MedChemComm

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/medchemcomm

MedChemComm

RSCPublishing

ARTICLE

Elucidation of shanorellin biosynthetic pathway and functional analysis of associated enzymes

Cite this: DOI: 10.1039/x0xx00000x

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

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Since fungal natural products biosynthesized by polyketide synthases frequently exhibit useful biological activities, identifying and understanding the mechanism of biosynthetic steps taken by PKSs is of great interest. One such compound isolated from *Chaetomium globosum* is shanorellin, whose biosynthetic gene cluster was activated by overexpressing the transcription factor, CgsG. Through targeted gene knockout in *C. globosum* and *in vitro* biochemical analyses, we determined for the first time the gene cluster and the pathway for the biosynthesis of shanorellin. We also identified that a cytochrome P450 CgsB is responsible for catalysing hydroxylation of a methyl group at C5 in the aromatic product biosynthesized by CgsA. Subsequently, a flavin-containing monooxygenase CgsF affects decarboxylation of the intermediate to form the quinone product, shanorellin. While this class of natural products exhibit notable biological activities, silencing of relevant gene clusters in the producing organisms results in trace-level production of those compounds under conventional culture conditions. Overexpression of the biosynthetic gene cluster's transcription regulator was achieved by incorporating a regulatable promoter into the native fungal genome. This approach allowed us to identify the shanorellin gene cluster and decipher its biosynthetic pathway effectively.

Introduction

In recent years, "mining" for secondary metabolite gene clusters, including polyketide biosynthetic genes, has been practiced through sequence analysis of fungal genomes.[1] Our attempt at abolishing transcriptional regulators associated with epigenetic silencing of secondary metabolite biosynthetic pathways in *Chaetomium globosum* resulted in the identification of the products generated by different gene clusters and isolation of novel secondary metabolites. We were able to identify six gene clusters that biosynthesize 11 natural products known to be produced by *C. globosum*, including one cytochalasan and six azaphilone-type compounds.[2-4] Transcription factor found within or near a gene cluster can govern the biosynthesis of its respective secondary metabolites, and this class of regulators is often to be blamed for the low gene expression and less-than-desirable level of production of the corresponding natural product.[5,6] Through overexpression of a transcription factor encoded in a putative polyketide biosynthetic gene cluster in *C. globosum*, we were able to activate the silent biosynthetic gene cluster and obtain a natural product shanorellin **4** and its precursor **1**, from the fungus.[7] In this study, we extended the study to confirm the identity of the shanorellin gene cluster through knock-out experiments, and validate our proposed shanorellin biosynthetic pathway with *in vivo* and *in vitro* experiments.

Results and discussion

Identification of the polyketide synthase responsible for biosynthesizing shanorellin (4)

The core structure of **4** suggested that it was formed by a polyketide synthase (PKS) that could assemble a triketide chain (Fig.1a). BLASTP[8] search of the *C. globosum* genome sequence identified *cgsA* which exhibited an amino acid sequence identity and similarity of 39% and 56%, respectively, to a PKS gene (accession number XP_002339967; see also Table S1 in the Supplementary Information).[7] The predicted domain organization (Fig. 1a) indicated CgsA to be an iterative non-reducing PKS known to produce aromatic cyclic compounds.[9] Thus, we speculated that *cgsA* might code for the PKS responsible for the formation of **4**. To verify the function of the gene, *cgsA*, we performed targeted gene deletion by double homologous gene replacement using an *hph* (hygromycin B phosphotransferase gene) cassette on the *C. globosum* Δ *CgpyrG*/ Δ *CgligD* strain, CGKW14.[2] The CGKW14 strain was generated in a previous study by knocking out *CgligD* in order to suppress random integration of DNA fragments into the genome of *C. globosum*.[2] Thus, like *Aspergillus nidulans*,[10-13] CGKW14 can serve as a convenient strain for analysis of natural product biosynthesis in *C. globosum*. Transformants having the *hph* gene integrated into *cgsA* were successfully obtained after confirming the deletion of the target gene by diagnostic PCR (Fig. 2). Seven-day culture of Δ *cgsA*/CGKW14 grown in MYG liquid medium was extracted with EtOAc, and the dried extract was subjected

Medicinal Chemistry Communications Accepted Manuscript

to metabolite analysis by liquid chromatography mass spectrometry (LC–MS). As expected, deletion of *cgsA* abolished the production of **4** as well as **1** in *C. globosum* (Fig. 3 i vs. vi).

Targeted Deletion of *cgsB*, *cgsC*, *cgsE* and *cgsF*, and Identification of Intermediates 1, 2 and 3

Bioinformatics analysis indicated that *cgsB*, *cgsC*, *cgsE* and *cgsF* likely code for a cytochrome P450 oxygenase (P450), a hydrolase, a hypothetical protein and a salicylate hydroxylase, respectively (Table S1).[7] To verify the actual function of those annotated genes in this cluster, we performed targeted deletion of these genes in CGKW14 overexpressing *cgsG*, to promote active expression of the remaining genes in the *cgs* gene cluster.[2] We initiated the functional analysis of these genes by looking for the biosynthesis of the starting precursor **1** and the absence of the final product **4**. The mutant *Δcgs*B/CGKW14 (*cgsG*) was successfully obtained, and the deletion of *cgsB* was confirmed by diagnostic PCR (Fig. 2). EtOAc extract of seven-day culture of *Δ cgsB*/CGKW14 (*cgsG*) was subjected to metabolite analysis by LC–MS. The result indicated that *ΔcgsB*/CGKW14 (*cgsG*) did not produce any of the compounds in the UV trace (Fig. 3 ii vs. vi). However, production of **1** by *ΔcgsB*/CGKW14 (*cgsG*) was confirmed in the extracted MS trace showing a clear peak for **1** with the m/z^+ of 195 (Fig. 4). As expected from our proposed biosynthetic pathway (Fig. 1a), deletion of *Δ cgsC* (Fig. 2) did not affect the production of **3** and **4** (Fig. 3B iii vs. iv), preventing us from assigning a specific function to CgsC, which also does not have a sequence-based annotated function. On the other hand, deletion of *cgsE*, a gene encoding a conserved hypothetical protein, (Fig. 2) led to the loss of production of **3** (Fig. 3 iv vs. vi). The chemical structure of **3** and **4** were characterized in our previous report.[7] Lastly, we analysed the *Δ cgsF*/CGKW14 (*cgsG*) strain (Fig. 2). The UV traces from the HPLC analysis of the extract showed a formation of another intermediate **2**, an increased accumulation of **1** and an abolishment of the production of **4** (Fig. 3 v vs. vi). The chemical structure of **2** was characterized by HRESIMS (see the Chemical Characterization section below) and UV spectrum (Fig. 5). While this compound tended to cyclize spontaneously to form **3** during the purification steps, the result from the *Δ cgsE* strain indicates that CgsE is responsible for the formation lactone in **3**.[14]

In vitro characterization of CgsB and CgsF for the formation of 3, 4 and 5 using 1 as a substrate

Based on the results from mutant strains described above, we next attempted to test the proposed biosynthetic pathway through *in vitro* assays using **1** prepared from *Δ cgsF*/CGKW14 (cgsG).

We began by preparing CgsB for the *in vitro* assay as a microsomal fraction of *Saccharomyces cerevisiae* due to difficulty of producing P450s heterologously. In this assay, we were able to observe that CgsB was able to convert **1** into **3** (Fig. 6 i). Next, to examine the activity of CgsF, we expressed *cgsF* in *Escherichia coli* and obtained an approximately 70% pure CgsF (Fig. S1 in the Supplementary Information). The sample was coloured in dark yellow, suggesting that CgsB is a flavoprotein. The *in vitro* assay on CgsF revealed that it was responsible for the formation of **5** from **1** (Fig. 6 ii). The chemical structure of **5** was characterized by HRESIMS, ^{1}H NMR and ^{13}C NMR to confirm it to be a known compound[15] (see the Chemical Characterization section below). We propose the reaction mechanism for the formation of **5** to involve a flavin-mediated oxidative decarboxylation as shown in Fig. 1b. When CgsF was added to the reaction mixture alone, it catalysed the formation of **5** in the presence of 1 mM FAD. On the other hand, when **1** was incubated with CgsB and CgsF simultaneously, three compounds **3**, **4** and **5** were identified in the reaction mixture (Fig. 6 iii). Those results indicate that CgsF exhibits relaxed substrate specificity, accepting both **1** and **2** as its substrates. We predict that the substrates are initially hydroxylated in a similar fashion by the flavin-containing monooxygenase (FMO) CgsF through a typical FAD-catalysed reaction, followed by a decarboxylation to yield the final products **4** and **5**.

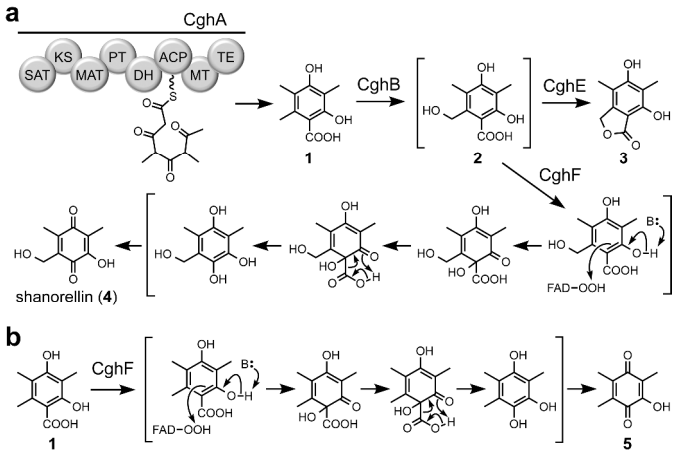


Fig.1 (a) Proposed mechanism for the biosynthesis of **3** and **4** via **1** and **2**. (b) Proposed reaction mechanism of CgsF-catalysed formation of **5** from **1**. Abbreviations: SAT, starter unit:ACP transacylase; KS: ketosynthase; MAT: malonyl-CoA acyltransferase; PT: product template; DH: dehydratase; ACP: acyl carrier protein; MT: methyltransferase; TE: thioesterase.

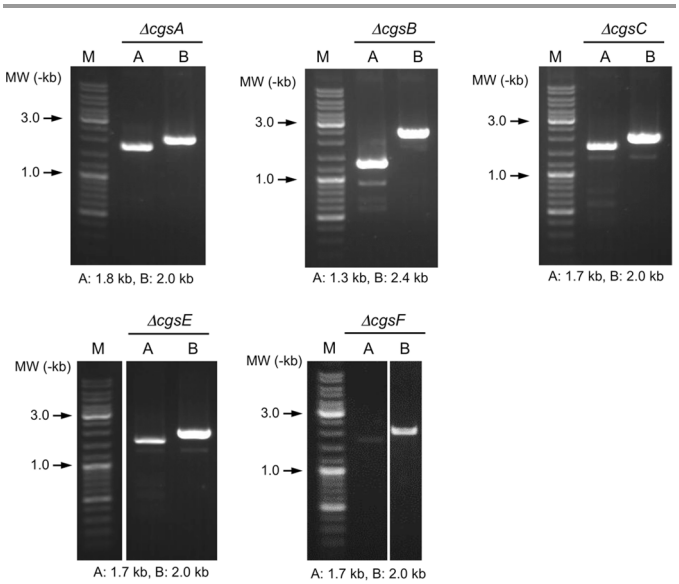


Fig.2 PCR analysis for confirming the deletion of *cgsA*, *B*, *C*, *E*, and *F* from the genome of the *Chaetomium globosum Δ CgpyrG*/*Δ CgligD* strain, CGKW14.[2] Genomic DNA from each of the deletion strain was used as the template for the PCR reaction. Details of the primers used are given in the Experimental section and Table S2 in the Supplementary Information. Abbreviations: M, molecular

Medicinal Chemistry Communications Accepted Manuscript

This journal is © The Royal Society of Chemistry 2012

marker; Lane A, PCR fragment A; Lane B, PCR fragment B. Expected DNA size were shown below the pictures.

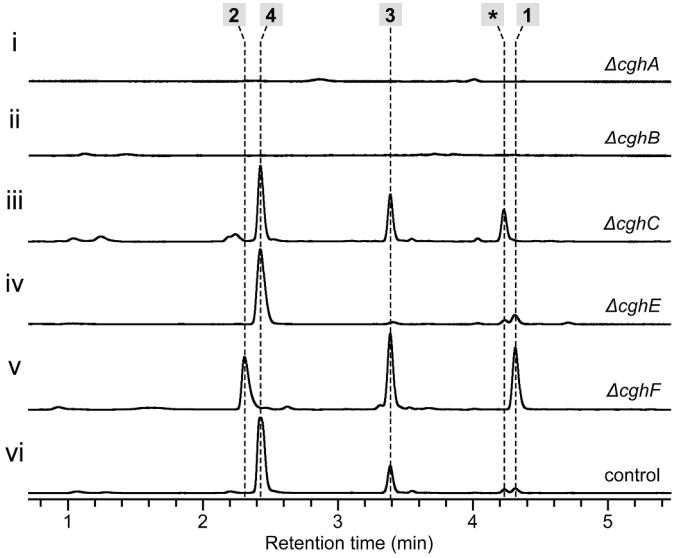


Fig.3 HPLC traces of metabolic extracts from the cultures of various *C. globosum* strains to assign the genes to the proposed biosynthetic steps involved in the formation of shanorellin. All deletions were carried out in CGKW14[2] by replacing the target gene with *CgpyrG* or *hph* via homologous recombination. All cultures were grown in MYG medium for 7 days. All HPLC traces were monitored at 280 nm. Plots are extract of the culture of (i) *ΔcgsA* strain, (ii) *ΔcgsB* of *cgsG* overexpression strain, (iii) *ΔcgsC* of *cgsG* overexpression strain, (iv) *ΔcgsE* of *cgsG* overexpression strain, (v) *ΔcgsF* of *cgsG* overexpression strain and (vi) *cgsG* overexpression strain. *: Peaks that were specifically present in mutant strains but were not structurally elucidated due to poor stability.

Chemical Characterization of 2 and 5

Detailed compounds purification and structural characterization were described in our previous literature.[7]

Compound **2**:
ESI-MS: *m/z* 211 $(M-H)^-$; HRESIMS: *m/z* 211.06075 $(M-H)^-$, calcd. for $C_{10}H_{11}O_5^-$, 211.06120, Δ = 0.5 mmu.

Compound **5**[15]:
1H NMR ($CDCl_3$, 500 MHz) δ: 5.29 (1H, s, OH), 2.04 (3H, s, 5-CH_3), 2.02 (3H, s, 6-CH), 1.91 (3H, s, 3-CH_3); ^{13}C NMR ($CDCl_3$, 125 MHz) δ: 187.8 (C-4), 183.6 (C-1), 150.8 (C-2), 143.6 (C-6), 136.2 (C-5), 116.8 (C-3), 13.0 (5-CH_3), 11.7 (6-CH_3), 8.3 (3-CH_3); ESI-MS: *m/z* 165 $(M-H)^-$; HRESIMS: *m/z* 165.05797 $(M-H)^-$, calcd. for $C_{10}H_{11}O_5^-$, 165.05572, Δ = 2.3 mmu.

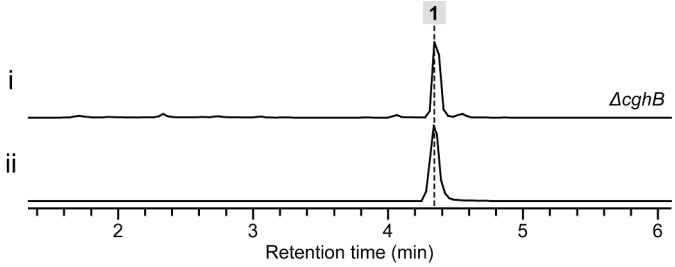


Fig.4 Analysis of the culture extract of the *ΔcgsB* strain for the production of **1**. Extracted MS trace of (i) the culture extract of the *ΔcgsB* strain and (ii) the authentic reference of **1**. Extracted MS traces corresponding to the m/z^+ for **1** (m/z^+ = 195, i and ii) are shown.

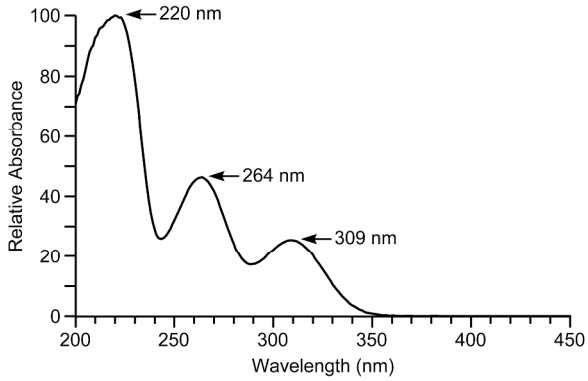


Fig.5 UV spectrum of **2** showing maxima at 220, 264 and 309 nm.

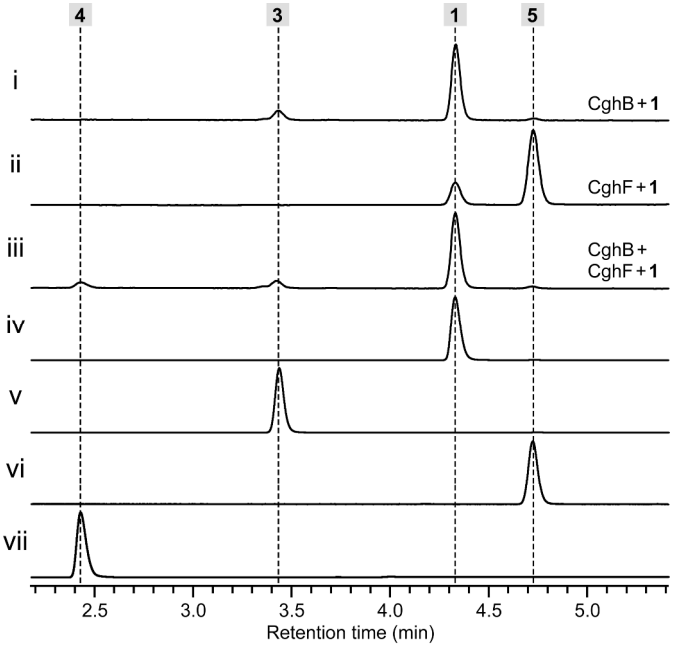


Fig. 6 *In vitro* analysis of the activities of CgsB and CgsF against **1** as a substrate. Detailed reaction conditions are given in the Experimental section. All traces were monitored at 280 nm. HPLC profiles of (i) the reaction mixture containing CgsB with **1**, showing the formation of **3** for 4 h reaction; (ii) the reaction mixture containing CgsF with **1**, showing the formation of **5** for 4 h reaction; (iii) the reaction mixture containing CgsB and CgsF with **1**, showing the formation of **3** and **4** for 4 h reaction; the authentic reference of (iv) **1**, (v) **3**, (vi) **5** and (vii) **4**..

Conclusion

Our study has elucidated the shanorellin biosynthetic pathway with gene knockout experiments using our engineered strain of *C. globosum* capable of site-specific homologous recombination. Furthermore, we demonstrated *in vitro* that CgsB and CgsF, a P450 and an FMO, respectively, are each responsible for a hydroxylation step in the biosynthesis of shanorellin. CgsF was also found to accept an earlier intermediate of the pathway as a substrate to form a dehydroxylated shanorellin, indicating the promiscuous nature

Medicinal Chemistry Communications Accepted Manuscript

of this FMO. Such relaxed substrate specificity of redox enzymes has been seen in other systems, and has been recognized as one of the key mechanisms of how nature quickly generates diversity in the natural products that originate from a single biosynthetic pathway.[16,17] The promiscuity is also a potentially valuable trait for engineered biosynthesis of novel natural products.

Experimental

Strains and general techniques for DNA manipulation

For the construction of disruption cassettes and confirmation of the modified genotype, the genomic DNA isolated from *C. globosum* or CGKW14 or the transformants[18] was used as template for PCR reactions. Genomic DNA from above strains was prepared using the CTAB isolation buffer at pH 8.0 (20 g L^{-1} CTAB, 1.4 M sodium chloride and 20 mM EDTA). The gene-specific primers are listed in Table S2 in the Supplementary Information. PCR was performed using KOD Plus Neo (TOYOBO Co., Ltd.). Sequences of PCR products were confirmed through DNA sequencing (Macrogen Japan Corporation). *E. coli* XL1-Blue (Stratagene) was used for plasmid propagation. DNA restriction enzymes were used as recommended by the manufacturer (Fermentas).

Spectroscopic analyses

NMR spectra were obtained with a JEOL JNM-ECA 500 MHz spectrometer (^{1}H 500 MHz, ^{13}C 125 MHz). ^{1}H NMR chemical shifts are reported in parts per million (ppm) using the proton resonance of residual solvent as reference: $CDCl_3$ δ 7.26. ^{13}C NMR chemical shifts are reported relative to $CDCl_3$ δ 77.16.[19] Mass spectra were recorded with a Thermo SCIENTIFIC Exactive liquid chromatography mass spectrometer by using both positive and negative ESI. LC–MS was conducted with a Thermo SCIENTIFIC Exactive liquid chromatography mass spectrometer by using positive and negative electrospray ionizations. Samples were separated for analysis on an ACQUITY UPLC 1.8 µm, 2.1 × 50 mm C18 reversed-phase column (Waters) using a linear gradient of 5–30% (v v^{-1}) MeCN in H_2O supplemented with 0.05% (v v^{-1}) formic acid at a flow rate of 500 µL min^{-1}. UV–visible spectrum of **2** in (solvent) was collected at room temperature with an ACCELA PDA Detector.

Preparation of the deletion strains

Deletion of target genes in *C. globosum* was carried out by homologous recombination using the *C. globosum* strain CGKW14 whose random nonhomologous recombination activity has been disabled by the disruptions of *CgligD* and *CgpyrG*.[2] A disruption cassette comprised of a selection marker *pyrG* flanked on both sides by a 1000-base pair fragment that is homologous to the site of recombination in the *C. globosum* genome was introduced to CGKW14 to replace the target gene with the selection marker. Disruption of the target gene was confirmed by amplifying the disrupted segment from the genomic DNA by PCR.

Transformation and cultivation for production of secondary metabolites

A mutant CGKW14 strain was initially cultured on oatmeal agar plates (60 g L^{-1} oatmeal and 12.5 g L^{-1} agar) at 30 °C for 14 days. Approximately 1×10^6 to 4×10^6 of sexual spores, or ascospores, collected from a single plate were used to inoculate 200 mL of MYG medium (10 g L^{-1} malt extract, 4 g L^{-1} glucose and 4 g L^{-1} yeast extract), which was shaken for additional 16 h at 30 °C. Grown cells were collected by centrifugation and washed with 1 M sorbitol. The cells were incubated with 4 mL of 1 M sorbitol, 50 mg mL^{-1} lysing enzyme (Sigma-Aldrich) and 1500 units of β-glucuronidase at 30 °C for 3 h. The resulting protoplasts were filtered and subsequently centrifuged at 1,500 × *g* for 5 min at room temperature. The collected protoplasts were washed with 1 M sorbitol and centrifuged to remove the wash solution. The cells were suspended in STC buffer at pH 8.0 (1 M sorbitol, 10 mM calcium chloride and 10 mM Tris-HCl) to approximately 5×10^7 mL^{-1} protoplast concentration. Then, 40 µL of PEG solution at pH 8.0 (400 mg mL^{-1} polyethylene glycol 8,000, 50 mM calcium chloride and 10 mM Tris-HCl) was added to 200 µL of the protoplast suspension. The mixture was subsequently combined with 4 µg of the DNA fragment with which the cells were to be transformed. The mixture was incubated on ice for 20 min to allow the transformation to proceed. After incubation on ice, 1 mL of the PEG solution was added to the reaction mixture, and the mixture was incubated at room temperature for additional 5 min. The resulting cells were plated on an MYG–sorbitol agar medium (MYG medium with 15 g L^{-1} agar and 1 M sorbitol) with a suitable selection agent. To select for hygromycin resistance, the DNA–protoplast mixture was plated initially on an MYG–sorbitol agar medium without marker antibiotics. After incubating the plate at 30 °C for 5 h, the cells were overlaid with an MYG–sorbitol agar medium supplemented with 200 µg mL^{-1} hygromycin B to perform the selection. To prepare for a protoplast-mediated transformation of strains with *CgpyrG* deficiency, cells were plated on a RM agar medium (2 g L^{-1} L-asparagine, 1.5 g L^{-1} ammonium chloride, 0.12 g L^{-1} magnesium sulphate, 1.4 g L^{-1} potassium phosphate monobasic, 1.5 g L^{-1} sodium phosphate dibasic, 3 g L^{-1} sodium sulphate, 10 g L^{-1} glucose, 1 µg L^{-1} thiamine and 15 g L^{-1} agar).[20] After 3 days of incubation at 30 °C, grown colonies were transferred onto a fresh MM (dextrose 10 g L^{-1}, L-asparagine 2 g L^{-1}, $MgSO_4$•$7H_2O$ 0.2 g L^{-1}, NH_4Cl 1.5 g L^{-1}, KH_2PO_4 1.35 g L^{-1}, Na_2SO_4 0.3 g L^{-1}, Na_2HPO_4 1.45 g L^{-1} and thiamine•HCl 1 µg mL^{-1}) or MYG containing hygromycin B agar medium and were incubated for several days at 30 °C. The resultant cells were transferred onto oatmeal agar plates and incubated for 3 days at 30 °C. The grown mycelia were inoculated into 100 mL of MYG liquid medium at 30 °C for 5 days shaken at 180 rpm. The broth was extracted with organic solvents for isolation of compounds.

Construction of Disruption Cassette and deletion strain from CGKW14

To prepare a deletion strain of CGKW14 missing a functional copy of a target gene, initially a disruption cassette of each of the target gene was constructed. A disruption cassette included a selection marker (orotidine-5'-phosphate decarboxylase gene *CgpyrG* or hygromycin B phosphotransferase gene *hph*) flanked on both sides by a 1000-base pair fragment that is homologous to the site of recombination at or near the ends of the target gene in the CGKW14 genome. The disruption cassette was introduced into CGKW14 (*ΔCgpyrG*/ *ΔCgligD*) to allow homologous recombination to take place through

Medicinal Chemistry Communications Accepted Manuscript

This journal is © The Royal Society of Chemistry 2012

which the target gene would be replaced by the selection marker. Disruption of the target gene was confirmed by amplifying the disrupted segment of the genomic DNA by PCR. For the construction of the disruption cassette, PCR was carried out using KOD Plus Neo DNA polymerase as recommended by the manufacturer. For the 1000-base pair homologous region that is to be appended to each terminal of the selection marker to knock out the target gene via homologous recombination, both regions were amplified from the CGKW14 genomic DNA by PCR. The following primer sets were used to prepare the required flanking homologous regions and selectable markers for each of the target genes:

Gene	Marker	Primer (PCR template)	Confirmation
cgsA knock-out	*hph*	19153-Fw1/19153-Rv1 (gDNA) hph_Fw/hph_Rv (pKW3202) 19153-Fw2/19153-Rv2 (gDNA)	19153-Fw3/Cg-KO-hph-Rv Cg-KO-hph-Fw/19153-Rv3
cgsB knock-out	*hph*	19154-Fw4/19154-Rv4 (gDNA) hph_Fw/hph_Rv (pKW3202) 19154-Fw5/19154-Rv5 (gDNA)	19154-Fw3/Cg-KO-hph-Rv Cg-KO-hph-Fw/19154-Rv3
cgsC knock-out	*hph*	19155-Fw1/19155-Rv1 (gDNA) hph_Fw/hph_Rv (pKW3202) 19155-Fw2/19155-Rv2 (gDNA)	19155-Fw3/Cg-KO-hph-Rv Cg-KO-hph-Fw/19155-Rv3
cgsE knock-out	*hph*	19156-Fw1/19156-Rv1 (gDNA) hph_Fw/hph_Rv (pKW3202) 19156-Fw2/19156-Rv2 (gDNA)	19156-Fw3/Cg-KO-hph-Rv Cg-KO-hph-Fw/19156-Rv3
cgsF knock-out	*hph*	19157-Fw1/19157-Rv1 (gDNA) hph_Fw/hph_Rv (pKW3202) 19157-Fw2/19157-Rv2 (gDNA)	19157-Fw3/Cg-KO-hph-Rv Cg-KO-hph-Fw/19157-Rv3

Amplified DNA fragments were visualized by agarose gel electrophoresis with ethidium bromide using UV (365 nm) transilluminator and purified with QIAquick Gel Extraction Kit (QIAGEN). Three purified fragments (two homologous regions and the selectable marker), each at 50 to 150 ng in a total volume of 45 μL, were mixed with the delivery vector pKW1810[3] (2 μg) pre-digested with *Sma* I (10 units) or pRS426[21] (0.1 μg) pre-digested with *Sac* I (10 units) and *Kpn* I (10 units) at 37 °C for 30 min. The mixture was transformed into *S. cerevisiae* BY4741 for *in vivo* homologous recombination. These four DNA fragments were joined *in situ* by the endogenous homologous recombination activity of *S. cerevisiae* through the 15-bp homologous sequences present at the ends of those DNA fragments. The desired transformants were selected for the presence of the selection marker *URA3* on a uracil-deficient plate. The resulting plasmid carrying the selection marker gene flanked by two 1000 bp homologous regions was recovered from the yeast transformant and transferred to *Escherichia coli*. The plasmid was amplified in *E. coli* for subsequent characterization by restriction enzyme digestion and DNA sequencing to confirm its identity. The PCR products were transformed into CGKW14. Transformants whose target gene was replaced by *CgpyrG* were selected on MM agar plates. Transformants whose target gene was replaced by *hph* were selected on MYG agar plates supplemented with 200 μg mL^{-1} hygromycin B. The selection confirmed the integration of the disruption cassette into the genome of the transformant.

Confirmation of targeted deletion for *cgsA*, *B*, *C*, *E* and *F* genes by PCR

To verify that the cassette was inserted into the target gene, the genomic DNA isolated from the transformants was analysed by PCR. Results of the PCR analyses are given in Fig. 2. Two sets of PCR primers were designed for this verification. For the first set, one primer that anneals to the selection marker and another primer that anneals at the 3' side of the homologous region were designed (lane A). For the second primer set, one primer that anneals near the 3' end of the target gene and another primer that anneals approximately 0.5 kb inside of the target gene (lane B). Combination of the results from those two separate PCR reactions ensured us that we had the targeted gene replaced by our desired selection marker.

Construction of pKW19192 for expression of *cgsB* in *S. cerevisiae*, and pKW19208 for expression of *cgsF* in *E. coli*

To construct the vector for expressing *cgsB* in yeast, *cgsB* ORF was amplified with a primer set 19192-Fw1/19192-Rv1 (Table S2) from cDNA that was synthesized from mRNA isolated from *C. globosum*. The amplicon was mixed with the delivery vector pKW1810[3] (0.1 μg), which was digested with *Sma* I (10 units) at 37 □C for 8 h, for *in vivo* homologous recombination. The mixture was transformed into *S. cerevisiae* BY4741. The two fragments were joined *in situ* by the endogenous homologous recombination activity of *S. cerevisiae* through the 15-bp homologous sequences present in both DNA fragments. The desired transformants were selected for the presence of the selection marker *HIS3* on a histidine-deficient plate. The resulting plasmid pKW19192 carrying the *cgsB* gene was recovered from the yeast transformant and transferred to *E. coli*. The plasmid was amplified in *E. coli* for subsequent characterization by restriction enzyme digestion and DNA sequencing to confirm its identity.

For expression of *cgsF* in *E. coli*, *cgsF* gene was amplified from cDNA that was synthesized from mRNA isolated from *C. globosum* with a primer set 19208-Fw1/19208-Rv1 (Table S2). A plasmid backbone was also amplified from pET28b vector with a primer set and 19205-Fw1/19205-Rv1 (Table S2). These DNA fragments were then simultaneously joined together using GeneArt Seamless Cloning and Assembly kit (Life Technologies). The resulting plasmid was amplified in *E. coli* for restriction digestion analysis by *Nco* I (10 units) and *Xho* I (10 units), and later sequenced to confirm its identity. This plasmid was named pKW19208. This vector allowed production of CgsF having a C-terminal His_6-tag. The plasmid carrying the desired ORF was recovered from the *E. coli* transformant. The plasmid was amplified for restriction digestion analysis, and later sequenced to confirm the identity of the plasmid.

***In vitro* assays of CgsB and CgsF for elucidating the biosynthesis of 3, 4 and 5 by using 1 as a substrate**

For elucidating the proposed shanorellin biosynthetic mechanism involving CgsB and CgsF, we prepared CgsB-containing microsomal fraction of *S. cerevisiae* and *E. coli*-produced CgsF that was partially purified through Ni-NTA Sepharose resin column chromatography. Samples of CgsB[20] and CgsF[22] were prepared following the methods reported previously. To examine the activity of CgsB, the assay mixture

Medicinal Chemistry Communications Accepted Manuscript

(250 µL) containing 200 µL of microsomal fraction containing CgsB from 200 mL yeast culture, 130 µM of **1**, 1 mM NADPH and 100 mM Tris-HCl (pH 7.4) was incubated at 30 °C for 4 h. To test the activity of CgsF, the assay mixture (250 µL) containing 0.24 µM of partially purified CgsF, 130 µM of **1**, 1 mM NADPH, 1 mM FAD and 100 mM Tris-HCl (pH 7.4) was incubated at 30 °C for 4 h. After the 4-h incubation, both reaction mixtures were quenched by addition of 250 µL of saturated potassium hydrogen sulfate to adjust the pH to 2.0 and extracted with 500 µL of EtOAc. The organic layer was concentrated *in vacuo*. The dried material was dissolved in 50 µL of *N*,*N*-dimethylformamide and subjected to LC–MS analysis.

Acknowledgements

We would like to express our appreciation to financial support from Japan Society for the Promotion of Science (JSPS) through the "Funding Program for Next Generation World-Leading Researchers," initiated by the Council for Science and Technology Policy (No. LS103) (K.W.). These works were also supported in part by Amano Enzyme Foundation (K.W.), by Mochida Memorial Foundation for Medical and Pharmaceutical Research (K.W.), by The Naito Foundation Japan (K.W.), by Nagase Science and Technology Foundation Japan (K.W.) and by The Tokyo Biochemical Research Foundation (K.W.).

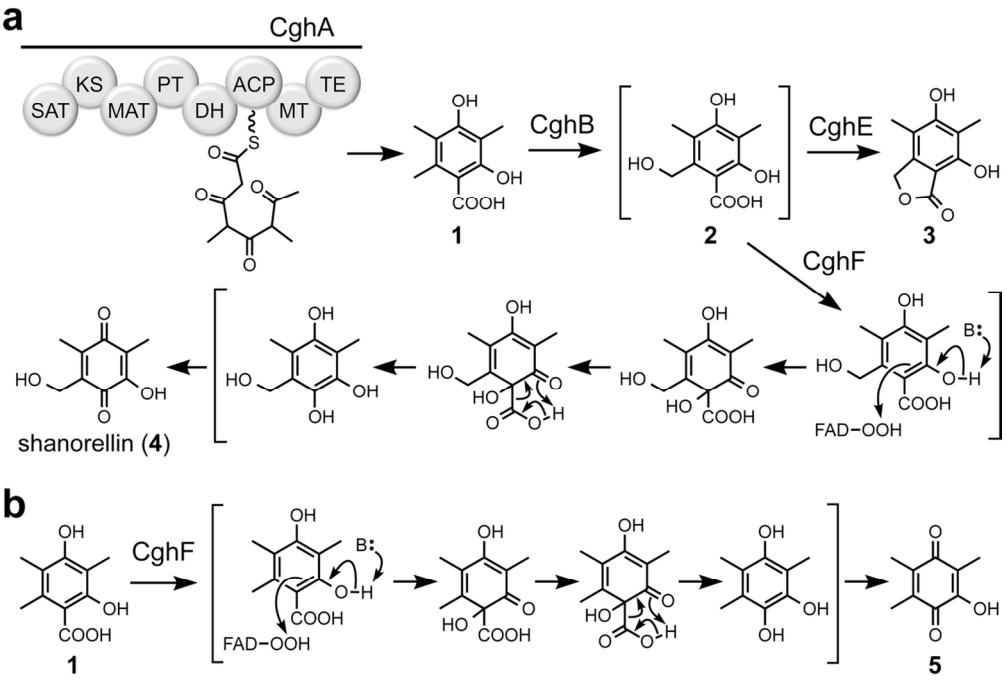
Notes and references

[a] Department of Pharmaceutical Sciences, University of Shizuoka, Shizuoka 422-8526, Japan
[b] School of Biosciences, The University of Nottingham Malaysia Campus, Selangor 43500, Malaysia

1 J. M. Winter, S. Behnken and C. Hertweck, *Curr. Opin. Chem. Biol.*, 2011, **15**, 22–31.
2 T. Nakazawa, K. Ishiuchi, M. Sato, Y. Tsunematsu, S. Sugimoto, Y. Gotanda, H. Noguchi, K. Hotta and K. Watanabe, *J. Am. Chem. Soc.*, 2013, **135**, 13446–13455.
3 K. Ishiuchi, T. Nakazawa, T. Ookuma, S. Sugimoto, M. Sato, Y. Tsunematsu, N. Ishikawa, H. Noguchi, K. Hotta, H. Moriya and K. Watanabe, *Chembiochem*, 2012, **13**, 846–854.
4 J. M. Winter, M. Sato, S. Sugimoto, G. Chiou, N. K. Garg, Y. Tang and K. Watanabe, *J. Am. Chem. Soc.*, 2012, **134**, 17900–17903.
5 A. A. Brakhage and V. Schroeckh, *Fungal. Genet. Biol.*, 2011, **48**, 15–22.
6 K. Scherlach and C. Hertweck, *Org. Biomol. Chem.*, 2009, **7**, 1753–1760.
7 Y. Tsunematsu, S. Ichinoseki, T. Nakazawa, N. Ishikawa, H. Noguchi, K. Hotta and K. Watanabe, *J. Antibiot. (Tokyo)*, 2012, **65**, 377–380.
8 M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis and T. L. Madden, *Nucleic Acids Res.*, 2008, **36**, W5–9.
9 Y. H. Chooi and Y. Tang, *J. Org. Chem.*, 2012, **77**, 9933–9953.
10 D. W. Brown, T. H. Adams and N. P. Keller, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 14873–14877.
11 J. W. Bok, D. Hoffmeister, L. A. Maggio-Hall, R. Murillo, J. D. Glasner and N. P. Keller, *Chem. Biol.*, 2006, **13**, 31–37.
12 J. W. Bok, Y. M. Chiang, E. Szewczyk, Y. Reyes-Dominguez, A. D. Davidson, J. F. Sanchez, H. C. Lo, K. Watanabe, J. Strauss, B. R. Oakley, C. C. Wang and N. P. Keller, *Nat. Chem. Biol.*, 2009, **5**, 462–464.
13 Y. M. Chiang, E. Szewczyk, A. D. Davidson, N. Keller, B. R. Oakley and C. C. Wang, *J. Am. Chem. Soc.*, 2009, **131**, 2965–2970.
14 B. G. Hansen, E. Mnich, K. F. Nielsen, J. B. Nielsen, M. T. Nielsen, U. H. Mortensen, T. O. Larsen and K. R. Patil, *Appl. Environ. Microbiol.*, 2012, **78**, 4908–4913.
15 M. Shimizu, H. Orita, T. Hayakawa and K. Takehira, *Tetrahedron Lett.*, 1989, **30**, 471–474.
16 P. Wang, X. Gao and Y. Tang, *Curr. Opin. Chem. Biol.*, 2012, **16**, 362–369.
17 K. Ishiuchi, T. Nakazawa, F. Yagishita, T. Mino, H. Noguchi, K. Hotta and K. Watanabe, *J. Am. Chem. Soc.*, 2013, **135**, 7371–7377.
18 D. M. Binninger, C. Skrzynia, P. J. Pukkila and L. A. Casselton, *EMBO J.*, 1987, **6**, 835–840.
19 H. E. Gottlieb, V. Kotlyar and A. Nudelman, *J. Org. Chem.*, 1997, **62**, 7512–7515.
20 T. Saruwatari, F. Yagishita, T. Mino, H. Noguchi, K. Hotta and K. Watanabe, *Chembiochem*, 2014, **15**, 656–659.
21 R. S. Sikorski and P. Hieter, *Genetics*, 1989, **122**, 19–27.
22 Y. Tsunematsu, N. Ishikawa, D. Wakana, Y. Goda, H. Noguchi, H. Moriya, K. Hotta and K. Watanabe, *Nat. Chem. Biol.*, 2013, **9**, 818–825.

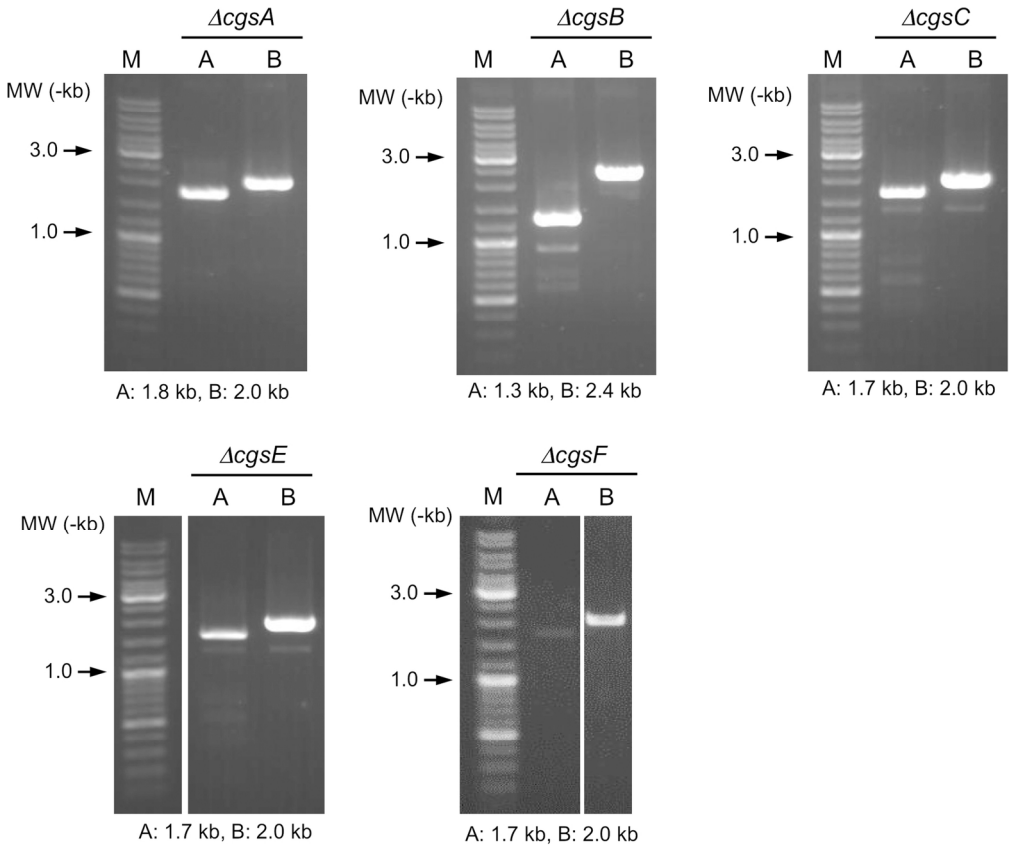
Medicinal Chemistry Communications Accepted Manuscript

This journal is © The Royal Society of Chemistry 2012



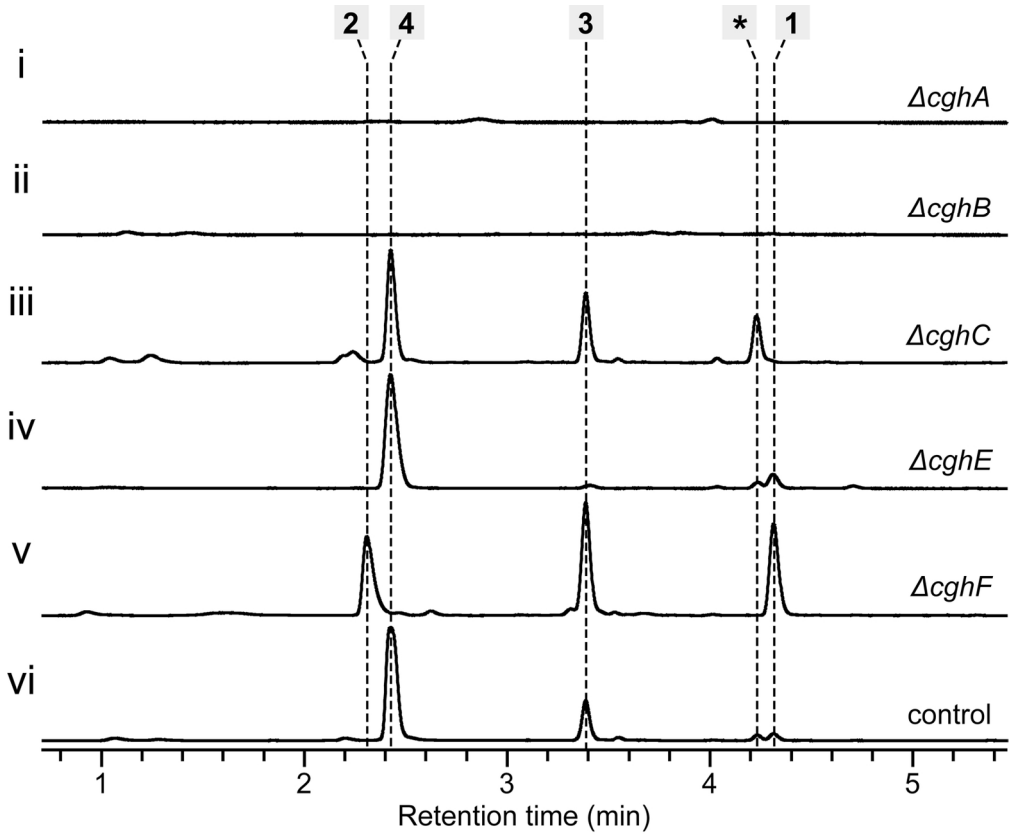
58x39mm (600 x 600 DPI)

Medicinal Chemistry Communications Accepted Manuscript



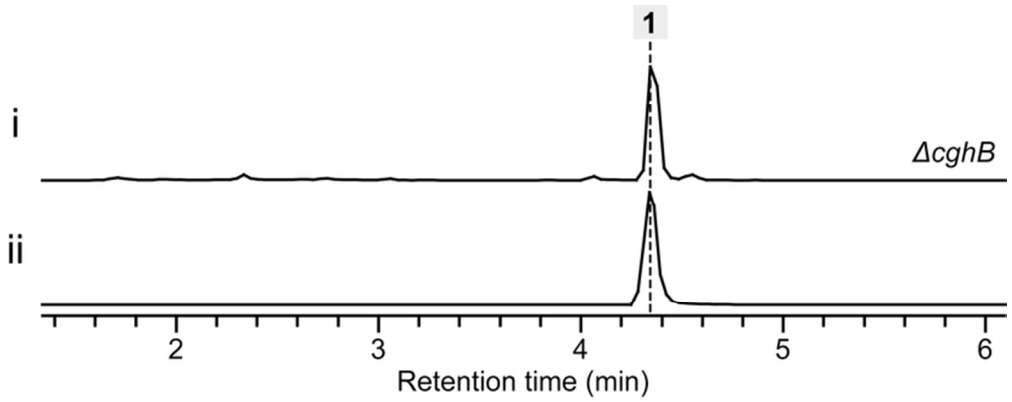
73x61mm (600 x 600 DPI)

Medicinal Chemistry Communications Accepted Manuscript



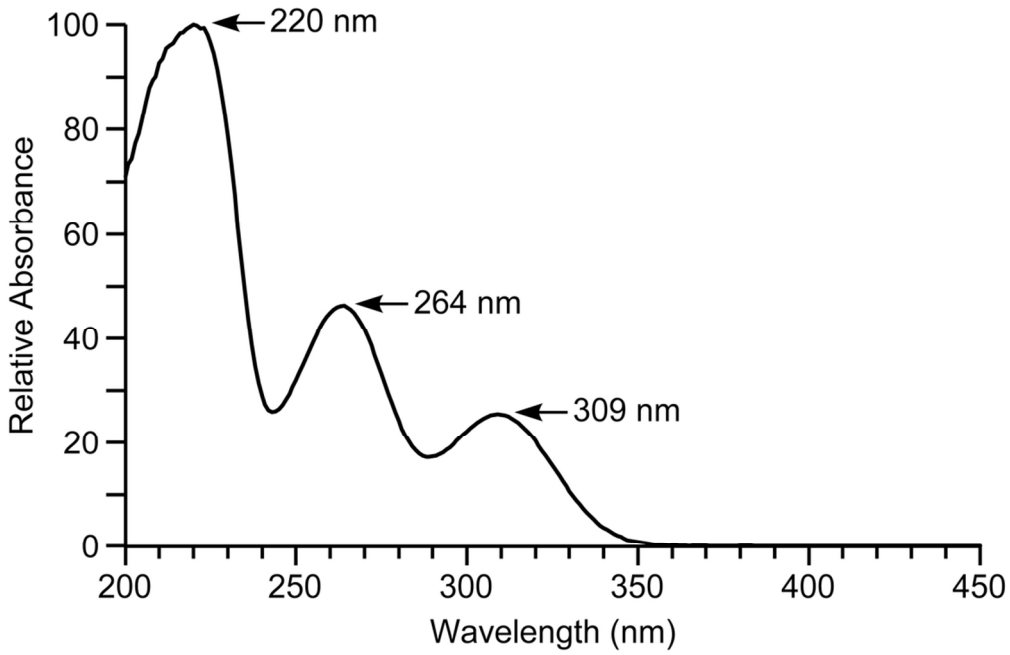
72x60mm (600 x 600 DPI)

Medicinal Chemistry Communications Accepted Manuscript



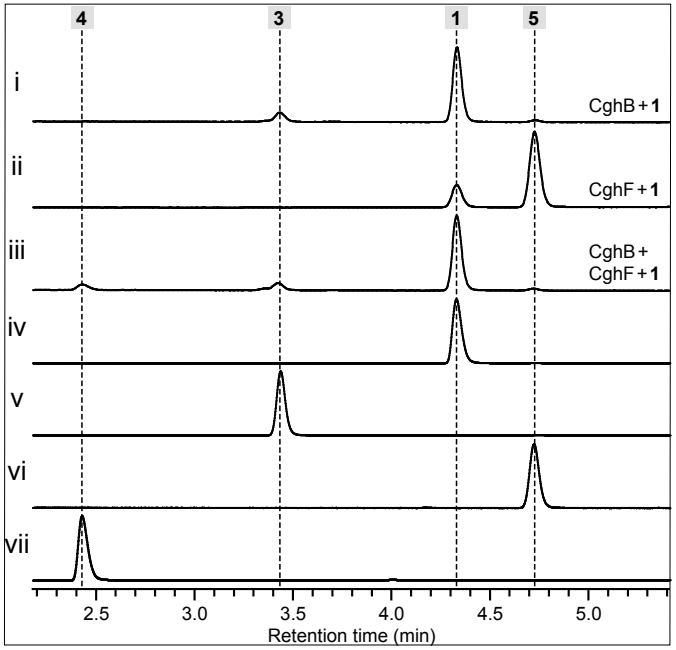
34x13mm (600 x 600 DPI)

Medicinal Chemistry Communications Accepted Manuscript



49x31mm (600 x 600 DPI)

Medicinal Chemistry Communications Accepted Manuscript



Medicinal Chemistry Communications Accepted Manuscript