

Supporting Information for

Elucidation of shanorellin biosynthetic pathway and functional analysis of associated enzymes

Michio Sato,^a Haruka Yamada,^a Kinya Hotta^b and Kenji Watanabe*^a

^a Department of Pharmaceutical Sciences, University of Shizuoka, Shizuoka 422-8526, Japan

^b School of Biosciences, The University of Nottingham Malaysia Campus, Semenyih, Selangor Darul Ehsan 43500, Malaysia

*Correspondence e-mail: kenji55@u-shizuoka-ken.ac.jp

Table of Contents

Figure S1. Picture of the SDS–PAGE gel of the partially purified CgsF.	S2
Table S1. The <i>cgs</i> gene cluster organization and predicted protein functions.	S3
Table S2. Oligonucleotide primer sequences.	S4

Figure S1. SDS-PAGE analysis of the *Escherichia coli*-produced CgsF enriched by metal-affinity chromatography using Ni-NTA Sepharose resin. Lane M: molecular weight marker; lane 1: CgsF (48 kDa).

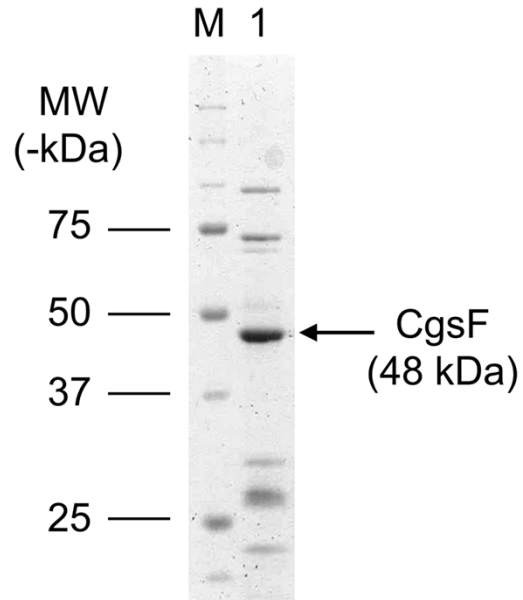
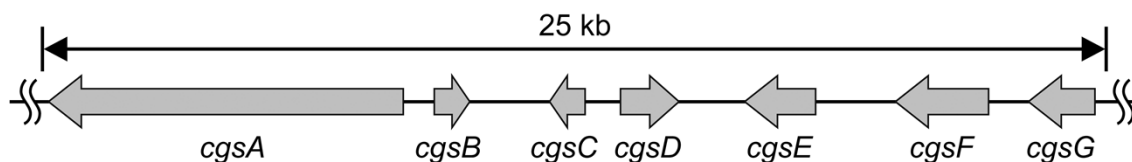


Table S1. Organization of the *cgs* gene cluster responsible for the biosynthesis of shanorellin 4, and predicted functions of corresponding translation products.



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

Table S2. Oligonucleotide primer sequences. DNA primers were designed on the basis of sequence data obtained from the *Chaetomium globosum* and *Streptomyces hygroscopicus* (for *hph*) genome sequence database.

Primer name	Sequence (5' to 3')
19140-Fw1	GAAATTGCACCAAGATGAGCCACAACATCGCGG
19133-Rv4	CAAAGGGCGTTTCCGTTGCCCG
19133-Rv3	CTAAAGGGAACAAAAGCTGGTCAAAGGGCGTTTCCGTTGCCCG
19133-Fw4	GATTTCCGTAGCCGACCTCGATTATGCC
19151-Fw2	CAGAAAGGCCACCAAGATGAGCCACAACATCGCGG
19151-Fw1	GATCGCAGGTCAGGACTTCAGGCTGGTATAAGATCCAAGGC
19151-Rv1	CATCTTGGTGGGCCTTTCTGTAATCCTCAAAGGAGCCG
19114-Rv3	CTCCTGACCTGCGATCTTCCCT
19133-Fw1	CTATAGGGCGAATTGGGTACATGATTTCCGTAGCCGACCTCGATTATG
19121-Fw2	GAAGATCGCAGGTCAGGACTTCAGGCTGGTATAAGATCCAAGGCTATCG
19121-Rv2	CACGGGTGGTGCCATTTGACGGCTGGAAAGGTGC
19121-Fw3	CAAATGGCACCACCCGTGAAACTCCGACCCGCAAGC
10121-Rv3	CAAGAGAGTAGGCGCAATTTTCATAGCCACTCAGTCTGCACTG
19121-Fw4	GAAATTGGCGCCTACTCTCTTGTAAAGCGTAAGTCAAGATCTCTGC
19121-Rv4	CTAAAGGGAACAAAAGCTGGGGAAATACAGGCCGGACTACAGTTGATATAC
19153-Fw1	GGGCGAATTGGGTACGTTCAAGGGTTCTGCATTGGTTTC
19153-Rv1	CAACAGGACTGTAGCGAGACCTGGAAGAAGCGGTGGTTCG
19153-Fw2	CTCCACCGTTCCATTGAGAGGAGTTTTGTGTTTGCGGGG
19153-Rv2	GAACAAAAGCTGGAGCTGGTGGAGTCTGCCTCGATGAAAGCCG
19153-Fw3	CGAACCGCCAAACAACGTAGTCCTC
19153-Rv3	CTCGTATCTGGAAGTCCCATGAGC
19154-Fw4	GGGCGAATTGGGTACCGTGAGGTGAGATCCAGTAAGGCG
19154-Rv4	CAACAGGACTGTAGCGGGAAGTGGCGGAGACGGTAGTATAC
19154-Fw5	CTCCACCGTTCCATTCCGGTCCCTTCGTGCCCCGTTTCAC
19154-Rv5	GAACAAAAGCTGGAGCTGGAGAAATCGCGAAAGGAAGTCGCCTC
19154-Fw3	CCAGAGCTTCGGTGAGGCTGCTTAG
19154-Rv3	CAAACACGCTTAGCACGGCACTTGTC
19155-Fw1	GGGCGAATTGGGTACGAACGGGGGAAAGGGGTGAGCCTTG
19155-Rv1	CAACAGGACTGTAGCGGGTTGCCGTGGGTGGAGAGCAGAC
19155-Fw2	CTCCACCGTTCCATTCCGTCAACGCCCTCATCGAGGAATAC
19155-Rv2	GAACAAAAGCTGGAGCTAGTGGCCGTCTCTGGAGGGAACGAC
19155-Fw3	GTTATATCCAAGTGAGGGTGCTCGTC
19155-Rv3	CTATCAAGAGACGATCTTGAGCGG
19156-Fw1	GGGCGAATTGGGTACCGCCCCCAAGTTCCAACGCAAAAC
19156-Rv1	CAACAGGACTGTAGCTATGGATCAGCCCATGCACATAGC
19156-Fw2	CTCCACCGTTCCATTATGTTTTAAGAACTTCTGCGGGGAG
19156-Rv2	GAACAAAAGCTGGAGCTCGAACTTTCAAGCGTTGATTATGGAGC

19156-Fw3	GCTACAGTTCAAGGAGGAGAACTC
19156-Rv3	CAAGCCAGCCGAGTAGAGTTACATG
19157-Fw1	GGGCGAATTGGGTACGCCATTATCGCCGGTTAATACAATG
19157-Rv1	CAACAGGACTGTAGCTCCACCCCCCTCCCCGTCCGAG
19157-Fw2	CTCCACCGTTCCATTATTAGTGAGTATTAAGCAAGGCTGAC
19157-Rv2	GAACAAAAGCTGGAGCTGTTGTCATTTCCCCAGCGACC
19157-Fw3	GTGGTCCGAGTCACTAAAGGTAG
19157-Rv3	CCCAACTTCTCTTTGCTCCGAGAATC
19140-Rv1	CAAGTCAGCAGCGGCAAGGATTTTCGGCGGGGTTGTATCTTGC
pyrG Fw	TGCCGCTGCTGACTTGAGGTG
pyrG Rv	CTCCTGACCTGCGATCTTCCCT
19121-Fw1	CTATAGGGCGAATTGGGTACGCTGGTTGTCAGTCTAGATCCGGCTGTTTC
19121-Rv1	CCTCAAGTCAGCAGCGGCAATCGAAGCGTGTATATAGACCGTCTTGC
hph_Fw	GCTACAGTCCTGTTGCGCCTTCCGG
hph_Rv	AATGGAACGGTGGAGAAAGTCCCTGGC
Cg-KO-hph-Rv	GCACTCGTCCGAGGGCAAAGGAATAG
Cg-KO-hph-Fw	CGACAGACGTCGCGGTGAGTTCAG
19140-Fw1	GAAATTGCACCAAGATGAGCCACAACATCGCGG
