A genomic island of *Streptomyces coelicolor* with the self-contained regulon of an ECF sigma factor

Camilla M. Kao,[a,□]# Nitsara Karoonuthaisiri,[a,□1] David Weaver,[a,□2] Jonathan A. Vroom,[a,□3] Shuning A. Gai,[a,□4] Mai-Lan Ho,[a,□5] and Kedar G. Patel[a]

[a]Department of Chemical Engineering, Stanford University, Stanford, California, USA

Running Head: GI 6 of *S. coelicolor* with ECF sigma factor regulon

#Address correspondence to Camilla M. Kao, ckao.science@gmail.com.

□ Present Address: School of Engineering, Stanford University, Stanford, California, USA

□1 Present Address: Biosensing Technology Research Unit, National Center for Engineering and Biotechnology, Pathum Thani, Thailand

□2 Present Address: ERM: Environmental Resources Management, San Francisco, California, USA

□3 Present Address: Molecular and Cellular Engineering, Codexis, Redwood City, California, USA

□4 Present Address: Molecular Biology, Prothena Biosciences, South San Francisco, California, USA

□5 Present Address: Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA

N.K. and D.W. contributed equally to this work.

Word count for the abstract: 400

Word count for the text: 2897

Abstract

Streptomycetes constitute the largest genus of actinobacteria, living predominantly in soil and decaying vegetation. The bacteria are widely known for their filamentous morphologies and their capacity to synthesize antibiotics and other biologically active molecules. More than a decade ago, we and others identified 22 genomic islands that *Streptomyces coelicolor* M145 possesses and other *Streptomyces* strains lack. One of these genomic islands, Genomic Island (GI) 6, encodes an extracytoplasmic function (ECF) sigma factor that we were characterizing in separate work. Here we report that artificial induction of the ECF sigma factor, which is encoded by SCO3450, causes the transcription of approximately one-fourth of GI 6, or ~26 mostly contiguous genes, to increase. More than half of the regulon encodes putative enzymes involved in small molecule metabolism. A putative haloacid dehalogenase is present. Genes encoding two putative anti-sigma factors flank SCO3450, the three genes residing within the

regulon. Our data suggest that the ECF sigma factor and its regulon are a self-contained transcriptional unit that can be transferred by horizontal gene transfer. To our knowledge, only one other example has been identified of an ECF sigma factor and its contiguous regulon appearing to be transferrable by horizontal gene transfer [18,19]. Because the regulon appears not to be induced by the 44 growth conditions recently examined by Byung-Kwan Cho and colleagues [20], if it confers fitness to *S. coelicolor*, the regulon likely does so in as-yet unknown situations. Those situations might range from scavenging to detoxification to even communication within microbial communities.

IMPORTANCE *Streptomyces* bacteria grow as hyphae that colonize soil and differentiate into spores when nutrients become scarce. In their terrestrial habitats, the bacteria encounter diverse conditions. Presumably so that the bacteria can cope with those conditions, the chromosomes of streptomycetes are highly dynamic, varying greatly in structure not only between species but also between closely related strains of a single species. The bacteria also have large numbers of extracytoplasmic function (ECF) sigma factors, which undoubtedly help the microorganisms respond to the plethora of challenges coming from the environment. This work illustrates these two threads of *Streptomyces* biology dovetailing: Genetic adaptability through horizontal gene transfer seems to have enabled *Streptomyces coelicolor* to acquire a self-contained transcriptional unit that consists of an ECF sigma factor and its regulon. The suggested facile movement of the regulon between microbial hosts indicates the value of the metabolism of small molecules possibly mediated by the regulon.

Introduction

Researchers began studying *Streptomyces* genetics in the second half of the 20th century because of the filamentous morphologies and biosynthetic capacities of the bacteria [1]. Many of the *Streptomyces* strains that were studied belong to a group of closely related "blue" streptomycetes that had been given various names, but probably should all have the same name [2]. During the early 2000's, we characterized the genomes of six of these "blue" strains by using DNA microarrays: the sequenced strain [3], *S. coelicolor* M145, which derives from *S. coelicolor* A3(2), which in turn derives from Waksman's strain 3443 [4]; Sermonti's SE1 (John Innes (JI) strain 1152) [5]; Bradley's S199 (JI 1153) [6]; *S. lividans* 66 (JI 1326) [7]; *S. lividans* ISP5434 (JI 2896), which contains the plasmid pIJ101 [8]; and *S. violaceoruber* SANK95570 (JI 3034), which contains the plasmid pSV1 that encodes the biosynthetic genes for the antibiotic methylenomycin [9]. That study was described in the doctoral dissertation of co-author David Weaver [10]. Similar findings were reported by Jayapal et al., who examined *S. coelicolor* M145 and *S. lividans* TK21 [11].

By comparing the genome of *S. coelicolor* M145 to the five wild type genomes by using DNA microarrays, we identified 22 sets of contiguous genes in *S. coelicolor* M145 that are absent in the wild type strains [10]. We designated these genes Genomic Islands (GIs) 1 to 22. Fig. S1, Fig. S1, and Fig. S3 show the GIs. The sizes of the GIs range from 3 kb to 150 kb and the number of ORFs within them from 3 to 148. *S. coelicolor* M145 likely acquired the 22 GIs by horizontal

gene transfer, because they have the following characteristics: Direct repeats flank most of the GIs; the GIs contain regions with low G + C content; the GIs encode transposable elements such as invertases, recombinases, and transposons, as well as plasmid-related proteins; some GIs appear to have inserted into tRNA sequences; and the GIs have slightly lower gene expression during exponential growth [10,11]. We found that different wild type strains lack different GIs and have different boundaries at the sites corresponding to the GIs in *S. coelicolor* M145. We also identified DNA that the wild type strains possess and that *S. coelicolor* M145 lacks. Some of that DNA (corresponding to GIs 8, 14, 17, 18, and 19 in *S. coelicolor* M145) might be large segments, because we were unable to amplify the DNA by PCR [10]. The genomic islands have a relatively uniform distribution across the chromosome of *S. coelicolor* M145, with a slight abundance in the right chromosome arm [10,11].

A sigma factor is a specialized unit of RNA polymerase that permits the multisubunit enzyme to initiate transcription selectively. Extracytoplasmic function (ECF) sigma factors constitute the most abundant, smallest, and most divergent group of sigma factors [12]. Their name derives from many of their members having roles in sensing and responding to signals generated outside of the cell or in the cell membrane [13]. The genome of *S. coelicolor* M145 encodes 63 sigma factors, of which 49 belong to the ECF group [3]. During the early 2000's, we sought to identify regulons of *S. coelicolor* sigma factors by overexpressing each one individually and analyzing the response of the transcriptome by using DNA microarrays. The ECF sigma factor encoded by SCO3450 was evaluated as part of that work. While we were unaware of this fact at that time, SCO3450 is located in GI 6 [10].

The ECF sigma factor encoded by SCO3450 belongs to the subgroup ECF01 as classified by Thorsten Mascher and colleagues in 2009 [14]. This subgroup includes RpoE-like sigma factors, of which the best studied is σ^W of *Bacillus subtilis*. Members of the ECF01 subgroup are widely distributed throughout bacterial phyla, but are absent in most actinobacteria, the phylum which includes the genus *Streptomyces* [15]. ECF01 sigma factors have been characterized experimentally to be involved in responses to envelope stress and the production and detoxification of antimicrobial compounds [14].

Results and Discussion

Genomic Island 6 encodes an ECF sigma factor and the regulon of the sigma factor

Table 1 lists the five wild type *Streptomyces* strains that we studied that lack GI 6 [10].

Fig. 1 shows the boundaries of GI 6 (the left-hand columns) and the levels of transcripts, as measured by microarrays, after artificial induction of SCO3450 in the strain *S. coelicolor* M600 growing exponentially in a liquid culture (the right-hand columns) [16]. Fifteen minutes after we induced the sigma factor, transcripts of approximately one-fourth of the genes in GI 6 increased in abundance. The genes numbered ~26: SCO3437 and SCO3442-3465; SCO3478 is an induced gene located 14 kb away from the other genes. Cross-hybridization does not explain the result for SCO3478 because the gene, which encodes a dehydrogenase, lacks sequence similarity to

97 other *S. coelicolor* genes. SCO3450 is located among the induced genes (Fig. 1, arrow; Fig. S4).
98 Together these data suggest that the ECF sigma factor and its regulon are a self-contained
99 transcriptional unit that can be transferred by horizontal gene transfer.

100

101 To our knowledge, the only other example of an ECF sigma factor and its regulon appearing to
102 be transferrable by horizontal gene transfer is the system for cobalt and nickel resistance in the
103 bacterium *Cupriavidus metallidurans* CH34 [18,19]. There, the plasmid pMOL28 contains six *cnr*
104 genes that are organized in two adjacent operons of three genes each. The gene *cnrH* encodes
105 the ECF sigma factor CnrH. Deletion and complementation of *cnrH* indicated that CnrH is
106 essential for the regulation of the *cnr* system. When CnrH was present but the periplasmic
107 sensor CnrX and the anti-sigma factor CnrY were absent, high-level constitutive expression was
108 observed for the *cnr* genes [18].

109

110 *The regulon of the ECF sigma factor contains conserved putative -35 and -10 promoter regions*

111 Byung-Kwan Cho and colleagues reported in 2016 transcription start sites of *S. coelicolor* genes
112 induced by 44 growth conditions [20]. Those conditions included growth in rich media and
113 many kinds of minimal media, growth in the liquid phase and on solid media, rapid growth to
114 stationary phase, and several kinds of shocks. That study did not identify transcription start
115 sites for the regulon of the ECF sigma factor encoded by SCO3450. This observation indicates
116 that the regulon is not induced by the growth conditions examined by Jeong et al. [20]. Perhaps

117 *S. coelicolor* uses the regulon for circumstances that are not replicated in the laboratory, such
118 as interactions in the natural environment among different microbial species.

119

120 We used the bioinformatics tool PromoterHunter [21], combined with visual inspection of
121 nucleotide sequences, to identify possible -35 and -10 promoter regions in the regulon of the
122 ECF sigma factor. The conserved sequences for the -35 and -10 promoter regions might be
123 AACGG and CG, respectively (Fig. 2). The promoters of 14 genes contain these sequences
124 exactly. The data suggest that operons might comprise SCO3445 to SCO3444, SCO3448 to
125 SCO3446, SCO3455 to SCO3452, and SCO3463 to SCO3465 (see Fig. S4). Alternatively, genes not
126 listed in Fig. 2 might have promoters with less conserved -35 and -10 regions. Note that
127 SCO3460 and SCO3478 do not possess the highly conserved promoter regions (see Fig. S4). In
128 addition, that SCO3438 has the conserved promoter regions supports the notion that this gene
129 belongs to the regulon of the ECF sigma factor. In Fig. 1, the lack of data for SCO3438 from all of
130 the microarray hybridizations indicates that its spot on the microarrays was functioning poorly.

131

132 *The regulon of the ECF sigma factor encodes putative enzymes for small molecule metabolism*

133 Recently we updated the annotations of the regulon of the ECF sigma factor by using
134 EnsemblBacteria [22] and UniProt BLASTS [23]. Of the approximately 26 genes in the regulon,
135 15 encode putative enzymes for small molecule metabolism. 11 genes encode putative
136 membrane and transport proteins. Four genes encode putative regulatory proteins. Four of the

137 26 genes fall into two categories. Only one gene lacks a putative function. Table S1 lists the
138 annotations of each gene in the regulon.

139

140 *Enzymes encoded by the regulon of the ECF sigma factor*

141 Table 2 lists the types of enzymes encoded by the regulon of the ECF sigma factor. The table
142 excludes SCO3461 and SCO3462, because they lack gene expression data (Fig. 1, right-hand
143 columns).

144

145 The putative enzymes fall into three general groups. One group possibly catalyzes the cleavage
146 of small molecules. The group includes a glycoside hydrolase, a nucleoside phosphorylase, and
147 a haloacid dehalogenase. A second group of enzymes might catalyze oxidation and reduction
148 reactions. The group includes a glutaredoxin-like protein, two oxidoreductases, and a
149 dehydrogenase. The third group of enzymes might catalyze transfer reactions. The group
150 includes two methyltransferases, perhaps a phosphatidyltransferase, and two nucleotide-
151 diphospho sugar transferases. A final gene, SCO3445, encodes a small membrane protein that
152 has 30-50% identity to small regions of seven larger proteins, a set which includes several
153 enzymes.

154

155 The composition of the enzymes in the regulon of the ECF sigma factor suggests that the
156 regulon might help to metabolize small molecules. A mixture of compounds might be being

157 degraded: sugars, nucleosides and nucleotides, and lipids for the purposes of energy and
158 biosynthesis; halogenated molecules for detoxification. Because the regulon encodes a set of
159 putative enzymes with diverse substrates, the small molecules might be coming from the
160 environment, for example, from neighboring cellular compartments or hyphae that are lysing
161 due to hostile conditions.

162

163 Haloacid dehalogenases (HADs) belong to a large superfamily of hydrolases with diverse
164 substrate specificity [24]. Type II HADs catalyze the hydrolytic dehalogenation of small L-2-
165 haloalkanoic acids to yield the corresponding D-2-hydroxyalkanoic acids [25]. Because many
166 *Streptomyces* bacteria produce halogenated antibiotics [26,27], the dehalogenase encoded by
167 SCO3446 might serve to defend *S. coelicolor* against competitors in the environment by helping
168 to catabolize antibiotics.

169

170 The following pairs of genes have no significant similarity, such that no cross-hybridization on
171 the microarrays should have occurred: SCO3452 and SCO3449, which encode the
172 methyltransferases; SCO3443 and SCO3460, which encode the oxidoreductases; and SCO3464
173 and SCO3465, which encode the nucleotide-diphospho-sugar transferases (Table 2).

174

175 *Anti-sigma factors within the regulon of the ECF sigma factor*

176 Two genes that possibly encode anti-sigma factors flank SCO3450, the gene encoding the ECF
177 sigma factor. The gene product of SCO3451 encodes a 103-amino-acid protein. A UniProt BLAST
178 shows that the protein has approximately 40% identity with varying coverages to putative anti-
179 sigma factors of other bacterial species, including putative transmembrane anti-sigma factors.
180 The Constrained Consensus TOPology (CCTOP) prediction server [28] predicts with a reliability
181 of 95.9069 that SCO3451 encodes a protein with one transmembrane segment (Fig. S5). The
182 gene product of SCO3451 also contains a putative zinc-finger found in some anti-sigma factor
183 proteins (Fig. S6). This zinc finger domain overlaps with the predicted transmembrane domain.

184

185 SCO3449 encodes a 106-amino-acid protein. A UniProt BLAST shows that the protein has
186 approximately 40% identity with coverages around 50% to putative anti-sigma factors in other
187 bacterial species, including putative transmembrane anti-sigma factors. However, according to
188 CCTOP, the protein does not contain a transmembrane domain. Like the putative anti-sigma
189 factor encoded by SCO3451, the putative anti-sigma factor encoded by SCO3449 contains a
190 possible zinc-finger found in some anti-sigma factor proteins (Fig. S7).

191

192 Microarray data are lacking for the expression of SCO3449 (Fig. 1, right-hand columns), such
193 that inferences about the role of this gene in comparison to those of the induced genes should
194 be made with caution. However, because microarray data are available for this gene in the
195 hybridizations that identified the genomic islands (Fig. 1, left-hand columns), the spot on the
196 microarrays that represented this gene was likely intact. It is possible that a low abundance of

197 transcripts of SCO3449 in both the reference and experimental samples of mRNA produced
198 poor signals from the microarray spot of the gene.

199

200 *Membrane and transport proteins encoded by the regulon of the ECF sigma factor*

201 Four contiguous genes in the regulon of the ECF sigma factor, SCO3453 to SCO3456, encode
202 proteins that likely constitute an ABC transporter system. UniProt BLASTS show that all of the
203 proteins have 45-50% identity with coverages greater than 95% to homologs with putative
204 functions in spermidine and putrescine transport in other bacterial species. In particular, the
205 species include *Geodermatophilus* and *Wenxinia*. Spermidine is a polyamine involved in cellular
206 metabolism that can be used to stimulate RNA polymerase. Putrescine attacks S-adenosyl
207 methionine and converts it to spermidine [29].

208

209 The putative transporters Sco3454 and Sco3455 have 34% identity with each other, the
210 coverage being 44% between the C-terminal portions of the proteins.

211

212 *Genomic Island 6 consists of four segments with transposases at their boundaries*

213 In addition to the regulon of the ECF sigma factor, we updated the annotations of the other
214 genes of Genomic Island 6 by using EnsemblBacteria and UniProt BLASTS (Table S2). In the
215 table, the location of the regulon is denoted by the words “REGULON HERE.”

216

The length of GI 6 is 108 kb. The genomic island consists of four segments of DNA bounded by transposases (Table S2, pink color). Three of the four segments have coherent putative functions: the oxidation and reduction of copper (Table S2, blue color); a characterized agarase encoded by *dagA* [30,31]; and the utilization of sugars (Table S2, yellow color). Included in the latter segment are a putative *lacI*-family transcriptional regulator and a putative β-galactosidase. A large fourth segment encodes putative enzymes and many hypothetical proteins.

The coherent functions of at least three of the four segments of GI 6 indicate that significant portions of the island might be "active." If so, the regulon of the ECF sigma factor is likely active as well. While a nucleotide blast of the regulon of the ECF sigma factor yielded no similar segments of DNA in other sequenced organisms, the value to *S. coelicolor* of the regulon for as-yet unknown reasons is indicated by the presence of the regulon on a genomic island. It would be interesting to determine how common the regulon is among natural populations of microbes and whether the regulon moves easily between hosts.

Materials and Methods

Construction of a S. coelicolor *strain that overexpresses SCO3450*

SCO3450, the gene which encodes the ECF sigma factor harbored by GI 6, was amplified by PCR and cloned into the conjugative plasmid pIJ6902 [32] by using the restriction sites *Nde*I and

237 *Bgl*II. These restriction sites placed the gene immediately downstream of the thiostrepton-
238 inducible promoter *tipA*p. The resulting plasmid was transformed into the methylation-
239 deficient strain ET12567 with a non-transmissable helper plasmid, pUZ8002, and conjugated
240 into *S. coelicolor* M600 as described by Kieser et al. [33]. The exconjugants were selected by an
241 overlay of 50 μg/mL of apramycin.

242

243 *Time course of the overexpression strain*

244 Supplemented minimal medium (SMM) was inoculated with approximately 5 x 10^7 spores/mL
245 of a spore stock. The culture was grown at 30°C to early exponential phase, which
246 corresponded to a cell density of OD_{450} ~ 0.5. A reference sample, designated "0 min," was
247 harvested. To induce transcription of SCO3450, thiostrepton was added to the culture to a final
248 concentration of 30 μg/mL. Samples of the culture were harvested 15, 30, 45, 60, and 90
249 minutes after induction of SCO3450.

250

251 *Extraction of total RNA from the overexpression strain*

252 Samples of cells from liquid cultures were recovered by filtration on Whatman filter paper (15-
253 20 mm diameter; catalog #1002 055). RNA was isolated using the modified Kirby mix protocol
254 as described previously [34] with the following modifications: Harvested volumes ranged
255 between 5 mL and 20 mL, depending on the cell densities. RNA samples were treated only once
256 with DNase I (50-70 units; RNase-free, Invitrogen) for 15 minutes at room temperature.

257

258 *DNA microarray experiments for the overexpression strain*

259 Samples of cDNA were synthesized from total RNA as described previously [34]. cDNA of the
260 reference sample was labeled with Cy3-CTP. cDNAs of the samples isolated at subsequent time
261 points were labeled with Cy5-CTP. The cDNAs were hybridized to microarrays as described
262 previously [34].

263

264 *Identification of -35 and -10 conserved promoter regions in the regulon of the ECF sigma factor*

265 The tool PromoterHunter [21] was used to examine DNA upstream of the genes in the regulon
266 of the ECF sigma factor. Weight matrices corresponding to AAC for the -35 promoter region and
267 CG for the -10 promoter region were used, because they reflect the conserved promoter
268 regions of the subgroup ECF01 of ECF sigma factors [14]. A global G + C content of 72% was
269 used. The space between the -35 and -10 regions was specified to be between 17 to 21 base
270 pairs. Initially DNA segments of 300 base pairs were examined for isolated genes and genes
271 located at the beginning of likely operons. Visual inspection of sequences returned by
272 PromoterHunter identified AACGG and CG as possible consensus sequences for the -35 and -10
273 promoter regions, respectively. PromoterHunter was used to search for these conserved
274 sequences within DNA segments of 1000 base pairs upstream of all of the genes in the regulon,
275 in order to ensure that all instances of the sequences were identified.

276

277 **Acknowledgements**

278 We thank Thomas M. Privalsky for informing us about bioinformatic tools and Mark J. Buttner
279 and David A. Hopwood for discussions about the data.

280

281 **Author Contributions**

282 C.M.K. conceived the experiments. D.W. designed the experiments that compared the genomes
283 of the six *Streptomyces* strains and performed those experiments with J.A.V., M.L.H., and K.G.P.
284 N.K. designed the experiment overexpressing the ECF sigma factor and performed the
285 experiment with S.A.G. C.M.K. conducted the recent bioinformatic analyses and wrote the
286 manuscript.

287

288 **References**

289 1. Hopwood DA. 2007. Streptomyces in nature and medicine: the antibiotic makers. Oxford
290 University Press, New York, NY.
291 2. Kutzner HJ, Waksman SA. 1959. Streptomyces coelicolor Mueller and Streptomyces
292 violaceoruber Waksman and Curtis, two distinctly different organisms. J Bacteriol
293 78:528—538.
294 3. Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris
295 DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M,
296 Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke

L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141—147.

4. Hopwood DA. 1999. Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. Microbiology 145:2183—2202.
5. Sermonti G, Spada-Sermonti I. 1955. Genetic recombination in *Streptomyces*. Nature 176:121.
6. Bradley SG. 1960. Reciprocal crosses in Streptomyces coelicolor. Genetics 45:613—619.
7. Lomovskaya ND, Mkrtumian NM, Gostimskaya NL, Danilenko VN. 1972. Characterization of temperate actinophage phi C31 isolated from Streptomyces coelicolor A3(2). J Virol 9:258—262.
8. Kieser T, Hopwood DA, Wright HM, Thompson CJ. 1982. pIJ101, a multi-copy broad host-range Streptomyces plasmid: functional analysis and development of DNA cloning vectors. Mol Gen Genet 185:223—228.
9. Aguilar A, Hopwood DA. 1982. Determination of methylenomycin A synthesis by the pSV1 plasmid from *Streptomyces violaceus-ruber* SANK 95570. J Gen Microbiol 128:1893—901.
10. Weaver D. 2005. PhD thesis. Stanford University, Stanford, CA. Genome plasticity in *Streptomyces coelicolor*.

11. Jayapal KP, Lian W, Glod F, Sherman DH, Hu WS. 2007. Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*. BMC Genomics 8:229.

12. Feklístov A, Sharon BD, Darst SA, Gross CA. 2014. Bacterial sigma factors: a historical, structural, and genomic perspective. Annu Rev Microbiol 68:357—376.

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

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

15. ECFfinder. http://ecf.g2l.bio.uni-goettingen.de:8080/ECFfinder/.

16. Takano E, White J, Thompson CJ, Bibb MJ. 1995. Construction of thiostrepton-inducible, high-copy-number expression vectors for use in *Streptomyces* spp. Gene 166:133—137.

17. Huang J, Lih CJ, Pan KH, Cohen SN. 2001. Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. Genes Dev 15:3183—3192.

18. Grass G, Grosse C, Nies DH. 2000. Regulation of the *cnr* cobalt and nickel resistance determinant from *Ralstonia* sp. strain CH34. J Bacteriol 182:1390—1398.

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

20. Jeong Y, Kim JN, Kim MW, Bucca G, Cho S, Yoon YJ, Kim BG, Roe JH, Kim SC, Smith CP, Cho BK. 2016. The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2). Nat Commun 7:11605.

21. Klucar L, Stano M, Hajduk M. 2010. phiSITE: database of gene regulation in bacteriophages. Nucleic Acids Res 38:D366—D370.

22. Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C, Kumar N, Liu Z, Maurel T, Moore B, McDowall MD, Maheswari U, Naamati G, Newman V, Ong CK, Paulini M, Pedro H, Perry E, Russell M, Sparrow H, Tapanari E, Taylor K, Vullo A, Williams G, Zadissia A, Olson A, Stein J, Wei S, Tello-Ruiz M, Ware D, Luciani A, Potter S, Finn RD, Urban M, Hammond-Kosack KE, Bolser DM, De Silva N, Howe KL, Langridge N, Maslen G, Staines DM, Yates A. 2018. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res 46:D802—D808.

23. The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158—D169.

24. Koonin EV, Tatusov RL. 1994. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity: application of an iterative approach to database search. J Mol Biol 244:125—132.

25. Janssen DB, Oppentocht JE, Poelarends GJ. 2001. Microbial dehalogenation. Curr Opin Biotechnol 12:254—258.

26. van Pée KH. 1996. Biosynthesis of halogenated metabolites by bacteria. Annu Rev Microbiol 50:375—399.

27. Rezanka T, Spizek J. 2003. Halogen-containing antibiotics from Streptomycetes, p 309—353. *In* Atta-ur-Rahman (ed), Studies in natural products chemistry, 1st ed, vol 29. Elsevier Science.

28. Dobson L, Reményi R, Tusnády GE. 2015. CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Res 43:W408—W412.

29. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto SC, Wu CH, Xenarios I, Yeh LS, Young SY, Mitchell AL. 2017. InterPro in 2017--beyond protein family and domain annotations. Nucleic Acids Res 45:D190—D199.

30. Temuujin U, Chi WJ, Lee SY, Chang YK, Hong SK. 2011. Overexpression and biochemical characterization of DagA from *Streptomyces coelicolor* A3(2): an endo-type β-agarase producing neoagarotetraose and neoagarohexaose. Appl Microbiol Biotechnol 92:749—759.

379 31. Bibb MJ, Jones GH, Joseph R, Buttner MJ, Ward JM. 1987. The agarase gene (*dagA*) of
380 *Streptomyces coelicolor* A3(2): affinity purification and characterization of the cloned
381 gene product. J Gen Microbiol 133:2089—2096.
382 32. Huang J, Shi J, Molle V, Sohlberg B, Weaver D, Bibb MJ, Karoonuthaisiri N, Lih CJ, Kao
383 CM, Buttner MJ, Cohen SN. 2005. Cross-regulation among disparate antibiotic
384 biosynthetic pathways of *Streptomyces coelicolor*. Mol Microbiol 58:1276—1287.
385 33. Kieser T, Bibb MJ, Buttner, MJ, Chater KF, Hopwood DA. 2000. Practical *Streptomyces*
386 genetics. John Innes Foundation.
387 34. Weaver D, Karoonuthaisiri N, Tsai HH, Huang CH, Ho ML, Gai S, Patel KG, Huang J, Cohen
388 SN, Hopwood DA, Chen CW, Kao CM. 2004. Genome plasticity in *Streptomyces*:
389 identification of 1 Mb TIRs in the *S. coelicolor* A3(2) chromosome. Mol Microbiol
390 51:1535—1550.

391

392 **FIG 1** Genomic Island 6 and the regulon of the ECF sigma factor encoded by SCO3450. GI 6 is
393 108 kb in length. The arrow indicates the location of SCO3450, which encodes a putative ECF
394 sigma factor. See Table 1 for a list of the strains used to identify GI 6. To obtain the data in the
395 figure, we used DNA microarrays that contained the 7825 predicted genes in the chromosome
396 of *S. coelicolor* M145 [17]. For the left-hand columns, the maximum intensity of green
397 corresponds to a 3-fold difference in signal on the microarrays for DNA in *S. coelicolor* M145
398 relative to DNA in the wild type strains. SCO3450 was overexpressed in an exponentially
399 growing liquid culture (SMM) of a *S. coelicolor* M600 derivative, through the use of a
400 thiostrepton-inducible promoter [16]. RNA was harvested from the culture immediately prior to

401 the addition of thiostrepton and every 15 to 30 minutes afterwards for 90 minutes. cDNA of the
402 initial RNA sample (OD_{450} ~ 0.5) was labeled with the green fluorescent dye Cy3. cDNA of RNA
403 isolated from subsequent time points was labeled with the red fluorescent dye Cy5. Yellow
404 microarray spots for a particular time point represented genes with transcript levels equal to
405 the levels of the initial time point. Red and green microarray spots represented genes induced
406 and repressed, respectively, after the addition of thiostrepton. The yellow color is shown here
407 as black for clarity. For the right-hand columns, the maximum intensity of red corresponds to 8-
408 fold induction of a given gene at a particular time relative to the beginning of the time course. A
409 control experiment with a M600 derivative that lacks SCO3450 in the induction plasmid
410 identified genes induced by thiostrepton. Labels on the right indicate ORF numbers and genes
411 with names.

412

413 **FIG 2** Alignment of putative promoters in the regulon of the ECF sigma factor encoded by
414 SCO3450. Red letters indicate the putative -35 and -10 promoter regions. On the right-hand
415 side, the numbers in parentheses indicate the distance between each -10 promoter region and
416 the start codon of the respective gene.

417

418 **Table 1. Strains that revealed the genomic islands of *S. coelicolor* M145.**

Strain	Comment	Ref.
M145	*S. coelicolor* sequenced “reference” strain, SCP1- SCP2-	[3]
1152	*S. coelicolor* Sermonti's SE1	[5]
1153	*S. coelicolor* Bradley's S199 strain	[6]
1326	*S. lividans* 66	[7]

2896	*S. lividans* ISP5434-, the strain from which pIJ101 was isolated	[8]
3034	*S. violaceroruber* SANK95570, harbors pSV1, which encodes methylenomycin biosynthetic genes	[9]

419

420 **Table 2. Types of enzymes encoded by the regulon of the ECF sigma factor.**

Type	Gene	AA	Putative Function	Other feature
CLEAVAGE				
Glycoside hydrolase	SCO3444	617	glycoside hydrolase family 15/phosphorylase b kinase regulatory chain family; six-hairpin glycosidase	
Nucleoside phosphorylase	SCO3463	262	nucleoside phosphorylase domain	
Dehalogenase	SCO3446	225	haloacid dehydrogenase (HAD)-like domain	
REDOX				
Glutaredoxin	SCO3442	114	glutaredoxin (DNA synthesis?)	
Oxidoreductase	SCO3443	454	pyridine nucleotide-disulphide oxidoreductase; dihydrolipoyl dehydrogenase?	
	SCO3460	505	pyridine nucleotide-disulphide oxidoreductase; dihydrolipoamide dehydrogenase?	
Dehydrogenase	SCO3478	344	D-isomer specific 2-hydroxyacid dehydrogenase	
TRANSFER				
Methyltransferase	SCO3452	359	S-adenosyl-L-methionine-dependent methyltransferase	
	SCO3459	287	S-adenosyl-L-methionine-dependent methyltransferase	
Phosphatidyltransferase?	SCO3457	205	CDP-alcohol phosphatidyltransferase; YnjF?	transmembrane
Nucleotide-diphospho-sugar transferase	SCO3464	210	nucleotide-diphospho-sugar transferases; transferase 1, rSAM/selenodomain-associated	
	SCO3465	236	nucleotide-diphospho-sugar transferase; transferase 2, rSAM/selenodomain-associated	
MISC				
Enzyme?	SCO3445	55	low to moderate homology to small regions of seven larger proteins, which include enzymes	membrane

421 AA = amino acids

Legends for Supplemental Material

FIG S1 Genomic Islands 1 to 10 of *S. coelicolor* M145. Only the genomic islands are shown and not the entire chromosome. Table 1 lists the strains used. To obtain these data, we used DNA microarrays that contained the 7825 predicted genes in the chromosome of *S. coelicolor* M145 [17]. Genomic DNA from *S. coelicolor* M145 was used as the reference sample and labeled with the green fluorescent dye Cy3. Genomic DNA samples from the five wild type strains each were labeled with the red fluorescent dye Cy5. The labeled DNA of each wild type strain was mixed with the labeled DNA of *S. coelicolor* M145 and hybridized to a microarray. Yellow spots on the microarrays represented genes at equal copy numbers between *S. coelicolor* M145 and the other strains. Green spots represented genes present in *S. coelicolor* M145 but absent in the other strains. Red spots represented genes at higher copy numbers in the strains relative to *S. coelicolor* M145. The yellow color is shown here as black for clarity. The maximum intensity of green corresponds to a 3-fold difference in signal on the microarrays for DNA in *S. coelicolor* M145 relative to DNA in the wild type strains. The green bars to the left of the GIs designate horizontally transferred genes in the genome sequence of *S. coelicolor* M145 that were predicted by Bentley et al. [3]. Regions of GIs not predicted by Bentley et al. lack a bar. Labels on the right indicate ORF numbers and genes with names.

FIG S2 Genomic Islands 11 to 18 of *S. coelicolor* M145. See the text of Fig. S1.

443 **FIG S3** Genomic Islands 19 to 22 of *S. coelicolor* M145. See the text of Fig. S1.

444

445 **FIG S4** The region of the *S. coelicolor* M145 chromosome that contains the regulon of the ECF

446 sigma factor. The figure was obtained from EnsemblBacteria [22]. SCO3478 is not shown.

447

448 **FIG S5** Result from CCTOP for the putative anti-sigma factor encoded by SCO3451. The putative

449 anti-sigma factor is predicted to contain one transmembrane domain. The data were obtained

450 from the Constrained Consensus TOPology (CCTOP) prediction server [28].

451

452 **FIG S6** Predicted zinc finger in the putative anti-sigma factor encoded by SCO3451. The data

453 were obtained from UniProt [23].

454

455 **FIG S7** Predicted zinc finger in the putative anti-sigma factor encoded by SCO3449. The data

456 were obtained from UniProt [23].

457

458 **Table S1. Annotations of the regulon of the ECF Sigma Factor encoded by SCO3450.**

459 Annotations were obtained from EnsemblBacteria [22] and UniProt [23]. For the UniProt

460 BLASTs, parentheses denote genera and percent identities of similar proteins. Red text denotes

461 putative enzymes. Blue text denotes proteins with homology to small regions of larger proteins.

462

463 **Table S2. Annotations of Genomic Island 6, excluding the regulon of the ECF sigma factor**

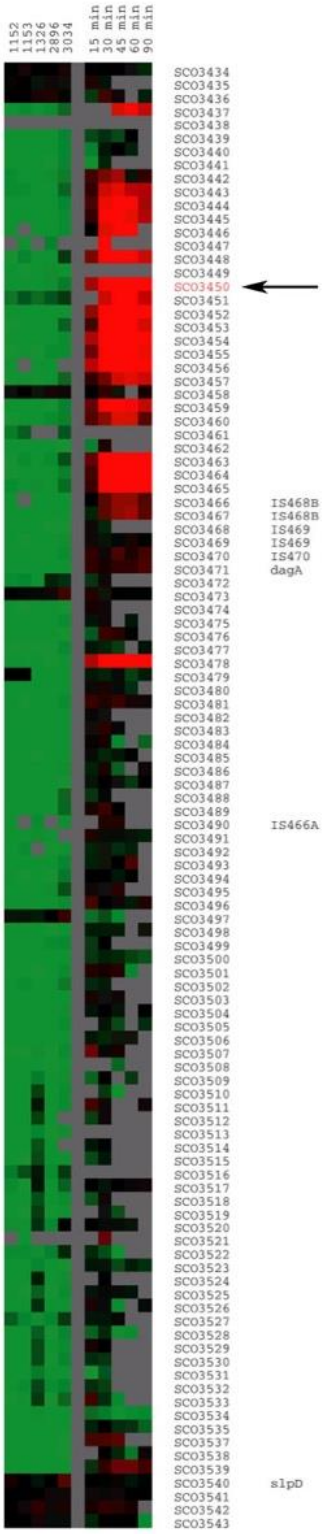
464 **encoded by SCO3450.** Annotations were obtained from EnsemblBacteria [22] and UniProt [23].

465 For the UniProt BLASTs, parentheses denote genera and percent identities of similar proteins.

466 Red text denotes putative enzymes. Blue text denotes proteins with homology to small regions

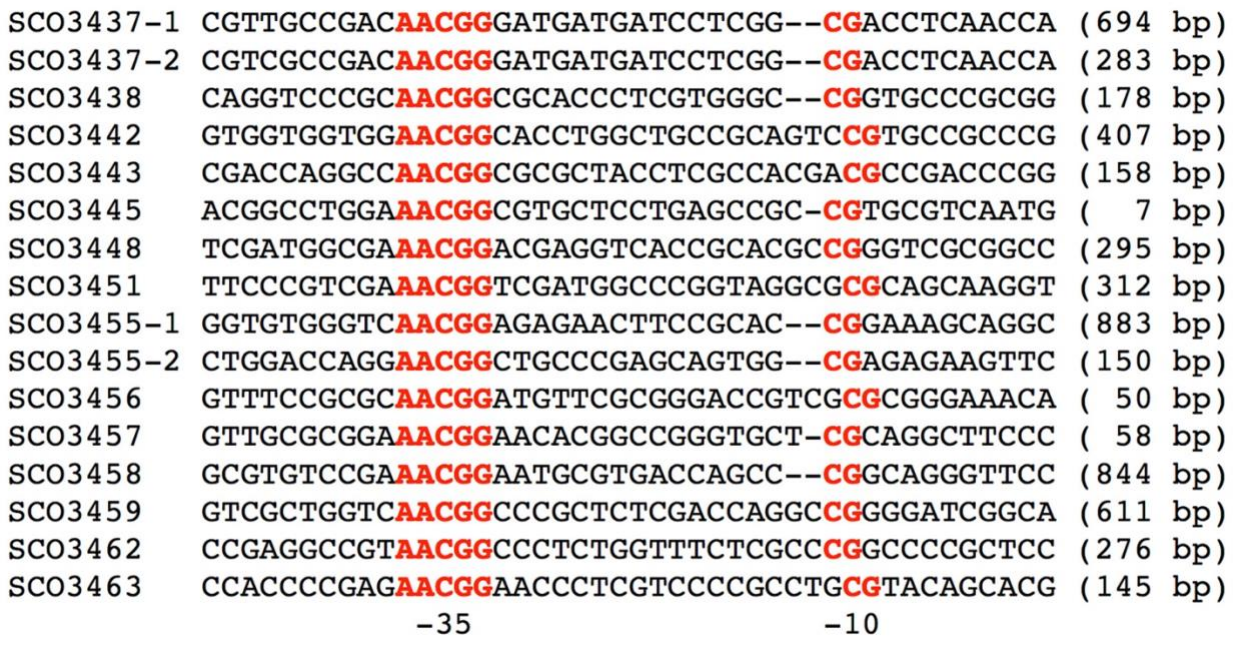
467 of larger proteins. See the text for additional details.

468



469

470 **FIG 1** Genomic Island 6 and the regulon of the ECF sigma factor encoded by SCO3450.



471

472 **FIG 2** Alignment of putative promoters in the regulon of the ECF sigma factor encoded by

473 SCO3450.