The novel ECF56 SigG1-RsfG system modulates morphological differentiation and metal-ion homeostasis in *Streptomyces tsukubaensis*

Rute Oliveira\(^1,2,3\), Matthew J. Bush\(^4\), Silvia Pires\(^2,6\), Govind Chandra\(^4\), Delia Casas-Pastor\(^5\), Georg Fritz\(^7\) & Marta V. Mendes\(^1,2\)

Extracytoplasmic function (ECF) sigma factors are key transcriptional regulators that prokaryotes have evolved to respond to environmental challenges. *Streptomyces tsukubaensis* harbours 42 ECFs to reprogram stress-responsive gene expression. Among them, SigG1 features a minimal conserved ECF \(\sigma_2-\sigma_4\) architecture and an additional C-terminal extension that encodes a SnoaL\(_2\) domain, which is characteristic for ECF \(\sigma\) 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 remains unknown. Our results show that in addition to predicted self-regulatory intramolecular amino acid interactions between the SnoaL\(_2\) domain and the ECF core, SigG1 activity is controlled by the cognate anti-sigma protein RsfG, encoded by a co-transcribed \(\text{sigG}1\) neighbour gene.

Characterisation of ∆\(\text{sigG}1\) and ∆\(\text{rsfG}\) strains combined with RNA-seq and ChIP-seq experiments, suggests the involvement of SigG1 in the morphological differentiation programme of *S. tsukubaensis*. SigG1 regulates the expression of alanine dehydrogenase, \(\text{ald}\) and the WhiB-like regulator, \(\text{wblC}\) required for differentiation, in addition to iron and copper trafficking systems. Overall, our work establishes a model in which the activity of a \(\sigma\) factor of group ECF56, regulates morphogenesis and metal-ions homeostasis during development to ensure the timely progression of multicellular differentiation.

Streptomyces are unique soil-dwelling bacteria with a prolific biosynthetic capacity responsible for hundreds of widely used bioactive compounds. The onset of streptomycetes specialized metabolism and the production of secondary metabolites is intimately related with the morphological differentiation process. *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. Previously, we demonstrated that differentiation in *S. natalensis* is tightly controlled by oxidative stress\(^1\). We illustrated that the transition from first mycelium stage (MI) to secondary metabolite producing differentiated mycelium (MII) is preceded by phenomena of cell death in the vegetative mycelia, which reroutes nutrients to differentiate MII hyphae into nascent aerial mycelium. Also, pivotal to this substrate to aerial mycelium transition is the production of a hydrophobic coat.

---

\(^1\)Bioengineering and Synthetic Microbiology Group, i3S- Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal. \(^2\)IBMC, Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal. \(^3\)Programa Doutoral em Biologia Molecular e Celular (MCBiology), ICBAS, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal. \(^4\)Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. \(^5\)Center for Synthetic Microbiology, Philippus-University Marburg, 35032 Marburg, Germany. \(^6\)Present address: Jill Roberts Institute for IBD Research, Weill Cornell Medicine, New York, NY 10021, USA. \(^7\)School for Molecular Sciences, University of Western Australia, Perth 6009, Australia. \(^*\)email: mvm@ibmc.up.pt
that consists of two families of proteins, the rodulins and the chaplins. Subsequently, cell division and chromosome segregation generate chains of uni-genomic spores. Fully mature spores frequently acquire a strain-specific colour derived from a polyketide spore pigment produced during the late stages of development. The process of differentiation in the Streptomyces genus 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. 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).

In order to survive complex niches, streptomycetes 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 the Streptomyces genus, the main mechanisms that orchestrate external signal-sensing and amplification include the autoregulatory DNA-binding transcription factors, two-component systems and the extracytoplasmic function (ECF) sigma factors. 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 respond to environmental fluctuations (reviewed in Lonetto, et al.).

ECFs form the largest family of sigma factors, the group IV, which derive from Escherichia coli. ECFs structure displays the minimal architectural requirement to bind DNA, the sigma σ2 region 2 (σ2) and σ4 region 4 (σ4) conserved domains, which bind to the -10 and -35 promoter regions, respectively. Canonical ECFs recognise conserved motifs in the -10 and the -35 boxes, often separated by an optimal spacer of 16-17bp to accommodate the binding of σ2 and σ4 to the promoters. 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. The residues within the ECF σ2-σ4 linker are important for maintaining the bond between σ2 and σ4 and stabilizing the binding of these regions to the RNAP. However, they do not seem to play a direct role in RNAP binding.

Despite the discovery of ECFs in 1994, detailed information regarding how these proteins evolved is still sparse. ECFs show great diversity and have been classified in 157 well-defined phylogenetic families. 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 respond to environmental fluctuations (reviewed in Lonetto, et al.).

Results

**sigG1 encodes an ECF sigma factor with a SnoaL_2 extension.** A gene encoding an ECF, STSU_11560, was identified in S. tsukubaensis located upstream of the gene that codes for OxyR—the major bacterial regulator of H₂O₂-induced stress response. STSU_11560 harbours the amino terminal σ70_r2 (PF04542,
σ2) and σ70_r4 (PF04545, σ4) conserved domains, and a C-terminal (CT) extension, encoding a SnoaL_2 domain (PF12680), which makes it distinctive from other typical ECF σ factors (Fig. 1a). Domains σ2 and σ4 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 phyla17,18. 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 (Supplementary Fig. S1)17. 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 subgroup 3 of the ECF56 family (Supplementary Fig. S1).

The pool of SnoaL_2-containing ECFs in S. tsukubaensis resembles the organisation of orthologous ECFs experimentally characterised in the Mycobacterium genus. Sequence homology revealed that the full-length proteins encoded by STSU_11560 and STSU_12530 are similar to SigG from Mycobacterium tuberculosis H37Rv (40% and 47% identity, respectively), while STSU_14518 and STSU_17474 showed identity to SigJ from M. tuberculosis H37Rv (43% and 38% identity, respectively). The closest homologous protein to STSU_11560 found in the Streptomyces genus was a sigma-70 family RNA polymerase sigma factor from Streptomyces katsurahmaniae (75% identity, E value: 2e-168). 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) predict intramolecular contacts between SigG1 ECF core domains and SnoaL_2. ECFs with additional domains have been proposed to promote intramolecular contacts that regulate ECF activity24,28,29. Making use of the statistical method called direct-coupling analysis (DCA) we assessed whether the SnoaL_2 domain is folded towards the ECF domains σ2 and σ4. 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 interaction34. 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 activity35. 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 σ2 and σ4 (Fig. 1b, area B and C). Additionally, DCA revealed a high probability of contacts between the N-terminal part of SnoaL_2 and the σ4 region (Fig. 1b, area A). These results suggest a potential self-regulatory function for SnoaL_2 (Fig. 1c).

SigG1 interacts with a putative anti-sigma factor, RsfG. Downstream from sigG1 we identified an additional uncharacterised gene, STSU_11555, that encodes an hypothetical protein with no conserved Pf domains. 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 or the sigI-rsgI pair in Bacillus subtillis.

To investigate whether sigG1 and STSU_11555 are co-transcribed, we assessed the transcription of both genes by RT-PCR using primers pairs that amplified at different locations in the mRNA. The RT-PCR revealed a longer transcript that included both sequences, indicating that sigG1 and STSU_11555 are co-transcribed (Fig. 2a).

The transcriptional start site (TSS) of sigG1 was mapped by 5′RACE PCR and sequencing. This result was confirmed by analysis of existing RNA sequencing (RNA-seq) data. 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.

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 (Fig. 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 σ2–σ4 region and the SnoaL_2 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 pRSPDuet1-1 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 (Fig. 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 (Supplementary Fig. S2).

**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 (α-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 (ΔsigG1 and ΔrsfG, 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 transcription levels by RT-qPCR confirmed expression of rsfG in the rsfG mutant, and SigG1 protein was similarly detected by automated western blot (WEST Simple) (Supplementary Fig. S3) 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 (ΔrsfG vs ΔsigG1) were considered significant for peak calling. Using this approach we identified six peaks dispersed throughout the genome (Fig. 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 (Fig. 3b). Strikingly, no clear peak was identified near the sigG1 promoter suggesting that this sigma factor might not be subject to autoregulation via a positive feedback loop, a common feature of ECFs. SigG1 binds at the promoters of two genes related with morphological differentiation in streptomycetes—alanine dehydrogenase, ald8—and the WhiB-like regulator, whLC39. We also identified three additional SigG1 target genes involved in metal-ion dependent homeostasis, STSU_23024, STSU_32197 and aconitase hydratase, acn. Furthermore, peak 3 is located in the intergenic region between STSU_22045 and STSU_22050 with no apparent promoter region (Fig. 3b).

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 ΔsigG1 by analysing gene expression in MGl-2.5 liquid cultures at mid-exponential phase of growth—72 h (Supplementary Fig. S5). Gene expression results were validated by RT-qPCR on a selection of candidate genes (Supplementary Fig. S6) and the pattern of expression correlated with the one observed in the RNA-seq data. The level of transcripts for hrdB and rspP genes, encoding the primary essential sigma factor (HrdB) and the 30S ribosomal protein S16 (RspP) respectively, was monitored as an internal normalization control. 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). In addition, a careful inspection of RNA-seq data allowed to identify a putative promoter region for peak 3 identified in the ChIP-seq experiment. The data suggests that SigG1 binds to the upstream region of a putative anti-sense RNA of unknown function at the STSU_22050 locus (Supplementary Fig. S4A). Surprisingly, inspection of both ChIP-seq and RNA-seq NGs data unveiled an unexpected genomic rearrangement in the ΔsigG1 mutant. Genomic instability resulting in excision of large fragments of DNA has previously been described in the Streptomyces genus and other bacteria40–44. 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 349,601 to 577,663 bp (Fig. 3a). No deletion was observed in the ΔrsfG strain.

Considering the central role of the ChIP-seq peaks, and the identification of the TSSs for each gene27, we were able to isolate and align promoter sequences harboring the six identified peaks. By inputting the sequences for the sigG1–dependent genes into the MEME Suite tool38, 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 (Fig. 3c), separated by 17 bp non-conserved residues. The spacer size and the -10 box are in accordance with what was previously proposed for ECFs in the ECF56 group families18.

**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 ald that codes for an alanine dehydrogenase enzyme, Ald1. Ald1 converts L-alanine residues, generated by protein turnover, into extra pools of pyruvate that help fuel metabolism for growth promotion in B. subtilis38. Moreover, Ald1 is required for maturation of spores in S. coelicolor86. 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 ΔsigG1 lacked the spore-associated pigmented appearance of the wild-type (Fig. 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 pC3 integration site, located within STSU_17107 in S. tsukubaensis (integration was confirmed by PCR). 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 (Fig. 4b).
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 2 kb region around the target genes. *S. tsukubaensis* ΔrsfG: blue. Negative control (ΔsigG1): 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. P1: STSU_23024p; P2: ald; P3: STSU_22050p; P4: wblCp; P5: STSU_32197p; P6: acnAp.

Table 1. SigG1-binding regions identified by ChIP-seq analysis. FC – fold change. *Distance to START codon. **Causal relationship upon sigG1 deletion, RNA-seq.
We conclude that SigG1 is required for the timely differentiation of *S. tsukubaensis*. Strikingly, ΔrsfG 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 ΔsigG1 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.

**Figure 4.** Phenotypes of the *S. tsukubaensis* wild-type, sigG1 and rsfG null mutants. (a) 10⁶ spores were streak on ISP4 and grown for 14 days. To evaluate the phenotype of the complemented strains, 10⁴ spores were plated on ISP4. The pIJ12333 vector was used to express sigG1 with 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 14 days of growth on ISP4. Scale bar: 2 µm.
Using the RNA-seq data, we evaluated transcription in the ΔrsfG mutant and compared it to transcript levels in the wild-type. The identification of 18 significantly deregulated genes (Table 2; Supplementary Fig. S5; FDR < 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 (GSEA47) of ΔrsfG showed an activation of genes involved in the TCA cycle (Supplementary Fig. S7).

**Table 2.** rsfG-dependent genes identified by RNA-seq analysis (FDR < 0.05). Volcano plot available in Supplementary Fig. S5 FDR false discovery rate, FC fold change.

| Gene ID   | Product                               | ΔrsfG/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 | EtfK/YybS/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, FlbB   | −1.8            | 2.1E−02|

**sigG1 is required for the maintenance of metal-ion homeostasis.** The SigG1 regulon includes two genes whose products are involved in the maintenance of metal-ion availability. STSU_23024 shares homology to genes that have been linked to metal-ion dependent drug efflux systems48. 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 (Supplementary Fig. S4B). 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-ion dependent transporter systems, we further asked if metal-ion homeostasis was compromised in the ΔsigG1 mutants. Using antiSMASH 4.049 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-siderophore, high affinity chelators that sequester ferric iron at very low concentrations to assist in metal-ion internalisation50. These small molecules are therefore secreted in iron-limiting conditions. Culture supernatants from cells grown in iron-limiting conditions exhibited substantially higher levels of siderophores when sigG1 was absent (as shown by the CAS assay in Fig. 5a) and, more strikingly, ΔsigG1 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 (Fig. 5b). Noticeably, the intracellular iron content was restored upon complementation of the mutant strain (Fig. 5b), corroborating the importance of SigG1 in the maintenance 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, ΔsigG1, ΔrsfG, ΔsigG1-rsfG and the ΔsigG1 strain complemented with sigG1, were inoculated on ISP4 copper-replete media (ISP4 supplemented with 100 µM CuSO4). After 6 days of growth, all sigG1/rsfG-related mutant strains exhibited an impaired growth when compared to the wild-type (Fig. 5c). Moreover, ΔsigG1 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 detritmetal concentrations of CuSO4. The dramatic arrest of development of aerial mycelium observed in ΔsigG1 was reverted by expressing the sigG1 wild-type allele in this mutant.

**The S. tsukubaensis 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 H2O2 through the Fenton reaction51—that results in harmful levels of oxidative stress. To examine the impact of the fluctuations in iron availability upon sigG1 dele-
We assessed the maintenance of the redox balance by addressing the response to H$_2$O$_2$. The OxyR regulon is at the core of the bacterial defence against H$_2$O$_2$ 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$_2$O$_2$. After 15 min of H$_2$O$_2$ exposure, we examined sigG1 transcript levels by RT-qPCR. Upon disruption of the major regulator of H$_2$O$_2$, oxyR, we observed an upregulation of sigG1 transcripts in standard conditions, indicating that sigG1 is expressed during the response to the oxidative stress damage (Fig. 6). Although transcription in the wild-type was not affected by the treatment, sigG1 levels were strongly induced by exogenous H$_2$O$_2$ in the absence of OxyR, indicating that sigG1 is activated in response to H$_2$O$_2$-mediated oxidative stress. The transcript levels of rsfG were unchanged upon the ROS insult in these strains.

Figure 5. Evaluation of metal-ion 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$_4$ 32 μM) conditions. (b) Intracellular iron content in S. tsukubaensis wild-type, ΔsigG1, ΔrsfG-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. Statistically significant differences were determined by one-way ANOVA followed by a Dunnet's test (GraphPad Prism).
Deletion of sigG1 impairs secondary metabolism in *S. tsukubaensis*. Morphological differentiation and the oxidative stress response are tightly linked to secondary metabolism in the *Streptomyces* genus. To evaluate if the morphological impairment observed on solid medium and the defective oxidative stress response displayed by ∆sigG1 had an impact on secondary metabolism we assessed the production of FK506. In FK506-producing conditions (MGm-2.5 liquid cultures), ∆sigG1 showed no significant differences in growth when compared to the wild-type (Supplementary Fig. S8A) however we detected a delayed onset of FK506 biosynthesis, which resulted in a significant reduction (> 30%) in the total production when compared to the wild-type (Supplementary Fig. S8B). Nevertheless, we cannot exclude that the effect observed upon the production of FK506 could be influenced by the genomic deletion displayed by the ∆sigG1 strain.

Discussion

ECFs with additional domains have been predicted to rely on the autoregulatory activity of the interaction of the ECF conserved core with domains at the C-termini. SigG1 lacks the conserved NPDKL motif at the CT extension predicted to interact with the YVGPWLPEP motif in the linker between σ2 and σ4 of the σECF41 (Supplementary Fig. S9). In fact, DCA analysis predicted that SigG1 conformation is established by pairs of amino acid residues that are distinct from the ones that are in contact in σECF41. This suggests an alternative regulation of SigG1 in respect to ECF41. In addition, 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. 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.

In previous studies in *S. natalensis*, we showed that the cell death of the vegetative mycelia that precedes the emergence of aerial mycelium results from an increase of the H2O2 intracellular pool. Like in *S. coelicolor*, in *S. tsukubaensis* the LysR regulator OxyR perceives the H2O2 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-ion homeostasis. In this study, we provide evidence that the antioxidant response in *S. tsukubaensis* 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 transcriptional regulator in the proximity of a SigG1 homologue-encoding gene. We hypothesize that SigG1 is likely to be an ECF that responds to H2O2 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 H2O2 exhibited by a ΔsigJ mutant, sigJ mRNA levels are not responsive to H2O2-induced stress.

The control of intracellular H2O2 homeostasis is linked with the homeostasis of metal-ion content as the cells face emergent toxicity due to the Fenton reaction. Characterisation of ΔsigG1 strain suggests that SigG1 participates in the control of fluctuations in iron and copper availability in *S. tsukubaensis*. Note that despite the genome gap displayed by the ΔsigG1 strain, complementation with a copy of sigG1 restored the wild-type phenotype, not only regarding intracellular iron and copper toxicity (Fig. 5) but also the morphological phenotype (Fig. 4). The SigG1 regulon includes genes encoding proteins with iron-sulphur centres that are known regulators...

**Figure 6.** Stress-responsive gene expression. Transcription analyses of oxyR and sigG1 expression in *S. tsukubaensis* mutant strains, upon a H2O2 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. Statistically significant differences to the wild-type were determined by one-way ANOVA followed by a Dunnet's test (GraphPad Prism).
of metal-ion homeostasis, including AcnA and WblC. In streptomycetes, WblC is involved in developmental cascades and, moreover, it regulates the expression of the gene encoding SigK, a well-studied sigma factor responsible for coordinating the response to oxidative stress induced by thiol agents67. Furthermore, SigG1 regulates the expression of two genes encoding putative metal-ion dependent efflux systems described in other bacteria. Together with a disrupted iron balance, we show that strains defective in rsfG were significantly more sensitive to exogenous copper than the wild-type. The mechanisms controlling copper uptake or secretion in Streptomyces species were recently identified68–70. The function of the product of STSU_32197 remains unknown since no clear homologues are identified for this protein. However, one possibility is that SigG1 mediates copper tolerance by activating an operon for the copper chaperone/P-type ATPase system CopA/CopZ (STSU_32197-STSU_32205) responsible for the secretion of these metal-ion in S. lividans. Similarly to ferrous iron, copper ions can also react with intracellular thiols to produce hydroxyl radicals (the Haber Weiss reaction). Moreover, by occupying a high position in the Irving-Williams series, copper can substitute iron as a cofactor for transcriptional regulators or enzyme activity68, which damages iron–sulphur centres due to displacement of the iron ions69. Our hypothesis is that to accompany the transcriptional activation of iron–sulphur clusters containing proteins, SigG1 activates the copA/copZ copper efflux locus to limit copper concentration in the cytoplasm and avoid toxicity. In the absence of sigG1, cells fail to export copper and to avoid an increase in oxidative stress, cells can trigger the export of iron to balance redox active metals in the cytoplasm. A low intracellular iron content prompts bacteria to recover environmental iron bound to siderophores, which in turn suppress copper-induced ROS toxicity65. Previous works have demonstrated that cuproproteins are recruited for the developmental switch from vegetative to aerial mycelium60,65, which is consistent with the timeframe of SigG1 activity. Overall, our results suggest that SigG1 might be involved in regulating iron and copper homeostasis, probably in an effort to counteract the metal-ion imbalance caused by oxidative damage.

Curiously, no SigG1 binding site was identified upstream of rsfG. 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 growth55,66. The potential BldN-dependent regulation of sigG1 could explain the iron limitation and the impaired aerial growth observed for S. tsukubaensis ΔsigG1.

Altogether, these findings provide an indication as to how Streptomyces could protect itself from oxidative stress generated endogenously at the colony core. We hypothesize that as the rate of primary metabolism slows down, nutrient depletion provides the oxidative trigger required to activate SigG1. SigG1 responds by catalysing the expression of Afd1, which detoxifies the alanine from protein turnover, leading to increased pyruvate production during sporulation. AfnA will then promote pyruvate conversion into energy through the TCA cycle. This extra energy, together with a balance in intracellular metal-ion 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 (Supplementary Fig. S7). Moreover, we observed the enhanced transcription of genes encoding other proteins that are directly involved in metal-ion homeostasis (soda, bfr) in the ΔrsfG mutant (p > 0.05).

Transitions between mycelia differentiation stages are 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 decreased the production of FK506 (Supplementary Fig. S8). 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. Although in the latter case we cannot completely exclude an indirect effect of the genomic deletion displayed by ΔsigG1.

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. Overall, the insights gained in this work strengthens the link between morphological development, metal-ion 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 and thrive in the complex environment of the soil.

Materials and methods

Bacterial strains and growth conditions. Bacterial strains used in this study are described in Table 3.

Escherichia coli strains were grown under aeration in LB (Lysogeny Broth) liquid medium or on LB agar at 30 °C or 37 °C. Streptomyces tsukubaensis strains were grown at 30 °C on Diço ISP4 solid medium (BD, NJ, USA) for spore production. Liquid cultures were grown in M6G-2.5 medium at 28 °C, under aeration at 220 rpm34.

Ampicillin at 100 μg/mL, kanamycin 50 μg/mL, hygromycin 25 μg/mL, apramycin 50 μg/mL, chloramphenicol at 25 μg/mL or thiostrepton at 40 μg/mL where added to the media when required. Phenotypic analysis of S. tsukubaensis was conducted on ISP4 solid media or in M6G-2.5 liquid cultures.

Construction of deletion mutant strains. In-frame deletions ΔsigG1::apr, ΔrsfG::apr and ΔsigG1-rsfG::apr were generated using REDIRECT PCR targeting technology67. Briefly, native sigG1 was replaced with the apr-oriT cassette on the 15C1 cosmid using primers red_sigG1_F/Rand red_rsfG_F/R (Supplementary Table S1). Modified cosmid was transferred to S. tsukubaensis WT by intergeneric conjugation68 and apramycin resistant recombinants selected. In-frame deletions were confirmed by PCR (primers red_conf_sigG1_F/R and red_conf_rsfG_F/R, Supplementary Table S1) and by Southern hybridization. For complementation of the sigG1 deletion mutant strain, 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 pJ12333. The construct was introduced into S. tsukubaensis ΔsigG1 by conjugation.
Table 3. Strains and plasmids used in this study.

### Strains or plasmids

| Strains or plasmids | Description | Source/Reference |
|---------------------|-------------|-----------------|
| **S. tsukubaensis** |            |                 |
| NRRL 18488          | Wild-type   | 62              |
| ΔsigG1              | ΔsigG1::apr; Apr<sup>+</sup> | This study |
| ΔrsfG               | ΔrsfG::apr; Apr<sup>+</sup> | This study |
| ΔsigG1-rsfG         | Δ sigG1-rsfG::apr; Apr<sup>+</sup> | This study |
| WT pIJ12333         | Wild-type expressing pIJ12333 empty vector | This study |
| ΔsigG1 pIJ12333 + sigG1 | ΔsigG1::apr expressing pIJ12333-sigG1; Tsr<sup>+</sup> Apr<sup>+</sup> (ΔC31-transduction) | This study |
| ΔsigG1-rsfG pIJ12333 + sigG1 | ΔsigG1-rsfG::apr expressing pIJ12333-stsu_11555-60; Apr<sup>+</sup> (ΔC31-transduction) | This study |
| ΔoxyR               | ΔoxyR::apr; Apr<sup>+</sup> | (Pires et al., unpublished results) |
| **E. coli**         |            |                 |
| BTH101              | F-; cyA99 araD139 galE15 galK16 rpsL11 (Srr); hsdr2 mcrA1 mcrB<sup>+</sup> | 75              |
| BW25113 [pIJ790]    | Δ(araD-araB)5676; ΔlacZ4787::ermB4; lacIq-4000(lacI<sup>Q</sup>), λ, rpsL369(Am<sup>R</sup>), rpp-1, Δ(rbsD-rbsA)5586, hsdr514; Cm<sup>+</sup> | 63              |
| DH5a                | F-::phoA ΔlacZ5807 Δ araC-araD-araB169 recA1 endA1 hsdr17 (rK<sup>+</sup> mK<sup>+</sup> phoA supE44 λ tht-1 gyrA96 relA<sup>A</sup>) | 64              |
| ET10267 [pUZ8002]   | dam, dcm, hsd, Kan<sup>A</sup>, Cm<sup>+</sup> | 64              |
| Nico21              | carc::CBD fhuA2 [λm Tm]<sup>+</sup> ompT gal (A DE3) [dcm] araBAD1; CBD slyD::CBD glmS66Ala Δxys<sup>+</sup> λ DE3=λ sdanH<sup>+</sup> ΔEcoRI-B int::(lacZ::PlacUV5::T7 genel) [s] Ani<sup>+</sup> | New England Biolabs |
| **Plasmids**        |            |                 |
| pIJ12333            | Vector for conjugal transfer of DNA from E. coli to Streptomyces sp.; integrative (ΔC31-transduction), Hyg<sup>B</sup> Tsr<sup>+</sup> | 65              |
| pIJ12333-sigG1      | pIJ12333 carrying sigG1 preceded by its native promoter, Tsr<sup>+</sup> | This study |
| pIJ773              | Plasmid template for amplification of the apr-orfT cassette for Redirect<sup>+</sup> PCR-targeting | 67              |
| pIJ790              | Modified IRED recombination plasmid [oriR101] [repA101(ts)] araBp-gam-be-exo, Cm<sup>+</sup> | 67              |
| pUZ8002             | RP4 derivative with defective oriT, Kan<sup>+</sup> | 64              |
| pET15b              | T7 expression vector, Amp<sup>+</sup> | Novagen |
| pET15b-sigG1        | pET15b-sigG1 full length coding sequence | This study |
| pRSFDuet-1<sup>+</sup> | T7 co-expression vector. Contains two multiple cloning sites (MCS), Kan<sup>+</sup> | Novagen |
| pRSFDuet1-1         | 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<sup>+</sup> | 75              |
| pUT18              | Two-hybrid vector, C-terminal cyAT18fusion, Amp<sup>+</sup> | 75              |
| pKT25              | Two-hybrid vector, N-terminal cyAT25fusion, Kan<sup>+</sup> | 75              |
| pKNT25              | Two-hybrid vector, C-terminal cyAT25fusion, Kan<sup>+</sup> | 75              |
| pKT25-zip           | Derivative of pKT25 carrying the leucine zipper of GCN4 fused in frame to T25 | 75              |
| pUT18C-zip          | Derivative of pUT18C carrying the leucine zipper of GCN4 fused in frame to T18 | 75              |

**Scanning electron microscopy.** Samples for scanning electron microscopy were obtained by growing strains on ISP4 agar for 15 days and prepared as previously described<sup>64</sup> with minor modifications. Isolated colonies were mounted on an aluminium stub using Tissue Tek (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 a FEI Nova NanoSEM 450 (FEI, Eindhoven, The Netherlands). Sublimation of surface frost was performed at − 95 °C for 3½ min before sputter coating the sample with platinum for 2½ min at 10 mA, 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 3 kV.

**S. tsukubaensis genomic DNA isolation.** Genomic DNA from *S. tsukubaensis* strains was obtained with the Master Pure Gram-positive DNA Purification Kit (Epicentre, WI, USA) or with the optimized procedure for Gram-positive bacteria from the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific, MA, USA) 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, WI, USA) or Q5 High-Fidelity DNA Polymerase (New England Biolabs, MA, USA). All oligonucleotides used are listed in Supplementary Table S1.

**RNA isolation and RT-qPCR analyses.** Samples for gene expression studies were harvested at 72 h (mid-exponential phase), 96 h (late exponential phase) or 120 h (early stationary phase) of growth. For H<sub>2</sub>O<sub>2</sub> induced stress experiments, samples were collected at 72 h (t<sub>0</sub>) and 15 min after the addition of 5 mM H<sub>2</sub>O<sub>2</sub> (t<sub>1</sub>). Cul-
ture aliquots were mixed with two volumes of RNA Protect Bacteria Reagent (Qiagen, Hilden, Germany) 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 (Qiagen) according to manufacturer instructions with modifications described in Beites, et al. RNA quality and integrity were evaluated in an Experion Automated Electrophoresis System (Bio-Rad, CA, USA). For cDNA synthesis, 1 μg 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 (Supplementary Table S1) and using, 10 μL of KAPA SYBR FAST RT-qPCR Master Mix (KAPA Biosystems, MA, USA) 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 non-specific products. Analysis included three biological replicates and technical triplicates for each cDNA. The data obtained was analysed using the method described by Pfaffl. For each analysis rpsP and hrdB mRNAs were used for normalization.

**Mapping of the 5’ terminus of mRNA by Rapid Amplification of cDNA Ends (5’ RACE).** Transcriptional start site (TSS) identification was performed using the 5’ RACE System for Rapid Amplification of cDNA Ends, Version 2.0 kit (Thermo Fisher Scientific), following the manufacturer’s instructions. First strand cDNA synthesis was carried out using the gene-specific primer GSP1 (Supplementary Table S1). PCR amplification of tagged cDNA was carried out using the 5’-RACE abridged anchor primer (AAP) with the GSP2 nested primer (Supplementary 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 72 h, 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 (Illumina, CA, USA). The generated DNA fragments (DNA library) were sequenced in the illumina Hiseq 4000 platform, using 150 bp paired-end sequencing reads (Stab-Vida, Portugal).

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

**Secreted siderophore detection.** Siderophore production was assessed using the Chrome Azurol S (CAS) assay. Cultures were grown in MGm-2.5 media with or without ferrous iron supplementation. Samples were harvested at 96 h 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 QuantiChrom Iron Assay Kit (BioAssay Systems, CA, USA) according to the manufacturer’s instructions. Intracellular iron levels were measured by flame atomic absorption spectrometry (F-AAS) following a procedure adapted from Yang et al. Samples were collected from liquid cultures by centrifugation and cell pellets were washed three times with TE buffer (20 mM Tris–HCl, 5 mM EDTA, pH 7.7), followed by one wash with metal-free double distilled water to remove salts. Cells were suspended in 65% (v/v) HNO3, 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 as previously described with minor modifications. Extraction of FK506 was carried out by mixing 1 mL of culture with 1 mL of 100% methanol, for 1 h at 30 °C. The mixture was centrifuged and the supernatant was recovered and analysed using a SunFire C18 column (4.6×150 mm, 3.5 μm; Waters, MA, USA) in a HPLC system (Hitachi, Tokyo, Japan). 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 6His-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 0.2 mM IPTG at 16 °C, 180 rpm. 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, 0.3 M NaCl, 10 μg/mL Dnaasel, 200 μg/mL lysozyme, protease inhibitor cocktail pH 7.4) and lysed using a French press homogenizer at 8000 psi. After centrifugation, the clear extract was loaded onto a 1 mL HisTrap Niquel Sepharose High Performance column (GE healthcare, IL, USA) pre-charged with Ni2+ and equilibrated with binding buffer (20 mM sodium phosphate, 300 mM NaCl, 20 mM 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.
Elution was carried out to an adequate level of purity with approximately 120 mM imidazole, and fractions were collected and analysed 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 (Millipore, MA, USA).

Production of SigG1-specific polyclonal IgY antibodies. The 6His-SigG1 purified protein and the V16-A39 epitope (SigG1V16-A39) were used to induce immunization of quail for the production of the anti-SigG1 polyclonal antibodies in egg yolks at HenBiotec (Coimbra, Portugal).

Immunoblot detection of SigG1 and 6His-RsfG. Total protein extracts were run on 10% SDS-PAGE and transferred to a nitrocellulose membrane. Membranes were incubated with blocking buffer (5% (w/v) dried milk in TPBS/Tris-buffered saline (TBS)) and further incubated with blocking buffer supplemented with anti-SigG1 polyclonal quail antibody (final dilution 1:1000) or tetra-his mouse monoclonal antibody (Qiagen), final dilution 1:5000. After incubation, the membranes were incubated twice in TPBS/Tris-buffered saline, and incubated with secondary antibody—rabbit anti-chicken IgY (IgG) coupled to peroxidase (A9046, Sigma-Aldrich, MI, USA) or anti-mouse IgG conjugated to HRP (Santa Cruz Biotechnology, CA, USA). Signals were revealed with Prime Plus ECL detection kit (Bio-Rad).

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 (Ndel 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 20 h incubation with 1 mM IPTG, at room temperature, with aeration. The molecular weights of the recombinant proteins were verified by nALC-MS/MS. Cells were suspended in lysis buffer (20 mM sodium phosphate, 0.15 M NaCl, 10 µg/mL DNasel, 200 µg/mL lysozyme, protease inhibitor cocktail pH 7.4) and disrupted by mechanical lysis through a FRENCH Press (Thermo Scientific) at 4000 psi. The soluble co-expressed proteins were batch purified in 0.15 M 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 analysed by SDS-PAGE. The protein fractions were concentrated in 10 kDa cut-off Amicon filters. Analytical size exclusion chromatography (SEC) was used to probe the molecular weight of the SigG1-6His-RsfG complex using a Superose12 10/300 GL analytical grade column (GE Healthcare) connected to an AKTA Purifier 10 system (GE Healthcare). Protein was eluted in 20 mM 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: RNase A (13.7 kDa), chymotrypsinogenA (25 kDa), ovalbumin (44 kDa) and BSA (66 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, CA, USA) with the Wes No Secondary Detection (12 to 230 kDa) Master kit. Protein samples were prepared in Water (25 kDa), ovalbumin (44 kDa) and BSA (66 kDa).

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 Gallagher, et al.²⁴ with the following modifications. For immunoprecipitation of SigG1 cross-linked DNA, the total extract was incubated overnight with polyclonal anti-SigG1V16-A39 antibody at 4 °C. Immunoprecipitation was carried out using a goat anti-chicken IgY agarose (ab76444, Abcam) for 4 h. Genomic DNA libraries enriched for SigG1 binding were produced from these samples, size selected to ~100–500 bp, 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, Souffleuwersheim, France) were performed as described previously²⁵. Briefly, the DNA fragments that encode the full-length RsfG and SigG1 proteins were cloned into BACTH T18 and T25 containing vectors. 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 supplemented with 0.3% lactose, 0.5 mM Isopropyl-b-D-1-thiogalactopyranoside (IPTG), the β-galactosidase chromogenic indicator 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-Gal, 40 µg/ml) and appropriate antibiotics, at 30 °C for 3–7 days. The positive clones were assessed using the β-galactosidase chromogenic indicator 5-bromo-4-chloro-3-indolyl-b-D-galacto-pyranoside (X-Gal, 40 µg/ml) and appropriate antibiotics, at 30 °C for 3–7 days. The positive clones were assessed using the β-galactosidase chromogenic indicator 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-Gal, 40 µg/ml) and appropriate antibiotics, at 30 °C for 3–7 days.
Bioinformatics procedures. ECF phylogenetic trees were built as described in Casas-Pastor et al.\(^{17}\). For Direct Coupling Analysis (DCA) protein sequences of members of groups ECF41 and ECF56 were retrieved from the most recent ECF classification\(^{17}\) and aligned using Clustal Omega 1.2.3\(^{17}\). DCA was performed using Gaussian DCA\(^{18}\) as described in Wu, et al.\(^{19}\). Calculation of differentially expressed genes: after trimming (ambiguous limit = 2 nt, quality limit = 0.01), RNA-seq generated reads were mapped onto the concatenated version of the reference genome (see ChIP-seq mapping part in this section). Gene expression levels were determined based on the transcripts per Million (TPM). Expression with log2 fold change ≥ log2 (1.5) and q-value ≤ 0.05 or log2 fold 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 20 (Qiagen). Functional enrichment analyses: regulators 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\(^{49}\) using Gene set collections obtained by searching the S.\(_{\text{tsukubaensis}}\) genome for homologues of proteins involved in sporulation and TCA through NCBI BLASTp analyses (https://blast.ncbi.nlm.nih.gov). The ΔsigG1 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: 10,000 permutations, classic scoring scheme and meandiv normalization. ChIP-seq data analysis: reads resulting from paired-end sequencing were aligned to the genome of S.\(_{\text{tsukubaensis}}\) using the bowtie2\(^{79}\) and further treated as described in Gallagher et al.\(^{74}\). Enrichment for the ΔsigG1 control samples was subtracted from the enrichment in ΔrsfG samples. Significance of enrichment values was calculated assuming normal distribution of the enrichment values. Results were visualized in the Integrated Genome Browser\(^{80}\), or in the CLC Genomics Workbench 20 (Qiagen). The reference genome was the concatenation of all contigs available on NCBI at the time (genome assembly ASM29715v2; WGS project AJSZ01). However, after this article was submitted for review, genome assembly ASM29715v2 was superceded by a new version (ASM29715v3) with an updated annotation (GenBank accession numbers: CP029157, CP029158 and CP029159). Whilst this does not affect the results or claims of this article, the locus tags that identify genes have changed. For instance, the locus tag for STSU_11560 has changed to STSU_011570 in the latest genome assembly version.

Statistical analyses. For each experiment, we assayed at least three independent biological replicates. Statistical significance was addressed through the GraphPad Prism 8 software, according to the requirements of each data set.

Data availability
Sequencing data was submitted in GEO under the following accession codes: GSE144815 for RNA-seq data and GSE144907 for ChIP-seq data.

Received: 24 July 2020; Accepted: 26 November 2020
Published online: 10 December 2020

References
1. Beites, T. et al. *Streptomyces natalensis* programmed cell death and morphological differentiation are dependent on oxidative stress. *Sci. Rep.* 5, 12887 (2015).
2. Claessen, D. et al. The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlinis and chaplinis. *Mol. Microbiol.* 53, 433–443 (2004).
3. Elliot, M. A. et al. The chaplinis: a family of hydrophobic cell-surface proteins involved in aerial mycelium formation in *Streptomyces coelicolor*. *Gene Dev.* 17, 1727–1740 (2003).
4. Flärdh, K. & Buttner, M. J. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* 7, 36–49 (2009).
5. McCormick, J. R. & Flärdh, K. Signals and regulators that govern *Streptomyces* development. *FEBS Microbiol. Rev.* 36, 206–231 (2012).
6. Bush, M. J., Tschowri, N., Schlimpert, S., Flärdh, K. & Buttner, M. J. c-di-GMP signalling and the regulation of developmental transitions in streptomycetes. *Nat. Rev. Microbiol.* 13, 749–760 (2015).
7. Ulrich, L. E., Koonin, E. V. & Zhulin, I. B. One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.* 13, 52–56 (2005).
8. Francis, V. I. & Porter, S. L. Multikinase networks: two-component signaling networks involving multiple stimuli. *Annu. Rev. Microbiol.* 73, 199–223 (2019).
9. Helmmann, J. D. Where to Begin? Sigma factors and the selectivity of transcription initiation in bacteria. *Mol. Microbiol.* 112, 335–347 (2019).
10. Lentonno, M. A., Donohue, T. J., Gross, C. A. & Buttner, M. J. Discovery of the extracytoplasmic function σ factors. *Mol. Microbiol.* 112, 348–355 (2019).
11. Helmmann, J. D. The extracytoplasmic function (ECF) σ factors. *Adv. Microb. Physiol.* 46, 47–110 (2002).
12. Lentonno, M., Gribkov, M. & Gross, C. A. The sigma 70 family: sequence conservation and evolutionary relationships. *J. Bacteriol.* 174, 3843–3849 (1992).
13. Lane, W. J. & Darst, S. A. The structural basis for promoter -35 element recognition by the group IV σ factors. *PloS Biol.* 4, 1491–1500 (2006).
14. Lin, W. et al. Structural basis of ECF-σ-factor-dependent transcription initiation. *Nat. Commun.* 10, 710–710 (2019).
15. Fang, C. et al. Structures and mechanism of transcription initiation by bacterial ECF factors. *Nucleic Acids Res.* 47, 7094–7104 (2019).
16. Gaballa, A. et al. Modulation of extracytoplasmic function (ECF) σ factor promoter selectivity by spacer region sequence. *Nucleic Acids Res.* 46, 134–145 (2018).
17. Casas-Pastor, D. et al. Expansion and re-classification of the extracytoplasmic function (ECF) σ factor family. Preprint at https://doi.org/10.1101/2019.12.11.873521v2 (2019).
18. Huang, X., Pinto, D., Fritz, G. & Mascher, T. Environmental sensing in Actinobacteria: a comprehensive survey on the signaling capacity of this phylum. *J. Bacteriol.* 197, 2517–2535 (2015).
63. Macomber, L. & Imlay, J. A. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8344–8349 (2009).

64. Krumshcabel, G., Manzl, C., Berger, C. & Hofer, B. Oxidative stress, mitochondrial permeability transition, and cell death in Cu-exposed trout hepatocytes. *Toxicol. Appl. Pharmacol.* **209**, 62–73 (2005).

65. Fujimoto, M. et al. Pleiotropic role of the Sco1/SenC family copper chaperone in the physiology of *Streptomyces*. *Microbiol. Biotechnol.* **5**, 477–488 (2012).

66. Traxler, M. F., Seyedsayamdost, M. R., Clardy, J. & Kolter, R. Interspecies modulation of bacterial development through iron competition and siderophore piracy. *Mol. Microbiol.* **66**, 628–644 (2012).

67. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. 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**, 1541–1546 (2003).

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

69. Bush, M. J., Chandra, G., Findlay, K. C. & Buttner, M. J. Multi-layered inhibition of *Streptomyces* development: BldO is a dedicated repressor of whiB. *Mol. Microbiol.* **104**, 700–711 (2017).

70. Pfaff, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).

71. Shin, S. H., Lim, Y., Lee, S. E., Yang, N. W. & Rhee, J. H. CAS agar diffusion assay for the measurement of siderophores in biological fluids. *J. Microbiol. Methods* **44**, 89–95 (2001).

72. Yang, J. H. et al. *Bradyrhizobium japonicum* senses iron through the status of haem to regulate iron homeostasis and metabolism. *Mol. Microbiol.* **60**, 427–437 (2006).

73. Ordonez-Robles, M., Santos-Beneit, F., Rodriguez-Garcia, A. & Martin, J. F. Analysis of the Pho regulon in *Streptomyces tsukubaensis*. *Microbiol. Res.* **205**, 80–87 (2017).

74. Gallagher, K. A. et al. c-di-GMP arms an anti-σ to control progression of multicellular differentiation in *Streptomyces*. *Mol. Cell** **77**, 586–599 (2019).

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

76. Miller, J. H. *Experiments in Molecular Genetics*. (Cold Spring Harbor Laboratory Press, 1972).

77. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

78. Baldassi, C. et al. Fast and accurate multivariate gaussian modeling of protein families: predicting residue contacts and protein–protein interaction partners. *PLoS ONE* **9**, e92721 (2014).

79. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

80. Nicol, J. W., Helt, G. A., Blanchard, S. G. Jr., Raja, A. & Loraine, E. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25**, 2730–2731 (2009).

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

82. Barreiro, C. et al. Draft genome of *Streptomyces tsukubaensis* NRRL 18488, the producer of the clinically important immunosuppressant tacrolimus (FK506). *J. Bacteriol.* **194**, 3756–3757 (2012).

83. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

84. Baldassi, C. et al. Fast and accurate multivariate gaussian modeling of protein families: predicting residue contacts and protein–protein interaction partners. *PLoS ONE* **9**, e92721 (2014).

85. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

86. Nicol, J. W., Helt, G. A., Blanchard, S. G. Jr., Raja, A. & Loraine, E. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* **25**, 2730–2731 (2009).

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

88. Barreiro, C. et al. Draft genome of *Streptomyces tsukubaensis* NRRL 18488, the producer of the clinically important immunosuppressant tacrolimus (FK506). *J. Bacteriol.* **194**, 3756–3757 (2012).

89. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).

90. Paget, M. S. B., Chamberlin, L., Atri, A., Foster, S. J. & Buttner, M. J. Evidence that the extracytoplasmic function sigma factor of is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **181**, 204–211 (1999).

91. Sherwood, E. J., Hesketh, A. R. & Bibb, M. J. Cloning and analysis of the planosporicin lantibiotic biosynthetic gene cluster of *Planomonospora alba*. *J. Bacteriol.* **195**, 2309–2321 (2013).

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 ECExpress (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 *S. tsukubaensis* samples, Mervyn Bibb (JIC, UK) for the plJ12333 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*.

Author contributions

R.O., M.J.B. and M.V.M conceived and designed the work; R.O., and S.P. performed the experiments; D.C.P. and G.F. designed and performed the DGA experiments and the phylogenetic analyses; R.O., G.C. and M.V.M analysed the NGS data; R.O., M.J.B., D.C.P., G.F. and M.V.M wrote and edited the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-78520-x.

Correspondence and requests for materials should be addressed to M.V.M.

Reprints and permissions information is available at www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020