npg

NOTE

Overexpressing transcriptional regulator in *Chaetomium globosum* activates a silent biosynthetic pathway: evaluation of shanorellin biosynthesis

Yuta Tsunematsu[1], Suguru Ichinoseki[1], Takehito Nakazawa[1], Noriyasu Ishikawa[1], Hiroshi Noguchi[1], Kinya Hotta[2] and Kenji Watanabe[1]

In recent years, many secondary metabolite gene clusters, including polyketide biosynthetic genes, have been uncovered through fungal genome sequencing. Although approximately 33 polyketide biosynthetic genes can be identified in the genome of a fungus *Chaetomium globosum*, only 11 fungal polyketide scaffolds[1] can be isolated from cultures grown under typical growth conditions.[2] Transcription factors (TFs) located within a gene cluster often determine whether its respective secondary metabolites can be biosynthesized, and it is this class of regulators that is responsible for the low expression level of biosynthetic genes and less-than-desirable yield of the corresponding natural products.[3,4] Here, we explored the feasibility of activating a silent gene cluster through overexpression of a TF encoded in a putative polyketide biosynthetic gene cluster in *C. globosum* for which a stably propagating expression vector is unavailable.

For the current study, we focused on a putative polyketide biosynthetic gene cluster from *C. globosum*. DNA sequence analysis of this 25-kilobase (kb)-long cluster (Figure 1a) revealed the presence of seven genes, *cgsA–G* (Table 1), that appeared to be responsible for a polyketide biosynthesis. In order to activate this silent gene cluster involving a large polyketide synthase (PKS) gene (>7.1 kb), a recombinant strain capable of overexpressing the transcriptional factor *cgsG* was constructed. Instead of activating the expression of the endogenous copy of *cgsG* within the *cgs* cluster, we chose to introduce a separate copy of *cgsG* under the control of a constitutively active promoter to achieve overexpression of the TF in *C. globosum* to avoid potential repression of the transcription of *cgsG* located within the cluster by some other regulation mechanism. For the overexpression of this regulator, which is highly homologous to a known GAL4-like DNA-binding protein, we chose to use pBluescript SK as the base for the delivery vector for transforming *C. globosum* with a second copy of constitutively expressing *cgsG*. For the construction of the delivery vector, PCR was performed with the *C. globosum* genomic

DNA as a template to amplify the *β*-tubulin (*tub*) promoter and terminator. Similarly, a *hph* gene encoding for a hygromycin phosphotransferase that confers hygromycin resistance in *C. globosum* was amplified from pPHT1 plasmid[5] as a selectable marker. These three fragments, *tub* promoter, *hph* and *tub* terminator, were assembled into a single fragment by overlap extension PCR. The resulting *hph* expression cassette was inserted into pBluescript using Geneart Seamless Cloning and Assembly Kit (Life Technologies, Grand Island, NY, USA) to yield the delivery vector, pKW3202. Also, the *actA* promoter, which is responsible for the transcription of the actin gene, was isolated from the *C. globosum* chromosomal DNA and inserted into pKW3202 to prepare pKW13000. Subsequently, the 1.4-kb DNA fragment containing the coding and terminator region of *cgsG* was inserted immediately downstream of the *actA* promoter in pKW13000 to yield pKW12203, a hygromycin resistance-conferring vector capable of overexpressing *cgsG* (Figure 1b). Transformation was carried out as described elsewhere[6] with the following modification: the recipient *C. globosum* strain was initially cultivated in MYG medium containing 0.8 M sorbitol and 200 μg ml^{-1} hygromycin. The plasmid pKW12203 was linearized by *Xba*I and transformed into *C. globosum*. Approximately 50 transformants were grown on MYG agar plates supplemented with 200 μg ml^{-1} hygromycin. The selection confirmed the integration of pKW12203 carrying *cgsG* into the genome of the transformant. This transformed *C. globosum* strain was termed CGKW1.

Next, we examined whether transcription of *cgsA* encoding for a PKS can be activated when the transcriptional regulator CgsG is overexpressed. Total RNA (1 μg) isolated from CGKW1 was subjected to reverse transcription-PCR for *cgsA* cDNA synthesis with a random hexamer primer (Applied Biosystems, Life Technologies). The cDNA formation was detected by agarose gel electrophoresis (Figure 1b, *cgsA*) following PCR amplification of cDNA using the following primer set: CHGG8793coreFw, 5′-TCCATGGATCCCCAGCAGAGG-3′ and

[1]Department of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan and [2]Department of Biological Sciences, National University of Singapore, Singapore
Correspondence: Dr K Watanabe, Department of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga-ku, Shizuoka 422-8526, Japan.
E-mail: kenji55@u-shizuoka-ken.ac.jp
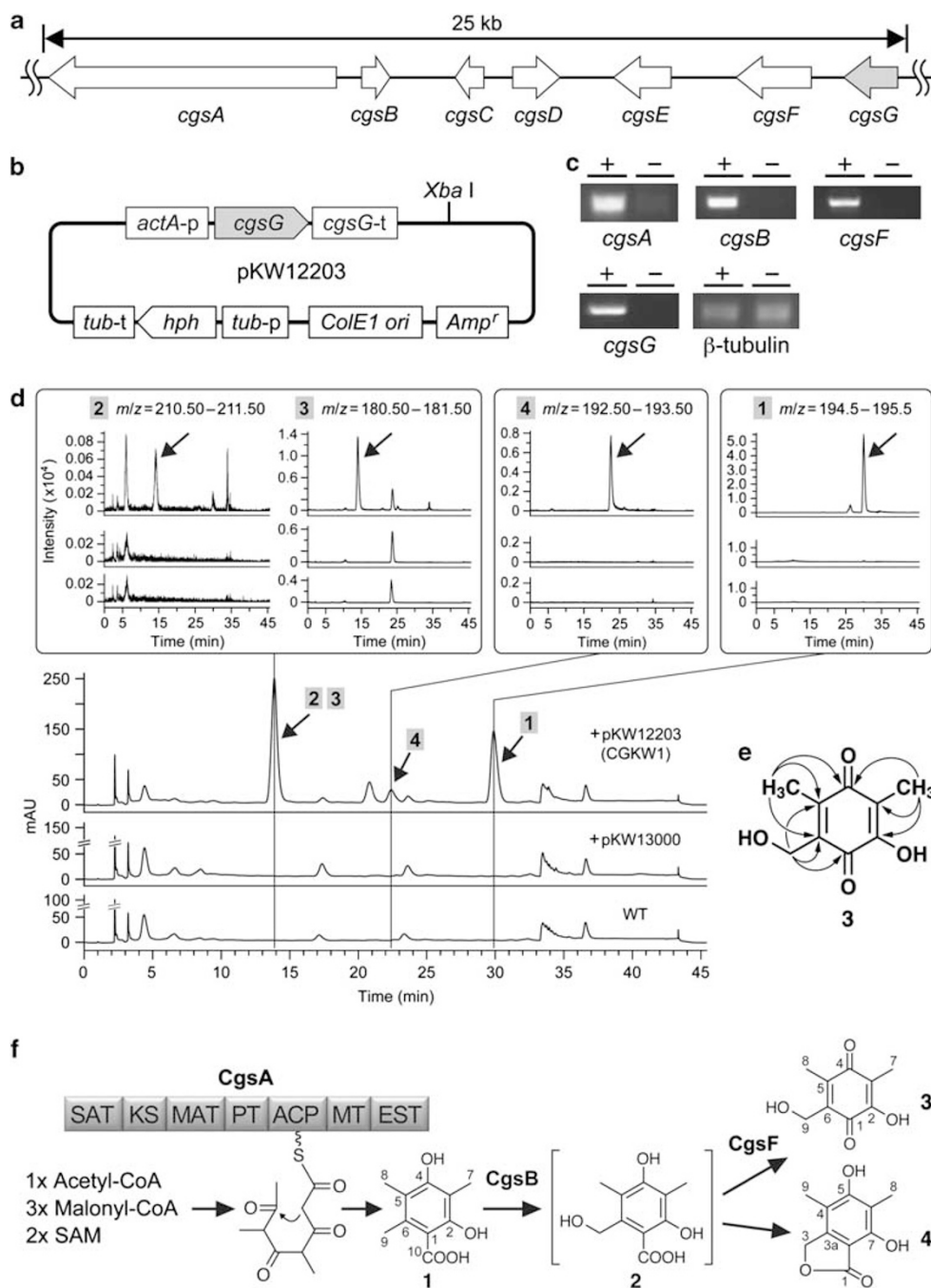
Figure 1 Characterization of the silent *cgs* gene cluster, activation of transcription from the cluster, and the products biosynthesized by the associated enzymes. (**a**) Organization of the *cgs* gene cluster involved in the biosynthesis of **3**. (**b**) Map of the plasmid pKW12203 carrying the hygromycin phosphotransferase (*hph*) gene under the control of a β-tubulin promoter (*tub-p*) and a β-tubulin terminator (*tub-t*), along with the transcription regulator *cgsG* gene under the control of an actin promoter (*actA-p*) and the original *cgsG* terminator (*cgsG-t*). (**c**) Reverse transcription-PCR analysis of the effect of overexpressing the transcriptional regulator *cgsG* on the expression of three other *cgs* genes, *cgsA*, *cgsB* and *cgsF* in *C. globosum*. Expression of the β-tubulin gene was used as a positive control. −: *C. globosum* transformed with an empty plasmid pKW13000 carrying no *cgs* gene; +: CGKW1, a *C. globosum* strain overexpressing the transcriptional regulator gene *cgsG* introduced by pKW12203. (**d**) Sets of three HPLC traces from the LC–MS analysis of the metabolic extracts from the wild-type and recombinant strains of *C. globosum* for determining the effect of overexpressing the transcriptional regulator CgsG on activating the silent *cgs* biosynthetic gene cluster. In each set of three chromatograms, the top one is from *C. globosum* strain CGKW1, the middle one is from *C. globosum* carrying pKW13000, and the bottom one is from the wild-type *C. globosum*. All traces were monitored at 254 nm. Boxed chromatograms are the extracted ion chromatograms for the range of *m/z* covering the compounds **1**–**4**. (**e**) Analysis of the HMBC data for **3**. (**f**) Proposed biosynthetic pathway of **3** from a polyketide product **1** by the activated *cgs* gene cluster. Predicted domain organization of the iterative PKS CgsA is also shown. The domains, depicted as shaded boxes, are: SAT, starter unit; ACP transacylase; KS, ketosynthase; MAT, malonyl-CoA acyltransferase; PT, product template; ACP, acyl carrier protein; MT, methyltransferase; EST, lipase-like esterase.

Table 1 Deduced functions of the open reading frames ORFs in the *cgs* biosynthetic gene cluster from *C. globosum*

Gene	Amino acids (no)	Proposed protein function (most homologous protein[a], NCBI accession number, species)	Identity/similarity (%)
cgsA (CHGG_08793)	2379	Polyketide synthase (polyketide synthase putative, XP_002339967, *Talaromyces stipitatus*)	39/56
cgsB (CHGG_08794)	238	Cytochrome P450 (unnamed protein product, BAE65443, *Aspergillus oryzae*)	68/85
cgsC (CHGG_08795)	262	SGNH_hydrolase (carbohydrate esterase family-3 protein, CCD55162, *Botryotinia fuckeliana*)	40/60
cgsD (CHGG_08796)	378	Reverse transcriptase (hypothetical protein, XP_001226257, *Chaetomium globosum*)	76/79
cgsE (CHGG_08797)	405	Hypothetical protein CHGG_08797 (conserved hypothetical protein, XP_002483593, *Talaromyces stipitatus*)	67/79
cgsF (CHGG_08798)	371	Salicylate hydroxylase (salicylate hydroxylase putative, XP_002483592, *Talaromyces stipitatus*)	43/57
cgsG (CHGG_08799)	472	GAL4 (C6 finger domain protein GliZ, XP_002151409, *Penicillium marneffei*)	29/44

[a]Deduced functions of the ORFs identified within the *cgs* biosynthetic gene cluster were based on the percentage sequence similarity/identity to known proteins as determined by Protein BLAST search against the NCBI non-redundant database.

CHGG8793coreRv, 5′-ACCCGTCCCATGAGCCTCCAC-3′. As a negative control, the absence of transcription of *cgsA* was confirmed in *C. globosum* transformed with pKW12203 lacking the *cgsG* insertion (Figure 1c, WT). Transcription of the β-tubulin gene was used as a positive control to confirm that general gene transcription remained similar in both the wild-type and the CGKW1 strain (Figure 1c, β-tubulin). Similar investigations revealed successful activation of *cgsB* and *cgsF* genes upon overexpression of an extraneous copy of *cgsG* (Figure 1c, *cgsB* and *cgsF*).

To isolate and characterize the polyketide compound(s) biosynthesized by the *cgs* cluster, CGKW1 was cultivated on an oatmeal agar for 7 days at 30 °C. For the control samples, the wild-type *C. globosum* and *C. globosum* carrying a blank plasmid pKW13000 were also cultured under the same condition. Reversed-phase LC–MS analysis of the crude organic extract of the culture broth revealed the presence of several nonpolar metabolites that were present only in the CGKW1 culture broth (Figure 1d). For a full characterization of those compounds, we cultivated the strain on MYG liquid medium on a large scale and purified the compounds using flash column chromatography and HPLC when necessary. NMR data, including ^{1}H and ^{13}C spectra, as well as HMBC analysis (Figure 1e), allowed determination of the chemical structure of the compound eluted at around 13.9 min to be shanorellin (**3**).[7] Another compound, 2,4-dihydroxy-3,5,6-trimethylbenzoic acid (**1**),[8] a presumed intermediate compound for the biosynthesis of **3** (Figure 1f), was also isolated from CGKW1. Furthermore, a phthalide, 5,7-dihydroxy-4,6-dimethylisobenzofuran-1(3*H*)-one (**4**),[9] was isolated from the culture of CGKW1. The results of the spectrometric analyses of these compounds are given below.

Compound **1**: Yellowish white powder; ^{1}H NMR (acetone-d_6, 500 MHz): $\delta = 2.50$ (3 H, s), 2.15 (3 H, s), 2.11 (3 H, s); ^{13}C NMR (acetone-d_6, 125 MHz): $\delta = 174.9$ (C-10), 161.6 (C-2), 159.0 (C-4), 138.6 (C-6), 116.4 (C-5), 108.8 (C-3), 105.7 (C-1), 18.9 (C-9), 12.3 (C-8), 8.7 (C-7); ESI-MS: *m/z* 195 [M–H]$^{-}$; HRESIMS: *m/z* 195.0654 ([M–H]$^{-}$, calcd. for $C_{10}H_{11}O_4$, 195.0663).

Compound **3**: Orange crystal; ^{1}H NMR (CDCl$_3$, 500 MHz): $\delta = 4.58$ (2 H, s, H-9), 2.14 (3 H, s, H-8), 1.94 (3 H, s, H-7); ^{13}C NMR (CDCl$_3$, 125 MHz): $\delta = 187.8$ (C-4), 184.1 (C-1), 150.9 (C-2), 145.4 (C-5), 136.8 (C-6), 118.0 (C-3), 56.7 (C-9), 12.5 (C-8),

8.4 (C-7); ESI-MS: *m/z* 181 [M–H]$^{-}$; HRESIMS: *m/z* 181.0503 ([M–H]$^{-}$, calcd. for $C_9H_9O_4$, 181.0506).

Compound **4**: Yellowish white powder; ^{1}H NMR (acetone-d_6, 500 MHz): $\delta = 8.42$ (1 H, brs, OH-5), 8.02 (1 H, brs, OH-7), 5.26 (2 H, s, H-3), 2.14 (3 H, s, H-8), 2.12 (3 H, s, H-9); ^{13}C NMR (acetone-d_6, 125 MHz): $\delta = 173.2$ (C-1), 161.3 (C-5), 154.1 (C-7), 145.3 (C-3a), 111.4 (C-6), 110.6 (C-4), 103.3 (C-7a), 70.3 (C-3), 11.3 (C-9), 8.3 (C-8); ESI-MS: *m/z* 195 [M + H]$^{+}$; HRESIMS: *m/z* 195.0652 ([M + H]$^{+}$, calcd. for $C_{10}H_{11}O_4$, 195.0652).

Our result revealed that the *cgs* gene cluster is responsible for the biosynthesis of compound **3**. This cluster contains *cgsA* that encodes a 255-kDa iterative PKS with seven domains as ascertained by *in silico* sequence analysis: starter unit ACP transacylase, ketosynthase, malonyl-CoA acyltransferase, product template, acyl carrier protein, methyltransferase and lipase-like esterase (Figure 1f). The analysis also identifies that CgsA can be categorized as a non-reducing PKS owing to its lack of domains capable of catalyzing a reduction and/or dehydration reaction.[10] Considering the relatively high sequence similarity between CgsA and the 5-methylorsellinic acid-synthesizing MpaC (37.6% identity, 52.7% similarity), it is highly plausible that CgsA synthesizes the benzoic acid core structure of **1**[8] from one molecule of acetyl-CoA and three molecules of malonyl-CoA. Then the methyltransferase domain in CgsA can transfer two methyl groups[7] from *S*-adenosyl-L-methionine to C7 and C8 to generate **1**.

BLAST[11] and FFAS03[12] sequence analyses of other enzymes encoded in the *cgs* cluster suggest that the polyketide backbone is modified by two more enzymes to afford **3** as the final product (Figure 1f). The first step is likely catalyzed by CgsB, which is predicted to be a cytochrome P450. CgsB could perform hydroxylation of the C9 methyl substituent in the PKS product **1** to afford the presumed intermediate **2**. Although **2** was not isolated from the culture directly, the LC–MS analysis identified that **2** co-elutes with the final product 3 as a minor constituent (Figure 1d). Also, a phthalide **4**, which can be formed by spontaneous lactonization of **2**,[13] was identified by the LC–MS analysis, providing further support for the involvement of **2** as an intermediate of the pathway. Then, the final step for the formation of **3** can be carried out by CgsF, which is highly homologous to salicylate hydroxylase. Salicylate hydroxylase belongs to a flavin-dependent mono-oxygenase family

and is responsible for the decarboxylative hydroxylation of salicylate to form catechol with a reduction of an enzyme-bound FAD by NADH under aerobic conditions.[14,15] Analogously, CgsF can perform a decarboxylative hydroxylation to convert a dihydroxybenzoic acid **2** to a hydroxyhydroquinone product, which can be readily oxidized into the final hydroxyquinone product, shanorellin **3**.

In this study, we successfully activated a silent polyketide biosynthetic gene cluster involving an iterative PKS present in *C. globosum* by constitutively overexpressing an extraneous copy of a TF associated with the cluster. Activation of the silent gene cluster was confirmed by reverse transcription-PCR. A strict correlation between the expression of the otherwise silent *cgs* genes and the production of several polyketide products strongly indicates that the identified polyketide products are indeed the products of the Cgs biosynthetic enzymes. This was the first successful activation of polyketide biosynthesis from a silent gene cluster by overexpressing its associated TF in *C. globosum*. Our results showed that CgsA is an iterative PKS capable of synthesizing 2,4-dihydroxy-3,5,6-trimethylbenzoic acid (**1**). Furthermore, we were able to identify two tailoring enzymes, a cytochrome P450 CgsB and a salicylate hydroxylase homolog CgsF in the gene cluster. Through this work, we were able to establish experimentally that the silent *cgs* cluster is responsible for shanorellin (**3**) biosynthesis. In addition, we were able to propose the shanorellin biosynthetic pathway involving CgsB hydroxylating the C6 methyl group of the initial polyketide product **1** to generate **2**, followed by CgsF converting **2** into shanorellin (**3**) via a decarboxylative hydroxylation. Our result clearly demonstrates that our approach of plasmid-based overexpression of TF is a practical approach toward mining for new natural products that can be biosynthesized by fungi. Furthermore, our methodology can help expand our knowledge of novel enzymes and their associated metabolic pathways that would facilitate our efforts toward engineering biosynthetic pathways for production of analogs with useful bioactivity.

ACKNOWLEDGEMENTS

1   Scherlach, K., Boettger, D., Remme, N. & Hertweck, C. The chemistry and biology of cytochalasans. *Nat. Prod. Rep.* **27,** 869–886 (2010).
2   Brakhage, A. A. & Schroeckh, V. Fungal secondary metabolites - strategies to activate silent gene clusters. *Fungal. Genet. Biol.* **48,** 15–22 (2010).
3   Flaherty, J. E. & Payne, G. A. Overexpression of *aflR* leads to upregulation of pathway gene transcription and Increased aflatoxin production in *Aspergillus flavus. Appl. Environ. Microbiol.* **63,** 3995–4000 (1997).
4   Bergmann, S. *et al.* Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans. Nat. Chem. Biol.* **3,** 213–217 (2007).
5   Cummings, W. J., Celerin, M., Crodian, J., Brunick, L. K. & Zolan, M. E. Insertional mutagenesis in *Coprinus cinereus*: use of a dominant selectable marker to generate tagged, sporulation-defective mutants. *Curr. Genet.* **36,** 371–382 (1999).
6   Shibuya, I. *et al.* Molecular cloning of the glucoamylase gene of *Aspergillus shirousami* and its expression in *Aspergillus oryzae. Agric. Biol. Chem.* **54,** 1905–1914 (1990).
7   Wat, C., McInnes, A. G., Smith, D. G. & Vining, L. C. Use of ^{13}C in biosynthetic studies. Location of isotope from labeled acetate and formate in shanorellin by ^{13}C nuclear magnetic resonance spectrometry. *Can. J. Biochem.* **50,** 620–624 (1972).
8   Hirota, A., Morimitsu, Y. & Hojo, H. New antioxidative indophenol-reducing phenol compounds isolated from the *Mortierella* sp. Fungus. *Biosci. Biotech. Biochem.* **61,** 647–650 (1997).
9   Achenbach, H., Mühlenfeld, A. & Brillinger, G. U. Phthalide und chromanole aus *Aspergillus duricaulis. Liebigs Ann. Chem.* **808,** 1596–1628 (1985).
10 Chiang, Y. M., Oakley, B. R., Keller, N. & Wang, C. C. Unraveling polyketide synthesis in members of the genus *Aspergillus. Appl. Microbiol. Biotechnol.* **86,** 1719–1736 (2010).
11 Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T. L. NCBI BLAST: A better web interface. *Nucleic Acids Res.* **36,** W5–W9 (2008) (http://blast.ncbi.nlm.nih.gov/Blast.cgi).
12 Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A. FFAS03: a server for profile–profile sequence alignments. *Nucleic Acids Res.* **33,** W284–W288 (2005) (http://ffas.burnham.org/ffas-cgi/cgi/document.pl).
13 Ichihara, A., Adachi, K., Hosokawa, K. & Takeda, Y. Enzymatic hydroxylation of aromatic carboxylic acids. *J. Biol. Chem.* **237,** 2296–2302 (1962).
14 Katagiri, M *et al.* Salicylate hydroxylase, a monooxygenase requiring flavin adenine dinucleotide. II. The mechanism of salicylate hydroxylation to catechol. *J. Biol. Chem.* **240,** 3414–3417 (1965).
15 Suzuki, K., Gomi, T. & Itagaki, E. Intermediate and mechanism of hydroxylation of *o*-iodophenol by salicylate hydroxylase. *J. Biochem.* **109,** 791–797 (1991).