Phytochemistry 103 (2014) 13–21

Contents lists available at ScienceDirect

Phytochemistry

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

Biochemical characterization of the castor bean *ent*-kaurene synthase(-like) family supports quantum chemical view of diterpene cyclization

Alana J. Jackson, David M. Hershey, Taylor Chesnut, Meimei Xu, Reuben J. Peters *

Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 19 February 2014
Received in revised form 4 April 2014
Available online 5 May 2014

Keywords:
Natural products biosynthesis
Diterpenoids
Terpene synthases

ABSTRACT

It has become apparent that plants have extensively diversified their arsenal of labdane-related diterpenoids (LRDs), in part via gene duplication and neo-functionalization of the ancestral *ent*-kaurene synthase (KS) required for gibberellin metabolism. For example, castor bean (*Ricinus communis*) was previously shown to produce an interesting set of biosynthetically related diterpenes, specifically *ent*-sandracopimaradiene, *ent*-beyerene, and *ent*-trachylobane, in addition to *ent*-kaurene, using four separate diterpene synthases, albeit these remain unidentified. Notably, despite mechanistic similarity of the underlying reaction to that catalyzed by KSs, *ent*-beyerene and *ent*-trachylobane synthases have not yet been identified. Given our interest in LRD biosynthesis, and the recent availability of the castor bean genome sequence, a synthetic biology approach was applied to biochemically characterize the four KS(-like) enzymes [KS(L)s] found in *Ricinus communis* [i.e., the RcKS(L)s]. In particular, using bacteria engineered to produce the relevant *ent*-copalyl diphosphate precursor and synthetic genes based on the predicted RcKS(L)s, although this ultimately required correction of a "splicing" error in one of the predicted genes, highlighting the dependence of such a synthetic biology approach on accurate gene sequences. Nevertheless, it is possible to assign each of the four RcKS(L)s to one of the previously observed diterpene synthase activities, providing access to functionally enzymes. Intriguingly, the product distribution of the RcKS(L)s seems to support the distinct diterpene synthase reaction mechanism proposed by quantum chemical calculations, rather than the classically proposed pathway.

1. Introduction

Plants are particularly prolific producers of terpenoids, of which there are more than 55,000 known (Köksal et al., 2011). Prominent among these are the labdane-related diterpenoids (LRDs), with ~7,000 known, and whose biosynthesis in plants can be traced back to the requisite production of gibberellin phytohormones (Peters, 2010). These natural products are characterized by their derivation from a sequential pair of terpene synthase (TPS) catalyzed reactions. First, (bi)cyclization of the general diterpenoid precursor (*E*,*E*,E)-geranylgeranyl diphosphate (GGPP), generally to the eponymous labdadienyl/copalyl diphosphate (CPP) intermediate, mediated by a class II diterpene cyclase. This is followed by the action of a more typical class I (di)terpene synthase. In the case of gibberellins, this pair of reactions yields *ent*-kaurene (**1**) via *ent*-CPP, as catalyzed by a CPP synthase (CPS) and subsequently acting *ent*-kaurene synthase (KS) (Fig. 1). Given their homology to the ancestral KS, those class I diterpene synthases acting on CPP (**5**) are often termed KS-like (KSL), and these fall into what has been designated the TPS-e sub-family (Chen et al., 2011). This sub-family is further distinguished by the almost universal presence of an additional/insertional γ-domain relative to other plant class I TPS.

It is now evident that monocots, particularly cereal crop plants, have significantly expanded their arsenal of LRD natural products (Schmelz et al., 2014). In part, this wide range of LRDs evolved via gene duplication of the ancestral KS and neo-functionalization of the resulting KSLs, which produce the multicyclic hydrocarbon backbone structures that characterize the resulting various families of LRDs (Peters, 2010). The evolutionary radiation of KSLs

Abbreviations: aa, amino acid; KS, *ent*-kaurene synthase; KSL, KS-like; RcKS(L), castor bean (*Ricinus communis*) KS(L) Euphorbiaceae; AtKS, *Arabidopsis thaliana* KS Cruciferae; CPP, copalyl diphosphate; CPS, CPP synthase; GC–MS, gas chromatography with mass spectrometric detection; GGPP, (*E*,*E*,*E*)-geranygeranyl diphosphate; nt, nucleotide; TPS, terpene synthase.

* Corresponding author. Tel.: +1 515 294 8580; fax: +1 515 294 0453.
E-mail address: rjpeters@iastate.edu (R.J. Peters).

http://dx.doi.org/10.1016/j.phytochem.2014.04.005
0031-9422/

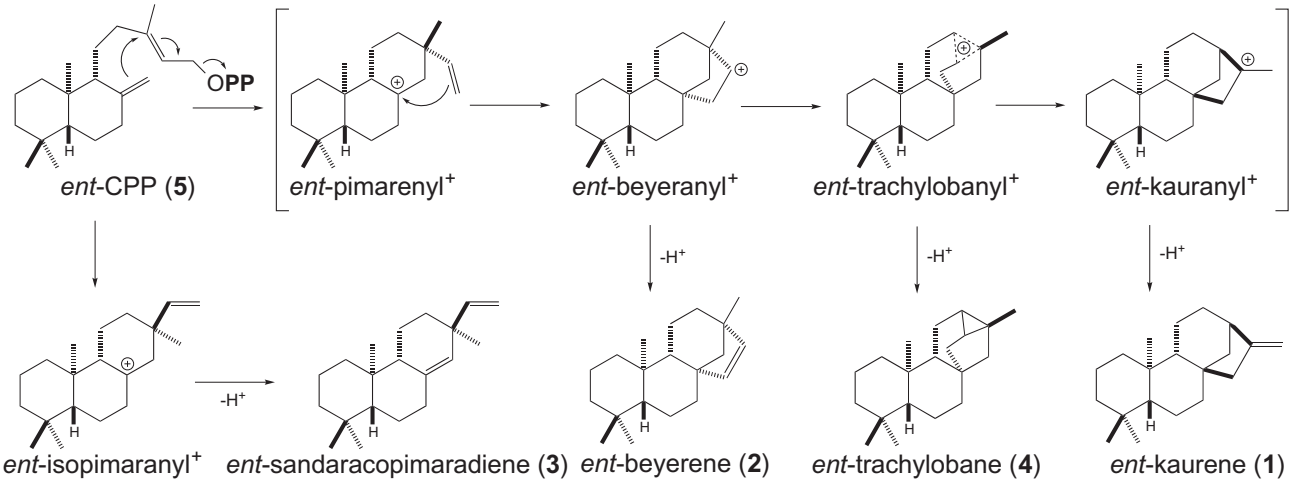

Fig. 1. Cyclization mechanisms for the diterpene synthase activities previously identified in castor bean by assays with cell-free extracts (Robinson and West, 1970b; Spickett et al., 1994), and their relationship to the classical mechanism for production of *ent*-kaurene (**1**). This classic mechanism proceeds via ionization of the allylic diphosphate ester bond in *ent*-CPP (**5**) to trigger initial cyclization to the depicted *ent*-pimarenyl$^+$ intermediate, followed by secondary cyclization to the depicted *ent*-beyeranyl$^+$ intermediate that rearranges via the depicted *ent*-trachylobanyl$^+$ intermediate en route to the *ent*-kauranyl$^+$ intermediate that is quenched by deprotonation to yield *ent*-kaurene. As shown, the production of *ent*-beyerene (**4**) and *ent*-trachylobane (**2**) similarly arises from deprotonation of the corresponding carbocations.

in monocots was first discovered in rice (Peters, 2006), but more recent work has demonstrated that this applies at least to other cereal crop plants such as wheat (Zhou et al., 2012), as well as maize (Schmelz et al., 2014). In particular, the identified monocot KSLs cluster with the monocot KSs, while the dicot KSs form a separate group. Nevertheless, such derivation of KSLs from KS appears to have occurred early in at least the *Poaceae* (grass) plant family, as suggested by clustering of KSLs from all three investigated species, which diverged early in evolution of the grasses (Schmelz et al., 2014), separate from the monocot KSs. The continuing nature of this gene family radiation is further indicated by the presence of KSLs in the KS containing cluster, as well as the number of KSLs in each species ($\geqslant$4). Indeed, comparison of functionally distinct alleles of a rice KSL led to identification of single residue "switch" for product outcome that applies in KSs as well (Xu et al., 2007).

By contrast, little is known about the production of LRDs in dicots (Zi et al., 2014), although a number are known to produce more specialized LRDs (i.e., other than gibberellins). While various species from the *Lamiaceae* plant family produce LRDs, the relevant KSLs identified to date generally do not contain the otherwise typical γ-domain (Caniard et al., 2012; Gao et al., 2009; Schalk et al., 2012), highlighting their unusual evolutionary origin (Hillwig et al., 2011), and complicating comparison of these to KSs – e.g., for analysis of catalytic mechanism (Zi et al., 2014). Although full-length KSLs have been identified from dicots (Sallaud et al., 2012; Zerbe et al., 2013), these are only distantly related to each other and have not yet led to any mechanistic insights.

Castor bean has been reported to produce an interesting set of four biosynthetically related diterpenes derived from *ent*-CPP. Specifically, *ent*-beyerene (**2**), *ent*-sandaraco-pimaradiene (**3**) and *ent*-trachylobane (**4**), as well *ent*-kaurene (**1**) (Robinson and West, 1970a). Critically, this has been shown to result from the activity of four distinct enzymes rather than being produced by a smaller number of promiscuous cyclases (Robinson and West, 1970b; Spickett et al., 1994). Notably, no KSL specifically producing either *ent*-beyerene (**2**) or *ent*-trachylobane (**4**), which can be envisioned as arising from deprotonation of plausible intermediates en route to *ent*-kaurene (**1**) (Fig. 1), has yet been reported. With the recent report of the castor bean genome sequence (Chan et al., 2010), it seemed possible to functionally identify the relevant RcKSLs via a synthetic biology approach. Here the potentially relevant enzymes from the predicted transcriptome were expressed, using synthetic gene constructs, in *Escherichia coli* engineered to produce the necessary *ent*-CPP precursor (**5**) via a previously described modular metabolic engineering system (Cyr et al., 2007). These studies not only led to functional identification of novel enzymes, but also provided insights into diterpenoid evolution and the catalytic mechanism of diterpene synthases.

2. Results

2.1. Initial identification of KS(L)s from castor bean

To begin investigating the interesting set of castor bean diterpene synthases, BLAST searches of the available *Ricinus communis* sequence information were carried out using the KS from *Arabidopsis thaliana* (AtKS) as a probe for full-length (i.e., γ-domain containing) KS(L)s, as well as the miltiradiene synthase from *Salvia miltiorrhiza* (SmMS) as a probe for shorter (i.e., non γ-domain containing) KSLs. From this bioinformatic search four full-length (but no shorter) KS(L)s were found among the predicted genes from the reported genome sequence (Chan et al., 2010), termed here RcKS(L)1-4 in the order in which they were listed in the BLAST results (i.e., similarity to AtKS). Initial molecular phylogenetic analysis indicated that the top hit was significantly more closely related to dicot KSs and was termed RcKS(L)1, while the other three clustered separately and were termed RcKSL2-4. *RcKSL2-4* were found in close proximity to each other, within a region of 65 kb, with *RcKSL2* and *4* occurring as a tandem gene pair. Rather than attempting to clone full-length cDNA for each of these predicted genes, synthetic open reading frames, codon optimized for expression in *E. coli*, were obtained.

2.2. Biochemical analysis of the RcKS(L)s

A modular metabolic engineering system was previously developed that enables facile production of terpenoids in *E. coli* (Cyr et al., 2007). Of particular interest here, this system enables co-expression of a GGPP synthase (GGPS) with potential diterpene synthases, including both a CPS and KS(L). Accordingly, the synthetic *RcKS(L)* were truncated to remove the N-terminal plastid-directing transit peptide sequences, individually sub-cloned into compatible expression vectors and each co-expressed with either just the GGPS, or the GGPS along with a CPS. This enabled analysis