Noda Institute for Scientific Research GRANT
2011 Research Grant

Establishing a new methodology for genome mining and biosynthesis of natural products and their analogs through a yeast molecular genomics

Kenji WATANABE

Department of Pharmaceutical Sciences, University of Shizuoka

Research aims

Fungal genome sequencing revealed the presence of many genes coding for natural product biosynthetic enzymes, including polyketide synthases and nonribosomal peptide synthetases. However, it remains a challenge to characterize these enzymes and identify the compounds they synthesize, either in their original hosts or more tractable heterologous hosts, such as yeast. Here, we have developed a streamlined method for isolating biosynthetic genes from fungal sources and producing bioactive molecules using an engineered *Saccharomyces cerevisiae* strain as a host. We used overlap extension PCR and yeast homologous recombination technology to clone a desired fungal polyketide synthase or nonribosomal peptide synthetase gene (5–20 kb) into a yeast expression vector quickly and efficiently. This approach was used successfully to clone five polyketide synthases and one nonribosomal peptide synthetase from various fungal species. Subsequent detailed chemical characterizations of the resulting natural products identified 6 polyketide and 2 nonribosomal peptide products, some of which were new compounds. Our system should facilitate the efforts in mining the uncharacterized fungal biosynthetic potential for identifying novel natural products and rationally engineering in the biosynthetic pathways for production of analogs possessing desired bioactivity.

Methods

Assembling the PKS and NRPS biosynthetic genes on plasmids

To construct the expression vector in yeast, GAL1 promoter–ADH1 terminator cassette was taken from the genome of *S. cerevisiae* and cloned into pTOWug2-836 vector to yield pKW1250. The HIS3 gene responsible for L-histidine biosynthesis was cloned out from pRS423 and integrated into pTOWug2-836 replacement with URA3, uracil biosynthetic gene, yielding pKW1113 by the homologous recombination activity. The TEF1 promoter–CYC1 terminator cassette was inserted into pKW1113(HIS3) vector to yield pKW1810[1,2]. In pKW1810, the expressed ORF will have at least one Flag-tag at the N terminus. A C-terminal hexahistidine tag-coding fragment and multiple cloning sites originating from pET32a plasmid were also present in pKW1810. Each ORF, CHGG_00542, CHGG_10027, CHGG_10128, CC1G_05377 and Afu6g12080, was predicted based on the sequence information available from the Broad Institute database, and their predicted functions were determined by comparison to known proteins using the BLAST peptide sequence database search program and the FFAS03 protein sequence profile-profile alignment and fold recognition program.

A full-length cDNA of the putative PKS or NRPS gene was synthesized by reverse transcriptase with a oligo$(dT)_{20}$ Primer using the corresponding mRNA in the isolated total RNA as a template. To construct the yeast expression vector carrying one of the target biosynthetic gene, the gene was amplified from the full-length cDNA by PCR in overlapping smaller fragments no more than 3 kb in length using multiple pairs of primers. Subsequently, these DNA fragments were reconstituted into a full-length double strand DNA coding for the desired ORF by the overlap extension PCR method. This amplified ORF was combined with pKW1810 linearized by restriction enzyme digestion. Both DNA fragments were then simultaneously introduced into *S. cerevisiae* BY4741. The two fragments were joined in situ by the endogenous homologous recombination activity of *S. cerevisiae* through the 25-bp homologous sequences present in both DNA fragments. The desired transformants were selected by identifying the plasmid vector marker, HIS3, on an L-histidine-deficient plate. The plasmid now carrying the desired ORF was recovered from the yeast transformant and transferred to *E. coli*. The plasmid was amplified for examining its structure by restriction enzyme digestion, and later sequenced to confirm identity of the plasmid.

Results

Characterization of polyketide products biosynthesized by Chaetomium globosum PKS CHGG_00542 in yeast[3)]

We cultivated the SCKW5 strains on an YPD liquid medium at a large scale and purified the compounds using column chromatography and HPLC when necessary. Fractions obtained from the HPLC eluent of the pKW14050(CHGG_00542)/SCKW5 culture extract were dried in vacuo to furnish pure solid **1**[3)]. Subsequently, **1** was

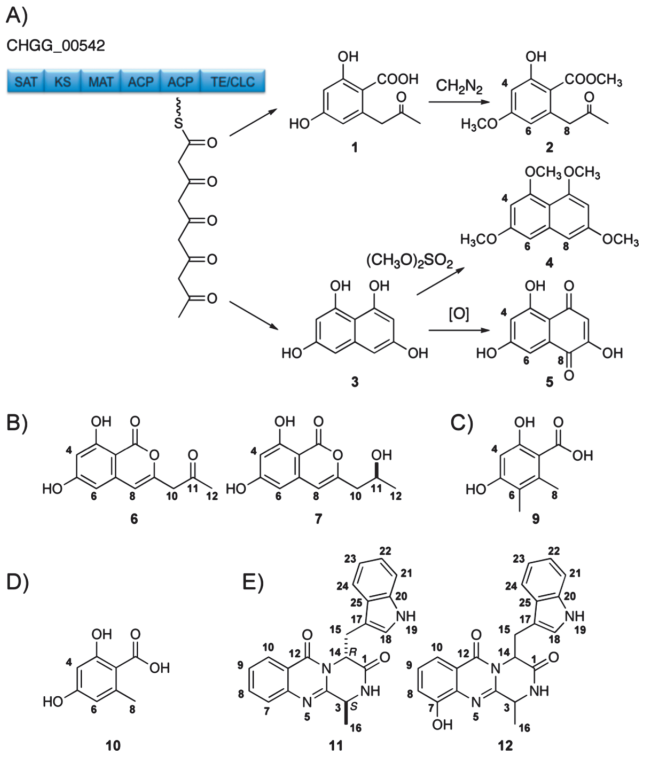

Figure 1. Chemical structures of the compounds biosynthesized by the engineered yeast. A) Compounds **1**–**5** were biosynthesized by the wild type CHGG_00542 PKS. B) Compounds **6** and **7** were biosynthesized by the S2010A mutant CHGG_00542 PKS. C) Compound **9** was biosynthesized by CHGG_10027 and CHGG_10128 PKSs. D) Compound **10** was biosynthesized by CC1G_05377 PKS. E) Compounds **11** and **12** were biosynthesized by Afu6g12080 NRPS.

reacted with diazomethane to give a methyl ester **2** at 1.3 mg/L (Figure 1A). Furthermore, the engineered yeast allowed us to biosynthesize **3**, which was converted into a methyl ester **4** by dimethyl sulfate at 32.5 mg/L of culture for further chemical structure characterizations. Additionally, **3** was presumably oxidized spontaneously to yield **5** at 25 mg/L as another biosynthetic product produced by CHGG_00542. Based on in vitro studies, Ebizuka *et al.* have proposed a biosynthetic pathway for **1** involving a PKS called PKS1 from *A. oryzae*[4]. Therefore, CHGG_00542 could play a functionally equivalent role as PKS1. Interestingly, the authors were able to rationally re-engineer the biosynthetic pathway to produce **6** (Figure 1B) by inactivating the TE/CLC domain of PKS1 via a site-directed mutagenesis of the catalytic Ser2009 to an Ala. Amino acid sequence comparison between PKS1 and CHGG_00542 indicates 61% identity over the entire sequence. On the other hand, 74% identity was observed between their TE/CLC domains. To determine if CHGG_00542 is functionally similar to PKS1, we prepared a S2010A mutant of CHGG_00542, where the residue S2010 corresponds to S2009 in PKS1. We designated this mutant as CHGG_00542* and named the yeast expression vector carrying this mutant gene pKW14051. As predicted, we were able to isolate **6** from the culture of pKW14051(CHGG_00542*)/SCKW5 that was prepared under essentially the same condition used to produce **1**. However, we also unexpectedly detected the presence of another compound **7** in this culture.

Characterization of polyketide products biosynthesized by Chaetomium globosum PKSs CHGG_10027 and CHGG_10128 in yeast [3]

We were also able to isolate nonpolar compounds from the cultures of pKW14051(CHGG_10027)/SCKW5 and pKW14051(CHGG_10128)/SCKW5 at less than 0.5 and approximately 1.7 mg/L of culture, respectively. These compounds were characterized by HR-ESI-MS and NMR following similar procedures as described earlier. Surprisingly, spectroscopic analysis of these products indicated that CHGG_10027 and CHGG_10128 biosynthesized an identical compound, 6-methylorsellinic acid **9** (Figure 1C). CHGG_10027 is a type I iterative PKS with seven modules as ascertained by *in silico* analysis: SAT (starter unit: ACP transacylase), KS (ketosynthase), MAT (malonyl-CoA acyltransferase), PT (product template), ACP (acyl carrier protein), MT (methyltransferase) and RED (reductase). On the other hand, CHGG_10128 is predicted to have eight modules: SAT, KS, MAT, PT, double ACP, MT and EST (esterase/lipase). The product structure suggests that the SAT domain in both CHGG_10027 and CHGG_10128 accepts an acetyl-CoA as the only type of starting unit for the synthesis of the polyketide core structure. Also, both of these PKSs would be categorized as a non-reducing PKS due the lack of domains capable of catalyzing a reduction and/or dehydration reaction, but neither has a conventional C-terminal thioesterase domain. CHGG_10027 has a reductase domain, and CHGG_10128 carries an esterase/lipase-like domain. The only example of a type I iterative PKS with a C-terminal esterase/lipase-like domain reported was MpaC from *Penicillium brevicompactum* involved in the biosynthesis of mycophenolic acid. In this recent report, MpaC was proposed to synthesize **9**, the same compound produced by CHGG_10128, as an intermediate during the mycophenolic acid biosynthesis. Comparison of the domain architecture of CHGG_10128 and MpaC shows that both PKSs are highly homologous to each other except for the absence of the second ACP domain in MpaC. Since the amino acid sequence between the corresponding domains of CHGG_10128 and MpaC are on average approximately 42% identical and 60% similar, it is plausible that CHGG_10128 produces the same polyketide product **9** that MpaC synthesizes. However, it was unexpected that both CHGG_10027 and CHGG_10128 would produce the same polyketide compound despite that these PKSs had clearly different domain architectures. In particular, CHGG_10027 carries a C-terminal RED domain, suggesting that CHGG_10027 is expected to generate an aldehyde-containing product. It is possible that the aldehyde is accidentally oxidized to a carboxylic acid by some yeast oxidoreductase.

Conclusion

In this study, we established an innovative approach for biosynthesizing bioactive compounds of fungal origin by focusing on previously uncharacterized biosynthetic gene clusters and using an engineered *S. cerevisiae* strain as a surrogate host. We exploited the overlap extension PCR method to quickly synthesize full-length PKS and NRPS genes, which are usually 5- to 20-kb or longer, using a pool of cDNA reverse-transcribed from a total RNA isolated

from the source fungus. The amplified gene was subsequently cloned into a yeast expression vector using the recombination capability of yeast. This series of procedures allowed fast and efficient establishment of a yeast system for expression of biosynthetic genes of unknown function and production of corresponding natural products. Our results clearly demonstrated successful expression of four type I iterative PKS genes and one NRPS gene from three different fungal species in *S. cerevisiae*, all of which led to the production of a total of six compounds whose identities were characterized spectroscopically and verified the speculated functions of these biosynthetic mega-enzymes. Our plasmid-based system provides an advantage over fungal systems in terms of ease and speed of cloning the target genes, and it also tolerates handling of substantially large genes. Also, the use of a plasmid-borne system simplifies the effort of engineering biosynthetic pathways for production of various analogs using traditional molecular biological techniques. By streamlining the process of translating uncharacterized fungal biosynthetic genes into structurally characterized compounds, our methodology should facilitate the efforts in isolating novel natural products and rationally engineering in the biosynthetic pathways for production of analogs possessing comparable or even more potent bioactivity.

References

1. Tsunematsu, Y. *et al.* (2012) Overexpressing transcriptional regulator in *Chaetomium globosum* activates a silent biosynthetic pathway: evaluation of shanorellin biosynthesis. *J. Antibiot*. **65**: 377–380.
2. Nakazawa, T. *et al.* (2012) Overexpressing transcriptional regulator in *Aspergillus oryzae* activates a silent biosynthetic pathway to produce novel polyketide. *ChemBioChem* **13**: 855–861.
3. Ishiuchi, K. *et al.* (2012) Establishing a new methodology for genome mining and biosynthesis of polyketides and peptides through yeast molecular genetics. *ChemBioChem* **13**: 846–854.
4. Ebizuka, Y. *et al.* (2004) Unprecedented mechanism of chain length determination in fungal aromatic polyketide synthases. *Chem. Biol.* **11**: 1101–1106.