Discovery of novel fungal terpene synthases and their production

Tomohisa KUZUYAMA

The University of Tokyo

Research objective

Terpenes are the largest group of natural products, with approximately 80,000 reported examples with diverse natural structures. Terpenes are biosynthesized by the action of terpene cyclase on acyclic substrates, resulting in a complex polycyclic skeleton and strictly regulated absolute stereochemistry.[1] Due to their skeletal diversity and stereochemistry-regulatory mechanism, terpenes exhibit a wide variety of biological activities. Their potential functions in industrial applications and drug discovery are immeasurable, including the aromatic activity of limonene, the antiparasitic activity of artemisinin, and the anticancer activity of taxol. These terpenes have been discovered through a process known as "natural product exploration". In addition, genome information has become readily available, and genome mining has been conducted to discover new enzymes from sequence information and new terpene compounds based on functional analysis of the enzymes. However, because of the extremely low sequence similarity among microbial terpene cyclases, and the fact that most of their genes are not expressed or are cryptic, the number of examples of such analyses is limited. Therefore, microbial terpene cyclase genes are considered a major unexploited genetic resource.

In this study, diterpene A (molecular formula, $C_{20}H_{34}O_3$), which has a complex skeleton and is of fungal origin, was set as the primary research target, with the aim of identifying the terpene cyclase gene, which is the key enzyme in its biosynthesis.

In addition, we employed the hidden Markov model (HMM) method, which can also search for short sequence motifs, to promote the discovery of terpene cyclases and research their biosynthetic mechanisms.[2] However, there are still many terpene cyclases with unknown functions. In this study, we developed an HMM to discover new terpene cyclases and aimed to discover terpene compounds that have not yet been obtained by conducting functional analysis of the newly discovered enzymes.

Methods

To identify the terpene cyclase gene, the key enzyme in the biosynthesis of diterpene A (molecular formula, $C_{20}H_{34}O_3$), the genomic DNA sequence of the producing fungus was obtained and analyzed in detail. In general, biosynthetic genes of secondary metabolites often form clusters. Therefore, if a gene encoding an enzyme of unknown function exists near the geranylgeranyl diphosphate (GGPP) synthase gene, which enables the synthesis of GGPP, a

substrate of diterpene A cyclase, the enzyme of unknown function may be involved in the production of diterpene A. The presence of cytochrome P450, which is involved in hydroxylation, is also an indicator of a terpene cyclase gene. The target terpene cyclase gene was identified by expressing the candidate gene in *Escherichia coli*, preparing a recombinant enzyme, and reacting it with GGPP to determine whether the desired diterpene backbone was formed.

Terpene cyclases are extremely sophisticated biocatalysts that act alone on acyclic substrates in the presence of metal ions, such as magnesium, to form multiple carbon-carbon bonds and create complex polycyclic skeletons through multistep reactions with strict control of absolute stereochemistry. Since several important sequence motifs have been found to catalyze this multistep reaction, we created a new HMM by adding important sequence motifs found in the amino acid sequences of newly identified terpene cyclases. Next, we aimed to discover amino acid sequences of new candidate terpene cyclase from publicly available amino acid sequence databases and perform functional analysis of the recombinant proteins, as described above, to further expand the repertoire of novel terpene cyclases and discover new terpene compounds.

Results

The genomic DNA sequences of the producing fungi were obtained and analyzed in detail to identify the terpene cyclase gene involved in diterpene A biosynthesis. The only gene cluster containing both the GGPP synthase and cytochrome P450 genes was identified. The gene cluster also contained a gene of unknown function. Next, this functionally unknown gene was expressed in *E. coli*, and the recombinant enzyme was prepared and reacted with GGPP, and the reaction was found to proceed. The reaction product was analyzed by gas chromatography-mass spectrometry and found to form the carbon skeleton of diterpene A. We identified a terpene cyclase gene involved in diterpene A biosynthesis. The structure of the terpene cyclase was predicted by AlphaFold2 and was found to be very different from the structure of CotB2,[3] a previously determined terpene cyclase (Fig. 1a).

We selected 110 newly identified amino acid sequences of class I terpene cyclases from publicly available amino acid sequence databases and created an HMM. Next, using this HMM, we retrieved candidate class I terpene cyclases from the amino acid sequence database of Cyanobacteria, which has increased in size in recent years, and identified several enzymes with unknown functions. Next, we excluded amino acid sequences that showed more than 30% identity and selected several candidates for class I terpene cyclases. Among these, an open reading frame (ORF; *cpt11*) encoding a protein of 560 amino acids was found upstream of the class I terpene cyclase candidate gene, *cts11*. This ORF was presumed to be a copalyl-diphosphate (CPP) synthase, based on a homology search. Therefore, recombinant *cts11* and

cpt11 were prepared and purified in *E. coli*, yielding soluble proteins of 40 kDa and 62.5 kDa, respectively. When these recombinant enzymes were incubated with GGPP in the presence of $MgCl_2$, *syn*-CPP (**2**-PP) was produced from the reaction mixture of CPT11, and unknown compound **3** was produced from the reaction mixture of CPT11 and CTS11 (Fig. 1b). Therefore, we purified **3** and analyzed various nuclear magnetic resonance spectral data in detail and found that **3** is a novel tricyclic diterpene compound (Fig. 1c).

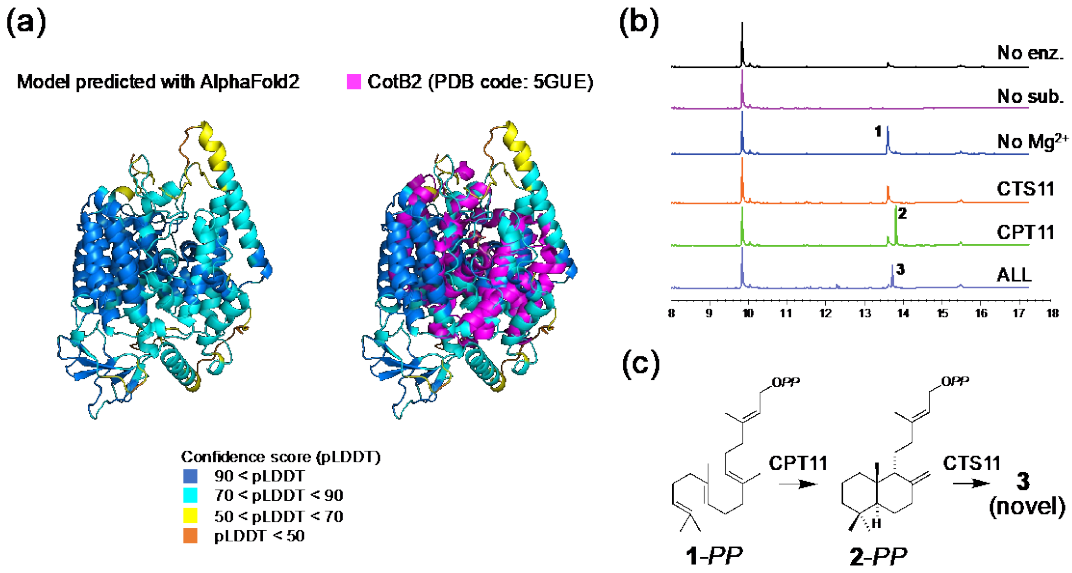

Figure 1. Predicted structure of the terpene cyclase and functions of CPT11 and CTS11.

Conclusion

Diterpene A was found to be biosynthesized by a new type of terpene cyclase that is structurally different from previously identified terpene cyclases.

The diterpene compound obtained by genome mining using the newly created HMM is a novel tricyclic diterpene compound from Cyanobacteria, and the created HMM is effective for discovering novel terpene cyclases. Novel terpene compounds are expected to be discovered using HMMs in the future.

References

(1) Meguro, A. *et al*. (2015) An unusual terpene cyclization mechanism involving a carbon-carbon bond rearrangement. *Angew. Chem. Int. Ed. Engl*. **54:** 4353-4356.

(2) Yamada, Y. *et al*. (2015) Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. USA*. **112:** 857-862.

(3) Tomita, T. *et al*. (2017) Structural insights into the CotB2-catalyzed cyclization of geranylgeranyl diphosphate to the diterpene cyclooctat-9-en-7-ol. *ACS Chem. Biol*. **12:** 1621-1628.