Metabolic Engineering 42 (2017) 1–8

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

Releasing the potential power of terpene synthases by a robust precursor supply platform

Guangkai Bian[a,1], Yichao Han[b,1], Anwei Hou[a], Yujie Yuan[a], Xinhua Liu[a], Zixin Deng[a,c], Tiangang Liu[a,c,*]

[a] Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education and Wuhan University School of Pharmaceutical Sciences, Wuhan 430071, PR China
[b] J1 Biotech Co., Ltd., Wuhan 430075, PR China
[c] Hubei Engineering Laboratory for Synthetic Microbiology, Wuhan Institute of Biotechnology, Wuhan 430075, PR China

ARTICLE INFO

Keywords:
Terpenoids
Terpene synthases
Promiscuous
Protein engineering
Precursor supply
Combinatorial biosynthesis
Skeleton reframing

ABSTRACT

Terpenoids represent the largest family of natural products. Their structural diversity is largely due to variable skeletons generated by terpene synthases. However, terpene skeletons found in nature are much more than those generated from known terpene synthases. Most promiscuous terpene synthases (*i.e.* those that can generate more than one product) have not been comprehensively characterised. Here, we first demonstrated that the promiscuous terpene synthases can produce more variable terpenoids *in vivo* by converting precursor polyisoprenoid diphosphates of different lengths (C_{10}, C_{15}, C_{20}, C_{25}). To release the synthetic potential of these enzymes, we integrated the engineered MVA pathway, combinatorial biosynthesis, and point mutagenesis to depict the comprehensive product profiles. In total, eight new terpenoids were characterised by NMR and three new skeletons were revealed. This work highlights the key role of metabolic engineering for natural product discovery.

1. Introduction

Nature is an excellent synthetic chemist, utilising two simple processes, *i.e.* mutations and combinations of biosynthetic genes, to create diverse natural products (Cane et al., 1998). By combining sets of genes, domains, and modules identified from different biosynthetic pathways or created by scientists in various ways, the concept of combinatorial biosynthesis has accelerated the discovery and expanded the diversity of products (Kim et al., 2015; Tsoi and Khosla, 1995). This concept has been applied extensively to two large families of complex natural products, namely, polyketides and nonribosomal peptides (Khosla et al., 2014; Sieber and Marahiel, 2005; Weissman and Leadlay, 2005). Compared with these two families, very few reports have described the application of this strategy to terpenoids, the largest and most structurally diverse family of natural products. Recently, combinatorial biosynthesis has been proposed to generate diverse labdane-based diterpenoids (Andersen-Ranberg et al., 2016; Jia et al., 2016; Mafu et al., 2016). However, many other skeletons in the big terpenoid family remain to be explored.

The diversity of terpenoids arises from three steps (Fig. S1): (i) prenyltransferases (PTs) assemble the C_5 precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) into polyisoprenoid diphosphates with various lengths, such as C_{10} geranyl diphosphate (GPP), C_{15} farnesyl diphosphate (FPP), C_{20} geranylgeranyl diphosphate (GGPP), and C_{25} geranylfarnesyl diphosphate (GFPP) (Kellogg and Poulter, 1997); (ii) terpene synthases cyclise these polyisoprenoid diphosphates to generate terpene products with a single ring or intricate multiple rings (Christianson, 2006); (iii) tailoring enzymes, including oxygenases, methyltransferases, acetyltransferases, and glycosyltransferases, add functional groups at different positions, further enhancing structural diversity. The possibility of obtaining linear products from the first step is very limited for the fixed length of C_5 precursors, and the flexibility of the final products from the third step strongly depends on the skeletons from the second step. Thus, variability in the skeleton is mainly introduced by the cyclisation step. Biosynthesis by terpene synthases in this step has inherent advantages over chemical synthesis (Maimone and Baran, 2007); however, the biosynthetic repertoire identified to date only covers a small proportion of the terpene skeletons discovered, which only represents a small proportion of potential terpene skeletons according to theoretical

* Corresponding author.
E-mail address: liutg@whu.edu.cn (T. Liu).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ymben.2017.04.006
Received 3 February 2017; Received in revised form 27 March 2017; Accepted 19 April 2017
Available online 23 April 2017

calculations (Tian et al., 2016). Additionally, most traditional methods to investigate terpene synthases only focus on their major functions under native conditions, ignoring many potential applications.

Combinatorial biosynthesis of terpenoid skeletons will be an effective solution when two prerequisites are satisfied: identification of terpene synthases that tolerate different lengths of isoprene-diphosphate as substrates and the generation of suitable polyisoprenoid diphosphate substrates for these enzymes. In this study, we successfully identified promiscuous terpene synthases in nature and developed a terpene-overproducing chassis to satisfy these two criteria. Six *Escherichia coli* variants, obtained by combining two terpene synthases and three PTs, generated 50 terpenoids. We also demonstrated that terpene synthases with substrate promiscuity are easily available in nature, and the protein engineering of theses enzymes can be further incorporated in our strategy.

2. Materials and methods

2.1. Materials

Linalool (PubChem CID: 6549; Sigma-Aldrich, St. Louis, MO, USA), α-terpineol (PubChem CID: 17100; Sigma-Aldrich), pyrophosphate reagent (P7275; Sigma-Aldrich), and α-farnesene (PubChem CID: 5281516; J & K Scientific, Beijing, China) were used in this study. Kinetic assays were recorded using a Multimode Plate Reader Enspire (PerkinElmer, Waltham, MA, USA). Proteins were purified using a Biologic DuoFlow Chromatography System (Bio-Rad, Hercules, CA, USA). Terpenoids were separated using a Dionex Ultimate 3000 UHPLC system (Thermo Scientific, Waltham, MA, USA). Optical rotations were measured on a PerkinElmer Model 341 polarimeter (PerkinElmer). ^{1}H NMR and ^{13}C NMR were performed on an Agilent 400 MHz or 600 MHz instrument (DirectDrive2; Santa Clara, CA, USA). Single-crystal data were measured using a Bruker Kappa APEX-Duo CCD Diffractometer (Bruker, Billerica, MA, USA). Electronic circular dichroism (ECD) was used in conjunction with time-dependent density functional theory (TDDFT) calculations.

2.2. Identification of class I terpene synthases

Class I terpene synthases were identified by screening all predicted proteins against a set of hidden Markov models constructed for each terpene synthase family. The model trained from the non-plant terpene cyclase alignment profile (cd00687 from the Conserved Domain Database) showed the best prediction coverage in our preliminary test data. The final identified proteins were manually curated on the basis of protein alignment and transcript assembly.

2.3. Phylogenetic analysis

A multiple sequence alignment was generated using MAFFT (Katoh et al., 2002) with the local pair iterative refinement method. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). The phylogenetic tree was inferred using the maximum likelihood method based on the Jones-Taylor-Thornton model. All positions containing gaps or missing data were eliminated. The final dataset included 211 positions.

2.4. Motif analysis

The active-site cavity was mapped according to existing crystal structures of fungal terpene synthases (PaFS, AtARS, and FsTDS) and refined alignment of all characterised fungal terpene synthases. The binding pocket of terpene synthases consisted of six α-helixes, *i.e.* C, D, F, G, H, and J. Strictly conserved and catalysis-related sites were excluded. Sequence logos of unconserved residues in the active-site cavity were generated using the WebLogo server (http://weblogo.berkeley.edu/).

2.5. Molecular modelling

The three-dimensional model of the TC domain of FgMS was built using the SWISS-MODEL web server (https://www.swissmodel.expasy.org/). The crystal structure of the TC domain of PaFS from *Phomopsis amygdali* was selected as the template. The resulting homology structures were visualised using UCSF Chimera.

2.6. In vitro assays and kinetic measurements

To test the *in vitro* activities of FgMS, its mutants, and FgGS, reactions were conducted using 10 μM purified proteins, 100 μM substrates (GPP, FPP, GGPP, or GFPP), and 2 mM Mg^{2+} in 200 μL of 50 mM PB buffer (pH 7.6) with 10% glycerol at 30 °C overnight. The sample was extracted with hexane (200 μL) and then analysed by GC-MS. For steady-state kinetics, 100-μL scale reactions were carried out in 50 mM Tris-HCl buffer (pH 7.6) with 10% glycerol and 50 μL of pyrophosphate reagent. The concentration of the enzyme (FgMS or FgGS) was kept constant at 1 mg/mL with 2 mM Mg^{2+} as the only metal ion, while substrates (GPP, FPP, GGPP, and GFPP) were added at difference concentrations ranging from 1 to 200 mM. Product assays were carried out by measuring the release of pyrophosphate (PPi), which were recorded using an Enspire Multimode Plate Reader following a previously described method (Agger et al., 2008).

2.7. Fermentation and purification of terpenoids

E. coli T7 (harbouring pMH1, pFZ81, and pGB310), T8 (harbouring pMH1, pFZ81, and pGB311), T9 (harbouring pMH1, pFZ81, and pGB312), T10 (harbouring pMH1, pFZ81, and pGB313), and T11 (harbouring pMH1, pFZ81, and pGB314) were cultivated in 2-L flasks containing 1 L LB medium at 37 °C with 100 mg/L ampicillin (AMP), 50 mg/L KAN, and 34 mg/L chloramphenicol (CM). When the OD_{600} reached 0.6–0.8, 0.1 mM IPTG was added to the cultures, protein expression was induced at 16 °C for 18 h, and the strains were then cultivated at 28 °C for an additional 3 days. The cell cultures were extracted twice with an equal volume of hexane. The organic layer was concentrated using a rotary evaporator, and terpenoids were purified using a Dionex Ultimate 3000 UHPLC system.

2.8. Detection of terpenoids

Terpenoids were extracted by hexane and detected by GC-MS (Thermo TRACE GC ULTRA combined with a TSQ QUANTUM XLS MS). The samples were injected into a TRACE TR-5MS (30 m × 0.25 mm × 0.25 μm). The oven temperature was set at 80 °C for 1 min, increased to 220 °C at a rate of 10 °C/min, and held at 220 °C for 15 min. The injector and transfer lines were maintained at 230 °C and 240 °C, respectively. Product structures were determined by comparison with authentic standards, comparison with mass spectra data from the NIST library, and NMR analysis.

3. Results

3.1. Terpene synthases with substrate promiscuity

To predict promiscuous terpene synthases from sequences, we collected sequences of terpene synthases in fungus genomes, a major source of terpenoids that has not yet been fully explored, and constructed a phylogenetic tree. Fifty-one genetically or biochemically characterised terpene synthases (Additional Data Table S1), including mono-, sesqui-, di-, and sesterterpene synthases, were mainly divided into five clades (Fig. 1a). Clade III was particularly interesting. Notably, as of the writing of this manuscript (August 2016), all characterised di-