). Design of single guide RNAs targeting the *sigG1* gene was performed using CRISPy-web (Blin et al., 2016). CRISPR plasmids assembly was performed according to (Tong et al., 2015) with minor modifications. In order to assemble the sgRNAs into pCRISPR-TT, PCR amplifications were carried out using forward primers containing the selected sgRNA (sgRNA sigG1 V2, Table S2) and a universal reverse primer (sgRNA-R) (Table S2), using pCRISPR-TT as DNA template. Amplified PCR fragments were cloned into the *SnaBI* and *NcoI* sites of pCRISPR-TT. We designed the homologous recombination template to delete the *sigG1* coding sequence, keeping the downstream and upstream regions intact. Two PCR fragments of approximately 1Kb, immediately upstream and downstream of *sigG1* were amplified using DNA from *S. tsukubaensis* NRRL 18488 as template and primers sigG1_F1/R1 and sigG1_F3/R3 (Table S2). The assembly of the homologous recombination template in pUC19 was performed by Gibson assembly (Gibson et al., 2009). Following confirmation, the assembled homology templates were recovered from pUC19 using *DraI*, and cloned into the *StuI* site in pCRISPR-TT vectors. The resulting pCRISPR-TT plasmid was conjugated into *S. tsukubaensis.* Spore suspensions of positive clones were streaked on Tryptone Soy Agar (TSA, Neogen, UK) containing apramycin 50 μg mL^{-1} and theophylline 10 mM to induce *cas9* expression. We selected null mutants for *sigG1* by colony PCR using oligonucleotides SigG1_*Spe*I_R and CR *sigG1*__conf_F in Table S2. Positive clones were streaked on ISP4 media (10 mM MgCl2) for spore suspensions. Plasmid curing was achieved by streaking spores on TSA and incubating the cultures at 35˚C for 72 hours. Loss of pCRISPR-TT plasmid was confirmed by colony PCR using the pUC_R and sgRNA-R primers (Table S2) and by plating each mutant on ISP4 media (10mM MgCl2) containing either no antibiotic or apramycin 50 μg mL^{-1}.

Table SII4. Primers and plasmids used to generate the Δ*sigG1_CR* markerless mutant-

Primer	Sequence 5´-3´	Use
sgRNA-R	ACGCCTACGTAAAAAAAGCACCGACTCGGTGCC	Markerless gene KO
sigG1_F1	AGCTTGCATGCCTGCAGGTCTTTAAAACCGAGTAGCACCGAGTAG	
sigG1_R1	ACCCGGGTCACATACCTCTCTTCGTATCAGC	
sigG1_F3	GAGAGGTATGTGACCCGGGTCACAACAG	
sigG1_R3	CCCGGGGATCCTCTAGAGTCTTTAAAACCGGCACGCTCTACCAG	

pUC_R	ACACAGGAAACAGCTATGAC
CR_sigG1_conf_F	GCAGCCGACACCGCAAGAGA

Strains	Relevant genotype	Reference
pCRISPR-Cas9	CRISPR-Cas9 backbone vector	(Tong et al., 2015)
pCRISPR-TT	Derivative from CRISPR-Cas9 carrying the theophylline riboswitch and the IS1 element, ApraR	Albuquerque et al., personal communication

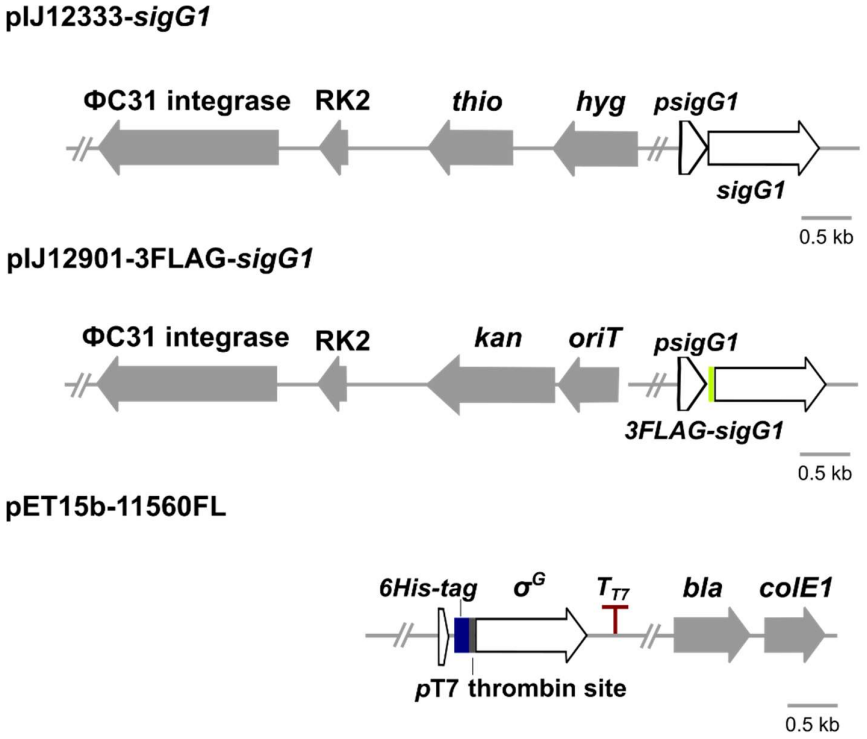


Figure SII22 Schematic representation of the constructs generated to introduce DNA into *S. tuskubaensis* and *E. coli.* Constructs used for complementation of *S. tsukubaensis* strains with the *sigG1* wild type allele under the regulation of its own dedicated promoter (pIJ12333-*psigG1-sigG1*); and a FLAG tagged *sigG1* allele under the regulation of its own dedicated promoter (pIJ12333-*psigG1-3FLAG-sigG1*) are depicted. For overexpression of SigG1 in *E. coli* for antibody production, sigG1 allele was introduced into the pET15b expression vector, under the control of a T7 promoter system (pET15b-11560FL).

Acknowledgements. We acknowledge the help of Pedro Albuquerque with the generation of the Δ*sigG1_CR* mutant. Information on SigG1 predicted structure (in Figure SII16), and the ELISA assays to test antibodies affinity to SigG (in Figure SII17) were provided as service by Ricardo Pires at HenBiotech, Proteins & Antibodies (Coimbra, Portugal).

References

Beites, T., Rodriguez-Garcia, A., Santos-Beneit, F., Moradas-Ferreira, P., Aparicio, J.F., and Mendes, M.V. (2014). Genome-wide analysis of the regulation of pimaricin production in Streptomyces natalensis by reactive oxygen species. Appl Microbiol Biotechnol *98*, 2231-2241.

Birch, A., Häusler, A., and Hütter, R. (1990). Genome rearrangement and genetic instability in Streptomyces spp. Journal of Bacteriology *172*, 4138-4142.

Blin, K., Pedersen, L.E., Weber, T., and Lee, S.Y. (2016). CRISPy-web: An online resource to design sgRNAs for CRISPR applications. Synthetic and Systems Biotechnology *1*, 118-121.

Cullum, J., Altenbuchner, J., Flett, F., and Piendl, W. (1986). DNA amplification and genetic instability in Streptomyces. Biotechnol Genet Eng Rev *4*, 59-78.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods *6*, 343-345.

Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F. (2003). PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proceedings of the National Academy of Sciences *100*, 1541-1546.

Leblond, P., and Decaris, B. (1994). New insights into the genetic of streptomyces instability. (FEMS Microbiology Letters), pp. 225-232.

Martinez-Castro, M., Salehi-Najafabadi, Z., Romero, F., Perez-Sanchiz, R., Fernandez-Chimeno, R.I., Martin, J.F., and Barreiro, C. (2013). Taxonomy and chemically semi-defined media for the analysis of the tacrolimus producer *Streptomyces tsukubaensis*. Appl Microbiol Biotechnol *97*, 2139-2152.

Rudolph, M.M., Vockenhuber, M.-P., and Suess, B. (2013). Synthetic riboswitches for the conditional control of gene expression in Streptomyces coelicolor. Microbiology *159*, 1416-1422.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research *28*, 33-36.

Tong, Y., Charusanti, P., Zhang, L., Weber, T., and Lee, S.Y. (2015). CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. ACS Synthetic Biology *4*, 1020-1029.

Volff, J.-N., and Altenbuchner, J. (1998). Genetic instability of the Streptomyces chromosome. Molecular Microbiology *27*, 239-246.

Zhang, Z., Du, C., de Barsy, F., Liem, M., Liakopoulos, A., van Wezel, G.P., Choi, Y.H., Claessen, D., and Rozen, D.E. (2020). Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. Science Advances *6*, eaay5781.

Chapter 3

Crystal structure of the SnoaL_2 ECF56 SigG1 and its cognate anti-sigma RsfG

Rute Oliveira, [a, b, c] José P. Leite, [a,b,c] Frederico Lourenço, [a,b] Delia Casas-Pastor, [e] Georg Fritz, [e,*] Luís Gales [a,b,d] and Marta V. Mendes, [a, b, d]

[a]Bioengineering and Synthetic Microbiology Group, i3S- Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

[b]IBMC, Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

[c]Programa Doutoral em Biologia Molecular e Celular (MCBiology), ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Rua de Jorge Viterbo Ferreira n. 228, 4050-313 Porto, Portugal

[d]ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Rua de Jorge Viterbo Ferreira n. 228, 4050-313 Porto, Portugal

[e]LOEWE-Zentrum für Synthetische Mikrobiologie, Philipps-Universität Marburg, 35032 Marburg, Germany

*Present address: School for Molecular Sciences, University of Western Australia, Perth 6009, Australia

In preparation

3.1 Abstract

The timely progression of morphological differentiation in *Streptomyces* requires the participation of the ECF56 sigma-factor SigG1, which is essential to activate regulators of the developmental cascades and mechanisms that maintain metal homeostasis. RsfG is the cognate antagonist for SigG1 that prevents its binding to the RNA polymerase to initiate runaway transcription. Here we sought to determine the crystal structure of the first ECF56 sigma factor and its cognate repressor RsfG, an uncharacterised protein with no obvious conserved domains. We obtained crystals of the purified proteins that diffracted to a maximum resolution of 1.9Å. Despite SigG1 having a few sequence similarities with ECF41 SigJ from *Mycobacterium tuberculosis*, sequence homology is limited to three conserved domains. Due to the structural arrangement of these segments, the structure of SigG1 could not be solved by molecular replacement. Consequently, the RsfG-SigG1 complex structure, for which diffraction data was also collected (2.3 Å), remains unsolved. Attempts to obtain phase information were pursued with heavy-atom derivatives, as well as selenomethionine (SeMet) expression, but no phase information was obtained so far (as an example, the best SeMet crystals diffracted only up to 7 Å). Our results indicate that the molecular basis for SigG1 conformation provides additional layers of complexity, despite the similarities with the SnoaL_2-containing ECF described in literature. We hypothesize that we might be in the presence of a new sigma/anti-sigma factor complex, with an undescribed mechanism of regulation that ensures the specific release of SigG1.

3.2 Introduction

As soil-dwelling bacteria, *Streptomyces* have evolved to cope with the constant fluctuations in the surrounding environment, while developing a complex multicellular lifecycle. A major instrument of tempering such fluctuations is the regulation of transcription initiation by an array of alternative extracytoplasmic sigma factors, the ECFs. Sigma factors are interchangeable subunits of the RNA polymerase (RNAP) that promote transcription in response to specific developmental or stress stimuli (Österberg et al., 2011). The binding and unbinding of the ECF to RNAP is a highly dynamic process, hence it requires a stringent regulation, to ensure the assembly of the correct ECF at the correct time for the necessary response. This regulation is often dependent on auxiliary proteins, like the anti-sigma factor that binds to the ECF and prevents its early activity.

ECFs harbouring CT extensions are often described in literature as auto regulatory proteins that lack cognate anti-sigma partners. In previous work, we identified *S. tsukubaensis* SigG1, an ECF sigma with a long CT extension that encodes the SnoaL_2 domain, and its cognate anti-sigma RsfG (Oliveira et al., 2020). The only

example of a SnoaL_2 ECF that has been experimentally characterised is a member of the ECF41 family, *M. tuberculosis* SigJ (Goutam et al., 2017). SigJ harbours a SnoaL_2 that promotes intramolecular contacts with the σ2-σ4 ECF core region to stabilize the ECF binding to promoter DNA. To date, no antagonist regulator was assigned to SigJ.

Previous Direct-Coupling Analysis (DCA) statistics indicated putative differences between SigG1 and SigJ, especially within the interface of contact between SnoaL_2 and the core ECF (Oliveira et al., 2020). Additionally, multiple sequence alignments of ECF41 and ECF56 sequences from different bacteria showed that both the conserved distal NPDKL consensus that links the SnoaL_2 domain to the σ2-σ4 linker and the ECF41-signature motif (YVGPWLPEP) in this region are absent in sigma factors of group ECF56. This indicated that key residues responsible for the compact conformation of the SigJ were missing in SigG1, suggesting a different interface of intramolecular contacts. Despite the information obtained with the different modelling efforts, the physiological ternary structure for members of the ECF56 family remains elusive. Structural data on the new SigG1 protein is of key importance to better characterise these intramolecular interfaces and to distinguish between what could be inhibiting or stabilizing compact conformations.

Conservation of proteins encoded in the same genomic region provide hints on putative regulatory partners of the ECF. In many families, the *ecf* genomic neighbourhood contains genes that code for homologues of known anti-sigma factors (Casas-Pastor et al., 2019). In the case of the families harbouring SnoaL_2 ECFs no obvious anti-sigma factors were found near the *ecf* gene (Huang et al., 2015). Strikingly, by narrowing the search to *Streptomyces* genera, we found genes that encode putative regulatory proteins both in *sigG1* and *sigG2* genomic contexts. For the SigG1 and SigG2 orthologues annotated in 188 Streptomyces genomes, the *ecf* genes were adjacent to genes encoding hypothetical proteins or STAS (Sulphate Transporter and Anti-Sigma factor antagonist) domain containing proteins in about 70% of the cases analysed. This could indicate that intramolecular regulation in these ECFs might not be sufficient to inhibit transcription. In line with this, we identified a cognate sigma antagonist, RsfG.

A number of sigma factors antagonists have been described, following four major classes of anti-sigma factor domains (ASD), each exhibiting different traits of binding to the sigma factor. Overall, ASDs combine a few disordered regions and 2-3 principal helices that are responsible for mediating the interaction with the sigma factor. This interaction might be through direct binding to the sigma factor domains or indirect interaction, requiring a small molecule, like c-di-GMP, to mediate the interaction interface (Gallagher et al., 2019). RsfG is a hypothetical protein with no evident conservation of

protein domains, which brings an extra layer of challenge in predicting the ternary structure for RsfG or the characteristics of the interaction with SigG1.

Following identification and cloning of SigG1 and RsfG, co-expression and co-purification studied revealed that the two proteins form a stable complex in *E.coli*, giving a strong indication that these proteins interact *in vivo.* Nevertheless, not much is known about the composition of RsfG and we could not identify conserved features between this protein and other well-studied anti-sigma factors. The determination of the structure of this enigmatic protein will enable to identify putative traits that are responsible for the inhibitory role over SigG1. Furthermore, it will be crucial to identify the conformation of SigG1 when in complex with its antagonist and better characterise the interface between the two regulators.

3.3 Results

3.3.1 SigG1 is a highly decorated sigma factor. SigG1 is a protein with 369 amino acids which comprises the σ2 and σ4 ECF core domains separated by a 56-residue linker, plus a 100aa long C-terminal extension that encodes the SnoaL_2 domain. Unlike other ECF sigma factors, SigG1 exhibits an additional 50aa long N-terminal. SigG1 has an expected molecular weight of approximately 40kDa and a theoretical isoelectric point of 5.46.

Previous over expression and IMAC purification of the recombinant SigG1 (see Chapter 2) showed two species of the protein as identified by Peptide mass fingerprinting (PMF): one protein with the predicted 40kDa and a C-terminally truncated species with a molecular mass of approximately 25kDa. In addition, SigG1 exhibits two species with an approximate size, close to the 40kDa expected mass for the full-length protein. The fact that these two proteins migrate above the expected molecular weight, and the small difference in size between both species (<5kDa), leaves room for speculation that there might be some posttranslational decoration of SigG1. While sigma factors are monomeric subunits of the RNA polymerase (RNAP), recent studies have shown that there is still a lot to uncover about the posttranslational regulation of these monomers. Posttranslational modifications (PTMs) identified for sigma factors include for instance lysine acetylation of σ^{70}, tyrosine phosphorylation of the EcfP σ2 domain or the phosphorylation of the N-terminal tyrosine, T8, of SigH (Iyer et al., 2020; Kim et al., 2020; Park et al., 2008).

Phosphorylation regulation of the ECF might be a widespread mechanism in bacteria and constitute an important trait in mediating the interaction of the sigma subunit with the β′-subunit of RNAP. Additionally, in EcfP the phosphorylation of the σ2 DAED motif seems to substitute the activity of a cognate anti-sigma factor. The DAED motif is

conserved but the T63 in EcfP degenerate DAED motif is not present in SigG1. Nonetheless, we found that the *S. tsukubaensis* genome encodes a serine/threonine protein phosphatase nearby the *sigG1-rsfG* operon *STSU_11530,* which could act as an accessory protein in a putative phosphorelay process for the SigG1/RsfG system. *In silico* analyses of the *sigG1* sequence predicted two serine/threonine sites with high probability of being subjected to phosphorylation, S12 and T133. By liquid chromatography (LC)-MS/MS proteomics analysis we detected phosphorylation of one threonine residue, T12, and four serine S2, S9, S218 and S236 residues (Table 3.1). In the same genomic context, we found a gene that codes for a peptidoglycan-binding

Table 3.1. SigG1 posttranslational modifications.

POSTTRANSLATION MODIFICATION	*IN SILICO* METHOD	PEPTIDE SEQUENCE	POSITION	SCORE | THRESHOLD	(LC)-MS/MS
THEORETICAL					
ACETYLATION	PAIL	(…)VASV**k**s**A**LQRAR(…)	217	2.31 | 0.20	Sequenced
O-GLYCOSYLATION	NetOGlyc	m**s**DQGGGDST(…)	2	0.80 | 0.50	Sequenced
			9	0.88 | 0.50	-
		MSDQGGGDS**t**(…)	10	0.89 | 0.50	Sequenced
		MSDQGGGDSTQ**t**(…)	12	0.99 | 0.50	Sequenced
			17	0.95 | 0.50	-
			27	0.93 | 0.50	-
			32	0.74 | 0.50	-
			130	0.69 | 0.50	-
			133	0.75 | 0.50	-
			136	0.76 | 0.50	-
			137	0.74 | 0.50	-
			143	0.63 | 0.50	-
			160	0.69 | 0.50	-
			165	0.80 | 0.50	-
			225	0.79 | 0.50	-
			226	0.55 | 0.50	-
			236	0.78 | 0.50	-
			242	0.64 | 0.50	-
			245	0.60 | 0.50	-
	O-GlcNAc	M**s**DQGGGDST(…)	2	0.39 | 0.34	Sequenced
		MSDQGGGDSTQ**t**(…)	12	0.42 | 0.40	Sequenced
			17	0.73 | 0.40	-
			27	0.40 | 0.37	-
			130	0.45 | 0.42	-
			133	0.42 | 0.40	-
			136	0.74 | 0.44	-
			143	0.49 | 0.46	-

			160	0.47 | 0.37	-
			212	0.55 | 0.46	-
			245	0.43 | 0.38	-
			324	0.63 | 0.47	-
			366	0.42 | 0.39	-
	YinOYang	MSDQGGGDSTQ**t**(…)	12	Na	Sequenced
			133	Na	-
N-GLYCOSYLATION	NetNGlyc		2	0.65 | 0.5	ns
PUPYLATION	GPS-PUP		217	3.24 | 2.45	ns
PHOSPHORYLATION	NEtPhos	MSDQGGGDSTQ**t**(…)	12	0.82 | 0.5	Sequenced
			133	0.95 | 0.5	-
EXPERIMENTAL (ADDITIONAL)					
(LC)-MS/MS	Phospho	M**s**DQGGGDSTQT(…)	S2		Sequenced
		MSDQGGGD**s**TQT(…)	S9		Sequenced
		(…)**s**EWGAATEPS(…)	S236		Sequenced
		(…)VASVk**s**ALQR(…)	S218		Sequenced

ns: not studied

-: not found

protein (PBP) glycosyltransferase (STSU_11540). This prompt us to further investigate the presence of glycosylated SigG1 residues within the (LC)–MS/MS data. These results corroborated the *in silico* analyses that predicted O-type glycosylation of SigG1 residues. We were able to obtain the sequence of three peptides with HexNAc O-linked glycosylic decorations of the serine/threonine residues S2, T10 and T12 (Table 3.1). Strikingly, the residues S2 and T12 were common to the glycosylation and the phosphorylation studies, which indicates an intricate regulation of SigG1 dependent on these particular amino acids.

In silico analyses further predicted that SigG1 might be acetylated in the lysine residue K217, which was corroborated by peptide sequencing (Table 3.1, Figure 3.1C). This same lysine might also be exposed to regulation by pupylation, a mechanism found in members of the phylum Actinobacteria, that is characterised by the binding of the ubiquitin-like Pup protein to lysine residues to target proteins to degradation; however, this possibility was not further addressed (Striebel et al., 2014).

3.3.2 *In silico* modelling of SigG1 ternary structure. To predict the ternary structure of the SigG1 full-length protein we carried out structure-based sequence alignment using Phyre 2 and Raptor X servers. The resulting model is consistent to other ECF sigma factors crystal structures available in the PDB database, in which σ2 and σ4 form two helices bundles well separated by a disordered linker. In addition in SigG1,

there is a region mainly composed of beta-sheets that enclose the CT SnoaL_2 domain (Figure 3.1). The sequence alignment retrieved the proteins that could potentially exhibit a similar ternary structure, from the total list of protein structures deposited in the PDB database. The highest scored hit in the alignment rank showed the predicted structure for SigG1 showed 24% identity to the that crystal structure of SigJ from *M. tuberculosis* (PDB: 5XE), probably due to a highest coverage of the amino acid sequence. However, within the hits with 100% confidence, the predicted structure for SigG1 shares highest identity (44%) with the BldN chain of the crystal structure of *S. venezuelae* BldN complex with RsbN (PDB: 6DXO, Schumacher et al. (2018)). This alignment, which covers the N-terminal ECF core region, indicates that SigG1 σ2-σ4 region is more similar to the σ2-σ4 core of BldN than the same region in SigJ.

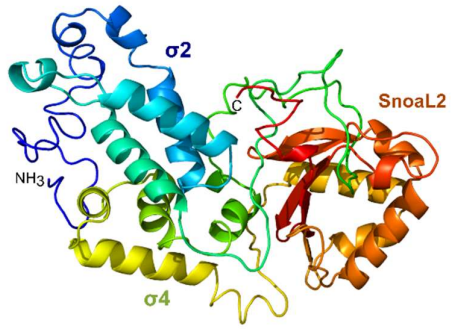


Figure 3.1. Model of SigG1 ternary structure as predicted by the Phyre2 server (Kelley et al., 2015). The σ2 domain is depicted in the blue α-helices; the σ4 domain is represented by the green and yellow α-helices; the SnoaL_2 is represented by the orange α-helices and the red β-sheets.

A more detailed study of the putative SigG1 ternary structure using the DCA statistical method we identified the specific predicted contacts between the SnoaL_2 extension and ECF core domains and compared it to the data available for ECF41 (see Figure 3.1C, in Chapter 2). Residue co-evolution modelling has given high-accuracy indications of protein conformations, as it unveils intramolecular contacts between distant residues in the protein sequences, which have systematically corroborated and complemented crystal structure data (Morcos et al., 2011; Weigt et al., 2009). DCA provided computational predictions of SigG1 conformation where residues in the SnoaL_2 domain contact with pairs of co-evolved residues in the σ2 and σ4 regions, or within the linker between the two. We have generated a structural homology model for SigG1 and mapped the pairs of residues with high predictive scores to this model (Figure 3.2). We found clear differences between the predicted ECF41 and ECF56 intramolecular interactions. ECF41 exhibits contacts between the distal part of the CT extension and the linker between σ2 and σ4 (Wu et al., 2019), while σ2 is relatively scarce in contacts with SnoaL_2. These particular residues are absent in ECF56 and main interactions occur between SnoaL_2 and the σ4 domain or the σ2-σ4 linker. We plotted the top seven predicted contacts in the homology model (Figure 3.3) and observed that the interface between the SnoaL_2, the σ2-σ4 linker and the first part of

the σ4 domain in this model seems was favourable for the interactions predicted by the DCA to occur. The equivalent region of ECF41's YVGPWLPEP linker motif, which is

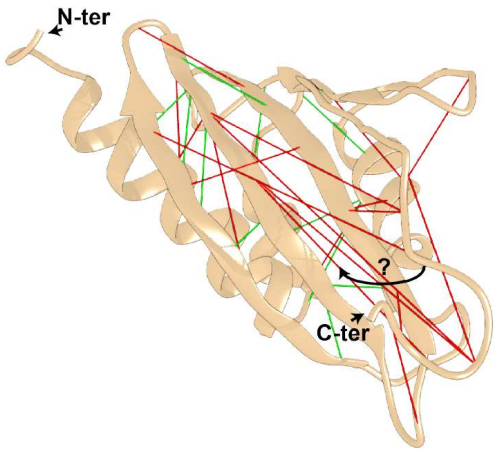


Figure 3.2. Top 30 contact predictions within SigG1 SnoaL_2 (lines) plotted in the homology model of the C-terminal extension of SigG1. Green links show strong predicted contacts with minimum atomic distances < 8Å. Interactions with reduced predictive power, separated by > 8Å (red lines), might still be possible. The C-terminus distal part of SnoaL_2 exhibits an alternative conformation. The CT tail needs to lay on top of the beta sheet, as indicated by the black arrow, in order to obtain have minimal distances in DCA.

essential for contact with the distal part of C-terminal extension in ECF41 (Wu et al., 2019), is predicted to contact the central area of SigG1 SnoaL_2 (Figure 3.3, in purple and grey). This contrasts with ECF41 organisation, where the NPDKL motif in the distal

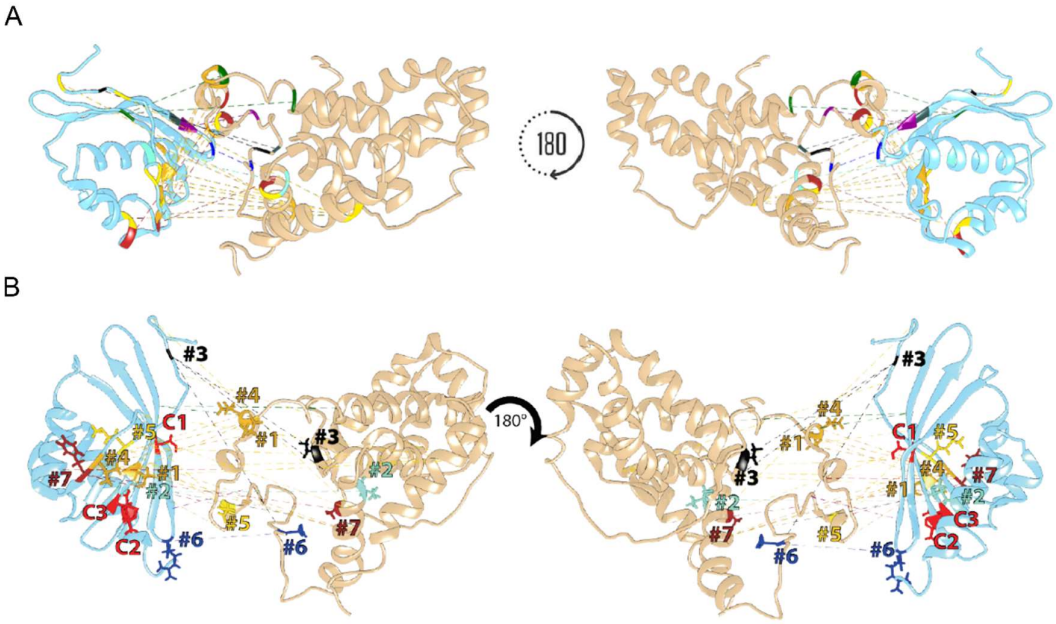


Figure 3.3. Predicted contacts between SigG1 SnoaL_2 (in blue) and the σ2 and σ4 linker, first and last helix of σ4 (in beige). The predictions are plotted on the modelled structure of SigG1. Predicted interactions are represented by dashed lines. Pairs of residues involved in the predicted contacts are highlighted with the same colour. (**A**). Top seven predicted contacts between SigG1 SnoaL_2 (in blue) and the ECF core region mapped into the homology-modelled SigG1 structure. Negative control positions are coloured in red (**B**).

part the C-terminal extension performs this function. Instead of an NPDKL motif of members of ECF41, the last part of the C-terminal extension of SigG1 contains FLGP, which is predicted to contact the last part of σ4 domain and the σ2 domain (Figure 3.4, in yellow and black, respectively). Interestingly, four different predicted contacts bind a single residue from the distal part of σ4 domain (Figure 3.3, in yellow). These regions are more distant in the homology model (Figure 3.3); however, the four residues in SnoaL_2 are in close three-dimensional distance, which could create an alternative conformation of SnoaL_2 with respect to the core ECF domains that allows this contact to occur. Additional experimental data on SigG1 structure is essential to complement the DCA data and confirm this hypothesis.

3.3.3 SigG1 expression and crystallization. In an effort to obtain the atomic resolution structure of SigG1 a recombinant histidine tagged SigG1 protein (6H-SigG1) was overexpressed according to the methods previously described (Figure 3.4 See the supplemental information in Chapter 2 for procedures). Analytical size exclusion chromatography (Superose 12) featured 6H-SigG1 with an experimental size of

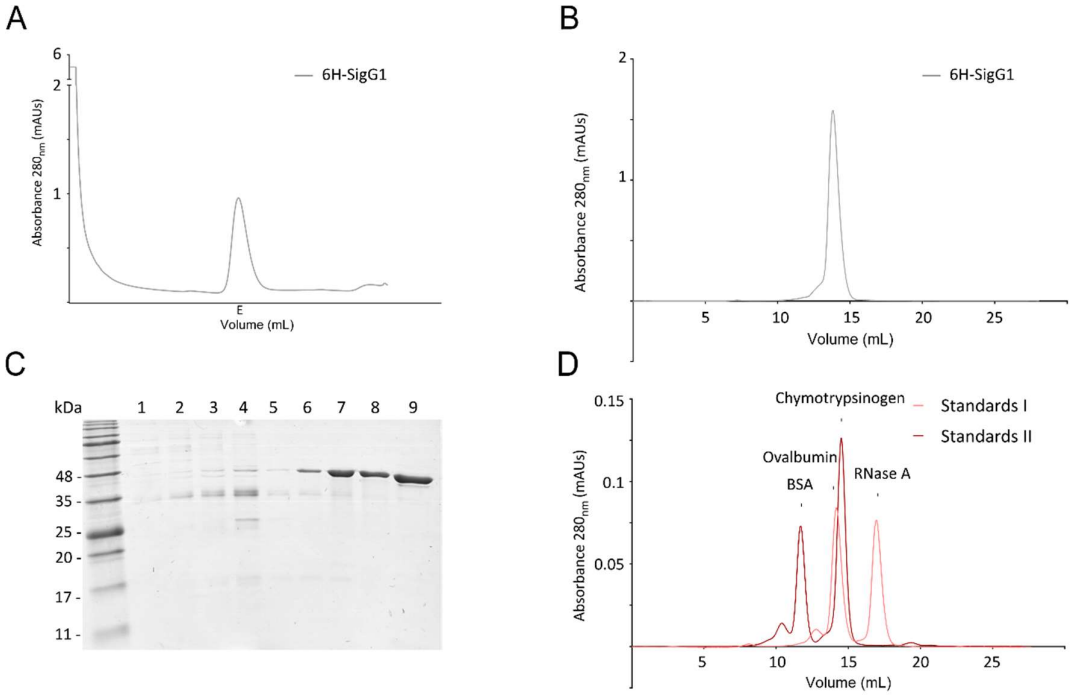


Figure 3.4. Overexpression of the recombinant SigG1. Absorbance spectra of 6H-SigG1 IMAC purification. E: 120mM imidazole (**A**). Size exclusion chromatography (SEC) analysis of the eluted 6H-SigG1 protein (**B**). Purified 6H-SigG1 in SEC fractions analysed by SDS-PAGE. 1-8: SEC fractions within the observed peak (**C**). SEC run of protein standards for molecular weight determination. BSA: 66 kDa; Ovalbumin: 44 kDa; Chymotrypsinogen: 25 kDa; Ribonuclease A: 13.7 kDa (**D**).

approximately 37kDa in solution. The identity of the recombinant protein was confirmed by Western blot and (LC)-MS/MS. The purified 6H-SigG1 protein was crystallized (Figure 3.5) and analysed through X-ray diffraction. Data collection and refinement statistics are shown in Table 3.2.

Table 3.2. Crystallographic data collection and processing statistics for diffracting 6H-SigG1 crystals (outer shell values in brackets)

Diffraction source	SOLEIL PROXIMA-1 beamline
Wavelength (Å)	0.9785
Temperature (K)	100
Detector	DECTRIS EIGER 16M
Crystal-to-detector distance (mm)	278.84
Rotation range per image (°)	0.1
Total rotation range (°)	360
Mean I half-set correlation CC1/2	0.997 (0.968)
Mosaicity (°)	0.08
Resolution range (Å)	49.79-2.50
Total No. of reflections	440779 (66132)
No. of unique reflections	31663 (4583)
Completeness (%)	100 (100)
Multiplicity	13.9 (14.4)
(I/σ(I))	12.4 (4.0)
Rmeas	0.146 (0.614)
Rmerge	0.140 (0.593)
a, b, c (Å)	92.64, 151.84, 128.18
α, β, γ (˚)	90.00, 90.00, 90.00
Space group	C 222_1

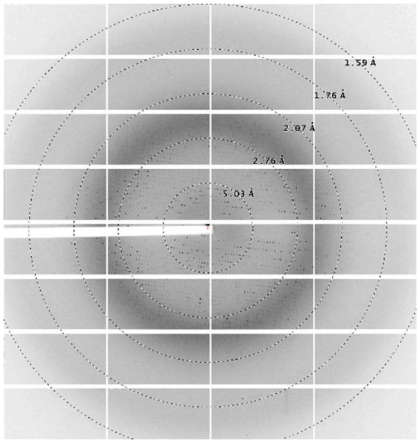


Figure 3.5. X-ray diffraction patterns from 6H-SigG1 crystals. Data from the higher-resolution diffracting crystal was collected at a synchrotron radiation source up to 1.9Å wavelength X-ray. The diffraction pattern is the result of merging 50 consecutive images of 0.1º rotation. The dashed-line circles correspond to the resolution limits indicated.

The crystal asymmetric unit indicated the presence of a homodimer. Making use of structural data available on the Protein Data Bank (PDB) for the closest SigG1 orthologue, *M. tuberculosis* SigJ (PDB: 5ex7), and in order to solve the phases of these crystals, we applied modelling studies for molecular-replacement. Moreover, we carried out molecular replacement attempts with the *C. metallidurans* CnrH σ2 domain (PDB: 4CXF, Maillard et al. (2014)); the *E. coli* σ^{70} σ4 domain (PDB: 2P7V, Patikoglou et al. (2007)); and the *S. venezuelae* BldN (PDB: 6DXO, Schumacher et al. (2018)). All the trials made with these templates were unsuccessful.

3.3.4 The cognate anti-sigma factor, RsfG. In previous work, we identified the cognate antagonist for SigG1, RsfG (Oliveira et al., 2020). RsfG is an uncharacterised protein with and theoretical molecular weight of 16.34kDa and an isoelectric point of 4.41. Like other studied anti-sigma factors, it did not share identity with other characterised proteins or any Pfam conserved domains, which makes it a very intriguing protein to study. When we purified the recombinant 6H-RsfG-SiG1 complex from *E.coli* (see Chapter 2) we found that RsfG was expressed in two species with a slight difference in the molecular weight (approximately 2kDa). To investigate if this was a consequence of a posttranslational modification of RsfG, we run the amino acid sequence in an array of *in silico* prediction tools and found high degree of probability for RsfG acetylation, glycosylation or phosphorylation (Table 3.3).

Table 3.3. RsfG posttranslational modifications.

POSTTRANSLATION MODIFICATION	*IN SILICO* METHOD	PEPTIDE SEQUENCE	POSITION	SCORE | THRESHOLD	(LC)-MS/MS
THEORETICAL					
ACETYLATION	PAIL		116	0.58 | 0.20	-
O-GLYCOSYLATION	NetOGlyc		4	0.62 | 0.50	-
			5	0.76 | 0.50	-
			11	0.57 | 0.50	-
	O-GlcNAc		4	0.46 | 0.30	-
			5	0.50 | 0.30	-
			11	0.62 | 0.30	-
			66	0.61 | 0.40	-
			127	0.41 | 0.40	-
			150	0.48 | 0.40	-
	YinOYang		150	na	-
N-GLYCOSYLATION	NetNGlyc		2	0.65 | 0.50	ns
PHOSPHORYLATION	NEtPhos		150	0.69 | 0.50	-

However, a liquid chromatography (LC)–MS/MS proteomics search for these specific PTMs did not confirmed the modifications predicted by the *in silico* individual tools.

Further *in silico* analyses and membrane topology mapping predicted that RsfG consists of an alpha helix at the amino-terminal (M1–V21) preceding the core region of a 9-stranded antiparallel β-barrel, and a disordered carboxy-terminal region (G135-V154) (Figure 3.7). In addition, it displays a predicted short hydrophobic region of 22 amino acids (A124-Y145) at the C-terminal region. The strongest model predicts that the N-terminal of the protein is positioned in the cytosol, while the majority of the protein seems to be non-cytosolic (Figure 3.6). *S. venezuelae* anti-sigma factor RsbN also contains a short region of about 20 residues, predicted to be a transmembrane domain. Although the function of this domain has not been addressed so far, it could point out to a location of RsbN and RsfG in close proximity to the cell membrane. In fact, the TM in RsbN appears to be right in the middle of a cytosolic part of the protein and a putative extracellular domain, while RsfG lacks an extracellular region. In the RsbN-BldN system, the activation signal is perceived outside of the cell, probably by the RsbN extracellular domain. As we have previously described, SigG1 activation is dependent on an intracellular signal, hence it does not require an extracellular domain of RsfG to trigger SigG1 release (Oliveira et al., 2020). Homology predictions by Phyre2 further retrieved similar structures for RsfG. However, all the hits in the rank were less than 96% confidence show very low identity.

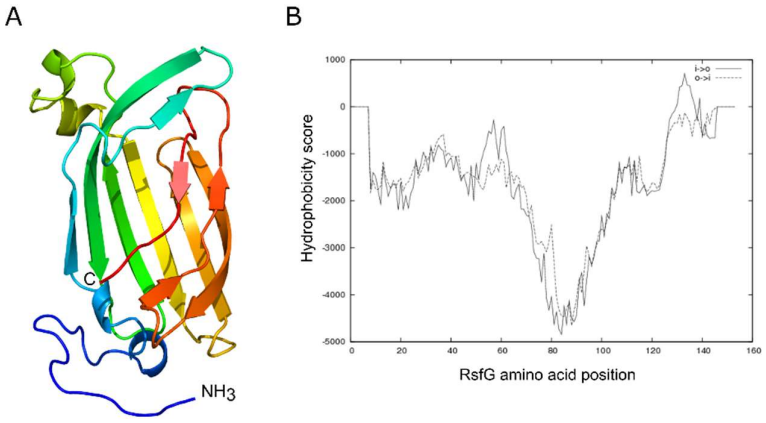


Figure 3.6. Ternary structure and membrane topology predictions for RsfG. Structure model for RsfG as predicted by the Phyre2 server (Kelley et al., 2015) (**A**). Hydropathy profile using the TMpred method. The strongest model indicated predicts that the N-terminal region is positioned in the cytoplasm (i>o) and is represented by the filled black line. A short hydrophobic region of 22 residues (A124-Y145) is predicted with a hydrophobicity score of 706 (**B**).

The first hit (37% identity) was a fatty acid binding protein from *M. tuberculosis,* Rv0813 (PDB: 2FWV, Shepard et al. (2007)). When we aligned the amino acidic sequence of Rv0813 to RsfG, we observed the identity shared between the two proteins was 17%. As it has been traditional for anti-sigma factors, we could not extract relevant information from these alignments.

3.3.5 Co-expression expression and co-purification of SigG1 and RsfG in *E. coli.*

Since RsfG shows no homology to known anti-sigma factors, it was unclear how it might bind and affect the function of SigG1. In order to better characterise SigG1 and RsfG, and obtain more information on their physiological structure in complex, we have produced the respective recombinant proteins in a heterologous system, using *E. coli*.

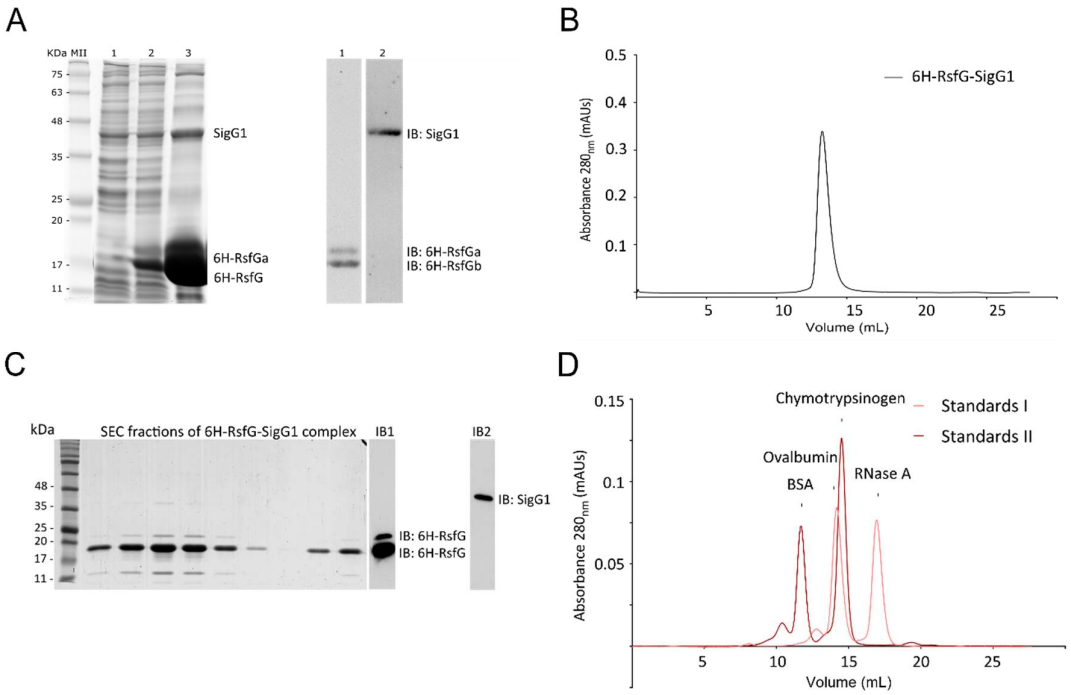


Figure 3.7. Production of 6H-RsfG-SigG1 complexes. Co-expression and co-purification of SigG1 and 6H-RsfG in *E. coli* Nico21 cells using the pRSFDuet-1 system. 1: non-induced total extract; 2: total extract of cells induced with 1mM IPTG; 3: SigG1 bound to 6H-RsfG eluted with 250mM imidazole. Immunoblot detection using an antibody against the His-tag (1) and a polyclonal antibody against SigG1 (2) confirmed the identity of the proteins (**A**). Size exclusion chromatography (SEC) analyses of 6H-RsfG-SigG1 complexes (**B**). SDS-PAGE of RsfG-SigG1 complexes SEC fractions with confirmation of the proteins identity by Western blotting. 6H-RsfG was detected using an anti-His-tag antibody (IB1). SigG1 was detected with a specific polyclonal anti-SigG1^{A16V39} antibody (IB2) (**C**). SEC run of protein standards for molecular weight determination. BSA: 66 kDa; Ovalbumin: 44 kDa; Chymotrypsinogen: 25 kDa; Ribonuclease A: 13.7 kDa (**D**).

We addressed the conformation of SigG1 bound to RsfG by purifying a 6H-RsfG-SigG1 native complex. IMAC purification of the 6H-RsfG-SigG1 complex unveiled, in addition to the untagged SigG1, two species of 6His-RsfG, as identified by Peptide Mass Fingerprinting (PMF): one protein with the theoretical molecular mass of 16kDa, and a second species with an approximate increase of 2kDa in the molecular mass (Figure 3.7). SigG1 was bound to RsfG in a complex of approximately 57kDa, as determined by the pull-down assays followed by analytical size exclusion chromatography (SEC).

Crystals of 6H-RsfG-SigG1 complexes achieved its maximum size after one week of growth at 20˚C, and were loaded on a SDS-PAGE gel to visualize the crystallized protein (Figure 3.8). X-ray diffraction data collection is summarized in Table 3.4. Despite of several attempts, we could not find a solution for the phases of these crystals. We pursued extensive molecular replacement searches using the structures available in PDB of proteins that shared high homology rates with SigG1 and RsfG in Phyre2 analyses. Searches were carried out using *M. tuberculosis* SigJ (PDB: 5xe7, Goutam et al. (2017)); *C. metallidurans* CnrH σ2 domain (PDB: 4CXF, Maillard et al. (2014)); *E. coli* σ^{70} σ4 domain (PDB: 2P7V, Patikoglou et al. (2007)); the human metal binding protein (PDB: 3IA8, Bianchetti et al. (2011)); and *M. tuberculosis* hypothetical protein Rv0813 (PDB: 2FWV, Shepard et al. (2007)). Since no single solution was obtained, phase

Table 3.4. Data collection and processing statistics of for a native 6H-RsfG-SigG1 diffracting crystals

Diffraction source	SOLEIL Proxima-1 beamline
Wavelength (Å)	
Temperature (K)	100
Detector	DECTRIS EIGER 16M
Crystal-to-detector distance (mm)	
Rotation range per image (°)	
Total rotation range (°)	
Exposure time per image (s)	
Mean I half-set correlation CC1/2	0.995 (0.661)
Mosaicity (°)	1.06
Resolution range (Å)	79.25-2.30
Total No. of reflections	108945 (14968)
No. of unique reflections	31236 (4500)
Completeness (%)	99.5 (99.4)
Multiplicity	3.5 (3.3)
(I/σ(I))	8.5 (2.2)
Rmeas	0.102 (0.667)
Rmerge	0.087 (0.563)

a, b, c (Å)	73.37, 61.13, 82.75
α, β, γ (˚)	90.00, 106.72, 90.00
Space group	P 1 2_1 1

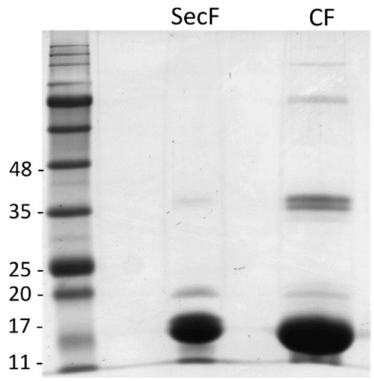


Figure 3.8. Crystallization of native 6H-RsfG-SigG1 crystals. SDS-PAGE of 6H-RsfG-SigG1 crystallized protein (**A**). Crystals were grown in the MembFac #29 (0.1 M Ammonium sulfate; 0.1 M HEPES sodium pH 7.5; 0.5 M sodium phosphate dibasic dihydrate; 0.5 M potassium phosphate dibasic) condition over one week. SecF: SEC fraction; CF: crystal fraction.

determination was carried out with heavy-atom derivatives with multiple combinations of metal concentrations (up to 100 mM). Both single wavelength anomalous diffraction (SAD) and multi-wavelength anomalous dispersion (MAD) measurements were collected. The presence of anomalous signal was screened with the SHELXC/SHELXD/SHELXE pipeline (Sheldrick, 2010), but so far no solution was found. In an additional effort to overcome this problem, we proceeded to the production of 6H-RsfG-SigG1 complexes containing selenomethionines (SeMet).

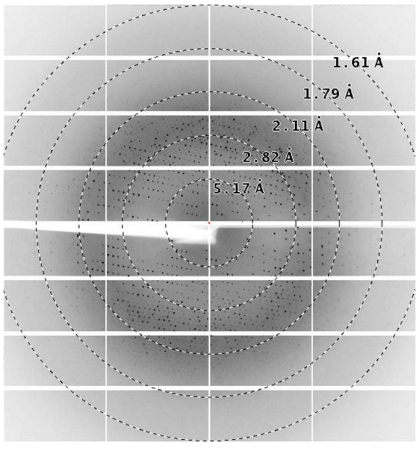


Figure 3.9. X-ray diffraction patterns from 6H-RsfG-SigG1 crystals. Data from the higher-resolution diffracting crystal was collected with synchrotron radiation source up to 2.3Å wavelength X-ray. The diffraction pattern is the result of merging 50 consecutive images of 0.1º rotation. The dashed-line circles correspond to the resolution limits indicated.

3.3.6 Production of selenomethionine-containing (SeMet) 6H-RsfG-SigG1 crystals. Since phase determination by molecular replacement or heavy-atom derivatives could not be achieved, we prepared SeMet-mutants, where selenium atoms replaced the sulphur groups in the methionine residues. Unexpectedly, through analysis

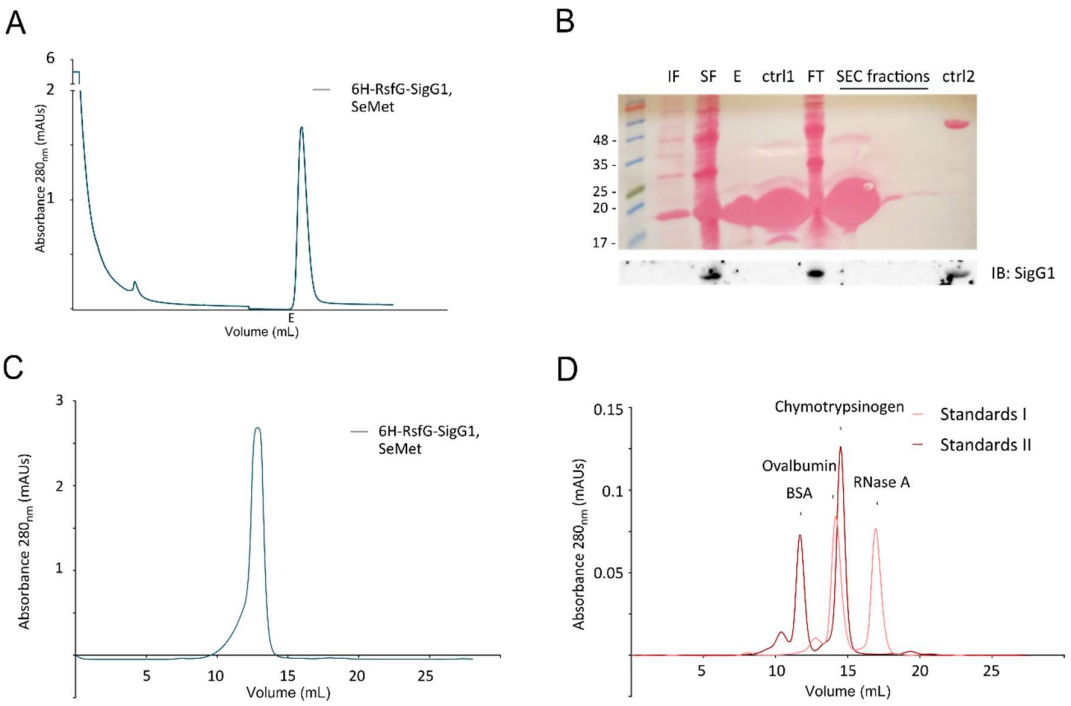


Figure 3.10. Production of selenomethione-containing (SeMet) 6H-RsfG-SigG1 complexes. IMAC spectra. E: 250mM imidazole (A). Expression of selenomethionine-containing 6H-RsfG-SigG1 complexes analysed by SDS-PAGE, and stained the nitrocellulose membrane with transferred protein with Ponceau S. Western blot using an anti-SigG1^{A16V39} polyclonal antibody confirmed the identity of SigG1 (B). Results show that the SigG1 soluble protein co-expressed with 6H-RsfG did not bind to the Ni^{2+} column and was discarded in the IMAC flow through fraction. Size exclusion chromatography (SEC) analyses of the eluted 6H-SeMet-RsfG protein (C). SEC run of protein standards for molecular weight determination. BSA: 66 kDA; Ovalbumin: 44 kDA; Chymotrypsinogen: 25 kDa; Ribonuclease A: 13.7 kDa (D).

of the total protein lysates by SDS-PAGE we observed that most of the soluble SeMet SigG1 was release from 6H-SeMet-RsfG and discarded in the flow through fraction of the IMAC purification (Figure 3.10). Nonetheless, we sought to further explore the function of RsfG by solving its three-dimensional structure – and also because it could be helpful for future determination of the RsfG-SigG1 complex structure – so we purified the 6H-SeMet-RsfG that remained bound to the chromatography nickel column (theoretical size: 17.6kDa) and used it in crystallization trials. Analytical SEC featured the purified protein with an experimental size of 48kDa, indicating the presence of homo or heterodimers in solution. (LC)-MS/MS confirmed the identity of 6H-SeMet-RsfG, but was elusive regarding the presence of SeMet SigG1 in this solution. Despite the numerous crystallization trials, only one condition yielded crystals of 6H-SeMet-RsfG. X-ray diffraction of these crystals was unsuccessful and data - obtained to the maximum

resolution of 7Å - was insufficient to proceed with further analyses (poor diffraction spots, Figure 3.11).

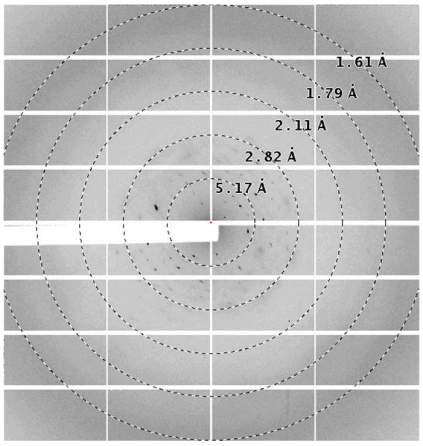


Figure 3.11. X-ray diffraction patterns from 6H-6H-SeMet-RsfG crystals. The diffraction pattern is the result of merging 50 consecutive images of 0.1º rotation. The dashed-line circles correspond to the resolution limits indicated.

3.3.6.1 The 6H-RsfG multimeric form is dependent on co-purification with SigG1. Sigma/anti-sigma factor complexes are traditionally composed of monomeric forms of both proteins that bind in solution in a 1:1 proportion. The co-expression of 6H-RsfG with SigG1 resulted in high yields of two species of RsfG when compared to the amount of SigG1 produced. To assess if the multimeric form of the 6H-RsfG recombinant protein was physiologically relevant or an artefact of a long expression protocol we expressed the RsfG coding sequence in the MCS1 of the pRSFduet-1™ vector used for co-expression. The MCS2 used to produce the untagged SigG1 was left unchanged in this experiment. The individual expression of RsfG in this system using the same

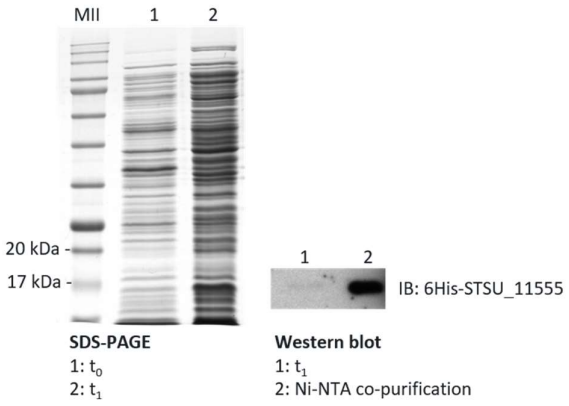


Figure 3.11. Production of the tagged version of RsfG (6H-RsfG). SDS-PAGE analyses of the single expression of RsfG using the pRSFduet-1™ MCS-1. The MCS-2 was left unchanged. Expression of 6H-RsfG alone produces a soluble single species of the protein. To confirm its identity by WB, RsfG was detected with an anti-His-tag antibody. SDS-PAGE 1:non-induced fraction 2: IPTG-induced total protein. Immunoblot 1: IPTG-induced total protein, 2: Ni-NTA purification eluate.

expression protocol was found to produce soluble RsfG in a much lower amount than the observed in the co-expression experiments. Additionally, Ni-NTA purification of the recombinant protein and identification by Western blot showed that the expression of 6H-RsfG alone produces a soluble single species of the protein, unlike what we found in the co-purification experiments. These results point out to a more complex and abundant production of the RsfG protein when in complex with SigG1, which might not occur when the cognate ECF regulator is absent.

3.3.6.2 6H-SeMet-RsfG and 6H-SigG1 in heterogeneous solution. In order to obtain samples of 6H-RsfG-SigG1 with a selenomethionine signal we attempted to promote the formation of the complex by adding the 6H-SeMet-RsfG and 6H-SigG1 proteins purified individually, in a 1:1/4:1 solution. The resulting sample was run on a size exclusion chromatography system and the SEC fractions were analysed by SDS-PAGE (Figure 3.13). The two proteins were separated throughout elution according to their size, indicating that they remain unbound in solution. Results show that the proteins purified individually did not bind in solution, indicating that a specific physiological context is required for a RsfG-SigG1 stable complex to be formed.

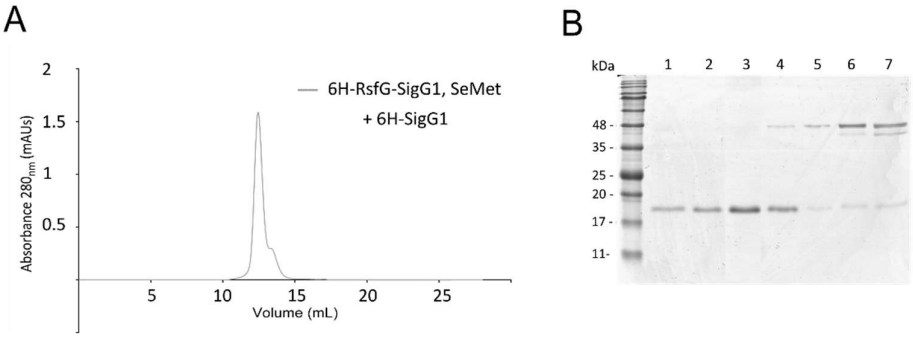


Figure 3.12. SDS-PAGE analyses of SEC fractions of a mixture of SigG1 and RsfG. Previously purified 6H-SeMet-RsfG and 6H-SigG1 were mixed in solution in a 4:1 proportion. The resulting heterogeneous solution was analysed by SEC. Upon elution, the two proteins separated according to their size indicate that they remain unbound in solution. 1-7: SEC fractions SEC fractions within the observed peak up to 14mL of elution volume.

3.4 Discussion

SigG1 is a novel stress-responsive sigma factor with unique features that most probably control its activity. One interesting observation in this work was that, although the experimental size of SigG1 recombinant protein meets the theoretical size of the monomeric protein, we found additional SigG1 species with slight changes on the apparent size. Therefore, beyond the presence of several layers of self and assisted regulation, we propose that SigG1 is heavily decorated with modifications associated to

its functional groups. PTMs that are essential for the optimal activity of ECFs have been described in both Gram-positive and Gram-negative bacteria (Iyer et al., 2020; Park et al., 2008). Perhaps the most studied mechanism is the regulation of ECF activity by kinases and phosphatases. Another study of the *M. tuberculosis* phosphoproteome predicted that SigG − which harbours a SnoaL_2 CT extension − might also be subjected to phosphorylation of serine 190 from the σ4 domain by an unknown kinase (Prisic et al., 2010). Here we pinpointed four serine residues, S2, S12, S218 and S236 as playing roles in phosphorylation-dependent posttranslational regulation. Two of these modifications were observed in the long N-terminal part of SigG1, and similarly to what was estimated for *M. tuberculosis* for SigG, SigG1 S218 might coordinate the phosphorylation of the SigG1 σ4 domain. The last phosphorylation sequenced in our studies was in S236 in the beginning of the SnoaL2 domain, representing an extra regulatory mechanism associated to this extension. Nonetheless, in their studies of the three-dimensional structure of *M. tuberculosis* SigJ, Goutam et al. (2017) propose that the presence of SigJ species with approximate sizes, are a result of the addition of an unknown ligand to the SnoaL2 domain of this protein. Hence, one cannot exclude that the different apparent sizes observed are due to the presence of small molecules that could act as specific ligands that trigger signal transduction from these ECFs.

The mode of action of the SigG1 protein gets more fascinating as we study further layers of regulation. From the calculations on the probability of intramolecular contact in SigG1, it appears that pairs of residues in the SnoaL_2 domain co-evolved with residues in the ECF core, possibility to maintain an optimal active/inactive compact conformation of SigG1. As Casas-Pastor et al. (2019) has shown before, ECFs with SnoaL_2 extensions have been included in phylogenetic groups that are not always proximate in the ECF tree, indicating that more factors determine the particular features of ECFs with these extensions. In fact, DCA results for *M. tuberculosis* SigJ also predicted a compact conformation of the SnoaL_2 domain towards the ECF core, but it shows different points of co-evolved residues when compared to SigG1, indicating that the extension interacts with the core ECF regions in a unique manner, depending on the type of ECF. This demonstrates that while sharing a high degree of conservation that enables the identification of Pfam SnoaL_2 domains, these extensions can carry their own specificities that result in dissimilar conformations of the ECFs. The study of the conformation of SigG1 in its different physiological states – active/inactive, compact/relaxed, isolated/in complex with its anti-sigma – will help unveil the purpose of having this array of regulatory features.

Despite the numerous efforts, molecular replacement using the *M. tuberculosis* SigJ structure in SigG1 and the RsfG-SigG1 complexes diffraction data failed, therefore,

and despite the good resolution of the information obtained, we could not solve the phases of the crystals grown in the different conditions. We will carry on with these searches using structures of the closest sigma/anti-sigma complexes as models. Examples of the structures that are in use are *S. venezuelae* BldN-RsbN (PDB: 6DXO, Schumacher et al. (2018)), *C. metallidurans* CnrH-CnrY (PDB: 4CXF, Maillard et al. (2014)). The difficulties in modelling with the structure of the most similar protein SigJ, both in homogeneous solution and in complex with RsfG, suggest that the structure of SigG1 might be unrelated to the structure encountered for SigJ. The fact that the SigJ structure could not be superimposed with SigG1 structural data, gives additional support to the idea that ECF41 and ECF56 proteins might enclose dissimilar conformations and consequently, different modes of action.

While we find high conservation of SigG1 homologues, RsfG is rarely found in genetic neighbourhoods of actinobacterial ECFs (Casas-Pastor, personal communication). In a search of 14483 actinobacterial genomes, (147 of which are *Streptomyces* genomes), *rsfG* is predominantly present in the vicinity of genes encoding ECFs from groups ECF59, ECF57, ECF43, ECF03 and ECF42. Within the ECF56 family, only 1% *ecf* genes were found nearby *rsfG* orthologues. Zooming into the genomic contexts of ECF56 subgroup 3 – which encloses SigG1 – we found *rsfG* genes in 10% of the genomes, 70% of these located in located in position +1 from the ECF coding sequence. These results suggest that RsfG might not always co-occur with ECF56 proteins and that it is unlikely to play a conserved function throughout the phylum. Nonetheless, and similarly to what was observed before for the conservation of LysR-type regulators close to ECF56 proteins (see Discussion section in Chapter 2), we hypothesize that enriching the pool of genomes to the *Streptomyces* genus would give a more clear insight of the conservation of RsfG and its function in bacteria.

In silico predictions, pinpointed candidate RsfG residues to contain PTMs that commonly regulate sigma/anti-sigma systems. Although we found two RsfG species that migrated with different apparent sizes, (LC)-MS/MS sequencing could not identify any of these modifications. We speculate that specific physiological conditions might determine the effective decoration RsfG, depending on whether it is isolated or when in complex with SigG1. The different relative abundance of RsfG recombinant proteins that were purified individually or in sigma/anti-sigma complexes adds support to these speculations and could indicate that PTMs might be lost when SigG1 unbound form RsfG. Nonetheless, evidence obtained with this work indicates strong plasticity of the features underpinning the sigma/anti-sigma complex, but also of posttranslational control of the individual SigG1 and RsfG proteins.

By co-purifying the recombinant SigG1 with 6H-RsfG, we were able to produce protein crystals where both proteins seem to co-crystalize. The native 6H-RsfG-SigG1 complex featured an experimental size of 57kDa, consistent with a 1:1 stoichiometry of SigG1 to 6H-RsfG. Surprisingly, SDS-PAGE analyses indicated a much higher proportion of 6H-RsfG in this solution. One possibility is that we have an additional homodimeric complex of 6H-RsfG in this solution with an approximate size to a 6H-RsfG-SigG1 complex that was impossible to distinguish by SEC analyses. However, with the novelty of these proteins and due to the complexity of this sigma/anti- sigma association, solving the phases of these crystals has been a major challenge of this work.

The attempts made to generate selenomethionine-substituted 6H-RsfG-SigG1 (6H-SeMet-RsfG-SigG1) to obtain phase information also proved ineffective. When cells were grown in selenomethionine-supplemented medium, SigG1 did not form a stable complex and was dissociated from RsfG in the eluates obtained. The purification of these proteins revealed a 48kDa protein in SEC runs. Since the theoretical size of the monomeric 6H-RsfG was 17.6kDa we speculate that we could have purified an homodimeric for of 6H-SeMet-RsfG. This would be consistent with the dimerization signal observed in the BACTH experiments for RsfG (see Chapter 2). Another possibility is that we purified a 6H-SeMet-RsfG-SigG1 complex formed by 6H-SeMet-RsfG bound to a residual amount of SigG1 that was undetectable by SDS-PAGE or (LC)-MS/MS. Overall, and despite having obtained large crystals, this indicates that these might not be the ideal conditions to produce such complexes. Not because of reduced solubility of SigG1, but possibly due to the loss of the optimal conformation of the proteins upon the incorporation of the heavy metal residues. One specificity of these experiments is the presence of DTT in the growth conditions. We speculate that changes in the redox state of the environment might affect the stability of the SigG1-RsfG interaction.

Gathering our results, we see a number of important questions that were left unanswered. For several ECFs described in previous studies, the linker between the sigma factor σ2 and σ4 domains is flexible to favour an “open” conformation of the ECF that allows binding of the ECF core to the -10 and -35 boxes in the promoter DNA. Do ECF56 proteins have such a relaxed conformation? Does a compact SigG1 form inhibit binding to the -10 and -35 boxes? What role does the SnoaL_2 extension play in the stabilization/repression of the ECF? Which are the active and inactive conformations of SigG1? Will SigG1 assume a compact conformation when in complex with RsfG? Does this interaction require the direct binding of SigG1 to RsfG? Are there other signalling molecules involved in the assembly of SigG1-RsfG complex? Are we in the presence of a new class of anti-sigma domains (ASD)?

In sum, this work leaves a few open avenues to explore regarding the mode of action of ECF56 sigma factors. A few observations point towards a regulation of these ECFs that goes beyond its intrinsic architecture. Perhaps the most obvious and primordial evidence is that ECFs with SnoaL_2 extensions are classified in distant phylogenetic groups, given the differences in their regions σ2 and σ4 (Casas-Pastor et al., 2019; Huang et al., 2015). This provides a strong indication that despite the preservation of the three conserved domains, these proteins have evolved to acquire new mechanisms of action, in order to meet more challenging output demands. Structural insights will be essential not only to dissect the binding determinants of the RsfG-SigG1, but also to define the key features that distinguish ECF41 from ECF56 sigma factors.

3.5 Materials and Methods

Bioinformatics procedures. *In silico* searches for posttranslational modifications were carried out using online tools. PAIL (http://bdmpail.biocuckoo.org/prediction.php) predicted acetylation of lysine residues (Li et al., 2006). Putative glycosylated residues were retrieved from NetOGlyc, NetNGlyc and YinOYang (http://www.cbs.dtu.dk/services) (Gupta and Brunak, 2002; Gupta, 2004; Steentoft et al., 2013). NetPhos (http://www.cbs.dtu.dk/services) retrieved putative phosphorylated residues (Blom et al., 1999). GPS-PUP (http://protein.cau.edu.cn/pbPUP/prediction.php) was used to search for pupylation (Liu et al., 2011). Structure homology searches were performed using SigG1 and RsfG protein sequences against the Protein Data Bank (PDB), using the Phyre 2 Server (Kelley et al., 2015). RsfG hydrophobic profile was obtained with TMPred tool (https://embnet.vital-it.ch/software/TMPRED) (Hofmann and Stoffel, 1993).

Production of recombinant proteins. Purification of 6H-SigG1 was performed by *E. coli* heterologous overexpression and chromatography as described in Chapter 2. The purity of the protein in the SEC fractions were analysed by SDS-PAGE. The fractions of interest were concentrated to 10 mg mL^{-1} using an Amicon Ultra-15 30kDa (Milipore). Expression of 6H-RsfG-SigG1 complexes was carried our using a single colony of Nico21 cells expressing the pRSFduet-1-6H-RsfG-SigG1 that was inoculated in LB medium, containing 1% (w/v) glycerol and 50 mg mL^{-1} kanamycin. Cultures were grown to OD_{600nm} of 0.3 at 30°C prior to 2% ethanol and cold-induced stresses. Induction was carried out with 1mM Isopropyl beta-D-1-thiogalactopyranoside (IPTG), for 20h at RT. Cells were then collected at 4000 rpm for 30 min at 4°C and suspended in lysis buffer (20 mM Sodium phosphate, 150 mM NaCl, pH7.4) containing 10 µg mL^{-1} DnaseI, 200 µg mL^{-1} lysozyme, 3mM 1,4-Ditiothreitol (DTT) and EDTA-free protease inhibitor cocktail.

Cells were lysed using a French press homogenizer at 8000 psi and the cell debris was removed by centrifugation at 17000 rpm for 30 min at 4°C. The supernatant was incubated with a nickel-based matrix (Ni-NTA Agarose, Qiagen) and eluted with 250 mM imidazole by centrifugation. Imidazole was removed during protein concentration using an Amicon Ultra-15 30kDa. The protein was eluted with 250 mM Imidazole. In order to produce selenomethionine labelled 6H-RsfG-SigG1 complexes a single colony of Nico21 cells expressing the pRSFduet-1-6H-RsfG-SigG1 was inoculated in SelenoMethionine medium from Molecular Dimensions, containing 50 mg mL^{-1} kanamycin. Expression conditions followed the procedure described for the native protein. The total protein extract was loaded onto a 1 mL HisTrap IMAC column (GE Healthcare) and equilibrated with 20 mM sodium phosphate buffer containing 150 mM NaCl, 3 mM DTT and 20 mM imidazole, pH 7.4. The protein was eluted with 250 mM Imidazole. SEC purified fractions of interest were concentrated to 10 mg mL^{-1} using an Amicon Ultra-15 3kDa (Milipore), exchanging the buffer to 10 mM HEPES, 150 mM NaCl, pH 7.5.

Size exclusion chromatography (SEC). Analytical SEC was used to increase the purity of the sample and to probe the molecular weight of the purified proteins. SigG1 in in 20 mM NaP e 150 mM NaCl, pH 7.4 buffer was injected at a concentration of 7.5 mg mL^{-1} in a Superose 12 column, at a flow rate of 0.2 mL min^{-1}. 6H-RsfG-SigG1 native protein was run in 20 mM NaP e 150 mM NaCl, pH 7.4 buffer at a concentration of 0.5 mg mL^{-1} in a Superose 12 column, at a flow rate of 0.2 mL min^{-1}. 6H-SeMet-RsfG-SigG1 in 10 mM HEPES, 150 mM NaCl, pH7.5 buffer was injected at a concentration of 10 mg mL^{-1} in a Superose 12 column, at a flow rate of 0.2 mL min^{-1}. For probing of the molecular weight, we used a calibration curve with the following standards: Chemotrypsinogen A 25 kDa, BSA 66 kDa and Aldolase 158 kDa.

(**LC)–MS/MS analyses.** The identity of the proteins was confirmed through liquid chromatography proteomics using an Ultimate 3000 liquid chromatography system coupled to a Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific).

Crystallization conditions. Initial 6H-SigG1 crystallization trials were performed with commercials kits using a crystallization robot (Oryx4), with each drop contained a 1:1 protein/precipitant solution ratio. The following commercially available kits were tested: PEGRx™ 1, Index™ HR2-144, SaltRx™ HR2 and MembFac™ (Hampton Research, USA) and Morpheus (Molecular Dimensions, UK). A crystallization hit was found with MembFac™ condition #33 (0.1 M Ammonium sulfate, 0.1 M HEPES sodium pH 7.5, 18% v/v Polyethylene glycol 400). Crystals were produced at 20 ˚C and grew to maximum size within two weeks (Table 3.6). The best crystals were soaked in mother

liquor containing 25% (v/v) glycerol and flash frozen in liquid nitrogen for diffraction experiments.

For crystallization trials of the purified RsfG-SigG1 complex, protein was used at a concentration of 10 mg ml^{-1} in 20 mM sodium phosphate pH 7.4, 10-15 mM NaCl, after diafiltration using an Amicon 3K cut-off membrane. Initial screening was performed with an Oryx4 crystallization robot (Douglas Instruments, UK), using a 1:1 protein/precipitant solution ratio and at 20°C. The following commercially available kits were tested: Crystal Screen™ 2, MembFac™, PEG/Ion Screen™, PEG/Ion 2 Screen™ (Hampton Research, USA); Morpheus (Molecular Dimensions, UK). A crystallization hit was found with MembFac™ condition #29 (0.1 M Ammonium sulfate; 0.1 M HEPES sodium pH 7.5; 0.5 M sodium phosphate dibasic dihydrate; 0.5 M potassium phosphate dibasic). Crystallization was then pursued with manually dispensed 24-well sitting-drop plates incubated at 20 °C, with different protein/precipitant solution ratios (1:1; 1:4) (Table 3.7). Crystals grew over a week. For heavy atom derivatives preparation, grown crystals were transferred to a freshly prepared drop of precipitant solution supplemented with one of the following metals: ammonium tetrachloroplatinate (II), sodium tetrachloroaurate (III) hydrate, sodium hexachloroiridate (III) hydrate, samarium (III) nitrate hexahydrate, mercury (II) bromide, lead chloride, potassium tetracyanoplatinate (II) hydrate, potassium osmate (VI) dehydrate, sodium tungstate dehydrate and potassium perrhenate. Multiple combinations of metal concentration (up to 100 mM) and incubation time (from minutes to several days) were tested. Before data collection, the crystals were transferred to the mother liquor supplemented with 20% (v/v) glycerol and flash frozen in liquid nitrogen.

Crystallization trials of the selenomethionine-substituted 6H-Rsfg-SigG1 complexes were performed using the MembFac™ kit (Hampton Research, USA) in a crystallization robot (Oryx4), each drop contained 1:1 protein/precipitant solution ratios. The plates were incubated at 20 °C. Crystals were obtained after 3 days in MembFac™ condition #29. Crystal optimization was performed using different protein concentrations and different protein:reservoir ratios. The best crystals were obtained with the protein concentration at 10 mg ml^{-1} with the drop ratio of 3:1 (protein: crystallization solution). Crystals were soaked in mother liquor containing 25% (w/v) glycerol and flash-frozen in liquid nitrogen for diffraction experiments.

Table 3.6 Crystallization conditions of diffracting SigG1 crystals

Method	Hanging drop
Temperature (K)	20 ˚C
Protein concentration (mg. mL^{-1})	7.5

Buffer composition of protein solution	20 mM sodium phosphate pH 7.4, 10-15 mM NaCl
Composition of reservoir solution	SaltRx™1 screen #1: 1.8 M Sodium acetate tri-hydrate pH 7.0 and 0.1 M BIS-TRIS propane pH 7.0
Volume and ratio of drop	0.3 µl 1:1
Volume of reservoir (µL)	350µl

Table 3.7 Crystallization conditions of diffracting 6H-RsfG-SigG1 crystals

Method	Hanging drop
Temperature (K)	20 ˚C
Protein concentration (mg. mL^{-1})	10
Buffer composition of protein solution	20 mM sodium phosphate pH 7.4, 10-15 mM NaCl
Composition of reservoir solution	MembFac™ #29: 0.1 M Ammonium sulfate; 0.1 M HEPES sodium pH 7.5; 0.5 M sodium phosphate dibasic dihydrate; 0.5 M potassium phosphate dibasic
Volume and ratio of drop	1:1/1:4
Volume of reservoir (µL)	350µl

Table 3.8 Crystallization conditions of diffracting 6H-SeMet-RsfG crystals

Method	Hanging drop
Temperature (K)	20 ˚C
Protein concentration (mg. mL^{-1})	10
Buffer composition of protein solution	10 mM HEPES and 150 mM NaCl, pH 7.5
Composition of reservoir solution	MembFac™ #29: 0.1 M Ammonium sulfate; 0.1 M HEPES sodium pH 7.5; 0.5 M sodium phosphate dibasic dihydrate; 0.5 M potassium phosphate dibasic
Volume and ratio of drop	3:1
Volume of reservoir (µL)	350µl

X-ray diffraction data collection and processing. Data sets of 6H-SigG1 crystals were collected at the Proxima 1 synchrotron beamline (SOLEIL, Gif-sur-Yvette, France). Crystals were processed up to 2.5 Å with unit cell dimensions of a = 92.64, b = 151.84, c = 128.18, α = 90°, β = 90° and γ = 90°, and were indexed to the C 222_1 space group. Data collection and processing statistics are presented in Table 3.2. Diffraction images were processed with the XDS (Kabsch, 1993) or iMosflm (Battye et al., 2011)

software packages and reduced with the CCP4 suite (Bailey, 1994). Molecular replacement was extensively pursued using Phaser MR (McCoy et al., 2007), with multiple search models prepared from PDB entries 6DXO (Schumacher et al., 2018), 5Xe7 (Goutam et al., 2017), 4CXF (Maillard et al., 2014) and 2P7V (Patikoglou et al., 2007). No single solution was found.

Data sets of 6H-RsfG-SigG1 native and heavy-atom soaked crystals were collected at Proxima 1, Proxima 2A (SOLEIL, Gif-sur-Yvette, France) and XALOC (ALBA, Barcelona, Spain) synchrotron beamlines. Native crystals were processed up to 2.3 Å with unit cell dimensions of a = 73.37, b = 61.13, c = 82.75, α = 90°, β = 106.72° and γ = 90°, and were indexed to the P1 21 1 space group. Data collection and processing statistics are presented in Table 3.4. Diffraction images were processed with the XDS (Kabsch, 1993) or iMosflm (Battye et al., 2011) software packages and reduced with the CCP4 suite (Bailey, 1994). Initially, molecular replacement was extensively pursued using Phaser MR (McCoy et al., 2007), with domains of PDB entries 3IA8 (Bianchetti et al., 2011) and 2FWV (Shepard et al., 2007) as search models. As no single solution was obtained, phase determination was carried out with heavy-atom derivatives, with both SAD and MAD measurements collected. The presence of anomalous signal was screened with the SHELXC/SHELXD/SHELXE pipeline (Sheldrick, 2010), but so far no solution was found.

Data sets of 6H-SeMet-RsfG crystals were collected at the Proxima 1 (SOLEIL, Gif-sur-Yvette, France) and XALOC (ALBA, Barcelona, Spain) synchrotron beamlines. However, data was rather poor (visible spots around 7Å). Hence, data indexing and processing could not be achieved. Further cycles of crystal quality improvement and data collection are required.

3.6 Acknowledgements

This work was partially funded by National Funds through FCT- Fundação para a Ciência e a Tecnologia, I.P., under the project ERA-IB-2/0001/2015. It was further supported by FEDER -Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation (POCI), Portugal 2020; and by Portuguese funds through FCT Fundação para a Ciência e a Tecnologia, I.P/Ministério da Ciência, Tecnologia e Ensino Superior POCI-01-0145-FEDER-007274 and NORTE-01-0145-782 FEDER-000012. R.O. was supported by the FCT fellowship SFRH/BD/107862/2015 and by the EMBO award ASTF438-2015, M.V.M. was supported by the FCT fellowship SFRH/BPD/95683/2013 and the FCT contract DL57/2016/CP1355/CT0023 and D.C.P. and G.F. were supported through the IMPRS-Mic and the ERASynBio project ECFexpress (BMBF grant 031L0010B). FL, JPL

and LG were supported by FCT fellowship SFRH/BD/129921/2017, xxx and xxx. This work was further funded by FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI); projects NORTE-01-0145-FEDER-000012—Structured Programme on Bioengineering Therapies for Infectious Diseases and Tissue Regeneration and NORTE-01-0145-FEDER-000008—Porto Neurosciences and Neurologic Disease Research Initiative at i3S, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement; and by Portuguese funds through FCT—Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Ensino Superior in the framework of the project "Institute for Research and Innovation in Health Sciences" (POCI-01-0145-FEDER-007274 and PTDC/BIA-MIC/28779/2017).. The authors are grateful to the staff of the European Synchrotron Radiation Facility (Grenoble, France), SOLEIL (Essonne, France) and ALBA (Barcelona, Spain) synchrotrons. We are grateful to Tatiana Cereija for helpful discussions. We further acknowledge the support of the i3S Scientific Platforms Biochemical and Biophysical Technologies and Proteomics. The authors have no conflict of interest to declare.

3.7 References

Bailey, S. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallographica Section D *50*, 760-763.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallographica Section D *67*, 271-281.

Bianchetti, C.M., Bingman, C.A., and Phillips Jr., G.N. (2011). Structure of the C-terminal heme-binding domain of THAP domain containing protein 4 from Homo sapiens. Proteins: Structure, Function, and Bioinformatics *79*, 1337-1341.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites1 1Edited by F. E. Cohen. Journal of Molecular Biology *294*, 1351-1362.

Casas-Pastor, D., Müller, R.R., Becker, A., Buttner, M., Gross, C., Mascher, T., Goesmann, A., and Fritz, G. (2019). Expansion and re-classification of the extracytoplasmic function (ECF) σ factor family. bioRxiv, 2019.2012.2011.873521.

Gallagher, K.A., Schumacher, M.A., Bush, M.J., Bibb, M.J., Chandra, G., Holmes, N.A., Zeng, W., Henderson, M., Zhang, H., Findlay, K.C.*, et al.* (2019). c-di-GMP Arms an Anti-σ to Control Progression of Multicellular Differentiation in *Streptomyces*. Molecular Cell *77*, 586-599.e586.

Goutam, K., Gupta, A.K., and Gopal, B. (2017). The fused SnoaL-2 domain in the *Mycobacterium tuberculosis* sigma factor σjmodulates promoter recognition. Nucleic Acids Research *45*, 9760-9772.

Gupta, R., and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput, 310-322.

Gupta, R., Jung, E. and Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. In preparation.

Hofmann, K., and Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler *374*, 166.

Huang, X., Pinto, D., Fritz, G., and Mascher, T. (2015). Environmental Sensing in Actinobacteria: a Comprehensive Survey on the Signaling Capacity of This Phylum. Journal of Bacteriology *197*, 2517-2535.

Iyer, S.C., Casas-Pastor, D., Kraus, D., Mann, P., Schirner, K., Glatter, T., Fritz, G., and Ringgaard, S. (2020). Transcriptional regulation by σ factor phosphorylation in bacteria. Nature Microbiology.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. Journal of Applied Crystallography *26*, 795-800.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols *10*, 845-858.

Kim, J.-E., Choi, J.-S., Kim, J.-S., Cho, Y.-H., and Roe, J.-H. (2020). Lysine acetylation of the housekeeping sigma factor enhances the activity of the RNA polymerase holoenzyme. Nucleic Acids Research.

Li, A., Xue, Y., Jin, C., Wang, M., and Yao, X. (2006). Prediction of Nε-acetylation on internal lysines implemented in Bayesian Discriminant Method. Biochemical and biophysical research communications *350*, 818-824.

Liu, Z., Ma, Q., Cao, J., Gao, X., Ren, J., and Xue, Y. (2011). GPS-PUP: computational prediction of pupylation sites in prokaryotic proteins. Molecular bioSystems *7*, 2737-2740.

Maillard, A.P., Girard, E., Ziani, W., Petit-Härtlein, I., Kahn, R., and Covès, J. (2014). The Crystal Structure of the Anti-σ Factor CnrY in Complex with the σ Factor CnrH Shows a New Structural Class of Anti-σ Factors Targeting Extracytoplasmic Function σ Factors. Journal of Molecular Biology *426*, 2313-2327.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of Applied Crystallography *40*, 658-674.

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., and Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proceedings of the National Academy of Sciences *108*, E1293-E1301.

Oliveira, R., Bush, M., Pires, S., Chandra, G., Casas-Pastor, D., Fritz, G., and Mendes, M. (2020). *Streptomyces tsukubaensis* differentiation is aided by the redox responsive σ/anti-σ pair, SigG1/RsfG. Applied and Environmental Microbiology.

Österberg, S., del Peso-Santos, T., and Shingler, V. (2011). Regulation of alternative sigma factor use. Annu Rev Microbiol *65*, 37-55.

Park, S.T., Kang, C.-M., and Husson, R.N. (2008). Regulation of the SigH stress response regulon by an essential protein kinase in Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences *105*, 13105-13110.

Patikoglou, G.A., Westblade, L.F., Campbell, E.A., Lamour, V., Lane, W.J., and Darst, S.A. (2007). Crystal Structure of the Escherichia coli Regulator of σ70, Rsd, in Complex with σ70 Domain 4. Journal of Molecular Biology *372*, 649-659.

Prisic, S., Dankwa, S., Schwartz, D., Chou, M.F., Locasale, J.W., Kang, C.M., Bemis, G., Church, G.M., Steen, H., and Husson, R.N. (2010). Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. Proceedings of the National Academy of Sciences *107*, 7521-7526.

Schumacher, M.A., Bush, M.J., Bibb, M.J., Ramos-León, F., Chandra, G., Zeng, W., and Buttner, M.J. (2018). The crystal structure of the RsbN–σBldN complex from Streptomyces venezuelae defines a new structural class of anti-σ factor. Nucleic Acids Research *46*, 7405–7417.

Sheldrick, G. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallographica Section D *66*, 479-485.

Shepard, W., Haouz, A., Graña, M., Buschiazzo, A., Betton, J.-M., Cole, S.T., and Alzari, P.M. (2007). The Crystal Structure of Rv0813c from *Mycobacterium tuberculosis* Reveals a New Family of Fatty Acid-Binding Protein-Like Proteins in Bacteria. Journal of Bacteriology *189*, 1899-1904.

Steentoft, C., Vakhrushev, S.Y., Joshi, H.J., Kong, Y., Vester-Christensen, M.B., Schjoldager, K.T.-B.G., Lavrsen, K., Dabelsteen, S., Pedersen, N.B., Marcos-Silva, L.*, et al.* (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. Embo J *32*, 1478-1488.

Striebel, F., Imkamp, F., Özcelik, D., and Weber-Ban, E. (2014). Pupylation as a signal for proteasomal degradation in bacteria. Biochimica et Biophysica Acta - Molecular Cell Research *1843*, 103-113.

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. Proceedings of the National Academy of Sciences *106*, 67-72.

Wu, H., Liu, Q., Casas-Pastor, D., Dürr, F., Mascher, T., and Fritz, G. (2019). The role of C-terminal extensions in controlling ECF σ factor activity in the widely conserved groups ECF41 and ECF42. Molecular Microbiology *112*, 498-514.

Chapter 4

The *S. venezuelae* ECF56

Rute Oliveira, [a, b, c], Matthew J. Bush, [e], Govind Chandra, [e], and Marta V. Mendes, [a, b, d]

[a]Bioengineering and Synthetic Microbiology Group, i3S- Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

[b]IBMC, Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal

[c]Programa Doutoral em Biologia Molecular e Celular (MCBiology), ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Rua de Jorge Viterbo Ferreira n. 228, 4050-313 Porto, Portugal

[d]ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Rua de Jorge Viterbo Ferreira n. 228, 4050-313 Porto, Portugal

[e]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, 13 Norwich NR4 7UH, United Kingdom

4.1. Abstract

While holding extraordinary levels of plasticity, bacteria also keep great conservation of regulatory mechanisms. In particular, the regulation of transcription initiation by extracytoplasmic sigma factors (ECFs) has been systematically described as highly conserved within the Actinobacteria phylum, especially in terms of genomic neighbourhood and regulation by accessory proteins (Casas-Pastor et al., 2019; Huang et al., 2015; Staroń et al., 2009). By studying the model organism *Streptomyces venezuelae,* we addressed the conservation of the physiological role and mode of action of SigG2, a novel ECF with a SnoaL_2 C-terminal (CT) extension (PF12680). In this work, we report that unlike *Streptomyces tsukubaensis* SigG1, and despite the remarkable protein domain conservation observed, SigG2 might not be essential for metal homeostasis during the progression of *Streptomyces* differentiation. We show how we have investigated the function of SigG2 and weather it has an accessory co-regulatory partner. Our evidence supported the existence of an interacting protein with no conserved domains that we established as a candidate for playing the antagonist role for SigG2.

4.2. Introduction

Whilst harbouring the most intriguing and unique molecules in nature, most industrially relevant species, like *S. tsukubaensis*, represent a challenge for cell and molecular biology studies, as the biosynthesis of the bioactive compounds in laboratory requires a strictly regulated environment. With more relaxed growth conditions and a variety of molecular tools available, other *Streptomyces* strains are frequently used as model organisms, in particular when we intend to study fundamental molecular mechanisms. *S. venezuelae* is a model organism with a fast growth rate and an accessible genetic tractability due to, for instance, its high transformation efficiencies. Moreover, *S. venezuelae* undergoes an entire developmental cycle in liquid culture, in contrast to *S. tsukubaensis* that, in the conditions tested, differentiates into complete chains of spores exclusively on solid medium. This feature is particularly valuable to facilitate the application of molecular biology techniques for the study of processes related to multicellular morphological differentiation.

The regulatory processes that culminate in *Streptomyces* differentiation have been widely studied in *S. venezuelae* (see Figure 1.3 in Chapter 1). A well-documented array of regulators control the developmental transitions that compose the multicellular lifecycle (Bush et al., 2015). It comprises, mainly, a network of transcription factors like the Bld and the Whi regulators, but it also includes sigma and anti-factors. For instance, the sigma factor WhiG (and its antagonist, RsiG) involved in controlling the trigger

sporulation, or BldN and its cognate anti-sigma factor, RsbN, which control the expression of spores hydrophobic sheath proteins, the chaplins and the rodlins (Bibb et al., 2012)(Gallagher et al., 2019).

In *S. venezuelae* NRRL B-65442 chromosome, we identified 45 sigma factors of which 32 are ECFs. To our knowledge, only two ECFs have been studied in these bacteria: ECF121 BldN and ECF42 in the ECF42 family. ECF42 has been shown to perform a positive auto-regulatory function through interaction of its CT TPR extension with the ECF N-terminal core, but no clear physiological role has been assigned to this ECF (Liu et al., 2018). While *S. tsukubaensis*, might benefit from possible redundancy among the four SnoaL_2 containing ECFS, *S. venezuelae* NRRL B-65442 harbours only one ECF56 with the SnoaL_2 CT extension, which is the orthologue of *S. tsukubaensis* SigG2.

ECFs are responsible for the activation of specific genes necessary for environmental stress responses. However, some ECFs have been characterised that react to developmental cues and have an important role in defining the mycelia structure (Bibb et al., 2012; Oliveira et al., 2020). This study focuses on the role and regulation of SigG2, the *S. venezuelae* orthologue of *S. tsukubaensis* SigG1, a developmental regulator that is triggered by intracellular oxidative stress to control the timely progression of differentiation. Here we report our investigations the role of this uncharacterised ECF in morphological differentiation. Although unique ECFs with CT extensions are thought to auto-regulate their activity, most ECFs rely on the arrest of its runaway activity by an antagonist regulator, the anti-sigma factor. We show that SigG2 might interact with accessory regulators to control its release and activity.

4.3. Results

4.3.1. *S. venezuelae* harbours one SnoaL_2 containing ECF, SigG2. We identified *vnz_RS22720* as the only gene encoding for a SnoaL_2-containing ECF in the *S. venezuelae* NRRL B-65442 chromosome. It encodes a 323 amino acid-ECF with a predicted molecular weight of 35521.40 Da and theoretical isoelectric point of 4.75. It shares homology with SigG1 (38%) and SigG2 (75%) identified in the *S. tsukubaensis* genome and therefore we named this ECF as SigG2. SigG2 contains the σ2 and σ4 ECF core conserved domains followed by a SnoaL_2 CT domain. Additionally, it exhibits a predict signal peptide, from residue L262 to residue A274, of unknown function.

Studies on the evolution of ECFs have been able to segregate hundreds of ECFs in different families and predict specific traits that might be characteristic of the members of each family (Casas-Pastor et al., 2019; Huang et al., 2015; Staroń et al., 2009). One of the most studied features is the genomic context in which a particular ECF is located, concerning, for example, if neighbouring genes encode candidate anti-sigma factors, or proteins of the molecular signature for a particular biological process. For instance, for the ECF42 family above mentioned, about 90% of the ECFs are neighbouring DGPF proteins (YciI superfamily, PF03795), and to a smaller extent, they can be located nearby genes that code for glyoxalase/bleomycin resistance proteins/dioxygenases, acyl-CoA dehydrogenases, ketosteroid isomerases or other uncharacterized conserved proteins. Corroborating an auto-regulatory activity of the TPR CT extension, no obvious anti-sigma factor genes were identified near ECF42 proteins. When analysing the genomic region of SigG2 orthologues in 98 *Streptomyces* genomes, we found that SigG2-like proteins were positioned in the same genomic region as LysR-type regulators of unknown function in 82% of the cases, (Figure S4.1). No obvious genes encoding for anti-sigma factors were found in this region, however, in the majority of the cases there was a hypothetical protein encoded within an antisense downstream gene. Making use of a genome-wide study that defined the transcription start sites (TSS) in *S. venezuelae*, we observed that *sigG2* is transcribed from a single TSS – located 48bp upstream the predicted start codon (Figure 4.1) – whereas the downstream gene, *vnz_RS22715,* might be transcribed from an independent TSS, since its positioned in the reverse orientation in the genome.

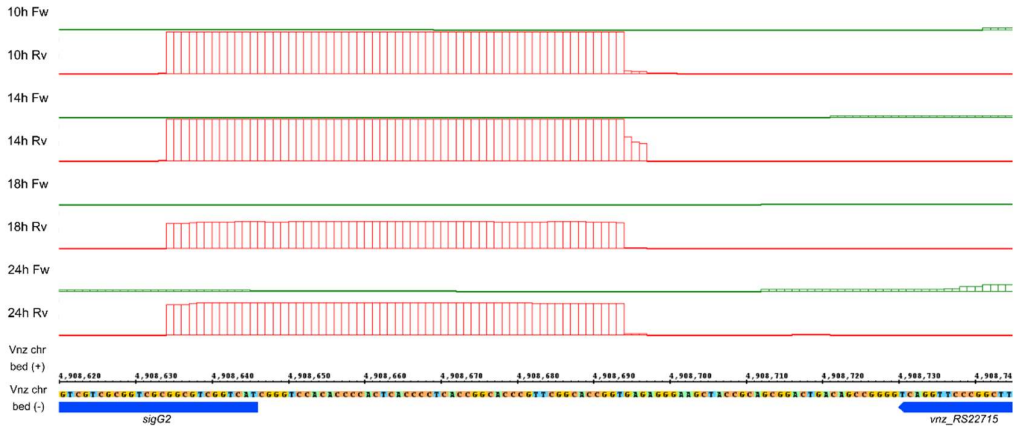


Figure 4.1. Genome browser viewer showing the *sigG2* TSS throughout growth, obtained with a genome-wide study for the determination of TSS *in S. venezuelae* genome. This information was retrieved from a larger dataset of TSS in *S. venezuelae,* from Mark Buttner's lab. Differentiation stages: vegetative growth (10-14h), mycelia fragmentation/onset of sporulation (14-18h) and sporulation (18-24h).

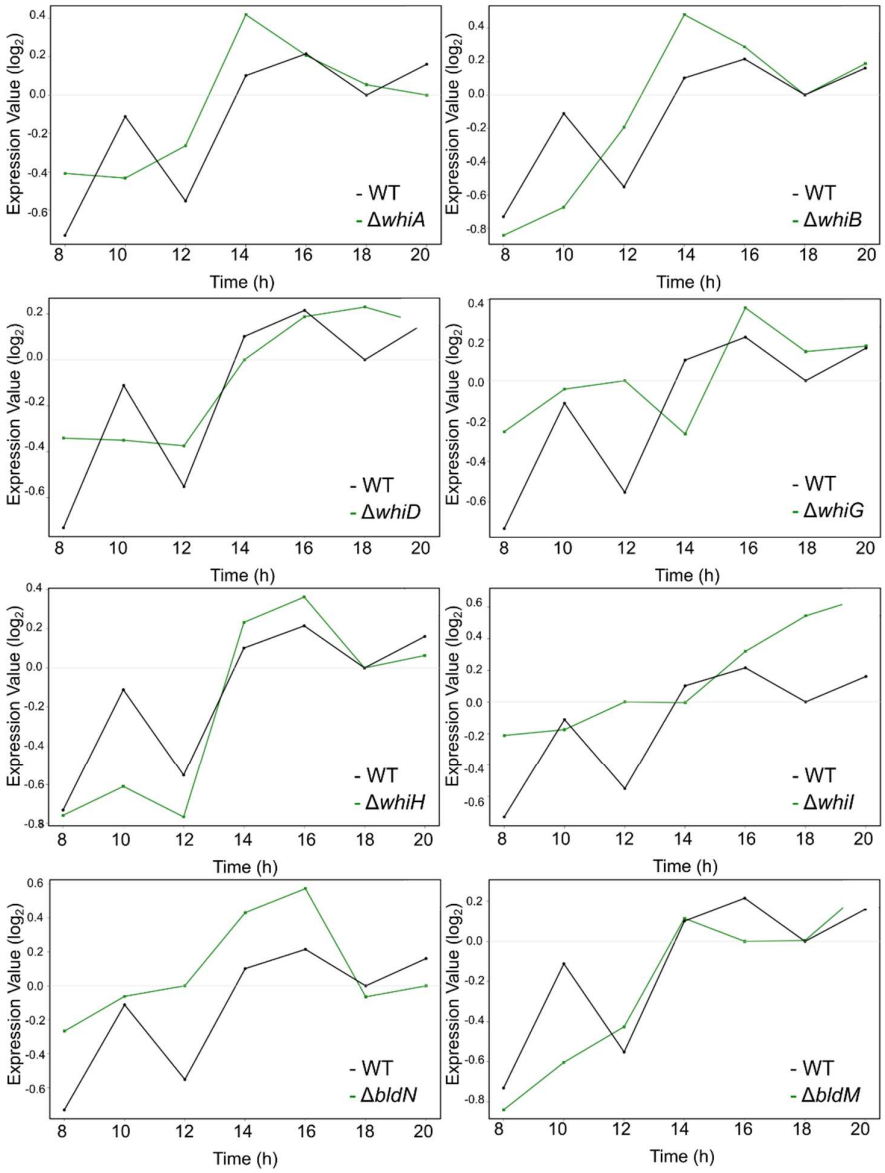


Figure 4.2. Transcriptional profiling experiments of *sigG2* in cells of *S. venezuelae* WT, Δ*whiA*, Δ*whiB*, Δ*whiG*, Δ*whiH,* Δ*whiI,* Δ*bldM* and Δ*bldN* Δ*whiH.* Expression was evaluated throughout growth in MYM liquid cultures. Data in these plots was extracted from experiments deposited in the ArrayExpress database as E-MTAB-4673, E-MTAB-8114, E-MTAB-5853, E-MTAB-2716 and E-MEXP-3612, respectively (Al-Bassam et al., 2014; Bibb et al., 2012; Bush et al., 2013; Bush et al., 2016; Gallagher et al., 2019). Bibb MJ, Buttner MJ and Chandra G obtained data concerning expression of *sigG2* in the Δ*whiD* background (unpublished results). Differentiation stages: vegetative growth (8-14h), mycelia fragmentation/onset of sporulation (14-16h) and sporulation (18-20h).

4.3.2. Expression of *sigG2* is strongly upregulated during the transitions that lead to aerial differentiation. We examined microarray expression data available

for *S. venezuelae* wild type and a set of different backgrounds of deletion mutants of characterised developmental regulators (Figure 4.2). We assessed the transcript levels of *sigG2* in differentiating cultures in different growth stages – vegetative growth (8-14h), mycelia fragmentation/onset of sporulation (14-16h) and sporulation (18-20h) – for 20h (when sporulation was near to completion). We observed two-steps of *sigG2* upregulation in wild type cells that seem to follow the transitions between growth stages: a first step of transcription activation during early vegetative growth, and a second step of activation in a later stage of development that could correspond to hyphae septation. Despite the overexpression seen in the Δ*whiA,* Δ*whiB* and Δ*bldN* backgrounds, the two-step pattern of expression disappears in the absence of important *whi* and *bld* classical regulators.

4.3.3. Physiological characterisation of *sigG2* deletion mutants. We generated *S. venezuelae* Δ*sigG2* mutants by gene replacement using the REDIRECT technology as described before (see Chapter 2 supplemental information II for details) and confirmed the deletion by PCR using primers listed in Table S4.3. Having in mind that the SnoaL_2 ECF we previously described had a role in the progression of *Streptomyces* differentiation, we characterized the mutant phenotypes focusing on possible effects on morphological development. Preliminary physiological characterisation of a *sigG2* deletion mutant has turned out inconclusive, since no clear differentiation phenotypes were observed in cultures in MYM liquid media or on MYM agar (Figure 4.3). Strikingly, there was a significant difference in the colouring of the mycelia when we compared cells grown in MYM-tap and cells grown in MYM-dH_2O in liquid media. Nevertheless, S. venezuelae hyphae failed to sporulated in MYM-dH_2O

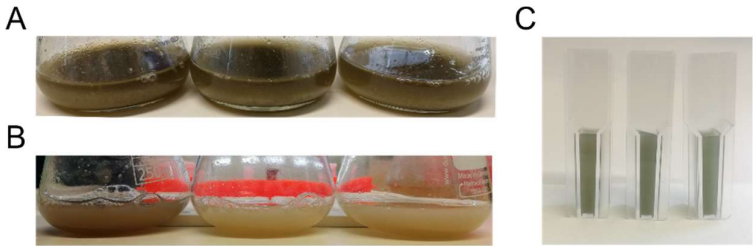


Figure 4.3. Phenotypes of *S. venezuelae* wild type and Δ*sigG2* deletion mutants in liquid cuçtures. Strains were cultured for 24h at 30˚C, in MYM-tap liquid medium (50% dH_2O and 50% tap water), as routinely done for *S. venezuelae* (**A**). Surprisingly, when cells were cgrown in MYM made with 100% dH_2O water, Δ*sigG2* cultures exhibited a lighter colour of the mycelia indicating the lack of the greenish pigmentation of the wild type strain. However, *S. venezuelae* hyphae failed to sporulated in these conditions (**B**). Photographs of *S. venezuelae* wild type and *sigG2*-depleted cultures showing no obvious differences in the phenotype when grown in MYM-tap. Cells were

grown for 16h at 30˚C. A small vh at olume of each culture was dispensed in 1.5 ml cuvettes before image acquisition (**C**).

cultures, hence we carried ou further experiments aiming to evaluate the progression of differentiation in MYM-tap medium. We additionally tested growth in different carbon and nitrogen sources using defined media with starch (ISP4); maltose, yeast extract and malt extract (MYM); soya flour (SFM) and (2PYE) (Figure 4.4). The Δ*sigG2* mutant resembled the wild type phenotype in the conditions used, indicating that loss of sigG2 might not result in a similar delay in morphological development observed for *S. tsukubaensis* Δ*sigG1.*

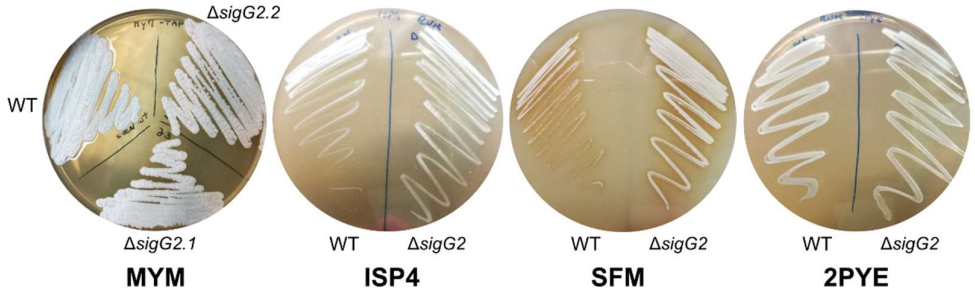


Figure 4.4. Phenotypes of *S. venezuelae* wild type and the *sigG2*-deletion strain on solid substrates with different carbon or nitrogen sources. Defined media with maltose, yeast extract and malt extract (MYM); starch (ISP4); soya-flour (SFM); peptone and yeast extract (2PYE) were used.

4.3.4. SigG2 is not involved in the maintenance of iron homeostasis. Given that SigG1 targets a regulon that controls metal metabolism, we sought to investigate if SigG2 was also involved in the maintenance of iron homeostasis. Evaluation of total iron content also showed no differences in the iron levels in Δ*sigG2* culture supernatants when compared to the wild type (Figure 4.5).

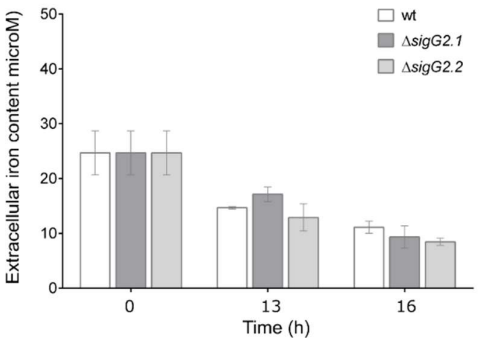


Figure 4.5. Total extracellular iron content of the wild type and *sigG2*-depleted strains. The availability of iron in the culture media was evaluated throughout growth in wild type and Δ*sigG2* cultures. Total iron levels were measured using the QuantiChrom™ assay during the initial stages of growth until a late exponential phase.

4.3.5. Vnz_RS21305 is a candidate SigG2 interacting partner. The downstream gene *vnz_RS22715* is located in the antisense orientation relatively to *sigG2* and therefore it cannot be co-transcribed with *sigG2.* When examining *vnz_RS22715* expression throughout growth it does not seem to fluctuate significantly with the transitions between growth stages. Moreover, we could not establish a correlation between *vnz_RS22715* and *sigG2* transcript levels that could point towards a repressor activity of Vnz_RS22715 over SigG2 (Figure 4.6).

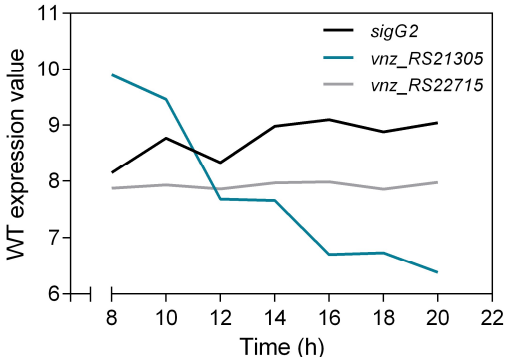


Figure 4.6. Transcript profilling of *sigG2*, *vnz_RS21305* and *vnz_RS22715* in *S. venezuelae* wild type differentiating cultures. Expression data was extracted from the transcription array data (E-MTAB-2716) in (Al-Bassam et al., 2014).

To identified a putative antagonist for SigG2 we sought to find an interacting protein that could have a regulatory effect over SigG2. We assessed protein-protein interaction, measured by bacteial adenylate cyclase (BACTH) assays (Figure 4.7). The full-length SigG2 protein was cloned into a bait vector of the BACTH system to assess the capacity for physical interaction with other members of the *S. venezuelae* proteome. Using the T25-SigG2 as bait against a *S. venezuelae* T18C-library we obtained an array of positive signals measured by the blue colour of the β-galactosidase chromogenic substrate. We isolated about 100 clones of positive clones, expanded each clone in liquid culture and performed minipreps to isolate the plasmids with the positive hit. The resultant vectors were then sequenced to obtain the identity of the positive preys. We retrieved a list of 60 fragments of the *S. venezuelae* library with sequences that were in frame with the T18 tag and therefore could code for annotated proteins. Selected candidates that could be of interest to this work are displayed in Table 4.1. We further tested 25 clones for the T18C-SigG2 construct against the T25-library but the results of these experiments only retrieved uncharacterised hypothetical proteins. Of the total positive hits, only the fragment corresponding to the coding sequence for Vnz_RS21305 was reproduced in a second independent screening experiment, indicating that SigG2 and Vnz_RS21305 physically interact *in vivo*. Data on gene expression in the wild type revealed a downregulation of the *vnz_RS21305* transcript levels throughout growth that could be consistent with a suppression of a putative anti-sigma factor repressor, at the culture stages when *sigG2* is activated (*e.g.* time points 10-12h in Figure 4.6). Among

the total candidates, we found several proteins of interest such as regulators of oxidative stress responses or proteases. In addition, we found regulators of differentiation like ParB, which is involved in the timely septation of the filamentous hyphae. When in

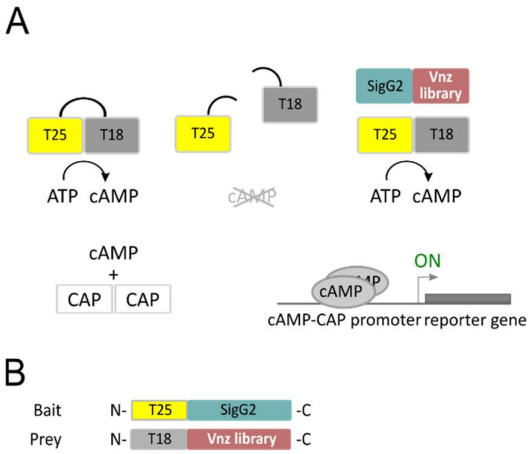


Figure 4.7. Schematic representation of the bacterial two-hybrid system used to screen the *S. venezuelae* library for proteins that interact with SigG2.

Table 4.1. Selected candidates for protein-protein interaction with SigG2 retrieved by BACTH

Prey Locus_ID	Description
Vnz_RS21305 [1]	LPXTG cell wall anchor domain-containing protein
Vnz_RS04905	Ankyrin-repeated domain containing protein
Vnz_RS08380	Phosphoribosyl-AMP cyclohydrolase
Vnz_RS08770	Phenazine-specific anthranilate synthase
Vnz_RS11930	Aminopeptidase N, PepN
Vnz_RS16410	dTMP kinase
Vnz_RS18150	ParB/RepB/Spo0J family partition protein, ParB
Vnz_RS18360	Serine-threonine protein kinase
Vnz_RS21185	Metal dependente hydrolase
Vnz_RS23825	Ferredoxin
Vnz_RS26940	Histidine kinase
Vnz_RS26970	Protein Implicated in DNA repair function with RecA and MutS, CinA
Vnz_RS30890	Phenylacetic acid degradation protein, PaaA
Vnz_RS38695	Hypotethical small protein, 6kDa

[1] Obtained in two independent screening experiments

complex with ParA, it distributes chromosomes evenly throughout the sporogenic hyphae that will later differentiate into chains of unigenomic spores (Donczew et al., 2016). In order to confirm these interactions, we sought to clone the full-length sequence

of relevant prey and test it in BACTH assays. In a first attempt to validate the screening results, and to test interaction with SigG2 we performed one-to-one BACTH the experiments using both N and C-terminal fusions of ParA and ParB with T18 and T25 fragments against the full-length constructs of SigG2. SigG2 did not show interaction with ParA or ParB. Further validation of the interaction of SigG2 with Vnz_RS21305, or with other putative interactors will give more insights on the SigG2 regulatory network.

4.4 Discussion

Like in other Streptomycetes genomes, we were able to identify a high number of ECFs in the *S. venezuelae* genome, which, in our opinion, could reflect its fine-tuned regulation of an intricate developmental cycle and fast adaptation to stressful environments. Of the total ECFs encountered, seven proteins have shown homology to proteins of the NTF2 superfamily (IPR032710) but only one exhibits the conserved Pfam domain SnoaL_2 (PF12680). The investigations described in this chapter focussed on the role of this new member of the ECF56 family in *S. venezuelae*, which is the closest orthologue to SigG1 described in *S. tsukubaensis* (Oliveira, 2020). After preliminary functional characterisation of a *sigG2*-deletion mutant, we concluded that this copy of a SnoaL_2-ECF might be important in different biological processes than the ones regulated by SigG1 in *S. tsukubaensis*, since it does not exhibit clear difference in morphology or iron availability, when compared to the wild type strain.

The similar protein domain composition prompted us to evaluate the mechanism of action of this ECF and see if it had any points of contact to the mode of action described for SigG1. Careful analysis of the SigG2 sequence revealed that the polyketide cyclase active residues of the SnoaL-like enzymes are absent in SigG2 SnoaL_2. This substantiates that the SnoaL_2 domain in ECF56 group might be a sensor domain with no catalytic activity. We further explored protein-protein interaction to search for a putative posttranslational regulator for this ECF. In order to investigate this we screened a bacterial-two hybrid (BACTH) library for potential SigG2 regulatory partners. Bacterial two-hybrid systems present the probability of protein-protein interaction in an *E. coli* heterologous host under standard growth conditions. Such system does not account for time/condition of expression, stability of the proteins, or the availability of specific ligands that would interfere with the interaction. A more targeted approach in *Streptomyces* would help to discard false positives and unveil stronger candidates with potential to interact with SigG2 in order to enhance our understanding of the SigG2 regulatory network. Nonetheless, we identified Vnz_RS21305 as a strong candidate SigG2 interaction partner. Vnz_RS21305 is an uncharacterised protein that contains a LPXTG cell wall anchor domain. Interstingly, like *S. tsukubaensis* RsfG, it contains a predicted

23 aa transmembrane domain between residues 435D and 457A, which could be consistent with the topology described for known anti-sigma factors. In addition, time-resolved transcription profiling showed a downregulation of *vnz_RS21305* expression that seems to follows the activation of *sigG2*. Together with the BACTH experiments, this indicates that this could be a good candidate SigG2 anti-sigma factor. For a functional characterisation, and to assess if this could be the antagonist for SigG2, further investigation is required through the generation of deletion mutant strain for this gene, and through addiotnal biochemical assays to characterise the interaction between SigG2 and the putative regulator.

The insights gained with this preliminary work suggest that we might be in the presence of a novel ECF56 that is regulated by accessory proteins, despite having a long C-terminal extension with a NTF2-like domain. Whether SigG2 works in concert with a cognate anti-sigma factor is still to be determined. Although the specific trigger and functional role of SigG2 remain elusive, through gene expression data we show that it is conceivable that is implicated in the molecular cascades that regulate developmental transitions in *Streptomyces*. This could indicate that, although these ECFs might carry different specific roles in the cell, signal-transducing cascades from SnoaL_2-containing sigma factors are important in morphological differentiation.

Previously, works have shown that the *M. tuberculosis* SigG2 orthologue SigG is indirectly involved in the response to DNA damage stress during *M. tuberculosis* macrophage infection (Gaudion et al., 2013; Lee et al., 2008; Smollett et al., 2011). More recently, one work on the activity of a truncated version of SigG2 (which lacks the C-terminus SnoaL_2 domain) has described that the SigG2 σ2-σ4 core is capable of activating transcription *in vitro* (Thompson, 2020). Through transcriptomics analysis of a *S. venezuelae* strain overexpressing the N-terminal SigG2 σ2-σ4 core, the authors predicted that SigG2 guides the transcription of a large regulon that comprises regulators of the DNA damage response. Specific stages of transcription activation after fading of the vegetative mycelia. Further examination of SigG2 loss or gain of function mutants, combined with studies on SigG2 direct binding to promoters would be essential to determine the direct cascades under the regulation SigG2. Moreover, additional investigation of the function of the SnoaL_2 domain would be crucial to define the mechanism of regulation, and define whether it works in sigma/anti-sigma pair, or if it relies exclusively on autoregulation trough the C-terminus SnoaL_2 domain.

4.5 Materials and Methods

Bacterial strains, growth conditions and preparation of *S. venezuelae* spores. The strain used in this work was *Streptomyces venezuelae* available in Mark

Buttner's lab at the Molecular Microbiology department of the John Innes Centre, UK. This strain is deposited in the NRRL culture collection under stock code NRRL B-65442. The 8kb genome was fully sequenced and the sequence was submitted to the NCBI database under reference NZ_CP018074.1 (Bush et al., 2019). It revealed a 158kb linear plasmid (NZ_CP018075.1) in addition to the linear chromosome. The origin of NRRL B-65442 is "unsure", and is supposed to derive from the original 1940's *Streptomyces venezuelae* soil isolates - the Caracas/Yale/Burkholder strain – or from the Illinois Agricultural Experiment Station at Urbana/Gottlieb strain. Strains were generally grown in MYM liquid medium or on MYM agar, unless otherwise stated. Spores suspensions were prepared from single colonies in 20% glycerol.

Construction of deletion mutant strains with the REDIRECT technology. *E. coli* BW25113 containing a λ RED plasmid, pIJ790, was used to replace the native alleles on the PL1-J3 (Δ*sigG2.1*) and PL1-D8 (Δ*sigG2.1*) cosmids with the apr-oriT cassette amplified from the pIJ773 template plasmid, using specific oligonucleotides. *E. coli* ET12567 containing pUZ8002 was used for λ RED-mediated recombineering (Gust 2002) upon conjugation experiments. Conjugations between *E. coli* and *S. venezuelae* were carried out as described in (Kieser et al., 2000). PCR analysis confirmed the expected deletions in the chromosome.

Transcriptional profiling data. Experimental *sigG2* transcription start sites were obtained from a Cappableseq library kindly provided by Matt Bush at the Molecular Microbiology Department of the John Innes Centre. Transcriptional microarray profilling data in cells of *S. venezuelae* WT, Δ*whiA*, Δ*whiB*, Δ*whiG*, Δ*whiH,* Δ*whiI,* Δ*bldM* and Δ*bldN* Δ*whiH* are deposited in the ArrayExpress database as E-MTAB-4673, E-MTAB-8114, E-MTAB-5853, E-MTAB-2716 and E-MEXP-3612, respectively (Al-Bassam et al., 2014; Bibb et al., 2012; Bush et al., 2013; Bush et al., 2016; Gallagher et al., 2019). Bibb MJ, Buttner MJ and Chandra G at the Molecular Microbiology Department of the John Innes Centre, obtained data concerning expression of *sigG2* in the Δ*whiD* background (unpublished results).

Quantification of extracellular iron content. Extracellular iron content was determined throughout growth in S. venezuelae differentiating cultures in MYM-tap medium. Supernatants of these cultures were to measure total iron using with the QuantiChrom™ Iron Assay Kit (DIFE-250).

Bacterial two-hybrid screening assays. The BACTH assays were performed ad described previously (see Chapter 2). BTH101 cells were chemically transformed with the T25/T18 bait vector containing the full-length sequence for SigG2. We prepared electrocompetent cells expressing the bait vector and transformed with the pUT18C or pKT25 S. venezuelae libraries. Cells were grown in M63 minimal media supplemented

with 0.3% lactose, 0.5 mM IPTG, 40 µg/ml 5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (X-Gal), 50 µg/ml carbenicillin, and 25 µg/ml kanamycin. To evaluate the range of the interaction, positive clones were streaked onto MacConkey agar plates supplemented with 0.3% lactose, 0.5 mM IPTG, 40 µg/ml 5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (X-Gal), 50 µg/ml carbenicillin, and 25 µg/ml kanamycin. Plates were incubated at 30˚C for 10 days. Positive clones were expanded in LB cultures for miniprep isolation of the plasmid containing the positive candidate. Sequencing of the isolated plasmids revealed the identity of the coding sequence interacting protein. Susan Schlimpert at the Molecular Microbiology Department of the John Innes Centre kindly provided BACTH vectors containing the ParA and ParB coding sequences.

4.6 Acknowledgements

We acknowledge the information on *sigG2* TSSs and the transcriptomics data on *vnz* genes in the different backgrounds (shown in figures and 4.1, 4.2 and 4.6) that were kindly provided by Mark Buttner's lab at the Molecular Microbiology Department of the John Innes Centre (see methods section for details). We thank Govind Chandra for making the plots displayed in figure 4.2. The *S. venezuelae* BACTH library was provided by Mark Buttner's lab. We also thank Susan Schlimpert at the Molecular Microbiology Department of the John Innes Centre for providing the BACTH vectors containing the ParA and ParB coding sequences.

4.7 References

Al-Bassam, M.M., Bibb, M.J., Bush, M.J., Chandra, G., and Buttner, M.J. (2014). Response Regulator Heterodimer Formation Controls a Key Stage in *Streptomyces* Development. PLoS Genet *10*, e1004554.

Bibb, M.J., Domonkos, Á., 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. Molecular Microbiology *84*, 1033-1049.

Bush, M.J., Bibb, M.J., Chandra, G., Findlay, K.C., and Buttner, M.J. (2013). Genes required for aerial growth, cell division, and chromosome segregation are targets of WhiA before sporulation in Streptomyces venezuelae. mBio *4*, e00684-00613.

Bush, M.J., Chandra, G., Al-Bassam, M.M., Findlay, K.C., and Buttner, M.J. (2019). BldC Delays Entry into Development To Produce a Sustained Period of Vegetative Growth in Streptomyces venezuelae. mBio *10*, e02812-02818.

Bush, M.J., Chandra, G., Bibb, M.J., Findlay, K.C., and Buttner, M.J. (2016). Genome-Wide Chromatin Immunoprecipitation Sequencing Analysis Shows that WhiB Is a Transcription Factor That Cocontrols Its Regulon with WhiA To Initiate Developmental Cell Division in Streptomyces. mBio *7*, e00523-00516.

Bush, M.J., Tschowri, N., Schlimpert, S., Flärdh, K., and Buttner, M.J. (2015). c-di-GMP signalling and the regulation of developmental transitions in streptomycetes. Nature Reviews Microbiology *13*, 749-760.

Casas-Pastor, D., Müller, R.R., Becker, A., Buttner, M., Gross, C., Mascher, T., Goesmann, A., and Fritz, G. (2019). Expansion and re-classification of the extracytoplasmic function (ECF) σ factor family. bioRxiv, 2019.2012.2011.873521.

Donczew, M., Mackiewicz, P., Wróbel, A., Flärdh, K., Zakrzewska-Czerwińska, J., and Jakimowicz, D. (2016). ParA and ParB coordinate chromosome segregation with cell elongation and division during *Streptomyces* sporulation. Open Biology *6*, 150263.

Gallagher, K.A., Schumacher, M.A., Bush, M.J., Bibb, M.J., Chandra, G., Holmes, N.A., Zeng, W., Henderson, M., Zhang, H., Findlay, K.C., *et al.* (2019). c-di-GMP Arms an Anti-σ to Control Progression of Multicellular Differentiation in Streptomyces. Molecular Cell *77*, 586-599.e586.

Gaudion, A., Dawson, L., Davis, E., and Smollett, K. (2013). Characterisation of the *Mycobacterium tuberculosis* alternative sigma factor SigG: its operon and regulon. Tuberculosis *93*, 482-491.

Huang, X., Pinto, D., Fritz, G., and Mascher, T. (2015). Environmental Sensing in Actinobacteria: a Comprehensive Survey on the Signaling Capacity of This Phylum. Journal of Bacteriology *197*, 2517-2535.

Kieser, T., Bibb, M., Buttner, M., Chater, K., and Hopwood, D.A. (2000). Practical *Streptomyces* Genetics (John Innes Centre, Norwich, United Kigdom).

Lee, J.-H., Geiman, D.E., and Bishai, W.R. (2008). Role of Stress Response Sigma Factor SigG in *Mycobacterium tuberculosis*. Journal of Bacteriology *190*, 1128-1133.

Liu, Q., Pinto, D., and Mascher, T. (2018). Characterization of the Widely Distributed Novel ECF42 Group of Extracytoplasmic Function σ Factors in Streptomyces venezuelae. Journal of Bacteriology *200*, e00437-00418.

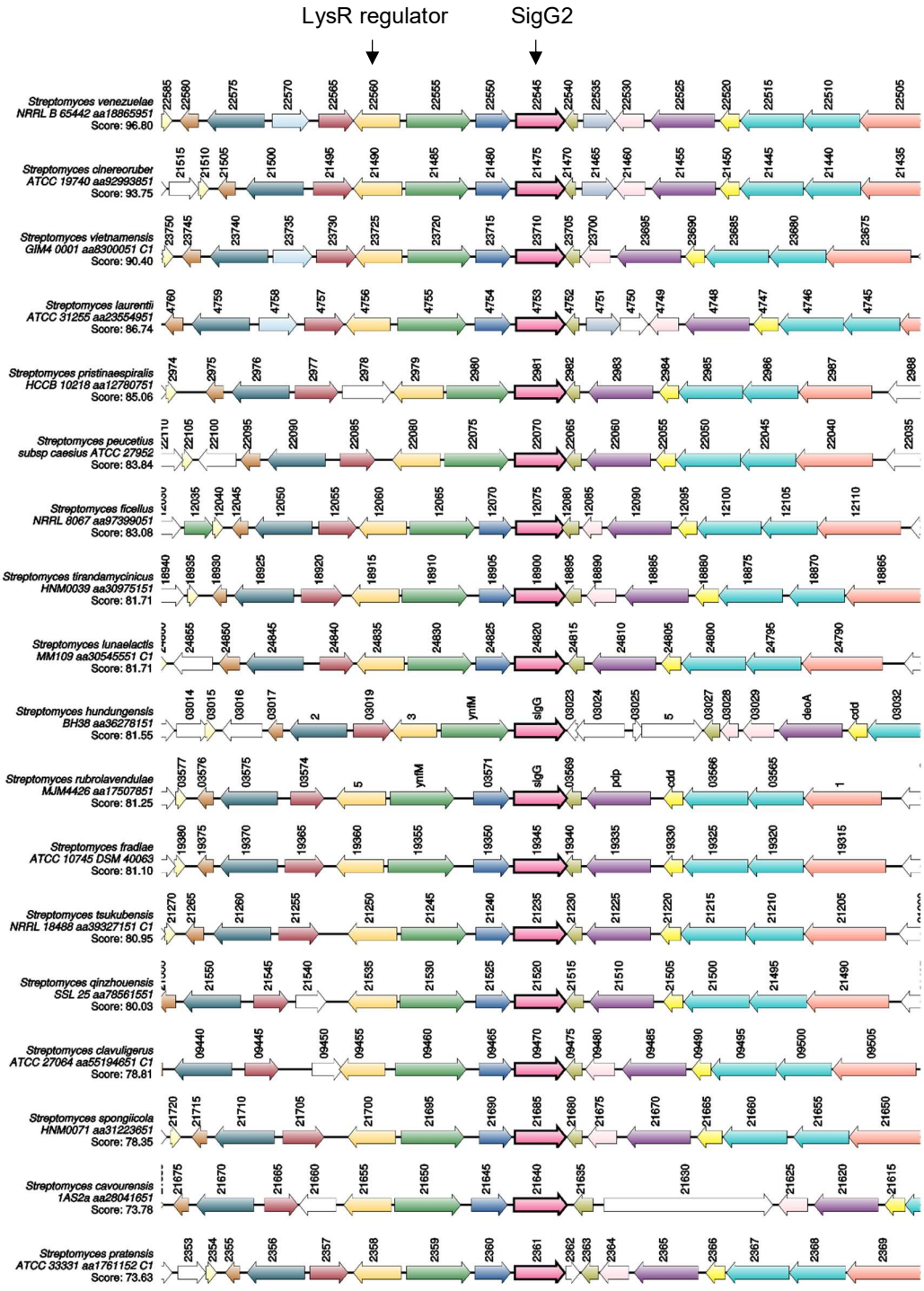
Oliveira, R., Bush, M.J., Pires, S., Chandra, G., Casas-Pastor, D., Fritz, G. and Mendes, M.V. (2020). *Streptomyces tsukubaensis* differentiation is aided by the redox responsive σ/anti-σ pair, SigG1/RsfG. Applied and Environmental Microbiology.

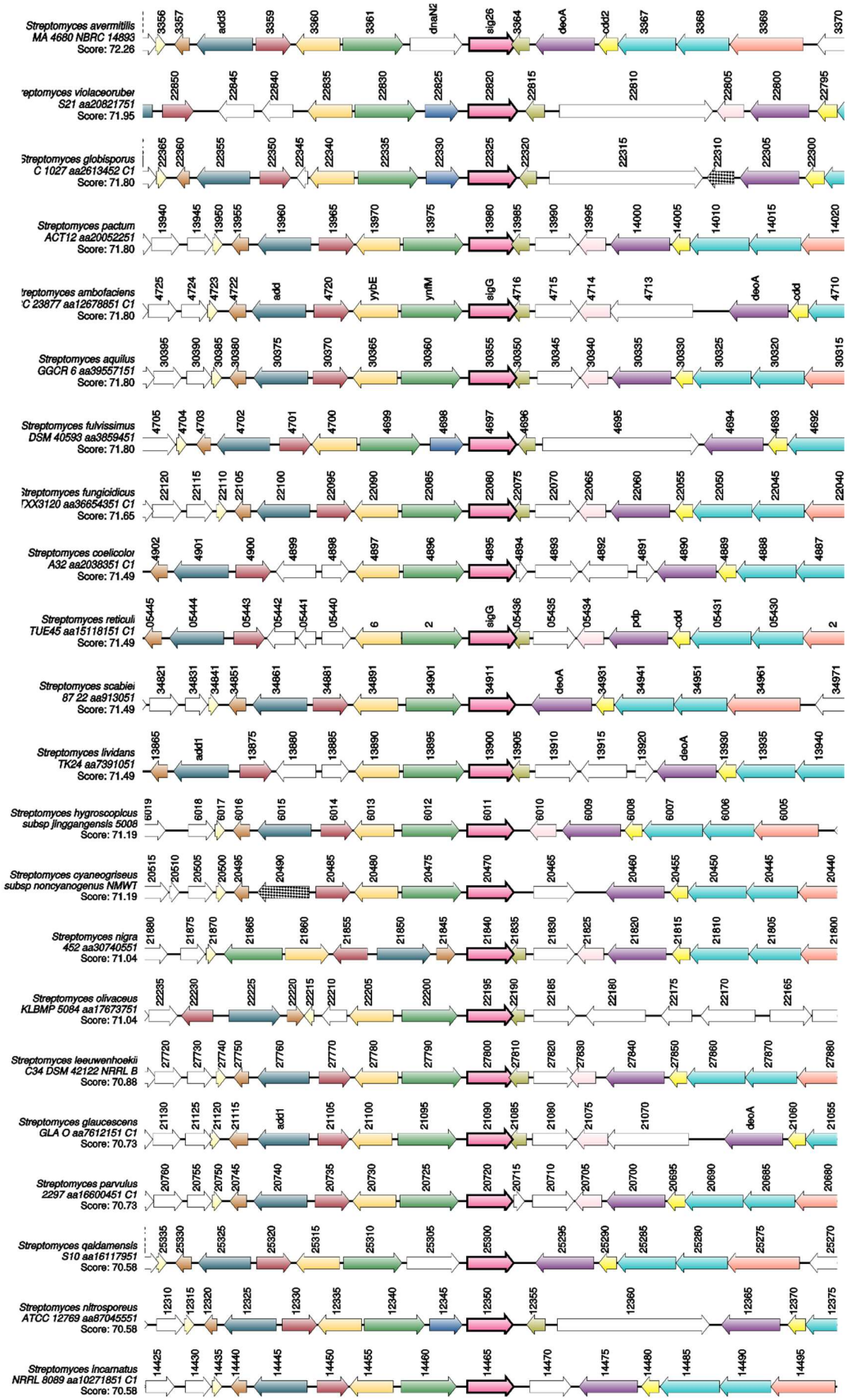
Smollett, K.L., Dawson, L.F., and Davis, E.O. (2011). SigG does not control gene expression in response to DNA damage in *Mycobacterium tuberculosis* H37Rv. J Bacteriol *193*, 1007-1011.

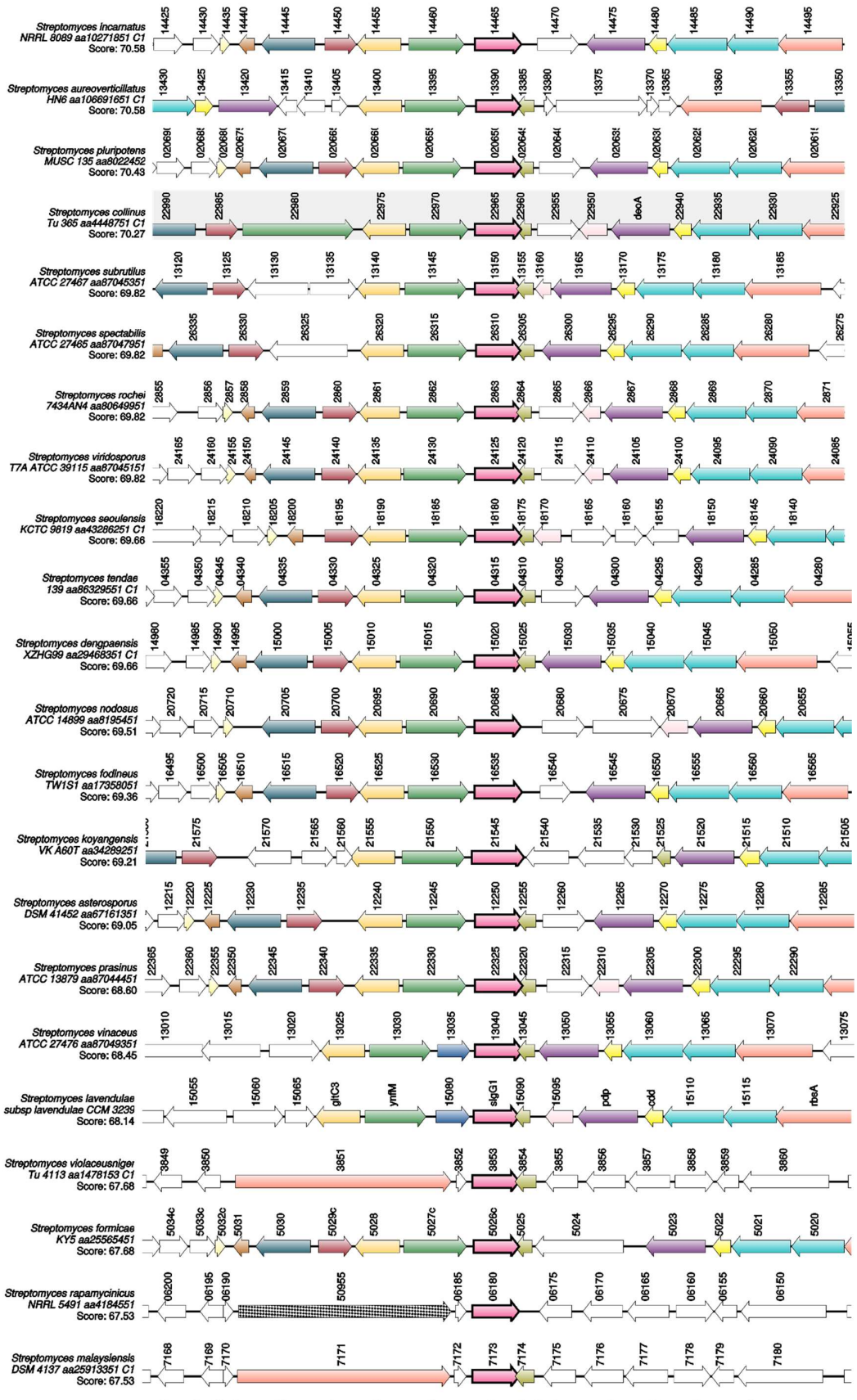
Staroń, A., Sofia, H.J., Dietrich, S., Ulrich, L.E., Liesegang, H., and Mascher, T. (2009). The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. Molecular Microbiology *74*, 557-581.

Thompson, M.G., Zargar, A., Cruz-Morales, P., De Rond, T., Chang, S., Pearson, A. N., Goyal, G., Barajas, J.F., Blake-Hedges, J.M., Phelan, R. M., Umana, V.R., Hernández, A. C., Hillson, N. J., Shih, P. M. and Keasling, J.D. (2020). Characterization of an ECF56-family sigma factor from *Streptomyces venezuelae* reveals a highly conserved regulome. bioRxiv.

4.8 Supplementary material




Streptomyces avermitilis MA 4680 NBRC 14893
Score: 72.26
3356 3357 add3 3359 3360 3361 dnaN2 sigZ6 3364 deoA cdd2 3367 3368 3369 3370
eptomyces violaceoruber S21 aa20821751
Score: 71.95
22850 22845 22840 22835 22830 22825 22820 22815 22810 22805 22800 22795
Streptomyces globisporus C 1027 aa2613452 C1
Score: 71.80
22365 22360 22355 22350 22345 22340 22335 22330 22325 22320 22315 22310 22305 22300
Streptomyces pactum ACT12 aa20052251
Score: 71.80
13940 13945 13950 13955 13960 13965 13970 13975 13980 13985 13990 13995 14000 14005 14010 14015 14020
reptomyces ambofaciens C 23877 aa12678851 C1
Score: 71.80
4725 4724 4723 4722 add 4720 yybE ymfM sigG 4716 4715 4714 4713 deoA cdd 4710
Streptomyces aquilus GGCR 6 aa39557151
Score: 71.80
30395 30390 30385 30380 30375 30370 30365 30360 30355 30350 30345 30340 30335 30330 30325 30320 30315
Streptomyces fulvissimus DSM 40593 aa3859451
Score: 71.80
4705 4704 4703 4702 4701 4700 4699 4698 4697 4696 4695 4694 4693 4692
Streptomyces fungicidicus TXX3120 aa36654351 C1
Score: 71.65
22120 22115 22110 22105 22100 22095 22090 22085 22080 22075 22070 22065 22060 22055 22050 22045 22040
Streptomyces coelicolor A32 aa2038351 C1
Score: 71.49
4902 4901 4900 4899 4898 4897 4896 4895 4894 4893 4892 4891 4890 4889 4888 4887
Streptomyces reticuli TUE45 aa15118151 C1
Score: 71.49
05445 05444 05443 05442 05441 05440 6 2 sigG 05436 05435 05434 pdp cdd 05431 05430 2
Streptomyces scabiei 87 22 aa913051
Score: 71.49
34821 34831 34841 34851 34861 34881 34891 34901 34911 deoA 34931 34941 34951 34961 34971
Streptomyces lividans TK24 aa7391051
Score: 71.49
13865 add1 13875 13880 13885 13890 13895 13900 13905 13910 13915 13920 deoA 13930 13935 13940
Streptomyces hygroscopicus subsp jinggangensis 5008
Score: 71.19
6019 6018 6017 6016 6015 6014 6013 6012 6011 6010 6009 6008 6007 6006 6005
Streptomyces cyaneogriseus subsp noncyanogenus NMWT
Score: 71.19
20515 20510 20505 20500 20495 20490 20485 20480 20475 20470 20465 20460 20455 20450 20445 20440
Streptomyces nigra 452 aa30740551
Score: 71.04
21880 21875 21870 21865 21860 21855 21850 21845 21840 21835 21830 21825 21820 21815 21810 21805 21800
Streptomyces olivaceus KLBMP 5084 aa17673751
Score: 71.04
22235 22230 22225 22220 22215 22210 22205 22200 22195 22190 22185 22180 22175 22170 22165
Streptomyces leeuwenhoekii C34 DSM 42122 NRRL B
Score: 70.88
27720 27730 27740 27750 27760 27770 27780 27790 27800 27810 27820 27830 27840 27850 27860 27870 27880
Streptomyces glaucescens GLA O aa7612151 C1
Score: 70.73
21130 21125 21120 21115 add1 21105 21100 21095 21090 21085 21080 21075 21070 deoA 21060 21055
Streptomyces parvulus 2297 aa16600451 C1
Score: 70.73
20760 20755 20750 20745 20740 20735 20730 20725 20720 20715 20710 20705 20700 20695 20690 20685 20680
Streptomyces qaidamensis S10 aa16117951
Score: 70.58
25335 25330 25325 25320 25315 25310 25305 25300 25295 25290 25285 25280 25275 25270
Streptomyces nitrosporeus ATCC 12769 aa87045551
Score: 70.58
12310 12315 12320 12325 12330 12335 12340 12345 12350 12355 12360 12365 12370 12375
Streptomyces incarnatus NRRL 8089 aa10271851 C1
Score: 70.58
14425 14430 14435 14440 14445 14450 14455 14460 14465 14470 14475 14480 14485 14490 14495


Streptomyces incarnatus NRRL 8089 aa10271851 C1 Score: 70.58
Streptomyces aureoverticillatus HN6 aa106691651 C1 Score: 70.58
Streptomyces pluripotens MUSC 135 aa80224521 Score: 70.43
Streptomyces collinus Tu 365 aa4448751 C1 Score: 70.27
deoA
Streptomyces subrutilus ATCC 27467 aa87045351 Score: 69.82
Streptomyces spectabilis ATCC 27465 aa87047951 Score: 69.82
Streptomyces rochei 7434AN4 aa80649951 Score: 69.82
Streptomyces viridosporus T7A ATCC 39115 aa87045151 Score: 69.82
Streptomyces seoulensis KCTC 9819 aa43286251 C1 Score: 69.66
Streptomyces tendae 139 aa86329551 C1 Score: 69.66
Streptomyces dengpaensis XZHG99 aa29468351 C1 Score: 69.66
Streptomyces nodosus ATCC 14899 aa8195451 Score: 69.51
Streptomyces fodineus TW1S1 aa17358051 Score: 69.36
Streptomyces koyangensis VK A60T aa34269251 Score: 69.21
Streptomyces asterosporus DSM 41452 aa67161351 Score: 69.05
Streptomyces prasinus ATCC 13879 aa87044451 Score: 68.60
Streptomyces vinaceus ATCC 27476 aa87049351 Score: 68.45
gltC3
ynfM
sigG1
pdp
cdd
rbsA
Streptomyces lavendulae subsp lavendulae CCM 3239 Score: 68.14
Streptomyces violaceusniger Tu 4113 aa1478153 C1 Score: 67.68
Streptomyces formicae KY5 aa25565451 Score: 67.68
Streptomyces rapamycinicus NRRL 5491 aa4184551 Score: 67.53
Streptomyces malaysiensis DSM 41137 aa25913351 C1 Score: 67.53

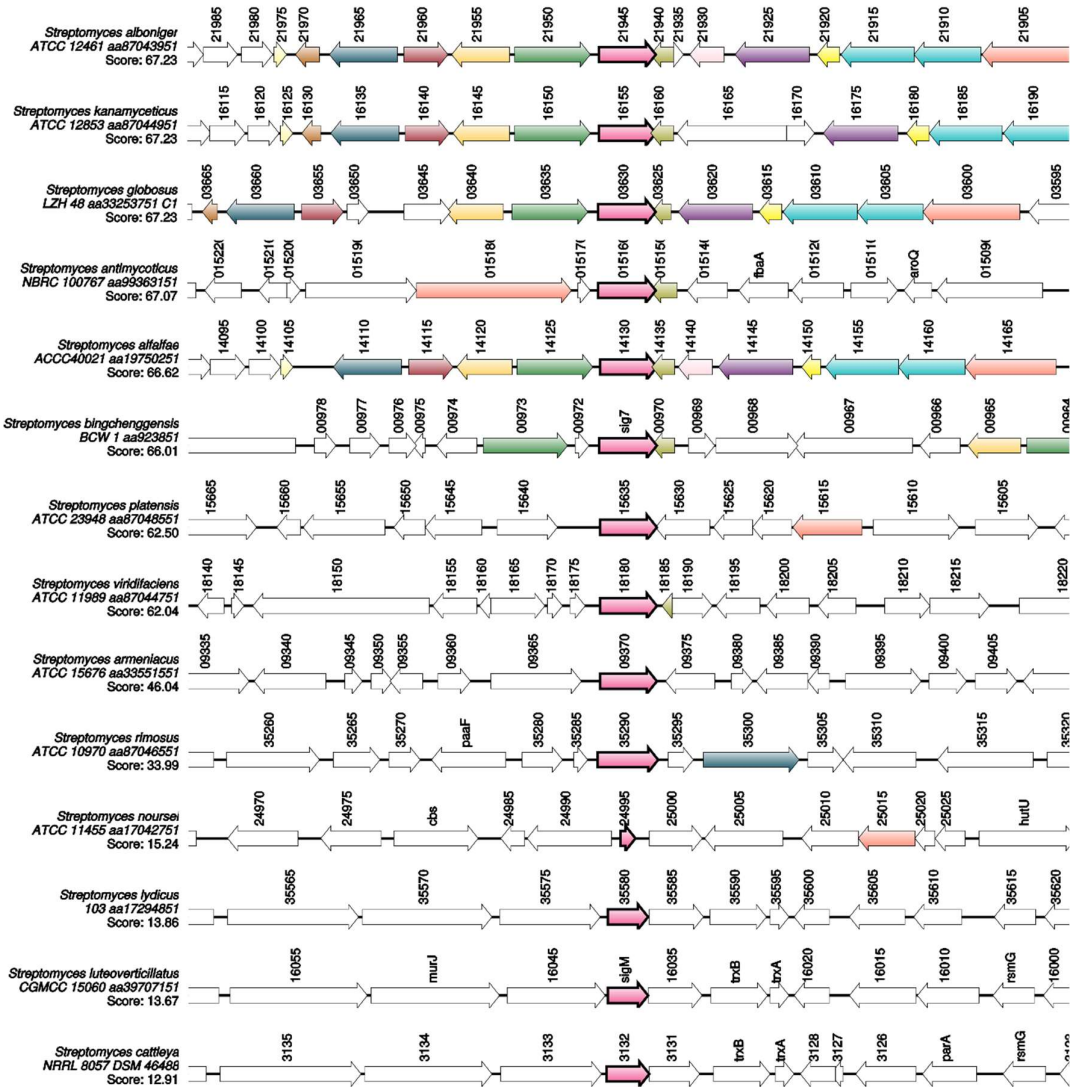


Figure S4.1 Selected *Streptomyces* genomic regions showing the genomic neighbourhood for orthologues of *sigG2*. The *sigG2* gene (22545 in *S. venezuelae* genomic region) and its orthologues are depicted at the centre in pink. Genes encoding LysR-type regulators are located upstream of the *sigG2* orthologues and are coloured dark yellow (22560 in *S. venezuelae* genomic region).

Table S4.1. Selection of antibiotic resistance cassette for plasmid integration in *Streptomyces venezuelae*.

Strain	**Antibiotics commonly used to select for *Streptomyces***				
S. *venezuelae*	Viomycin^{S}	Hygromycin^{S}	Kananamycin^{S}	Thiostrepton^{S}	Apramycin^{S}

Table S4.2. Integration sites for plasmid integration in *Streptomyces venezuelae.*

attB site / Strain	ϕC31	ϕBT1	TG1	SV1	pSam2	VWB	ϕJoe
S. *venezuelae*							
No integration (plasmid)				[2]	[1]	[1]	[1]
Integration (plasmid)	pIJ12551	pMS82	pRF11	[2]	[1]	[1]	[1]

[1]: Not tested experimentally in this work

[2]: No attB site found in genome.

Table S4.3 Oligonucleotides used in this study

Primers	Sequence 5' - 3'	Use
Red SigG2 F	CGGGTGCCGGTGAGGGGTGAGTGGGGTGTGGACCCGATGATT CCGGGGATCCGTCGACC	Redirect™ gene KO
Red SigG2 R	CCTGCTCACCGGCCTCGACGGACCGCTGGGAAACCCCTATGTA GGCTGG	Redirect™ gene KO
KpnI SigG2	GTGGTACCGATGACCGACGCCGCGACC	Bacterial two-hybrid
EcoRI SigG2	ACGGCGAATTCACCCCTAGCCCTCCAGCC	Bacterial two-hybrid
T18seq_F	GTGTGGAATTGTGAGCGGAT	Bacterial two-hybrid
T18seq_R	TTCCACAACAAGTCGATGCG	Bacterial two-hybrid
T25seq_F	CGGTGACCAGCGGCGATT	Bacterial two-hybrid
T25seq_R	GGCGATTAAGTTGGGTAACGCC	Bacterial two-hybrid
NT25seq_F	CCCCAGGCTTTACACTTTATGC	Bacterial two-hybrid
NT25seq_R	TTGATGCCATCGAGTACGGCT	Bacterial two-hybrid
T18Cseq_F	GTGCCGAGCGGACGTTCGA	Bacterial two-hybrid
T18Cseq_R	CTTAACTATGCGGCATCAGAGC	Bacterial two-hybrid

Chapter 5

General discussion and future directions

The ubiquitous distribution throughout diverse ecosystems exposes actinobacterial species to highly distinct environmental conditions. While some have evolutionary advantage as human intracellular prokaryotes, others can be obligate aerobes and thrive in highly oxidizing environments. Hence, bacteria within this phylum harbour a degree of high transcriptional plasticity acquired through the evolution of a plethora of RNAP sigma subunit, the sigma factors. In contrast, other prokaryotes that colonize only one type of environment seem to be simpler and have only a scarce number of these regulators. Among the sigma factors, the members of the ECF family confer high plasticity to the cells, since they are responsible for programming the transcriptional response to the different signals encountered.

Significant progress has been made in understanding the mechanisms behind ECF-dependent transcription initiation. However, we keep finding novel ECFs, with distinct domain architectures, although their functional roles remain elusive. Here we described a new sigma/anti-sigma pair that comprises a SnoaL_2 extension at the C-terminus of the ECF SigG1, and its antagonist RsfG.

5.1. Numerous layers of ECF regulation

Transcription complexes in Actinobacteria are particularly less stable than in other bacteria, like *E. coli*, and thus harbour additional players that ensure an optimal transcription (Davis et al., 2014; Hu et al., 2012). In fact, a substantial part of the transcription complexity of stress responses is achieved with the participation of auxiliary transcriptional regulators, or other layers of posttranslational regulation of the ECF. The latest expansion on the ECF classification unveils that ECFs that are controlled by several regulatory layers are more common than originally thought (Casas-Pastor et al., 2019). Accordingly, we proposed that both SigG1 and its *S. venezuelae* orthologue, SigG2, might be subjected to several layers of regulation that generate stringent mechanisms of ECF56-dependent transcription inhibition.

We found that *sigG1* was transcribed independently or in co-transcription with its cognate anti-sigma factor gene *rsfG.* We also found that *sigG1* expression was activated when *rsfG* is not present. However, protein levels did not correlate with these transcription rates, since the SigG1 protein was produced in lower amounts in Δ*rsfG,* as compared to the wild type. One possibility is that a shorter transcript comprising solely *sigG1* might be targeted for degradation. This would prevent the polar effects that might derive from the unrestrained overexpression of a sigma factor that regulates restricted targets. It is therefore conceivable that, although a fast response is required when cells are challenged by H_2O_2-stress, the abundant transcripts produced in the presence of this agent might have been achieved from the longer, more stringently controlled species.

Although reprogramming transcription is important, posttranslational regulation seems to be the predominant source of stringency in ECF systems. In this work, we report that the ECF alone might harbour posttranslational modifications with a potential regulatory role. However, an important finding of this study is the existence of a SigG1 antagonist, RsfG, which binds to SigG1 to prevent its runaway activity during the differentiation process. What mediates the binding or the dissociation of SigG1 from RsfG is yet unknown, but we show that the stability of the complex might depend on specific conditions of the environment. For instance, when we replaced the methionine residues in these proteins by selenomethiones, we observed the release of SigG1 from RsfG. One explanation is that this was due to a change in the redox state of the culture in the presence of DTT − which could result in an alteration of the optimal conformation to form the complex − or the absence of specific ligands to mediate interaction.

The crystallographic studies on SigG1 and RsfG ternary structures could give important insights on the mode of action of sigma factors of the ECF56 family. In addition, knowledge gathered with such studies could help answering the long lasting question of whether or not CT extensions play an inhibitory role through intramolecular regulation, or if ECF56 sigma factors unequivocally need an anti-sigma factor to control their activity in any physiological response. Another major goal was to define what distinguishes ECF56 from ECF41 sigma factors. One strong clue that remains from our studies is that the σ2-σ4 core is more likely to determine the ECF family and each family would be assigned its own type of regulon. If this was the case, the SnoaL_2 domain, which is a less conserved region, would participate in the control the ECF release. Hence, we propose that the ECF56 architecture, which is conserved in other phylogenetically distant ECF families, might play a more significant role in its own regulation, rather than participating in selecting the target regulons. If the ECF56 autoregulatory capacity derives exclusively from its unique architecture, or if is it associated to a specific molecule that mediates the intramolecular contacts, is still to be determined.

These studies on the three-dimensional structure of SigG1 and RsfG could also shed light on what dissociates the sigma factor from the cognate anti-sigma factor. Interestingly, *sigG1* neighbouring genes might participate in this process and conceal additional layers of regulation, as has been commonly observed for other ECFs. For instance, genes adjacent to *rsfG* predictably encode a DNA-binding protein, a membrane protease and a peptidoglycan-binding protein (PBP). Additional experiments to evaluate the timing of expression of these genes and protein-protein interactions of their products with SigG1 and RsfG would indicate if these could act as auxiliary proteins to this sigma/anti-sigma system, in the stabilization of SigG1 binding to the promoters or in coordinating regulated proteolysis of RsfG, for example.

5.2. Transduction of the trigger signal

Some known sigma antagonists display several transmembrane domains that place it at the membrane, with an exposed sensor domain oriented towards the extracellular space to perceive the stress signal. In turn, RsfG contains a single predicted short transmembrane domain and therefore it is unlikely that it integrates in the cell membrane. In addition, neither SigG1 nor RsfG contain evident redox sensors like the Rex regulator conserved motif, or cysteine residues that could coordinate metal ions or catalyse the formation of disulfide bonds. Hence, it is conceivable that the signal that triggers SigG1 activity is perceived elsewhere by a novel sensory mechanism, or an auxiliary protein of the SigG1-RsfG system. Given the PTMs identified in the recombinant SigG1, one possibility is that its activation could be dependent on a phosphorylation signal by a sensor kinase.

Another hypothesis is that SnoaL_2 could be the domain responsible for transducing the stress signal. Comparative genome analysis showed that SigG1 homologues are represented in several different bacteria, some species harbouring more than one copy of these proteins. This suggest that the presence of a SnoaL_2 domain fused to an ECF core reflects a widespread conserved mechanism of ECF regulation. Previously described SnoaL_2 domains are important players in the biosynthesis of specialised compounds, like the antibiotic nogalamycin in *Streptomyces nogalacter,* because they catalyse an intermediate hydroxylation step of the biosynthesis of this compound as polyketide cyclases. The active residues in these enzymes are absent in SigG1 or SigG2 SnoaL_2 domains, indicating that this could be a non-catalytically active domain when fused to an ECF core region.

The S. tsukubaensis strain harbours other SnoaL_2 domains with the active site residues, which stand alone in protein sequences, and therefore it seems to have the machinery required for the hydroxylation of secondary metabolites. Hence, we propose the SnoaL2 domain in SigG1 plays a role that is not related to the biosynthesis of bioactive compounds. Additional site-directed mutagenesis and *in vitro* assays are essential to determine if SigG1 SnoaL_2 can compensate for the hydrolase function, in the absence of the SnoaL-like polyketide cyclases. While we did not address the biochemical properties of the SnoaL_2 domain, by aligning the sequences of SnoaL-like enzymes and SnoaL_2-containing ECFs, it seems like at least two residues modifications - that favour alterations to hydrophobic aminoacids - are conserved in ECF sequences (Figure 5.1).

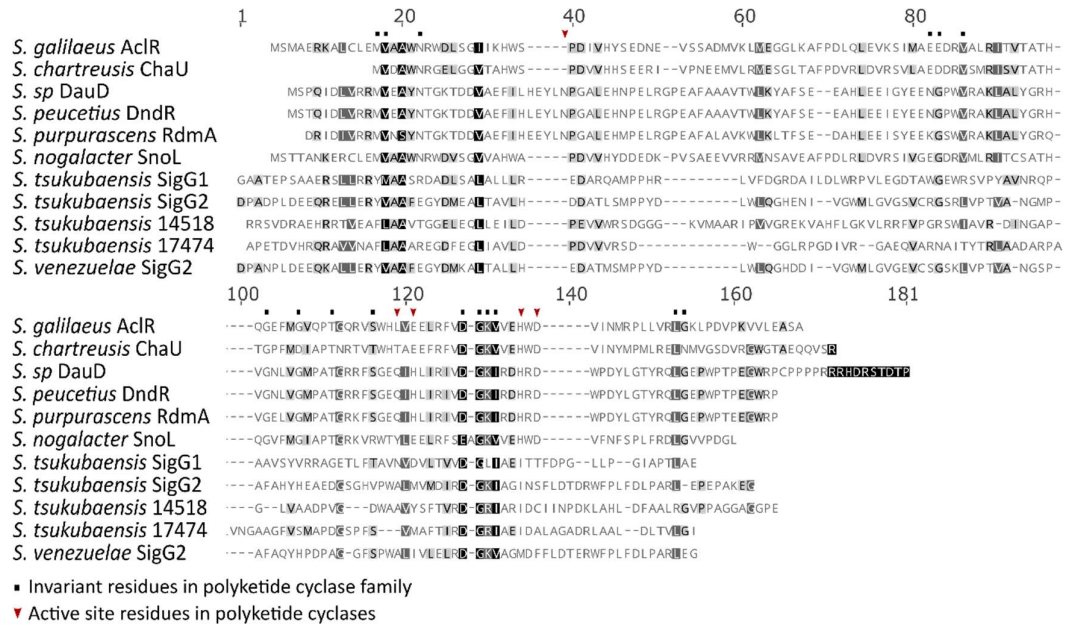


Figure 5.1. Sequence alignment of SnoaL-like cyclases and SnoaL2-containing ECFs from *Streptomyces.* AclR: Q1XDX7; ChaU: Q4ROL2; DauD: Q55215; DnrD: Q54808; RdmA: Q54526; SnoL: Q9RN64; SigG1: I2N5H1; 12530: I2N4W0; 14518: I2N3T7; 17474: I2N230; SigG2: F2RL89. Catalytic residues in the hydroxylase sequences are marked with arrows. Invariant residues in polyketide cyclases are marked by dots.

The SnoaL_2 domain in SigG1 exhibits a modification of a cyclase histidine that became an isoleucine, SigG1 I352, and a cyclase glutamine that was substituted by an alanine, A288 in SigG2 (not conserved in SigG1, which keeps a polar uncharged residue instead, N338). However, experimental studies are crucial to better define residues that distinguish catalytic from non-catalytic SnoaL_2 domains. To support this idea, a previous work on *M. tuberculosis* SigJ crystal structure has proposed that SigJ was bound to a small ligand, still to be identified (Goutam et al., 2017). Overall, one hypothetical model is that the "inactive" SnoaL_2 domain of SigG1 recognises an effector molecule and transduces this signal to promote SigG1 binding to DNA. In such case, SigG1 does not bind to DNA in the absence of an effector molecule.

5.3. The SigG1 regulon

While a filamentous lifecycle confers great advantage in accessing nutrients and dispersing its reproductive structures, building such a complex structure demands the contribution of an arsenal of regulators. Several reports have demonstrated that *Streptomyces* differentiation benefits from sacrificing part of the colony by inducing controlled cell death (Beites et al., 2015; Miguélez et al., 1999; Oliveira et al., 2020; Tenconi et al., 2018). The RsfG-SigG1 system is active upon H_2O_2-induced stress

generated by the fading vegetative mycelia to modulate the stress and help treading through the various differentiation stages that will lead to the erection of aerial hyphae. What triggers the release of SigG1 from RsfG at this stage is yet unknown. Nonetheless, the activation of the sigma/anti-sigma system seems to follow *sigG1* expression during these developmental transitions. In line with this, the free SigG1 initiates the expression of both developmental regulators and the energetic powerhouses to fuel the metabolic shift from vegetative to aerial differentiation (from primary to secondary metabolism).

Genes within the SigG1 regulon are intimately related to the control of metal imbalance, which suggests that SigG1 is responsible for alleviating the cell from the damage caused by oxidative stress and reflects the complexity of ECF-mediated stress responses in these bacteria. In fact, we not only find cases where the same ECF responds to different environmental factors, but also, a series of ECFs are responsible for initiating overlapping regulons that will mediate the response to one type of stress. There is good example of this behaviour in *B. subtilis* where different ECF sigma factors respond and regulate cell envelope stresses (Luo et al., 2010; Mascher et al., 2007). In order to avoid polar effects of unintended crosstalk from the simultaneous activation of stress-responsive machineries, bacteria appears to insulate individual ECF-dependent signalling systems. This could explain the mild recovery of phenotypes of the Δ*rsfG-sigG1* double mutant in a rescue setting, possibly due to the activation of alternative compensation mechanisms.

Although we found a high degree of synteny of the presence of LysR-type regulators near the genomic regions where *S. tsukubaensis sigG1* and *sigG2* are located, SigG2 does not follow SigG1 transcriptional landscape. Moreover, we did not find ECF56 or BldN promoter boxes in *sigG2* promoter, which suggests its activation might not be dependent on the same sigma factors that activate *sigG1*. By addressing the function of *S. venezuelae* SigG2 we sought to give insights on how SnoaL_2-dependent signal transduction works in *Streptomyces*. Results obtained in this model fall short in elucidating the regulation of SigG2 activity, and future work is still needed to properly characterise SigG2 in particular, and in a broader context to better define the role of the SnoaL_2 extension in Actinobacterial ECFs.

When studying the initiation of transcription by an ECF a frequent question arises. Does the ECF56 bind first to RNAP or the target promoter? As a subunit of the RNAP it is likely that the ECF binds to the RNAP once it becomes available, and only then recruits the promoters of the genes that will produce a response. Traditionally in transcription regulation, many of the promoter regulatory elements overlap the sequence recognised by other transcriptional regulators. Therefore, a scenario of a priority binding of the ECF to the promoter, without having the other subunits of the RNAP ready to transcribe would

probably cost more time and energy to the cell. Another question would be: Does SigG1 bind to target promoters hierarchically, depending on the current developmental phase? In the bacterial strains and growth conditions tested, no correlation between the transcriptional activation observed in ChIP-seq experiments and RNA-seq transcript levels was found for some of the targets. Hence, we speculate that SigG1 indeed binds hierarchically to recruit promoter activity as specific needs emerge. To prove this hypothesis we would need additional studies on the individual targets in SigG1 regulon, including a thorough screening for the signals that activate each target, in order to evaluate its influence in the physiological conditions where they are highly expressed.

5.4. SigG1 as a synthetic biology building block

Our understanding on how to control bacterial transcription has largely improved throughout time. Nowadays, the modulation of mRNA levels has become a powerful tool in synthetic biology, for instance, to fine-tune the synthesis of clinically important specialized metabolites like antimicrobial or anticancer agents. In this context, ECF sigmas have been successfully implemented *in vivo* as building blocks for synthetic biology applications, as they provide responsive-switches ready to activate more effective signalling circuits (Pinto et al., 2018; Rhodius et al., 2013). Through the recruitment of RNAP to specific −10 and −35 promoter regions, ECFs enable the activation of targeted sequential regulatory cascades.

Comprehensive studies on the mechanisms of activation of SigG1 can make it suitable for engineering pathways of interest in bacteria that do not express ECF56 sigma factors, such as *E. coli*, *Bacillus*, *Shewanella*, *Listeria* or *Syneccocystis*. Responsive gene induction through SigG1, in particular, can be useful in providing energetic support to fuel metabolism, or to redirect metabolism to increase other resources like metals to improve growth rate. Moreover, by modulating the transition between developmental stages, SigG1 might be useful in controlling the time of activation of secondary metabolism pathways, in order to achieve a more robust production of selected bioactive compounds.

Computational statistical methods help to predict the activation and the stable behaviour of sigma cascades, however, several physiological issues might arise from expressing a sigma factor ectopically. In many cases, overexpressing a sigma factor translates into a fitness disadvantage, probably due to the triggering of high rates of runaway promoter activities. In our work, we show that it is possible to minimize this detrimental effect and achieve greater stability of SigG1 by limiting the number of plasmid copies in the cell − and therefore the metabolic burden induced by high replication rates of the vector − using integrative systems instead of replicative vectors. In addition, we

sought to ensure the controlled fold-induction of SigG1 expression by using the *sigG1* native promoter, as opposed to a constitutive strong promoter. An additional aspect that makes the biosynthetic application of the expression of a responsive sigma factor more difficult is the unpredictability of the cognate antagonist behaviour. In previous work, signal amplification through heterologous expression of anti-sigma factors resulted in a toxicity burden to *E. coli* cells (Rhodius et al., 2013), probably due to promiscuous binding of the anti-sigma to other sigma factors. Another study reported that even after ECF release, the intended switch might face additional time constrains that result in *"heterogeneous timing"* of transcription. Fritz et al. (2019) described that genetically identical cells in the same culture conditions might trigger the synthetic switches at different times, with subsequent unintended phenotypic heterogeneity. Another problem that might arise is the crosstalk among signalling cascades that might be activated upon expression of a regulator from a heterologous system. Despite the architectural resemblance, dissimilarities in SigG1 and SigG2 regulon showed that SnoaL_2-ECFs might target different sets of genes. This will be important an important consideration when thinking about a synthetic use of this type of ECFs, as it might avoid potential crosstalk with co-expressed endogenous ECFs, and insulates the stimulation of pathways of interest.

SnoaL_2 ECFs constitute major advantageous responsive switches to overcome these constrains, as they rely on several layers of regulation that provide high stringency to the ECF activation process. We and others proposed that the timely activation of a SnoaL_2 ECF might require 1) a redox sensing mechanism, 2) a small molecule binding, 3) intramolecular interactions and 4) the sequestration by an anti-sigma factor. Therefore, refining our understanding of the numerous triggers underpinning the activation of SigG1 is crucial to implement a high valuable tool that has the potential to be broadly applicable and confer plasticity to strains that lack these ECFs. Here we show that the activation of a target promoter by SigG1 can be artificially stimulated using a genetically introduced promoter sequence, and that it can be implemented in a heterologous host like *E. coli* (Figure 5.2). We generated a reporter-probe system with the *gusA* gene − which codes for the β-glucuronidase (GUS) enzyme − under the control of candidate SigG1 target promoters; and together with an IPTG-inducible system expressing the 6H-SigG1 (see pET15b-6H-SigG1 in Chapter 3 for details) we were able to detect activity of the alanine dehydrogenase promoter (*aldp*), upon expression of 6H-SigG1 (Figure 5.2A). We generated a reporter-probe system with the *gusA* gene − which codes for the β-glucuronidase (GUS) enzyme − under the control of candidate SigG1 target promoters; and together with an IPTG-inducible system expressing the 6H-SigG1 (see pET15b-6H-SigG1 in Chapter 3 for details) we were able to detect activity of the

alanine dehydrogenase promoter (*aldp*), upon expression of 6H-SigG1 (Figure 5.2A). An experiment with cells expressing only the p*ald-gusA* system induced with IPTG, without adding SigG1, could be included in these studies as a negative control. We then introduced the reporter-probe system in *S. tsukubaensis* and observed an activation of the alanine dehydrogenase promoter that could be dependent on the overexpression of the endogenous SigG1 (*aldp* activity is upregulated in the Δ*rsfG* mutant where SigG1 is overexpressed, when compared to the wild type) (Figure 5.2B). However, in other bacteria it was shown that a different transcriptional regulator, AldR, activates *aldp* (Lodwig et al., 2004) and therefore we cannot discard that this activation might not be fully dependent on SigG1 in *S. tsukubaensis*. The same happens with the *23024p* target

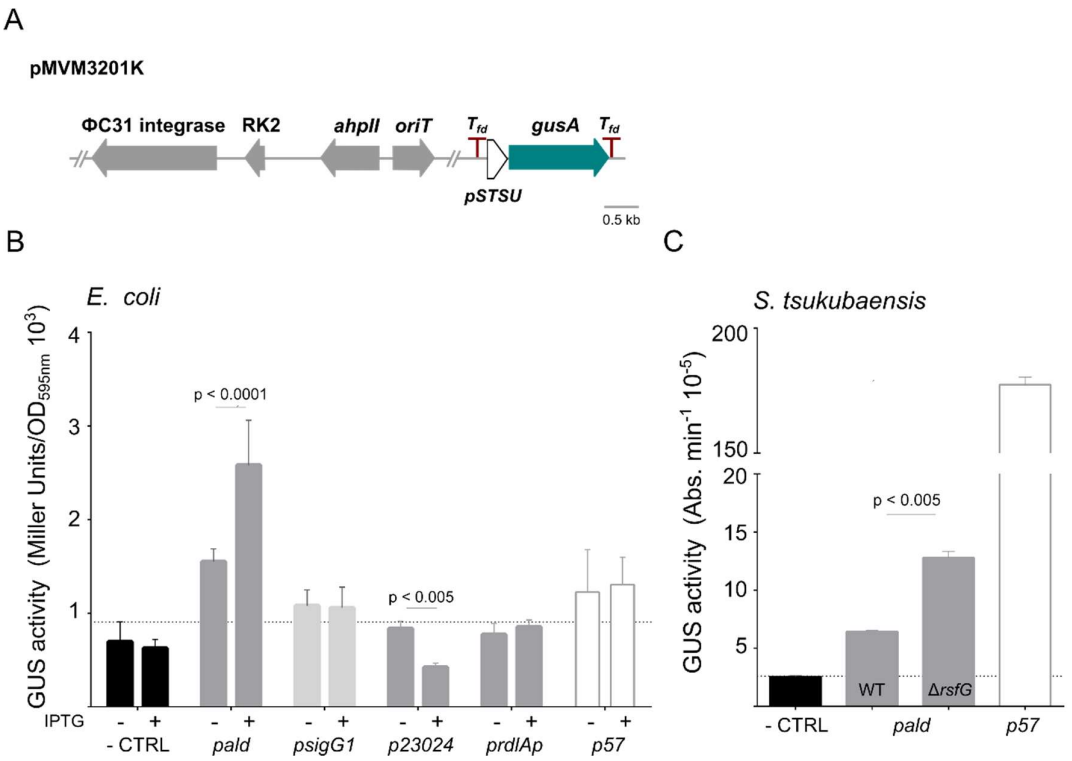


Figure 5.2. Promoter-probe expression directed by SigG1. Putative target promoters were fused to a β-glucuronidase (Gus, *gusA*) reporter system and their activity was evaluated by measuring the activity of the Gus enzyme in the presence of its chromogenic substrate X-Gluc. Schematic representation of the promoter-*gusA* constructs generated for promoter-probe assays (**A**). Gus activity driven by each promoter in *E.coli.* Activation of the target promoter *pald* was achieved upon expression of 6H-SigG1 – under the control of T7 inducible promoter (pET15b-SigG1) – induced by IPTG (**B**). *In vivo* Gus activity driven by the *pald* target promoter in *S. tsukubaensis* cells. Activation of *pald* was achieved upon deletion of the SigG1 antagonist (Δ*rsfG*), which released SigG1 to bind to its target. A construct without the *gusA* gene was used as a control for putative basal levels of endogenous Gus. Expression of *gusA* driven by the p57 promoter was used as a positive control. Results are representative of three independent experiments.

promoter, which can be activated by metal-dependent proteins in other bacteria (Ahmed et al., 1994; Ahmed et al., 1995). More experiments to assess the downregulation of this activity in the Δ*sigG1* mutant are required to confirm these observations. Regarding this application there is more work ahead. In particular, addressing how to improve the efficacy of ECF56 in providing energetic support to fine-tune metabolic shifts, would be crucial in order to provide a better knowledge-based tool to be used in synthetic biology.

5.5. Final remarks

Despite the knowledge gathered with this work, unanswered questions remain with respect to the mode of action of SigG1 and more studies are required in order to be able to securely modulate this system.

With the genome-wide screenings in this work, we were able to retrieve large collections of information. However, they shed limited light on the evolution of the biological processes and often leave a number of open questions. What are the evolutionary advantages to *S. tsukubaensis* of having multiple ECF56 sigma factors with different regulons? How does this ECF/anti-sigma system relate with the different lifestyles, such as the marine, terrestrial, and pathogenic systems? Do bacteria need more effective ROS-responsive or detoxification machineries to produce complex secondary metabolites?

In fact, persisting in an oxygenic environment conveys several stress challenges that can be circumvented through the reprogramming of transcription by a plethora of sigma factors. Overall, this work demonstrates that the expression of a family of sigma factors with a unique architecture can ameliorate stress and maintain the reductive-oxidative balance. Moreover, such regulators play structural roles on the cell as they activate genes that are essential to build the multicellular mycelia. Mostly generated by electron transfer mechanisms, reactive oxygen species (ROS) were traditionally seen as harmful molecules with severe detrimental effects for the cells like DNA damage, disruption of protein viability and lipid oxidation. Collectively, our work brings an additional support to the idea that this paradigm has changed and that ROS may have a bacteriostatic role, and become bactericidal when the internal defence mechanism is blocked.

In sum, our studies gathers evidence that contributes to a number of paradigm shifts on regulatory processes. Our work demonstrates that functional roles are not exclusively encoded in conserved sequences, by showing that proteins with highly similar architectures can have different modes of action. We contributed to the recent efforts to oppose the idea that ROS are compulsory harmful molecules, and propose that they can be key signalling triggers. We substantiate that ECF sigma factors can respond

to intracellular signals, unlike classical ECFs that are regulators of extracytoplasmic cues. We also show that ECF with C-terminal extensions are not exclusively orphan ECFs as has been proposed, and can be controlled by cognate antagonists. Understanding the biology behind SigG1 release and activity will be key, not only to establish a powerful regulatory tool, but also to open new avenues towards unveiling the function of novel ECFs that evolved with complex domain organisations.

5.6. References

Ahmed, M., Borsch, C.M., Taylor, S.S., Vázquez-Laslop, N., and Neyfakh, A.A. (1994). A protein that activates expression of a multidrug efflux transporter upon binding the transporter substrates. Journal of Biological Chemistry *269*, 28506-28513.

Ahmed, M., Lyass, L., Markham, P.N., Taylor, S.S., Vázquez-Laslop, N., and Neyfakh, A.A. (1995). Two highly similar multidrug transporters of Bacillus subtilis whose expression is differentially regulated. Journal of Bacteriology *177*, 3904-3910.

Beites, T., Oliveira, P., Rioseras, B., Pires, S.D.S., Oliveira, R., Tamagnini, P., Moradas-Ferreira, P., Manteca, Á., and Mendes, M.V. (2015). Streptomyces natalensis programmed cell death and morphological differentiation are dependent on oxidative stress. Scientific Reports *5*, 12887.

Casas-Pastor, D., Müller, R.R., Becker, A., Buttner, M., Gross, C., Mascher, T., Goesmann, A., and Fritz, G. (2019). Expansion and re-classification of the extracytoplasmic function (ECF) σ factor family. bioRxiv, 2019.2012.2011.873521.

Davis, E., Chen, J., Leon, K., Darst, S.A., and Campbell, E.A. (2014). Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. Nucleic Acids Research *43*, 433-445.

Fritz, G., Walker, N., and Gerland, U. (2019). Heterogeneous Timing of Gene Induction as a Regulation Strategy. Journal of Molecular Biology.

Goutam, K., Gupta, A.K., and Gopal, B. (2017). The fused SnoaL-2 domain in the *Mycobacterium tuberculosis* sigma factor σjmodulates promoter recognition. Nucleic Acids Research *45*, 9760-9772.

Hu, Y., Morichaud, Z., Chen, S., Leonetti, J.-P., and Brodolin, K. (2012). Mycobacterium tuberculosis RbpA protein is a new type of transcriptional activator that stabilizes the σ A -containing RNA polymerase holoenzyme. Nucleic Acids Research *40*, 6547-6557.

Lodwig, E., Kumar, S., Allaway, D., Bourdes, A., Prell, J., Priefer, U., and Poole, P. (2004). Regulation of l-Alanine Dehydrogenase in *Rhizobium leguminosarum* bv. *viciae* and Its Role in Pea Nodules. Journal of Bacteriology *186*, 842-849.

Luo, Y., Asai, K., Sadaie, Y., and Helmann, J.D. (2010). Transcriptomic and Phenotypic Characterization of a *Bacillus subtilis* Strain without Extracytoplasmic Function σ Factors. Journal of Bacteriology *192*, 5736-5745.

Mascher, T., Hachmann, A.-B., and Helmann, J.D. (2007). Regulatory overlap and functional redundancy among Bacillus subtilis extracytoplasmic function sigma factors. Journal of bacteriology *189*, 6919-6927.

Miguélez, E.M., Hardisson, C., and Manzanal, M.B. (1999). Hyphal Death during Colony Development in Streptomyces antibioticus: Morphological Evidence for the Existence of a Process of Cell Deletion in a Multicellular Prokaryote. Journal of Cell Biology *145*, 515-525.

Oliveira, R., Bush, M., Pires, S., Chandra, G., Casas-Pastor, D., Fritz, G., and Mendes, M. (2020). *Streptomyces tsukubaensis* differentiation is aided by the redox responsive σ/anti-σ pair, SigG1/RsfG. Applied and Environmental Microbiology.

Pinto, D., Vecchione, S., Wu, H., Mauri, M., Mascher, T., and Fritz, G. (2018). Engineering orthogonal synthetic timer circuits based on extracytoplasmic function σ factors Introduction. 1-15.

Rhodius, V.A., Segall-Shapiro, T.H., Sharon, B.D., Ghodasara, A., Orlova, E., Tabakh, H., Burkhardt, D.H., Clancy, K., Peterson, T.C., Gross, C.A*., et al.* (2013). Design of orthogonal genetic switches based on a crosstalk map of sigmas, anti-sigmas, and promoters. Mol Syst Biol *9*, 702.

Tenconi, E., Traxler, M.F., Hoebreck, C., van Wezel, G.P., and Rigali, S. (2018). Production of Prodiginines Is Part of a Programmed Cell Death Process in Streptomyces coelicolor. Frontiers in Microbiology *9*, 1742-1742.